



GIS-IDEAS 2024



Organized by

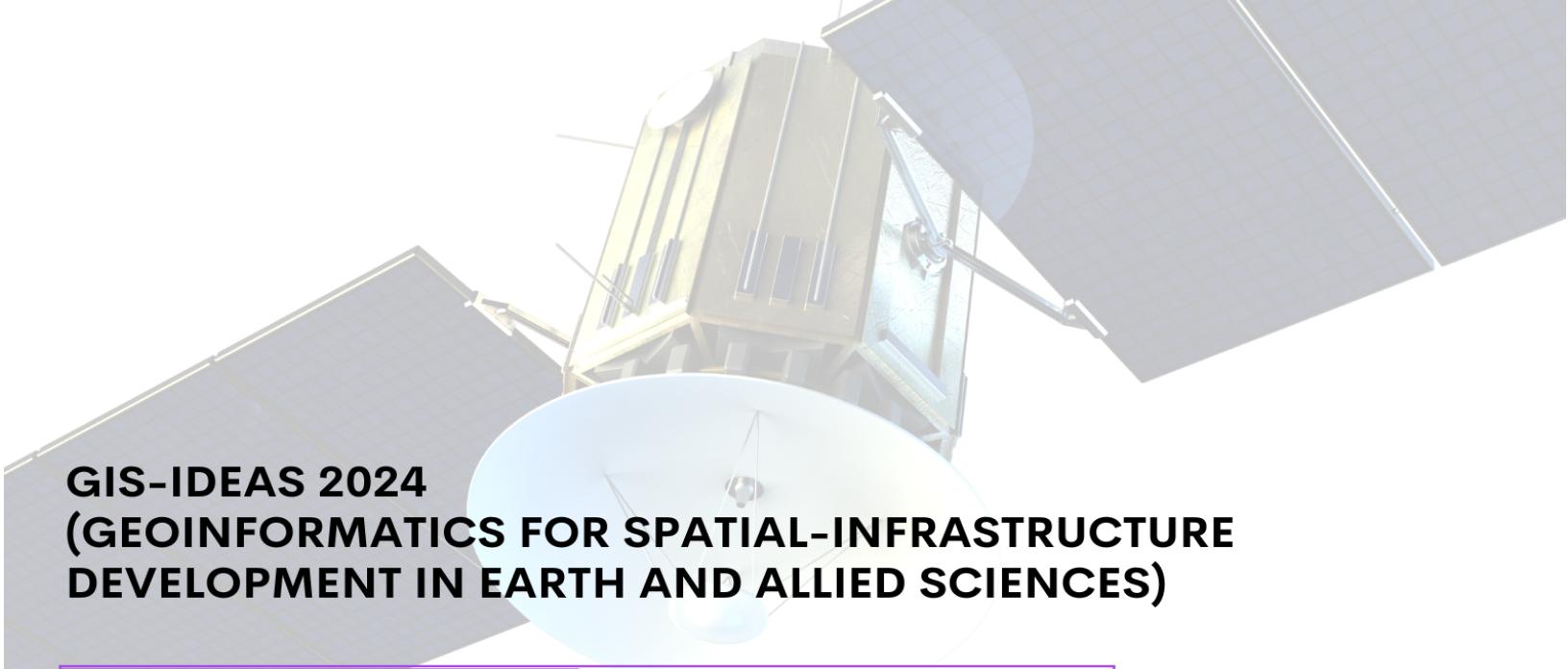
School of ICT, University of Phayao (UP)
Naresuan University (NU)

Osaka Metropolitan University (OMU)
Association of GeoInformatics Laboratories for Earthsciences (AGILE)



11 – 13 DECEMBER 2024





GIS-IDEAS 2024

(GEOINFORMATICS FOR SPATIAL-INFRASTRUCTURE DEVELOPMENT IN EARTH AND ALLIED SCIENCES)

11-13 December, 2024

**Grand Vista Hotel
Chiang Rai, THAILAND**



**The International Conference on GeoInformatics for Spatial-Infrastructure
Development in Earth & Allied Sciences (GIS-IDEAS) 2024**

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved. Copyright ©2024 by IEEE.

CATALOG NUMBERS

Media Type Requested	Part Number	ISBN	Print ISSN	Online ISSN	Media Qty Requested
XPORE COMPLIANT	CFP24VO2-ART	979-8-3315-0424-3			1

Welcome Message from General Chair of GIS-IDEAS 2024

Dear Distinguished Guests and Colleagues,

Welcome to the International Conference on GeoInformatics for Spatial-Infrastructure Development in Earth & Allied Sciences (GIS-IDEAS) 2024! It is with great enthusiasm and honor that we gather here in the beautiful city of Chiangrai province, a place renowned for its rich history and vibrant culture.

This year marks a significant milestone as we celebrate the 22th anniversary of the GIS-IDEAS2024 since the year of 2002. Over the decades, our community has grown and evolved, driven by a shared passion for innovation and excellence. Today, we stand on the shoulders of giants, ready to explore new horizons and push the boundaries of what is possible.

Our conference theme, "Geoinformatics for Sustainable Development Goals" reflects our commitment to addressing the global challenges of our time. From cutting-edge research in HealthGIS, GIS & Remote Sensing for environment majors, Web-GIS, Geospatial Artificial Intelligence and Machine Learning, AI and IoTs to breakthroughs in biomedical engineering and smart cities, GIS-IDEAS 2024 promises to be a melting pot of ideas and collaborations that will shape the future.

We have an exciting program lined up, featuring keynote addresses from world-renowned experts, technical sessions showcasing the latest research, and interactive workshops designed to foster learning and networking. Whether you are a seasoned researcher or a newcomer to the field, there is something here for everyone.

I would like to extend my heartfelt thanks to our sponsors, volunteers, and organizing committee for their tireless efforts in making this event possible. Your dedication and hard work are truly appreciated.

As we embark on this journey together, let us embrace the spirit of curiosity and collaboration. May this conference inspire you, challenge you, and provide you with new perspectives and opportunities.

Welcome to GIS-IDEAS 2024! Let's make this a memorable and impactful event.

Warm regards,

General Chair, GIS-IDEAS 2024

Welcome message from the Technical Program Chair GIS-IDEAS 2024

Welcome to the GIS-IDEAS 2024 Conference. This conference brings together to drive and facilitate collaboration in Geoinformatics and Information and Communication Technology (ICT) to encourage the development of Geoinformatics applications in natural and social environments.

The technical program comprises eight sessions that highlight the breadth and depth of geospatial technology today. All sessions offer significant advantages, benefits, and value by showcasing the diverse applications of geoinformatics across various fields. I hope that all our works will make a meaningful impact on global challenges and offer a comprehensive understanding of geoinformatics and ICT potential, equipping attendees with the skills and knowledge to drive innovation and positive change.

I want to extend my heartfelt thanks to each of you. To our technical paper authors, your innovative contributions form the backbone of our event. To keynote speakers, presenters, and all participants, your enthusiasm brings energy to our conversations. And to the lovely staff, committee and volunteers, your hard work makes everything possible.

Let's make the most of our time together—learning, sharing, and building lasting connections. I'm excited for the inspiring conversations ahead and hope you will enjoy the conference.

Warm regards,

Sawarin Lerk-u-suke

Technical Program Chair

Committees

CONFERENCE FOUNDERS

- Late Prof. Takashi FUJITA (JP)
- Dr. Nghiem Vu KHAI (VN)

GENERAL CHAIRS

- Phaisarn JEEFOO (TH)
- Venkatesh RAGHAVAN (JP)

GENERAL CO-CHAIRS

- Pornthep ROJANAVASU (TH)
- Krittika KANTAWONG (IEEE SPS Thailand Chapter, IEEE#94103612)
- Sittichai CHOOSUMRONG (TH)
- Chaiiwat VANSAROCHANA (TH)

FINANCIAL CHAIR

- Jirabhorn CHAIWONGSAI (TH)

TECHNICAL PROGRAM CHAIR

- Sawarin LERK-U-SUKE (TH)

TECHNICAL PROGRAM CO-CHAIRS

- Punnarumol Temdee (IEEE SPS Thailand Chapter, IEEE#94471090)
- Pradorn Sureephong (IEEE SPS Thailand Chapter, IEEE#98063511)
- Kampanart PIYATHAMRONGCHAI (TH)

PUBLICATION CHAIRS

- Sakkayaphop Pravetjit (TH)
- Paween Khoenkaw (IEEE SPS Thailand Chapter, IEEE#94052729)

ORGANIZING SECRETARIES

- Tatsuya NEMOTO (JP)
- Truong Xuan QUANG (TH)
- Tanayaluck CHANSOMBAT (TH)

WEB AND MEDIA COMMITTEE CHAIRS

- Natraj VADDADI (IN)
- Pongsakorn SIRIKAMNOI (TH)

SPECIAL SESSION CHAIR

- Saksit IMMAN (TH)
- Sitthisak PINMONGKHONKUL (TH)

STEERING COMMITTEE

- Ho Dinh DUAN (VN)
- Huynh Thi Lan HUONG (VN)
- Yasuyuki KONO (JP)
- Nguyen Kim LOI (VN)
- Truong Xuan LUAN (VN)
- Lam Dao NGUYEN (VN)
- Le Thi TRINH (VN)
- Nitin TRIPATHI (TH)
- Yasushi YAMAGUCHI (JP)
- Song XIANFENG (CN)
- Chalermchai PAWATTHANA (TH)
- Teerawong LAOSUWAN (TH)
- Kanchana NAKHAPAKORN (TH)
- Ornprapa P. ROBERT (TH)
- Arisara CHAROENPANYANET (TH)
- Kritchayan INTARAT (TH)
- Satith SANGPRADID (TH)
- Wirote LAONGMANEE (TH)
- Phurith MEEPROM (TH)
- Supaporn MANAJITPRASERT (TH)
- Pattana RACHAWONG (TH)

SCIENTIFIC COMMITTEE

- Tran Thi AN (VN)
- Tran Van ANH (VN)
- Sanit ARUNPLOD (TH)
- Nakarin CHAIKAEW (TH)
- Niti IAMCHUEN (TH)
- Atsushi KAJIYAMA (JP)
- Rangsan KET-ORD (TH)
- Jiraporn KULSOONTORNRAT (TH)
- Surenda Rama SITARAMAN (US)
- Susumu NONOGAKI (JP)
- Sarawut NINSAWAT (TH)
- Wipop PAENGWANGTHONG (TH)
- Vinayaraj POLIYAPRAM (JP)
- Bowonsak SRISUNGSISSUNTI (TH)
- Boonsiri SUKPROMSUN (TH)
- Pham Thi Mai THY (VN)
- Go YONEZAWA (JP)

Keynote Speakers



Dr. Pakorn Patchprayoon

Dr. Pakorn Patchprayoon is the Director of the Geo-Informatics Product Innovation Office at the Geo-Informatics and Space Technology Development Agency (GISTDA). His research focuses on understanding the physical processes of energy exchange between the land and water surfaces and the atmosphere by integrating satellite data with direct field measurements. Dr. Patchprayoon has dedicated 23 years to GISTDA, contributing in various research and leadership roles. He has authored and co-authored several publications and was a lecturer on GEOG 4093 Remote Sensing of the Environment at the University of Colorado, Boulder, USA. Dr. Patchprayoon holds a B.S. from Burapha University, M.S. from Mahidol University, and M.A. and Ph.D. from the University of Colorado-Boulder, USA.



Dr. Susumu Nonogaki

Susumu Nonogaki is a Geo-informaticist currently holding the position of Chief Senior Researcher at Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology (AIST). He obtained his PhD in Geosciences from Osaka City University in 2009. His work primarily focuses on three-dimensional modeling and analysis of shallow subsurface geological structure in urban areas in Japan. His expertise extends to GIS analysis, spatial interpolation, machine learning, geo-visualization, web mapping, and from 2005 to 2012 he lectured at training courses organized by Japan International Cooperation Agency (JICA) on GIS techniques for sustainable management of natural resources and agricultural product. Since 2010, he has been an active member of the Scientific Committee for GIS-IDEAS. His work continues to shape our understanding of the shallow subsurface geological conditions in urban areas and provides invaluable insights into urban development and planning, contributing to safer and more sustainable cities.



Dr. Tran Van Anh

Dr. Tran Van Anh is a lecturer at Hanoi University of Mining and Geology (HUMG). She obtained her Master's degree in Surveying and Mapping Engineering from HUMG, Vietnam, in 2001 and her PhD degree in GeoInformatics from Osaka City University (Japan) in 2007. Her field of study is remote sensing and GIS. She has working interests in Radar Interferometry (InSAR) for land deformation detections and optical images for air pollution (PM10) determination. Besides that, she also works on geospatial data research and builds predicting models. She has had more than 50 works published in prestigious domestic and international journals.



Dr. Natraj Vaddadi

Natraj Vaddadi is an Executive Member of the Governing Council, Centre for Education & Research in Geosciences. Natraj holds a Master's degree in Geology from the University of Pune and a Doctorate in Natural Resources and Environment from the Naresuan University, Thailand, specializing in Urban Groundwater Recharge through Rainwater Harvesting.

He is also an adjunct faculty in Petroleum Technology at the Nowrosjee Wadia College, University of Pune and has been teaching post-graduate courses in Drilling engineering and Production Operations at the University of Pune for over three decades. He has conducted several workshops in Open-Source GIS and GIS data collection tools in India, Thailand and Vietnam. He is also the author of an internationally acclaimed book "An Introduction to Oil Well Drilling".

As one of the founding members of the Centre for Education & Research in Geosciences, Vaddadi evangelises importance of Geoscience and Geospatial education at the grass roots and actively promotes the integration of Earth science knowledge in environmental conservation and sustainability efforts.



GIS-IDEAS



Tentative programme GIS-IDEAS 2024



December 10, 2024

Time (BKK Time)

- 14.00 – 17.00** Registration, Poster preparation
17.00 – 20.00 Committee meeting

December 11, 2024

Time (BKK Time)

- 08.00 – 16.00** Registration
08.30 – 09.30 Scientific program (IEEE track)
09.30 – 10.30 Workshop 1 session 1
Disaster Management System: Check Nam Application
10.15 – 10.30 Coffee break
10.30 – 11.30 Workshop 1 session 2
Technical Section: Air pollution monitoring System and
Demonstration: Application “Check Phoon”
11.30 – 12.00 Opening Ceremony
12.00 – 13.00 Lunch
13.00 – 13.30 Keynote 1 : Dr.Pakorn Petchprayoon
13.30 – 14.05 Keynote 2 : Dr.Susumu Nonogaki
14.05 – 14.15 Coffee break
14.15 – 15.40 Scientific program (IEEE track)
15.40 – 18.00 Scientific program (IEEE track)
18.00 – 21.00 Banquet

CO-ORGANIZERS



Osaka
Metropolitan
University





GIS-IDEAS



Tentative programme GIS-IDEAS 2024



December 12, 2024

Time (BKK Time)

08.00 - 12.00	Registration
09.00 - 10.30	Scientific program (GIS-IDEAS track)
10.30 - 11.00	Coffee break
11.00 - 11.30	Keynote 3 : Dr.Tran Van Anh
11.30 - 12.00	Keynote 4 : Dr.Natraj Vaddadi
12.00 - 13.00	Lunch
13.00 - 14.00	Workshop 2 : Digital Image Processing using DroneBox (Mapedia Co., Ltd.)
13.00 - 14.00	Poster Presentation
14.00 - 14.10	Coffee break
14.10 - 14.50	Closing ceremony, best paper announcement and next host presentation

December 13, 2024

Time (BKK Time)

08.00-17.00	1 Day trip at Chiang Rai Province
-------------	-----------------------------------

CO-ORGANIZERS



Osaka
Metropolitan
University



GIS-IDEAS: Schedule

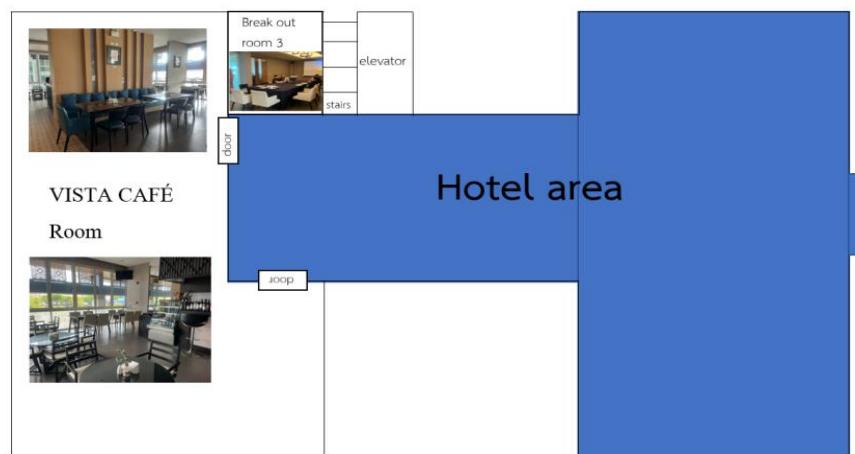
Date: December 11, 2024 Scientific program (IEEE track) Oral Presentaion				
Time	S1: Health GIS & Geoinformatics Room1: VISTA BALLROOM 1 (3rd floor) Session Chair: Assoc.Prof.Dr.Teerawong LAOSUWAN, Co-chair: Asst.Prof.Dr.Niti IAMCHUEN	S2: GIS for Spatial Analysis Room2: VISTA BALLROOM 2 (3rd floor) Session Chair: Assoc.Prof.Dr.Chaiwiwat VANSAROCHANA, Co-chair: Asst.Prof.Dr. Wipop PAENGWANGTHONG	S3: Sustainable Environment and GIS Room3: VISTA PRIVATE (2nd floor) Session Chair: Prof. Dr. Murray Alan POTTER, Co-chair: Dr.Sawarin LERK-U-SUKE	
8:30 – 9:30	Paper ID: 6506	Paper ID: 0134	Paper ID: 0420	
	Paper ID: 2223	Paper ID: 1354	Paper ID: 6313	
	Paper ID: 7115	Paper ID: 1558	Paper ID: 7521	
	Paper ID: 4118	Paper ID: 1664	Paper ID: 2380	
9:30 – 9:45			Paper ID: 9946	
14:15 – 15: 45	S4: GIS for Spatial Analysis Room1: VISTA BALLROOM 1 (3rd floor) Session Chair: Asst.Prof.Dr.Kampanart PIYATHAMRONGCHAI, Co-chair: Asst.Prof.Dr. Boonsiri SUKPROMSUN		S5: GIS & RS Room2: VISTA BALLROOM 2 (3rd floor) Session Chair: Prof.Dr.Venkatesh RAGHAVAN, Co-chair: Dr.Jiraporn KULSOONTORNRAT	
	Paper ID: 2115	Paper ID: 1148	Paper ID: 6317	
	Paper ID: 2599	Paper ID: 1366	Paper ID: 8706	
	Paper ID: 5067	Paper ID: 2691	Paper ID: 8935	
	Paper ID: 2608	Paper ID: 6289	Paper ID: 5907	
	Paper ID: 3425	Paper ID: 2898	Paper ID: 6240	
	Paper ID: 5470	Paper ID: 3110	Paper ID: 8194	
15:50 – 17: 05	S7: A Critical Network Perspective (SDGs) Room1: VISTA BALLROOM 1 (3rd floor) Session Chair: Asst.Prof.Dr. Surachai WONGCHAREE, Co-chair: Chutpong PACHANAPARN		S8: A Critical Network Perspective (SDGs) Room2: VISTA BALLROOM 2 (3rd floor) Session Chair: Assoc.Prof.Dr. Saksit IMMAN, Co-chair: Dr.Sawarin LERK-U-SUKE	
	Paper ID: 0536	Paper ID: 2112	Paper ID: 9141	
	Paper ID: 0587	Paper ID: 4793	Paper ID: 3619	
	Paper ID: 0667	Paper ID: 9512	Paper ID: 6990	
	Paper ID: 1358	Paper ID: 3602	Paper ID: 9425	
	Paper ID: 1687	Paper ID: 6544	Paper ID: 6244	
17:05 – 17: 20			Paper ID: 2384	

GIS-IDEAS 2004: VENUE



CONFERENCE LAYOUT

2nd floor



3rd floor

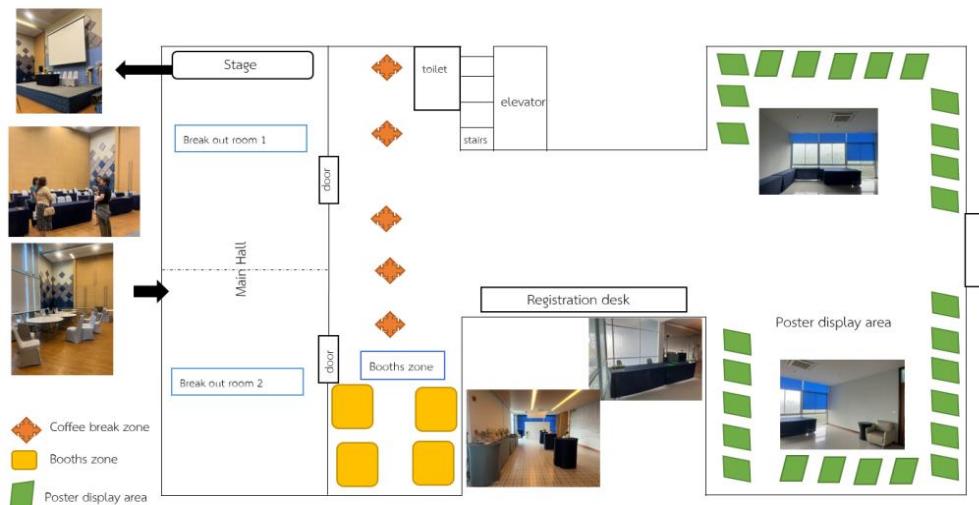


Table of Contents

Comparative Analysis of Evapotranspiration in Sugarcane Fields Using MODIS Satellite Data and Ground-Based Weather Stations	1
<i>Boonyasith Khobkhun, Pornthep Rojanavasu, Chanika Sukawattanavijit, Chompunut Chayawat, Krittika Kantawong and Jirabhorn Chaiwongsai</i>	
Spatial distribution of Fireflies (Coleoptera: Lampyridae) A Case Study of Kwan Phayao Area, Phayao Province, Thailand.....	6
<i>Panupong Chaiwongsae, Sitthisak Pinmongkhonkul, Arthit Nuntakwang, Boonchuang Boonsuk, Manas Titayavan, Warin Boonriam and Sorasak Nak-Eiam</i>	
Utilizing Unmanned Aerial Vehicles to assess carbon sequestration in trees conducted in response to plant genetic conservation project under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn at Chiang Rai Rajabhat University.....	11
<i>Krittawit Suk-Ueng, Sawarin Lerk-U-Suke, Pensri Malithong, Sutti Malithong and Pemmanee Senasutthiphap</i>	
Characterization of Biochar Produced from Sawdust and Polypropylene Plastic Waste Composite via Slow Pyrolysis	16
<i>Aitsara Chanthakhot, Poramed Aungthitipan, Pornmongkol Tansomros, Pattaranun Maskhunthod, Athicha Janthakot, Sukanya Hongthong, Torpong Kreetachat, Kowit Suwannahong, Surachai Wongcharee and Saksit Imman</i>	
Evaluating the Travel Carbon Footprint Associated with Using Digital Platforms for Accessing Social Welfare Services.....	22
<i>Wanawan Pragot, Kornganok Thongta, Chaturawit Chanchai, Sukhuma Chitapornpan, Somanat Somprasert, Supreeda Homklin, Chaiwat Photong, Pimsiri Suwanpat and Chattip Chaichakan</i>	
Satellite subsidence and sea-level data assimilation for mean high-water line assessment in the Bangkok area.....	27
<i>Sommart Niemnil, Femke Vossepoel, Marc Naeije, Milan Lazecky, Jenny Soonthornrangsar, Mark Bakker, Aimée Slangen, Wim J.F. Simons, Andy Hooper and Anuphaao Aobpaet</i>	
Geospatial Evaluation of Remote Working Hubs: A Graph Convolutional Network Model.	31
<i>Junke Xu, Alireza Dehghani and Yilin Li</i>	
Kinetics Evaluation of Total Solids Content Affects on Biohydrogen Production from Food Waste by Dark Fermentation	37
<i>Sorrrachat Intana, Sukhuma Chitapornpan, Somanat Somprasert and Satawat Tanarat</i>	
Climate Modeling in the Gulf of Thailand to Predict Beach Loss Due to Sea Level Rise...	43
<i>Meshra Singharath, Thanakrizt Peebkhunthod and Phaisarn Jeefoo</i>	
Enhancing Urban Heat Island Analysis through Indices-Based Prediction of Land Surface Temperature in Khon Kaen City, Thailand	49
<i>Kritchayan Intarat, Saruda Chuenkamol, Nithima Nuangjamnong, Areewan Hussadin, Chayapol Promao and Tuvachit Chalamkate</i>	
DEVELOPMENT OF WEB ATLAS FRAMEWORK OF CAT TIEN NATIONAL PARK	55
<i>Ngoc Truc Phuong Van</i>	

Efficiency of Liquid Hot Water Pretreatment of Water Hyacinth on enzymatic hydrolysis..	65
<i>Saksit Imman, Nopparat Suriyachai and Torpong Kreetachat</i>	
Enhancing Air Quality with an Innovative IoT-Integrated Hybrid Air Pollution Control System.....	69
<i>Wattana Loungon, Torpong Kreetachat, Saksit Imman, Nopparat Suriyachai, Surachai Wongcharee, Preut Thanarat and Kowit Suwannahong</i>	
A DSSAT Model Approach for Yield Optimization: A Case Study of Mae Na Rua, Phayao, Thailand	74
<i>Pongsak Jindasee, Prattana Deeprasertkul, Krittika Kantawong, Sawarin Lerk-U-Suke, Napa Rachata and Jirabhorn Chaiwongsai</i>	
Application of HEC-RAS and Satellite Imagery in Flood Risk Estimation for Rural Area in Thailand.....	80
<i>Nattaphon Rangsaritvorakarn and Niti Iamchuen</i>	
Evaluating Models for Urban LULC Classification by Integrating Earth Engine and Python Ecosystem	86
<i>Sitthisak Moukomla, Krudsakron Auynirundrunkool, Phurith Meeprom and Supaporn Manajitprasert</i>	
Assessing the Performance of Machine Learning Models using Satellite Dataset for Classification of Various Crop Types	90
<i>Montree Pinkaeo, Prattana Deeprasertkul, Phaisarn Jeefoo, Surachai Chantee, Pornthep Rojanavasu and Jirabhorn Chaiwongsai</i>	
Spatio-Temporal Analysis of Road Accident on Highway 2 between Nakhon Ratchasima Intersection and Joho Intersection	96
<i>Athiwat Phinyoyang, Tinn Thirakultomorn, Wilawan Prasomsup, Kankaew Kanya, Bantita Jaemklang and Buachomphu Phaensanthia</i>	
Geoinformatics-based Flood Extent Mapping and Analysis Using Sentinel-1 Imagery: A Case Study of Phayao Province	102
<i>Pannatorn Woottipriyatorn, Niti Iamchuen, Nakarin Chaikaew, Thidapath Anucharn, Thitisorn Sriprom and Phongsakorn Hongpradit</i>	
Characteristics of Infrasound Wave Background Noise on the Coast of Jakarta, Indonesia .	108
<i>Mario Batubara, Musthofa Lathif, Ibnu Fathrio, Asep Sulaeman, Heru Subarkah, Yopie Kristiana and Poki Agung Budiantoro</i>	
The Systematic Effectiveness Comparison of Spatiotemporal Carbon Dioxide Emission with Spatial Interpolation Methods: Case Study in Thailand	114
<i>Suriyawate Boonthalarath, Nakarin Chaikaew and Phaisarn Jeefoo</i>	
Research on developing an automated tool for evaluating spatial data quality according to the national geographic data standard technical model.....	125
<i>Yen Phan Quo, Nga Nguyen Thi Thu and Hanh Tong Thi</i>	
Integrating Sentinel-2 Imagery and Machine Learning for Crop Damage Evaluation in Thai Agriculture	139
<i>Viphada Boonlerd, Woranut Chansury, Panu Nuangjumnong, Pakorn Petchprayoon, Sansita Rattanasupa, Sirilux Noikeaing, Apantri Yutthapan and Pheeraphong Rattanaburi</i>	

GIS-based Hybrid Machine Learning for Landslide Susceptibility Assessment in Thai Nguyen Province, Vietnam	144
<i>Thuong Tran, Hoa Trieu and Nathaniel Bantayan</i>	
Test	150
<i>Krittika Test Submission</i>	
Identification of Spatial Patterns of Dengue Hemorrhagic Fever Using Spatial statistics ...	152
<i>Danh-Tuyen Vu, Anh-Huy Hoang and Tien-Thanh Nguyen</i>	
Acid Precipitation of Lignin from Alkali Black Liquor: A Study on the Optimization of Lignin Recovery Using H ₂ SO ₄ and HCl	159
<i>Khatiya Weerasai, Saksit Imman, Torpong Kreetachat and Nopparat Suriyachai</i>	
Satellite-based Nitrogen Uptake Evaluation in Rice Cultivation	162
<i>Porutai Thianthai, Phaisarn Jeefoo, Nuntikorn Kitratporn, Sukij Skawsang, Panu Nuangjamnong and Sawarin Lerk-U-Suke</i>	
Accuracy Assessment of UAV-Derived Data from Multi-Flight Plans with RTK GNSS Surveys	168
<i>Tinn Thirakultomorn, Athiwat Phinyoyang, Wilawan Prasomsup, Sumitta Khosungnoen and Kodchakon Phunak</i>	
Service area of Elderly care facility in the Perimeter Region of Thailand, using GIS	174
<i>Kulapramote Prathomchai, Chitsanu Anurakjaturong, Nitis Boonanun, Kitthiya Pooriwitayawattana, Supattra Tookhokkrud, Sopita Ountuang and Wilawan Khrutsawan</i>	
Enhancing Thailand's Cultural Diplomacy: A GIS-Based Analysis of Global Sentiments Toward Thai Food and Tourism	179
<i>Kednipa Prasongsri and Teerayut Horanont</i>	
Development of IoT-Based Compact Mushroom Cultivation Monitoring System	185
<i>Noor Ropidah Bujal, Shaizan Jusoh, Mokhtar Hashim, Syazana Md Dali, Nor Sukor Ali and Mohamed Noor Azman Bidin</i>	
Development of a Time Series Visualization Application for PM2.5 Level Changes Using Google Earth Engine.....	191
<i>Natima Udon, Kampanart Piyathamrongchai and Sittichai Choosumrong</i>	
Sex determination in leave of date palm (<i>Phoenix doactylifera</i>) on the juvenile vegetative phase between tissue culture and seeding by using LC- QTOF-MS technique ...	196
<i>Pranorm Khruueawan, Siriwat Boonchaisri, Achara Kleawkla, Chattakan Jantaput, Nakarin Chaikaew, Winai Wiriyaaalongkorn, Watcharin Jantawan and Ekawit Threenet</i>	
Land Use Changes Prediction in the Phra Phimon Irrigation Scheme using Multi-temporal Landsat Imagery and CA-ANN Model.....	201
<i>Kasidet Sirisakorn, Ketvara Sittichok and Chuphan Chompuchan</i>	
Estimating Monthly PM2.5 Levels using Integrated Satellite and Meteorological Data: A Case Study of Suphan Buri Province, Thailand	206
<i>Marut Phuphaniat, Phaisarn Jeefoo and Wipop Paengwangthong</i>	

Visualizing Awareness and Behavior in Reducing Single-Use Plastic among High School Students Using the Interactive ArcGIS Storymaps.....	210
<i>Thi Viet Huong Do, Phuong Nghi Nguyen and Thi Thanh Ninh Nguyen</i>	
Landslide Risk Mapping and development of landslide database and mobile app for Bhutan	220
<i>Thongley Thongley and Younten Tshering</i>	
Deep Residual Neural Networks with Self-Attention for Landslide Susceptibility Mapping in Uttaradit Province, Thailand	226
<i>Kritchayan Intarat, Nithima Nuangjamnong, Jojinda Sae-Jung, Woraman Jangsawang, Patimakorn Yoomee and Teerapong Panboonyuen</i>	
The Development of Satellite-based Active Fire Database and Hotspot Clustering to Enhance Forest Fire Management Practices in Thailand.....	232
<i>Woranut Chansury, Nuntikron Kitratporn, Anusorn Rungsipanich, Teansiri Moolchan, Chatchaya Buaniam, Itthirit Pintong and Nattanon Chaisak</i>	
Performance evaluation of Mini-UAV Photogrammetry for Rooftop Area Measurement ...	237
<i>Theerasak Ooppakarn and Sawarin Lerk-U-Suke</i>	
Optimizing Building Footprint Extraction from UAV Imagery using Pretrained Deep Learning Models: A Case Study in University of Phayao, Thailand	243
<i>Kannika Maomool, Samniang Suttara and Sawarin Lerk-U-Suke</i>	
An Implementation of the NOAA Satellite Data Reception Ground Station for Weather and Climate Enthusiasts	248
<i>Wichapol Utumporn, Suwabhat Somboonsub, Watchara Amasiri, Dahmmaet Bunnjaweh, Warinthon Kiadtkornthaweeeyot Evans and Pirapat Waritkraikul</i>	
Monitoring Land Use Changes and Comparing Watershed Quality Classes: A Case Study of Mueang Phetchabun District, Phetchabun Province	254
<i>Kittikun Nupat, Nakarin Chaikaew, Niti Iamchuen, Phongsakorn Hongpradit, Thitisorn Sriprom and Thidapath Anucharn</i>	
Perspective on mortality associated with particulate matter in upper northern Thailand over the 2014–2023 period.....	260
<i>Patsanun Lawongyer, Sittichai Pimonsree and Patipat Vongruang</i>	

Comparative Analysis of Evapotranspiration in Sugarcane Fields Using MODIS Satellite Data and Ground-Based Weather Stations

Boonyasith Khobkhun
*School of Information and Communication Technology,
 University of Phayao,
 Phayao, Thailand.
 67025280@up.ac.th*

Chompunut Chayawat
*Geo-Informatics Space Technology Development Agency (Public organization) : GISTDA
 Bangkok, Thailand.
 chompunut.cha@gistda.or.th*

Pornthep Rojanavasu
*School of Information and Communication Technology
 University of Phayao,
 Phayao, Thailand.
 pornthep.ro@up.ac.th*

Krittika Kantawong
*School of Information and Communication Technology
 University of Phayao,
 Phayao, Thailand.
 krittika.ka@up.ac.th*

Jirabhorn Chaiwongsai
*School of Information and Communication Technology
 University of Phayao,
 Phayao, Thailand.
 jirabhorn.ch@up.ac.th*

Chanika Sukawattanavijit
*Geo-Informatics Space Technology Development Agency (Public organization) : GISTDA
 Bangkok, Thailand.
 chanika@gistda.or.th*

Abstract—The objective of this study is to explore the feasibility of using satellite imagery to monitor evapotranspiration (ET). Spatial evapotranspiration estimates may differ from crop evapotranspiration, either being overestimated or underestimated. The approach of this assessment involves using satellite imagery products to monitor evapotranspiration and comparing the results with data from weather stations. MOD16A2GF is a satellite-based evapotranspiration product, was selected for comparison with weather station data. The study found that MOD16A2GF can be used to assess evapotranspiration effectively during Thailand's dry season, as the trend of evapotranspiration aligns with weather station data. However, during the rainy season, MOD16A2GF estimates were lower than those from the weather stations. Over the entire year, MOD16A2GF underestimated reference evapotranspiration by 3.90 mm/8 days in 2020 and 8.21 mm/8 days in 2021. Similarly, crop evapotranspiration was underestimated by 6.72 mm/8 days in 2020 and 11.17 mm/8 days in 2021. The correlation between ET from satellite data and ET from weather station shows that 87% of the variability in one dataset can be explained by the other, indicating a strong relationship. A prediction model using the Random Forest algorithm achieved the highest R^2 value of 0.76 and an MSE value of 132.47. Therefore, this study concludes that ET values obtained from MOD16A2GF products, using PET multiplied by the crop coefficient, can effectively monitor evapotranspiration in the study area.

Keywords—Evapotranspiration, ET, MODIS, weather station, spatial evapotranspiration, sugarcane

I. INTRODUCTION

In Thailand, agriculture is a key sector of the economy [1]. According to the 2023 agricultural census, agricultural land occupies approximately 44 percent of Thailand's total area [2]. As a result, agriculture represents the highest percentage of water consumption, amounting to 75 percent of the country's total water use, or approximately 113.960 million cubic meters per year [3]. For effective agricultural management, Thailand's farmland is divided into two zones: irrigated areas with water management and non-irrigation areas that rely on rainwater groundwater and runoff for farming. As climate change continues to progress, water allocation has become increasingly challenging. The adoption of technology can enhance the quality and efficiency of water

management, leading to increased agricultural productivity in the future.

Water demand data is an important factor in ensuring effective agricultural water management. Evapotranspiration (ET) represents the combination of soil evaporation (E) and water demand for plant transpiration (T) [4] both of which vary depending on the specific conditions of each area. Traditionally, direct measurements using lysimeters [5] and evaporation pans [6] have been used to determine evaporation rates by measuring the amount of water lost in millimeters per day. However, these methods are limited by data continuity, and scale. Currently, weather station data is used to monitoring meteorological conditions in agricultural area and store information based on user demand. This data is continuous and available in near real-time. Direct measurement is suitable for specific areas, but expanding the use of equipment to cover country is impractical due to the high cost of each station [7]. Satellite imagery is one of remote sensing technique offers an effective alternative for monitoring evapotranspiration. It is widely recognized for its efficiency in measuring evapotranspiration over large areas [8]. Although the limitations of direct measurement for evapotranspiration monitoring have been addressed with satellite imagery [9], but the resolution of these images can still affect the accuracy of evapotranspiration estimates [10]. Therefore, it is necessary to assess the accuracy of satellite data for estimating evapotranspiration before applying it to monitor evapotranspiration across various regions of Thailand.

The crop coefficient (K_c) is a parameter that represents the ratio between crop evapotranspiration (ET_c) and reference evapotranspiration (ET_0). The reference evapotranspiration is calculated using meteorological data from the area following the Penman-Monteith equation recommended by the Food and Agriculture Organization (FAO) [11]. Crop evapotranspiration is determined by multiplying the crop coefficient with reference evapotranspiration. The purpose of this study was to compare evapotranspiration from MOD16A2GF satellite imagery with GISTDA station 7-Kamphaeng Phet weather station to monitor evapotranspiration and to assess the feasibility of expanding evapotranspiration monitoring across the country using satellite imagery.

II. STUDY AREA

The study area is a sugarcane field located in Mahachai Sub-District, Sai-Ngam District, Kamphaeng Phet Province, Thailand, as shown in figure 1. Kamphaeng Phet is situated in the lower northern region of Thailand. The economy is mainly based on agriculture.

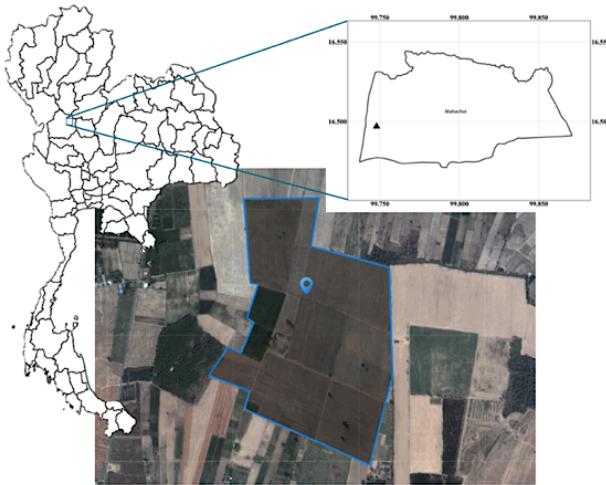


Fig. 1. Study area : location of GISTDA station7-Kamphaeng Phet.

Kamphaeng Phet is one of Thailand's key sugarcane-producing provinces with large tracts of flat and fertile land that make it ideal for large-scale cultivation. This province has a tropical savanna climate characterized by three main seasons: the summer (March to May), the rainy season (May to November) and winter (November to March). The combination of seasonal rains and warm temperatures creates optimal conditions for growing sugarcane. Sugarcane (*Saccharum officinarum* L.) is commonly cultivated under rainfed conditions with planting starting in the rainy season and harvesting during winter.

III. MATERAIL AND METHOD

A. MOD16A2GF global ET

MOD16A2GF is a dataset that measures evapotranspiration derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard the Terra satellite. This dataset has been storing in NASA's Distributed Active Archive Center (DAAC) and has been available from 2000 to the present. The current version of MOD16A2GF is 6.1. It is generated by the Penman-Monteith equation, incorporating Leaf Area Index (LAI), Fraction of Photosynthetically Active Radiation (FPAR), land cover data, and meteorological inputs from MOD15A2H [12]. The enhanced product offers an 8-day composite with a spatial resolution of 500 meters and is evaluated annually. MOD16A2GF is well-suited for assessing evapotranspiration throughout the year, which is supported by the annual production of global LAI and FPAR data from the MOD15A2H product, which are essential inputs for the MOD16A2GF product as they directly influence the accuracy of evapotranspiration estimates. By incorporating LAI and FPAR data from MOD15A2H, the MOD16A2GF dataset can more precisely account for variations in plant cover and photosynthetic activity, making it well-suited for year-round monitoring of evapotranspiration across various vegetation types and climates.

Thailand's agriculture is heavily reliant on seasonal rainfall. As a result, continuous monitoring is essential to assess water consumption and ensure effective water management [13]. MOD16A2GF plays a vital role in water resource management in Thailand [14]. It can provide consistent evapotranspiration data that is crucial for monitoring water usage in agriculture. It helps optimize irrigation practices, ensuring efficient water use and supporting strategic decision-making in water resource planning. Its large-scale monitoring capability is particularly beneficial in areas with limited ground-based measurements, making it a critical tool for promoting sustainable agriculture.

This study utilizes Google Earth Engine to generate time-series data on water consumption in the study area, which is then exported for comparison with direct measurements. These values are adjusted by the scale factor for Potential evapotranspiration (PET) from MOD16A2GF. Data from 2020 to 2021 includes 94 values used to compare plant water consumption during the study period.

B. Weather Station

GISTDA station7-Kamphaeng Phet is field measurements site located at latitude 16° 29' 44.5" N and longitude 99° 44' 50.3" E. It is equipped with 10 types of sensors for collecting meteorological data that influence crop growth including a Pyranometer, Net Radiometer, Spectroradiometer, Temperature and Humidity Probe, RM Young Wind Sensor, Rain Gauge, Soil Water Content Profile Probe, Standard Field of View Infrared Radiometer, 3D-Anemometer, Krypton Hygrometer, and network cameras show in figure 2. Measurements from these devices were recorded every 5 minute, 24 hours a day.

Complete data was collected from 2020 to 2021, during which sugarcane was grown year-round with one crop per year. There are 2 crops for analyze trends of evapotranspiration with 288 records per day. In total 210,528 records were gathered providing a sufficient dataset for studying evapotranspiration in the area. The data recorded 5 minute was first aggregated into daily values and then further into 8-day intervals to align with MOD16A2GF satellite data for comparison.



Fig. 2. GISTDA weather station and sensors

C. Reference Evapotranspiration : ET_o

According to the FAO, the Penman-Monteith equation [15] is used to calculate reference evapotranspiration (ET_o). This calculation incorporates factors including albedo, average temperature, wind speed, and saturated vapor pressure, as shown in:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

ET_o represents the reference evapotranspiration (mm/day), Δ is the slope of the steam pressure curve (kPa/ $^{\circ}$ C), R_n is the net radiation at the crop surface (MJ/m 2 /day), G is the soil heat flux density (MJ/m 2 /day), T is the average temperature ($^{\circ}$ C), u_2 is the wind speed (m/s), $e_s - e_a$ is the difference between saturated and actual vapor pressure (kPa) and γ is the psychrometric constant which accounts for the variation in atmospheric pressure with temperature and altitude. Both MOD16A2GF and weather station data are used to generate PET and ET_o data, respectively.

The difference in usage between PET and ET_o depends on the objectives of water planning and management in specific scenarios. PET or ET_p is suitable for large-scale study areas while ET_o is more appropriate for assessing water requirements in specific, localized areas [16]. Therefore, this concept was applied to investigate data usages by utilizing PET data from MOD16A2GF and ET_o data from weather stations to create comparative graphs of PET and ET_o . This approach was used to evaluate reference evapotranspiration based on both data sources.

D. Crop Evapotranspiration : ET_c

The evaluation of evapotranspiration can be conducted in various forms, such as using a Lysimeter or using a Class "A" pan. The calculation of evapotranspiration can be computed by equation, as shown in:

$$ET_c = ET_o \times Kc \quad (2)$$

ET_c represents crop evapotranspiration (mm/8day). ET_o is the reference evapotranspiration obtained from MOD16A2GF and weather station measurements, both computed using the Penman-Monteith equation. Kc is the crop coefficient that represents the plant's water consumption. The study area focuses on sugarcane cultivation which can be divided into four growth stages according to the Department of Agriculture, Thailand: the initial stage (0-30 days), tillering stage (31-170 days), grand growth stage (171-295 days), and harvest stage (296-330 days). The Kc values for each stage are 0.21, 0.73, 1.70, and 1.17, respectively [17] which are used to calculate the crop evapotranspiration. This calculation is based on PET from MOD16A2GF and ET_o from weather stations.

IV. RESULT AND DISCUSSION

In this study, evapotranspiration was assessed using satellite imagery and data from weather stations by comparing PET data from MOD16A2GF with ET_o data from weather monitoring stations, presented as 8-day evapotranspiration graphs.

A. Comparison between PET(MOD16A2.006) product and ET_o (GISTDA station7-Kamphaeng Phet)

The comparison shows that PET and ET_o tend to follow the same trend at the beginning of the year, before the rainy season (May-October). However, during the rainy season, rain began in May in 2020, whereas in 2021, it started earlier, in mid-April, resulting in lower PET and ET_o values that persisted throughout the rainy season. After October, the graph trends for both years align in the same direction as shown in figure 3. Reference evapotranspiration based on PET (MOD16A2GF) compared to ET_o (weather station), the PET values are lower than the average measurement 3.90 mm/8day in 2020 and 8.21 mm/8day in 2021.

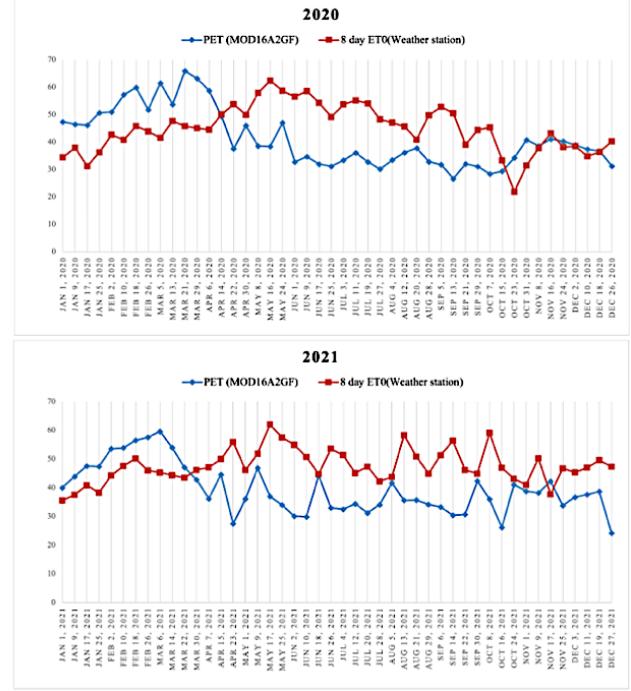


Fig. 3. Comparing of ET_o with PET(MOD16A2GF) and ET_o (weather station7-Kamphaeng Phet) between 2020-2021

B. Comparison between ET(MOD16A2.006) product and ET_o (GISTDA station7-Kamphaeng Phet)

The PET data was multiplied by the crop coefficients, and ET_o were used to create comparison graphs for each growth stage. The trends were similar overall. In the initial stage, evapotranspiration is comparable, while in the tillering stage, it differs. During the grand growth stage, evapotranspiration follows the same trend, and in the harvest stage, it becomes similar again, as shown in figure 4. Evapotranspiration based on PET compared to ET_o , the ET from PET values are lower than the average measurement 6.72 mm/8day in 2020 and 11.12 mm/8day in 2021.

C. Discussion

In the graphs used to study the correlation, the trends and relationships between the relevant variables become apparent.

The analysis of the correlation between "ET (PET x Kc)" and "ET (ET_o x Kc)" shows a value of 0.87. This indicates that 87% of the variability in one variable can be explained by the other, reflecting a strong relationship. The correlation coefficient is calculated using the following equation:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3)$$

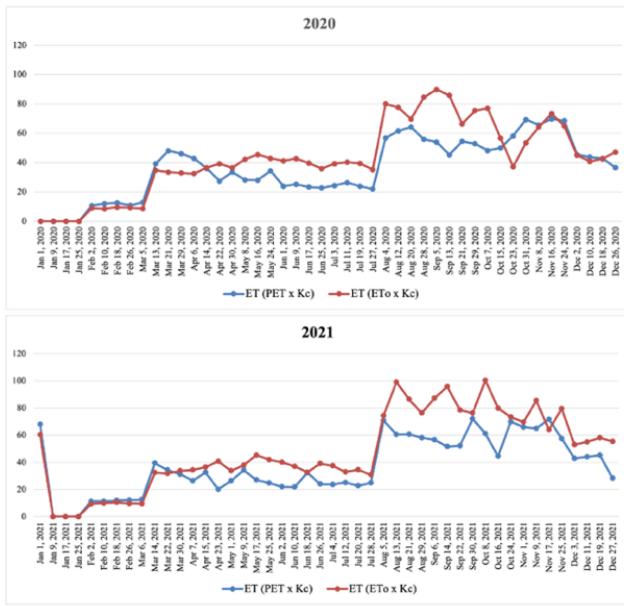


Fig. 4. Comparing of ET with ET(MOD16A2GF) and ET_c(weather station7-Kamphaeng Phet) between 2020-2021

The correlation coefficient (r) is calculated, where x and y represent the two variables and n is the number of data points.

The strong correlation in the data makes it suitable for ET prediction using machine learning techniques such as linear regression, random forests and decision trees. The dataset is divided into 80% for training and 20% for testing to evaluate and compare the performance of each model. These models will be evaluated and compared based on performance metrics including the coefficient of determination (R^2) and mean squared error (MSE) to measure the accuracy of each model. This approach assesses both linear and non-linear relationships by determining the compatibility of each model.

Firstly, Regression analysis to predict ET (PET x Kc) from ET (ET_o x Kc) showed that the model was fairly accurate as shown in figure 5 with an R^2 value 0.73. This indicates that the model could explain about 73% of the variance in ET (PET x Kc). The Mean Squared Error (MSE) was 147.63 suggesting the mean deviation between the predicted and actual values was within an acceptable range. These results suggest that the regression model effectively predicts ET (PET x Kc) from ET (ET_o x Kc), which is valuable for evaluating water evaporation in the study area and can be used to plan water management more efficiently.

Secondly, the Random Forest model was used to predict ET (PET x Kc) from ET (ET_o x Kc), the model was found to be highly efficient as shown in figure 6 with an R^2 value of 0.76. This indicates that the model can explain about 76% of the variance in ET (PET x Kc). The Mean Squared Error (MSE) was 132.47, suggesting that the mean deviation between the predicted and actual values is within an acceptable range.

Finally, the Decision Tree model was used to predict ET (PET x Kc) from ET (ET_o x Kc) as shown in figure 7, the model had an R^2 value of 0.64, indicating that it could explain about 64% of the variance in ET (PET x Kc). The Mean Squared Error (MSE) was 196.27 suggesting that the mean deviation between the predicted and actual values was within an acceptable range.

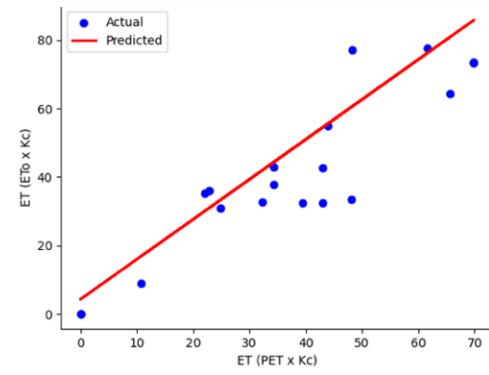


Fig. 5. Regression Analysis: Predicting ET (PETxKc) from ET (ET_oxKc)

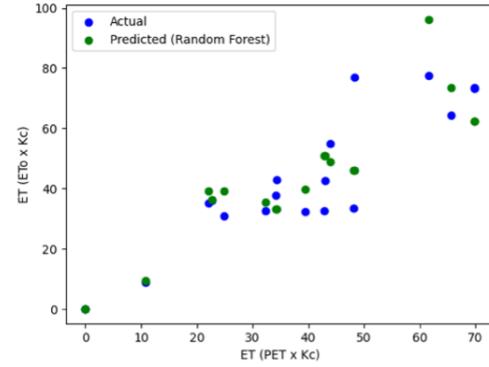


Fig. 6. Random Forest: Predicting ET (PETxKc) from ET (ET_oxKc)

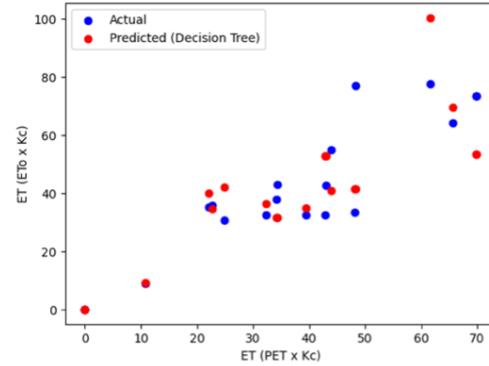


Fig. 7. Decision Tree: Predicting ET (PETxKc) from ET (ET_oxKc)

This approach evaluates both linear and non-linear relationships to determine the compatibility of each model. The analysis of the comparison results will reveal the effectiveness of each model and support more accurate predictions of plant water consumption. The comparison for the predictions from all models are shown in table 1.

TABLE I. THE COMPARISON RESULTS FOR THE PREDICTIONS FROM ALL MODELS

Model	R ²	MSE
regression analysis	0.73	147.63
random forest	0.76	132.47
decision tree	0.64	196.27

The results showed that the Random Forest model achieved the highest R^2 value of 0.76 and the lowest MSE value of 132.47 among the models tested, while the Decision Tree model had the lowest R^2 value of 0.64 and the highest MSE value of 196.27. Random Forest is robust to noise and overfitting makes it more suitable for handling the complex

and non-linear relationships in environmental data. By averaging the predictions of multiple trees, the Random Forest model minimizes errors and enhances reliability, consistently demonstrating better performance in capturing variable relationships compared to Decision Trees. This robustness makes Random Forest a more effective choice for predicting evapotranspiration in this study.

D. Limitations

The limitations of study revealed that the data resolution was not fully aligned with the context of the area, as well as the continuity and scale of measurements from the station. This inconsistency may lead to discrepancies in the comparison process. The results indicate that while satellite image data can be compared with station measurements, and both show consistent trends, the quantities may differ. These differences can be attributed to environmental factors, weather conditions, cloud cover, and discrepancies in the measuring equipment. As a result, the evapotranspiration measured by the station and the satellite imagery varied. For future studies, it is crucial to refine methods for estimating evapotranspiration from satellite images, enabling the use of continuous satellite data across Thailand to predict water consumption in different regions effectively.

V. CONCLUSION

The assessment of evapotranspiration from satellite imagery, when compared with ET and PET data from MOD16A2GF, indicates that the data is primarily applicable at the beginning of the year, as air humidity during this period is conducive to soil evaporation and plant transpiration. As humidity increases, weather station measurements show higher values, while evapotranspiration from MOD16A2GF follows the opposite trend. After the rainy season, the values converge again as reduced humidity creates more favorable conditions for soil evaporation and plant transpiration.

The study revealed a strong correlation in the data and applied machine learning techniques to predict ET using three models. The Random Forest model performed the best, achieving an R^2 value of 0.76 and MSE of 132.47, indicating high predictive accuracy. This suggests that ET values derived from MOD16A2GF, using PET multiplied by the crop coefficient, can effectively monitor and predict evapotranspiration in the study area. The findings highlight the potential of satellite-based data and machine learning for enhancing water resource management.

Future work could involve testing the method on other crops and in different regions for a broader assessment. Additionally, applying this approach to areas with varying climates such as arid or tropical regions, will help evaluate the model's validity across diverse environmental contexts. This will contribute to the development of a comprehensive and effective model that aligns with the specific needs and context of the country's agricultural water management.

ACKNOWLEDGMENT

This research has received funding support from the NSRF via the Program Management Unit for Human Resource & Institutional Development, Research and Innovation [grant number B13F670082-8]. We wish to thank the Geo-Informative Space Technology Development (Public Agency) for recommend and weather station data used in this study.

REFERENCES

- [1] "Report: Agricultural Statistics of Thailand 2023," Office of Agricultural Economics, 2024.
- [2] "Census of Agriculture 2023 whole kingdom," National Statistical Office Thailand, 2024.
- [3] "Master Plan on Water Resource Management 20 Years (2018-2037)," Office of the National Water Resources Thailand, 2018.
- [4] Guangyao Gao, Di Wang, Tianshan Zha, Lixin Wang, Bojie Fu, "A global synthesis of transpiration rate and evapotranspiration partitioning in the shrub ecosystems," Journal of Hydrology 606: 127417, 2022.
- [5] Mansoor Al Tamimi, Steve Green, Zied Hammami, Khalil Ammar, Mouza Al Ketbi, Ali M. Al-Shrouf, Mohamed Dawoud, Lesley Kennedy, Brent Clothier, "Evapotranspiration and crop coefficients using lysimeter measurements for food crops in the hyper-arid United Arab Emirates," Agricultural Water Management 272: 107826, 2022.
- [6] Hong Wang, Fubao Sun, Fa Liu, Tingting Wang, Wenbin Liu, Yao Fenga, "Reconstruction of the pan evaporation based on meteorological factors with machine learning method over China," Agricultural Water Management 287: 108416, 2023.
- [7] Yang Yang, Yufeng Luo, Conglin Wu, Hezhen Zheng, Lei Zhang, Yuanlai Cui, Ningning Sun, Li Wang, "Evaluation of six equations for daily reference evapotranspiration estimating using public weather forecast message for different climate regions across China," Agricultural Water Management 222: 386-399, 2019.
- [8] Radoslaw Guzinski, Hector Nieto, Juan Manuel Sánchez, Ramón López-Urrea, Dalenda Mahjoub Boujnah, Gilles Boulet, "Utility of Copernicus-Based Inputs for Actual Evapotranspiration Modeling in Support of Sustainable Water Use in Agriculture," IEEE Journal of selected topic in applied earth and remote sensing vol.14: 11466, 2021.
- [9] G.W. Kite, P. Droogers, "Comparing evapotranspiration estimates from satellites, hydrological models and field data," Journal of Hydrology 229: 3-18, 2020.
- [10] A.L. Ruhoff, A.R. Paz, L.E.O.C. Aragao, Q. Mu, Y. Malhi, W. Collischonn, H.R. Rocha, S.W. Running, "Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin," Hydrological Sciences Journal 58 (8): 1685, 2013.
- [11] Luis S. Pereira, Paula Paredes, Dalila Espírito-Santo, Maher Salman, "Actual and standard crop coefficients for semi-natural and planted grasslands and grasses: a review aimed at supporting water management to improve production and ecosystem services," Irrigation Science 42: 1139-1170, 2023.
- [12] Steven W. Running, Qiaozhen Mu, Maosheng Zhao, Alvaro Moreno, "User's Guild MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3 and Year-end Gap-filled MOD16A2GF/A3GF) NASA Earth Observation System MODIS Land Algorithm (For Collection 6.1)," Lpdaac USGS, Mar. 2021. [Online]. Available: https://lpdaac.usgs.gov/documents/931/MOD16_User_Guide_V61.pdf
- [13] Muhammad Wagas, Adila Naseem, Usa Wannasingha Humphries, Phyoe Thandar Hlaing, Muhammad Shoaib, Sarfraz Hashim, "A comprehensive review of the impacts of climate change on agriculture in Thailand," Farming System column 3: 100114, 2024.
- [14] Nutchanart Sriwongsitanon, Thienchart Suwawong, Sansarit Thianpopirug, James Williams, Li Jia, Wim Bastiaanssen, "Validation of seven global remotely sensed ET products across Thailand using water balance measurements and land use classifications," Journal of Hydrology: Regional Studies 30: 100709, 2020.
- [15] Richard G. ALLEN, Luis S. PEREIRA, Dirk RAES, Martin SMITH, "Crop Evapotranspiration. (Guidelines for Computing Crop Water Requirements)," FAO Irrigation and Drainage Paper No. 56, 1998.
- [16] Keyu Xiang, Yi Li, Robert Horton, Hao Feng, "Similarity and difference of potential evapotranspiration and reference crop T evapotranspiration – a review," Agriculture Water Management 232: 106043, 2020.
- [17] "Evapotranspiration of plant," Nakhon Sawan Field Crops Research Center Department of Agriculture Thailand, 2016.

Spatial distribution of Fireflies (Coleoptera: Lampyridae) A Case Study of Kwan Phayao Area, Phayao Province, Thailand.

Panupong Chaiwongsen*

School of science

University of Phayao

Phayao, Thailand

phanuphong.ch@up.ac.th

Sitthisak Pinmongkhonkul

School of science

University of Phayao

Phayao, Thailand

sitthisak.pi@up.ac.th

Arthit Nuntakwang

School of science

University of Phayao

Phayao, Thailand

arthit.nu@up.ac.th

Boonchuang Boonsuk

School of science

University of Phayao

Phayao, Thailand

boonchuang.bo@up.ac.th

Manas Titayavan

School of Agriculture and Natural

Resources

University of Phayao

Phayao, Thailand

manas.ti@up.ac.th

Warin Boonriam

Faculty of Environment and Resource

Studies

Mahidol University

Nakhon Pathom, Thailand

warin.boo@mahidol.edu

Sorasak Nak-eiam

Department of Biology, Faculty of Science and Technology

Rambhai Barni Rajabhat University

Chanthaburi, Thailand

sorasak.n@rbru.ac.th

Abstract— This study examined the richness of fireflies and their ecological features in Kwan Phayao, contrasting disturbed community areas with woodland regions from January to November 2024. The research employed random sampling over a 100-meter line transect to examine the interaction between natural water sources, plant variety, and firefly populations. The investigation identified 157 adult fireflies from the Luciolinae subfamily, encompassing three genera: Sclerotia (44 specimens), Asymmetricata (5 specimens), and Pygoluciola. Species classification identified Sclerotia aquatilis (86.27%), Asymmetricata circumdata (9.80%), and Pygoluciola sp. (3.92%). Station 1 had the largest average number of adult fireflies (75), connected with vast plant variety and substantial floating plant coverage, notably water hyacinths. The geospatial investigation of Kwan Phayao revealed significant biological variances across several places. Northwestern regions, typified by limited precipitation and scant vegetation, exhibited dramatically less biodiversity compared to other places. The geographical inequalities highlight the urgent necessity for focused conservation initiatives.

Keywords— Firefly diversity, habitat, Kwan Phayao, biological indicators, Thailand

I. INTRODUCTION

Thailand is a major hotspot for biodiversity, having a high species richness and quantity of insects [1]. Since insects are located all throughout the nation, the biodiversity and natural balance have been greatly influenced by the quick changes in the natural and man-made environments. Fireflies in Thailand serve as significant bioindicators of ecological health and natural balance. Terrestrial habitats, especially in locations

with rich forests, as well as freshwater and brackish water habitats are all home to these insects. Fireflies are important in both medical and agricultural settings because their larvae eat a variety of snails, including the apple snail (*Pila ampullacea*) and the golden apple snail (*Pomacea canaliculata*), which are known to transmit intestinal flukes and meningitis, among other diseases that affect humans and animals. Phayao Province has a high biodiversity due to its largely wooded landscapes and natural water sources such as Kwan Phayao, which acts as a food supply and nursery for firefly larvae. Given the province's abundant ecological and physical resources, it is projected to be another hotspot for firefly biodiversity.

Firefly populations had a significant reduction of 70% in 2009 [4], attributable to many reasons such as habitat degradation, light pollution, and pesticide application [9]. Research in the agricultural region of Ban Bua, Ban Tun Sub-district, Mueang Phayao District, revealed that *Sclerotia seriata* was the predominant firefly species in zones with elevated relative humidity near tropical woods [2]. Artificial illumination has a substantial influence on firefly mating behavior. Both male and female mating behaviors are impacted by light disturbances from human sources, such as street lights and domestic lighting [15].

Spatial Pattern Analysis of Object Values According to the First Law of Geography: All Entities are Interconnected, Yet Proximate Entities Exhibit Stronger Relationships than Distant Ones [17]. Spatial phenomena consistently reveal that analogous qualities are often present in proximate regions. Rainfall quantities are generally clustered, with regions of significant precipitation located near one another, while places undergoing drought are frequently situated in close proximity. This phenomenon is referred to as Spatial Autocorrelation.

Spatial autocorrelation encompasses statistical techniques employed to examine the correlations between variables in proximate regions relative to those in more distant places. It illustrates how the distribution of variable values is contingent upon the spatial arrangement of items. The Moran's I statistic is among the most used statistical techniques employed for this investigation.

This research examines the impact of environmental circumstances on firefly populations, species diversity, and plant species in both disturbed and undisturbed natural environments. The results will serve as a great resource for developing conservation strategies for fireflies and their natural habitats, offering vital information for habitat protection and preservation.

II. METHOD

A. Study area

The wetlands around Kwan Phayao Lake (coordinates 19.155874, 99.890957) and chosen regions with abundant adult firefly populations were researched across ten research stations as follows: Station 1: Nong Bua Daeng Ban, San Klang, Station 2: Wetlands near the restaurant, Station 3 - Ban Thung Kio Ancient Pier, Station 4: Wetlands near the coffee shop, Station 5: Wetlands inside the fish protection zone at Ing Kwan Beach, Ban San Nong Niao, Station 6: Wetlands near Khun Det Bridge, Station 7: Ban Than Dok Mai Pier, Station 8: Fish protection zone in Ban Pratu Prasat, Station 9: Wetlands near Hong Hian Kwan Phayao, Station 10: Phayao Inland Fisheries Research and Development Centre, as illustrated in Fig. 1



Fig. 1. Firefly study areas: Kwan Phayao study site

B. Establishment of study area

Using systematic random sampling techniques, adult firefly specimens were gathered in the research locations once a month on average between January 2024 and November 2024 [6]. The collection employed insect sweep nets from 18:00 to 24:00 hours, aligning with peak firefly activity periods [14]. The sweep nets were deployed to collect flying firefly and promptly closed to avert specimen escape. Specimens were put in plastic containers with damp cotton to preserve humidity. Data gathering involved utilizing a GPS receiver to record the date, time, and geographical coordinates as well as the weather, plant types, and habitat features in regions where firefly were discovered. This method acted as the first preparation for further laboratory analysis.

C. Determination of physico-chemical parameters

Temperature and relative humidity were quantified utilizing a digital thermo-hygrometer throughout the examination of physical characteristics. Water quality indicators, such as pH and dissolved oxygen (DO), in the research regions were assessed using a Multiparameter Waterproof Meter. Average precipitation data for the study regions were acquired from the Thai Meteorological Department.

D. Study of Biological Parameters

Examination of biological parameters plant specimens, encompassing trees, riparian flora, aquatic vegetation, and floating plants, were documented and gathered for species identification. The research also encompassed an examination of firefly alimentary sources by the sampling of flora from both riparian zones and aquatic regions within the study locations.

E. Statistical analyses

The species diversity index was assessed utilizing the Shannon-Wiener function (Krebs, 1999). Species richness was determined utilizing Margalef's index. The evenness index was assessed utilizing Pielou's evenness index. The relative abundance was assessed with Pettingill's technique (1950). Sorenson's Similarity Coefficient was used to determine the species similarity between stations [5]

F. Analysis of Geographical Information System (GIS) Technology Data

Automated Spatial Autocorrelation Analysis of Species DistributionThe geographical distribution of species in ecosystems frequently reveals spatial linkages influenced by environmental variables, including climate and topography. This investigation might be performed utilizing Python programming with the Folium module for spatial data viewing and examination.

III. RESULTS AND DISCUSSION

Classification of mature firefly

The study revealed that Kwan Phayao contained 155 adult fireflies. They can be categorized under the subfamily Luciolinae, which includes three genera: *Sclerotia* (142 specimens), *Asymmetricata* (8 specimens), and *Pygoluciota* (5 specimens). The classification at the species level was as follows: *Sclerotia aquatilis* (142 specimens: 91.62%), *Asymmetricata Circumdata* (8 specimens: 5.16%), and *Pygoluciola* sp. (5 specimens: 3.23%) (Table I).

Species composition

No adult fireflies were detected in the research area from January 2024 to May 2024, based on the survey of specimens collected between January 2024 and November 2024. Adult fireflies were initially documented in June 2024, with a total of 238 specimens identified. Subsequently, there were 198 specimens in July, 172 in August, 160 in September, 161 in October, and 165 in November. (Fig. 2)

Table I. Distribution of Firefly Populations Across Various Research Stations in Kwan Phayao

Species	Number of individuals/study site										total
	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	
<i>Abscondita chinensis</i>											
<i>Asymmetricata circumdata</i>		1			4			3	8		
<i>Asymmetricata ovalis</i>											
<i>Luciola curitibana</i>											
<i>Medeoteryx</i> sp.											
<i>Pyrocoelia</i> sp.							5			5	
<i>Sclerotia aquatilis</i>	35	23	24	15	19	19			7	142	
total	35	23	25	15	19	23	5	0	0	10	155
Richness	-	-	0.31	-	-	0.32	-	-	-	-	0.40
Evenness	-	-	0.05	-	-	0.15	-	-	-	-	0.26
Shannon-Winer diversity	-	-	0.17	-	-	0.46	-	-	-	-	0.61

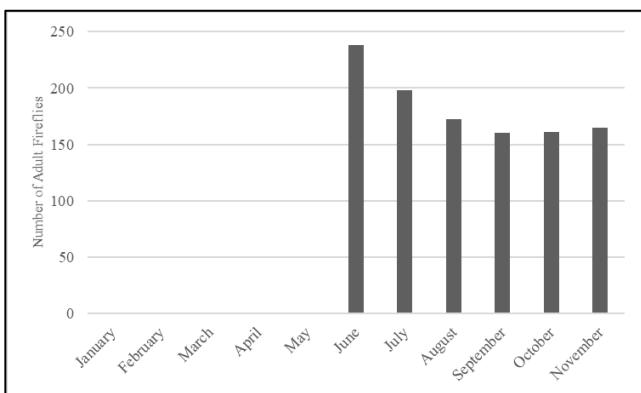


Fig. 2. Comparison of Counted Adult Fireflies Across Months from January to November 2024 in the Kwan Phayao Study Area

The extensive survey of adult fireflies, carried out from January to November 2024, including ten separate study sites to ascertain population dispersal across various areas. The study uncovered a diverse population density, highlighting notable findings about the presence and lack of fireflies. Stations 8 and 9 had a total lack of adult fireflies over the whole survey duration. The other sites displayed differing degrees of firefly populations. Station 1 emerged as the most populated station, boasting an average of 60.33 specimens, markedly surpassing other locations. Station 5 ranked as the second most populous location, accommodating an average of 41.00 fireflies. The remaining sites had more moderate populations: Station 2 averaged 23.25 specimens, Station 3 recorded 20.50, while Station 4 displayed 21.00 firefly. Reduced population densities were noted in the other sites, with Station 7 averaging 6.50 specimens, Station 10 documenting 10.00 firefly, and Station 6 exhibiting the lowest population at just 3.60 specimens. Fig. 3 clearly depicts the firefly distribution over the examined areas. The study provides significant insights into the geographical distribution of adult firefly populations over the designated survey period.

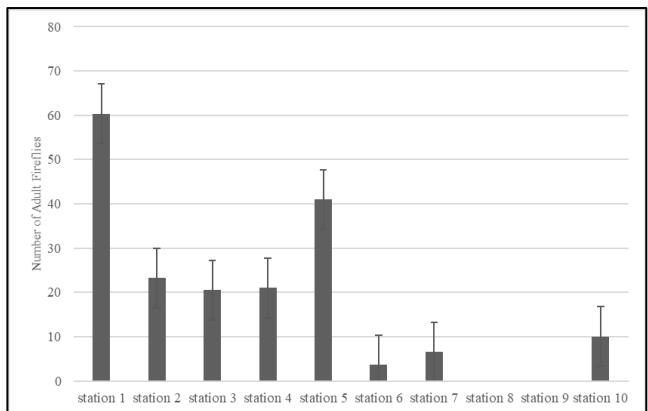


Fig. 3. The mean quantity of adult fireflies seen at each station

A survey of plant communities in the research sites

The examination of plant data as firefly habitats yielded the following conclusions across multiple stations:

- Station 1: Water primrose (*Jussiaea repens*) was the most abundant (13.55%), followed by yellow bladderwort (*Utricularia aurea*) (12.30%).
- Station 2: The most common plant was giant reed (*Arundo donax*) (27.50%), followed by Asian watermoss (*Salvinia cucullata*).
- Station 3: The most prevalent species was giant reed (*Arundo donax*), followed by Asian watermoss (*Salvinia cucullata*) (21.70%).
- Station 4: Asian watermoss (*Salvinia cucullata*) had the highest density (27.15%), followed by water fern (*Azolla pinnata*) (22.88%).
- Station 5 has the highest abundance of Asian watermoss (*Salvinia cucullata*) (40.95%), followed by para grass (*Brachiaria mutica*).
- Station 6: Yellow velvetleaf (*Limnocharis flava*) had the highest density (24.32%), followed by water primrose (16.44%).
- Station 7: The most abundant plants were giant reed (*Arundo donax*) and water fern (*Azolla pinnata*), with water primrose (*Jussiaea repens*) coming in second at 13.59%.
- Station 8: Asian watermoss (*Salvinia cucullata*) had the highest density (27.15%), followed by water fern (*Azolla pinnata*) (22.88%).
- Station 9: Water primrose (*Jussiaea repens*) was the most abundant (13.19%), followed by water fern (*Azolla pinnata*) (13.15%).
- Station 10: Asian watermoss (*Salvinia cucullata*) was the most common (24.82%), followed by gigantic reed (22.27%) in Table II.

Water quality assessments in the study

Station 4 had the highest average water pH level (7.99 ± 1.30), while Station 11 had the lowest average of 6.52 ± 0.41 . Station 6 had the greatest dissolved oxygen (DO) level at $4.35 \pm 0.65 \text{ mg/l}$, whereas Station 4 had the lowest average at $3.01 \pm 0.68 \text{ mg/l}$. Water conductivity measurements revealed Station 4 had the greatest value at $825.41 \pm 0.50 \mu\text{s/cm}$, while Station 10 had the lowest at $274.21 \pm 41.42 \mu\text{s/cm}$. TDS levels were highest at Station 3 ($342.14 \pm 41.25 \text{ mg/l}$) and lowest at Station 9 ($123.90 \pm 12.45 \text{ mg/l}$). Table III some Common Mistakes

Table II. Plant species of adult fireflies in the Kwan Phayao

Family	Scientific Name	Relative density									
		1	2	3	4	5	6	7	8	9	10
CYPERACEAE	<i>Cyperus alburnifolius</i>	0.66	1.26	1.25		2.31	0.47		1.50		1.74
BALSMINACEAE	<i>Hydrocoris triflora</i>	0.38				2.49					0.87
ARACEAE	<i>Colocasia esculenta</i>	0.09					0.51				0.20
TYPHACEAE	<i>Typha angustifolia</i>	0.61					2.13	3.05			1.73
FABACEAE	<i>Nepenthes oliverae</i>	0.50			2.71		0.46	3.64	2.71	0.35	2.90
POLYGONACEAE	<i>Polygonum tonentense</i>	4.34	6.46	7.00	9.95	14.72	9.57	12.15	9.95	7.42	9.82
MIMOSACEAE	<i>Mimosa pigra</i>	6.03	1.82	2.08	2.91		7.60	11.52	2.91	9.27	
POACEAE	<i>Bracharia mutica</i>	12.18	15.74	13.64	2.07	19.06	4.80	2.50	2.07	3.70	17.47
POACEAE	<i>Aruno donax</i>	6.66	27.50	25.04	14.54		25.27	7.60	17.82	14.54	10.45
SALVINIACEAE	<i>Salvinia cucullata</i>	5.80	22.05	21.70	27.15	40.95	9.12	13.21	27.15	12.12	24.82
PONTEDERIACEAE	<i>Eichornia crassipes</i>	13.55	11.77	12.27	15.51		16.44	13.59	15.51	13.19	
AZOLLACEAE	<i>Azolla pinnata</i>	9.69	6.96	9.56	22.88		10.64	17.82	22.88	13.05	20.99
LENNACEAE	<i>Spiriodela polyrrhiza</i>	3.26	6.46	7.47	0.90		0.52	0.51	0.90	0.43	
LIMBULARIACEAE	<i>Utricularia aurea</i>	12.30									
BUTOMACEAE	<i>Limnocharis flava</i>	5.83					24.32				22.75
NYMPHAEACEAE	<i>Nymphaea lotus</i>	3.52									
Total		100	100	100	100	100	100	100	100	100	100

Table III. Water parameters at each station (Mean \pm SE) of Kwan Phayao

Parameter	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10
pH	7.52 \pm 0.45	7.82 \pm 1.12	7.98 \pm 0.09	7.99 \pm 1.30	6.87 \pm 0.29	6.88 \pm 1.20	6.91 \pm 0.23	6.95 \pm 0.11	6.97 \pm 0.14	6.77 \pm 0.55
Dissolved oxygen (mg/L)	4.35 \pm 0.65	4.07 \pm 0.79	3.79 \pm 0.12	4.07 \pm 0.42	4.02 \pm 0.42	3.24 \pm 0.43	4.03 \pm 0.17	3.01 \pm 0.68	3.03 \pm 0.75	4.02 \pm 0.42
Conductivity (µmho/cm)	387.93 \pm 12.46	400.16 \pm 13.45	701.12 \pm 32.25	825.41 \pm 0.50	294.50 \pm 11.28	284.20 \pm 10.25	274.50 \pm 10.50	281.21 \pm 11.24	286.41 \pm 12.23	274.21 \pm 41.42
Total dissolved solids (mg/L)	286.79 \pm 19.87	289.74 \pm 12.47	342.14 \pm 41.23	142.21 \pm 21.30	228.13 \pm 12.45	142.52 \pm 65.41	178.13 \pm 42.15	329.50 \pm 10.50	123.90 \pm 12.45	192.74 \pm 1.85

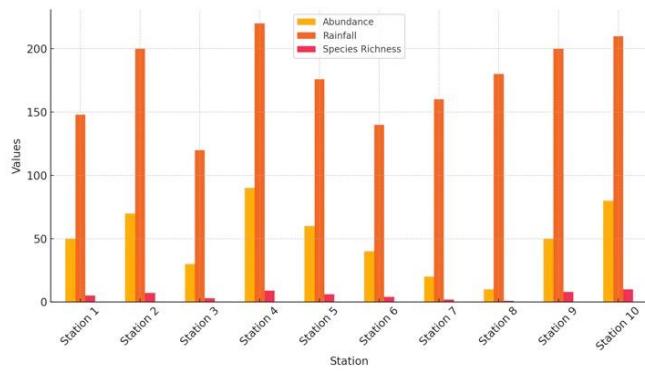


Fig. 4. Chat: Comparison of Abundance, Rainfall, and Species Richness



Fig. 5. Comprehensive Map: Abundance, Rainfall, and Species Richness



Fig. 6. Tooltips: Hover over circles for detailed station data. Circle Color and Size of circle represent abundance and proportional of species richness, respectively

III. RESULTS AND DISCUSSION

The survey data from January to November 2024 presents an intricate interplay of ecological factors in Kwan Phayao. Spatial distributions of species abundance indicate high-density clusters near stations 1 and 5. This is explained in Fig. 4, as evidenced by monthly abundance data. These clusters align closely with water quality indicators such as dissolved oxygen (DO) and pH levels, which were highest near these stations. Temporal trends reveal that species like *Sclerotia aquatilis* proliferated significantly during the rainy season (June to August), coinciding with a spike in rainfall (up to 220 mm). This suggests a positive correlation between rainfall and habitat quality enhancement in these areas, which is explained in Fig. 4.

Where dissolved oxygen levels above 4.35 mg/L and consistent pH readings around 7.5 contributed to favorable conditions for aquatic life. Meanwhile, temperature and humidity also showed moderate impacts, especially during summer months, limiting species diversity in some regions. Locations with denser vegetation, as noted in the plant diversity dataset, provided essential shelter and food resources, creating microhabitats that foster species richness, particularly in the southern parts of Kwan Phayao, which is explained in Fig. 5.

The investigation discovered no mature fireflies at stations 8 and 9 throughout the research period. Station 8 had the lowest dissolved oxygen (3.01 ± 0.68) and Total dissolved solids (329.50 ± 10.50) among all stations, according to statistical analysis. This was owing to station 8's high concentration of dissolved organic matter, which resulted in an abundance of microbes, aquatic insects, phytoplankton, and algae. Furthermore, Station 8's closeness to hospital facilities and hotels influenced adult firefly mating behavior through light from buildings and dorms. Human activities, automobile lights, and street lighting all had an influence on adult firefly mating behavior, as females respond to males by flashing lights [15]. Environmental light pollution (0.05-0.3 lux) increased courting duration and male-female mounting time, suggesting that artificial light inhibited natural firefly mating [16]. Station 1 had the largest firefly population, averaging 60.33 mature fireflies per station. The survey discovered abundant floating plant cover, including

water hyacinth (*Eichhornia crassipes*), Asian watermoss (*Salvinia cucullata*), and water fern (*Azolla pinnata*), which act as egg-laying locations for adult fireflies. The lush riparian vegetation offered ideal environment for firefly pupae [15]. Because of the distance from buildings, the lack of light pollution allowed for uninterrupted firefly mating, resulting in the greatest firefly population [15][16]. Kwan Phayao collects wastewater from buildings, residences, hospitals, and hotels, contaminating the water with suspended particles, sediments, algae, and bacteria [5]. The firefly species *Sclerotia aquatilis* thrives in contaminated water (Thancharoen, 2007) and may be used as a biological water quality indicator [6].

Mapping rainfall, vegetation types, and species distribution with geospatial coordinates highlights strong spatial variability. Between station 5 and station 6, characterized by minimal rainfall and sparse vegetation, had significantly lower abundance levels. This spatial heterogeneity emphasizes the importance of targeted conservation efforts in regions where environmental quality and biological abundance diverge. By integrating these multi-faceted factors, the study not only underscores the dynamic nature of Kwan Phayao's ecosystem but also provides a robust framework for future ecological management.

ACKNOWLEDGMENT

This research recognizes the undergraduate biology student for their and contributions to sample collection and extends gratitude for the considerable support from their families. I would like to express my gratitude to Dr. Swarin Rerkusuk for his support in geographical data analysis. This research received financing from the University of Phayao and the Thailand Science Research and Innovation Fund (Fundamental Fund 2024).

REFERENCES

- [1] Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., and Hilton-Taylor, C. (2002), "Habitat loss and extinction in the hotspots of biodiversity," *Conservation biology*, 16, 4, 900-923.
- [2] Chaiwongsaen, P. (2020), "Diversity of fireflies in the agricultural areas of bann-bua, bann-tun sub-district, mueang district, phayao province," The master of science degree in biology, University of phayao. Phayao.
- [3] Fu, X., Meyer-Rochow, V.B., Tyler, J., Suzuki, H., and De Cock, R. (2012), "Structure and function of the eversible organs of several genera of larval firefly (Coleoptera: Lampyridae)," *Chemoecology* 19, 155-168.
- [4] Hadley, D. 2009, "Habits and Traits of Fireflies, Family Lampyridae," Retrieved March 1, 2024, Retrieved from <https://www.thoughtco.com/fireflies-family-lampyridae-1968148>
- [5] Institute for Innovative Learning, Mahidol University, "Water Pollutants," Retrieved March 25, 2021, from https://il.mahidol.ac.th/e-media/ecology/chapter3/chapter3_water7.htm
- [6] Krebs, C.J. (1999), "Ecological Methodology," Benjamin/Cummings. Menlo Park, CA.
- [7] Konsue, W., Jaisukul, C., & Priyanont, P. (2014), "Biodiversity of insect groups as environmental indicators and benefits for agro-tourism and eco-tourism," Research report funded by RDPB for fiscal year 2014. Department of Biology, Faculty of Science, Chulalongkorn University, 20-22.
- [8] Lawrence, J. F. (1982), "Order Coleoptera," 114-170.
- [9] Marks, K. (2009), "Fireflies in decline as natural habitats are destroyed," Retrieved March 1, 2024, Retrieved from <https://www.abacademies.org/articles/guidelines-for-promoting-tourism-from-biodiversity-of-fireflies-in-agricultural-areas-of-banbua-phayao-province-thailand-13319.html>
- [10] McDermott, F. A. (1966), "Lampyridae Coleopterorum Catalogus Supplementa," *Secunda* 9, 1-149.
- [11] Nak-eiam, S. (2015), "Taxonomy and species distribution of fireflies (Coleoptera; Lampyridae) in the North of Thailand," Doctoral dissertation, Ph.D., Naresuan University, Phitsanulok
- [12] Phongchan, N. (2013), "Species, abundance and light emission of fireflies in Welu Estuary mangrove forest," Chanthaburi Province. Master's thesis, Chulalongkorn University, Bangkok.
- [13] Sara M. Lewis and Ona T. Wang (1991), "Reproductive Ecology of Two Species of *Photinus* Fireflies (Coleoptera: Lampyridae)," *Psyche A Journal of Entomology* 98(4).
- [14] Sommitr, W. (2004), "Ecological Population of *Pyrocoelia praetexta* Olivier in Peach Botanical Garden," Chiang Mai: Graduate School, Chiang Mai University.
- [15] Takeda, M., Amano, T., Kato, K., and Higushi, H. (2004), "The habitat requirement of the Genji-firefly *Luciola cruciata* (Coleoptera: Lampyridae)," a representative endemic species of Japanese rural landscapes. *Biodiversity and Conservation* (2006) 15, 191–203
- [16] Thancharoen, A. (2007), "The Biology and mating behavior of an aquatic firefly species, *Luciola aquatilis* sp. nov. Thancharoen (Coleoptera: Lampyridae)," Doctoral dissertation, Ph.D., Mahidol University, Bangkok.
- [17] Tobler, W. (1974), "Spatial autocorrelation and the archaeologist: Introductory comments presented to the Society for American Archaeology." Washington, D.C.

Utilizing Unmanned Aerial Vehicles to assess carbon sequestration in trees conducted in response to plant genetic conservation project under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn at Chiang Rai Rajabhat University.

1st Krittawit Suk-ueng
Faculty of Science and Technology
Chiang Rai Rajabhat University
Chiang Rai, Thailand
nsukung@gmail.com

2nd Sawarin Lerk-u-suwe
Department of Geographic Information
Science, School of Information and
Communication Technology
University of Phayao
Phayao, Thailand
sawarin.l1@gmail.com

3rd Pensri Malithong
Technology and Innovation Transfer
Institute (TTI)
Chiang Rai Rajabhat University Chiang
Rai, Thailand
pensrimalithong@gmail.com

4th Sutti Malithong
Technology and Innovation Transfer
Institute (TTI)
Chiang Rai Rajabhat University
Chiang Rai, Thailand
kor1084@gmail.com

5th Pemmanee Senasutthiphaphan
Faculty of Science and Technology
Chiang Rai Rajabhat University
Chiang Rai, Thailand
631459003@crru.ac.th

Abstract— This study aimed to evaluate carbon sequestration derived from field surveys and unmanned aerial vehicles (UAV) in the plant genetic protection area established under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn at Chiang Rai Rajabhat University (RSPG-CRRU), encompassing 4.66 hectares. Four sample plots of 50 by 20 meters were randomly created. We assessed the diameter and height of trees within the sample plots. The above-ground biomass of trees was 95.78 metric tons per hectare. Carbon sequestration amounted to 45.01 tons of carbon per hectare, whereas total carbon dioxide absorption reached 165.22 tons of CO₂ equivalent per hectare. The utilization of UAVs to evaluate above-ground carbon sequestration shown a strong correlation ($r = 0.97$) with field data, and the root mean square error was ± 0.38 meters. UAV data indicates that the above-ground biomass was 29.02 tons per hectare. The carbon sequestration of trees was 13.65 tons of carbon per hectare, while the carbon dioxide absorption was 106.55 tons of CO₂ equivalent per hectare.

Keywords—carbon, environment, forest, greenhouse gases, sustainability

I. INTRODUCTION

Emissions of greenhouse gases from anthropogenic activities, including the industrial revolution, are an important cause of the global climate change issue. Forests, on the other hand, sequester natural carbon in the form of biomass, which is stored in various parts of the tree such as leaves, roots, and trunk. [1], [2], [3], [4]. Gathering spatiotemporal data on the physical attributes of forests, derived from tree biomass, is a preferable option over field surveys for assessing carbon sequestration. Using UAVs to gather these physical characteristics of forests aligns with the objectives of reducing emissions from deforestation and

forest degradation, as well as the role of conservation, sustainable forest management, and enhancing the forest carbon stock (REDD+) of the United Nations Framework Convention on Climate Change (UNFCCC) and the United Nations Sustainable Development Goals (SDGs). These goals emphasize urgent action to combat climate change, minimize its disruptions, slow down the rate of global warming, and reap the benefits of carbon trading in the future [5], [6], [7], [8].

The development of geospatial technologies has transformed natural science research, enabling more precise and dynamic analysis of ecological systems. This alignment with cutting-edge geospatial technology allows for high-resolution mapping and real-time monitoring. The rapid development of geospatial technologies, especially satellite remote sensing, machine learning via Google Earth Engine (GEE) platform and Unmanned Aerial Vehicles (UAVs), has significantly enhanced forest carbon storage. Recent advancements in geospatial technologies, particularly the integration of Unmanned Aerial Vehicles (UAVs), have revolutionized natural resources monitoring and analysis, offering unprecedented spatial and temporal resolution. This alignment with modern geospatial tools significantly enhances data accuracy and analytical depth for supporting natural resources management and conservation strategies [9], [10].

Thailand has strategies to reduce carbon dioxide emissions by increasing green spaces in organizations, public and private sectors, educational institutions, and communities. Universities, both domestic and international, have emphasized the importance of these efforts by implementing Green University projects [11], [12]. Chiang Rai Rajabhat University has developed a 5-year strategic

plan (2022-2026), aiming to have higher education institutions rank among the world's green universities under Strategy 5 (Education for enhancing quality of life in an environmentally friendly manner). Chiang Rai Rajabhat University has a protected forest area committed to 4.66 ha of the Plant Genetic Protection Area under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn (RSPG-CRRU) [13], [14]. This study aims to assess forest carbon sequestration obtained from field surveys and unmanned aerial vehicles in RSPG-CRRU for Chiang Rai Rajabhat University as a basis for future Green University projects.

II. METHODS

A. Study area and sample plots

The plant genetic protection area at Chiang Rai Rajabhat University, under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn, spans 4.66 ha. Four sample plots (estimated 8.6% of the study area) of 50×20 meter, namely plot A, plot B, plot C and plot D, were determined from May to June 2023 (see Fig. 1). We measured the diameter at breast height (DBH) and height of all trees in the sample plots to be approximately 4.5 centimeters [15], [16].



Fig. 1. Sample plots in the study area.

B. Unmanned Aerial Vehicle Image Processing

The UAV images, achieved from DJI Phantom 4 Pro V2.0 supported by School of Information and Communication Technology, University of Phayao, Thailand, were permitted from Airports of Thailand to fly legally on 16 August 2022 and captured on 19 August 2022. The flight plan, a flying altitude of 80 meters, for the study area is shown in Figure 2 before image processing in Dronebox platform, resulted in Digital Surface Model (DSM), Digital Terrain Model (DTM) and Canopy Height Model (CHM) [17] (see Fig. 3)

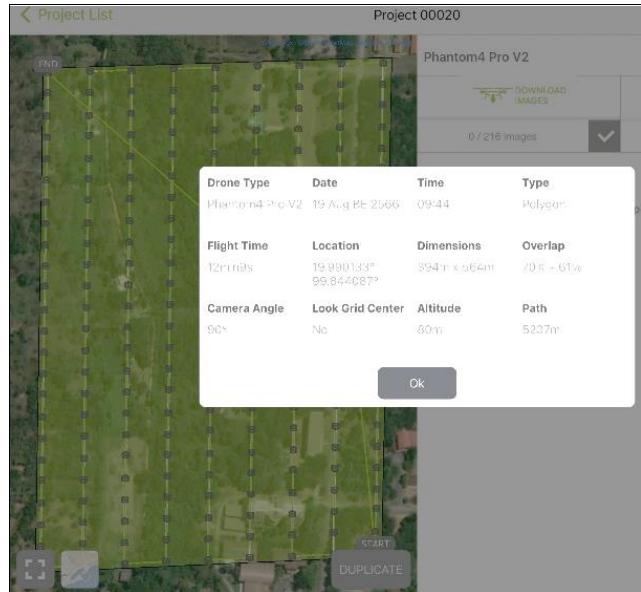


Fig. 2. Flight plan (70% overlap and 60% sidelap).

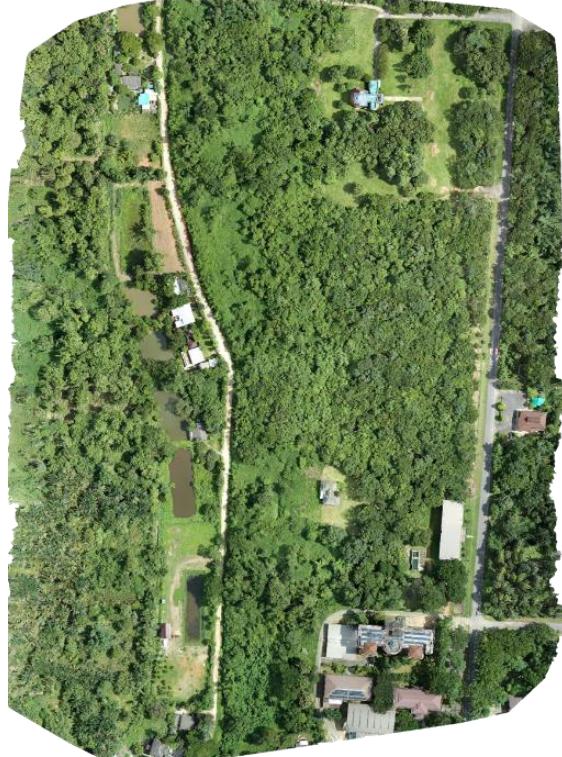


Fig. 3. UAV images processed from Dronebox platform.

C. Above-ground biomass and carbon sequestration assessment methods

1) Above-ground biomass and carbon sequestration assessment archived from field measurement: Allometric equation of mixed deciduous and deciduous dipterocarp forests was used for calculating above-ground biomass following Ogawa et al. [18]. The research estimated carbon stocks using above-ground biomass with a carbon fraction of 0.47 and calculated carbon absorption through multiplication by 3.67 following Gibbs et al. [19].

2) Tree heights derived from field surveys and those obtained from UAV data, based on a selection of 30 trees, were analyzed for root mean square error (RMSE). The correlation between DBH derived from field data and the tree height obtained from UAV was analyzed using Pearson's

correlation. Subsequently, DBH derived from tree heights acquired via UAV was predicted by using simple regression analysis at a 95% confidence level with the R statistical software [20]. Equation (1) is commonly used to estimate the average carbon stock (C_{ABG}) in each grid (creating a grid with dimensions of 0.1×0.1 meters to correspond with the positions of trees observed in the field) by multiplying the average above-ground biomass of the area per tree in each grid (kg) by the number of trees presented in each grid (T) and the carbon fraction value (CF) of 0.47 [21]. Subsequently, importing the grid and the canopy height model into Zonal Statistics in QGIS version 3.32 [22], [23].

$$\overline{C_{ABG}} = \overline{W_{tc}} \times T \times CF \quad (1)$$

III. RESULTS AND DISCUSSION

Table I revealed the differences between tree heights derived from field surveys and those acquired from UAV data based on 30 trees. Figure 4 displayed CHM derived from UAV data.

TABLE I Tree heights derived from field surveys and UAV data.

<i>Tree heights derived from field surveys (meters)</i>	<i>Tree heights derived from UAV data (meters)</i>
7.016	6.562
7.016	6.750
7.016	6.515
4.820	4.196
7.016	6.441
3.248	2.695
6.109	5.665
6.109	5.665
2.975	3.192
5.570	5.652
5.854	6.304
7.283	7.257
5.145	5.546
6.109	6.323
3.820	3.628
4.204	3.915
5.145	4.946
4.457	4.452
5.765	5.733
5.570	5.746
5.145	5.124
8.498	8.781
8.498	8.056
6.605	7.092
5.765	6.311
2.624	2.238
7.016	6.562

<i>Tree heights derived from field surveys (meters)</i>	<i>Tree heights derived from UAV data (meters)</i>
7.016	6.750
7.016	6.515
4.820	4.196

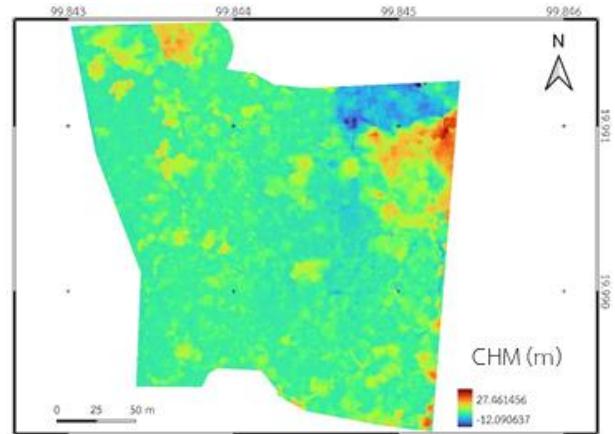


Fig. 4. Canopy Height Model derived from UAV data.

Figure 5 showed a high level of correlation ($R^2 = 0.933$) between tree heights recorded in field surveys and tree heights acquired by UAV data. Moreover, the correlation coefficient ($r = 0.97$) was quite high, and the RMSE was ± 0.38 meters. From Pearson's correlation between the DBH derived from field data and tree height collected from UAV for 30 trees, it was found that the correlation was 0.252, not statistically significant at the 0.05 level.

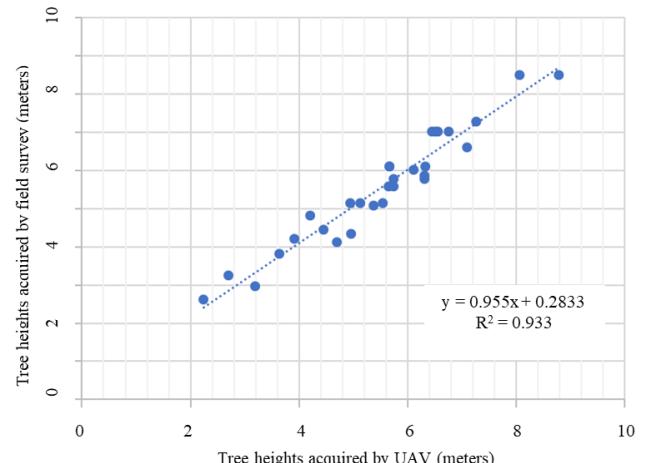


Fig. 5. Relationship between tree heights acquired by field survey and UAV data presented by R^2 .

Above-ground biomass of trees derived from field survey was 95.78 tons/ha. Moreover, the carbon sequestration was 45.01 tonne C/ha and the total carbon dioxide absorption was 165.22 tCO₂-eq/ha. Above-ground biomass obtained from UAV data was 29.02 tons/ha. The carbon sequestration of trees was 13.65 tonne C/ha and the carbon dioxide absorption was 106.55 tCO₂-eq/ha. These results may depend on tree species and related DBH [24]. The tree heights found from UAV data were lower than those found from field surveys (see Table I). In addition, the tree heights found from UAV data and DBH derived from field data were not significantly different. These differences affect the above-ground biomass, carbon stock, and carbon dioxide

absorption found from field surveys and UAV data (see Table II).

TABLE II Above-ground biomass (AGB), carbon stock (C-stock) and carbon dioxide (CO_2) absorption achieved from field survey and UAV data

Data	AGB (tons/ha)	C-stock (tonne C/ha)	CO_2 absorption (tCO ₂ -eq/ha)
Field	95.78	45.01	165.22
UAV	29.02	13.65	106.55

This study recommended that UAV images, achieved from DJI Phantom 4 Pro V2.0, derived from field surveys can be used for evaluating forest carbon sequestration in RSPG-CRRU. Optimal physical parameters, including altitude, image overlap and weather conditions should be considered for estimating carbon sequestration in forests [25], [26], [27]. Frontal overlap and sidelap should be raised to 80-90% and 60-75%, respectively. The picture trigger modes of UAV's flight should be stopping to capture [28], [29]. In addition, UAVs with LiDAR sensors may be useful for improving DSM and DTM accuracy [19].

IV. CONCLUSION

The above-ground biomass of trees in RSPG-CRRU was 95.78 tons/ha. The total carbon dioxide absorption was 165.22 tCO₂-eq/ha, and the carbon sequestration was 45.01 tons C/ha. The use of UAV to evaluate above-ground carbon sequestration has shown a strong correlation ($r = 0.97$) with field data, with a RMSE of ± 0.38 meters. According to UAV data, the above-ground biomass was 29.02 tons/ha. In addition, the carbon sequestration of trees was 13.65 tons C/ha, while the carbon dioxide absorption was 106.55 tCO₂-eq/ha.

ACKNOWLEDGMENT

The research was supported by Environmental Science and Technology Program, Faculty of Science and Technology, Chiang Rai Rajabhat University. In addition, School of Information and Communication Technology has provided crucial support during this research work.

REFERENCES

- [1] Intergovernmental Panel on Climate Change (IPCC), "Climate change 2021: The physical science basis," 2021, IPCC Sixth Assessment Report. [Online]. Available: <https://www.ipcc.ch/report/ar6/wg1/>
- [2] United Nations Environment Programme (UNEP), Emissions Gap Report 2022: The Closing Window. Nairobi: United Nations Environment Programme, 2022.
- [3] J. Rogelj, P. M. Forster, E. Kriegler, C. J. Smith and R. Séférian, "Estimating and tracking the remaining carbon budget for stringent climate targets," *Nature*, vol. 571, pp. 335-342, July 2019.
- [4] B. W. Griscom, J. Busch, S. C. Cook-Patton, P. W. Ellis, J. Funk, S. M. Leavitt, G. Lomax, W. R. Turner, M. Chapman, J. Engelmann, N. P. Gurwick, E. Landis, D. Lawrence, Y. Malhi, L. S. Murray, D. Navarrete, S. Roe, S. Scull, P. Smith, C. Streck, W. S. Walker and T. Worthington, "National mitigation potential from natural climate solutions in the tropics," *Phil. Trans. R. Soc. B.*, vol. 375, 20190126, Jan. 2020.
- [5] A. Hashem, "Estimation of aboveground biomass/carbon stock and carbon sequestration using UAV imagery at Kebun Raya Unmul Samarinda Education Forest, East Kalimantan, Indonesia," 2019, The degree of Master of Science in Geo-Information Science and Earth Observation. [online]. Available: <https://essay.utwente.nl/83721/1/hashem.pdf>
- [6] E. D. Vivar-Vivar, M. Pompa-Garcia, J. A. Martínez-Rivas and L. A. Mora-Tembre, "UAV-based characterization of tree-attributes and multispectral indices in an Uneven-aged mixed conifer-broadleaf forest," *Remote Sens.*, vol. 14, no. 12, 2775, June 2022.
- [7] X. Shi, T. Wang, S. Lu, K. Chen, D. He and Z. Xu, "Evaluation of China's forest carbon sink service value," *Environ. Sci. Pollut. Res. Int.*, vol. 29, pp. 44668-44677, Jan. 2022.
- [8] S. Taugourdeau, A. Diedhiou, C. Fassinou, M. Bossoukpe, O. Diatta, A. N'Goran, A. Auderbert, O. Ndiaye, A. A. Diouf, T. Tagesson, R. Fensholt and E. Faye, "Estimating herbaceous aboveground biomass in Sahelian rangelands using structure from motion data collected on the ground and by UAV," *Ecol. Evol.*, vol. 12, no. 5, e88677, May 2022.
- [9] F. Cheng, G. Ou, M. Wang and C. Liu, "Remote sensing estimation of forest carbon stock based on machine learning algorithms," *Forests*, vol. 15, no. 4, 681, April 2024.
- [10] S. Sharma, S. Dhal, T. Rout and B. S. Acharya, "Drones and machine learning for estimating forest carbon storage," *Carbon res.*, vol. 1, no. 21, pp. 1-4, October 2022.
- [11] P. Intarasaksit and J. Chaiyarat, "Creating a green university in Thailand: A case study of Srinakharinwirot University," *Indian J. Public Health Res. Dev.*, vol. 12, no. 4, pp. 291-298, July 2021.
- [12] P. Aregarot, K. Kubaha and S. Chiarakorn, "A study of sustainability concepts for developing green university in Thailand," *Sustainability*, vol. 16, no. 7, 2892, Mar. 2024.
- [13] Chiang Rai Rajabhat University, "Strategic Development Plan 5 year period (2022-2026)," 2023, Personal Administration Division [Online]. Available: <https://personnel.crru.ac.th/page%20/psn/download/detail/545>
- [14] Chiang Rai Rajabhat University, "Plant genetic conservation project under the royal initiative of her Royal Highness Princess Maha Chakri Sirindhorn," 2023, RSPG-CRRU. [Online]. Available: <http://www.rspg.crru.ac.th/>
- [15] D. W. Gibbons and R. D. Gregory, in *Ecological census techniques*, W. J. Sutherland, Ed. Cambridge: Cambridge University Press, 2006, pp. 190-192.
- [16] U. Chanlabout and B. Nahok, "Forest structure and carbon stock of Suan Phueng Nature Education Park in Ratchaburi Province, Western Thailand," *Biodiversitas*, vol. 23, no. 8, pp. 4314-4321, Aug. 2022.
- [17] Dronebox, "ODM Quality Report," 2023, Mapedia. [Online]. Available: <https://dronebox.io/>
- [18] H. Ogawa, K. Yoda, K. Ogino and T. Kira, (1965), "Comparative ecological studies on three main type of forest vegetation in Thailand II. Plant Biomass," *Nature and Life in Southeast Asia*, vol. 4, pp. 49-80, 1965.
- [19] H. K. Gibbs, S. Brown, J. O. Niles and J. A. Foley, "Monitoring and estimating tropical forest carbon stocks: making REDD a reality," *Environ. Res. Lett.* vol. 2, no. 4, 045023, Dec. 2007.
- [20] R Core Team, "R: A language and environment for statistical computing," 2023, R Foundation for Statistical Computing. [Online]. Available: <https://www.R-project.org/>
- [21] H. Ngernin, "Quantification of carbon sequestration of teak areas with UAV (Unmanned Aerial Vehicle) in Wang Nok Aen, Wang Thong District, Phisanulok," 2018, Thesis B.S. in Geography, Naresuan University. [online]. Available: https://ww2.agi.nu.ac.th/nred/Document/is-PDF/2561/geo_2561_08_FullPaper.pdf 2018
- [22] QGIS Development Team, "QGIS Geographic Information System," 2023, Open Source Geospatial Foundation Project. [Online]. Available: <http://www.qgis.org>
- [23] M. Wegmann, B. Leutner and S. Dech, "Remote sensing and GIS for ecologists: Using open source software (Data in the wild)," London: Pelagis Publishing, 2016.
- [24] H-K. Jo, J-Y. Kim and H-M. Park, "Carbon reduction and planning strategies for urban parks in Seoul," *Urban For. Urban Green.*, vol. 41, pp. 48-54. May 2019.
- [25] W. Ni, G. Sun, Y. Pang, Z. Zhang, J. Liu, A. Yang, Y. Wang and D. Zhang, "Mapping three-dimensional structures of forest canopy using UAV stereo imagery: Evaluating impacts of forward overlaps and image resolutions with LiDAR data as reference," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 10, pp. 3578-3589, Oct. 2018.
- [26] J. P. Dandois, M. Olano and E. C. Ellis, "Optimal altitude, overlap, and weather conditions for computer vision UAV estimates of forest structure," *Remote Sens.* vol. 7, no. 10, pp. 13895-13920, Oct. 2015.

- [27] E. Seifert, S. Seifert, H. Vogt, D. Drew, J. van Aardt, A. Kunneke and T. Seifert, "Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images," *Remote Sens.* vol. 11, no. 10, 1252, May 2019.
- [28] H. Ngernin, "Quantification of carbon sequestration of teak areas with UAV (Unmanned Aerial Vehicle) in Wang Nok Aen, Wang Thong District, Phitsanulok," 2018, Undergraduate Thesis. [Online]. Available: https://ww2.agi.nu.ac.th/nred/Document/is-PDF/2561/geo_2561_08_FullPaper.pdf.
- [29] P. Suknak, N. Petchthoong, A. Phasada, K. Wangyao and T. Laphitchayangkul, "The study of flight elevation and speed of UAV photogrammetric accuracy for mapping," 2020, Vol. 25 (2020): Proceeding: The 25th National Convention on Civil Engineering. [Online]. Available: <https://conference.thaince.org/index.php/ncce25/article/view/196>

Characterization of Biochar Produced from Sawdust and Polypropylene Plastic Waste Composite via Slow Pyrolysis

1st Aitsara Chanthakhot*Field of Civil Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

65010352004@msu.ac.th

2nd Poramed Aungthitipan*Field of Civil Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

65010352001@msu.ac.th

3rd Pornmongkol Tansomros*Field of Civil Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

65010352002@msu.ac.th

4th Pattaranun Maskhunthod*Field of Civil Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

65010352007@msu.ac.th

5th Athicha Janthakot*Field of Civil Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

65010352003@msu.ac.th

6th Sukanya Hongthong*Faculty of Engineering and Industrial Technology**Chaiyaphum Rajabhat University*

Chaiyaphum, Thailand

sukanya.ho@cpru.ac.th

7th Torpong Kreethachat*School of energy and environment**University of Phayao*

Phayao, Thailand

torpong.envi@gmail.com

8th kowit Suwannahong*Department of Environmental Health**Faculty of Public Health, Burapha University*

Chonburi, Thailand

kowit.su@gobuu.ac.th

9th,* Surachai Wongcharee*Field of Environmental Engineering**Faculty of Engineering, Mahasarakham University*

Mahasarakham, Thailand

surachai.w@msu.ac.th

10th Saksit Imman*School of energy and environment**University of Phayao*

Phayao, Thailand

saksit.im@up.ac.th

I. INTRODUCTION

Abstract— The combination of plastic waste and sawdust can be used to create a composite material, specifically biochar, which leverages the properties of both components. The formulation and manufacturing process depend on the desired characteristics and application of the biochar composite. Various ratios and techniques can be optimized to enhance the properties of the composite, and additional additives or modifiers can also be incorporated to achieve specific results. In this context, composite sawdust and biochar have been developed to study the structure of charcoal with the goal of improving wastewater treatment using waste materials as precursors via slow pyrolysis process that heats biomass in an oxygen-limited environment at lower temperatures (300-500°C) and over longer periods.

Through chemical and physical analysis, it was observed that the biochar derived from the sawdust and plastic waste composite (SP biochar) consists of approximately 80-90% carbon. The physical structure of SP biochar is smooth and honeycomb-like, with a porous surface and a specific surface area of around 0.6656 m²/g. The isotherm analysis classifies it as a Type III isotherm. Several functional groups were identified in the SP biochar, including hydroxyl groups (-OH, 1990-2140 cm⁻¹), carboxyl groups (-COOH, 1040-1050 cm⁻¹), phenolic groups (-OH attached to an aromatic ring, 1566-1650 cm⁻¹), and ether (-O) groups. These functional groups contribute to the stability of the biochar's structure and influence its hydrophobic properties.

In conclusion, SP biochar demonstrates potential as an effective sorbent for water and wastewater treatment, making use of waste materials for a valuable environmental application.

Keywords— Sawdust, Plastic waste, Biochar, Slow pyrolysis

Sawdust waste refers to the leftover sawdust generated as a byproduct in industries or activities involving wood processing [1]. Proper management of sawdust waste is important to minimize its environmental impact and explore its potential beneficial uses. Sawdust can often be recycled or reused in various ways, including being utilized as a raw material in the production of particleboard, oriented strand board (OSB), or other composite wood products [2]. Additionally, it can be processed into wood pellets or briquettes for use as a fuel source in biomass boilers, stoves, or power plants. Sawdust waste can also be repurposed for small-scale uses, such as in baskets for gardening or insulation in certain applications. If sawdust waste cannot be recycled or repurposed, it should be properly disposed of in accordance with local waste management regulations to avoid environmental harm, as improper disposal may contribute to pollution and negatively impact ecosystems [3]. The appropriate management and utilization of sawdust waste depend on factors such as industry, local regulations, and the volume of waste generated [4].

Plastic waste refers to discarded plastic materials that are no longer useful. This type of waste poses significant environmental challenges due to its durability and low biodegradability [5]. Plastics are widely used in various industries, and the waste generated includes packaging materials, appliances, furniture, and electronic devices. If not properly managed, plastic waste accumulates in landfills or enters water bodies and natural environments, causing harm

to wildlife and ecosystems. Plastic waste is particularly concerning because it can take hundreds of years to decompose, leading to persistent pollution and microplastic contamination [6]. Proper management practices are essential to mitigate the impacts of plastic waste, and this often involves reducing plastic consumption, promoting recycling, and supporting the development of sustainable alternatives such as biodegradable materials [7]. Addressing plastic waste requires collaboration between individuals, businesses, governments, and the recycling industry [8]. Biochar is a carbon-rich material produced through pyrolysis, which involves heating biomass, such as wood chips or agricultural residues, in the absence of oxygen [9]. This process, typically conducted at temperatures between 300 and 700 degrees Celsius, converts biomass into biochar, while also producing byproducts like bio-oil and syngas. Biochar has a porous structure, making it an effective soil amendment that improves soil health by enhancing water retention, increasing nutrient availability, and improving soil structure [10]. It also supports beneficial microorganisms, contributing to better soil aeration. Biochar's environmental benefits extend beyond soil health; it can also contribute to carbon sequestration and reduce the need for synthetic fertilizers [11]. However, sustainability is key, and the biomass feedstock used for biochar production should be responsibly sourced, considering factors like land use and biodiversity [12].

Slow pyrolysis is a thermochemical process that heats biomass in an oxygen-limited environment at lower temperatures (300–500°C) and over longer periods [12]. This method prioritizes the production of biochar over bio-oil, unlike fast pyrolysis. In slow pyrolysis, the controlled environment prevents complete combustion, leading to the formation of a carbon-rich biochar that retains much of the structure of the original biomass [14]. While the process takes longer and may be less energy-efficient compared to fast pyrolysis, it results in a higher-quality biochar with stable carbon properties [15]. Biochar produced from slow pyrolysis has applications in agriculture, horticulture, land reclamation, and carbon capture technologies. Additionally, it can be used in wastewater treatment, where biochar's adsorption properties help remove contaminants from water [16, 17, 18].

This research focuses on the introduction of sawdust and plastic waste into the slow pyrolysis process to produce biochar for use in water and wastewater treatment. Physical and chemical analyses of the biochar are essential for designing appropriate engineering processes for use in water treatment plants. In summary, this study explores the integration of sawdust and polypropylene plastic waste to produce biochar through slow pyrolysis, which aligns well with sustainability goals. Using plastic waste, especially polypropylene, in biochar production reduces plastic pollution, a major global environmental challenge. Polypropylene, like many plastics, is persistent in the environment and resists natural decomposition, often leading to ecological harm. By converting this waste into biochar, the study demonstrates an innovative approach to upcycling plastic waste, potentially contributing to reduced landfill use and decreased pollution in natural ecosystems.

II. METHODOLOGY

A. Preparation of SP Biochar (Sawdust and plastic waste biochar composites)

Raw materials, including sawdust and plastic waste (non-moldable material for recycling, specifically polypropylene (PP)), were sourced from a factory in Chaiyapoom Province, Thailand, as shown in Fig. 1. The size of both raw materials was approximately 0.5 cm to 2.0 cm. The sawdust and plastic waste were dried at 105°C for 24 hours. Subsequently, a mixture with a ratio of 4 g of sawdust to 1 g of plastic waste, totaling about 1 kg, was placed in a modified reactor (OD=60 cm, H=1 m) under an anaerobic environment using a slow pyrolysis process with LPG fuel. The temperature for the pyrolysis was increased at a rate of 12°C per minute and then maintained at 300°C, 400°C, and 500°C for 1 hour for each run. After 1 hour of slow pyrolysis, the samples were allowed to cool down to room temperature. Once cooled, the prepared mixed biochar, as shown in Fig. 1, was milled using a marble mortar to obtain powdered biochar (Fig. 1) and stored for analysis, including FTIR, SEM-EDX, and BET [19].

B. Specifications Fourier Transform Infrared Spectroscopy (FTIR)

Fourier Transform Infrared Spectroscopy (FTIR) is a widely used analytical technique that provides information about the chemical composition and molecular structure of a sample. FTIR spectroscopy is based on the interaction between infrared (IR) radiation and the sample [20]. The sample is exposed to a broad range of IR wavelengths, and the resulting absorption or transmission of the radiation is measured. The obtained spectrum represents the unique molecular fingerprint of the sample [21]. FTIR spectroscopy is a versatile technique for chemical analysis and characterization [22]. Its non-destructive nature, high sensitivity, and broad application range make it a valuable tool in research, quality control, and material analysis.

C. Scanning Electron Microscopy-Energy Dispersive X-ray spectroscopy (SEM-EDX)

Scanning Electron Microscopy (SEM): SEM is an imaging technique that uses a focused electron beam to scan the surface of a sample. The interaction between the electrons and the sample generates various signals, including secondary electrons and backscattered electrons, which can be detected to form high-resolution images of the sample's surface [23]. SEM provides detailed information about the topography, morphology, and surface features of the sample.

Energy Dispersive X-ray Spectroscopy (EDX): EDX is an analytical technique that detects characteristic X-rays emitted by the sample when it is bombarded with electrons [24]. These X-rays are produced due to the interaction between the incident electrons and the atoms in the sample. EDX measures the energy and intensity of the emitted X-rays, allowing for the identification and quantification of the elements present in the sample. EDX provides information about the elemental composition and distribution within the sample [25].

SEM-EDX is a versatile and powerful technique that combines imaging and elemental analysis capabilities. It provides valuable information about the surface morphology, topography, and elemental composition of a wide range of samples, making it an essential tool in materials characterization, research, and quality control.

D. Porosity analysis and surface area analysis

The Brunauer Emmett Teller (BET) theory is a widely used method for determining the surface area of porous materials, especially solids [26]. It provides a quantitative measure of the specific surface area based on the physical adsorption of gas molecules onto the material's surface. Gas Adsorption Isotherms: BET analysis relies on measuring the gas adsorption isotherms, which describe the amount of gas adsorbed by the material as a function of the gas pressure at a constant temperature [27]. The isotherms are typically obtained using techniques such as nitrogen adsorption/desorption measurements using a device called a gas sorption analyzer [28]. It has become a standard method for characterizing the porosity and specific surface area of a wide range of materials. Please correct the grammar.

III. RESULTS AND DISCUSSION

A. Element analysis

Fig. 3 presents the elemental compositions of SP biochar produced at 300°C, 400°C, and 500°C. It can be observed that the carbon content of SP biochar increased with rising temperatures. The SP biochar produced at 500°C exhibited the highest carbon content, at approximately 87%, which is similar to the carbon content of SP biochar produced at 400°C (86.80%). In contrast, the SP biochar produced at 300°C had the lowest carbon content compared to the other temperatures. This indicates that SP biochar primarily consists of carbon as its main constituent, comprising 80-90%, while other elements are present in smaller amounts, contributing to its porosity and surface area.

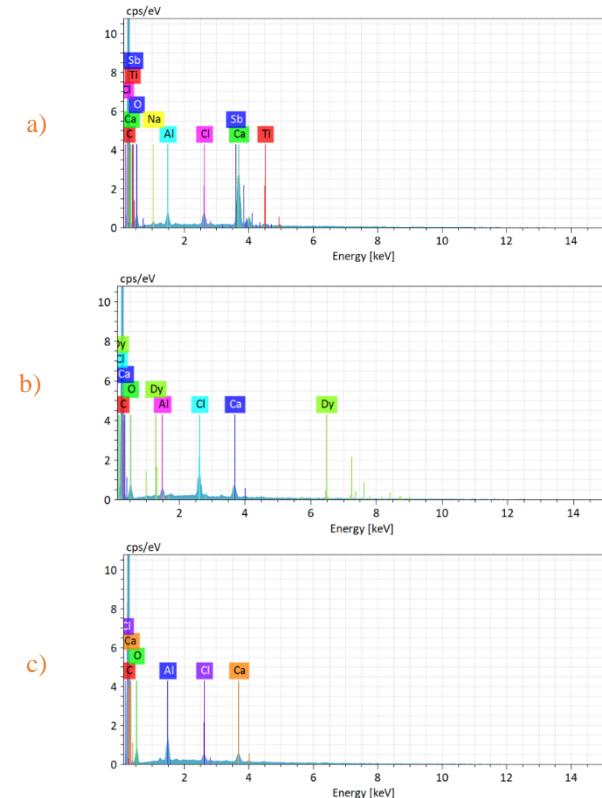


Fig. 3. Elemental compositions of SP biochar: a) SP biochar 300°C, b) SP biochar 400°C and c) SP biochar 500°C

B. SEM

SEM-EDX was used to provide valuable information about the surface morphology, topography, and elemental composition of various SP biochar samples. Fig. 4 a) shows the SEM image of SP biochar at 300°C. It can be seen that after pyrolysis for 1 hour, the structure of SP biochar at 300°C exhibited a porous distribution with a plastic cover on the surface. In contrast, SP biochar at 400°C displayed a clear porosity (Fig. 4 b). However, at 500°C, the materials subjected to high temperatures showed a breakdown in porosity or a decrease in porosity, as illustrated in Fig. 4 c. The selection of raw materials and the control of activation parameters, such as temperature, time, and method, are crucial in determining the surface structures of activated materials. These factors help tailor the pore size distribution and specific surface area to meet specific requirements for various applications, such as adsorption, catalysis, and energy storage.

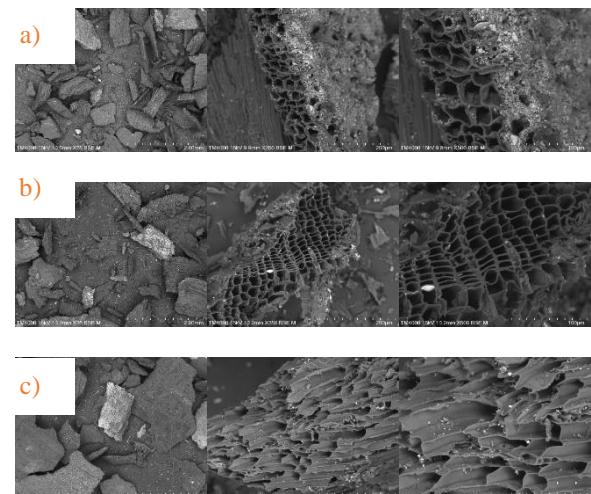


Fig. 4. SEM-EDX for SP biochar: a) SP biochar 300°C, b) SP biochar 400°C and c) SP biochar 500°C

C. Porosity, pore volume and pore size distribution

Table I presents the specific surface area, pore size, and total pore volume of the SP biochar obtained. It can be concluded that the SP biochar produced at 500°C had the highest specific surface area, approximately 0.6656 m²/g, followed by that produced at 400°C (0.3462 m²/g) and 300°C (0.3020 m²/g), respectively. In other words, the structure changed with increasing temperature, along with variations in porosity and time. The total pore volume and average pore diameter also followed the same trend as the specific surface area. This indicates that the properties of SP biochar can be modified using different manufacturing techniques to suit its intended applications.

TABLE I. SPECIFIC SURFACE AREA, PORE SIZE AND TOTAL PORE VOLUME OF SP BIOCHAR

Temperature	SP biochar 300°C	SP biochar 400°C	SP biochar 500°C
Surface Area (m ² /g)	0.3020	0.3462	0.6656
Total pore volume (m ³ /g)	0.001818	0.004851	0.004006
Average pore diameter (nm)	24.075	56.0534	24.075

D. SP biochar N_2 gas adsorption isotherm

In the BET analysis, a known quantity of an adsorbate gas, typically nitrogen (N_2), is introduced to the sample at different relative pressures. The gas molecules are adsorbed onto the surface of the material, and the amount of gas adsorbed is measured. By plotting the quantity of adsorbed gas versus the relative pressure, a specific type of isotherm known as a BET isotherm is obtained. The BET isotherm can be extrapolated to determine the quantity of gas adsorbed at a relative pressure close to zero. This value represents the monolayer adsorption capacity. By knowing the cross-sectional area of the adsorbate gas molecule, the specific surface area of the material can be calculated using the BET equation. The specific surface area calculated using the BET method is often expressed in square meters per gram (m^2/g). It provides information about the external and internal surface area of the material and can be used to compare different materials or evaluate the effect of various treatments or modifications on the surface properties.

The BET equation also provides information about the specific surface area of the adsorbent. Figures 5a, 5b, and 5c show the adsorption isotherm of WP biochar at 300°C, 400°C, and 500°C for 1 hour of pyrolysis time. The results indicate that the WP biochar isotherm is classified as Type III, which represents physical adsorption onto the surface. The characteristic observed in the Type III adsorption isotherm occurs when the formation of monolayers, bilayers, trilayers, and other layers takes place simultaneously, resulting in an almost exponential increase in the amount of adsorption. Type III is an isotherm without a pivot point and is shaped like a concave curve. This type is rare and typically occurs with weak adsorption. It is primarily observed in non-porous solids and micro-porous solids, where the attraction between the adsorbent and the weak sorbent is minimal. Adsorption layers primarily form only at low relative pressures, but as multilayer adsorption occurs, the attraction between the adsorbed layers increases, enhancing absorbance at high relative pressures.

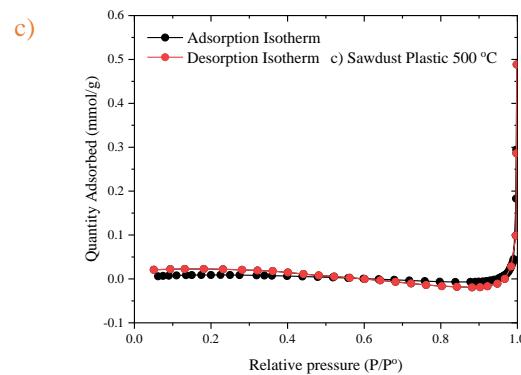
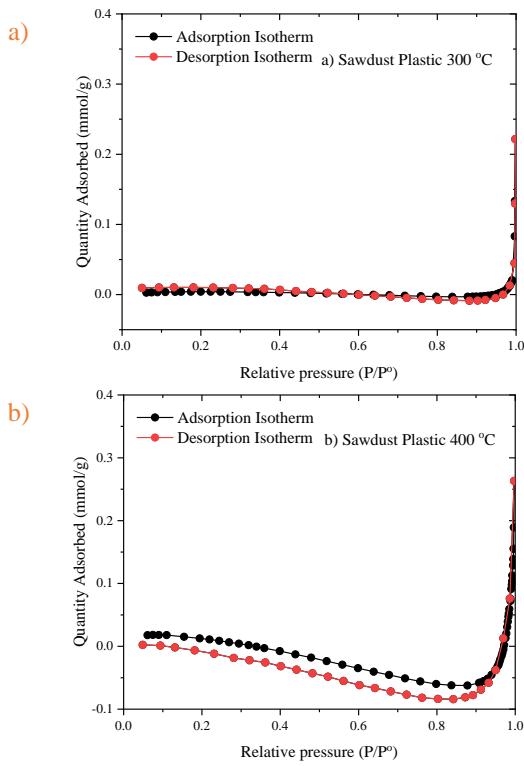
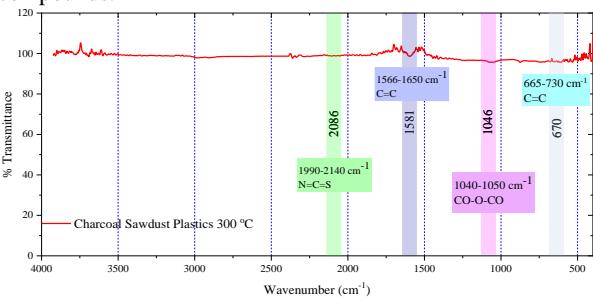


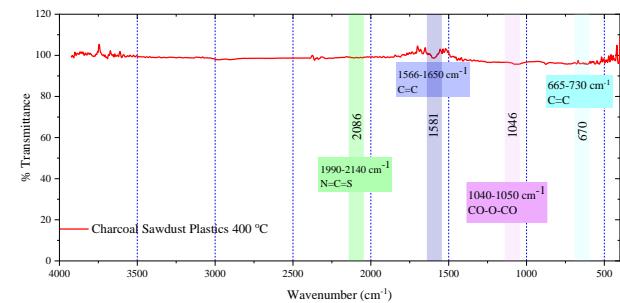
Fig. 5. Isotherm of SP biochar: a) SP biochar 300 °C, b) SP biochar 400°C and c) SP biochar 500°C

E. FTIR analysis-Functional groups study of SP biochar

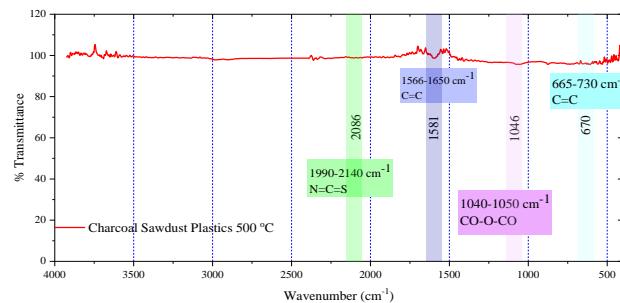
Fourier Transform Infrared (FTIR) spectroscopy (Fig. 6) was used to analyze the functional groups of SP biochar. The SP biochar possesses various functional groups on its surface, which can influence its chemical reactivity and interactions with other substances. The functional groups found in SP biochar include hydroxyl groups (-OH, 1990-2140 cm⁻¹), carboxyl groups (-COOH, 1040-1050 cm⁻¹), phenolic groups (-OH attached to an aromatic ring, 1566-1650 cm⁻¹), and ether groups (-O). These functional groups contribute to the stability of the biochar structure and can influence the biochar's hydrophobicity and interactions with organic compounds.



a)



b)



c)

Fig. 6. FTIR SP biochar: a) SP biochar 300°C, b) SP biochar 400°C and c) SP biochar 500°C

It can be concluded that the study tested biochar at 300°C, 400°C, and 500°C, with findings indicating that higher temperatures increase carbon content and porosity, which positively impact adsorption capacity. Specifically, biochar produced at 500°C had the highest surface area and a balanced pore size, making it ideal for adsorptive applications in water treatment. However, 400°C also yielded a highly effective structure for water treatment, combining high porosity with structural stability. Including recommendations for optimal pyrolysis temperature could provide practical guidance for producing biochar tailored to water purification applications.

IV. CONCLUSION

The findings of this study, which investigated slow pyrolysis at temperatures of 300°C, 400°C, and 500°C, indicate that the biochar derived from plastic sawdust is predominantly carbon rich. Qualitative characterization of the elemental composition of SP biochar at these temperatures reveals a significant increase in carbon content with rising pyrolysis temperatures. Specifically, the SP biochar produced at 500°C exhibited the highest carbon content, approximately 87%, closely followed by the sample at 400°C, which contained about 86.80%. In contrast, the SP biochar obtained at 300°C demonstrated the lowest carbon content among the tested temperatures. This suggests that SP biochar primarily comprises carbon, constituting 80-90% of its total composition, while other elements are present in minimal amounts. SEM images reveal how the structure and porosity of biochar vary at different pyrolysis temperatures, with a notable increase in porosity and surface area at 400°C, which optimizes adsorption properties. The BET analysis supports this by showing that higher temperatures lead to increased surface area and pore volume, crucial for adsorbing contaminants from water. The composite's distinct mix of sawdust and plastic offers a high surface area and chemical functionality (e.g., hydroxyl and carboxyl groups), enhancing its water treatment effectiveness compared to conventional biochar. This combination potentially makes the biochar better at adsorbing diverse contaminants, a significant advantage for applications in water purification. Fourier Transform Infrared (FTIR) spectroscopy was employed to analyze the functional groups present in SP biochar. The results indicate that SP biochar contains various functional groups on its surface, which can significantly influence its chemical reactivity and interactions with other substances. The identified functional groups include hydroxyl groups (-OH, 1990-2140 cm⁻¹), carboxyl groups (-COOH, 1040-1050 cm⁻¹), phenolic groups (-OH attached to an aromatic ring, 1566-1650 cm⁻¹), and ether groups (-O). Given the physical and chemical properties of SP biochar, composite materials incorporating biochar present promising applications in water treatment. The porous structure of biochar contributes to its high surface area and exceptional adsorption capacity, making it effective for the removal of contaminants and pollutants from water.

ACKNOWLEDGMENT

The authors would like to thank, acknowledge and express gratitude that this research was financially supported by the Faculty of Engineering, Mahasarakham University in providing an opportunity to pursue this research work. The authors would like to extend their gratitude to the UOE Fundamental Fund (UOE219/2567) at the University of Phayao for granting them access to experimental equipment and enabling them to conduct this remarkable research.

REFERENCES

- [1] Adegoke, K.A., et al., Sawdust-biomass based materials for sequestration of organic and inorganic pollutants and potential for engineering applications. *Current Research in Green and Sustainable Chemistry*, 2022. 5: p. 100274.
- [2] Alvarez-Chavez, B.J., et al., Physical, chemical, thermal and biological pre-treatment technologies in fast pyrolysis to maximize bio-oil quality: A critical review. *Biomass and Bioenergy*, 2019. 128: p. 105333.
- [3] Srivastava, R.K., et al., Valorization of biowastes for clean energy production, environmental depollution and soil fertility. *Journal of environmental management*, 2023. 332: p. 117410.
- [4] Zhang, F., et al., Current technologies for plastic waste treatment: A review. *Journal of Cleaner Production*, 2021. 282: p. 124523.
- [5] Sharuddin, S.D.A., et al., A review on pyrolysis of plastic wastes. *Energy conversion and management*, 2016. 115: p. 308-326.
- [6] Li, W.C., H.F. Tse, and L. Fok, Plastic waste in the marine environment: A review of sources, occurrence and effects. *Science of the total environment*, 2016. 566: p. 333-349.
- [7] Qureshi, M.S., et al., Pyrolysis of plastic waste: Opportunities and challenges. *Journal of Analytical and Applied Pyrolysis*, 2020. 152: p. 104804.
- [8] Wang, J. and S. Wang, Preparation, modification and environmental application of biochar: A review. *Journal of Cleaner Production*, 2019. 227: p. 1002-1022.
- [9] Weber, K. and P. Quicker, Properties of biochar. *Fuel*, 2018. 217: p. 240-261.
- [10] Atkinson, C.J., J.D. Fitzgerald, and N.A. Hipps, Potential mechanisms for achieving agricultural benefits from biochar application to temperate soils: a review. *Plant and soil*, 2010. 337: p. 1-18.
- [11] Delucchi, M.A., Impacts of biofuels on climate change, water use, and land use. *Annals of the New York Academy of Sciences*, 2010. 1195(1): p. 28-45.
- [12] Vieira, F.R., et al., Optimization of slow pyrolysis process parameters using a fixed bed reactor for biochar yield from rice husk. *Biomass and Bioenergy*, 2020. 132: p. 105412.
- [13] Qambrani, N.A., et al., Biochar properties and eco-friendly applications for climate change mitigation, waste management, and wastewater treatment: A review. *Renewable and Sustainable Energy Reviews*, 2017. 79: p. 255-273.
- [14] Kambo, H.S. and A. Dutta, A comparative review of biochar and hydrochar in terms of production, physico-chemical properties and applications. *Renewable and Sustainable Energy Reviews*, 2015. 45: p. 359-378.
- [15] Libra, J.A., et al., Hydrothermal carbonization of biomass residuals: a comparative review of the chemistry, processes and applications of wet and dry pyrolysis. *Biofuels*, 2011. 2(1): p. 71-106.
- [16] Bridgwater, A., Principles and practice of biomass fast pyrolysis processes for liquids. *Journal of analytical and applied pyrolysis*, 1999. 51(1-2): p. 3-22.
- [17] Fryda, L. and R. Visser, Biochar for soil improvement: Evaluation of biochar from gasification and slow pyrolysis. *Agriculture*, 2015. 5(4): p. 1076-1115.
- [18] Nanda, S., et al., The progressive routes for carbon capture and sequestration. *Energy Science & Engineering*, 2016. 4(2): p. 99-122.
- [19] Khan, S.H., B. Pathak, and M. Fulekar, A study on the influence of metal (Fe, Bi, and Ag) doping on structural, optical, and antimicrobial activity of ZnO nanostructures. *Advanced Composites and Hybrid Materials*, 2020. 3: p. 551-569.
- [20] Chen, Y., et al., Applications of micro-fourier transform infrared spectroscopy (FTIR) in the geological sciences—a review.

- International journal of molecular sciences, 2015. 16(12): p. 30223-30250.
- [21] Dutta, A., Fourier transform infrared spectroscopy. Spectroscopic methods for nanomaterials characterization, 2017: p. 73-93.
- [22] Tucureanu, V., A. Matei, and A.M. Avram, FTIR spectroscopy for carbon family study. Critical reviews in analytical chemistry, 2016. 46(6): p. 502-520.
- [23] hou, W., et al., Fundamentals of scanning electron microscopy (SEM). Scanning Microscopy for Nanotechnology: Techniques and Applications, 2007: p. 1-40.
- [24] Abd Mutualib, M., et al., Scanning electron microscopy (SEM) and energy-dispersive X-ray (EDX) spectroscopy, in Membrane characterization. 2017, Elsevier. p. 161-179.
- [25] Kourkoumelis, N., I. Balatsoukas, and M. Tzaphlidou, Ca/P concentration ratio at different sites of normal and osteoporotic rabbit bones evaluated by Auger and energy dispersive X-ray spectroscopy. Journal of biological physics, 2012. 38: p. 279-291.
- [26] Naderi, M., Surface area: brunauer–emmett–teller (BET), in Progress in filtration and separation. 2015, Elsevier. p. 585-608.
- [27] Thommes, M., et al., Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report). Pure and applied chemistry, 2015. 87(9-10): p. 1051-1069.
- [28] Sinha, P., et al., Surface area determination of porous materials using the Brunauer–Emmett–Teller (BET) method: limitations and improvements. The Journal of Physical Chemistry C, 2019. 123(33): p. 20195-20209.

Evaluating the Travel Carbon Footprint Associated with Using Digital Platforms for Accessing Social Welfare Services

1st Kornganok Thongta

School of energy and environment
University of Phayao
Phayao, Thailand
64140050@up.ac.th

2nd Chaturawit Chanchai

School of energy and environment
University of Phayao
Phayao, Thailand
64140083@up.ac.th

3rd Sukhuma Chitapornpan

School of energy and environment
University of Phayao
Phayao, Thailand
sukhuma.ch@up.ac.th

4th Somanat Somprasert

School of energy and environment
University of Phayao
Phayao, Thailand
somanat.so@up.ac.th

5th Supreeda Homklin

School of energy and environment
University of Phayao
Phayao, Thailand
supreeda.ho@up.ac.th

6th Chaiwat Photong

School of energy and environment
University of Phayao
Phayao, Thailand
chaiwat.ph@up.ac.th

7th Pimsiri Suwanpat

School of energy and environment
University of Phayao
Phayao, Thailand
pimsiri.su@up.ac.th

8th Chatthip Chaichakan

Faculty of political science and public administration
Chiangmai University
Chiangmai, Thailand
chatthip.ch@cmu.ac.th

9th Wanawan Pragot

School of energy and environment
University of Phayao
Phayao, Thailand
wanawan.pr@up.ac.th, 0000-0001-5610-6240

Abstract— This study evaluates the travel carbon footprint associated with using digital platforms for accessing social welfare services. The research focuses on the 'Me Sitthi' mobile application, designed to enhance access to welfare rights and benefits while potentially reducing travel-related emissions. A survey of 352 participants across ten sub-districts in Lamphun (Muang Jie, Ban Klang, and Umong), Chiang Mai (Mae Kue), and Phayao (Sri Thoi, Wiang, Kue Wiang, and Ngim) provinces was conducted, coupled with GIS-based distance calculations and regression analysis. Results show an average travel carbon footprint of 1.72 ± 2.02 kgCO₂eq/capita across the studied areas, with significant variations between locations (ranging from 0.32 to 7.10 kgCO₂eq/capita). Regression analysis revealed that travel distance had the strongest positive correlation (0.7565) with carbon footprint, while age showed a negative correlation (-0.1527). The use of digital platforms resulted in an average cost reduction of 31.2 baht per person in transport expenses. Motorcycles were the most common mode of transportation (62% of respondents), aligning with national trends. This research provides insights into the potential environmental benefits of digital platforms in social welfare service delivery, highlighting the complex interplay of factors influencing travel-related carbon emissions.

Keywords— travel carbon footprint, social welfare services, digital platform, regression analysis

I. INTRODUCTION (HEADING I)

Digital platforms are technology-driven infrastructures that facilitate interactions, exchanges, and co-creation of value among users, organizations, or different groups. These platforms enable connections between various participants, such as buyers and sellers, service providers and consumers, or developers and users, often through digital tools, services, and data [1]. Most digital platforms act as private regulators of their ecosystems. They establish the rules through which their various users (be they individuals or organizations) interact, decide what behaviours to encourage or discourage on the platform, and choose how to enforce them [2]. The digital platforms in this research was designed and applied in

the creation of a mobile application called 'Me Sitthi'. This application is expected to serve as a digital platform where welfare service providers and users can interact, allowing users to better access and understand their welfare rights and benefits [3].

The travel carbon footprint is based on ISO14064 standards [4]. Equally important impacts related to the use of vehicles promote the consumption of fuel and energy. Road transport is a significant contributor to total greenhouse gas emissions [5] from engine combustion process [6]. Mobile vehicles emit direct greenhouse gas (GHG) emissions of carbon dioxide (CO₂), methane (CH₄) and nitrous oxide (N₂O) from the combustion process of various fuel types, as well as several other pollutants such as carbon monoxide (CO), Non-methane Volatile Organic Compounds (NMVOCs), sulphur dioxide (SO₂), particulate matter (PM) and oxides of nitrate (NOx), which cause or contribute to local or regional air pollution. The GHG calculation is based on the amount and type of fuel combusted and its carbon content [7].

The primary fuels used in Thailand including gasoline, diesel, compressed natural gas: Gasoline (CNG), liquefied petroleum gas (LPG), and Ethanol [8]. Gasoline is the main fuel used in personal cars and motorcycles. It is widely used and comes in several types, such as Gasoline 91, Gasoline 95, and E85 (gasoline mixed with 85% ethanol). Diesel is primarily used in trucks, vehicles requiring high torque, and agricultural machinery. There are various types of diesel, such as Diesel B7 (diesel mixed with 7% biodiesel) and Diesel B10 (diesel mixed with 10% biodiesel). Compressed Natural Gas (CNG) is used in certain types of vehicles, such as taxis and buses. It offers benefits in terms of reducing emissions and operational costs. Liquefied Petroleum Gas (LPG) is used in some vehicles and also for domestic and industrial purposes, such as cooking. Ethanol is a biofuel produced from plants like corn or sugarcane and is used as a blend in gasoline to improve

combustion and reduce emissions. It is used in various blends, such as E20 (20% ethanol) and E85 (85% ethanol).

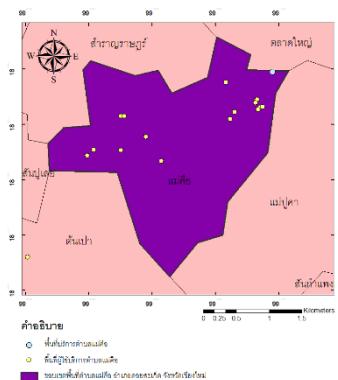
The most popular vehicles in Thailand, as recorded by the Department of Land Transport, updated for 2024[9], are as follows: 22,722,999 motorcycles (52.55%), 12,060,533 of sedan (Not more than 7 passengers) (27.88%), 7,077,302 van & pick up (16.37%), and 1,378,496 public buses (3.19%). These different vehicle categories have varying fuel combustion rates. Vehicles with higher fuel combustion rates typically have larger engines designed for heavy-duty applications [7]. Greenhouse gas emissions from the freight transportation sector are a significant contributor to climate change, pollution, and negative health impacts because of the common use of heavy-duty diesel vehicles (HDVs) [10]. The heavy-duty vehicles account for approximately 23% of total greenhouse gas (GHG) emissions.

This research aims to evaluate the travel carbon footprint associated with using digital platforms for accessing social welfare services called “Mesithi”. The study areas were selected based on the willingness of the participants to be the case study for using this social welfare services. Data was collected through a survey conducted from October to December 2023, with a total of 352 participants across ten sub-districts in Lamphun (Muang Jie, Ban Klang, and Umong), Chiang Mai (Mae Kue), and Phayao (Sri Thoi, Wiang, Kue Wiang, and Ngim) provinces in the North of Thailand. The collected data will be converted into kilograms of carbon dioxide equivalent. Regression analysis is used to find the correlation between the travel carbon footprint and other variables such gender, education, age, income, and type of vehicle. The details of the study will be explained in the next section.

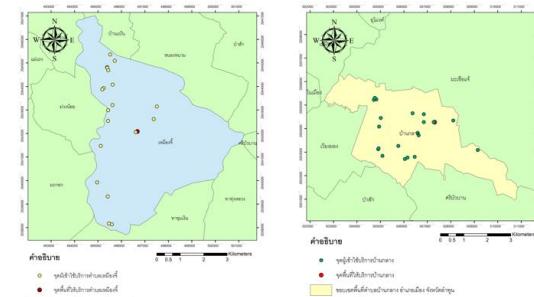
II. METHODOLOGY

A. Survey and questionnaire

Data was collected through a survey conducted from October to December 2023, with a total of 352 participants across ten sub-districts in Lamphun (Muang Jie, Ban Klang, and Umong), Chiang Mai (Mae Kue), and Phayao (Sri Thoi, Wiang, Kue Wiang, and Ngim) provinces in the North of Thailand as shown in Fig. 1. The dots in the map represented the sampling location.

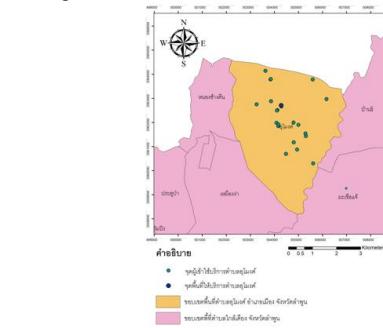


Mae Kue, Chiang Mai
a. A study area in Chiang Mai

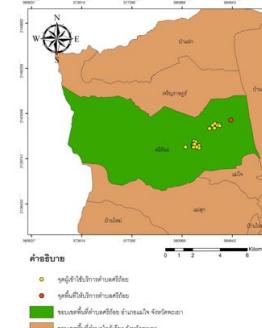


Muang Jie, Lamphun

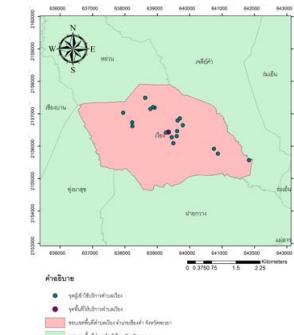
Ban Klang, Lamphun



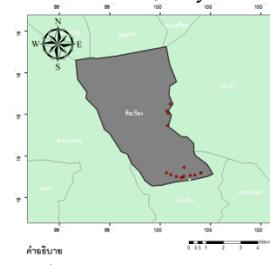
Umong, Lamphun
b. Three study areas in Lamphun



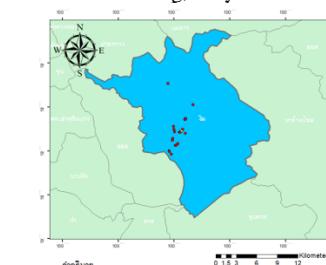
Sri Thoi, Phayao



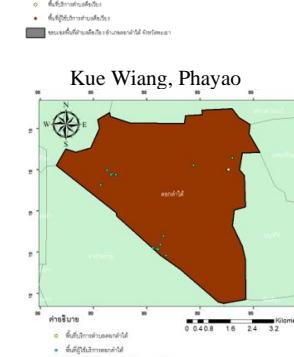
Wiang, Phayao



Kue Wiang, Phayao



Ngim, Phayao



Dok Khamtai, Phayao

Ban Pin, Phayao
c. Six Study area in Phayao

Fig. 1. Ten study areas in Lamphun, Chiang Mai, and Phayao provinces in the North of Thailand

The questionnaire includes various factors expected to impact greenhouse gas emissions: distance, age, gender, type of vehicle, education, income, and type of fuel used. Before using the questionnaire, we sought approval from the Human Research Ethics Committee in the Humanities and Social Sciences for the study titled "The Development of a Personalized Social Welfare System and Mechanism in Local Communities, Leaving No One Behind, Using a Digital Platform and the Game 'Mesitthi'" Contract number: N74A660496.

B. Travel carbon footprint calculations

The geographic information system (GIS) (ArcGIS, Geoprocessing tool) was used to map local agencies' locations and the respondents' house locations. Then, distance between the respondents' home and their local services are determined to estimate greenhouse gas emissions reductions from reduced travel when using the platform. The assumption was that using the platform would reduce travel from respondents' homes to local agencies across all 9 service areas. Travel distances are then used to analyze fuel consumption based on the fuel efficiency rate according to the type of vehicle and type of fuel. The fuel consumption is calculated using equation (1) and the consumption rate by type of vehicle adapted from Thailan greenhouse gas organization [11] in Table I.

$$\text{Fuel or energy consumption} = \frac{\text{Distance}}{\text{Consumption rate}} \quad (1)$$

Where *Distance* refers to the travel from respondents' homes to local agencies across all 9 service areas; *Consumption Rate* is the coefficient quantifying the amount of Fuel or energy used per unit of distance [11].

TABLE I. CONSUMPTION RATE BY TYPE OF VEHICLE

Vehicle Type	Consumption Rate	Unit
Average Sedan (All Sizes)	14.763	km/L
Average Pickup Truck	6.369	km/L
Natural Gas Vehicle (NGV)	11.905	km/kg
Liquefied Petroleum Gas Vehicle (LPG)	8.929	km/L
Average 4-Stroke Motorcycle (All Sizes)	37.640	km/L
Average 2-Stroke Motorcycle (All Sizes)	32.435	km/L
Electric vehicle (EV)	0.5986	kWh/km

The travel carbon footprint is calculated using the following formula.

$$CO_2 \text{ Emission} = \text{Activity data} \times \text{Emission factor} \quad (2)$$

Where *Activity data* refers to the data related to activities that lead to the emission of greenhouse gases; *Emission factor* is the coefficient quantifying the amount of greenhouse gases emitted per unit of activity provided in Table II [11].

TABLE II. EMISSION FACTORS

Fuel Type	Emission Factor (kg CO ₂ /unit)	Unit
Gasoline	2.2394	liter
Diesel	2.7406	liter
NGV	2.2609	kilogram
LPG	1.7306	liter

C. Regression analysis

Regression analysis was used to determine the optimal combination of variables influencing travel carbon footprints using the two-way Pearson correlation coefficient, with a 95% confidence interval of ± 0.1046 for a sample size of 352 ($N=352$). The variables considered included vehicle type, distance traveled, gender, age, income, and education

III. RESULTS AND DISCUSSION

A. Survey results

Survey was used to collect the data of 352 participants across ten sub-districts in Lamphun (Muang Jie, Ban Klang, and Umong), Chiang Mai (Mae Kue), and Phayao (Sri Thoi, Wiang, Kue Wiang, and Ngim) provinces in the North of Thailand.

TABLE III. EMISSION FACTORS

Variable		N	%
<i>Vehicle</i>	Sedan	48	14%
	Pickup car	73	21%
	Electric vehicle	3	1%
	Motorcycle	250	63%
	Taxi	0	0%
	Public bus	5	1%
<i>Gender</i>	Sedan	48	14%
	Male	100	29%
	Female	248	71%
<i>Age</i>	Unspecified	3	0%
<i>Age</i>	0-15	1	0%
	15-18	8	2%
	19-26	0	0%
	26-40	52	15%
	41-60	189	54%
	61 years old up	98	28%
<i>Income</i>	0-15	1	0%
	0-30,000	124	36%
	30,000-60,000	101	29%
	60,000-150,000	84	24%
<i>Education level</i>	0-30,000	124	36%
	Primary school	89	17%
	Lower secondary	52	31%
	Upper secondary	95	17%
	Diploma	33	3%
	Bachelor's degree	69	29%
	Postgraduate	3	3%

The results showed that motorcycles are the most common vehicle, used by 62% of respondents (215 users), followed by pickup trucks at 21% (73 users), and sedans in third place at 14% (48 users). This trend aligned with the 2024 statistics from the Department of Land Transport [9], which also indicated that motorcycles were the most popular vehicles in Thailand. There were more female respondents over male for around 40%. The data showed a significant representation of older adults among the participants. The predominant age group was 41-60 years old (54%, 189 respondents), followed by those 61 years and older (28%, 98 respondents). Together, these two age groups accounted for 82% of the participants. It is noted that the old age group were available during survey period in workday while others were rarely present. Income levels were evenly distributed, with a slight lean toward lower incomes. Educational levels were diverse, with notable representation from both secondary education and bachelor's degree holders.

B. Travel carbon footprint

Fig.1 provides a comparative view of carbon footprints across different locations, highlighting areas that may need more attention in terms of reducing travel-related carbon emissions.

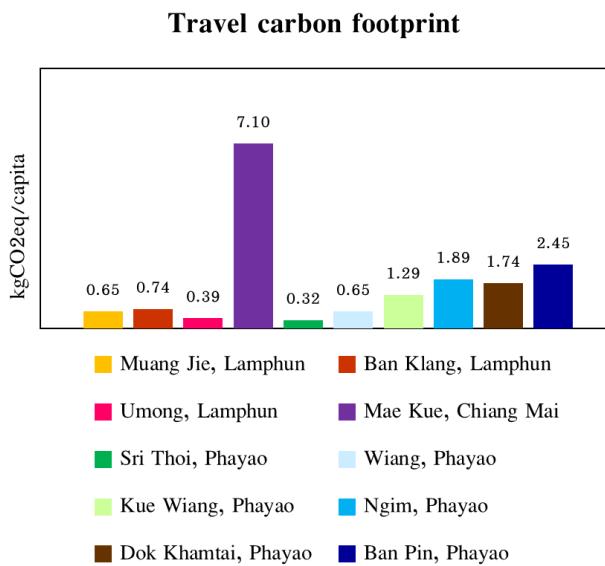


Fig. 2. Travel carbon footprint across ten study areas in Lamphun, Chiang Mai, and Phayao provinces in the North of Thailand

Mae Kue, Chiang Mai had the highest carbon footprint at 7.10 kgCO₂eq/capita, followed by Ban Pin, Phayao average 2.45 kgCO₂eq/capita. The others had significantly lower footprints, ranging from 0.32 to 1.89 kgCO₂eq/capita with Sri Thoi, Phayao reporting the lowest at 0.32 kgCO₂eq/capita. The average travel carbon footprint across sampling sites was 1.72 ± 2.02 kgCO₂eq/capita. There's considerable variation between locations, which could be due to factors like transportation methods, distance travel, or local infrastructure. The diesel engine vehicle like pickup car CO₂ emission per 1 km distance is higher than the motor cycle around 56% and the sedan around 80%, respectively [11]. The distance is directly effect to the travel carbon footprint due it is the variable in the calculation equation. The local infrastructure like efficient public transportation systems (buses, trains, subways) reduce the need for private car usage. By offering frequent and reliable service, these networks encourage people to opt for public transit over driving, thereby reducing overall emissions.

Calculating the traveling savings used the data of the average annual fuel prices for Phayao Province in 2023 [12]; 33.08 baht per liter for diesel, 36.87 baht per liter for gasoline, and 4.42 baht per kWh for EV charger. Using digital platforms resulted in an average cost reduction of 31.2 baht per person in transport expenses from the reduce of the travel between the respondents' home and their local services. This demonstrated that digital platforms like 'Me Sithi' have the potential to reduce travel-related carbon emissions and the travel cost by minimizing the need for physical trips to local service agencies.

C. Regression analysis

The two-way Pearson correlation coefficient, with a 95% confidence interval of ± 0.1046 for a sample size of 352 was

used. The variables considered include vehicle type, distance, gender, age, income, and education level.

TABLE IV. REGRESSION ANALYSIS OF THE TRAVEL CARBON FOOTPRINT CONSIDERED VEHICLE TYPE, DISTANCE, GENDER, AGE, INCOME, AND EDUCATION LEVEL VARIABLES

	GHG	Distance	Vehicle	Gender	Age	Income	Education
GHG	1						
Distance	0.7565	1					
Vehicle	0.0452	0.0533	1				
Gender	0.0164	0.0758	0.1423	1			
Age	-0.1527	-0.1379	0.0051	-0.0452	1		
Income	-0.0380	-0.0410	-0.2045	-0.1085	-0.1730	1	
Education	-0.0002	0.0310	-0.0799	0.0865	-0.3364	0.3278	1

The significant factor from regression analysis using two-way Pearson correlation coefficient, with a 95% confidence interval of ± 0.1046 for a sample size of 352 ($N=352$) were represented in bold letter in Table IV. The results indicated that the travel distance (0.7565) had the strongest positive correlation with travel carbon footprint. It suggests that as the distance increases, the carbon footprint also increases significantly. Age (-0.1527) showed a negative correlation, implying that the travel carbon footprint tends to decrease for older participants. The reason behind this could be possible to the old age tend to transport in short distance and might be possible to behavioral or lifestyle differences between age groups. Other factors such as vehicle type, gender, income, and education level had no impact on travel carbon footprint. These results suggested that efforts to reduce travel carbon footprints should primarily focus on minimizing travel distances. The negative correlations with age and income indicated that older and higher-income individuals tend to have lower travel carbon emissions, possibly due to different travel patterns or vehicle choices. The analysis also revealed complex interrelationships between socioeconomic factors that influence travel behavior and resulting emissions.

IV. CONCLUSION

This research provides valuable insights into the potential environmental benefits of digital platforms in the context of social welfare service delivery, while also highlighting the complex interplay of factors that influence travel-related carbon emissions. The conclusions for this study are provided as follows.

1. Motorcycles were the most common mode of transportation among respondents (62%), aligning with national trends in Thailand. This had implications for carbon footprint calculations and potential emission reduction strategies.
2. The study sample was predominantly older adults (82% aged 41 and above) and female (71%).
3. The average travel carbon footprint was 1.72 ± 2.02 kgCO₂eq/capita ranging from 0.32 to 7.10 kgCO₂eq/capita.
4. Using digital platforms resulted in an average cost reduction of 31.2 baht per person in transport expenses.
5. Regression analysis revealed that travel distance had the strongest positive correlation (0.7565) with carbon footprint, while age showed a negative correlation (-0.1527). Other

factors such as vehicle type, gender, income, and education level had no impact.

6. The study demonstrated that digital platforms like 'Me Sitthi' have the potential to reduce travel-related carbon emissions by minimizing the need for physical trips to local service agencies.

The study explained the carbon footprint reduction and cost saving from using welfare digital platforms and this could promote the future welfare policy or digital platform development.

ACKNOWLEDGMENT

This research was funded by the National Research Council of Thailand under the project title "First Steps of Collaboration Towards Innovation for Social Welfare" with the project number of HREC-UP-HSS 2.2/120/66.

REFERENCES

- [1] G. Hu *et al.*, "The impact of urban digital platforms on entrepreneurial activity: Evidence from China," *Journal of Innovation and Knowledge*, vol. 9, no. 1, Jan. 2024, doi: 10.1016/j.jik.2024.100468.
- [2] A. Gawer, "Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age," *Innovation: Organization and Management*, vol. 24, no. 1, pp. 110–124, 2022, doi: 10.1080/14479338.2021.1965888.
- [3] C. Chaichakan, "Using agile-scrum methodology in co-creating digital innovation for social welfare services in Thailand: lessons and challenges," 2023.
- [4] ISO 14064-1:2018, "ISO 14064-1:2018 - Greenhouse gases — Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals," Edition 2, 2018. Accessed: Jul. 14, 2024. [Online]. Available: <https://www.iso.org/standard/66453.html>
- [5] M. Clairotte *et al.*, "Exhaust emission factors of greenhouse gases (GHGs) from European road vehicles," *Environ Sci Eur*, vol. 32, no. 1, Dec. 2020, doi: 10.1186/s12302-020-00407-5.
- [6] P. Wiśniowski, A. Borucka, M. Menes, and A. Świderski, "Evaluation of the CO₂ emission from motor vehicles in the context of sustainable transport development," *Combustion Engines*, Jul. 2021, doi: 10.19206/ce-140405.
- [7] IPCC, "2006 IPCC Guidelines for National Greenhouse Gas Inventories Chapter 3: Mobile Combustion," 2006.
- [8] Ministry of Energy of Thailand, "Department of Energy Business." Accessed: Aug. 30, 2024. [Online]. Available: <https://www.doeb.go.th/en>
- [9] "Department of Land Transport." Accessed: Aug. 30, 2024. [Online]. Available: <https://www.dlt.go.th/en>
- [10] C. Cunanan, M. K. Tran, Y. Lee, S. Kwok, V. Leung, and M. Fowler, "A Review of Heavy-Duty Vehicle Powertrain Technologies: Diesel Engine Vehicles, Battery Electric Vehicles, and Hydrogen Fuel Cell Electric Vehicles," 2021. doi: 10.3390/cleantechnol3020028.
- [11] TGO, "Carbon Label Project." Accessed: Aug. 30, 2024. [Online]. Available: <https://thaicarbonlabel.tgo.or.th/>
- [12] "PTT Oil and Retail Business Public Company Limited." Accessed: Aug. 30, 2024. [Online]. Available: https://www.pttor.com/en/oil_price

Satellite subsidence and sea-level data assimilation for mean high-water line assessment in the Bangkok area : Preliminary result

Femke Vossepoel

Dept. of Geoscience and Engineering
Delft University of Technology
Delft, Netherlands
F.C.Vossepoel@tudelft.nl

Marc Naeije

Astrodynamics and Space Missions
Delft University of Technology
Delft, Netherlands
m.c.naeije@tudelft.nl

Milan Lazecky

School of Earth and Environment
University of Leeds
Leeds, UK
M.Lazecky@leeds.ac.uk

Jenny Soonthornrangsang

Dept. of Geoscience and Engineering
Delft University of Technology
Delft, Netherlands
J.T.Soothornrangsang@tudelft.nl

Mark Bakker

Dept. of Geoscience and Engineering
Delft University of Technology
Delft, Netherlands
Mark.Bakker@tudelft.nl

Aimée Slangen

Dept. of Estuarine and Delta Systems
Royal Netherlands Institute for Sea
Research
Yerseke, Netherlands
Aimee.Slangen@nioz.nl

*Sommart Niemnil

Council for Scientific and
Technological Association of Thailand
Bangkok, Thailand
sniemnil@gmail.com

Wim J.F. Simons

Astrodynamic and Space Missions
Delft University of Technology
Delft, Netherlands
w.j.f.simons@tudelft.nl

Andy Hooper

School of Earth and Environment
University of Leeds
Leeds, UK
A.hooper@leeds.ac.uk

Anuphao Aobpaet

Dept. of Civil Engineering
Kasetsart University
Bangkok, Thailand
anuphao_a@hotmail.com

Abstract— The Bangkok, Thailand is undergone subsidence from land subsidence, tectonic motion and sea level rise. INSAR result from 2015-2022 show that Bangkok subsided from 10-30 mm/yr. GPS results show that Bangkok is still undergoing post-seismic tectonic deformations from the 2004 Mw 9.2 Andaman-Sumatra earthquake, with a significant change in vertical motion: from quasi-linear uplift at 6 ± 1 mm/yr to (temporary) non-linear subsidence rates up to 5 mm/yr from 1994-2022. Based on satellite altimetry data, the absolute sea-level rise for Gulf of Thailand is estimated at 3-5 mm/yr. Ground water pumping also play an important role on land subsidence. This paper presents results from an ongoing study.

Keywords— Land Subsidence, Tectonic motion, Sea Level Change

Land subsidence combined with sea-level rise threatens the inhabitants of the coastal area of Bangkok, Thailand; the main contributor to subsidence is groundwater extraction for drinking water, industry and food. Subsidence in this region is being monitored through in-situ levelling, Global Positioning System (GPS) observations and Interferometric Synthetic Aperture Radar (InSAR) analyses. An extensive network of tide gauges is combined with satellite altimetry to monitor sea-level changes; groundwater levels are measured in observation wells. Forecasting subsidence is complicated by uncertainties in model state and parameters, secondary effects

such as consolidation of clay, and the lack of in-situ data coverage.

HISTORICAL BACKGROUND

River deltas provide a living place to more than 500 million people, and many of these deltas are sinking[1-3]. This subsidence increases their susceptibility to flood hazard² which is further amplified by global sea-level rise[4]. The societal impact of flood hazard became very clear with the 2011 Thailand flooding of the Chao Phraya, the main river flowing through Bangkok into the Gulf of Thailand. Human-induced changes in land use and groundwater extraction have been identified as likely contributors to this flood event[5]. By studying how groundwater extraction affects subsidence and how this combined with the sea-level changes affects the mean high water line, we expect to better understand the influence of groundwater extraction and expected sea-level change on flood hazard. The dramatic event of the 2011 flood, which led to a total of 815 deaths and a total economic damage and loss estimated to be 45.7 billion US dollars[6], underpinned the importance of understanding flood hazard and monitoring sea-level and subsidence, especially in densely populated areas such as Bangkok. This equally holds for Ho Chi Minh City, Jakarta, Guangzhou, and other coastal cities. Land subsidence can occur because of natural processes such as sediment loading, compaction, and tectonics. It can be augmented as a

result of human activities such as groundwater extraction [3,7]. Historically, studies of land subsidence rely on point measurements, but with the availability of interferometric satellite aperture radar images (InSAR, see Fig. 1),

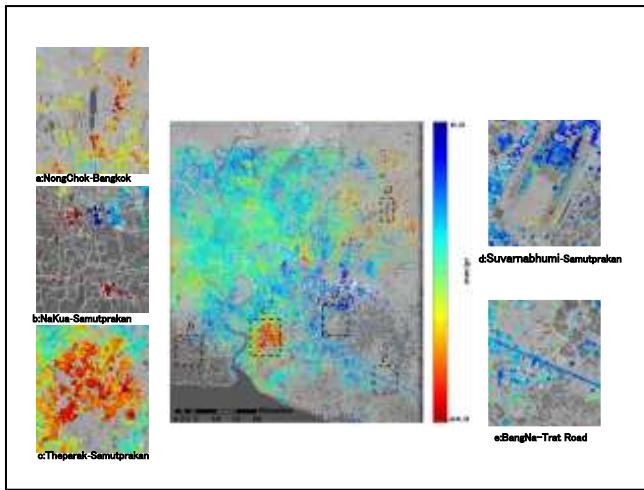


Fig. 1. the most recent published InSAR results (Aobpaet et al, 2013) for the Bangkok area using RADARSAT SAR images in between 2005 and 2010. It shows land-subsidence in Bangkok both at the east and west Chao Phraya river banks, with the maxima occurring mostly in new urban areas and coastal area

the combination of scattered GPS observations, borehole extensometer observations and levelling data with InSAR can provide meaningful analyses e.g.,[7]. Surge-driven water levels are likely to increase as sea level rises [8]. World tide-gauge records show that global mean sea level (GMSL) rose 3.2 ± 0.4 mm/yr between 1993 and 2010, with a strong impact on coastal zones [4]. For the Gulf of Thailand, the 1993-2009 sea level rising rates are in the range of 3 to 5.5 mm/yr [10]. By the year 2100, GMSL are expected to reach levels of 0.43 m to 0.84 m higher than today [9]. Any land subsidence in Bangkok, currently sinking at a rate of 2-3 cm/yr [11,12] will locally increase this number further. The high-water or high-tide line, the line that marks the limit of the rise of the medium tides of the sea between the spring and neap tides, can be used to locate the coastal boundary [13]. This demarcation will also increase with an increase in sea-level. To analyse the mean high-water lines of a large set of coastal areas, a global digital elevation model (DEM) can be used in combination with sea level estimates combined with the local one-year return level heights (RL1) from the Global Tides and Surge Reanalysis [8,14]. Even with deep cuts to carbon emissions, by 2100, areas in Thailand now home to 17 % of its people may face high-water lines higher than land, not considering episodic flooding events [14]. Since the 1970s, the main contributor to subsidence in the Bangkok area is considered to be the pumping of groundwater for drinking water, industry and food [15]. In addition to these human causes, crustal deformation can also cause vertical land motion¹⁵. While Global Isostatic Adjustment (GIA) amounts from +0.50 to +0.65 mm per year depending on location, much larger vertical displacements are associated with the 9.1 Mw 2004 Sumatra-Andaman earthquake [10,16]. In addition to the vertical land motions, coastal erosion occurs as a result of the complex interaction of human-induced shoreline changes and natural processes [2,17]. In theory, changes in

atmospheric variability and increased rainfall can also increase the likelihood of a flooding event, but in the case of the 2011 Thailand flood, the changes in hydrography of the river Chao Phraya, as well as human-induced changes in land use, specifically groundwater extraction, have been identified as more likely contributors [5].

Satellite and In-situ observations of land subsidence

Land subsidence, ranging in amplitude from several mm to several cm annually, can be accurately measured with geodetic techniques that make use of Global Navigation Satellite Systems (GNSS), such as the Global Position System (GPS). In the period 1994-1998, land motion in Thailand has been monitored consistently with GPS stations of the Geodynamics of South and Southeast Asia (GEODYSSSEA) project [18,19]. In particular the CHON GPS point of this campaign is of relevance for subsidence in the Bangkok area. It has been re-occupied at least annually since 2001 by the Royal Thai Survey Department (RTSD) after the GEODYSSSEA ended in 1998. After 1999 the BNKK station provided continuous GNNS data, and the same in 2005, 2008 and 2011 respectively, the RTSD, CUSV, SPKN stations followed to provide continuous monitoring in Bangkok. Additional GPS stations in the region provide complimentary data [16,20]. In addition to these observations, annual levelling campaigns have been conducted since 1978 by the Royal Thai Survey Department (RTSD) [12,21]. Since 2009, these levelling campaigns have been combined with Interferometric Synthetic Aperture Radar (InSAR) analyses for a spatial coverage of subsidence.

Satellite and in-situ observations of local sea level change

Satellite radar altimeters have observed the height of the sea surface of the global oceans continuously since the 1980's. It has led to studies of a wide spectrum of scientific issues ranging from ocean circulation to marine plate tectonics. In particular studies have focused on planetary-scale waves, interannual climate variations such as El Niño, oceanic boundary currents, coastal circulation, tides, surface waves, mean sea level variations, marine gravity and geoid, bathymetric features, lithospheric processes, etc. The processing of altimeter observations has advanced tremendously leading to cm accuracy, and since global coverage has increased with the advent of new altimeter missions, the technique has become very valuable for the study of coastal areas, such as Bangkok. Tide gauges observe sea level and fluctuations thereof. As most gauges are installed on land, vertical land motion will need to be corrected for when using tide-gauge observations for sea-level analysis [22] (see Figure 1b). In addition to in-situ observations of sea level, satellite-borne altimetry has been instrumental in the global analysis of sea level variations since the early 1990s. The data set has well-distributed precise sea-level measurements, unaffected by vertical land motion. Based on a decade of multi-satellite altimetry short-term variability in sea-level is not always geographically uniform [23]. In order to evaluate local sea-level changes, the combination of altimetry and tide-gauge observations can thus prove to be valuable. A 1940-2004 analysis [10] of tide gauges combined with satellite altimetry for the Gulf of Thailand suggests significantly higher trends of sea level change in the northern part of the Gulf of Thailand at Sattahip (5.0 ± 1.3 mm/yr) and Ko Sichang (4.5 ± 1.3 mm/yr) than the global rate from tide-gauge data of 1.8 ± 0.3 mm/yr over a comparable observation period (1950-

2000) [24]. Two other tide-gauges stations are located in the vicinity of Bangkok: the Pom Phrachunlachomkla and Bangkok Bar. These tidal stations do not have a foundation on the bedrock, and trend lines of the position of these gauges strongly suggest an influence of the settling of the sand layers on which the supporting piles stand. While the influence of vertical land motion of these tide gauges caused them to be excluded from the earlier sea-level analysis of the Gulf of Thailand [10], their observations may be valuable for monitoring ground motion. Observational analyses of sea-level change can be complimented with dynamic models of the ocean circulation. While regional models for tides [25] and waves [26] exist, for long-term sea-level rise, projections are generally taken from ocean climate models [27].

Groundwater Studies

Subsidence is often caused by groundwater pumping. Accurate subsidence simulation is influenced by limited data availability, especially input forcings/drivers and calibration data. In this study, an ensemble-based data-assimilation method is used to improve the estimates of land subsidence in Bangkok, Thailand, which is simulated by a linked data-driven and physics-based modeling approach. This approach is applied with limited groundwater and subsidence observations and only an estimate of basin-wide pumping. The assimilation of sparse and noisy head and subsidence observation data into the Ensemble Smoother with Multiple Data Assimilation (ESMDA) algorithm results in improved estimates and uncertainty quantification of groundwater heads, localized groundwater pumping and parameters.

PRELIMINARY RESULT

A. Land subsidence

InSAR result also shows relative land-subsidence takes places at rates up to 3 cm/yr in the past 20 years (Fig.2).

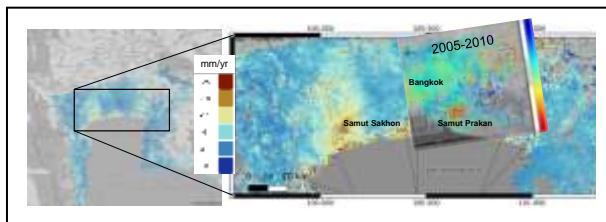


Fig. 2. show subsidence rate in (mm/yr) from 2015-2020 InSAR (Sentinel-1) time series analysis (preliminary / low resolution / based on line-of-sight velocities – vertical component still needs to be fully extracted) so subsidence of Bangkok seems to be stabilizing, details from high resolution analysis.

B. Land motion

GNSS result shows absolute subsidence rates (below 20 m) up to 5 mm/yr in the past 25 years (Fig.3).

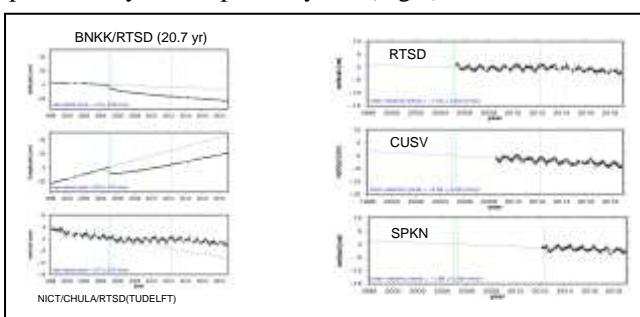


Fig. 3. Station position time series (weekly averaged / 1998-2021 / IGS14) and linear velocity estimates for the stations RTSD, CUSV and SPKN (Bangkok) in Thailand. Shown are the velocities in N (latitude), E (longitude) and U (vertical) direction. The time span is 3.3 to 4.1 yr, and position/velocity uncertainties are 1-sigma (68% confidence level) values. For station SPKN a position jump (green line) was estimated in March 2019 as the antenna pillar slightly moved (~1 cm to the southwest) after being hit upon at the pier.

C. Sea level Changes

Tide Gauge near Bangkok show relative sea-level rise rates up to 15 mm/yr in the past 60 years (combined effect of sea-level rise plus land-subsidence). Based on satellite altimetry data, the absolute sea-level rise in the Gulf of Thailand is estimated at 3-5 mm/yr (Fig.4).

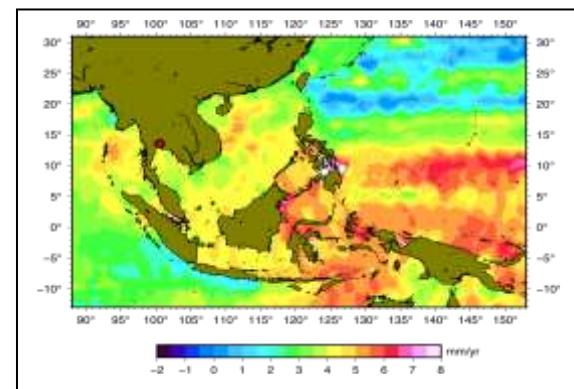


Fig.4. Absolute Sea level Change in Southeast Asia

D. Groundwater studies

ESMDA constrains subsidence estimates

The compaction of each layer is based on groundwater drawdown and layer properties. Land subsidence at a well nest is the sum of the compaction of all model layers. ESMDA is used to estimate three multiplier terms for each property and well nest.

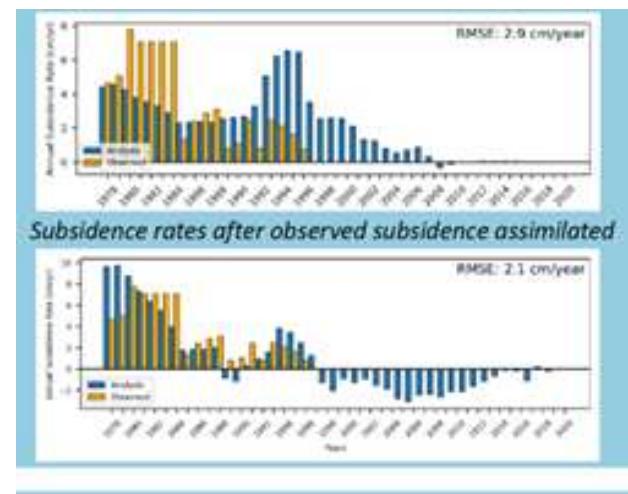


Fig.5. Simulated (blue) annual rates of land subsidence at well nest BKK006 compared to benchmark levelling (orange)

CONCLUSION

Absolute sea level rise due to climate change is undeniable and SALT shows Bangkok is presently facing a rate of up to 5 mm/yr in the Gulf of Thailand. Whether this is an even bigger threat is depending on what the land is doing:

- GNSS result shows absolute subsidence rates (below 20 m) up to 5 mm/yr in the past 25 years.

- InSAR result shows relative land-subsidence takes places at rates up to 3 cm/yr in the past 20 years.

- TG near Bangkok show relative sea-level rise rates up to 15 mm/yr in the past 60 years (combined effect of sea-level rise plus land-subsidence)

- Subsidence : RMSE of well nest BKK006 decrease from 2.9 to 2.1 cm/yr after implementing ESMDA

Having multiple observation techniques at the same time, these can be used to calibrate the different measurement systems. Tectonic subsidence due to the 2004 Mw 9.2 mega-thrust earthquake may increase the relative sea level in the coming decades by up 10 cm for the Bangkok area. Groundwater pumping and sea-level rise increases flood risk in this area. If no appropriate measures are taken, the Greater Bangkok Region faces progressive coastal land loss and increased/permanent flooding during the next decades. Mostly the ground soil level is affected as buildings with deeper foundations subside at lower rates.

ACKNOWLEDGMENT

We would like to extend our special thanks to the Port Authority of Thailand (PAT) and the Royal Thai Naval Dockyard for their permission and assistance in installing the SPKN GNSS station at their Pom Prachunlachomklao tide-gauge, the Hydrographic Department of the Royal Thai Navy (HDRTN) for their permission and assistance in installing the KTPH GNSS station at their KTN tide-gauge, the Thai Marine Department (TMD) for providing their TG data in South Thailand, the Department of Groundwater Resources (DGR) for providing their water well data and to the Royal Thai Survey Department (RTSD) for providing additional GPS data from the GEODYSSSEA PHUK point. Part of the GPS data used in this report was collected as part of our cooperation in the EU-ASEAN SEAMERGES (2004-2006) and GEO2TECDI-1/2 projects (2009-2013). We also want to thank the Phuket Marine Biological Centre (PMBC) for their continued excellent hosting of the PMBC GNSS station on their premises and by again providing excellent logistical (boat/staff) and technical support in setting up and inspecting the new KTPH GNSS station. We thank Remko Scharroo (EUMETSAT) and Eric Leuliette (NOAA) for their continued maintenance of the Radar Altimeter Database System.

REFERENCES

- [1] Erkens, G., Bucx, T., Dam, R., Lange, G. & Lambert, J. Sinking coastal cities. *Proc. Int. Assoc. Hydrol. Sci.* 372, 189–198 (2015).
- [2] Ericson, J. P., Vörösmarty, C. J., Dingman, S. L., Ward, L. G. & Meybeck, M. Effective sea-level rise and deltas: Causes of change and human dimension implications. *Glob. Planet. Change* 50, 63–82 (2006).
- [3] Syvitski, J. P. M. et al. Sinking deltas due to human activities. *Nat. Geosci.* 2, 681–686 (2009).
- [4] Nicholls, R. J. & Cazenave, A. Sea-Level Rise and Its Impact on Coastal Zones. *Science* (80-.). 328, 1517 LP – 1520 (2010).
- [5] Peterson, T. C., Stott, P. A. & Herring, S. Explaining Extreme Events of 2011 from a Climate Perspective. *Bull. Am. Meteorol. Soc.* 93, 1041–1067 (2012).
- [6] The World bank. <https://www.worldbank.org/en/news/feature/2011/12/13/world-bank-supports-thailands-post-floods-recovery-effort>
- [7] Higgins, S. A. et al. InSAR measurements of compaction and subsidence in the Ganges-Brahmaputra Delta, Bangladesh. *J. Geophys. Res. Earth Surf.* 119, 1768–1781 (2014).
- [8] Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H. & Ward, P. J. A global reanalysis of storm surges and extreme sea levels. *Nat. Commun.* 7, 11969 (2016).
- [9] Rhein, M. et al. Observations: Ocean. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change <https://eprints.soton.ac.uk/362480/> (2013).
- [10] Trisirisatayawong, I., Naeije, M., Simons, W. & Fenoglio-Marc, L. Sea level change in the Gulf of Thailand from GPS-corrected tide gauge data and multi-satellite altimetry. *Glob. Planet. Change* 76, 137–151 (2011).
- [11] Aobpaet, A., Caro Cuenca, M., Hooper, A. & Trisirisatayawong, I. Land Subsidence Evaluation using InSAR time series analysis in Bangkok Metropolitan Area. (2009).
- [12] Aobpaet, A., Cuenca, M. C., Hooper, A. & Trisirisatayawong, I. InSAR time-series analysis of land subsidence in Bangkok, Thailand. *Int. J. Remote Sens.* 34, 2969–2982 (2013).
- [13] NOAA. http://tidesandcurrents.noaa.gov/datum_options.html
- [14] Kulp, S. A. & Strauss, B. H. New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. *Nat. Commun.* 10, 5752 (2019).
- [15] Phien-wej, N., Giao, P. H. & Nutalaya, P. Land subsidence in Bangkok, Thailand. *Eng. Geol.* 82, 187–201 (2006).
- [16] Vigny, C. et al. Insight into the 2004 Sumatra–Andaman earthquake from GPS measurements in southeast Asia. *Nature* 436, 201–206 (2005).
- [17] Vongvisessomjai, S., Polsi, R., Manotham, C., Srisaengthong, D. & Charulukkana, S. Coastal Erosion in the Gulf of Thailand BT - Sea-Level Rise and Coastal Subsidence: Causes, Consequences, and Strategies. in (eds. Milliman, J. D. & Haq, B. U.) 131–150 (Springer Netherlands, 1996). doi:10.1007/978-94-015-8719-8_7
- [18] Wilson, P. et al. Study provides data on active plate tectonics in Southeast Asia region. *EOS Trans.* 79, 545–549 (1998).
- [19] Simons, W. J. F. et al. Vertical motion of Phuket Island (1994–2018) due to the Sumatra-Andaman mega-thrust earthquake cycle: impact on sea-level and consequences for coral reefs. *Mar. Geol.* 414, 92–102 (2019).
- [20] Simons, W. J. F. et al. A decade of GPS in Southeast Asia: Resolving Sundaland motion and boundaries. *J. Geophys. Res. Solid Earth* 112, (2007).
- [21] Bontenbal, M. Land subsidence in Bangkok -an overview of changes in land subsidence over the last 23 years. (VU Amsterdam, The Netherlands, 2001).
- [22] Wöppelmann, G. & Marcos, M. Vertical land motion as a key to understanding sea level change and variability. *Rev. Geophys.* 54, 64–92 (2016).
- [23] Cazenave, A. & Nerem, R. S. Present-day sea level change: Observations and causes. *Rev. Geophys.* 42, (2004).
- [24] Church, J. A., White, N. J., Coleman, R., Lambeck, K. & Mitrovica, J. X. Estimates of the Regional Distribution of Sea Level Rise over the 1950–2000 Period. *J. Clim.* 17, 2609–2625 (2004).
- [25] Cui, X., Fang, G. & Wu, D. Tidal resonance in the Gulf of Thailand. *Ocean Sci.* 15, 321–331 (2019).
- [26] Wannawong, W. & Ekkawatpanit, C. Tropical Cyclone Wind-Wave, Storm Surge and Current in Meteorological Prediction. in Natural Disasters(ed. Cheval, S.) (InTechOpen, 2012). doi:10.5772/32252.
- [27] Ritphring, S., Somphong, C., Udo, K. & Kazama, S. Projections of Future Beach Loss due to Sea Level Rise for Sandy Beaches along Thailand's Coastlines. *J. Coast. Res.* 85, 541–545 (2018).

Geospatial Evaluation of Remote Working Hubs: A Graph Convolutional Network Model

Junke Xu

CeADAR, Ireland's Centre for AI
University College Dublin
Dublin, Ireland
junke.xu@ucd.ie

Yilin Li

CeADAR, Ireland's Centre for AI
University College Dublin
Dublin, Ireland
yilin.li@ucd.ie

Alireza Dehghani

CeADAR, Ireland's Centre for AI
University College Dublin
Dublin, Ireland
alireza.dehghani@ucd.ie

Abstract—Remote Working Hubs (RWHs) are emerging as an alternative to conventional office spaces, offering shared environments that foster collaboration and reduce the isolation, often associated with working from home. To better understand the role of RWHs in urban planning and their impact on surrounding amenities, this study introduces a novel scoring model that leverages Graph Convolutional Networks (GCNs) for geospatial analysis. The new model evaluates the connectivity and influence of various urban amenities on RWHs, providing a comprehensive assessment framework. A case study in Dublin demonstrates the model's capability to support sustainable urban planning by identifying optimal locations for RWHs based on proximity to essential amenities. The results show that the new model outperforms traditional distance-based methods, offering a more nuanced understanding of the interactions between RWHs and the urban environment. This research provides valuable insights for urban planners and policymakers aiming to enhance the integration of RWHs into smart city infrastructures.

Index Terms—remote working hubs, geospatial analysis, graph convolutional networks, machine learning, visualisation.

I. INTRODUCTION

Due to progress in technology and changes in city life trends, RWH venues have become an alternative to conventional workspaces. Remote work has its benefits like emissions, personal time organisation and improved mental well-being. However, it doesn't completely overcome the challenges of working from home such as solitude and the absence of areas. RWH spaces tackle these issues by providing office setups with networking resources and motivating designs that boost both individual and team efficiency. The close availability of facilities near work hubs makes them appealing choices for workers and adds to their increasing popularity worldwide [1] [2].

In recent years, the number of RWHs has increased significantly, leading to a growing recognition of their role in enhancing social interactions and advancing smart city development. However, there remains a need for a comprehensive analysis of their performance, particularly in terms of flexibility, user satisfaction, and their broader impact on urban ecosystems. Understanding how RWHs interact with surrounding amenities is essential for sustainable urban planning. Traditional methods of evaluating these factors require substantial time and financial investment, making them impractical for large-scale implementation. To address these challenges, machine learning

methods offer a more efficient approach for processing large datasets and evaluating RWH performance [3].

This paper introduces the new scoring model, referred to as the S4 model, which integrates GCNs with a silhouette index to provide a more refined evaluation of RWHs and their surrounding amenities. The S4 model builds upon the previous S3 model by incorporating GCN-predicted relationships between amenities and RWHs, thereby offering a more comprehensive and accurate assessment of hub performance.

II. RELATED WORK

A. Impact of Remote Working

Changes in workplace practices can greatly influence both work-related and non-work-related travel patterns [4]. Employees often coordinate their personal activities, such as shopping, leisure, and healthcare, around their work commutes [5]. As a result, modifications in work arrangements can reduce the separation between work and personal activities, leading to more complex travel behaviours.

Although the effects of remote work on quality of life and transportation have been explored, limited research has examined the attractiveness of telecommuting from remote work centres. Additionally, evidence suggests that working from home has contributed to higher productivity levels [6]. Some companies have also indicated intentions to maintain remote work options for certain employees even after the COVID-19 pandemic [7]. The post-pandemic scenario, however, points to a potentially greater role for telecommuting. RWHs are a relatively new concept in telework, and their impact on quality of life, transportation systems, and environmental sustainability has yet to be fully analysed. With the reopening of cities and the return to traditional office work, combined with ongoing concerns about using public transportation, there may be a rise in car commuting for work purposes. This trend could conflict with current carbon reduction and sustainability goals. Therefore, it is essential to reconsider the future of teleworking, particularly in the context of RWHs, as a viable alternative to both home-based and office-based work.

B. Models Based on Geospatial Analysis

Geospatial analysis is essential for understanding spatial relationships in urban environments. By combining machine

learning methods like GCNs with geospatial data, researchers have developed advanced models for urban planning and infrastructure management. These models are particularly useful for optimising the placement and performance of RWHs. GCNs have been widely applied to analyse urban infrastructure and spatial relationships. One study developed a site selection model for Singapore using geospatial data and machine learning, focusing on the proximity of transportation hubs and amenities—an approach directly relevant to RWH assessment [8]. GCNs have also been used to predict pollution sources, demonstrating flexibility in modelling spatial distributions, which can help evaluate the environmental impact of RWHs [9].

Spatiotemporal models are key to understanding how urban areas evolve. A review of spatiotemporal graph models highlighted GCNs' ability to capture both spatial and temporal dynamics, making them ideal for long-term urban planning and identifying locations for RWHs [10]. Predictive models of urban functionality, including GCNs, have been used to analyse building footprints and mobility patterns, offering insights into the distribution of amenities and helping optimise RWH placement [11]. Another model predicted metro station vibrancy, focusing on relationships between transportation hubs and urban functions—an approach adaptable for RWHs [12].

Hybrid GCN models combining machine learning techniques have effectively addressed urban challenges. These models predict human activity intensity and identify congestion points, both of which are useful for optimising RWH locations and improving accessibility [13], [14]. Geospatial data also supports environmental monitoring, crucial for sustainable urban planning. A graph convolutional self-organising map method was used to identify pollution sources, offering insights into the environmental sustainability of RWHs [15].

C. Applications of GCN for Geospatial Networks

GCNs have been widely applied to geospatial networks for modelling complex relationships in graph-structured data. These models are effective in fields like urban planning, transportation, and environmental analysis, where spatial dependencies are critical.

One study used GCNs to recognise drainage patterns, developing a dual graph of drainage networks to assess spatial patterns, which proved useful for infrastructure planning [16]. Another study applied GCNs to site selection in Singapore, combining transport and land use data to improve urban planning, demonstrating their utility in smart city development [17]. GCNs have also imputed spatial origin-destination flows, capturing interactions in urban transportation systems and aiding mobility planning [18].

For urban land use classification, GCNs were integrated with Points of Interest (POIs) to extract spatial context, providing insights for urban development [19]. Another study applied spatiotemporal GCNs to forecast multi-scale traffic flows, improving traffic management by capturing both spatial and temporal dependencies [20]. GCNs have also been used for

urban land-use inference, where heterogeneous GCNs modelled spatial interactions in complex geospatial environments, offering a powerful tool for analyzing land-use patterns and informing urban planning decisions [21].

III. METHODOLOGY

A. Dataset and Pre-processing

Data on RWHs and their surrounding amenities in the Dublin region were obtained from two primary sources: the National Hub Network [22] and OpenStreetMap [23]. The National Hub Network dataset provides comprehensive details about each hub, including its location, capacity, ownership, and available services. OpenStreetMap data was leveraged to extract information about nearby urban amenities, such as cafes, restaurants, public transportation stops, and healthcare facilities, which enabled the analysis of their proximity to RWHs. Additionally, the geographical boundaries of Dublin were sourced from Zenodo [24], ensuring the accurate generation of random residential coordinates within the city. These boundaries are essential for ensuring that all simulated individuals are correctly located within the Dublin region.

Before applying the datasets in our model, we conducted a comprehensive data pre-processing procedure to ensure the quality and reliability of the information. Initially, we removed incomplete or duplicate entries from the National Hub Network dataset, discarding any hub lacking essential information such as location coordinates or service availability. This ensured that our analysis was based on consistent and accurate data.

Next, we geocoded the datasets using latitude and longitude to standardise the spatial information. All location data were transformed into the EPSG:4326 coordinate reference system, which is a widely accepted global standard for geospatial data. This ensured that all points were positioned uniformly on the map, allowing for seamless spatial analysis. To integrate the amenities data from OpenStreetMap with the RWH dataset, we calculated the geographical distance between hubs and nearby amenities using the OpenStreetMap API.

The final integrated dataset comprised 74 RWHs and over 1800 unique urban amenities, categorised into 14 different types. This dataset was used as the input for the GCN model to evaluate the relationships between RWHs and their surrounding amenities.

B. Graph Convolutional Network for Prediction

In this GCN model, RWHs and urban amenities are represented as nodes within a graph. The nodes are connected based on geographical proximity, with edge weights determined by the inverse of the distance between them, reflecting the assumption that the farther an amenity is from an RWH, the lesser its impact on the hub. To compute these distances accurately, the OpenStreetMap API is utilised, providing precise distances between each RWH and nearby amenities, ensuring that proximity is appropriately weighted.

TABLE I
LIST OF AMENITIES

No.	Amenity
1	Bank
2	Bar
3	Bicycle rental
4	Building supermarkets
5	Café
6	Child Care Provider
7	Fast food
8	Internet café
9	Parking entrance
10	Parking space
11	Post office
12	Restaurant
13	Shop convenience
14	Shop supermarkets

Each node in the graph is assigned a feature vector to capture the attributes relevant to predicting connections between RWHs and amenities. For the RWH nodes, features include attributes such as hub name, address, Eircode, hub classification, local authority, services provided, ownership, geographical coordinates, operational status, capacity, hub size, and broadband connectivity. For the amenity nodes, attributes such as the amenity's name, geographical coordinates, Eircode, and broader Eircode area are considered, reflecting characteristics that may influence the RWHs in their vicinity.

Table I lists the specific types of amenities used, covering 14 different categories, representing a variety of amenities that may have different impacts on RWH. As a result, 14 separate graphs are constructed, each focused on one type of amenity and its relationship to RWHs. By applying a GCN for link prediction, the likelihood and strength of connections between RWHs and amenities are assessed, providing insights into which types of amenities have the most significant influence on RWHs. This approach quantifies the varying levels of influence different amenities exert on the functionality and attractiveness of RWHs, offering a detailed analysis of how proximity and availability of urban amenities impact RWHs in urban environments.

C. New Scoring Formula

The new scoring model represents an enhancement over the previous S3 model by refining the way RWHs are evaluated in terms of their proximity to and connectivity with surrounding urban amenities [25]. While the S3 model accounts for the influence of various amenities on RWHs, the new model introduces a more balanced and normalised approach to scoring, making it a more comprehensive tool for evaluating hub performance. The new model will be referred to as S4. In the S3 model, the score for each RWH i is calculated as follows:

$$S3_i = \sum_{j=1}^N SI(j) \cdot F^j(i) + \sum_{l=1}^M SI(l) \cdot \text{Normal} \left(\sum_{k=1}^{H_l} \frac{H_l}{d(k,i)} \right) \quad (1)$$

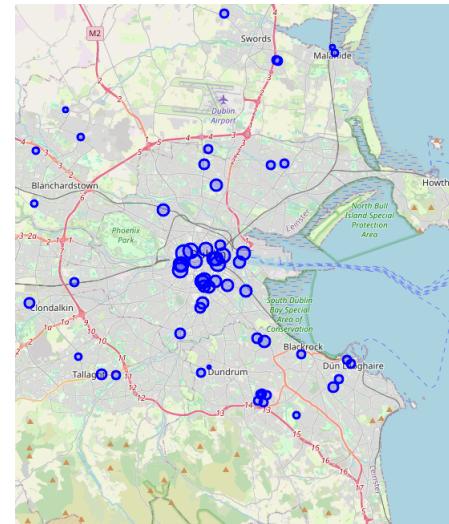


Fig. 1. Visualisation of S4 Score.

Here, the first part of the equation represents the impact of facilities on the RWH score. The Silhouette Index (SI) for each facility j indicates how well these facilities are clustered around the hub, while the feature function measures the quality or utility of each facility for RWH i . This part remains unchanged in both the S3 and S4 models, as it effectively captures the contribution of facilities to the hub's overall performance.

The second part of the S3 formula accounts for the impact of urban amenities and their proximity to the hub. However, the S3 model uses a basic distance-based weighting method, where the influence of each amenity is weighted by the SI and the inverse of its distance from the hub, normalised for comparability. While effective, this approach is limited in its ability to capture the full complexity of interactions between hubs and amenities.

To address this limitation, the S4 model introduces a more sophisticated method for evaluating amenity influence by leveraging GCNs. The second part of the S4 formula replaces the distance-based weighting in S3 with GCN-predicted results, providing a more comprehensive and accurate assessment of how amenities affect RWH performance. The formula for calculating the S4 score is as follows:

$$S4_i = \frac{1}{2} (\text{Normal} \sum_{j=1}^N SI(j)) \cdot F^j(i) + \text{Normal}(\sum_{k=1}^M S_k) \quad (2)$$

The second part is where the S4 model introduces its key improvement. Instead of relying on simple distance-based measures, this term uses a GCN to predict the overall influence of amenities on the hub. The GCN captures the complex spatial relationships between amenities and hubs, modelling not only their proximity but also the structure of the urban environment and the interconnectedness of the amenities themselves. The GCN output S_k is then normalised to ensure comparability across hubs.

TABLE II
DESCRIPTIVE STATISTICS FOR S3 AND S4

Statistic	S3	S4
mean	0.434102	0.600515
std	0.127305	0.178291
min	0.032913	0.167746
max	0.836892	0.968984

By averaging these two parts, the S4 model delivers a balanced evaluation of both the direct facility impact and the broader urban connectivity of each hub. The inclusion of GCN predictions in the second part enables a more nuanced and accurate representation of the role that amenities play in supporting RWH functionality and attractiveness. This makes the S4 model a more powerful tool for decision-makers looking to optimise the placement and development of Remote Working Hubs.

D. Visualisation

After developing the S4 formula, its visualisation on the map shown in Fig.1. provides a visual analysis of RWH across the Dublin area, where each circle represents a hub and the size of the circle reflects the S4 score of the hub. A cluster of RWHs in the city centre features larger circles, indicating higher scores likely due to a rich presence of amenities and superior support facilities in the urban core. In contrast, RWHs located in suburban and rural outskirts are depicted as smaller circles, denoting lower scores possibly stemming from an absence of nearby amenities. This distribution suggests a disparity in remote working infrastructure, with a concentration of higher-rated RWHs in densely populated commercial centres, and highlights development opportunities in outer regions with fewer, lower-scoring hubs.

IV. EVALUATION

A. Statistic Analysis

In comparing our scoring models, we analysed how the S3 and S4 models perform, as shown in Table II. On average, the scores for remote working hubs increased with the S4 model, suggesting a more favourable evaluation of amenities and their connectivity with the introduction of S4. The increase in average score indicates that the S4 model better captures the overall value of amenities. Additionally, the scores from the S4 model showed greater variability, which may reflect its increased sensitivity to the diverse amenity landscapes surrounding each hub.

The analysis reveals significant improvements in the S4 model's scoring system compared to the S3 model. The minimum score has notably risen, indicating a more favourable assessment of hubs with modest amenities. Moreover, the highest score in the S4 model surpasses that of the S3 model, suggesting that hubs with the most amenities are valued more highly. These findings highlight the S4 model's superior ability to discriminate and precisely evaluate the amenity value and connectivity of remote working hubs, reflecting a more comprehensive and nuanced approach to urban analysis.

B. New Evaluate Guideline

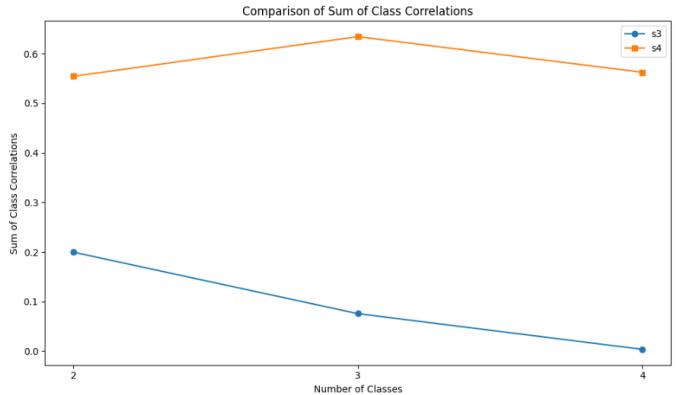


Fig. 2. Evaluation Cumulative Sum of Class Correlations.

Fig.2. presents a detailed comparison between the S3 and S4 scoring models, orange line means S4 and blue line means S3, focusing on the cumulative sum of class correlations across different clusters derived through k-means clustering of RWHs. Hubs were grouped into 2, 3, and 4 clusters based on their S3 and S4 scores to assess how well each model correlates with the amenities within these clusters. The correlations between amenities and the overall scores were then normalised by cluster size to ensure that larger clusters did not skew the results, allowing for a fair comparison of model performance.

The results reveal that the S4 model consistently outperforms the S3 model, exhibiting higher cumulative correlation sums across all cluster groupings. For example, in the case of two clusters, the S4 model demonstrates a significantly higher cumulative correlation sum than the S3 model, indicating that it better captures the relationship between hub scores and the quality of amenities. This trend continues as the number of clusters increases, with the S4 model maintaining stable correlation sums, suggesting that it is robust and effective in capturing the relationship between amenities and hub scores at both broad and granular levels.

In contrast, the S3 model shows a marked decline in cumulative correlation sums as the number of clusters increases. While the S3 model can capture some degree of relationship between amenities and hub scores when hubs are grouped into two clusters, its ability to reflect this relationship diminishes with more detailed classifications, indicating that the S3 model is less robust and precise in its evaluation of hubs.

The analysis further emphasises that normalising the correlations relative to cluster size was essential to eliminating potential biases from clusters of different sizes. This normalisation ensured that the observed differences between the models were due to their performance, rather than external factors like cluster size.

Overall, the S4 model's higher and more stable cumulative correlation sums across clusters indicate its superior capacity to evaluate RWHs based on amenities. This finding has significant implications for urban planners and stakeholders who

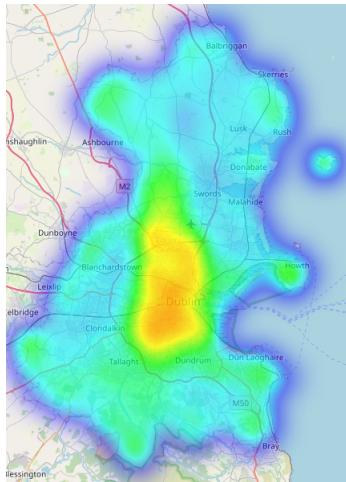


Fig. 3. The Heatmap of Predicted RWH Scores.

rely on such scoring models for making decisions related to infrastructure development, resource allocation, and identifying areas for improvement. The S4 model's robustness ensures that it can be effectively used for evaluations, regardless of how hubs are categorised.

Finally, the total cumulative score, obtained by summing the normalised correlations across all clusters, serves as a clear metric of each model's efficiency. The S4 model's higher total score confirms its greater ability to reflect the quality and impact of amenities in RWHs, underscoring its value for providing a more comprehensive and nuanced assessment of hub quality compared to the S3 model.

V. EXPERIMENT

A. Geospatial Prediction of RWHs Impact on Amenity Scores in Dublin

The study employs a multifaceted approach to visualise the potential influence of new RWHs within the Dublin area, as illustrated in Fig.3. This predictive heatmap integrates geospatial analysis and the K-Nearest Neighbors (KNN) algorithm, we developed predictive models for various amenity scores across the city, trained on a dataset of existing hubs. These models were then applied to a dense grid of geospatial points, randomly generated within the urban confines, as delineated by the Dublin General Area GeoJSON file. The amenity predictions were normalised and subjected to logarithmic scaling to enhance visual discernibility on the heatmap, ensuring that subtle variations in the scores were perceptible. The visualisation was rendered via the Folium library, which facilitated the overlay of these data points on an interactive map, marked by a colour gradient where warmer tones indicate areas of high potential impact from new hubs. This gradient serves as a crucial decision-making tool for urban planners and policymakers, providing a strategic overview of where the establishment of new remote working hubs could yield the most substantial benefits in terms of amenity access and urban balance.

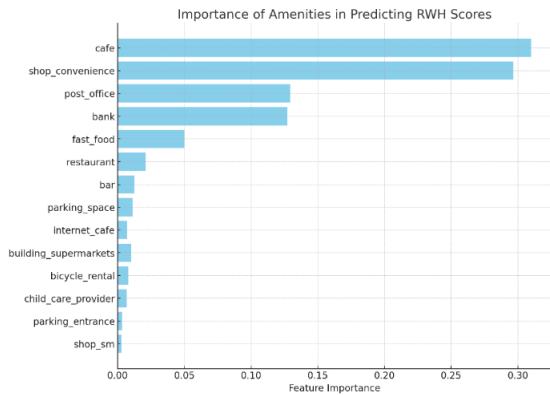


Fig. 4. The Importance of Amenities.

B. Urban Amenity Influence on RWHs

In this study, the Random Forest Regressor from the sklearn.ensemble module was applied to enhance the analysis, trained on a portion of the dataset excluding identifiers and overall scores. The objective was to predict the normalised total scores for Remote Working Hubs (RWHs), representing their overall amenity value. Geospatial data tools such as geopandas and shapely were employed to handle geographic shapes and calculate distances within the study's geographical context, enabling accurate simulations.

A geospatial simulation was conducted by generating 100 random points within the Dublin area, each representing hypothetical bank locations. These points were created using Python's random module to ensure wide variability across the study area. After simulating the addition of a bank and calculating its impact on the RWH scores, a broader analysis was carried out by introducing various types of amenities. For each random point, different amenities were simulated at the same location, and their influence on the RWH scores was visualised in Fig. 4. The Random Forest model was then used to predict the updated scores based on the modified amenity data.

Validation was performed using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared, provided by the sklearn.metrics module. The results demonstrated strong predictive performance, with an MAE of 0.032, MSE of 0.001, and R-squared of 0.9748. This process—from data preparation and model training to geospatial simulation and statistical evaluation—offered a comprehensive understanding of the potential impact of various amenities on RWHs within an urban environment.

VI. LIMITATIONS AND FUTURE RESEARCH

The findings of this study highlight the important role that RWHs play in shaping urban landscapes, especially within the context of smart city development and sustainable urban planning. However, several limitations must be acknowledged. A key constraint is the reliance on the quality and completeness of datasets, where inaccuracies or omissions—such as those

from OpenStreetMap or the National Hub Network—could affect model predictions. Additionally, this study offers a static view of the urban environment, omitting temporal dynamics that could reveal how RWHs and amenities interact over time in response to urban development or shifting work patterns.

Another consideration is the potential for biases in using GCNs across different urban contexts. Distinct characteristics, including infrastructure density, socio-economic factors, and cultural norms, may influence model performance in diverse cities. Expanding the model to incorporate data from varied urban settings would help validate its generalizability, while the integration of socio-economic factors and real-time data analytics could enhance its adaptability and relevance for dynamic urban planning scenarios.

VII. CONCLUSION

This study presents the S4 scoring model, an innovative framework designed to evaluate and integrate Remote Working Hubs (RWHs) into urban geospatial environments. Compared to the previous S3 model, the S4 model offers a more nuanced and comprehensive analysis of the relationship between RWHs and surrounding urban amenities. By incorporating Graph Convolutional Networks (GCNs), the model effectively captures both the proximity and connectivity of amenities, providing a more accurate reflection of their impact on RWH functionality. Through a case study in Dublin, the S4 model demonstrates its ability to highlight optimal locations for new RWHs, offering valuable insights for sustainable urban planning and strategic infrastructure development. Additionally, the statistical analysis confirms that the S4 model achieves higher accuracy and more stable correlations than the S3 model, underscoring its value in providing a deeper understanding of how urban amenities influence RWH performance. These findings contribute to the growing body of research on smart city development, suggesting that the S4 model can serve as a powerful decision-making tool for urban planners and policymakers, aiming to enhance the integration of RWHs into the urban ecosystem.

ACKNOWLEDGMENT

This research has been done in the Lighthouse project of CeADAR, Ireland's National Centre for AI.

REFERENCES

- [1] B. Caulfield and A. Charly, "Examining the potential environmental and travel time saved benefits of remote working hubs," *Transport Policy*, vol. 127, pp. 139–147, 2022.
- [2] I. Kojo and S. Nenonen, "Evolution of co-working places: drivers and possibilities," *Intelligent Buildings International*, vol. 9, no. 3, pp. 164–175, 2017.
- [3] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.
- [4] S. N. Kim, S. Choo, and P. L. Mokhtarian, "Home-based telecommuting and intra-household interactions in work and non-work travel: A seemingly unrelated censored regression approach," *Transportation Research Part A: Policy and Practice*, vol. 80, pp. 197–214, 2015.
- [5] E. D. V. Cerqueira, B. Motte-Baumvol, L. B. Chevallier, and O. Bonin, "Does working from home reduce CO2 emissions? An analysis of travel patterns as dictated by workplaces," *Transportation Research Part D: Transport and Environment*, vol. 83, pp. 102338, 2020.
- [6] D. A. Hensher, M. J. Beck, and E. Wei, "Working from home and its implications for strategic transport modelling based on the early days of the COVID-19 pandemic," *Transportation Research Part A: Policy and Practice*, vol. 148, pp. 64–78, 2021.
- [7] J. M. Sahut and R. Lissillour, "The adoption of remote work platforms after the Covid-19 lockdown: New approach, new evidence," *Journal of Business Research*, vol. 154, pp. 113345, 2023.
- [8] T. Lan, H. Cheng, Y. Wang, and B. Wen, "Site selection via learning graph convolutional neural networks: A case study of Singapore," *Remote Sensing*, vol. 14, no. 15, pp. 3579, 2022.
- [9] L. Gao, W. Zhang, Q. Liu, X. Lin, Y. Huang, and X. Zhang, "Machine learning based on the graph convolutional self-organizing map method increases the accuracy of pollution source identification: A case study of trace metal (loid)s in soils of Jiangmen City, south China," *Ecotoxicology and Environmental Safety*, vol. 250, pp. 114467, 2023.
- [10] A. Zeghina, A. Leborgne, F. Le Ber, and A. Vacavant, "Deep learning on spatiotemporal graphs: a systematic review, methodological landscape, and research opportunities," *Neurocomputing*, 2024, vol. 127861.
- [11] Y. Li, W. Huang, G. Cong, H. Wang, and Z. Wang, "Urban region representation learning with OpenStreetMap building footprints," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1363–1373, Aug. 2023.
- [12] L. Xiao, S. Lo, J. Zhou, J. Liu, and L. Yang, "Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China," *Environment and Planning B: Urban Analytics and City Science*, vol. 48, no. 8, pp. 2363–2384, 2021.
- [13] Y. Wang and D. Zhu, "A hypergraph-based hybrid graph convolutional network for intracity human activity intensity prediction and geographic relationship interpretation," *Information Fusion*, vol. 104, pp. 102149, 2024.
- [14] K. Qin, Y. Xu, C. Kang, and M. P. Kwan, "A graph convolutional network model for evaluating potential congestion spots based on local urban built environments," *Transactions in GIS*, vol. 24, no. 5, pp. 1382–1401, 2020.
- [15] L. Gao, W. Zhang, Q. Liu, X. Lin, Y. Huang, and X. Zhang, "Machine learning based on the graph convolutional self-organizing map method increases the accuracy of pollution source identification: A case study of trace metal (loid)s in soils of Jiangmen City, south China," *Ecotoxicology and Environmental Safety*, vol. 250, pp. 114467, 2023.
- [16] H. Yu, T. Ai, M. Yang, L. Huang, and J. Yuan, "A recognition method for drainage patterns using a graph convolutional network," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, pp. 102696, 2022.
- [17] T. Lan, H. Cheng, Y. Wang, and B. Wen, "Site selection via learning graph convolutional neural networks: A case study of Singapore," *Remote Sensing*, vol. 14, no. 15, pp. 3579, 2022.
- [18] X. Yao, Y. Gao, D. Zhu, E. Manley, J. Wang, and Y. Liu, "Spatial origin-destination flow imputation using graph convolutional networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7474–7484, 2020.
- [19] Y. Xu, B. Zhou, S. Jin, X. Xie, Z. Chen, S. Hu, and N. He, "A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method," *Computers, Environment and Urban Systems*, vol. 95, pp. 101807, 2022.
- [20] Y. Wang and C. Jing, "Spatiotemporal graph convolutional network for multi-scale traffic forecasting," *ISPRS International Journal of Geo-Information*, vol. 11, no. 2, pp. 102, 2022.
- [21] Z. Gong, C. Wang, Y. Chen, B. Liu, P. Zhao, and Z. Zhou, "Learning spatial interaction representation with heterogeneous graph convolutional networks for urban land-use inference," *International Journal of Geographical Information Science*, pp. 1–37, 2024.
- [22] Connected Hubs The Connected Hubs Network offers people the opportunity to work comfortably and securely in professional environments all over Ireland. [Online]. Available: <https://connectedhubs.ie/>
- [23] OpenStreetMap. [Online]. Available: <https://www.openstreetmap.org/#map=7/53.465/-8.240>
- [24] Zenodo. [Online]. Available: <https://zenodo.org/>
- [25] M. H. S. Jahromi, M. Yazdi, A. Dehghani, and M. Rasouli, "New scoring model for remote working hubs analysis and decision making," in *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, vol. 1, pp. 1–4, 2023.

Kinetics Evaluation of Total Solids Content Affects on Biohydrogen Production from Food Waste by Dark Fermentation

Sorchat Intana
School of Energy and Environment
University of Phayao
 Phayao, Thailand
 biohydrogenffw@gmail.com

Sukhuma Chitapornpan
School of Energy and Environment
University of Phayao
 Phayao, Thailand
 sukhuma.ch@up.ac.th

Somanat Somprasert
School of Energy and Environment
University of Phayao
 Phayao, Thailand
 somanat.so@up.ac.th

Satawat Tanarat
School of Energy and Environment
University of Phayao
 Phayao, Thailand
 satawat.ta@up.ac.th

Abstract— The impact of initial total solids content on biohydrogen production from food waste by dark fermentation was investigated in the present study. Batch experiment was taken into account with different TS concentration in the range of 10-50 g/L under the starting pH of 5.5. The optimum concentration of solids was identified to lie between 10-30 g/L, with the highest hydrogen yield and production rate of 90.7 mL/g TS_{added} and 15.7 mL/g VS-h attained at TS concentration of 30 g/L. Detrimental conditions that may significantly affect the long-term stability of the system include a final pH reading beneath 4.18, increased levels of propionic acid accumulation, and a reduction in the rate of TS solubilization, which was observed at TS concentrations of 40 g/L and higher. The Gompertz model was employed to describe the kinetics of hydrogen production. The maximum hydrogen yield (H_{max}), maximum production rate (R_{max}), and lag phase duration were found to be 10.17 mL H₂, 56.02 mL H₂/day, and 12.2 hours, respectively. These results were obtained at a total solids (TS) concentration of 30 g/L.

Keywords - Biohydrogen, Food waste, Dark fermentation, Solids content, Kinetics

I. INTRODUCTION

Food waste accounts approximately 8-10% of total global Greenhouse gas (GHG) emissions. When food is discarded, it frequently finds up in landfills, where it decomposes anaerobically, emitting methane, a strong GHG. The environmental impact of food waste is aggravated by rising urbanization and population increase, necessitating appropriate waste management strategies. Food waste in Thailand is a major environmental and economic issue, accounting for 40-70% of municipal solid trash, particularly in metropolitan and tourist areas such as Bangkok and Phuket [1]. The management of this waste presents issues since incorrect disposal methods, such as landfilling, result in substantial greenhouse gas emissions, mainly methane. Effective food waste management solutions, such as separation for animal feed or organic fertilizer production, can result in significant economic gains worth up to \$960,000 USD per year. Furthermore, Thailand intends to recycle 50% of organic waste by 2026, supporting a Bio-Circular-Green (BCG) economy model.

Anaerobic digestion (AD) is recognized as a promising method for food waste management and energy recovery. The process involves the breakdown of organic matter in the absence of oxygen, producing biogas, primarily methane, which can be used as a renewable energy source. Co-digestion

process, which involves combining food waste with other substrates, can enhance biogas production. For instance, the co-digestion of food waste with sewage showed that a higher proportion of food waste led to increased biogas yield [2-4]. Similarly, co-digestion with animal manure has been shown to significantly increase biogas production compared to food waste alone, demonstrating a synergistic effect that enhances substrate biodegradation, microbial load reduction and achieving a higher hydrolysis rate [5-7]. Moreover, integrating anaerobic digestion with other processes, such as pyrolysis, can further improve environmental sustainability. The anaerobic digestion-co-pyrolysis system has been shown to significantly reduce the global warming potential compared to standalone AD or pyrolysis, by effectively treating the solid fraction and impurities in food waste [8].

Hydrogen is considered as a clean energy source because of its ability to generate electricity without emitting pollutants or greenhouse gasses during use. Hydrogen combustion produces solely water as a byproduct due to its high specific energy content (33.3-39.4 kWh/kg) making it a viable alternative to fossil fuels, which contribute considerably to environmental degradation and climate change. Biohydrogen, which is made from renewable waste feedstocks such as agricultural and municipal trash, reducing pollution while simultaneously improving energy security and sustainability. Biomass-based hydrogen production has a significantly lower global warming potential (GWP) and acidification potential than fossil fuel-based processes, which exhibit high GWP values due to emissions from coal mining and transportation. Dark fermentation is one of potential process for producing biohydrogen from various organic substrates. This process is characterized by its light-independent nature and involves complex interactions among diverse microbial communities, which operate under anaerobic conditions to convert organic matter into biohydrogen and volatile fatty acids (VFAs). Research works have demonstrated the satisfactory outcomes of dark fermentation in producing biohydrogen from various organic substrates, such as the organic fraction of municipal solid waste (OFMSW) [9-10], biomass [11-13], slurry [14] and cheese whey [15].

Food waste is also being utilized for biohydrogen production. Most research has focused on dark fermentation and explored different strategies to optimize the process. Literatures cover many physicochemical factors to consider when producing biohydrogen, including waste chemical nature, moisture content, substrate concentration, carbon to nitrogen (C/N) ratio, pH, volatile fatty acids, hydrogen partial

pressure, inoculum sources, and pretreatment for the substrate and mixed culture [16-18]. Total solids (TS) concentration is also playing a crucial role in dark fermentation of food waste in many ways as maintaining appropriate TS level is essential for maximizing hydrogen yield. A high initial TS level can impede mass transfer between substrate and bacteria, resulting in less hydrogen production rate. The appropriate initial TS concentration for hydrogen production relies on waste composition, reactor configuration, and microorganism activity. There have been reports of low TS concentrations, which taken into account by diluting with water or mixing with other wastes, employed as several authors utilized different beginning TS values, ranging from 1.3 to 50 g/L [19]. To reduce the need for fresh water when diluting food waste prior to the dark fermentation, greater TS concentrations is worthwhile to be investigated compared to previous studies.

Although there have been studies on various important factors in hydrogen production in the past, more investigation is needed to fully grasp the potential of the process for improving bioenergy recovery efficiency. Certainly, kinetic analysis is a vital tool to analyze the dynamics of fermentative operation and, thus, achieve improved control and design of the process. Kinetics assessment has received insufficient attention to date as few research works identified kinetically analysis of the process [20-21]. To better comprehend the complexities of the biohydrogen production from food waste via dark fermentation, whereas kinetic parameters, process parameters, and operating conditions are closely related, a thorough kinetics evaluation corresponding with an appropriate model should be taken into account.

The impact of starting total solids content on biohydrogen production from food waste by dark fermentation had been investigated in the present study. Besides, process kinetics by the Gompertz model was also evaluated. The first objective was to identify the highest initial total solids content to be applied for biohydrogen production using as fewer diluted water as possible, whereas the second aim was to assess the best fit kinetics for dark fermentation of food waste.

II. MATERIALS AND METHODS

A. Substrate and Inoculum

Food waste was gathered from the main cafeteria at University of Phayao, Phayao, Thailand. It was mostly composed of rice, grains, vegetables, fruits, and meat. Bones and non-degradable particles were removed from the sampling before being ground in a food blender. Finally, the waste was refrigerated at 4°C prior the usage. The characteristics of the food waste used as substrate in this study are presented in Table 1.

TABLE 1. THE CHARACTERISTICS OF THE FOOD WASTE USED IN THIS STUDY

Parameter	Unit	Value
Moister content	%	85.4 ± 1.9
pH	-	5.74 ± 0.21
COD	g/L	149.6 ± 10.8
TS	g/L	56.8 ± 1.4
TVS	g/L	51.4 ± 1.0
C	%	63.2 ± 0.9
N	%	2.0 ± 0.2

Anaerobic seed sludge was collected from a full-scale channel digester at a poultry slaughterhouse in the northern part of Thailand. The sludge was pre-heated at 105°C for 3 hours in a drying oven to suppress the methane forming bacteria activity. The sludge had a pH and a volatile solids (VS) content of 7.0 and 9.65 g/L, respectively. For inoculum preparation, pre-heated sludge was grown in food waste at 20 g-COD/L and supplemented with nutrient solution at a rate of 0.5 mL/L. The composition of the nutrient solution was as follows: NH₄HCO₃ 200 g/L; KH₂PO₄ 100 g/L; MgSO₄.7H₂O 10 g/L; NaCl 1 g/L; Na₂MoO₄.2H₂O 1 g/L; CaCl₂.2H₂O 1 g/L; FeCl₂ 0.278 g/L [22]. The culture was incubated for 36 hours at 150 rpm before being applied as inoculums in a batch experiment.

B. Experimental design

The biohydrogen production was undertaken in 300 mL serum bottles with a working volume of 150 mL by batch experiment. For creating anaerobic condition, nitrogen gas was used to eliminate oxygen from batch reactors and were then capped with rubber stoppers. The bottles were kept at ambient temperature and shaken in an orbital shaker with a speed of 150 rpm. The bottles were filled with pre-heated anaerobic sludge with an initial VS concentration of 15 g/L_{reactor} and food waste, which was diluted with tap water, at different TS concentration in the range of 10-50 g/L to assess the impact of the initial TS concentration. Each batch bottle was adjusted with 0.1 N HCl to achieve an initial pH of 5.5 according to the suggestion from the literatures [23]. The pH was set at the start of the test and then reduced while fermentation progressed in the batch bottles.

C. Kinetics analysis

The experimental procedures were executed in triplicate. A kinetic examination of the cumulative hydrogen generation in relation to the initial TS concentration was conducted utilizing the experimental data acquired during specific batch cycles where the hydrogen output per cycle reached its peak. The modified Gompertz, as in (1), was employed to model the dynamics of biohydrogen production [24].

$$H(t) = H_{max} * \exp \left[-\exp \left(\frac{2.71828 * R_{max}(\gamma-t)}{H_{max}} + 1 \right) \right] \quad (1)$$

where, H(t) (mL) is the total amount of hydrogen produced at culture time, t (h); H_{max} (ml) is the maximal amount of hydrogen produced; R_{max} (mL/h) is the maximum hydrogen production rate; γ (h) is the lag time before the exponential hydrogen production.

D. Analytical analysis

Total Organic Carbon (TOC) was quantified using a Shimadzu 5050 TOC analyzer, while Volatile Suspended Solids (VSS) concentration, Chemical Oxygen Demand (COD), and Ammonia Nitrogen (NH₃-N) were assessed in accordance with standard methods [25]. Hydrogen, carbon dioxide, and methane were examined utilizing a gas chromatograph (Agilent 6890 N) outfitted with a thermal conductivity detector and a 30 m long (0.53 mm id) Carboxen 1010 Plot column. The temperatures employed for the injection port, column, and detector were set at 200, 100, and 230 °C, respectively. Nitrogen served as the carrier gas, maintaining a flow rate of 4 mL min⁻¹. At the conclusion of the run, liquid samples were collected for the assessment of volatile fatty acids (VFA) and solvents. To evaluate VFA,

alcohols, and acetone, 1 mL of the sample underwent centrifugation at 600g for a duration of 5 minutes. The concentrations of these by-products were quantified using a chromatograph (Varian 3300) equipped with a flame ionization detector (FID) and a 15 m long (0.53 mm id) Zebron ZB-FFAP column. The temperatures of the injector and detector were regulated at 190 and 210 C, respectively. The column's temperature was held constant at 45 C for a duration of 1.5 minutes; subsequently, it was raised to 135 C at a rate of 8 C/min. Nitrogen served as the carrier gas at a flow rate of 9.5 mL/min.

III. RESULTS AND DISCUSSION

A. pH variation and VFA production

The effect of TS concentration on initial pH, final pH and VFA production is illustrated in Fig. 1. The initial pH, which was fixed at 5.5, and final pH values recorded after a duration of 90 hours. Results indicate a negative relationship between the TS concentration and final pH, suggesting that an increase in TS concentration is associated with a decrease in final pH. The reactor exhibited a lower final pH at elevated TS concentration as the highest final pH values of 4.91 was achieved at 10 g/L TS level. The decrease in pH from commencement to conclusion exhibited a slight trend (0.59-0.71) across TS concentrations from 10 to 30 g/L. While, significant pH declines ranging from 1.32 to 1.44, were noted at TS concentrations exceeding 30 g/L. As fermentation progressed, the pH gradually decreased, primarily influenced by the accumulation of organic acids within the reactor. An increase in VFAs production (990-1,419 mg/L) was observed when the TS concentration in a fermentation process was elevated from 10 to 50 g/L.

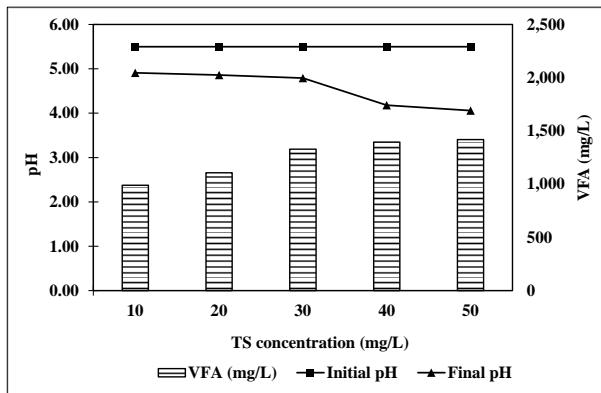


Fig. 1. Effect of TS concentration on initial pH, final pH and VFA production

The analysis of the composition of VFAs production is crucial as it can provide valuable insights into the extent of fermentation and hydrolysis processes. This investigation revealed that there were no statistically significant differences were observed in the VFAs compositions across the TS concentrations in the range of 10-30 mg/L. Butyric acid and acetic acid represented 68.0-75.0% and 21.3-30.2% of total VFAs, respectively, were identified as the primary fermentation products together with minor quantities of lactic acid and propionic acid. The proportion of butyric acid and acetic acid exhibited a notable decline, whereas lactic acid (5.3-6.2%) and propionic acid (21.9-24.9%) demonstrated a substantial rise over the TS concentrations of 40-50 mg/L (Fig. 2).

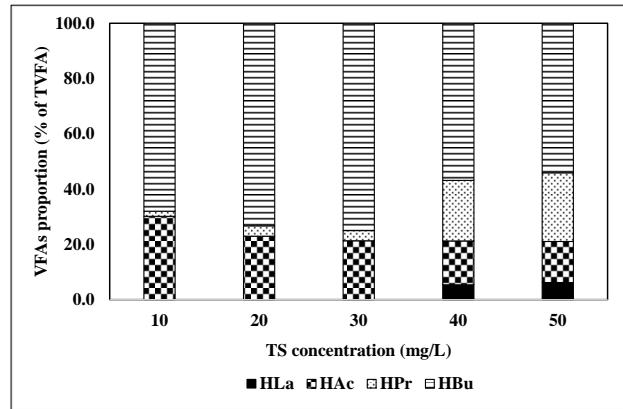


Fig. 2. The proportion of individual VFAs between different TS concentrations

The TS concentration apparently influences pH variation in dark fermentation. Higher TS concentrations can enhance the production of intermediates being consumed by H_2 -producing bacteria but also affect pH levels, which are crucial for maintaining optimal microbial activity. This observation is consistent with previous findings. Jariyaboon et al. found that higher solid concentration led to increased VFA production at 26.4 g VS/L. Additionally, the highest VFA concentration of 4.8 g/L was obtained with 19.25 g/L initial total sugar concentration in the study of dark fermentation of acid hydrolyzed waste ground wheat for hydrogen gas production [26]. These findings suggest that higher TS concentrations can facilitate more efficient fermentation processes, leading to increased VFA yields. It was also claimed that there had been a notable shift in the VFA composition when TS concentration had been higher, characterized by a rise in butyrate production and a concomitant decrease in acetate production [5, 27]. Besides, relative high concentrations of propionic acid were assessed at the 40 g/L and above of TS concentrations. Propionic acid is known to have an adverse effect on dark fermentation processes, primarily due to its inhibitory impact on microbial activity and metabolic pathways as it can penetrate microbial cell membranes leading to intracellular acid forming process. This acidification disrupts cellular homeostasis and inhibits microbial growth and activity, which are important for dark fermentation [28]. While higher TS concentrations generally improve VFA production, they can also lead to challenges such as increased viscosity and potential inhibition of microbial activity. Therefore, it is crucial to balance TS levels to optimize both VFA production and process stability for downstream processing and utilization strategies.

B. COD and TS removal

Fig. 3 illustrates the relationship between initial TS concentration and the removal of both TS and COD. Results indicate an inverse correlation between TS concentration and the efficiency of both TS and COD removal. The highest rates of substrate transformation were observed at 10 g/L starting TS concentration, with removal efficiencies of 72% for TS and 51% for COD. Previous studies have indicated that the hydrolytic-acidogenic phase can be a rate-limiting step in the degradation of organic waste as it is often attributed to the particulate nature of the waste material, which can hinder the accessibility of hydrolytic enzymes to the organic matter. Wang and Wang [29] highlights that the optimal range for liquor concentration, which correlates with total solids, is between 3% and 6% for effective COD removal in cattle

manure anaerobic fermentation. Within this range, the COD removal initially increases with the concentration of solids, suggesting enhanced microbial activity and substrate availability. Higher initial TS concentrations tend to decrease TS solubilization rate and prolong the fermentation period. This is attributed to the lower metabolite production from microbiological activity, which can slow down the overall process.

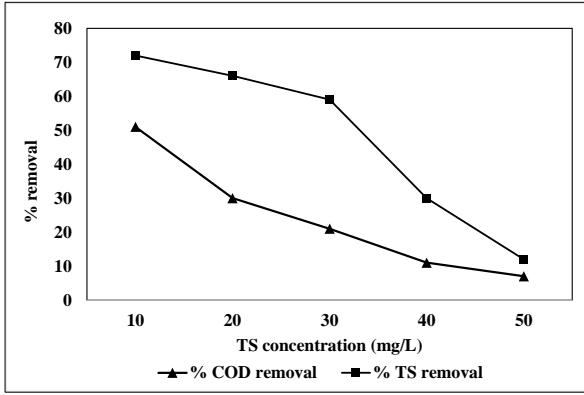


Fig. 3. The relationship between the TS concentration and TS and COD removal efficiency.

C. Biohydrogen production

A positive correlation was found between the hydrogen production and initial TS concentration, with a clear linear relationship observed, as a maximum hydrogen production of 369 mL was achieved at a TS concentration of 50 g/L. Nevertheless, an exclusive reliance on the volumetric hydrogen production for evaluating hydrogen production potential is insufficient. The hydrogen yield, which was determined based on the total solids input, and hydrogen production rate is substantially influenced the overall effectiveness which organic matter to be hydrolyzed into fermentable substrates. A negative correlation, in contrary, was detected for the hydrogen yields and hydrogen production rate as the highest value of those was achieved at low TS level. A maximum hydrogen yield and hydrogen production rate of 90.7 mL/g TS_{added} and 15.7 mL/g VS-h was achieved at TS concentration of 30 g/L while it diminished at 40 g/L and above TS level. Fig. 4 illustrated a function of hydrogen production, hydrogen yields and hydrogen production rates with TS concentrations.

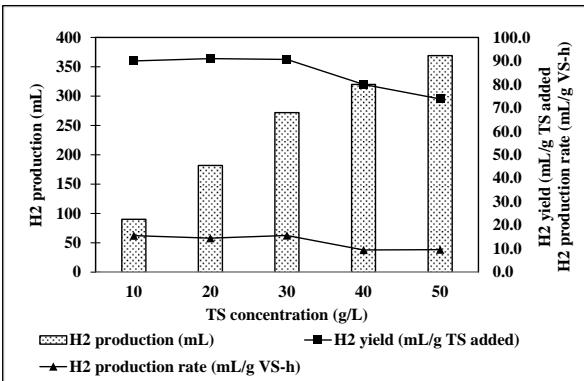


Fig. 4. The hydrogen production, hydrogen yield and hydrogen production rate at different TS concentrations.

Based on the yield and production rate of biohydrogen obtained, the optimal solids concentration range for biohydrogen production efficiency, in this study, appeared to

be in the range of 10-30 g/L. The highest hydrogen yield and production rate achieved at the lowest TS concentration may be attributed to a more uniform distribution of inoculum and substrate, resulting in minimal localized nutrient deficiencies and reduced toxicant dilution [30]. Research works indicated that a volatile solids (VS) concentration between 28 and 75 g/L is effective for enhancing hydrogen yields by dark fermentation. Nam [31] found that a VS concentration of 40 g/L is optimal for maximizing the specific hydrogen production rate, achieving 163.8 mL H₂/g VS-h when food waste was co-digested with sewage sludge under alkali pretreatment conditions. Kazemi et al. [32] likewise conducted the empirical investigation into the dark fermentation of food waste and revealed that an optimal TS concentration of approximately 30 g/L consistently yielded the most efficient hydrogen production rates when co-fermented with pruning wastes from pine, cypress, and berry trees, resulted in the highest bio-H₂ production yield of 84±6 mL/g VS under initially neutral conditions.

D. Kinetics

To ascertain the hydrogen production rate, the kinetic parameters of the Gompertz model were analyzed. Figure 5 illustrates an exemplar of the fitted data obtained, while Table 2 encapsulates the derived values. Overall, the model demonstrated a high degree of correlation with the experimental data. No lag phases were noted for the concentrations of 10 g TS/L; however, a lag-time ranging from 10 to 48 hours was identified for initial TS concentrations exceeding 10 g/L.

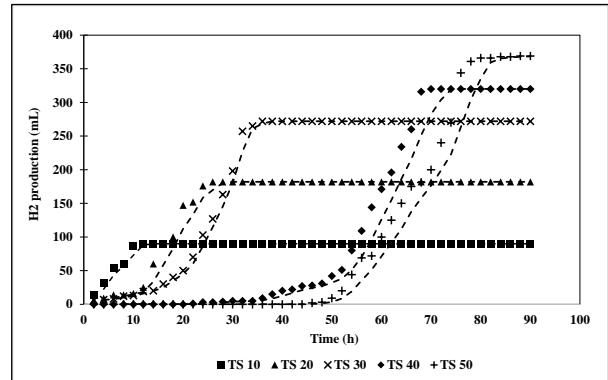


Fig. 5. Kinetic fitting for selected values and Gompertz model adjustment at different TS concentrations.

TABLE 2. GOMPERTZ MODEL PARAMETERS IN THIS STUDY

TS concentration (g/L)	H ₂ max (mL H ₂)	R _{max} (mL H ₂ /d)	γ (h)
10	2.63 ± 0.12	49.52 ± 2.60	0.0 ± 0.0
20	2.34 ± 0.18	47.08 ± 2.25	8.1 ± 0.2
30	10.17 ± 0.29	56.02 ± 3.10	12.2 ± 0.3
40	12.63 ± 0.42	37.98 ± 0.87	23.6 ± 0.6
50	14.28 ± 0.31	26.82 ± 0.53	45.4 ± 0.5

The Gompertz model has been represented a prominent mathematical framework for characterizing microbial proliferation and the kinetics of biohydrogen production derived from organic waste. This model proves to be especially advantageous for assessing the lag phase, maximum production rate, and theoretical yield of hydrogen. The kinetic parameters obtained from the present work exhibit a correlation with the findings derived from previous studies.

Pecorini et al. [33] founded that a duration of the lag phase as defined as a period required to achieve 95% of the peak hydrogen yield estimated to be roughly one day (29.3 h). The Gompertz model was similarly employed to evaluate the potential for hydrogen generation from diverse food waste substrates, including rice, fish, and vegetables and indicated that a mixture of waste exhibited the highest hydrogen production rate at 60 mL/d [34]. Important kinetics including lag time (λ), and maximum H₂ production rate (R_{max}) also accompanied with ones obtained from the investigation of hydrogen production from fruit and vegetable waste using mixed anaerobic cultures via dark fermentation [35]. The utilization of the Gompertz model across diverse research investigations underscores its efficacy in forecasting hydrogen generation from an assortment of food waste types under a range of conditions.

For further investigation, the economic assessment of byproducts, such as VFA production, which shows higher revenues compared to the biogas and supports the circular economy, will be taken into account for making it a financially viable option.

IV. CONCLUSIONS

The impact of the initial TS concentration on biohydrogen production from food waste by dark fermentation was examined. Adverse circumstances that could potentially influence the enduring stability of the system as the final pH below 4.18, higher propionic acid accumulation and TS solubilization rate decreasing were detected at TS concentrations of 40 g/L and above. A maximum hydrogen production of 369 mL was achieved at the highest TS concentration of 50 g/L but there was a negative correlation for the hydrogen yields and hydrogen production rate with TS level. The optimal solids concentration range for biohydrogen production appeared to be in the range of 10-30 g/L as a maximum hydrogen yield and hydrogen production rate of 90.7 mL/g TS_{added} and 15.7 mL/g VS-h was achieved at TS concentration of 30 g/L. The kinetic parameters were also evaluated by the Gompertz model. A robust correlation between the theoretical model and empirical findings was attained with H_{max}, R_{max}, and lag phase time designated as 10.17 mL H₂, 56.02 mL H₂/d, and 12.2 h, respectively at 30 g/L TS concentration.

ACKNOWLEDGMENT

This research was supported by “University of Phayao” and “Thailand Science Research and Innovation Fund (Fundamental Fund 2024)”. The authors also acknowledge the technical supports of the Environmental Testing Laboratory Center (ETLC), School of Energy and Environment, University of Phayao.

REFERENCES

- [1] A. Saiphet and K. Kunta, “Municipal Food Waste Management and Its Greenhouse Gas Emission in Thailand: A Case Study of 8 Municipalities”, *Environmental Research, Engineering and Management*, vol. 79(3), pp. 60–74, 2023.
- [2] M. Li, Z. Youcui, G. Qiang, Q. Xiaoqing, N. Dongjie, “Bio-hydrogen production from food waste and sewage sludge in the presence of aged refuse excavated from refuse landfill”, *Renewable Energy*, vol. 33, pp. 2573-2579, 2008.
- [3] C. Sreela-or, P. Plangklang, T. Imai, A. Reungsang, “Co-digestion of food waste and sludge for hydrogen production by anaerobic mixed cultures: statistical key factors optimization”, *International Journal of Hydrogen Energy*, vol. 36, pp. 14227-14237, 2011b.
- [4] H. Zhu, W. Parker, D. Conidi, R. Basnar, P. Seto, “Eliminating methanogenic activity in hydrogen reactor to improve biogas production in a two-stage anaerobic digestion process co-digesting municipal food waste and sewage sludge”, *Bioresource Technology*, vol. 102, pp. 7086-7092, 2011.
- [5] M. Wang, X. Sun, P. Li, L. Yin, D. Liu, Y. Zhang, W. Li, G. Zheng, “A novel alternate feeding mode for semi-continuous anaerobic co-digestion of food waste with chicken manure”, *Bioresour. Technol.*, vol. 164, pp. 309–314, 2014.
- [6] C. Dennehy, P.G. Lawlor, Croize, T. , Y. Jiang, L. Morrison, G.E. Gardiner, X. Zhan, “Synergism and effect of high initial volatile fatty acid concentrations during food waste and pig manure anaerobic co-digestion”, *Waste Manage.*, vol. 56, pp. 173–180, 2016.
- [7] R. Rajagopal, D. Bellavance, M.S. Rahaman, “Psychrophilic anaerobic digestion of semi-dry mixed municipal food waste: for North American context”, *Process Saf. Environ. Prot.*, vol. 105, pp. 101–108, 2017.
- [8] C. Park, N. Lee, J. Kim, J. Lee, “Co-pyrolysis of food waste and wood bark to produce hydrogen with minimizing pollutant emissions”, *Environmental Pollution*, vol. 270, pp. 116045, 2021.
- [9] I. Valdez-vazquez, E. Riosleal, F. Esparzagarcia, F. Cecchi, H. Poggivarnaldo, “Semicontinuous solid substrate anaerobic reactors for H₂ production from organic waste: mesophilic versus thermophilic regime”, *Int J Hydrogen Energy*, vol. 30, pp. 1383–91, 2005.
- [10] L.M. Alzate-Gaviria, P.J. Sebastian, A. Perez-Hernandez, D. Eapen, “Comparison of two anaerobic systems for hydrogen production from the organic fraction of municipal solid waste and synthetic wastewater”, *Int J Hydrogen Energy*, vol. 32, pp. 3141–6, 2007.
- [11] Y.T. Fan, Y.H. Zhang, S.F. Zhang, H.W. Hou, B.Z. Ren, “Efficient conversion of wheat straw wastes into biohydrogen gas by cow dung compost”, *Bioresour Technol*, vol. 97, pp. 500–5, 2006.
- [12] M.L. Zhang, Y.T. Fan, Y. Xing, C.M. Pan, G.S. Zhang, J.J. Lay, “Enhanced biohydrogen production from cornstalk wastes with acidification pretreatment by mixed anaerobic cultures”, *Biomass Bioenergy*, vol. 31, pp. 250–4, 2007.
- [13] C.C. Chen, Y.S. Chuang, C.Y. Lin, C.H. Lay, B. Sen, “Thermophilic dark fermentation of untreated rice straw using mixed cultures for hydrogen production”, *Int J Hydrogen Energy*, vol. 37, pp. 15540–6, 2012.
- [14] T. Kotsopoulos, I. Fotidis, N. Tsolakis, G.G. Martzopoulos, “Biohydrogen production from pig slurry in a CSTR reactor system with mixed cultures under hyper-thermophilic temperature (70 °C)”, *Biomass Bioenergy*, vol. 33, pp. 1168–74, 2009.
- [15] F. Kargi, N.S. Eren, S. Ozmihci, “Bio-hydrogen production from cheese whey powder (CWP) solution: comparison of thermophilic and mesophilic dark fermentations”, *Int J Hydrogen Energy*, vol. 37, pp. 8338–42, 2012.
- [16] E. Elbeshbishy, H. Hafez, B.R. Dhar, G. Nakhla, “Single and combined effect of various pretreatment methods for biohydrogen production from food waste”, *Int. J. Hydrogen. Energy*, vol. 36, pp. 11379-11387, 2011b.
- [17] J.H. Hwang, J.A. Choi, R.A.I. Abou-Shanab, B. Min, H. Song, Y. Kim, E.S. Lee, B.H. Jeon, “Feasibility of hydrogen production from ripened fruits by a combined two-stage (dark/dark) fermentation system”, *Bioresour. Technol.*, vol. 102, pp. 1051-1058, 2011.
- [18] H.M.N. Yasin, A.R. Nor'aini, C.M. Hasfalina, M.Z.M. Yusoff, M.A. Hassan, “Microbial characterization of hydrogen-producing bacteria in fermented food waste at different pH values”, *Int. J. Hydrogen. Energy*, vol. 36, pp. 9571-9580, 2011.
- [19] C. Ramos, G. Buitrón, I. Moreno-Andrade, R. Chamy, “Effect of the initial total solids concentration and initial pH on the bio-hydrogen production from cafeteria food waste”, *International Journal of Hydrogen Energy*, vol. 37(8), pp. 13288-13295, 2012.
- [20] I. Rivera, P. Bakonyi, G. Buitron, “H₂ production in membraneless bioelectrochemical cells with optimized architecture: the effect of cathode surface area and electrode distance”, *Chemosphere*, vol. 171, pp. 379-385, 2017.
- [21] K.B. Prajapati, R. Singh, “Bio-electrochemically hydrogen and methane production from co-digestion of wastes”, *Energy*, vol. 198, p. 117259, 2020.
- [22] J.J. Lay, Y.J. Lee, T. Noike, “Feasibility of biological hydrogen production from organic fraction of municipal solid waste”, *Water Res*, vol. 33, pp. 2579-86, 1999.
- [23] W. Zong, R. Yu, P. Zhang, M. Fan, Z. Zhou, “Efficient hydrogen gas production from cassava and food waste by a two-step process of dark

- fermentation and photo-fermentation. Biomass”, Bioenerg., vol. 33, pp. 1458-1463, 2009.
- [24] G. Bruiton and C. Carvajal, “Biohydrogen production from Tequila vinasses in an anaerobic sequencing batch reactor: Effect of initial substrate concentration, temperature and hydraulic retention time”, Bioresource Technology, vol. 101, pp. 9071–9077, 2010.
- [25] American Public Health Association/American Water Works Association/Water Environment Federation, “Standard methods for the examination of water and wastewater”, 21th ed. Baltimore: Port city press, 2005.
- [26] S. Ozmihiçi, F. Kargi, A. Cakir, “Thermophilic dark fermentation of acid hydrolyzed waste ground wheat for hydrogen gas production”, International Journal of Hydrogen Energy, vol. 36(3), pp. 2111-2117, 2011.
- [27] J. Hao, H. Zhuo, X. Shen, H. Wang, “Unveiling the biotic and abiotic mechanisms of solids concentration affecting hydrolysis and acidogenesis during sludge thermophilic anaerobic fermentation”, Journal of Cleaner Production, vol. 252, 2020.
- [28] J. Xu, J. Bao, X. Su, H. Zhang, X. Zeng, L. Tang, K. Wang, J. Zhang, C. Xusheng, Z. Mao, “Effect of propionic acid on citric acid fermentation in an integrated citric acid-methane fermentation process”, Bioprocess and Biosystems Engineering, vol. 39(3), pp. 391-400, 2016.
- [29] L. Wang and Z. Wang, “Study on the Factors Influencing COD Removal Rate of Cattle Manure Anaerobic Fermentation”, 2010 Asia-Pacific Power and Energy Engineering Conference, Chengdu, China, 1-4, 2010.
- [30] B. Zhang, P.J. He, F. Lu, L. Shao, P. Wang, “Extracellular enzyme activities during regulated hydrolysis of high-solid organic wastes”, Wat Res, vol. 41, pp. 4468-78, 2007.
- [31] J.Y. Nam, “Optimum Conditions for Enhanced Biohydrogen Production from a Mixture of Food Waste and Sewage Sludge with Alkali Pretreatment”, Energies, vol. 16(7), pp. 3281-3281, 2023.
- [32] R. Kazemi, S. Mirmohamadsadeghi, H. Amiri, “Efficient Biohydrogen Production by Dark Co-fermentation of Food-rich Municipal Solid Waste and Urban Pruning Wastes of Pine, Cypress, and Berry”, Process Safety and Environmental Protection, 2024.
- [33] I. Pecorini, F. Baldi, E. Albini, G. Galloppi, D. Bacchi, A. Baldi, A. Bianchini, A. Figini, P. Rossi, P. Paoli, C. Carcasci, G. Ferrara, E. Carnevale, “Hydrogen production from food waste using biochemical hydrogen potential test”, Procedia Environmental Science, Engineering and Management, vol. 4(3), pp. 155-162, 2017.
- [34] W. Han, M. Ye, A. Zhu, J. Huang, H. Zhao, Y. Li, “A combined bioprocess based on solid-state fermentation for dark fermentative hydrogen production from food waste”, Journal of Cleaner Production, vol. 112, pp. 3744-3749, 2016.
- [35] A. Dwivedi, V. Vidyadhar, M. Gedam, K. Suresh, Sustainable hydrogen production from fruit and vegetable waste (FVW) using mixed anaerobic cultures via dark fermentation: kinetic aspects, international journal of energy and environmental engineering, 11(3):341-349, 2020.

Climate Modeling in the Gulf of Thailand to Predict Beach Loss Due to Sea Level Rise

Mesharath Singharath*
*Applied Geoinformatics
 School of Information and
 Communication Technology,
 University of Phayao
 Phayao, Thailand
66025166@up.ac.th /
mesharath.singharath@yahoo.com*

Thanakrit Peebkhunthod
*Applied Geoinformatics
 School of Information and
 Communication Technology,
 University of Phayao
 Phayao, Thailand
65026458@up.ac.th /
thanakrit.p@msu.ac.th*

Phaisarn Jeefoo
*Geographic Information Science
 School of Information and
 Communication Technology,
 University of Phayao
 Phayao, Thailand
phaisarn.jeefoo@up.ac.th /
<https://orcid.org/0000-0001-8504-4026>*

Abstract— The world's sea levels have been rising during the 20th and early 21st centuries, following thousands of years of little change. Based on two Representative Concentration Pathway (RCP) scenarios, namely RCP2.5 and RCP8.5, this study estimated future SLR for all of Thailand's coastlines and found that between 2081 and 2100 using Coupled Model Intercomparison Project 5 (CMIP5). The results were presented of RCP 2.6 found that 13 sub districts in Rayong province had beach loss rates ranging from 70 to 92% and more than 92% in Choeng Noen, Noen Phra, Pak Nam, Taphong, and Tha Pa-Du and then Beach loss rates were identified in six sub districts of the fourth Suratthani province (54–61% and 100–145%) according to RCP 8.5, especially in Don Sak sub district.

Keywords— Climate Modeling, Sea Level Rise, CMIP5, the Gulf of Thailand

I. INTRODUCTION

One coastal hazard that is made worse by storm surges, wave action, shoreline erosion, wetland loss, and saltwater intrusion is sea level rise [1]. Sea level rise poses a threat to human infrastructure as well as the natural environment. Of special concern are low-lying human settlements and productive coastal wetlands. A large number of coastal communities across the world are dangerously densely inhabited. According to McGranahan, Balk, and Anderson (2007), the "low elevation coastal zone" is the contiguous territory along the coast that is less than 10 meters above sea level. It makes up only 2% of the world's land area but is home to 10% of its population. Nearly two-thirds of urban places with more than 5 million inhabitants are located at least partially in the low elevation coastal zone, and the majority of these settlements are found in developing nations [2].

Since the end of the 19th century, the estimated rate of rise in sea level worldwide has been between 1 and 2 mm/year [3]. Future sea level rise estimates vary; the IPCC 2007 report predicts that sea level would rise between 19 and 59 centimeters between the base period of 1980–1999 and the century, or 2090–1999. Current predictions place the upper range of sea level rise at 100 cm or more in just three years [4]. If the West Antarctic ice sheet collapses or the Greenland ice sheet melts more quickly than expected, sea levels would rise more quickly [5].

Based on the highest Representative Concentration Pathway (RCP) scenario, the IPCC Fifth Assessment Report (AR5) has proposed a likely rising rate of 8 mm/yr – 16 mm/yr for sea-level forecasts for the late 21st century (2081–2100)

[6]. Accordingly, the average SLR rate during the 20th century may be up to ten times higher in the future [7]. By passive submergence, long-term SLR produces shore line recessions [8–9] and may even cause low-lying coastal areas to flood. Concerns about the effects on the world have already been raised by the acceleration of SLR. For example, it is predicted that between 0.2% and 4.6% of the world's population will experience flooding each year, and that by 2100, flood losses might amount to a maximum of US\$210 trillion under RCP8.5. If sufficient measures are not taken, the future SLR could result in damages related to the environment and society. As a result, researching techniques for projecting sea level impacts and creating remedies to stop the loss are crucial [10].

The sea level in Thailand has not yet been thoroughly studied by many scholars. In their analysis of water level records from 22 tide gauge sites in the Gulf of Thailand and the Andaman Sea between 1972 and 2011, Putcharapitchakon and Ritphring (2012) found that the sea level had increased in Thailand at an overall averaged rate of 6.5 mm/yr [11]. Similarly, Sojisuporn et al. (2013) looked at the change in sea level between 1985 and 2009 and found that, during the previous 25 years, there has been a linear trend of about 5 mm/yr around the Gulf of Thailand [12]. In order to evaluate the projections of future beach loss along Thailand's coastlines against sea level rise scenarios of the Coupled Model Intercomparison Project Phase 5 (CMIP5) in 2081–2100, relative to a reference period 1986–2005, Sompratana et al. (2018) assembled a database of beach characteristics, including grain size diameter, beach slope, and beach width using the Bruun rule [13]. Based on these studies, it appears that Thailand's coasts are already experiencing a greater rate of sea level rise (SLR) than the world average. If coastal management planning is not given immediate attention, this could result in serious erosion in the future. Nevertheless, anticipated beach erosion or socioeconomic impacts were not covered in any of these studies.

Sea level rise brought on by climate change has resulted in coastline recession, which is now one of the biggest problems facing the planet. Because of sea level rise, Thailand's coastline is also expected to erode, particularly in low-lying places. Future projections of this erosion are also required.

Based on two Representative Concentration Pathway (RCP) scenarios, namely RCP2.5 and RCP8.5, this study estimated future SLR for Thailand's whole coastline and found

that, between 2081 and 2100, Thailand may lose 46–72% of its current beach areas.

Thailand's coastline spans around 3,148 km, of which 2,055 km are in the Gulf of Thailand and 1,053 km are in the Andaman Sea. It is situated close to the South China Sea. Based on physical attributes, the Department of Marine and Coastal Resources (DMCR) of Thailand has divided the beaches into 64 zones, of which 51 have sandy beaches. Sandy beaches in Thailand typically have a narrow width of 34.8 meters. The average size of the sand particles on these beaches is 0.3 mm, with over 80% of the beach covered by sand particles 0.2–0.5 mm in size. The shoreline slope of the beach varies from 1° to 14°.

II. METHOD

A. Bruun Rule Model

- The model used in the projection is called the Bruun Rule Model [8].

Based on Fig. 1, the Bruun Rule postulates that the equilibrium profile shape of a beach is preserved by its very nature (Fig. 1 a. top). The beach raises itself to equalize itself when the tide rises, either by lifting the whole cross-section or the water's bottom (Fig. b. top photo) As a result, it is assumed that the beach is eroded in accordance with the Bruun Rule. Bruun hypothesized that sand sediments used to raise the bottom of the water will come from the beach sand dunes (berm) (Fig. 1 a. bottom picture). The loss of beach sand will lead to the replacement of the sand under the water to adjust the balance according to the rising tide.

The Bruun Rule may have a lot of drawbacks. The Bruun Rule's primary drawback, however, is that it expects the sand to balance the sediment within a single cross-section (in-out of the beach). In actuality, however, sand sediments can move between cross-sections along the current (see bottom picture of Fig. 1 a., b.), a process known as long shore sediment transport.

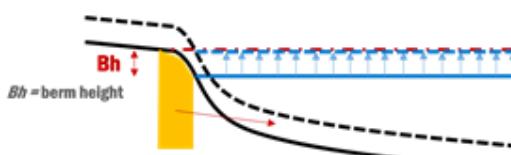
But the Bruun Rule is also often used to evaluate beach erosion at high tide. Its simplicity makes it easy to understand and has basic settings.

It assumes that over the long term the active part of a profile maintains a constant form as an equilibrium profiles.



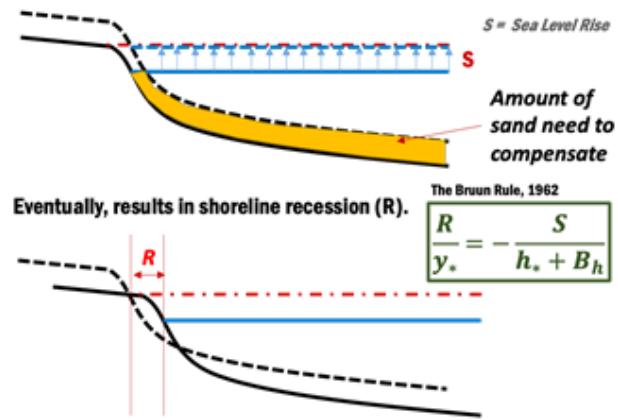
h is the water depth,
A is scaling parameter based on sediment size *d*
and *y* is the distance in the seaward direction

all of the sand is transported from shoreline.



(a.)

When sea level rises (*S*) the profile move upward to maintain its Equilibrium form



(b.)

Fig. 1 Bruun Rule Model.

B. Data

- Data on sea level rise sourced from GMSLR: This calculation's global mean sea level rise data came from WG1AR5_Chapter13_FINAL.pdf (Figure 13.11). In [14]. Mid-range was utilized in all RCP computations.

The mean significant wave height, which is greater than 12 hours every *t* years, as well as the maximum and average significant wave height for the past 30 years.

The mean, maximum, and mean significant wave periods in a 30-year period are all greater than 12 hours every *t* years.

Wave data is reanalyzed every three hours and is available from the European Center for Medium-Range Weather Forecasts (ECMWF) for the years 1980 to 2010. Wave data reanalyzed for three hours, with a resolution of one decimal place in latitude and longitude.

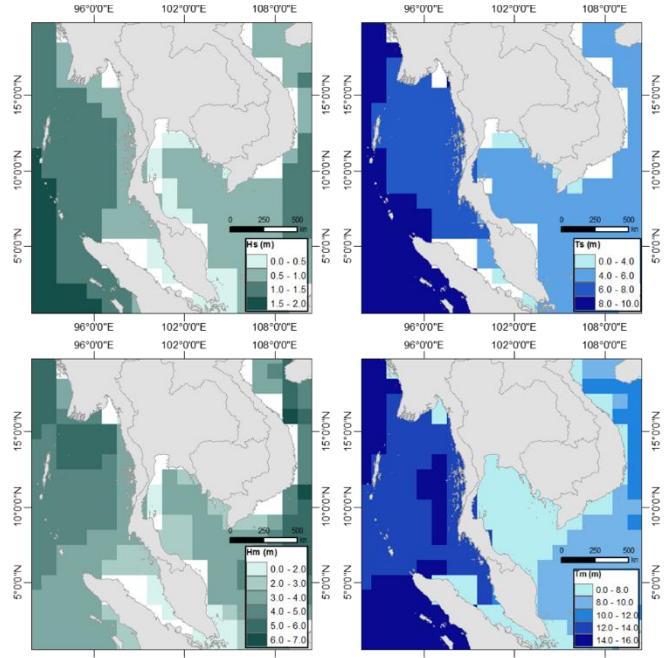


Fig. 2. The top row is the highest significant wave height and the highest significant wave period in 30 years, and the bottom row is mean significant wave height and mean significant tidal period over 30 years.

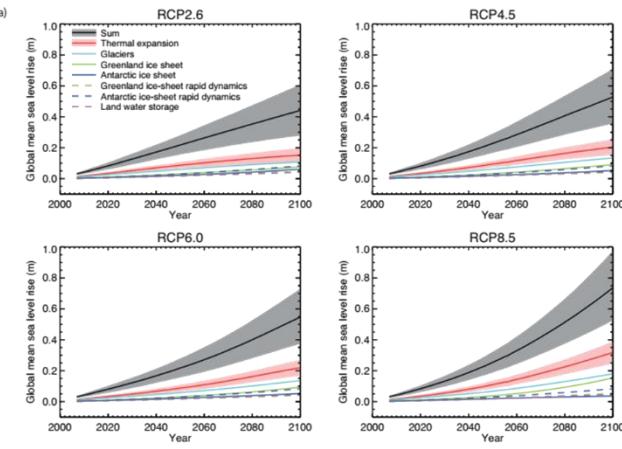


Fig. 3. Projections from process-based models of global mean sea level (GMSL) rise relation to 1986-2005 [14]

The ensemble-mean regional sea level rise data of 21 CMIP5 models for the RCP2.6 (0.37 m.), RCP 4.5 (0.45 m.), RCP6.0 (0.46 m.), and RCP 8.5 (0.61 m.) scenarios.

To generate the beach loss model, Geospatial databases were combined consist of data on sea level rise, beach data, the mean significant wave height, the mean, maximum, and mean significant wave periods in a 30-year period, wave data, and the ensemble-mean regional sea level rise respectively using raster calculator method in ArcGIS software.

Validating a beach loss model against historical data involves several key steps to ensure the model's accuracy and reliability. This study was investigated gather historical data on beach profiles, erosion rates, sea levels, and other relevant environmental factors and collect recent data from DMCR to compare against historical trends.

III. RESULTS

Future Beach Loss Projections (2081–2100): Based on RCP 2.6 (Fig. 4) and RCP 8.5 (Fig. 5) scenarios, these future projections (2080–2100) were derived from Ritphring et al.'s study, which used the Bruun rule assumption to calculate beach loss.

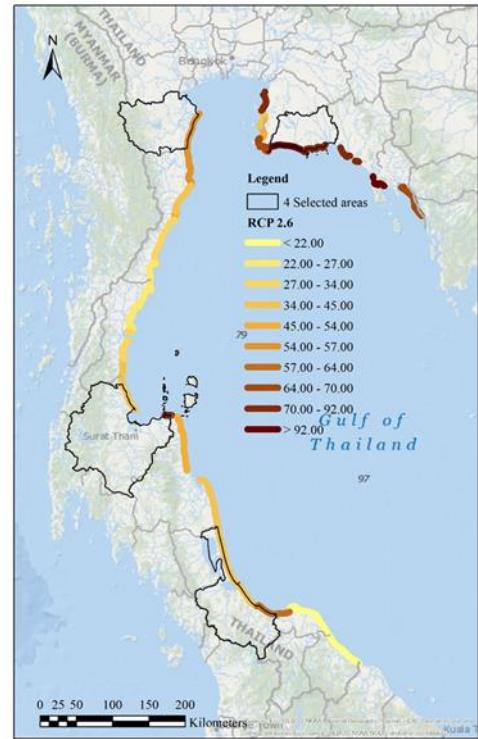


Fig. 4. Projection of beach loss for the RCP 2.6.

RCP 2.6 (Fig. 4), beach loss rate was found 13 sub districts of Rayong province (70 – 92% and more than 92%), the sub district that is affected by the sea level rise are Choeng Noen, Nuen Phra, Pak Nam, Taphong, and Tha Pa-Du (more than 92% of beach loss rate) and Chak Phong, Kram, Klaeng, Phe, and Noen Kho (70 – 92% of beach loss rate).

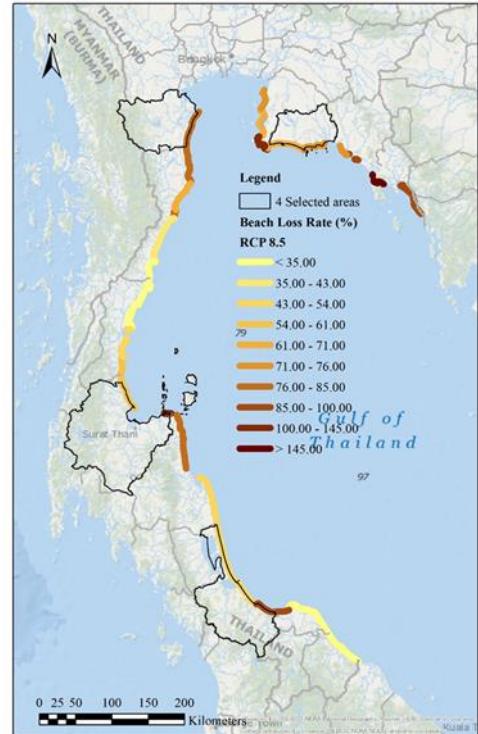


Fig. 5. Projection of beach loss for the RCP 8.5.

RCP 8.5 (Fig. 5), beach loss rate was found 6 sub districts of 4) Suratthani province (54 – 61% and 100 - 145%), the sub

district that is affected by the sea level rise are Don Sak (100 – 145% of beach loss rate) and Khan Thu Li, Phum Riang, Takrob, Tha Chana, and Wang (54 – 61% of beach loss rate).

Fig. 4-5 is adapted from [15] study and shows the present beach width and the projections of the future beach width caused by sea level rise evaluated from Coupled Model Intercomparison Project 5 (CMIP5) as per the 2 RCP scenarios.

IV. CONCLUSION

Based on two Representative Concentration Pathway (RCP) scenarios, namely RCP2.5 and RCP8.5, this study estimated future SLR for Thailand's whole coastline and found that, between 2081 and 2100, Thailand may lose 46-72% of its current beach areas.

Around 3,148 km make up Thailand's coastline, of which 1,053 km are in the Andaman Sea and 2,055 km are in the Gulf of Thailand. It's located not far from the South China Sea. 51 of the 64 zones that the DMCR of Thailand has classified as having sandy beaches are based on physical characteristics. The average width of Thailand's sandy beaches is 34.8 meters. These beaches have sand particles that range in size from 0.2 to 0.5 mm, making up over 80% of the beach. The average size of the particles is 0.3 mm. The beach's slope changes from 1° to 14° along its shoreline.

RCP 2.6 found that 13 sub districts in Rayong province had beach loss rates ranging from 70 to 92% and more than 92%. The sub districts that are most impacted by sea level rise are Chak Phong, Kram, Klaeng, Phe, and Noen Kho (70 – 92% of beach loss rate) and Choeng Noen, Noen Phra, Pak Nam, Taphong, and Tha Pa-Du (more than 92% of beach loss rate).

Beach loss rates were identified in six sub districts of the fourth Suratthani province (54–61% and 100–145%) according to RCP 8.5. Don Sak (100–145% of beach loss rate) and Khan Thu Li, Phum Riang, Takrob, Tha Chana, and Wang (54–61%) are the sub districts that are impacted by sea level rise.

The study's findings show the predicted loss of beach-tourism resort revenue as a result of SLR, and stakeholders and coastal management in Thailand should take action to stop the DMCR, Thailand from suffering large losses. Droughts and other natural calamities can also be modeled using climate data [16-17].

Coastal management strategies to mitigate beach erosion and protect vulnerable areas can be broadly categorized into structural and non-structural approaches. Here are some key strategies:

- Structural Approaches;

Seawalls and Revetments: These are solid structures built parallel to the coastline to protect the shore from wave action. While effective in the short term, they can lead to increased erosion in adjacent areas.

Groins and Breakwaters: These structures are designed to interrupt water flow and limit the movement of sediment. They can help build up beaches but may also cause erosion downstream.

Beach Nourishment: This involves adding sand or sediment to beaches to combat erosion. It's a temporary solution that requires regular maintenance.

- Non-Structural Approaches

Living Shorelines: These use natural elements like plants, sand, and rock to stabilize the shoreline. They provide habitat for wildlife and can adapt to changing conditions.

Dune Restoration: Planting vegetation on dunes helps to stabilize them and protect inland areas from storm surges.

Wetland Restoration: Wetlands act as natural buffers against storm surges and help to absorb wave energy.

- Holistic and Adaptive Management

Ecosystem-Based Approaches: These strategies focus on restoring and maintaining natural coastal ecosystems, which can provide long-term protection and resilience.

Community Involvement and Planning: Engaging local communities in planning and implementing coastal management strategies ensures that solutions are tailored to specific needs and conditions.

Thus, each of these strategies has its advantages and limitations, and often a combination of approaches is used to achieve the best results. It's important to consider the local context and involve stakeholders in the decision-making process to ensure sustainable and effective coastal management.

The Gulf of Thailand is facing significant challenges due to sea level rise, which is projected to cause substantial beach loss. This has several socioeconomic consequences:

Tourism

Beach Erosion: As beaches erode, the attractiveness of coastal areas for tourists diminishes. This can lead to a decline in tourism revenue, which is a crucial part of the local economy.

Infrastructure Damage: Rising sea levels can damage hotels, resorts, and other tourism-related infrastructure, leading to increased maintenance costs and potential loss of business.

Fisheries

Habitat Loss: Coastal erosion and rising sea levels can destroy critical habitats for fish and other marine life, affecting local fisheries.

Economic Impact: Reduced fish populations can lead to lower catches, impacting the livelihoods of those who depend on fishing.

Coastal Communities

Displacement: Rising sea levels can lead to the displacement of communities living in low-lying coastal areas.

Health Risks: Increased flooding and erosion can lead to water contamination and other health risks for coastal residents.

Economic Strain: The combined effects of reduced tourism and fisheries can strain the local economy, leading to higher unemployment and reduced income for residents.

Uncertainties in Sea Level Rise Predictions and Beach Loss Estimates [18];

Ice Sheet Dynamics: The behavior of ice sheets, particularly in Greenland and Antarctica, is a major source of uncertainty. Processes like marine ice sheet instability (MISI) and marine ice cliff instability (MICI) are not fully understood and can significantly impact sea level rise projections.

Climate Models: Different climate models produce varying predictions based on their assumptions and inputs. This variability can lead to a wide range of sea level rise estimates.

Geophysical Data: The choice of geophysical datasets can contribute significantly to the variance in coastal land loss projections. For example, uncertainties in topographic and bathymetric data can affect the accuracy of beach loss estimates.

Emission Scenarios: Future greenhouse gas emissions scenarios (e.g., RCP 2.6 vs. RCP 8.5) lead to different projections of sea level rise, adding another layer of uncertainty.

This study was organized Bruun Rule model for prediction beach loss and Table I are the comparison of predictive models with process-Based Models and Semi-Empirical Models.

TABLE I. COMPARISON OF PREDICTIVE MODELS

Titles	Model		
	Bruun Rule	Process-Based Models	Semi-Empirical Models
Overview	The Bruun Rule is a simple, two-dimensional model that predicts shoreline retreat in response to sea level rise by assuming a conserved equilibrium beach profile.	These models use detailed physical processes to predict sea level rise and beach loss. They consider factors like wave dynamics, sediment transport, and coastal geomorphology.	These models use statistical relationships between observed sea level rise and temperature changes to make predictions.
Strengths	It is computationally efficient and widely used for large-scale assessments due to its simplicity.	They provide more detailed and site-specific predictions, which can be more accurate for local applications.	They can provide broader estimates and are useful for understanding general trends.
Limitations	The Bruun Rule has been criticized for its lack of validation against observations, robustness, and general applicability. Many beach segments do not meet the	They require high-resolution data and are computationally intensive, making them less practical for large-scale assessments.	They may lack the detailed process understanding of physical models and can be less accurate for specific locations.

Titles	Model		
	Bruun Rule	Process-Based Models	Semi-Empirical Models
	assumptions required for its application.		

ACKNOWLEDGMENT

The facilities needed to complete this work were provided by the School of Information and Communication Technology (ICT), University of Phayao, Thailand, which the authors thank for that. We are grateful to the Geographic Information Science for financing our PhD research in 2024.

REFERENCES

- [1] NRC (National Research Council), "A Geospatial Framework for the Coastal Zone: national needs for coastal mapping and charting," Washington, DC: national Academies Press. 2004, pp.151
- [2] McGranahan, G., Balk, D. and Anderson, B., "the rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones," Environment and Urbanization, vol.19, 1, 2007, pp.17-37.
- [3] J.A. Church, and N.J. White, "Sea-Level Rise from the Late 19th to the Early 21st Century," Survey in Geophysics, vol.32, 2011, pp.585-602.
- [4] Climate Office, University of Copenhagen, "Rising sea level set to have major impacts around the world," Press release from Climate Change Copenhagen Conference, 10-12 March 2009.
- [5] R.J. Nicholls, P.P. Wong, V. Burkett, J. Codignotto, J.E. Hay, R.F. McLean, S. Ragoonaden, C.D. Woodroffe, "Coastal systems and low-lying areas. In Climate Change 2007: impacts, adaptation and vulnerability," 2007, pp.315-356, edited by M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, C.E. Hanson. Cambridge, UK: Cambridge University Press.
- [6] Intergovernmental Panel on Climate Change (IPCC), 2013. Climate Change 2013. The Physical Science Basis. Working Group I Contribution to the IPCC 5th Assessment Report – Changes to the Underlying Scientific/Technical Assessment (IPCC-XXVI/DOC.4).
- [7] J.A. Church, P.U. Clark, J.M. Gregory, S. Jevrejeva, A. Levermann, M.A. Merrifield, G.A. Milne, R.S. Nerem, P.D. Nunn, A.J. Payne, W.T. Pfeffer, D. Stammer, and A.S. Unnikrishnan, "Sea-Level Rise by 2100," Science, 342, 2013, pp.1445-1447.
- [8] P. Bruun, "Sea-Level rise as a cause of shore erosion," Journal Waterways and Harbors Division. ASCE, 88, 1962, pp.117-132.
- [9] M.J.F. Stive, S.G.J. Aarninkhof, L. Hamm, H. Hanson, M. Larson, K.M. Wijnberg, R.J. Nicholls, and M. Capobianco, "Variability of shore and shoreline evolution," Coastal Engineering, 47(2), 2002, pp.211-235.
- [10] J. Hinkel, D. Lincke, A.T. Vafeidis, M. Perrette, R.J. Nicholls, R.S.J. Tol, B. Marzeion, X. Fettweis, C. Ionescu, and A. Levermann, "Coastal flood damage and adaptation costs under 21st century sea-level rise," Proceeding of the National Academy of Sciences of the United States of America, 111(9), 2014, pp.3292-3297.
- [11] K. Putcharapitchakon, and S. Ritphring, "Sea Level Change in Thailand," Ladkrabang Engineering Journal, 29(3), 2021, pp.55-60 (in Thai).
- [12] P. Sojisuporn, C. Sangmanee, and G. Wattayakorn, "Recent estimate of sea-level rise in the Gulf of Thailand," Marjo International Journal Science Technology, 7(2), 2013, pp.106-113.
- [13] R. Sompratana, S. Chatuphorn, U. Keiko, and K. So, "Projections of future beach loss due to sea level rise for sandy beaches along Thailand's coastlines," Journal of Coastal Research, 85(1), 2018, pp.541-545.
- [14] https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter1_3_FINAL.pdf
- [15] S. Ritphring, C. Somphong, K. Udo, and S. Kazama, "Projections of future beach loss due to sea-level rise for sandy beaches along Thailand's coastlines," Proceedings from the International Coastal Symposium (ICS) 2018 (Busan, Republic of Korea). J. Coast. Res. 2018, 85, pp.16-20.

- [16] P. Jeefoo, "Thai Eastern Economic Corridor Drought Monitoring using Terra/MODIS Satellite-Based Data," *Geographia Technica*, 2023, vol.18, 2, pp.123-131.
- [17] C. Pachanaparn, P. Jeefoo and P. Rojanavasu, "Application of Remote Sensing for Drought Monitoring with NDVI-based Standardized Vegetation Index in Nan Province, Thailand" 26 – 28 January 2022 the 7th International Conference on Digital Arts, Media and Technology (DAMT) and 5th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (NCON), School of Computer and Information Technology Chiang Rai Rajabhat University, Thailand., 2022, pp. 330-335,
- [18] R. S. W. van de Wal, X. Zhang, S. Minobe, S. Jevrejeva, R. E. M. Riva, C. Little, K. Richter, and M. D. Palmer, "Uncertainties in Long-Term Twenty-First Century Process-Based Coastal Sea-Level Projections," *Surveys in Geophysics*, 2019, vol. 40, pp. 1655–1671.

Enhancing Urban Heat Island Analysis through Indices-Based Prediction of Land Surface Temperature in Khon Kaen City, Thailand

1st Kritchayan Intarat

Center of Excellence in Digital Earth Applications and Emerging Technology (CoE: DEET), Sirindhorn International Institute of Technology / Department of Geography, Faculty of Liberal Arts Thammasat University Pathumthani, Thailand, 12120 intaratt@tu.ac.th

<https://orcid.org/0000-0002-0728-753X>

2nd Saruda Chuenkamol

Department of Geography / Research Unit in Geospatial Applications (Capybara Geo Lab, Faculty of Liberal Arts Thammasat University Pathumthani, Thailand, 12121 saruda.chu@dome.tu.ac.th

<https://orcid.org/0009-0005-4680-2898> <https://orcid.org/0000-0002-5665-0737>

3rd Nithima Nuangjamnong

Department of Geography / Research Unit in Geospatial Applications (Capybara Geo Lab, Faculty of Liberal Arts Thammasat University Pathumthani, Thailand, 12121 nithima.n@arts.tu.ac.th

<https://orcid.org/0000-0002-5665-0737>

4th Areewan Hussadin

Department of Social Sciences Faculty of Liberal Arts Rajamangala University of Technology Thanyaburi Pathumthani, Thailand, 12110 areewan_h@rmutt.ac.th

<https://orcid.org/0009-0007-7763-5206>

5th Chayapol Promaoch

Research Unit in Geospatial Applications (Capybara Geo Lab, Faculty of Liberal Arts Thammasat University Pathumthani, Thailand, 12121 chayapol.palmwork@gmail.com

<https://orcid.org/0009-0007-1753-1536>

6th Tuvachit Chalamkate

Water Distribution Information Technology Division Provincial Waterworks Authority (Head office)

Bangkok, Thailand, 10210 tuvachitchalamkate@gmail.com

<https://orcid.org/0000-0003-4813-6460>

Abstract—Urban Heat Island (UHI) presents significant environmental concerns, exacerbated by urbanization and climate change, resulting in elevated temperatures in urban areas compared to rural ones. This study focuses on Khon Kaen City, Thailand, to enhance UHI analysis through machine learning models, leveraging satellite-derived indices to predict Land Surface Temperature (LST). Employing eXtreme Gradient Boost Regressor (XGBR) and Random Forest Regressor (RFR), we used Normalized Difference Built-Up Index (NDBI), Enhanced Built-up and Bareness Index (EBBI), and Soil-Adjusted Vegetation Index (SAVI) from Landsat 8 images for the years 2013, 2018, and 2023. Our results demonstrate that XGBR outperforms RFR in estimating LST, with higher R^2 values and lower mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Specifically, XGBR showed R^2 values of 0.7788, 0.8039, and 0.7574 for 2013, 2018, and 2023, respectively. The study emphasizes the crucial influence of NDBI and EBBI on predicting LST, showcasing the necessity of including building index variables in UHI research. This method provides important information for urban planners and policymakers. It can facilitate more precise evaluations of UHI trends and aid in developing sustainable urban strategies to counteract the effects of UHI in Khon Kaen City.

Index Terms—urban heat island, eXtreme gradient boost regressor, random forest regressor, indices-based predictor, Khon Kaen City

I. INTRODUCTION

Urban and industrial activities are the main aspects of the Urban Heat Island (UHI), a major environmental issue in urban areas. It can be noticed as the temperature in the city area is higher than in the surroundings. Such an issue may significantly impact energy usage, public health, and quality of life. It is essential to predict and analyze the effects of UHI for effective urban planning and sustainable development.

The land surface temperature (LST) is suggested to be a reliable metric for quantifying the UHI effect across different areas [1]. It can be described by employing remote sensing-derived indices, e.g. the Normalized Difference Built-up Index (NDBI), the Enhanced Built-up and Bareness Index (EBBI), and the Soil Adjusted Vegetation Index (SAVI).

Prior studies have proposed the efficacy of utilizing remote sensing indices in UHI research. For example, NDBI is commonly employed to detect urban areas and their thermal characteristics due to its sensitivity to built-up structures [2]. Meanwhile, EBBI integrates thermal and shortwave infrared bands to improve the differentiation between developed and undeveloped areas. It results in yielding a more nuanced insight into urban thermal properties [3]. Additionally, SAVI adjusts for changes in soil brightness [4]. Such advancement

over the conventional NDVI makes it particularly valuable for urban vegetation analysis.

Machine learning (ML) techniques have recently emerged as powerful tools for environmental modeling. In UHI estimation, the random forest regressor (RFR) advantages multiple decision trees to model the relationship between land LST and various predictors such as land cover, vegetation indices, and built-up area indices [5]. Each tree in the ensemble is trained on a bootstrap sample of the data. The final prediction is obtained by averaging the predictions of all individual trees to enhance the model's robustness and accuracy. The eXtreme gradient boost regressor (XGBR) has gained prominence. It reveals efficiency and superior performance in predictive modeling tasks [6]. Such a model is well-suited for handling large datasets and complex variable interactions. Due to its ability to automatically handle missing data and overfitting, it is ideal for modeling UHI regarding the selected indices.

This article applies multispectral satellite data and cutting-edge ML techniques to analyze UHI dynamics in Khon Kaen City, Thailand. The experiment will be conducted over a 5-year, specifically for 2013, 2018, and 2023. Models' inputs are constructed from the Landsat 8 images. Results will be tested and compared using the coefficient of determination (R^2), mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). By leveraging these advanced methodologies, this study aims to contribute to the existing knowledge of urban climatology and offer practical insights for urban planners and policymakers.

II. STUDY AREA: KHON KAEN CITY, THAILAND

Khon Kaen City is located in the northeastern region of Thailand (Fig. 1). Known for its rapid urbanization and economic development, Khon Kaen is a significant hub for the region's commerce, education, and healthcare. The Muang Khon Kaen reveals 980 km², with the city itself spanning an area of approximately 48 km² and comprises a diverse landscape that includes urban centers, agricultural lands, and water bodies [7].

The area experiences a tropical savanna climate with distinct wet and dry seasons. The average annual temperature ranges from 24°C to 32°C, with the hottest months typically being April and May. The monsoon season, which lasts from May to October, brings most of the annual rainfall, averaging around 1,200 mm annually. This climatic pattern influences the urban heat island effect, making Khon Kaen an ideal location for studying urban temperature variability.

Khon Kaen City has experienced rapid growth and development, becoming one of Thailand's fastest-growing cities over the past few decades. This growth has led to increased land surface temperatures, driven by factors such as the conversion of green spaces to built-up areas and the proliferation of impervious surfaces. The province's rapid development provides a compelling context for analyzing the impact of urbanization on local climate conditions.

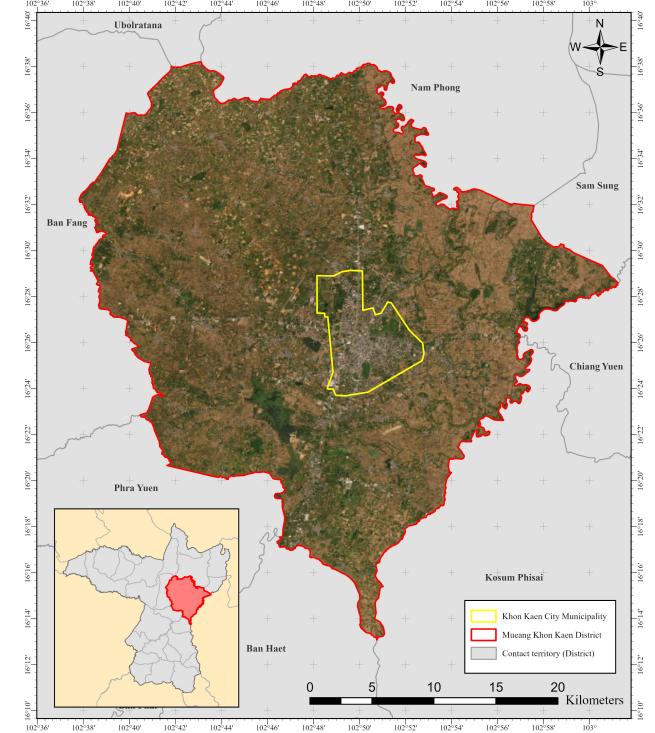


Fig. 1. The yellow boundary presents the study location (Khon Kaen City, Thailand).

III. DEPENDENT AND INDEPENDENT VARIABLES

A. Land Surface Temperature (LST)

The LST was selected as a dependent variable representing the UHI intensity in this experiment. LST can be calculated in three steps, starting from the digital number conversion to radiance in (1).

$$L_\lambda = \frac{L_{\max} - L_{\min}}{Q_{\max} - Q_{\min}} \times (DN - Q_{\min}) + L_{\min} \quad (1)$$

where L_λ refers to spectral radiance at the sensor's aperture (W/(m² sr μm)), L_{\max} and L_{\min} represent the spectral radiance that is scaled to Q_{\max} and Q_{\min} , Q_{\max} and Q_{\min} refer to the maximum and minimum quantized calibrated pixel value (255 and 1), and DN is the digital number value. Next, the spectral radiance was then converted to brightness temperature using the thermal constants from the satellite's metadata as in (2).

$$T_B = \frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)} \quad (2)$$

where T_B refers to at-satellite brightness temperature (K), K_1 and K_2 are the calibration constants 1 (W/(m²·sr·μm)) and 2 (K), and L_λ is a spectral radiance (W/(m²·sr·μm)). Lastly, converting T_B to LST is described in (3).

$$LST = \frac{T_B}{1 + \left(\frac{\lambda \cdot T_B}{\rho}\right) \ln(\epsilon)} \quad (3)$$

where L_λ is the spectral radiance ($\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \mu\text{m})$).

B. Normalized Difference Built-Up Index (NDBI)

NDBI, as defined in (4), is a spectral index used in remote sensing to identify and map built-up (urban) areas [1], [2]. It is beneficial for distinguishing such areas from other land cover types, such as vegetation or water bodies, by utilizing the reflective properties of different surfaces in specific spectral bands.

$$\text{NDBI} = \frac{\text{SWIR} - \text{NIR}}{\text{SWIR} + \text{NIR}} \quad (4)$$

where SWIR is the reflectance value in the shortwave infrared band, and NIR is the reflectance value in the near-infrared band. The NDBI value is ranging between -1 to 1. A value closer to 1 exhibits a build-up area.

C. Enhanced Built-up and Bareness Index (EBBI)

EBBI is a remote sensing index that applies to reflect the built-up and bare land areas. The computation incorporates with NIR, SWIR, and thermal infrared (TIR) bands [8]. The index is explained in (5). It can improve the detection of built-up areas separated from bare lands. It is essential in urban mapping and environmental monitoring [3].

$$\text{EBBI} = \frac{\text{SWIR} - \text{NIR}}{10\sqrt{\text{SWIR} + \text{TIR}}} \quad (5)$$

where TIR refers to the reflectance of the thermal band. A value less than 0.1 reveals bare land, a value between 0.1 and 0.3 represents build-up area, and a value greater than 0.3 exhibits other areas except previously mentioned.

D. Soil-Adjusted Vegetation Index (SAVI)

The SAVI is a modification of the NDVI designed to minimize soil brightness's influence in areas with low vegetative cover [9]. It introduces a soil brightness correction factor, L , to account for the effect of soil reflectance, making it particularly useful in arid and semi-arid regions [4]. SAVI is explained in (6).

$$\text{SAVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red} + L)} \times (1 + L) \quad (6)$$

where L refers to the soil brightness correction factor, typically set to 0.5. and Red represents the reflectance of the red band. SAVI can reduce the discrepancy from NDVI. A high value of SAVI represents dense vegetation cover.

IV. MACHINE LEARNING REGRESSION MODELS AND STATISTICAL TESTING

A. eXtreme Gradient Boost Regressor (XGBR)

XGBR is a model developed from gradient boosting [6], learning from the combination of multiple weak classification and regression tree (CART) learning sets until it evolves into an intense learning set through a boosting learning strategy [10]. XGBR is widely used because it can process very quickly, has high accuracy, and can effectively prevent overfitting through simple functions by combining a loss function

to indicate model compatibility level and a regularization function to control the model's complexity [6], [11]. The adjustment function is defined in (7).

$$\text{Obj}(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (7)$$

where $\text{Obj}(\theta)$ is the model's loss function, L represents loss function between actual values from the experiment (y_i) and estimated values (\hat{y}_i) of each single tree, $\Omega(f_k)$ refers to the model regularization, which can be derived from (8).

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

where γ represents a regularization parameter that controls the number of leaves, λ is a regularization parameter for the leaf weights, T refers to the number of leaves in the tree, and w_j^2 is the sum of the squared weights of the leaves. Besides the regression, such a model is also appropriate to every supervised learning function [6].

B. Random Forest Regressor (RFR)

RFR, one of the machine learning algorithms, divides large datasets into smaller or multiple decision trees using bootstrap samples from the training dataset and randomly selects subsets of independent variables and dependent variables for each node, allowing each tree to operate independently [5]. The final result of RFR is predicted from the averaging outcomes of all decision trees. RFR is an efficient machine learning algorithm as it can prevent overfitting [5], [12], has high accuracy when dealing with complex data, and handles non-linear relationships [13]. The model can be explained in (9) and employed as the baseline of this experiment.

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(\mathbf{x}) \quad (9)$$

where \hat{y} is the predicted value, T refers to the total number of trees in the forest, and $h_t(\mathbf{x})$ represents the prediction of the t^{th} decision tree for the input \mathbf{x} .

C. Statistical Testing

R^2 , MAE, MSE, and RMSE were incorporated to validate the efficiency of each ML model. These metrics help quantify the difference between the actual and predicted values. They also provide insights into the models' performance. Statistical metrics are explained in (10) to (13).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

where n is the total number of observations, y_i refers to the actual value, \hat{y}_i represents the predicted value.

V. RESULTS AND DISCUSSIONS

In the experiment, the models were employed to predict the dynamics of UHI for the years 2013, 2018, and 2023. In the training procedure, hyperparameter tuning was associated with both models using the Python GridSearch library. According to XGBR, hyperparameters were 'learning_rate', 'n_estimators', 'max_depth', 'subsample', and 'colsample_bytree'. In the case of RFR, 'n_estimators', 'max_depth', and 'min_sample_split' were associated. The best combination of hyperparameters and statistical testing are reported in Table I.

The dataset was split into 70:30 for training and testing. The 5-fold cross-validation was employed in the training session to reduce the bias and retrieve the model validation score. After the training, XGBR and RFR results were compared and evaluated using R^2 , MAE, MSE, and RMSE. The analysis from Table II clearly shows that the XGBR model is the most effective for estimating LST. This model consistently delivers better estimation results than RFR, with higher R^2 values observed across all the analyzed years, particularly 0.7788, 0.8039, and 0.7574. The XGBR model demonstrates a significantly greater ability to explain the variance in the dependent variable.

The XGBR and RFR handle data, complexity, and overfitting differently, but both methods are effective [6], [12], [14]. XGBR builds trees sequentially, providing each tree with correcting previous errors and using hyperparameters like learning rate and regularization terms to manage complexity and prevent overfitting. It also uses subsampling and early stopping to enhance generalization [6]. In contrast, RFR relies on a bagging approach, where multiple trees trained on random data samples and feature subsets produce an averaged prediction. Its averaging naturally reduces overfitting without pruning. Therefore, RFR is robust and suitable for noisy data [12].

Additionally, an assessment of the MAE, MSE, and RMSE for 2013 and 2018 indicates that XGBR consistently outperforms RFR, yielding lower yearly values. Such statistical report suggests a more effective reduction in the discrepancy between the actual and predicted LST values, except for 2023, where XGBR shows a slightly higher value. Nevertheless, upon an overall evaluation of both models' performance, it is evident that XGBR is more accurate than RFR, as it adeptly handles the intricacies associated with its loss function, enhancing the model's flexibility and adaptability to the characteristics of the training data [14]. Although RFR operates within statistically acceptable limits, preference should be given to XGBR for improved operational efficiency, demonstrating greater accuracy. Fig 2 displayed the predictions

TABLE I
BEST MODELS' HYPERPARAMETERS TUNING

Model	Hyperparameters	Year 2013	Year 2018	Year 2023
XGBR	'learning_rate'	0.01	0.01	0.01
	'n_estimators'	700	1,000	1,000
	'max_depth'	4	1	1
	'subsample'	1	1	1
	'colsample_bytree'	1	1	1
RFR	'n_estimators'	100	100	500
	'max_depth'	5	5	5
	'min_sample_split'	2	20	10

TABLE II
XGBR AND RFR MODELS' EVALUATION

Years	Models	R ²	MAE	MSE	RMSE
2013	XGBR	0.7788	0.3877	0.2194	0.4684
	RFR	0.7742	0.3780	0.2240	0.4733
2018	XGBR	0.8039	0.4198	0.2728	0.5223
	RFR	0.7549	0.4261	0.2829	0.5318
2023	XGBR	0.7574	0.4835	0.3703	0.6085
	RFR	0.7393	0.4775	0.3619	0.6016

of UHI (2013, 2018, and 2023) of the XGBR. It can be observed that the heat issue in urban areas has expanded.

The feature importance is demonstrated in Fig 3, showing that the NDBI significantly impacts the model's estimation of LST values for 2013, 2018, and 2023, followed by EBBI and SAVI. A strong correlation exists between LST, NDBI, and EBBI, emphasizing the importance of using building index variables in LST prediction. An expansion in building area notably contributes to increased LST, as urban surfaces can absorb solar radiation higher than other surface types [15]. Conversely, SAVI has a relatively minor influence on LST variation due to its unique characteristics that set it apart from the other two indices.

Despite the effectiveness and widespread use of NDBI, it has inherent limitations in accurately distinguishing between built-up areas and open spaces. In response to this challenge, the EBBI, which incorporates the TIR band, has been introduced to enhance the differentiation of albedo and emissivity values between built environments and open spaces. Moreover, the spectral differences captured by SAVI can aid in vegetation classification, thereby improving the identification of built area characteristics [16]. By leveraging these capabilities, the XGBR model performs better at predicting LST values than any single index.

The accuracy of using the Landsat-8 satellite imagery presents some limitations. The spatial resolution may only partially capture fine-scale temperature variation in urban areas with complex structures. Besides, temporal limitations arise because of the interval availability of image acquisition. It may not align with peak urban heat occurrences. Furthermore, several factors constrain the model's performance. The accuracy of both models depends on the quality and relevance of input data (NDBI, EBBI, and SAVI). They may not capture all influential variables related to the UHI.

From the above limitations, additional satellite data would

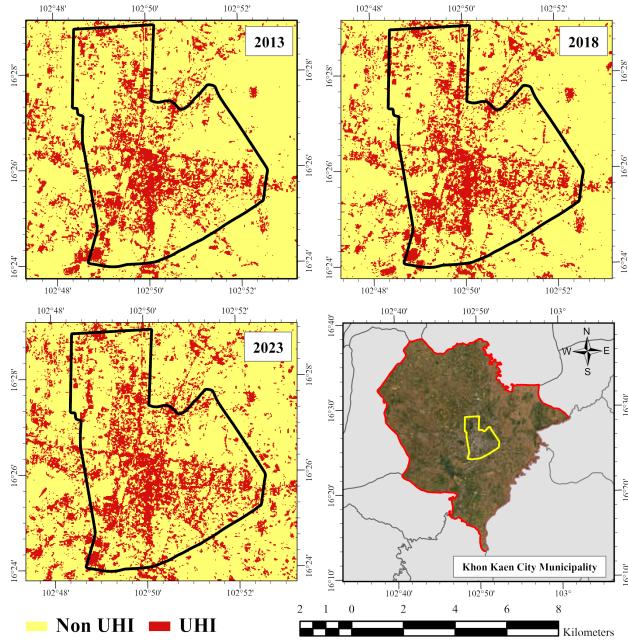


Fig. 2. UHI prediction map using XGBR model

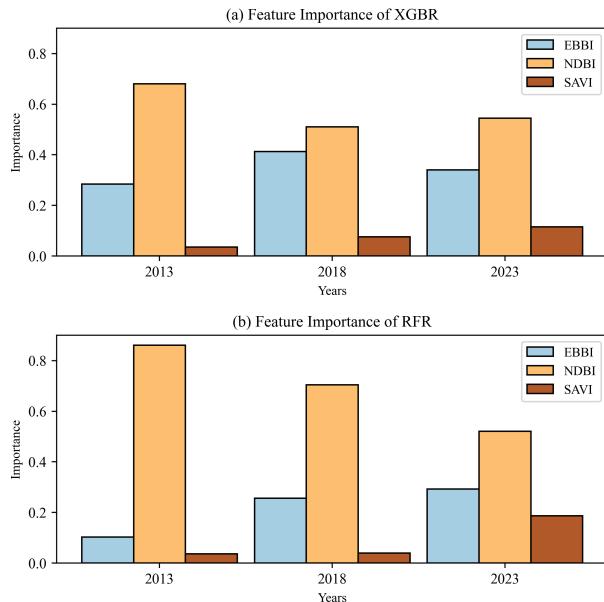


Fig. 3. Feature importance of (a) XGBR and (b) RFR.

be considered to leverage the quality of data input. Exploring more comprehensive ML models may yield and improve the prediction of UHI. Moreover, considering additional coverage indicators or factors such as socio-economic factors or land-use changes could provide a more comprehensive insight into the UHI phenomenon from various perspectives. This thorough understanding could guide local decision-making in the planning and development of Khon Kaen Province, helping to reduce the potential spread of urban heat island effects in the future.

VI. CONCLUSION

This study aims to utilize machine learning techniques in conjunction with indices derived from satellite imagery to estimate LST. The findings reveal that the XGBR model is the most effective in estimating LST in Khon Kaen City when associated with NDBI, EBBI, and SAVI. The integration of spectral capabilities improves the accuracy of LST estimations significantly. Therefore, employing the XGBR model enables a more accurate assessment of UHI formations and supports effective planning for future development in Khon Kaen City.

ACKNOWLEDGMENT

This research was supported by the Center of Excellence in Digital Earth and Emerging Technology (CoE: DEET), Thammasat University, Pathumthani, Thailand, 12120. It was also partially supported by the Faculty of Liberal Arts, Thammasat University, Research Unit in Geospatial Applications (Capybara Geo Lab).

REFERENCES

- [1] M. Jothimani, J. Gunalan, R. Duraisamy, and A. Abebe, "Study the relationship between lulc, lst, ndvi, ndwi and ndbi in greater arba minch area, rift valley, ethiopia," in *3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021)*. Atlantis Press, 2021, pp. 183–193.
- [2] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from tm imagery," *International journal of remote sensing*, vol. 24, no. 3, pp. 583–594, 2003.
- [3] A. R. As-Syakur, I. W. S. Adnyana, I. W. Arthana, and I. W. Nuarsa, "Enhanced built-up and bareness index (ebbi) for mapping built-up and bare land in an urban area," *Remote sensing*, vol. 4, no. 10, pp. 2957–2970, 2012.
- [4] H. Ar, "A soil adjusted vegetation index (savi)," *Remote Sens Environ*, vol. 25, pp. 295–300, 1988.
- [5] D. Jato-Espino, C. Manchado, A. Roldan-Valcarce, and V. Moscardo, "Arcuhi: A gis add-in for automated modelling of the urban heat island effect through machine learning," *Urban Climate*, vol. 44, p. 101203, 2022.
- [6] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [7] C. Keeratisikorn and S. Bonafoni, "Satellite images and gaussian parameterization for an extensive analysis of urban heat islands in thailand," *Remote Sensing*, vol. 10, no. 5, p. 665, 2018.
- [8] R. Y. Indriarti, H. Z. Hadibasyir, and A. T. Sasmi, "Analysis of the relationship between the percentage of residential density and the built-up land index in community environments," in *International Conference of Geography and Disaster Management (ICGDM 2022)*. Atlantis Press, 2023, pp. 392–406.
- [9] H. Ar, "A soil adjusted vegetation index (savi)," *Remote Sens Environ*, vol. 25, pp. 295–300, 1988.
- [10] I. B. Mustapha, M. Abdulkareem, T. M. Jassam, A. H. AlAteah, K. A. A. Al-Sodani, M. M. Al-Tholaia, H. Nabus, S. C. Alih, Z. Abdulkareem, and A. Ganiyu, "Comparative analysis of gradient-boosting ensembles for estimation of compressive strength of quaternary blend concrete," *International Journal of Concrete Structures and Materials*, vol. 18, no. 1, p. 20, 2024.
- [11] C. Xu, G. Huang, and M. Zhang, "Comparative analysis of the seasonal driving factors of the urban heat environment using machine learning: Evidence from the wuhan urban agglomeration, china, 2020," *Atmosphere*, vol. 15, no. 6, p. 671, 2024.
- [12] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.
- [13] L. Han, J. Zhao, Y. Gao, and Z. Gu, "Prediction and evaluation of spatial distributions of ozone and urban heat island using a machine learning modified land use regression method," *Sustainable Cities and Society*, vol. 78, p. 103643, 2022.

- [14] S. Fatima, A. Hussain, S. B. Amir, S. H. Ahmed, and S. M. H. Aslam, "Xgboost and random forest algorithms: an in depth analysis," *Pakistan Journal of Scientific Research*, vol. 3, no. 1, pp. 26–31, 2023.
- [15] P. J. Omar and V. Kumar, "Land surface temperature retrieval from tirs data and its relationship with land surface indices," *Arabian Journal of Geosciences*, vol. 14, no. 18, p. 1897, 2021.
- [16] R. Kaur and P. Pandey, "A review on spectral indices for built-up area extraction using remote sensing technology," *Arabian Journal of Geosciences*, vol. 15, no. 5, p. 391, 2022.

Development of web atlas framework of Cat Tien National Park

Van Ngoc Truc Phuong

*Department of Geodesy, Cartography and Geomatics
HoChiMinh City University of Natural Resources and Environment
HCMC, Vietnam
vntphuong@hcmunre.edu.vn*

Abstract: *Atlas has widely become a means for providing information and raising public awareness in conservation management and practices. Since 2000s, atlases for several protected areas (PAs) have been developed for targeted users and purposes. Each of thematic atlases with different contents, purposes and users, needs its own theoretical foundation for design and creation. This article is to build framework for atlas of Cat Tien National Park (NP) – one of the largest PAs in Vietnam. Based upon socio-ecological system (SES) model, the study firstly synthesizes national and local legislation on management of PAs, and research projects on Cat Tien NP management and sustainable development in buffer zones to localize the SES model. The study begins with secondary data collection and stakeholders interview to define target users, general specification for atlas, and atlas structure. Then experiment was conducted to select GIS technology. Cartographic and GIS analysis were used to develop mathematical principles for pilot maps, create pilot maps and web atlas interface. Simultaneously, in-depth interviews with stakeholders of Cat Tien NP were conducted for verification. The web atlas framework for Cat Tien NP consists of list of atlas users, atlas requirement, a proposal of the atlas structure with 17 chapters, a proposal of map content for 2 pilot chapters, 78 pilot maps (81%), and web atlas interface (consisting of 4 modules to demonstrate the overall atlas structure). The pilot maps and interface are built with ESRI technology. Results show that atlas framework is scientific grounds for building Cat Tien atlas, contributing to raising stakeholder awareness, supporting participatory management at Cat Tien NP. Further studies could be done with other chapters of the frame work of Cat Tien NP or apply the frame work to other PAs. Also, data should also be collect from indigenous ethnic minorities. Lastly, although it is not a focus in this study, multi-criteria analysis in selection of GIS technology is also essential. Although there are limitations, the framework should be applicable to other PAs.*

Keywords: *atlas framework, atlas design, Cat Tien National Park, protected areas, web atlas, socio-ecological system model*

I. INTRODUCTION

Atlas has been recently focused a means for providing spatial information and raising public awareness in conservation management and practices. It is a set of systematic arranged maps associated with supporting elements (such as tables, charts, images and texts). All contents of an atlas are inter-related, composing a unified product [1]. During the last two decades, digital atlases have bloomed, allowing users to interact with maps more easily, including customizing map content and symbology, querying and conducting spatial analysis [2], [3]. During that time, a number of atlases of protected areas (PAs) have been developed, such as Atlas of Pálava protected landscape area [4], [5], Atlas of Tourism Potential of Serbian PAs [6], Atlas of Yellowstone [7], Sequoia and Kings Canyon Park Atlas [8], Marine Protected Areas Atlas of Belize [9], National Atlas of PAs of Iran [10], [11], The bay of Bengal large marine ecosystem [12], Atlas of Swiss National Park (NP) [13], [14], Atlas of Community-Based MPAs in the Philippines [15]. In

these atlas, theoretical foundations are set to optimize atlas usages.

As a country with rich biodiversity, Vietnam has 155 PAs, among which include 33 NPs [16] – the highest priority category of the national PAs system, with typical natural ecosystems, largest area and largest number of species under protection [17]. Efforts have been made to improve participatory management mechanism of the national PAs system, among which is providing information and raising awareness among stakeholders [17], [18]. During the last decade, initiatives on providing spatial information have been kicked-off in forms of electronic atlas (i.e. atlas of Phong Nha – Ke Bang NP [19]) or GIS databases (i.e. Information system for management and monitoring of PAs resources in the Northwest region [20], National Biodiversity Database System [21]). Cat Tien NP is one of the largest PAs in Vietnam, which has great achievements ever since it is applied participatory management mechanism [22]. Therefore, this article is to build the atlas framework of the park.

II. ASPECTS OF ATLAS DESIGN

Procedures of atlas creation can principally be divided into two stages: conceptual design, and production and distribution [1], [23], [24]. The first stage is critical since it shapes the atlas framework. It begins with defining target users of atlas and analyzing their needs (Fig.1). This step influences the form of atlas publication, its content, graphical user interface (GUI) and interactions. Each user group with their common and specific needs requires specific settings including technical level of the content, thematic orientation of the whole atlas and its components, cartographic processing, graphic representation of the whole atlas... Next step is proposing the schedule, budget and structure of the team of executives who run the atlas project. Although the step is vital in production practices, it is outside the scope of this study. A demanding step of the whole atlas project is proposing the concept of the atlas, which states the procedures and rules of the atlas production. The proposal of the concept contains sub-steps which is not necessary to follow one after another:

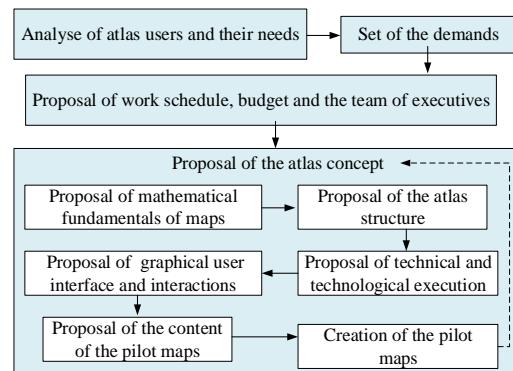


Fig. 1. Steps of atlas design (Source: adapted from [24], [25], [26])

- Mathematical fundamentals of maps - includes defining coordinate system, scale range and the organization of individual components.
- Atlas structure - propose overall structure of the atlas determined by a list of the topics, containing lists of atlas chapters and their sections. To develop atlas structure, it is essential to have an approach for the research objects.
- Technical and technological execution – assesses and choose technique/platform/ software (hereinafter collectively referred to as technique) to create and publish the atlas and its elements, in terms of methods and procedures. It includes GIS method used for creating, updating and entering data entry.
- GUI – sets rules for programming, graphically processing, atlas layout and users' interaction among elements. Main layout contains homepage, content pages, digital maps and other pages. Interaction for all pages and page elements are set.
- Content of maps – elaborates the content of maps in each section, mapping methods and symbology and other components.
- The pilot maps - are created for the verification of the proposal of atlas concept.

III. SOCIO-ECOLOGICAL SYSTEM MODEL

A. A quick review of models of protected areas

The management of PAs have been investigated under several models which are related to integrating PAs with their surroundings. Early in the historical evolution of conservation strategies (1872–1980), PAs were considered as islands. They were untouched natural areas and were established for the sake of preservation. This model can be still found in current conservation conventions. PAs are ecological islands, isolated from the surrounding socio-economic and ecological context. There is no relationship among PAs. The involvement of people is proscribed. Local and indigenous people, their values, knowledge and management systems of the conservation of natural resources are excluded from PAs management PAs policy [25], [26].

Since 1990s, network model has evolved and shifted from static to participatory management in which locals are included in some processes. PAs are connected to each other through the emergence of the concept ecological corridors which foster the movement of species among PAs [27]. Thereby, networks of PAs and ecological corridors were established. However, PAs remain no benefit on local economy and no social connection with surroundings. There are still limitation of local people on accessing ecosystem services, which affects their livelihoods, poverty and leads to social conflicts [25].

Ecosystem model emerged ever since mid-2000s. It encompasses PA conservation, local development and poverty reduction concerns. Human beings are considered as ecosystem components, have significant influence on nature and play crucial role in achieving sustainable ecosystem management. In this model, locals, especially the poor, and their livelihoods should always be considered alongside PAs conservation measures [28]. The concept of buffer zones enables the interaction between a PA (physical, ecological and socio-economic factors) and surrounding landscapes. This

model of management made a radical change from short-term to long-term management which emphasizes the involvement of stakeholders in decision-making [25]. Ecosystem services can provide great values to society. PAs strategies deal with monetizing social values of ecosystem services [29]. The main themes are expressed through the concepts of biodiversity conservation, systems thinking, multiple spatial scales, sustainable development and human component. However, the basic purpose is straight forward: integrated management of natural resources, therefore ecosystem model still places more emphasis on the natural over social aspects [30].

To highlight human intervention and use of PA resources, social-ecological systems (SESSs) model emerged as a modification ecosystem model and is considered as one of the most potentially useful model for PA around the world, particularly in developing countries [31]. SESSs place more emphasis on social factors. On the one hand, human being is the most powerful factor that impact on PA ecosystem. Humans depend on PA ecosystems for population expansion and increase demand for ecosystem-derived products, thereby reducing habitat diversity, harvesting animals or plants, altering disturbance regimes, striving to keep natural populations within pre-defined limits, or constructing fences that limit movement and population expansion [32]. On the other hand, ultimate goals of ecosystem conservation are aiming at and bringing about human well-being. The model reconnects nature and humans by considering not only intrinsic but also instrumental values (ecosystem services). It increases social supports for PA conservation based on values and benefits from PA ecosystem services, by means of laws, policies and other effective measures. That is the concept of ecological consolidations which acknowledges that human beings depend on and benefit from PAs. Hence, social benefits obtained from ecosystem services could enhance current conservation efforts. SESSs focus the need to understand local communities and other stakeholders, and require robust and thorough engagement activities to enhance communication between decision makers, managers and users of PAs resources. Ecosystem services are a common voice for involving stakeholders in decision making process related to PA and surrounding landscapes, thereby reducing social conflicts between conservation and local sustainable development [25]. In SESSs, beneficiaries of ecosystem services are stakeholders directly or indirectly involve (including future beneficiaries) in the valuation, use, or enjoyment of ecosystem services. The incorporating of ecosystem beneficiaries reveals the flow from ecosystem service providers to beneficiaries, which helps to detect the mismatches between the provision and use of ecological services, promote sustainable use of ecological services, and better manage ecological services [29].

The review of models shows that SESSs is most suitable, which helps to balance PA conservation and sustainable local development in countries with limited human resources, high rural population ratio and their livelihood heavily depending on ecosystem services like Vietnam.

B. Socio-ecological system model

1) Definition of socio-ecological system

There are several definitions of SES which generally show that SES is integrated and interlinked system of nature and humans. It is essential that the human part is not a part from but a part of nature. Social and ecological components interact in a constantly evolving and interdependent manner which

makes it impossible to distinguish their boundaries. Nature is the outcome of ecosystem and social system. Depending on the emphasis on one of the two systems, SES is also called coupled system or coupled human-natural system, coupled human-environment system and eco-social system [31], [33], [34],[35]. According to Binder et al (2013), when the emphasis is balancedly placed on the two systems, the complex system is termed SES [36]. Meanwhile, Truong (2014) considers SES is a modification of human systems, emphasizing social and institutional factors [37]. In this study the definition of SES are adopted from Soba and Dwyer (page 4) [31] as: “*A coherent system of biophysical and social factors that regularly interact in a resilient, sustained manner; A system that is defined at several spatial, temporal, and organizational scales, which may be hierarchically linked; A set of critical resources (natural, socio-economic and cultural resources) whose flow and use is regulated by a combination of ecological and social systems; and A perpetually dynamic, complex system with continuous adaptation.*”

2) Structure and functions of socio-ecological system

SESSs are hybrid systems evolved from reciprocal interactions of natural and social systems (green circle of Fig.2). Accordingly, the ecosystem is not simply the overlap between the two systems. Although the natural and social structures and processes are parts of SESSs, its core is formed by the hybridization of the two systems - the result of intricate interactions between different components over space and time (termed as “societal relations to nature”) – has its own ecological-social structures and processes [38]. The natural structures and processes compose of physical environment and their cycles (soil, water, air, climate regime, geology, hydrology) and of biological components (flora, fauna, and microorganisms). These components are intertwined, forming complex ecosystems, from which geographical location and hydro-climate regime shape different types of biological components, creating different ecosystems [31], [38], [39]. Societal structure composes of [31], [38], [39], [40]:

- Population – is the most important component, which includes characteristics of the population)
- Institutions - constitute management mechanism of biodiversity and ecosystem services, representing socially established rules of action. They are self-structuring and structured by political, economic, and social interactions, thereby, forming framework for the use of ecosystem services, including legal and customary laws, management regime...
- Technology used for conservation and ecosystem services - comprises all manmade material structures for specific purposes, intervening ecosystem functions. It refers to physical infrastructures, logistics and other technical elements that are used by stakeholders. Technology plays a critical role in determining the limits of the types of goods and services that can be produced from the existing natural resources of SESSs, thereby, determining the production form and production relations.
- Economic sectors – have special correlation with population. Tiny population are unable to form an economy with full range of economic sectors. The number of economic sectors profoundly affects the exploitation of natural resources of SESSs.



Fig. 2. Components of socio-ecological system (Source: [38], p. 175)

The process of using PA resources does not only include direct “societal relations to nature”. An arbitrary interaction is defined by contextual factors that determine the allocation of natural resources, vulnerability, adaptation, resilience, size and service provision capacity of a PA [38], [39]. Inside the SESSs’ structure and process (the core), knowledge, practices, institutions and technology are contextual factors for the interaction between nature and society.

3) Characteristics of socio-ecological system

Literature review ([25], [29], [31], [32], [33], [34],[35], [36], [37], [38], [39], [40], [41]) shows that different characteristics of SESSs have been mentioned, of which emergent properties of SESSs are commonly analyzed are ecosystem services, spatial and temporal multi-scale, adaptation, resilience, stability, and vulnerability. According to Cumming and Allen (2017) [32] and Bollettin et al. (2017) [41] adaptation, stability and vulnerability can be considered as sub-properties of the resilience. Therefore, the characteristics to be included in this paper are ecosystem services, spatial and temporal multi-scale, and resilience. Ecosystem services are the benefits that people obtain from ecosystems, including provisioning (food, water, raw materials, fuels...), supporting (soil formation and nutrient cycles, primary productivity, nutrient regeneration, species habitats...), regulating (flood and drought regulation, soil erosion and disease prevention), tourism and cultural services (tourism, entertainment, research, religious values). Through ecosystem services, “societal relations to nature” is explicitly expressed, and the management factor obviously becomes the driving force of the interaction between social agents and ecosystem functions (Fig.2).

The PA itself is a spatial entity with clearly defined boundaries on map. Within its boundary, the components of a SES could be separated into elemental structures and processes across spatial and temporal scales: trees/animals by years or decades; habitats by centuries. Outside the PA boundary, different spatial scales are delineated according to the connection between a PA and outside areas that the PA provides ecological services. That could be buffer zones or further areas. Hence, delineation of PA boundaries is in accordance with the management of buffer zones or the socio-economic exchange between PAs [32]. In a word, PAs are networks of SESSs influencing and being influenced by structures and processes at lower or higher spatial and temporal scales (Fig.3). Given the explicitly spatial nature of PAs, intricate interactions are often spatially defined, and require linkages between spatiotemporal scales and institutional levels of PAs management. Their responses can be changed which are either amplify or disrupt the system stability. A small change in a component at a given spatial and temporal scale can lead to non-linear responses to others, or

even to a complete reorganization of the whole system when control thresholds exceed acceptable limits.

Resilience includes stability (the ability of a system to absorb disturbances while maintaining its structures and functions without changing its state of existence), self-organization (mutual supports among the processes and structures of systems, giving the systems the ability to resist being moved away from its stable state), and adaptive capacity (the ability of an ecosystem to adapt to environmental changes through the use of its or other thresholds related to its characteristics). When adaptive capacity is exceeded their limits, systems lose the ability to self-adjust, degrade, or even be destroyed. The adaptation process is complicated by human intervention and cultural factors, through ecosystem services. reciprocal interactions can either reduce or increase the resilience of the ecosystem and the entire SES. Rapid changes can generate new components and dynamics, or eliminate adverse components and interactions to the systems [32]. To measure and evaluate the resilience of the SESSs, huge data set with more than 50 indicators and many different thresholds is needed [32], [41]. These indicators are incomplete. Analysis of the resilience of SESSs in developing countries is not feasible [34]. Therefore, in this paper, the proposal of atlas structure does not fully reflect the aspects of resilience. This property are also not selected for creation of the pilot maps.

IV. METHODS

A. Study areas

Cat Tien NP is the fifth largest NPs in Vietnam. It located in Dong Nai, Lam Dong, and Binh Phuoc provinces and covers the areas of 82,597 ha (Fig.4). The NP has reached great achievements during its application of participatory management mechanism. It is one of the PAs that have reached highest income from ecosystem services (so called payment of forest ecosystem services - PFES) since 2010. Among 80% of forest area of Cat Tien NP are applied PFES, 40% of which is contracted with buffer zones communities, contributing to local socio-economic development in the buffer zones communities including the poor and ethnic minorities [42]. There are 12,541 households participating PFES. Average incomes per household from 2019 PFES in the 3 provinces were 16.2, 12.3 and 3.6 million VND, respectively [43].

B. Methodology

This research is based on interdisciplinary approach since thematic atlas design require knowledge from many different disciplines, including cartography, GIS technology, and disciplines related to the atlas content which is knowledge on PA and related disciplines (Fig. 5). Particularly, it is based on SESSs model. To localize the model, policies, research projects and practices in Vietnam in general and at Cat Tien NP in particular are also integrated. Moreover, opinion of atlas users – that is stakeholders of Cat Tien NP – is critical during developing the web atlas framework as well as its verification.

C. Research methods and procedures

Main research methods are secondary data method as well as interviews. Beside, experiment research method is used to choose GIS technology for atlas of Cat Tien NP. Research procedure is as follows (Fig. 6):

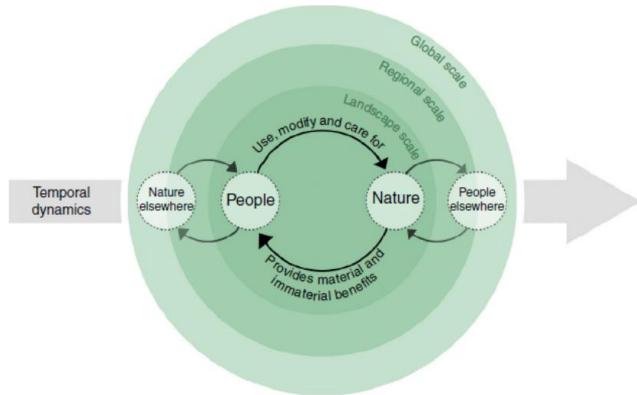


Fig. 3. Spatial and temporal scales of socio-ecological system (Source: [39], p. 18)

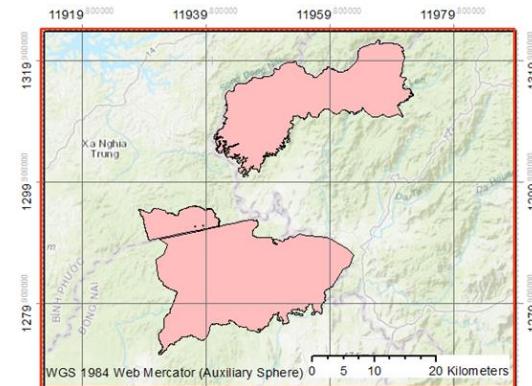


Fig. 4. Study area (source: [44], p 4)

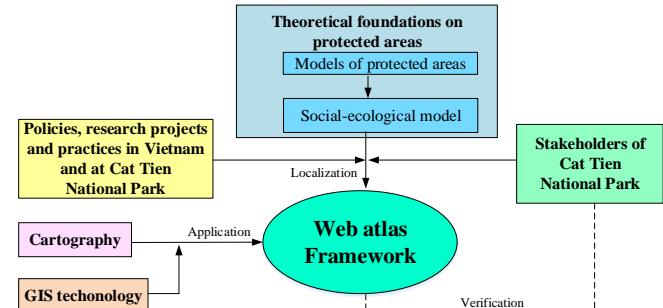


Fig. 5. Research framework

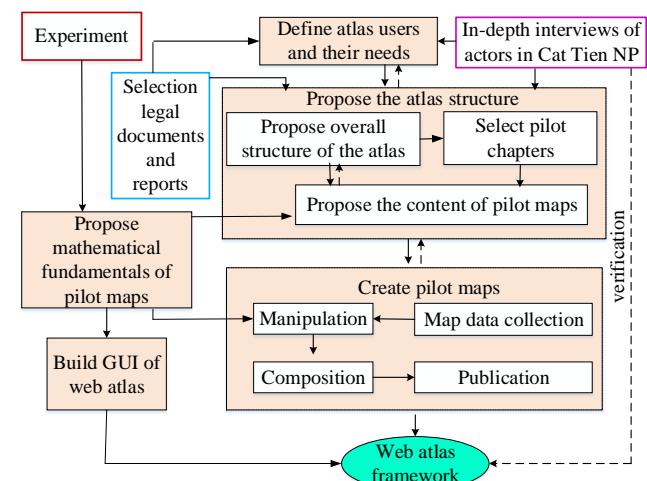


Fig. 6. Research procedure

First, atlas users needed to be defined. Defining atlas users means defining stakeholders of Cat Tien NP, which was based on legal documents and project reports on the organization, management, conservation and sustainable development of the park and its buffer zone. Then, interviews were conducted with actors including 03 officers at Cat Tien NP (Division of Science and International Relations, Division of Forest Ranger), 01 officer at a Commune People's Committee, 01 village's elder in a community, and a member of contracted community group on forest allocation and protection.

Experiment was simultaneously conducted. Technology selection is a multi-criteria analysis issue. In this article, indicators were limited to the feasibility of GIS technologies, i.e. mapping methods and interaction of web map with no-code or low-code added. Three popular GIS technologies in Vietnam as well as in Vietnam forestry sector were used for the experiment, including MapInfo Pro 16.0, QGIS (various versions), and ArcGIS (ArcGIS Desktop10.5, ArcGIS for Server 10.5, ArcGIS Pro 2.5, and ArcGIS online).

The overall structure of the atlas was proposed based upon SES model. Then, pilot maps are selected. According to atlas design principles, pilot maps need to be diversified in terms of content and mapping methods. In this study, we purposely selected pilot maps based on its content, because mapping methods were conducted in the experiment. Hence, two chapters were selected for piloting: Chapter Introduction (to portray an overview of Cat Tien NP), and Chapter Ecosystem Services (to illustrate most prominent emergent properties of SES). For proposal of content of the pilot maps, legal documents, policies on PA as well as research projects in Cat Tien NP and other Vietnam PAs as well as statistical year books of buffer zone districts were used. Legal documents were used to propose the content of pilot chapter (Tab. I). Moreover, actors at Cat Tiên NP were interviewed to verify the proposal of content of the pilot maps. Interviewee included 02 officers at Division of Science and International Relations, 02 officers at Division of Forest Range, 01 officer at a Commune People's Committee, and 01 at Provincial Forest Protection Fund.

Next step was proposing mathematical fundamentals of pilot maps. Scale ranges were defined by spatial multi-scale properties of SESs. The choice of coordinate systems of the maps was one that best compatible with provided base maps of the selected GIS technology. Interaction between webpages and their components had to follow the proposal of atlas structure. Then GUI of web atlas was developed. Web atlas was deployed and managed in a single place. Next, pilot data were collected and manipulated. Pilot maps were then composed and published as webpages and deployed into web atlas. For Chapter Introduction, multi-time GIS data and statistical data were collected (Tab. II and Tab. III). The creation of Chapter Ecosystem Services was described in a separated article ([44]). Finally, the web pilot maps and the GUI were verified through actor interviews. Interviewee were those who had been interviewed for proposing content of the pilot maps.

TABLE I. LEGAL DOCUMENT FOR PILOT CHAPTERS

No.	Documents
1	Biodiversiy Laws (2008, 2018)
2	The Laws of Forest Protection and Development (1994, 2004)

No.	Documents
3	Forestry Law 2017
4	Fishery Laws (2003, 2017)
5	Decision no. 192/2003/QĐ-TTg on the approval of strategy for management of Vietnam PA systems through 2010
6	Decision no. 218/2014/QĐ-TTg on the approval of strategy for management of Vietnam PA systems through 2020, with a vision towards 2030
7	Decision no. 41/TTg/1977 on the approval of establishment of 10 prohibited forests
8	Decision no. 194/CT/1986 on the approval of the category of 73 prohibited forests
9	Decision no. 1171/QĐ/BLN/1986 on the approval of regulations on production forests, protection forests and special-use forests
10	Decision no. 845 -TTg/1995 on the approval of Vietnam's Action Plan for Biodiversity Conservation
11	Decision no. 08/2001/QĐ-TTg on the approval of regulations on management of natural production forests, protection forests and special-use forests
12	Decision no. 62/2005/QĐ-BNN on the approval of regulations on indicators for classification of special-use forests
13	Decision no. 186/2006/QĐ-TTg the approval of regulations on forest management
14	Decree no. 117/2010/NĐ-CP on organization and management of the special-use forests
15	Decree no. 156/2018/NĐ-CP on detailing enforcement of a number of articles of the Forestry Law
16	Decree no. 109/2003 on conservation and sustainable development of wetlands
17	Decree no. 66/2019/NĐ-CP on conservation and sustainable development of wetlands
18	Decree no. 27/2005/NĐ-CP on detailing enforcement of a number of articles of the Fishery Law
19	Decree no. 57/2008/NĐ-CP on regulations on management Vietnam PAs with national and international importance
20	Circular no. 33/2018/TT- BNNPTNT on forest investigation, inventory and monitoring
21	Decree no. 582/QĐ-TTg/2017 on the approval of the category of under poverty hamlets, communes in zone III, II, I of mountainous and ethnic minority areas, period 2016 – 2020
22	Handbook for inspection, monitoring and evaluation of forest ecosystem services for the system of Forest Protection and Development Funds (2020)

TABLE II. GIS DATA FOR CHAPTER INTRODUCTION

No.	Data	Sources
1	Administrative boundaries ^a	Database of global administrative areas
2	PAs boundaries (multi-time) ^a	UNEP-WCMC and IUCN (2021)
3	Functional zones of Cat Tien NP (2010, 2020, 2030) ^b	Sub-institute for Southern Forest Investigation and Planning (2021)

^a GCS, ellipsoid WGS84, shapefiles, ^b VN-2000, Tab files

TABLE III. STATISTICAL DATA FOR CHAPTER INTRODUCTION

No.	Data	Sources
1	Forbidden forests	Decision no. 41-TTg/1977 on regulating forbidden forests; Decision no. 194/CT/1986 on regulating forbidden forests
2	Vietnam PAs until December, 2002	International Centre for Environmental Management, 2003, Vietnam national report on PAs and development
3	Vietnam PAs until March, 2008	Institute for Forest Investigation and Planning, 2008, Category of special-used forest system reviewed after planning to 2020
4	Vietnam planned PAs to 2020	Decision no. 1976-TTg/2014 on approval of special-used forest system planning

No.	Data	Sources
5	Vietnam PAs to 2014	Ministry of Natural resources and Environment, 2014, Vietnam's fifth national report to the United Nations CBD - Reporting period: 2009–2013
6	Vietnam PAs to 2015	Decision no. 1107/QĐ-BTNMT/2015 on category of PAs
7	Management and organization of PAs (multi-time)	Decisions of Prime Ministers or Provincial People Committees; websites of PAs
8	Buffer zone communes of Cat Tien NP	Cat Tien NP, 2010, Planning for conservation and sustainable development of Cat Tien NP from 2010 to 2020.

V. RESULTS AND DISCUSSIONS

A. Results

1) Users of Cat Tien NP web atlas

Atlas users are stakeholders in management of the park, which are divided into three groups:

- Primary stakeholders:

- Management board of the park: includes director board, divisions (Technology and International Relations, Forest Ranger, Finance and Planning, Organization and Administration), centers (Environmental Education and Services; Wildlife Rescue, Conservation and Development), Forest Ranger Stations. Division of Technology and International Relations could be the admin of the web atlas. Others could be analysts.

- Communal agencies in buffer zones (24 communes and 01 town): could be analysts (including Army, People's Committee, Forestry and Police) and viewers (including Women's Union, Youth Union and Farmers' Association).

- Management organizations in buffer zone hamlets: could be viewers of web atlas, including hamlet heads, village elders, forest protection teams, and Women's Union branches

- Forest protection and management teams could be viewers. Forest protection contracted organizations including military command and police of Cat Tien district could be analysts.

- Communities in 71 buffer zone hamlets could be viewers.

- Enterprises benefited from natural resources in the NP could be analysts (including Dong Nai and Saigon water supply companies, and 04 hydroelectric plants) and viewers (including small- and mid-sized enterprises such as travel companies, homestays, sand mining on Dong Nai river)

- Secondary stakeholders: could mostly be analysts

- National government agencies: include direct management agencies in forestry sectors (Ministry of Agriculture and Rural Development, Vietnam Administration of Forestry; Vietnam Forest Ranger Department, Vietnam CITIES management Authority, Vietnam Forest Protection and Development Fund), biodiversity management agencies (Nature Conservation and Biodiversity Administration, Vietnam Environment Administration), management agencies of related issues (Ministry of Finance, Ministry of Science and Technology, Ministry of Science and Technology, Ministry of Culture - Sports and Tourism).

- Provincial government agencies in the 3 provinces: include Provincial People Committees, Departments of Agriculture and Rural Development, Forest Ranger Divisions, Provincial Forest Protection and Development Funds, Dong Nai Culture and Nature Reserve, La Nga Forestry Company, Management board of Protection Forest 600, Departments of Natural Resources and Environment, Departments of Science and Technology, Departments of Culture - Sports and Tourism, Provincial Courts and Procuracies.

- District agencies of 8 districts in buffer zone: include District People Committees, Divisions of Agriculture and Rural Development, Divisions of Natural Resources and Environment, Forest Ranger Sub-Divisions, District Courts and Procuracies. Especially, this group has sub-groups in the roles of viewer including Farmers' Associations, Women's Associations, Youth Unions, Ethnic Affairs

- Research institutes, universities, non-governmental organizations, funding agencies: include Free the Bears, UNDP/GEF/SGP, Join Vietnam – Russia Tropical Science and Technology Research Center, Ramsar Research Center of East Asia, Asian Forest Cooperation Organization, Center for Nature Conservation and Development, Southern Institute of Ecology, Vietnam National University of Forestry at Dong Nai, Monkey World, Pingtung Rescue Center for Endangered Wild Animals, Trust Fund for Endangered Species

- Tertiary stakeholders: could be viewers
- Nearby tourist attraction: Cat Tien Archaeological Site

- Schools and religious institutions in buffer zone

2) Requirement/specification of the web atlas

- Completeness: The atlas needs to fully represent all components of SES and its important characteristics (ecosystem services, multi- spatial and temporal scales, and resilience). The map contents need to be appropriate to atlas purposes and audiences, to be combined both scientific, indigenous and local knowledge.

- Systematic connection of atlas contents: chapters, sections and maps are inter-linked each other, forming a unified product. The content of each map is complementary and interdependent, and closely related to the core content of atlas topics, forming a unified product.

- Simplicity and ease of use: web atlas needs to timely navigate users to the content they want to view, via menus or hotlinks, thereby help them make faster and better decisions. Atlas interface needs to be simple. Interactions are in Vietnamese and are not limit users.

- Accuracy of attribute data: it is a basic requirement of a media product and is an crucial requirement for management purposes.

- Updateability: The atlas update depends on the update of the input data. Most of the contents have input data updated annually. Some are updated on the investigation, monitoring or inventory or census basis. Others are based on scientific research and projects (e.g. number of species, ecotourism, of media publications) would be hard to update.

- Information security: security requirements should follow the principles of providing information in participatory management of biodiversity. Security are conducted via map generalization and applied to each users' group (where necessary).

3) *Proposal of atlas structure:*

Cat Tien NP web atlas divided into three parts (Part 1. Vietnam PAs's system (national scale), Part 2. PAs in Southeast region (regional scale) and Part 3. Cat Tien NP (PA scale)), describing multi-spatial characteristics of SES. Multi-time maps describe the multi-temporal characteristics of SES.

Each part of the web atlas composes of 17 chapters and associated sections, representing SES structure and outstanding emergent properties. Chapters 1 to 7 represent natural system, including (1) Introduction (History evolution of the park, management authorities, functional zones, compartments and buffer zones); (2) Geology – geomorphology (geology, mineral resources, terrain configuration and slope); (3) Climate (general climate, temperature, rainfall and humidity); (4) Hydrology (River Network, Surface Water and Groundwater Resources), (5) Soil (soil taxonomy); (6) Flora (Current status of forest by forest types, forest vegetation types/ forest ecosystems, checklist of flora, floral species, endemic species, rare species and biodiversity indicators); (7) Fauna (species diversity, endemic species, rare species, biodiversity indicators). Chapter 8 to 12 portrait societal structure of the SES, including (8) Population of the buffer zone (size, structure, living standards and poverty); (9) infrastructure of buffer zone (transportation, electricity, water, post and telecommunications, health, culture and education and environmental sanitation); (10) Institutions (legislation, customary law and relevant management agencies); (11) Science and technology of buffer zone (primary production, processing, storage and transport, health care and nature care); (12) Economy of buffer zone (land use, agriculture, forestry, fisheries, industry, cottage industry and services). Chapter 13 is Ecosystem Services which is called forest environmental services (FES) (PAs with FES, FES beneficiaries, Watershed of FES beneficiaries, payment of FES, social and economic impacts of payments of FES). Chapter 14 is Resilience (Environmental Impact, Tourism Carrying Capacity, Water Carrying Capacity, Climate Change Adaptation). Chapter 15 is Ecotourism of the park, including PA self-organizing, leasing and joint ventures. Chapter 16 is Forest management, protection and fire prevention, including forest management; forest protection; forest contract; support for development of the buffer zone; fire watch; control of invasive species; floral conservation and development; fauna rescue, conservation and development; Social media of PA legislation). Lastly, chapter 17 is Natural disasters and Damage.

4) *Experiment results*

The experiment shows the capability of the three technology, among which, ArcGIS meet most of indicators at high level (Tab. 4). Hence, it is use for creation of pilot maps and GUI of Cat Tien web atlas.

5) *Proposal of mathematical fundamentals of pilot maps*

Biodiversity conservation is a complex mission, requiring consideration of ecological, economic, social parameters and the interactions between them. Analysis and evaluation of the SES dynamics, their impacts and relationships for decision making challenge scientists and policy makers. The challenge

is not only related to the context of identification and analysis of the economic - cultural - social - environmental dynamics but also the rational choice of spatial dimension.

TABLE IV. ACCESSMENT OF GIS TECHNOLOGY CAPABILITY

No.	Indicators	GIS technologies		
		MapInfo	ArcGIS	QGIS
Mapping methods				
1	Diversity	Low	High	Medium
2	Aesthetics	Medium	High	Low
3	Time efficiency	Medium	High	Low
4	Data classification	Medium	High	Low
5	Data standadization	High	Medium	Low
6	Symbology (font symbol)	High	Medium	Low
7	Symbology (Color)	Low	High	Medium
Development of Webmap				
8	Diversity of mapping methods	N/A ^a	High	Low
9	GUI		High	Low
10	Interaction		High	Low

^aExperiment cannot be conducted due to copyright

The dynamics of SES exist at different spatial scales according to (1) the functions of the ecological system - the boundary distribution of the objects/phenomena (tree, habitat, ecological system, PA, river basin); (2) administrative boundaries; (3) the participation of spatially discrete locations in the supply chain of PAs' ecological services, reflecting the connection between PAs and outsides; (4) the components of sharing (movement of animals between ecological corridors, ecological linkages of PAs, socio-economic exchanges between PAs...). Among these spatial scales, the scale according to administrative units is important because they are where the contextual elements of SES, especially policies and management tools, are built and implemented. Therefore, administrative boundaries are view extents for thematic multi-scale map. Other spatial scales would be described in the atlas as thematic layers through mapping methods.

The management of Cat Tien NP is decentralized from central to the provincial, thereby, pilot multi-scale maps have 3 view extents. National extent described Vietnam PAs' system with statistics by ecological – economic regions. Southeast extent represents PAs in Southeast Region by provinces. Cat Tien NP extent describes thematic maps within CatTien NP boundary. As such, scale range for pilot map is 1: 30,000,000 – 1:1, 250,000 – 1: 400,000 – 1:40,000 (Fig. 7).

All GIS data are reprojected to WGS 1984 Web Mercator (Auxiliary Sphere) EPSG:3857 – the desired projection of basemap in ArcGIS online.

6) *Proposal of pilot map content*

Thematic layers and their symbology is described at all spatial scales. Symbology transformation is also determined. Thematic layers limited access to users are generalized or replaced by texts, to ensure information security, focusing in Chapter FES. Themes whose data are unavailable are also turned into texts. Details of proposed pilot map content can be viewed in the pilot maps and web atlas interface.

7) *Pilot maps*

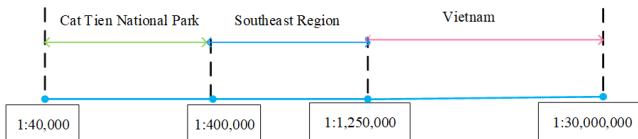


Fig. 7. Scale range of pilot maps

Pilot maps were built by ESRI platform (ArcMap, Arc Pro and ArcGIS Online). Seventy-eight out of 98 proposed pilot maps (81%) were created, due to lack of data (Tab. V). Part 1 and 2 are multi-scale maps. Pilot maps are illustrated in Fig. 8 and Fig. 9. Map interactions include legend, layer adding and ordering, drawing, charting, measuring, query, spatial analysis, sharing and swiping. Some are illustrated in Fig. 10.

8) Web atlas interface

The web atlas interface was developed by ArcGIS Experience. It has 4 modules, including homepage, Part 1 (15 pages), Part 2 (16 pages) and Part 3 (19 pages). Each module (except homepage) includes a map directory page and map pages corresponding to atlas chapters (Fig. 11 and Fig. 12). They are all interlinked for user navigation via menu or interactions.

B. Discussion

Web atlas frame work of Cat Tien NP is the integration of knowledge of PA (including SES model, legislation on or related PA management), cartography and GIS technology.

The process of defining the atlas users and proposal of atlas concept shows that management legislation have changed frequently, so the atlas framework also need to be updated overtime. Besides, in this study GIS technology was chosen through experiment and focus on indicators related to capability of GIS functions. In fact, technology selection could be improved by expert methods with more indicators (operations, information security, costs, human resources...).

The pilot chapters show that proposal of atlas concept is reasonable and feasible. ArcGIS platform is enable mapmakers to create and publish as web maps fast, efficient and intuitive. In terms of web atlas interface, they are all interlinked for user navigation via menu or interactions. Map interaction are customized and in Vietnamese. However, this study is unable to authorize users' groups because the web GUI was developed by ArcGIS personal account.

TABLE V. NUMBER OF PILOT MAPS

Chapters	Sections	No. of pilot maps/ proposed		
		Part 1	Part 2	Part 3
Introduction	Evolution of PA(s)	5/5	1/1	1/1
	Management authority	4/4	1/1	-
	Functional zones	1/4	1/1	-
	Compartments	1/4	1/1	1/1
	Buffer zones	1/4	1/1	1/1
FES	NPs with FES	7/9	-	-
	Beneficiaries of FES	7/7	5/5	0/4
	Payment of FES	7/7	7/7	7/7
	Social impacts of FES	0/2	0/2	0/2
	Economic impacts of FES	2/3	2/3	3/3

Chapters	Sections	No. of pilot maps/ proposed		
		Part 1	Part 2	Part 3
	Environmentl impacts of FES	4/4	5/5	2/3
Total		39/53		24/27
		15/18		

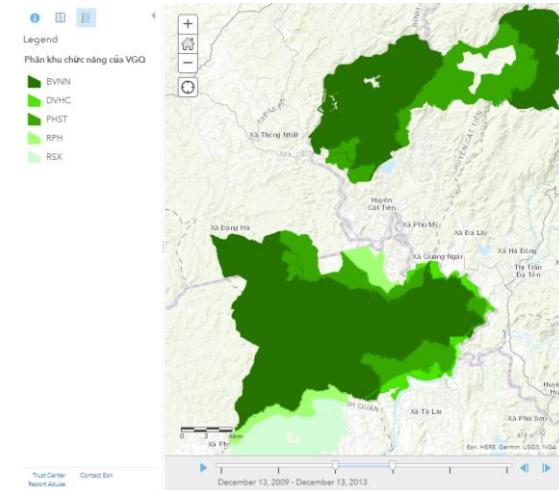


Fig. 8. Map of functional zones of Cat Tien NP

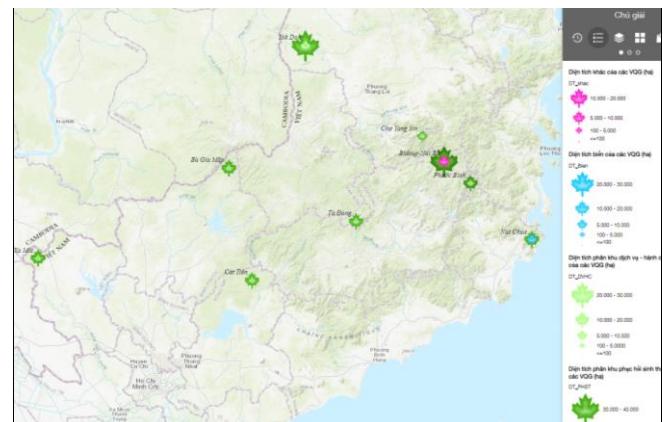


Fig. 9. Map of functional zones of PAs in SouthEast Region

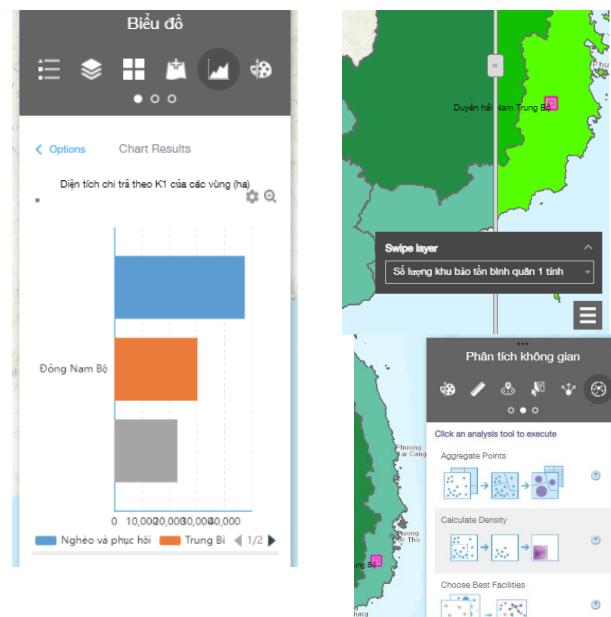


Fig. 10. Illustration of map interaction



Fig. 11. A map directory page

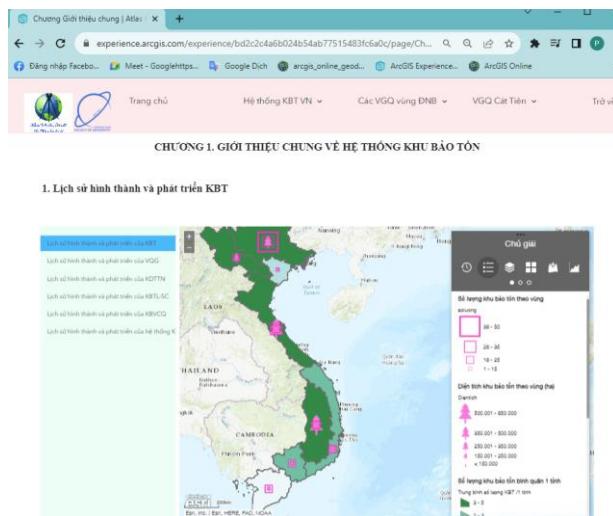


Fig. 12. A map page

In a word, although its limitation, the web atlas framework of Cat Tien NP is scientific grounds, contributing to raising stakeholder awareness, supporting participatory management at Cat Tien NP. It can be also considered as a general framework and applicable for making atlas in other PAs with adjustment.

VI. CONCLUSIONS

Web atlas framework of Cat Tien NP compose of a detailed list of users' groups, specification of atlas, a proposal of atlas structure (17 chapters) and two proposal of content of 2 pilot chapters, 78 pilot maps and a web atlas interface (4 modules and 51 pages). Further studies could be done with other chapters of the framework of Cat Tien NP or apply the framework for other PAs. Also, data should also be collect from indigenous ethnic minorities. Lastly, although it is not a focus in this study, multi-criteria analysis in selection of GIS technology is also essential.

ACKNOWLEDGMENT

I would like to thank Department of Geography – University of Social Sciences and Humanity (HCMC Vietnam National University) for granting ArcGIS account. I also appreciate CatTien NP stakeholders for the interviews and data collection.

REFERENCES

- [1] Lê Minh Vĩnh, “Cơ sở lý luận của atlas điện tử phục vụ quản lý hành chính cấp tỉnh/thành,” Đại học Bách Khoa, 2005.
- [2] A. Briney, “What Is an Atlas?” 2018, [Online]. Available: <https://www.thoughtco.com/what-is-an-atlas-1435685>.
- [3] W. Cartwright, G. Gartner, and A. Lehn, *Lecture Notes in Geoinformation and Cartography*. Verlag Berlin Heidelberg: Springer, 2009.
- [4] J. Miklín and R. Dušek, “Atlas of pálava protected landscape area,” *J. Maps*, vol. 8, no. 4, pp. 492–498, 2012, doi: 10.1080/17445647.2012.749002.
- [5] J. Miklín, “Atlas of Palava protected landscape area,” 2008. [Online]. Available: http://www.janmiklin.cz/atlas_chko_palava/obsah.htm.
- [6] UNDP Serbia, *atlas of tourism potential of Serbian protected areas*. United Nation Development Programme (UNDP), Belgrade, Serbia, 2013.
- [7] W. A. Marcus, J. E. Meacham, A. W. Rodman, and A. Y. Steingisser, “Atlas of Yellowstone,” *University of California Press*, 2017. www.atlasofyellowstone.com (accessed May 12, 2016).
- [8] U.S National Park Service, “Sequoia and Kings Canyon Park Atlas,” 2015. <http://maps2.itc.nps.gov/nps/parkatlas/jsp/atlas.jsp> (accessed Jul. 17, 2018).
- [9] Wildtracks, *Marine Protected Areas Atlas of Belize*. 2016.
- [10] Darvishsefat et al., “Concept of the National Atlas of Protected Areas of Iran and its Realization, Integrating generations: FIG working week,” no. June, pp. 1–9, 2008.
- [11] A. Darvishsefat, “Atlas of Protected Areas of Iran,” 2008.
- [12] “BOBLME-2014-Ecology-15,” 2014.
- [13] H. Haller, A. Eisenhut, and R. Haller, “Atlas du parc national suisse. Les 100 premières années,” *Mt. Res. Dev.*, vol. 35, no. 1, p. 98, 2015.
- [14] C. Schmid, “Atlas du Parc National Suisse - L’extension numérique,” *swisstopo*, 2013. <http://www.atlasnationalpark.ch/fr> (accessed Jul. 17, 2018).
- [15] Haribon Foundation, *Atlas of Community-Based Marine Protected Areas in the Philippines*. Haribon Foundation for the Conservation of Natural Resources, 2005.
- [16] Tổng cục Lâm nghiệp, *Các khu rừng đặc dụng Việt Nam*. NXB Nông Nghiệp, 2021.
- [17] Chính Phủ, “Nghị định 156/2018/NĐ-CP quy định một số điều của luật lâm nghiệp ngày 16 tháng 11 năm 2018.” Hà Nội, 2018, [Online]. Available: <https://luatvietnam.vn/nong-nghiep/nghi-dinh-156-2018-nd-cp-quy-dinh-chi-tiet-thi-hanh-luat-lam-nghiep-169023-d1.html>.
- [18] Tổng cục Lâm nghiệp, “Rừng đặc dụng và phòng hộ Việt Nam 2017 - 2018,” Hà Nội, 2019.
- [19] V. V. Trí, T. X. Mùi, and L. T. P. Lan, “Atlas điện tử Phong Nha - Kẻ Bàng,” 2015.
- [20] H. Q. Quỳnh, L. Q. Tuấn, and T. A. Tuấn, “Nghiên cứu xây dựng hệ thống thông tin quản lý, giám sát tài nguyên ở VQG và một số KBTTN khu vực Tây Bắc bằng công nghệ viễn thám và GIS có sử dụng ảnh VNREDSat-1,” 2016.
- [21] Vietnam Environment Administration, “National biodiversity database systems.” <https://nbds.ceid.gov.vn/> (accessed Aug. 01, 2023).

- [22] G. Giz, "Hồ sơ khu vực Vườn quốc gia Cát Tiên: Quản lý, quản trị và tài chính," 2021.
- [23] T. P. Bui, Q. H. Truong, T. X. Nhữ, and V. T. Nguyễn, "Xây dựng Atlas điện tử dân cư thành phố Hà Nội," *Tạp chí Khoa học Đại học Quốc gia Hà Nội Các khoa học trái đất và Môi trường*, vol. 3, no. Tập 21, pp. 26–37, 2015.
- [24] M. J. Šmídá, "Aspects of electronic atlas creation," 2008.
- [25] I. Palomo *et al.*, "Incorporating the social-ecological approach in protected areas in the anthropocene," *Bioscience*, vol. 64, no. 3, pp. 181–191, 2014, doi: 10.1093/biosci/bit033.
- [26] M. Pimbert and J. Pretty, "Parks, people and professionals: Putting 'participation' into protected area management," Geneva, 1995. doi: 10.4324/9781315066189.
- [27] N. Dudley, *Guidelines for applying protected area management categories*, vol. 46, no. 2. Gland: IUCN, 2008.
- [28] G. Shepherd, *The Ecosystem Approach - Learning from Experience*, no. 5. Gland: IUCN, 2008.
- [29] J. C. Ingram, F. DeClerck, and C. Rumbaitis Del Rio, *Integrating ecology and poverty reduction: The application of ecology in development solutions*. Springer, 2012.
- [30] UNEP, "Ecosystem Approach." 2014, [Online]. Available: <http://www.biodiversitya-z.org/content/ecosystem-approach>.
- [31] M. Perez-Soba and J. Dwyer, "The social-ecological system concept," 2016. doi: 10.1007/978-1-84996.
- [32] G. S. Cumming and C. R. Allen, "Protected areas as social-ecological systems: Perspectives from resilience and social-ecological systems theory," *Ecol. Appl.*, vol. 27, no. 6, pp. 1709–1717, 2017, doi: 10.1002/eap.1584.
- [33] I. Palomo *et al.*, "Incorporating the social-ecological approach in protected areas in the anthropocene," *Bioscience*, vol. 64, no. 3, pp. 181–191, 2014, doi: 10.1093/biosci/bit033.
- [34] G. S. Cumming, *Spatial Resilience in Social-Ecological Systems*. Dordrecht Heidelberg London New York: Springer, 2011.
- [35] Trương Quang Học, "Nghiên cứu và triển khai cách tiếp cận dựa trên hệ sinh thái tại Việt Nam," in *Ký yếu Hội thảo khoa học quốc gia "Tài nguyên thiên nhiên và tăng trưởng xanh,"* 2015, pp. 108–116.
- [36] C. R. Binder *et al.*, "Comparison of Frameworks for Analyzing Social-ecological Systems," *Ecol. Soc.*, vol. 18, no. 4, p. 26, 2013, [Online]. Available: <http://dx.doi.org/10.5751/ES-05551-180426>.
- [37] Trương Quang Học, "Cơ sở sinh thái học cho phát triển bền vững và ứng phó với biến đổi khí hậu." Digital Repository of Vietnam National University, Hanoi, 2014.
- [38] M. Mehrling, B. Bernard, D. Hummel, S. Liehr, and A. Lux, "Halting biodiversity loss: How social-ecological biodiversity research makes a difference," *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.*, vol. 13, no. 1, pp. 172–180, 2017, doi: 10.1080/21513732.2017.1289246.
- [39] M. Hamann, "Exploring connections in social-ecological systems: The links between biodiversity, ecosystem services and human well-being in South Africa," Stockholm University, 2016.
- [40] W. W. L. Cheung, "Chapter 1: Predicting the future ocean: Pathways to global ocean sustainability," in *Predicting Future Oceans*, Elsevier Inc., 2019, pp. 3–15.
- [41] V. Bollettino, T. Alcayna, P. Dy, and P. Vinck, "Introduction to Socio-Ecological Resilience," in *Oxford Research Encyclopedia of Natural Hazard Science*, vol. 1, Oxford University Press, 2017, pp. 1–32.
- [42] T. T. Phạm, T. L. C. Đào, T. L. Nguyễn, T. L. Hoàng, Đ. T. Nguyễn, and H. L. Phạm, "Tác động kinh tế và xã hội của chi trả dịch vụ môi trường rừng tại Vườn Quốc Gia Cát Tiên," 2020.
- [43] Vườn Quốc Gia Cát Tiên, "Ứng dụng công nghệ 4.0 trong chi trả dịch vụ môi trường rừng," Đồng Nai, 2020.
- [44] Van Ngoc Truc Phuong, "Web Mapping for Forest Environmental Services of Protected Areas," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1028, no. 1, pp. 0–8, 2022, doi: 10.1088/1755-1315/1028/1/012009.

Efficiency of Liquid Hot Water Pretreatment of Water Hyacinth on enzymatic hydrolysis

Saksit Imman

*School of Energy and Environment
University of Phayao
Phayao, Thailand
saksit.im@up.ac.th*

Nopparat Suriyachai

*School of Energy and Environment
University of Phayao
Phayao, Thailand
nopparat.su@up.ac.th*

Torpong Kreetachat

*School of Energy and Environment
University of Phayao
Phayao, Thailand
torpong.kr@up.ac.th*

Abstract— Liquid hot water (LHW) pretreatment is a promising chemical-free method for improving the enzymatic digestibility of lignocellulosic biomass, facilitating its conversion into fuels and chemicals within a biorefinery setting. This study investigates the impact of LHW pretreatment on the removal of hemicelluloses and lignin from water hyacinth under various temperature (140–180°C). At a temperature of 160 °C for 15 minutes, LHW pretreatment achieved optimal hemicellulose solubilization into the liquid phase, leading to a maximum pentose yield of 63.5% in the liquid and over 56.1% lignin removal from the solid residue. Additionally, enzymatic hydrolysis of the pretreated biomass using 10 FPU/g of Celluclast™ resulted in a 78.2% glucose yield. Scanning electron microscopy (SEM) showed that the pretreatment disrupted the biomass structure, enhancing the enzyme's ability to access cellulose microfibers. X-ray diffraction (XRD) analysis revealed a higher crystallinity index in the pretreated biomass compared to the untreated, and BET surface area measurements indicated a notable increase in surface area. This study provides valuable insights into how LHW modifies the physicochemical properties of water hyacinth and demonstrates an effective method for processing biomass in the biorefinery industry.

Keywords—Pretreatment, Water hyacinth, Biomass, Enzymatic hydrolysis, Liquid hot water

I. INTRODUCTION

The volatility of crude oil prices and the growing concerns about greenhouse gas emissions driving global warming are major factors prompting the search for renewable resources for fuel and chemical production [1]. Lignocellulosic biomass is emerging as a sustainable and renewable feedstock for the biorefinery industry. This biomass primarily consists of three biopolymers: (1) cellulose, a linear homopolymer of D-glucose organized into a highly crystalline fibrous structure; (2) hemicellulose, an amorphous branched heteropolymer composed of pentoses, hexoses, and sugar acids, serving as a binding agent; and (3) lignin, a heteropolymer of phenolic alcohols (including p-coumaryl, coniferyl, and sinapyl alcohols) that provides structural support and protection to plant cells [2]. These biopolymers form a complex, multi-component structure in the plant cell wall, which is resistant to physical, chemical, and enzymatic breakdown [3].

Pretreatment is a crucial step in the biorefinery process, aimed at removing hemicelluloses and/or lignin, decrystallizing cellulose, and reducing its degree of polymerization. These changes enhance the surface area and digestibility of cellulose microfibers, leading to increased sugar yields during enzymatic hydrolysis. Various chemical, thermal, and biological pretreatment methods have been studied extensively for woody biomass and agricultural by-products, each with its own set of advantages and limitations concerning technical and economic factors. Among these methods, autohydrolysis using liquid hot water (LHW) is considered highly effective and cost-efficient compared to other pretreatment technologies like diluted acid, lime, ammonia fiber explosion, and ammonia recycling percolation [4]. LHW uses water as the sole solvent at high temperatures and pressures. Under these conditions, water ionizes to produce hydronium ions, which facilitate the release of acetic acid from hemicelluloses [5]. This acid environment auto-catalyzes hemicellulose solubilization and carbohydrate degradation, enhancing cellulose accessibility and separating hemicelluloses from the biomass. Water hyacinth, a prevalent aquatic plant, offers significant potential as a biomass feedstock. Despite its abundance and rapid growth, it often causes environmental issues due to its invasive nature. Various pretreatment methods, including chemical (e.g., acid and alkali) and physicochemical techniques (e.g., steam explosion and CO₂ explosion), have been explored to improve enzymatic digestibility, each varying in efficiency and cost-effectiveness related to chemical and energy consumption [6]. However, the application of chemical-free LHW pretreatment to water hyacinth has not been reported. This study investigates the use of LHW to enhance the digestibility of water hyacinth and examines its effects on the biomass's physical and chemical properties. The findings provide valuable insights into the potential of LHW as an environmentally friendly pretreatment strategy for this promising biorefinery feedstock.

II. MATERIALS AND METHODS

A. Materials

Water hyacinth was sourced from Phayao Province, Thailand. Water hyacinth was initially processed using a cutting mill (Retsch SM2000, Hann, Germany) and then sieved to achieve a particle size range of 250–420 µm (0.21–0.35 mesh). This processed biomass was subsequently used for experimental studies. Analysis of the biomass revealed the

following composition on a dry-weight basis: 40.6% cellulose, 26.0% hemicellulose, 17.3% lignin, and 8.7% ash, as determined by the standard NREL method [7].

B. Liquid Hot Water Pretreatment

Pretreatment was carried out using a multi-reactor system consisting of six 50-mL reactors housed in a temperature-controlled jacket equipped with a vertical shaking mechanism to ensure optimal mixing. Each stainless steel reactor was fitted with a thermocouple to monitor the internal temperature. Water hyacinth was subjected to pretreatment with deionized water under various conditions, including temperatures ranging from 140 to 180 °C, a heating rate of 20 °C/min, and residence times of 5, 10, and 20 minutes. The substrate concentration was maintained at 10% (w/v), and the initial pressure was set at 25 bar using nitrogen. After reaching the desired conditions, the reactor was rapidly cooled in a water bath. The pretreated solid was then separated from the liquid phase through filtration and thoroughly washed with tap water using a Büchner funnel. The resulting sample was dried at 60 °C before being subjected to enzymatic hydrolysis to assess the efficiency of the pretreatment.

C. Enzymatic Hydrolysis

The efficiency of pretreatment was evaluated by examining the digestibility of the pretreated water hyacinth residues using a commercial cellulase enzyme. Hydrolysis reactions were conducted with a total volume of 1 mL containing 5% (w/v) of the pretreated substrate. The reactions were carried out with 10 FPU/g of Trichoderma reesei cellulase (Celluclast™ 1.5L; Novozymes AS, Bagsværd, Denmark), supplemented with 330 IU/g of Aspergillus niger β-glucosidase (Novozym 188; Novozymes AS) and 120 IU/g of Humicola insolens endoxylanase (Optimash® BG; Danisco AS, Copenhagen, Denmark), all in a 50 mM sodium acetate buffer at pH 5.0. The reactions were incubated at 50 °C for 72 hours with vertical mixing at 30 rpm. Each experiment was performed in triplicate. Cellulase activity was measured in terms of filter paper units (FPU) following a standard method [8]. β-Glucosidase activity was determined using p-nitrophenyl-β-D-glucopyranoside as the substrate [9], with one international unit (IU) defined as the amount of enzyme that produces 1 μmol of reducing sugar or p-nitrophenolate per minute. The fermentable sugar profiles were analyzed using a Waters e2695 high-performance liquid chromatograph equipped with a differential refractometer and a Bio-Rad Aminex HPX-87P column [10]. Sugar yields were calculated based on the sugars obtained from the enzymatic hydrolysis of the solid residues, expressed on a dry-weight basis. Sugar recoveries were reported as the percentage of glucose or pentose recovered relative to the available cellulose ($\times 1.11$) and hemicellulose ($\times 1.13$) in the native biomass.

D. Scanning Electron Microscopy Analysis

The native and pretreated water hyacinth microstructures were investigated via scanning electron microscope (SEM) using a JSM-6301F Scanning Electron Microscope (JEOL, Tokyo, Japan). The samples were dried and coated with gold for analysis. An electron beam energy of 5 kV was applied for analysis.

E. X-Ray Diffraction Analysis

Crystallinity of the native and pretreated water hyacinth was explored via x-ray diffraction using an X' Pert PRO diffractometer (PANalytical, Almelo, The Netherlands). The

pretreated samples were scanned in a range of $2\theta = 10-30^\circ$ with a step size of 0.02° at 500 kV, 30 mA and radiation at Cu K α ($\lambda = 1.54 \text{ \AA}$). Crystallinity was determined according to the following equations for crystallinity index (Eq. 1) [11].

$$\text{CrI} (\%) = \frac{I_{002} - I_{\text{amorphous}}}{I_{002}} \times 100 \quad \text{Eq. 1}$$

in which I_{002} is the intensity for the crystalline portion of biomass (i.e., cellulose) at $2\theta = 22.4$ and $I_{\text{amorphous}}$ is the peak for the amorphous portion (i.e., cellulose, hemicellulose, and lignin) at $2\theta = 18.0$.

F. BET Surface Area Measurement

The method of Brunauer, Emmett, and Teller (BET) was selected to exhibited the total surface area of pretreated water hyacinth. Raw and pretreated water hyacinth were determined for the BET surface area using a Belsorp-max TPDpro (BEL Japan, Tokyo, Japan) with thermal conductivity detector (Semi-diffusion type, 4-element W-Re filament) at the University of Phayao, Thailand.

III. RESULTS AND DISCUSSION

A. Composition of water hyacinth

The native composition of water hyacinth before pretreatment were determined according to the standard methods of the National Renewable Energy Laboratory (NREL, USA) [17]. The chemical compositions of water hyacinth used as starting material in this study was found to contain $40.6\% \pm 0.25\%$ cellulose, $26.0\% \pm 0.28\%$ hemicellulose, $17.3\% \pm 0.26\%$ lignin, $8.7\% \pm 0.26\%$ ash and others on dry basis (Fig. 1). Carbohydrates accounted for about 66.6% (as cellulose and hemicellulose) of the dry material was found for the water hyacinth. The hemicellulose part consisted of 95% xylose and small amounts of arabinose. There were also 8.2% other components which are chemically bound water and ethanol soluble materials and proteins [18]. The differences in physical structures and chemical composition thus make water hyacinth a biorefinery feedstock likely to show unique characteristics in response to pretreatment and susceptibility to enzymatic hydrolysis.

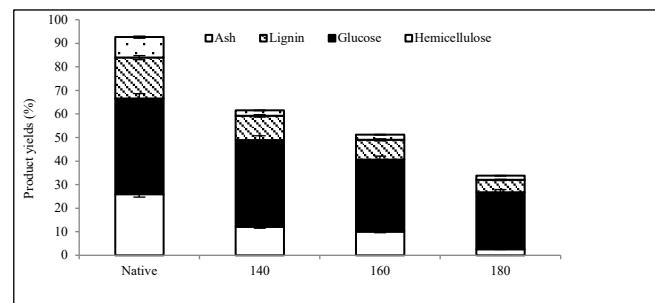


Fig. 1 Composition of water hyacinth

B. Liquid hot Water Effect on Sugar Releases in Aqueous phase

The effects of liquid hot water pretreatment of water hyacinth were explored by analysis of sugars and degraded products in the aqueous phase after pretreatment. Liquid hot water pretreatment showed extensive solubilization of hemicellulose in the native water hyacinth. An observed trend indicated that as pretreatment temperature and time increased, the concentration of sugars, especially pentoses like xylose and arabinose, also rose in the aqueous phase. The

highest pentose yield was observed at 140 °C, with a residence time of 15 minutes. These conditions produced 12.1% of hemicellulose yield from the native substrates under the conditions that maximized pentose release (Fig. 2).

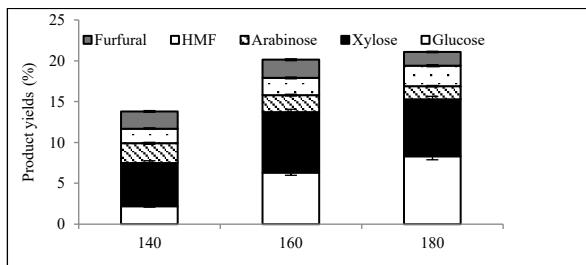


Fig. 2 Product yield obtained in aqueous phase

C. Enzymatic hydrolysis

The impact of LHW pretreatment on the digestibility of water hyacinth was assessed by measuring the sugars released following enzymatic hydrolysis. The findings revealed that both the temperature and duration of pretreatment positively influenced the digestibility of the solid residues. Maximum glucose yields were observed at 160 °C, with residence times of 15 minutes. These conditions resulted in glucose yields of 425.4 mg/g compared to pretreated water hyacinth, after 72 hours of enzymatic hydrolysis. These yields corresponded to 82.3% of the glucose yield from the native water hyacinth and were 5.6 times higher than those from non-pretreated biomass. For pentoses, higher yields were obtained under milder conditions (160°C), indicating a greater hydrolysis potential of hemicelluloses. These conditions produced pentose yields of 122.1 mg/g of pretreated substrate, translating to 38.2%, of pentose yield from native water hyacinth. These results were 1.8 times higher than those from non-pretreated biomass. Overall, higher temperatures and longer pretreatment times generally enhanced the digestibility of the biomass.

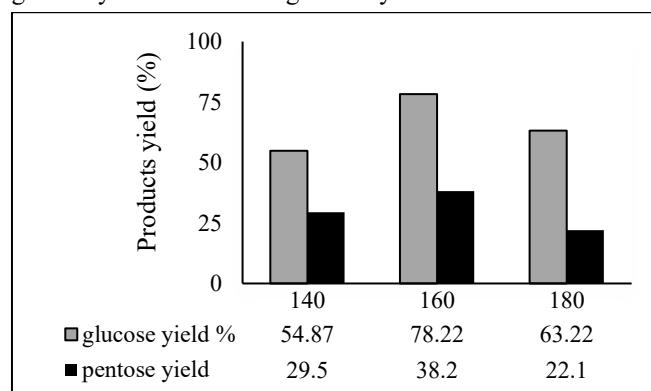


Fig. 3 Sugar released from enzymatic hydrolysis

D. Characterization of pretreated water hyacinth

The physical structure of water hyacinth pretreated with LHW under optimal conditions for maximum sugar yields was examined using SEM. In their native state, water hyacinth exhibited a smooth surface without visible cracks or cavities. However, after LHW pretreatment, the microstructure of the water hyacinth became disrupted, leading to the removal of hemicellulose and surface lignin. This disruption caused the surface lignin to peel away, increasing the accessibility of the inner cellulose microfibers, which is linked to enhanced enzymatic digestibility [19]. These structural changes are consistent with those observed

in various agricultural residues subjected to different pretreatment methods [20].

Crystallinity and surface area of the pretreated water hyacinth are crucial factors that influence the effectiveness of enzymatic hydrolysis. A significant increase in crystallinity (CrI) (63.7%) was observed in water hyacinth pretreated with LHW at optimal conditions (160 °C for 15 minutes) compared to the native biomass (59.3%). The rise in CrI is attributed to the LHW pretreatment's ability to remove amorphous xylan and lignin fractions while having a lesser impact on the disruption of highly crystalline cellulose. Similar increases in crystallinity have been reported in various biomasses pretreated with LHW, both without [13] and with acid/alkali promoters [21], as well as other methods like aqueous ammonia pretreatment [22]. LHW pretreatment also led to an increase in surface area compared to the native water hyacinth (2.3 m²/g). The maximum surface area observed was 4.9 m²/g under the conditions of 160 °C for 15 minutes (Table 1). This increase in surface area enhanced enzyme accessibility to cellulose fibers, resulting in greater enzymatic susceptibility of the solid fraction and the highest glucose yield from hydrolysis of the biomass treated under these conditions [23].

TABLE I. BET SURFACE AREA AND CRSTALLINITY INDEX OF NATIVE AND PRETREATED WATER HYACINTH

Biomass	Surface area (m ² /g)		Degree of crystallinity (%)	
	Native	Pretreated	Native	Pretreated
Water Hyacinth	2.3	4.9	59.3	63.7

E. Conclusion

The Liquid hot water pretreatment has proven to be an effective approach for processing water hyacinth. This method resulted in pentose and glucose yields of 38.2% and 82.3%, respectively, when combining the sugars obtained from enzymatic hydrolysis of the solid fraction and the liquid from the pretreatment step. Additionally, the formation of inhibitory by-products was minimal, remaining below levels that would inhibit ethanogenic yeasts. The improvement in water hyacinth digestibility was linked to the removal of non-cellulose components and an increase in the accessible surface area of the biomass. This study highlights the potential of Liquid hot water pretreatment as an effective method for enhancing the conversion efficiency of water hyacinth in sugar platform biorefineries.

ACKNOWLEDGMENT

Saksit Imman was financially supported by Unit of Excellent (UOE219/2567) from University of Phayao.

REFERENCES

- [1] P. Alvira, E. Tomás-Pejó, M. Ballesteros Perdices, and M.J. Negro, "Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review," *Bioresour. Technol.*, vol. 101, pp. 4851–4861, July 2010.
- [2] T. Tianwei, and H.X. Jian, "Biorefinery: an Efficient Way to Sustainable Development of Chemical Industry—a Special Issue for International Conference on Biorefinery (ICB 07) and the 5th International Conference on Separation Science and Technology

- (ICSST2007)," *Appl. Biochem. Biotechnol.*, vol. 160, pp. 301-302, December 2009.
- [3] B.C. Saha, B.I. Loren, A.C. Michael, and Y.V. Wu, "Dilute acid pretreatment, enzymatic saccharification, and fermentation of rice hulls to ethanol," *Biotechnol. Prog.*, vol. 21(3), pp. 816-822, May 2005.
- [4] J.T. Mohammad, and K. Keikhosro, "Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review," *Int. J. Mol. Sci.*, vol. 9(9), pp. 1621-1651, October 2008.
- [5] F.M. Girio, C. Fonseca, F. Carvalheiro, L.C. Duarte, S. Marques, and R. Bogel- Lukasil, "Hemicelluloses for fuel ethanol: a review," *Bioresour. Technol.*, vol. 101(13), pp. 4475-800, July 2010.
- [6] A.T.W.M. Hendriks, and G. Zeeman, "Pretreatments to enhance the digestibility of lignocellulosic biomass," *Bioresour. Technol.*, vol. 10(1), pp. 10-18, January 2009.
- [7] A. Sluiter, B. Hames, B. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, and D. Crocker, (2004). Laboratory analytical procedure (LAP). Golden: National Renewable Energy Laboratory (NREL).K. Eves and J. Valasek, "Adaptive control for singularly perturbed systems examples," Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
- [8] J.A. Perez, I. Ballesteros, M. Ballesteros, F. Saez, M.J Negro, and P. Manzanares, "Optimizing Liquid Hot Water pretreatment conditions to enhance sugar recovery from wheat straw for fuel-ethanol production," *Fuel*, vol. 87, pp. 3640-3647. December 2008.
- [9] G.L. Miller, "Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar," *Anal. Chem. Lab.*, vol. 31(3), pp. 426-428. March 1995.
- [10] H. Piyanun, C. Verawat, S. Warasirin, and E. Lily, "A thermotolerant β -glucosidase isolated from an endophytic fungi, *Periconia* sp., with a possible use for biomass conversion to sugars," *Protein Expr. Purif.*, vol. 67(2), pp. 61-69. October 2009.
- [11] R. Sindhu, M. Kuttiraja, P. Binod, K. U. Janu, R. K. Sukumaran, and A. Pandey, "Dilute acid pretreatment and enzymatic saccharification of sugarcane tops for bioethanol production," *Bioresour. Technol.*, vol. 102(23), pp. 10915-10921, December 2011.
- [12] T.-C. Hsu, G.-L. Guo, W.-H. Chen, and W.-S. Hwang, "Effect of dilute acid pretreatment of rice straw on structural properties and enzymatic hydrolysis," *Bioresour. Technol.*, vol. 101(13), pp. 4907-4913, July 2010.
- [13] S. Kim, and M. T. Holtzapple, "Effect of structural features on enzyme digestibility of corn stover," *Bioresour. Technol.*, vol. 97(4), pp. 583-591, March 2006.
- [14] W. E. Kaar, and M. T. Holtzapple, "Using lime pretreatment to facilitate the enzymic hydrolysis of corn stover," *Biomass. Bioenergy*, vol. 18(3), pp. 189-199. March 2000.
- [15] M. L. Soto, H. Domínguez, M. J. Núñez, and J. M. Lema, "Enzymatic saccharification of alkali-treated sunflower hulls," *Bioresour. Technol.*, vol. 49(1), pp. 53-59, 1994.
- [16] F.A. Patricia, and G. Rosana, "Fractionating process of lignocellulosic biomass for the enzymatic production of short chain cello-oligosaccharides," *Industrial. Crops. and Products*, vol. 178, pp. 114671, April 2022.
- [17] Q. Chu, X. Li, Y. Xu, Z. Wang, J. Huang, S. Yu, and Y. Qiang, "Functional cello-oligosaccharides production from the corncobs residues of xylo-oligosaccharides manufacture," *Process. Biochem.* Vol 49(8), pp.1271-1222, August 2014.
- [18] K. Yoshinori, K. Yutaka, O. Gentaro, and S. Takashi, "Production of cello-oligosaccharides by enzymatic hydrolysis in the presence of activated carbon," *Enzyme. Microb. Technol.*, vol. 12(1), pp. 72-75. Jauary 1990.
- [19] G.C. Alana, L.Eduardo, C. Karina, and P. Patricia, "Bacterial cellulose from kombucha: Assessing inoculum age and concentration, and its conversion via enzymatic hydrolysis into cellobiose and glucose," *Biocatal. Agric. Biotechnol.*, vol. 59, pp.103244. July 2024.
- [20] S. Chakravorty, S. Bhattacharya, A. Chatzinotas, W. Chakraborty, D. Bhattacharya, and R. Gachhui, "Kombucha tea fermentation: Microbial and biochemical dynamics," *Int. J. Food Microbiol.*, vol. 220, pp. 63-72. March 2016.
- [21] D. Sartika, A. P. Firmansyah, I. Junais, I. W. Arnata, F. Fahma, and A. Firmanda, "High yield production of nanocrystalline cellulose from corn cob through a chemical-mechanical treatment under mild conditions," *Int. J. Biol. Macromol.* Vol. 240, pp. 124327. June 2023.
- [22] P. Patchiya, G. Guaqing, M. Yufei, H. Xiaozgang, and A. Abuliti, "Effect of ball milling on the production of nanocellulose using mild acid hydrolysis method," *J. Taiwan Inst. Chem. Eng.* Vol. 60, pp. 617-622. March 2016.
- [23] J. Nurain, A. Ishak, and D. Alain, "Extraction, preparation and characterization of cellulose fibres and nanocrystals from rice husk," *Industrial. Crops. and Products*, vol. 37(1), pp.93-99, May 2012.

Enhancing Air Quality with an Innovative IoT-Integrated Hybrid Air Pollution Control System

Wattana Loungon

School of Energy and Environment

University of Phayao

Phayao, Thailand

wattana6446@gmail.com

Torpong Kreetachat

School of Energy and Environment

University of Phayao

Phayao, Thailand

Torpong.kr@up.ac.th

Saksit Imman

School of Energy and Environment

University of Phayao

Phayao, Thailand

Saksit.im@up.ac.th

Nopparat Suriyachai

School of Energy and Environment

University of Phayao

Phayao, Thailand

Nopparat.su@up.ac.th

Surachai Wongcharee

Faculty of Engineering

Mahasarakham University

Mahasarakham, Thailand

surachai.w@msu.ac.th

Preut Thanarat

Faculty of Informatics

Mahasarakham University

Mahasarakham, Thailand

Preut.t@msu.ac.th

Kowit Suwannahong *

Faculty of Public Health

Burapha University

Chonburi, Thailand

kowit.su@gobuu.ac.th

*Corresponding author

Abstract—

Particulate matter (PM) of 2.5 is the most severe air pollution problem facing the waste processor industry. Integrating Internet of Things (IoT) technology into hybrid wet scrubbers in combustion processes represents a significant leap forward in environmental monitoring and management. IoT enables the instantaneous tracking and regulation of waste processor factories in air pollution control, with the wet scrubber's adsorption system allowing for swift adjustments to changing conditions, thereby enhancing pollutant elimination's effectiveness. The removal efficiency of particulate matter of 2.5 microns (PM2.5) using an innovative air pollution system with an IoT system for PM2.5 real-time monitoring was investigated. IoT's remote monitoring and management capabilities allow for the flexible control of air pollution processes from any location with internet access, offering a promising future for air quality control. This research aimed to an effective IoT system for air pollution control and to create an effective air pollution control system, define objectives and requirements, select appropriate air quality sensors, deploy them strategically, establish a robust communication infrastructure, collect and store data, develop real-time monitoring and alerts, analyze data, and create visualizations. Integrate external data sources like weather information and traffic patterns to understand pollution levels. Implement control mechanisms based on analysis to mitigate pollution. Develop a user interface and public awareness to educate the public about air quality. Ensure compliance with local and national air quality monitoring and control regulations. Regularly maintain and calibrate sensors, plan system upgrades, and establish a feedback loop for continuous improvement with sustainable Health.

Keywords— *Air Quality, IOT, PM 2.5, Hybrid Air Pollution Control system*

I. INTRODUCTION

Increasing economic growth due to population growth has led to the extravagant use of limited resources in production and consumption processes, resulting in significant waste.

Additionally, this has impacted global temperature changes, causing a climate crisis or Climate Change. Campaigns encouraging consumers to change their behavior using the 3Rs principle—"Reduce, Reuse, and Recycle"—are insufficient for change. Therefore, producers also need to adjust.

The national development plan of Thailand has specified the "20-year national strategy", a long-term national development strategy. The objective is to improve the quality of the country in all sectors. The pilot area of the Eastern Economic Corridor (EEC) development project has been established, focusing on the development of 3 provinces in the eastern region: Rayong, Chonburi, and Chachoengsao, as these provinces were already one of the country's top industrial zones. This is to improve the country's competitiveness based on modern technology and innovation (Eastern Economic Corridor Policy Committee, 2019). Thailand has prepared the 13th National Economic and Social Development Plan, which has the main development goals for five years (2023-2027). The goals are clear and consistent with the national development plan. The sufficiency economy principle has been used to determine the direction of national administration and link it to the world's sustainable development (Sustainable Development Goals: SDGs). However, it cannot be denied that when the economy and technology grow, a large amount of waste follows. In many manufacturing industries, the principles of the BCG Model are applied, which is a holistic economic development that focuses on using biological resources to create value (Bioeconomy) together with the circular economy (Circular Economy), which will be under the green economy (Green Economy), a policy that the industrial sector jointly promotes to create economic development and enhance sustainable development, as well as develop society together with environmental conservation to be balanced (National Science and Technology Development Agency, 2021). One of the crucial principles of the BCG Model used in this study is the

circular economy system. The principle of the circular economy system, according to the Circular Economy, is a circular economy system that focuses on preserving the value of resources by using them for maximum benefit and reducing the release of waste and pollution. In the manufacturing industry, such principles focus on Zero Waste, which is reducing the amount of waste to zero. There will be an adjustment of the production process structure that relies on each other between industrial plants to create a balance between humans and resources.

During the cutting process, it was found that a significant amount of smoke and sparks were generated. Although the working environment is an open building, it was observed that the smoke ventilation was insufficient compared to the amount of smoke produced. Therefore, the objective of the research. To design and install an air pollution treatment system for PM2.5 generated using the Internet of Things system during the steel-cutting process. (IOT) for control and monitoring of the air pollution control system. The air pollution control treatment system is a hybrid wet scrubber with an activated carbon adsorption system enhanced to remove PM2.5 produced during processes. It ensures that the air released outside meets the standards set by the government and laws while also preventing disturbances to the community.

II. MATERIAL AND METHOD (AS SEEN IN FIG 1.)

A. Designing the Layout and Equipment for Installing an Air Pollution Treatment System

The initial phase involves designing the layout and selecting appropriate equipment for installing an air pollution treatment system. The design process focuses on ensuring the optimal configuration of the Hybrid wet scrubber with an adsorption system [1,2], considering factors such as site-specific pollution levels, space availability, and operational requirements. Special attention is paid to selecting sensors and actuators that will effectively monitor and regulate the system's performance, particularly in controlling PM2.5 emissions. Integrating IoT-enabled components during this stage is critical for real-time monitoring and system optimization.

B. Installation and Operation of the Air Pollution Treatment System

Once the design is finalized, the equipment is installed according to the planned layout. This includes mounting sensors for gas, particulate matter, temperature, and flow rate at the wet scrubber system's air inlet and outlet points. The installation also involves setting up IoT gateways that facilitate communication between the sensors and the cloud server for continuous data transmission with Air Quality Index calculations. [3,4,5] post-installation, the system undergoes an operational check to ensure it functions as designed. This check focuses on maximizing efficiency in treating air pollutants and excellent particulate matter like PM2.5.

The Air Quality Index formula, as in:

$$AQI = \frac{I_{max} - I_{min}}{C_{max} - C_{min}} \times (C_i - C_{min}) + I_{min} \quad (1)$$

C. IoT Protocol Workflow for Wet Scrubber Monitoring

The IoT procedure is implemented in several key steps. First, real-time data is collected via sensors positioned throughout the wet scrubber system. The data is transmitted to an IoT gateway using lightweight communication protocols such as MQTT or CoAP, ensuring reliable and scalable data handling. Next, this data is processed in a cloud server where machine learning algorithms or preset thresholds assess the scrubber's efficiency. Real-time monitoring is achieved through dashboards that track pollutant levels, pressure drop, and system performance. If anomalies are detected, notifications are triggered, and automated control feedback is sent to adjust system parameters like water flow rate or blower speed.

D. Testing and Implementing IoT-Enhanced Pollution Control Systems

The air pollution treatment system is tested under various conditions to verify its efficiency after installation. IoT-enabled features are tested to ensure accurate data collection, transmission, and processing. The system's ability to detect real-time pollutant levels, initiate alarms in case of threshold breaches, and perform corrective actions through automated control feedback is evaluated. The system's capacity to maintain optimal performance with minimal human intervention is confirmed, ensuring that it meets operational standards for PM2.5 mitigation [6,7,8].

E. Monitoring and Long-Term Analysis

The final phase involves long-term system performance monitoring through IoT [9,10,11]. Continuous real-time analysis of the air quality released into the external atmosphere ensures compliance with environmental standards. Historical data is stored in the cloud for advanced analytics, allowing for predictive maintenance and performance improvements. The system's capability for adaptive control, coupled with IoT monitoring, enables sustained efficiency in pollution treatment while minimizing operational downtime (As illustrated in Fig 1).

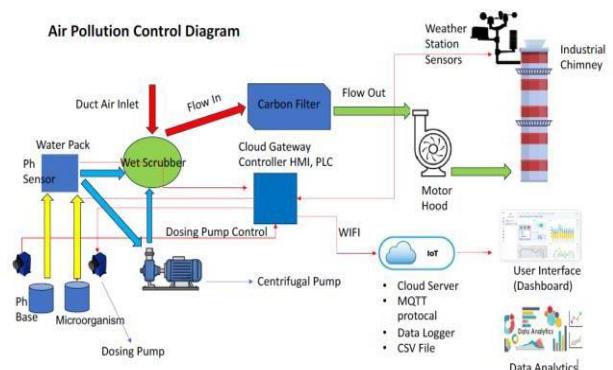


Fig. 1. Schematic of the IOT System for Air Pollution System

III. RESULTS AND DISCUSSION

The result demonstrates the integration of intelligent air pollution control systems with cloud-based technologies, focusing on using the Haiwell Cloud platform for monitoring, control, and data management in a networked environment.

The architecture connects Haiwell PLCs (Programmable Logic Controllers), HMIs (Human-Machine Interfaces), and various sensors and devices (such as photoelectric and proximity switches) through local and remote communication protocols. Data from the intelligent air pollution control system is transmitted to the Haiwell Cloud via the Internet and MQTT Servers for real-time monitoring and remote management. Critical components like (Supervisory Control and Data Acquisition) Systems, PLC [12,13], and other control systems are integrated into this cloud platform to optimize operations. The system supports multiple interfaces and device types, including mobile devices, that enable remote access to machine and air quality data through different protocols (Modbus, Inverter Freedom protocol, etc.). This lets operators and engineers control and monitor processes locally and remotely via cloud-connected HMIs and other interfaces. From an environmental perspective, as air pollution control, such an integration enhances the ability to monitor emissions and energy use in real time. This data could be used to optimize machine performance, reducing air pollution by minimizing waste and emissions during industrial operations as a cutting process. Cloud connectivity also allows for centralized data collection, which could support environmental compliance and reporting efforts (As in Fig 2).

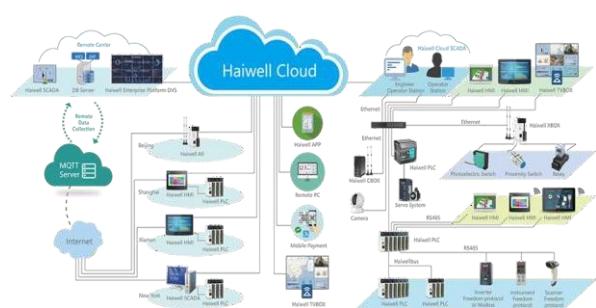


Fig. 2. Comprehensive IoT integrated with a hybrid air pollution control system

The process for determining and interpreting the Air Quality Index (AQI) based on real-time measurements of PM_{2.5} is shown in detail in Fig 3. The AQI is a standardized instrument that communicates air quality levels and their related health consequences to the public. The procedure commences with acquiring real-time PM_{2.5} data from the monitoring apparatus. The data calculates the AQI value through defined conversion procedures that link pollutant concentrations to an index from 0 to 500. The AQI is classified into various categories, each reflecting distinct air quality conditions and possible health effects. The categories span from 'Good' (AQI 0-50) to 'Hazardous' (AQI 301-500), each delineated by the health risks presented to vulnerable populations and the general populace [14].

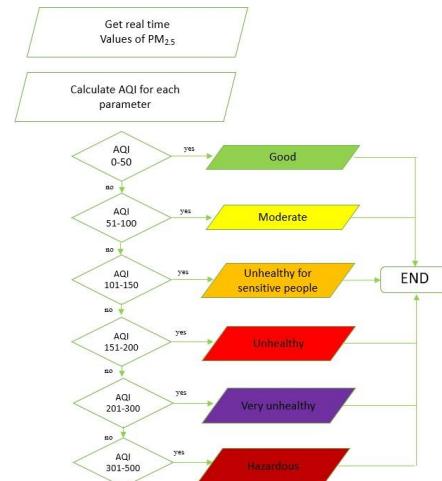


Fig. 3. Flowchart of the Air Quality Index

The design of the smart wet scrubber pollution system is illustrated in Fig 4-6 [15]. This system has many components designed to manage emissions and maintain air quality efficiently during industrial activities. These include continuous monitoring and automatic changes to ensure operational efficiency and adherence to air quality regulations.

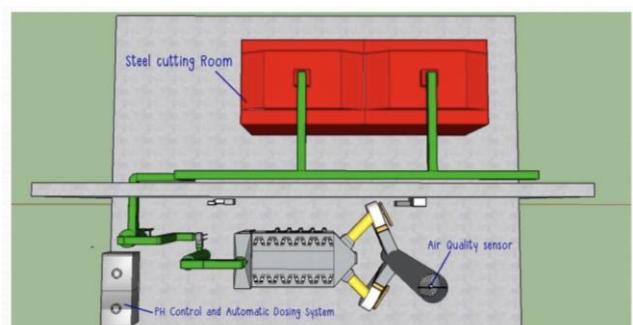


Fig. 4. Design of the air pollution control system of hybrid wet scrubber

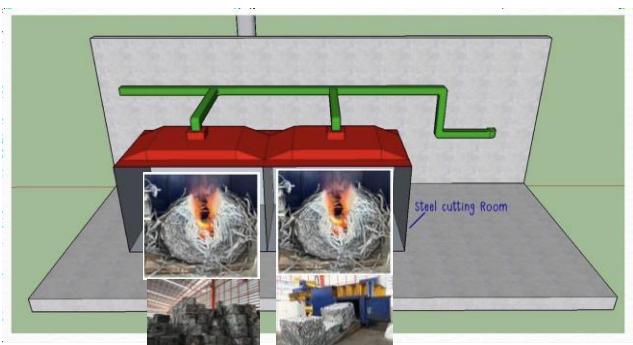


Fig. 5. Source activity of the air pollution

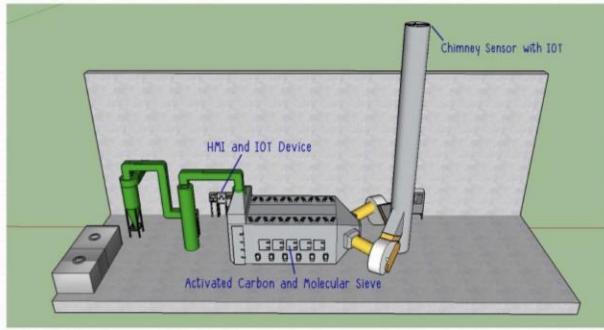


Fig. 6. Design of the Smart hybrid wet scrubber pollution control

After installing the factory's wet scrubber pollution control system [16, 17], depicted in Fig 7, and the operational image illustrated in Fig 8-9, the PLC equipped with an IoT gateway (JavaScript software) acquired the concentration of PM2.5 over 5 hours presented in Fig 10.



Fig. 7. Installation of Smart hybrid air pollution control system



Fig. 8. Operational graphic depicting the pollution control system of wet scrubbing



Fig. 9. Real-time graphic result for the Hybrid air pollution control

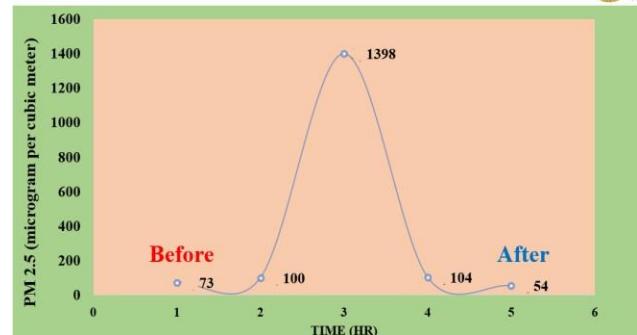


Fig. 10. The concentration of PM2.5 over 5 hours

As illustrated in Fig 10, The study reveals significant changes in air quality over 5 hours, with PM2.5 concentrations varying between "Before" and "After." PM2.5 concentrations are minimal in the "Before" period (0–2). At the 1-hour interval, the PM2.5 concentration measures 73 $\mu\text{g}/\text{m}^3$, rising slightly to 100 $\mu\text{g}/\text{m}^3$ at the 2-hour interval. Despite this, pollutant concentrations remain under 150 $\mu\text{g}/\text{m}^3$, indicating sustainable air quality. However, during the second and third hours, the PM2.5 concentration reaches a maximum of 1,398 $\mu\text{g}/\text{m}^3$. This dramatic rise suggests a sudden emission event or a significant escalation in pollutant sources. The peak far exceeds acceptable air quality standards, indicating a serious air pollution incident that could have severe health impacts, especially for those sensitive to the community. The concentration of PM2.5 decreases significantly after the peak, with a decrease to 54 $\mu\text{g}/\text{m}^3$ after five hours. This significant reduction in particulate matter is due to successfully implementing an air pollution control system, potentially involving a smart hybrid wet scrubber or other adsorption technique designed to mitigate high PM2.5 levels. The "After" label reflects the system's effectiveness in reducing particulate matter to safer levels, well below the earlier hazardous peak [18].

The researchers subsequently assessed the air quality levels every 5 minutes for two months to ascertain the distribution of AQI values. Fig 11 demonstrates that a higher AGI value indicates higher air quality. Particularly noteworthy is that the 'Good' category achieved an AGI of 27,814, indicating that the air conditions were primarily favorable. On the other hand, the result also shows substantial periods during which the air quality was classified as Moderate (AGI 17,496) and even Unhealthy (AGI 3,422). There is a possibility that vulnerable populations, such as children and those with pre-existing health concerns, could be affected by moderate and lousy air quality levels, as indicated by these changes. This type of data highlights the significance of ongoing monitoring of air quality and the adoption of mitigation methods that aim to reduce the incidence of poor air quality conditions to protect public health, particularly the health of vulnerable groups [19, 20].

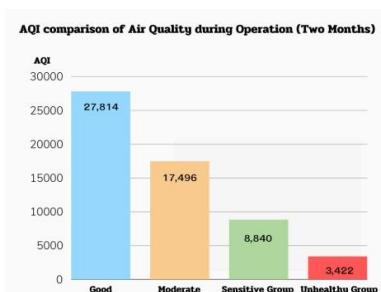


Fig. 11. AQI comparison of Air Quality during Operation (Two Months)

IV. CONCLUSION

In conclusion, the Internet of Things (IoT) via PLC board (Haiwell Cloud) has revolutionized the industrial sector by introducing connectivity, automation, and data-driven decision-making. IoT sensors and analytics enable proactive equipment maintenance, reducing downtime and optimizing asset performance for AQI in the environmental section of slitter processing. These IOT applications empower industries to operate more efficiently, improve decision-making, and drive innovation, paving the way for a more connected and intelligent industrial landscape.

ACKNOWLEDGMENT

The authors sincerely thank the University of Phayao for providing the research funding (UOE220/2567) that made this study possible. Additionally, a special thanks to Burapha University and Siam Wattana Waste Management for their invaluable support in supplying equipment and research facilities, which were essential to completing this work. Their generosity and collaboration are greatly appreciated.

REFERENCES

- [1] Wansom, Akasit, Pisit Maneechot, Nattagit Jiteurtragool, and Tharapong Vitidsant. 2023. "PM2.5 Collection Enhancement in a Smart Hybrid Wet Scrubber Tower" Processes 11, no. 12: 3306. <https://doi.org/10.3390/pr11123306>
- [2] Danzomo, Bashir Ahmed, M. J. E. Salami, Sani Jabirin, Md. Raisuddin Khan And Iskandar Mohamad Nor. "PERFORMANCE EVALUATION OF WET SCRUBBER SYSTEM FOR INDUSTRIAL AIR POLLUTION CONTROL." (2012).
- [3] S. K. Nguang and P. Shi, "Fuzzy H-infinity output feedback control of nonlinear systems under sampled measurements," *Automatica*, vol. 39, no. 12, pp. 2169-2174, 2003.
- [4] P. Shi, E. K. Boukas and R. K. Agarwal, "H1 control of discrete-time linear uncertain systems with delayed-state," Proc. of 37th IEEE Conference on Decision & Control, Tampa, Florida, pp. 4551-4552, 1998.
- [5] W. Panprasith, "Understanding indoor air quality," *Journal of Safety and Environment*, vol. 7, no. 1, pp. 57-60, 1997.
- [6] S. M. S. D. Malleswari and T. Krishna Mohana, "Air pollution monitoring system using IoT devices: Review," *Materials Today: Proceedings*, vol. 51, Part 1, pp. 1147-1150, 2022.
- [7] Pollution Control Department, "Pollution in Thailand," Ministry of Natural Resources and Environment, Bangkok, 2003.
- [8] Department of Health, Ministry of Public Health, "Technical manual: Volatile organic compounds in the atmosphere," Bangkok, 2012.
- [9] G. Mani, J. Kumar Viswanadhapalli, and P. Sriramalakshmi, "AI-powered IoT-based real-time air pollution monitoring and forecasting," *Journal of Physics: Conference Series*, vol. 2115, 012016, 2021.
- [10] J. Jo, B. Jo, J. Kim, S. Kim, and W. Han, "Development of an IoT-based indoor air quality monitoring platform," *Research Article, Open Access*, vol. 2020, Article ID 8749764.
- [11] Ministry of Interior, "Announcement on safety in working environments (chemical substances)," *Government Gazette*, May 30, 2007.
- [12] P. Shi, "Limited Hamilton-Jacobi-Isaacs equations for singularly perturbed zero-sum dynamic (discrete time) games," *SIAM J. Control and Optimization*, vol. 41, no. 3, pp. 826-850, 2002.
- [13] National Environment Board, "Announcement No. 30 (2007) on the standard for volatile organic compounds in the general atmosphere for one year," *Government Gazette*, 2007.
- [14] V. E. Alvear-Puertas, Y. A. Burbano-Prado, P. D. Rosero-Montalvo, P. Töziün, F. Marcillo, and W. Hernandez, "Smart and portable air-quality monitoring IoT low-cost devices in Ibarra City, Ecuador," *Sensors*, vol. 22, no. 7015, 2022.
- [15] Chapter 20 - Air Pollution Control. In *Environmental Engineering* (Fourth Edition), Weiner, R.F., Matthews, R.A., Eds.; Butterworth-Heinemann: Burlington, 2003; pp. 385-409.
- [16] Cheremisinoff, N.P.; Rosenfeld, P.E. Chapter 6 - Sources of air emissions from pulp and paper mills. In *Handbook of Pollution Prevention and Cleaner Production*, Cheremisinoff, N.P., Rosenfeld, P.E., Eds.; William Andrew Publishing: Oxford, 2010; pp. 179-259.
- [17] Said, M.; Manyele, S. Assessment of a Multi-Stage Wet Scrubber Performance Based on Characteristics of Scrubbing Solution. *Tanzania Journal of Engineering and Technology* 2023, 42, 206-223, doi:10.52339/tjet.v42i2.836.
- [18] Dominski, F.H.; Lorenzetti Branco, J.H.; Buonanno, G.; Stabile, L.; Gameiro da Silva, M.; Andrade, A. Effects of air pollution on health: A mapping review of systematic reviews and meta-analyses. *Environmental Research* 2021, 201, 111487, doi:<https://doi.org/10.1016/j.envres.2021.111487>.
- [19] Department of Industrial Works, "Air Pollution Treatment System Manual," 2nd ed., Bangkok: Ministry of Industry, 2007.
- [20] W. Panprasith, "Industrial ventilation for hygienists and occupational safety practitioners: Principles, calculations, and design," 3rd ed., Department of Occupational Health and Safety, Mahidol University, 2014.

A DSSAT Model Approach for Yield Optimization: A Case Study of Mae Na Rua, Phayao, Thailand

Pongsak Jindasee
*Geospatial Data Science,
School Of Information And Communication
Technology,
University of Phayao,
Phayao, Thailand.
67025257@up.ac.th*

Krittika Kantawong
*School of Information and Communication
Technology,
University of Phayao
Phayao, Thailand.
krittika.ka@up.ac.th*

<https://orcid.org/0000-0001-8006-7931>

Sawarin Lerk-U-Suke
*School of Information and Communication
Technology,
University of Phayao
Phayao, Thailand.
sawarin.le@up.ac.th*

Napa Rachata
*School of Information and Communication
Technology,
University of Phayao
Phayao, Thailand
napa.ra@up.ac.th*

Jirabhorn chaiwongsai
*School of Information and Communication
Technology,
University of Phayao
Phayao, Thailand
jirabhorn.ch@up.ac.th*

Prattana Deeprasertkul
*Researcher in Technology and Information
Research Section,
Digital and Technology Development
Division, Hydro-Informatics Institute (Public
Organization)
Bangkok, Thailand.
Prattana@hii.or.th*

Abstract— Rice production in northern Thailand, particularly in the Mae Na Ruea Sub-district, has been significantly affected by climate change, with increasing temperatures and decreasing water availability leading to reduced yields. This study aims to determine the optimal rice planting periods to maximize yields under changing climatic conditions by using the Decision Support System for Agrotechnology Transfer (DSSAT) model. The research utilizes the CERES-Rice module within DSSAT to simulate rice growth and yields across 15 planting periods from mid-April to late July 2023. Weather, soil, and crop management data were input into the model to predict yields under various environmental conditions. The results indicate that mid-May (W5) is the optimal planting time, with a predicted yield of 4294 kg/ha, 41.95% higher than the average. In contrast, early May (W3) showed the lowest yield at 1889 kg/ha. The simulated yield for weeks (W9) and (W10) were 3,018 kg/ha and 2,987 kg/ha, respectively, which were only 0.23% and 1.26% lower than the average. This demonstrates the model's a high degree of precision. The study highlights the significant impact of planting time on yield outcomes, emphasizing the need for adaptive strategies to cope with climate variability. However, the model's limitations, including underestimation of early-season drought stress and lack of consideration for factors such as irrigation and pest pressures, suggest the need for further research to enhance prediction accuracy and address long-term environmental variability.

Keywords— Rice, DSSAT Model, Climate, planting period

I. INTRODUCTION

Thailand has the potential to become a world leader in rice production, but current rice production is still lower than it should be when compared to competitors. In the 2020/2021 crop year, Thailand had a rice plantation area of 10.99 million

hectares [1]. Rice is the main economic crop of Mae Na Rua Sub-District, with the cultivation area accounting for approximately 40% of the total agricultural area. The jasmine rice cropping calendar plans in early May to June.

Global climate change is progressing continuously. In 2023, the global average temperature increased by 1.4°C, the highest level ever recorded [2]. This change has led to increased variability in rainfall and a higher frequency of extreme weather events. Additionally, it has impacted hydrological cycles in various regions, with some areas experiencing reduced rainfall and increased evaporation rates, resulting in decreased water availability [3]. Consequently, water management has become more complex and unpredictable [4], [5]. Therefore, preparing for the unavoidable impacts of climate change is essential. Enhancing strategies to adapt to these conditions is a crucial aspect of effective preparedness.

Northern Thailand is clearly impacted by climate change, with significant reductions in rainfall and rising temperatures, leading to decreased water availability. This has had a major effect on agriculture, particularly rice cultivation, causing a continuous decline in rice production across the country. Research in Thailand has shown that adopting a range of agricultural practices and utilizing crop modeling are essential for developing strategies to adapt to changing water conditions. These models help estimate yields, plan production, and select suitable crop varieties for various conditions. Implementing these strategies can effectively mitigate the impacts of climate change and enhance productivity. [6], [7], [8], [9].

This research aims to use the DSSAT (Decision Support System for Agrotechnology Transfer) model as a tool to identify

the optimal planting periods for rice in Mae Na Ruea Sub-district, Mueang District, Phayao Province. The model will assist in estimating expected yields under various conditions by utilizing data on weather, soil quality, and other agricultural factors. Additionally, the accuracy of the model's yield predictions will be assessed by comparing them with actual yield data. This will help evaluate the effectiveness of DSSAT in providing useful recommendations for effective rice management and decision-making [10], [11]. The subsequent phase entails the preparation of requisite data for the purpose of model evaluation. Upon the completion of data preparation, the model will be assessed utilizing the aforementioned data. The outcomes of this evaluation will then be presented for comparative analysis against the observed data.

II. DATA GATHERING AND MODEL

A. Study area: Mae Na Rua Sub-District, Mueang Phayao District, Phayao Province, Thailand.

The study area, located in the Mae Na Rua (MNR) sub-district of Phayao Province, northern Thailand, covers approximately 117 square kilometers, with coordinates between 19°00' N to 19°08' N and 99°45' E to 99°54' E. The topography features lowland areas in the north and mountainous regions to the west. Two main rivers, the NamRum and NamSai, traverse the area as shown in Fig.1. In 2023, the region recorded 683 mm of precipitation, a maximum temperature of 43.8°C, a minimum of 12.9°C, and an average relative humidity of 76.13% as reported by weather station from hydro-informatics institute (HII).

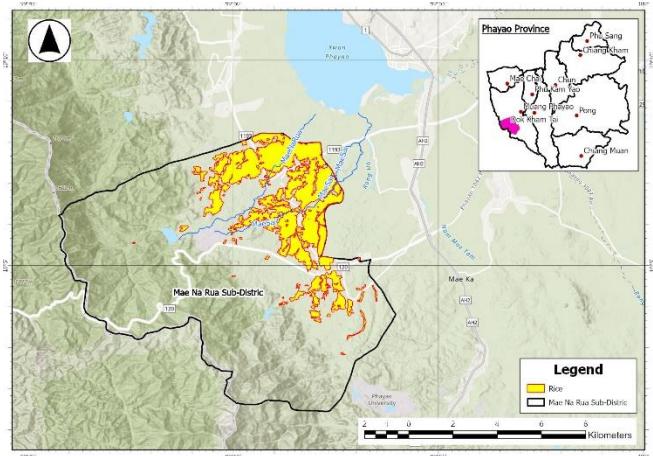


Fig.1 Geographical location of the study area.

B. Data Gathering

The data used in this study are comprehensively summarized in TABLE I, which provides an overview of the various datasets collected and utilized throughout the research. Productivity data (PD) is the Observed Yield from harvesting that is averaged to obtain central data for the area. This is done by agricultural agencies which collect data during harvest seasons. Based on the data collected and studied on rice planting periods in the study area, the research team used this information to calibrate the model. This includes details such as the type of data, source, temporal coverage, and relevance to the study objectives.

TABLE I. Summary of data collected for this study.

Data	Detail	Unit	Frequency data	Period	Reference
WD	Minimum temperature	°C	Daily	2023	HII
	Maximum temperature	°C	Daily	2023	HII
	Rainfall	mm.	Daily	2023	HII
	Solar Radiation	(MJ/m ² / day)	Daily	2023	DSSAT WeatherMan
Soil	Soil series	Series	Area	2010	Land develop department
Plant	Genetic coefficients	Genotyp e	Area	1959	Department of Agriculture
	Plant date	day	Daily	2023	Subdistrict Administrative organization
PD	Jasmine rice	kg/ha	Year	2023	office of agricultural economics

C. Weather Data(WD)

The DSSAT model's standard climate dataset (minimum data set) requires data on Solar Radiation (MJ/m²/day), Minimum Temperature (°C), Maximum Temperature (°C), and Rainfall (mm) [12]. All Climate datasets use daily data from 2023 at the Mae Sai Sub-district Administrative Organization (WBYN) station. The Weatherman program TABLE II., a DSSAT model climate data preparation module, can simulate Solar Radiation. Thailand experiences abundant rainfall during the southwest monsoon, which begins in May and can last until October. Warm, moist air from the Indian Ocean is carried inland, resulting in frequent showers and thunderstorms [13], [14].

TABLE II. Example of monthly weather data.

Year-Month	Tmax (°C)	Tmin (°C)	Rain (mm)	Solar (MJ/m ²)
2023-01	31.09	17.19	0.27	18.03
2023-02	34.08	20.2	1.35	18.88
2023-03	35.97	23.5	0.15	19.5
2023-04	40.02	27.36	0.76	21.84
2023-05	38.58	27.25	2.93	21.75
2023-06	36.27	26.95	2.39	18.69
2023-07	36.05	26.23	6.43	18.51
2023-08	33.17	25.59	3.39	14.73
2023-09	34.21	25.35	6.39	16.08
2023-10	34.27	24.13	0.3	15.32
2023-11	32.6	21.38	0.4	16.84
2023-12	32.86	20.24	0.06	15.45

D. Decision Support System for Agrotechnology Transfer (DSSAT) Model

To estimate crop yield, we employed the Decision Support System for Agrotechnology Transfer (DSSAT), specifically version v4.8.2. [15] For the assessment of rice growth rates and yields in the model, the necessary fundamental data for this model include climate data, soil data, crop management data, and plant data. The climate data include rainfall, solar radiation energy, maximum and minimum temperatures [16], [17]. The planting start dates are divided into 15 periods, which include: 17-Apr-23(W1), 24-Apr-23 (W2), 1-May-23(W3), 8-May-23(W4), 15-May-23(W5), 22-May-23(W6), 29-May-23(W7),

5-Jun-23(W8), 12-Jun-23(W9), 19-Jun-23(W10), 26-Jun-23(W11), 3-Jul-23(W12), 10-Jul-23(W13), 17-Jul-23(W14), 24-Jul-23(W15). In this simulation Fig.2, the CERES-Rice module was used. CERES-Rice is a component of DSSAT, a widely used crop simulation model designed to predict crop

growth, yield, [18] and responses to environmental conditions and management practices such as fertilizer, water, and climate[19].

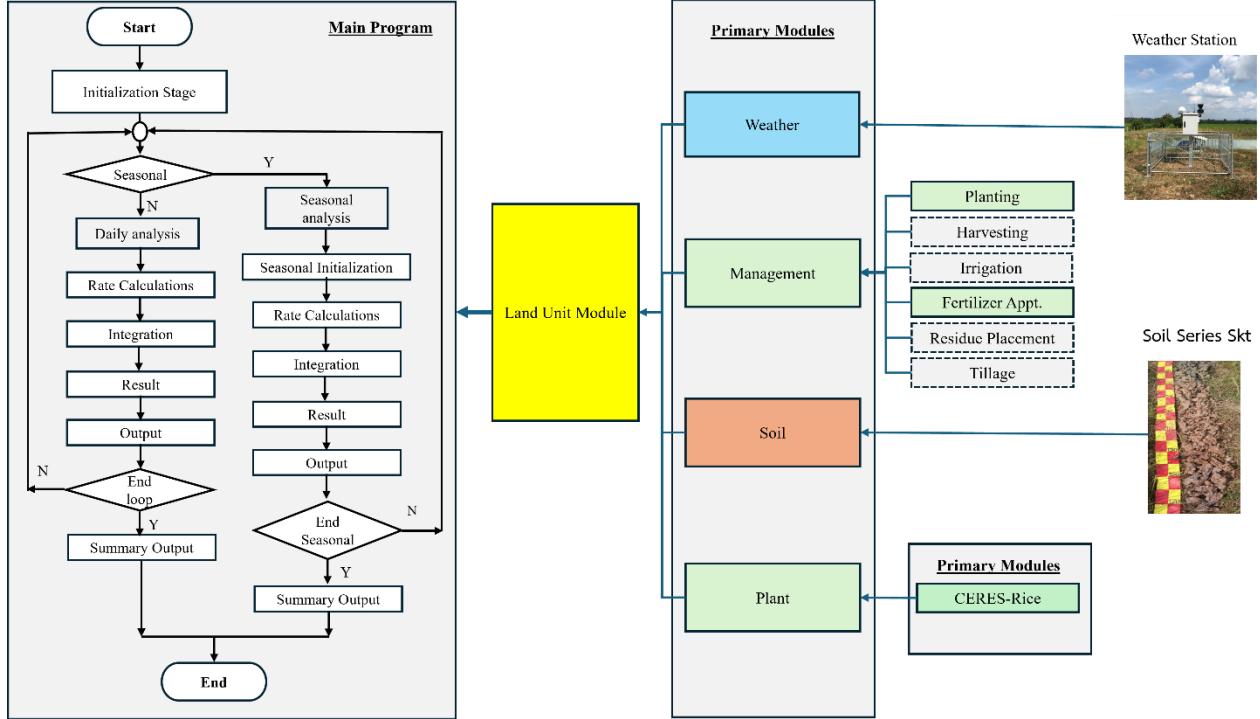


Fig.2 Modular structure of the DSSAT.

III. METHODOLOGY

Each module encompasses six operational steps, as depicted in Fig.2 run initialization, seasonal initialization, rate calculations, integration, daily output, and summary output. The central program governs the activation timing of these steps and the execution of tasks within each module. This modular architecture empowers each module to independently read inputs, initialize itself, compute rates, integrate its state variables, and generate outputs without reliance on other modules. Solely a select few 'interface' variables are exchanged between modules, enabling the interchangeability of modules with minimal disruption to the overall system. State variables are recorded post-integration to capture the system's status at the end of the day, while initial values are documented during initialization for the commencement of the simulation, eliminating the need for code modifications in other modules. The application driver exclusively interacts with the Land Unit module, as illustrated in Fig.2. The Land Unit Module serves as the

interface between the application driver (main program) and all components within the designated land area.[17]

IV. RESULTS AND DISCUSSION

The CERES-Rice model in DSSAT was used to evaluate suitable rice planting times for 2023, with harvesting based on the crop's maximum maturity. Fig. 3 presents a comparison of simulated rice yields for various planting periods. As shown in Fig. 4, adjusting planting dates by just 7 days can result in significant yield differences. In 2023, planting between April and mid-June showed the greatest variation in yield potential, with the possibility of achieving both the highest and lowest yields. In contrast, planting between mid-June and July produced yields closer to the average. According to Fig. 3, the average yield across all 15 periods was 2,925 kg/ha. The highest yield was recorded in W5 (mid-May) at 4,294 kg/ha, while the lowest yield was in W3 (early May) at 1,889 kg/ha. The model suggests that selecting the optimal planting time could increase yields by as much as 1,369 kg/ha above the average.

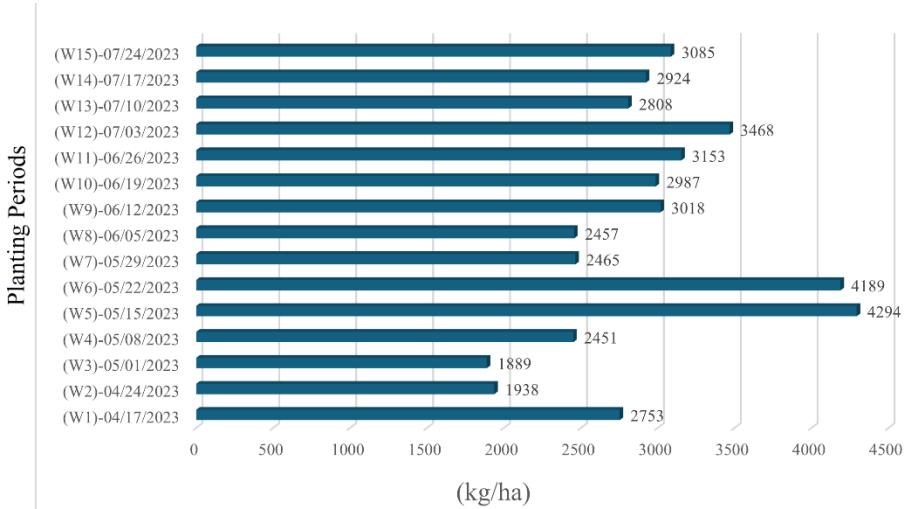


Fig. 3 Yield Evaluation Over Time for Different Planting Dates from the CERES-Rice Model.

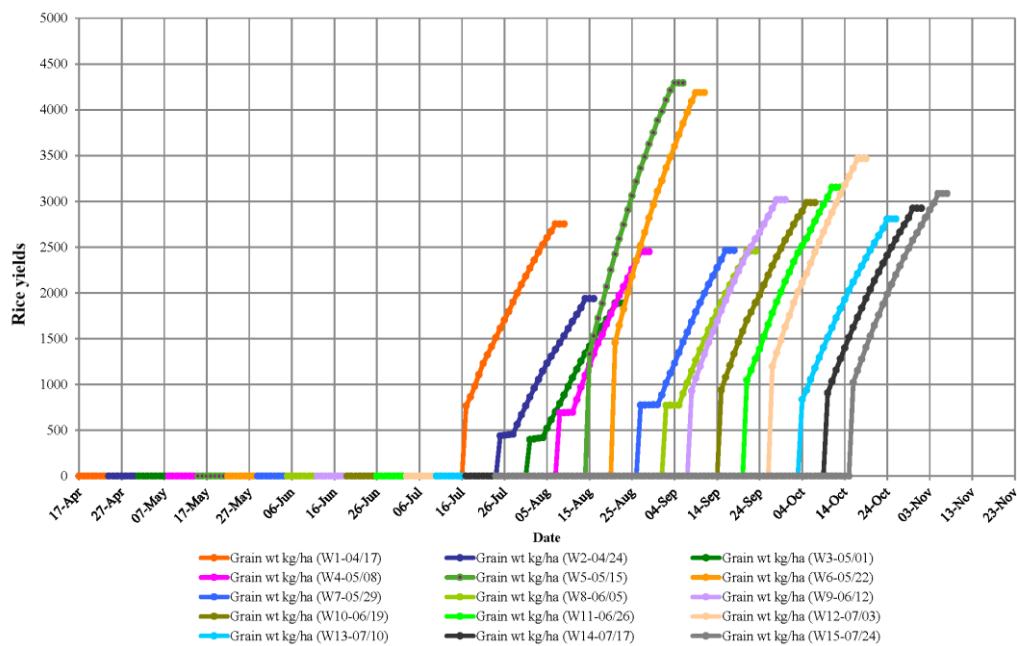


Fig. 4 Productivity and growth from the CERES-Rice Model.

The comparison between the average observed yield throughout the year (3,025 kg/ha) and the yield predicted by the CERES-Rice model for each planting period in 2023 shows percentage differences (Difference %) to highlight which planting times result in higher or lower yields. In W5 (May 15), the predicted yield was 4,294 kg/ha, 41.95% higher than the average. In contrast, W3 (May 1) had a predicted yield of 1,889 kg/ha, a decrease of 37.55%. Additionally, W6 (May 22) had a predicted yield of 4,189 kg/ha, 38.48% above the average, while W2 (April 24) dropped to 1,938 kg/ha, a decrease of 35.93%. Meanwhile, in W9 (June 12) and W10 (June 19), the simulated yield were 3,018 kg/ha, which is 0.23% lower than the average, and 2,987 kg/ha, which is 1.26% lower than the average, respectively. This indicates that during this period, the model

was able to predict yields that closely matched the actual observed values.

TABLE III. highlights significant discrepancies between observed and simulated crop yields, indicating both overestimation and underestimation by the CERES-Rice model across different treatments. Treatments W1 to W6 correspond to the summer season, a period characterized by low rainfall in the study area. The suboptimal growing conditions during this time, particularly water stress, may have contributed to reduced observed yields. It is possible that the model did not fully account for the adverse effects of early-season drought stress, leading to divergence between simulated and actual yields. Enhanced calibration of the model to better reflect these environmental constraints could improve its accuracy. Early season drought can significantly

affect plant germination and root formation, which are crucial factors for later growth. If the model underestimates drought stress, plant growth results may appear overly optimistic in the later season, as the root system was not properly affected by drought in the early period [20]. This can be improved by using growth data from each period obtained through remote sensing, which enhances prediction accuracy in crop management[21], [22] . In addition, other limiting factors e.g., pest pressure or unforeseen weather disturbances, may have reduced the actual yield. These results are consistent with studies on identifying optimal planting times for maximizing rice yield and adapting to climate conditions [23].

TABLE III. Comparison of Observed and Simulated Rice Yields from the CERES-Rice Model.

Planting Dates (Treatments)	Average Observed Yield (kg/ha)	Simulated Yield (kg/ha)	Difference (%)
(W1)	3,025	2,753	-8.99
(W2)	3,025	1,938	-35.93
(W3)	3,025	1,889	-37.55
(W4)	3,025	2,451	-18.98
(W5)	3,025	4,294	41.95
(W6)	3,025	4,189	38.48
(W7)	3,025	2,465	-18.51
(W8)	3,025	2,457	-18.78
(W9)	3,025	3,018	-0.23
(W10)	3,025	2,987	-1.26
(W11)	3,025	3,153	4.23
(W12)	3,025	3,468	14.64
(W13)	3,025	2,808	-7.17
(W14)	3,025	2,924	-3.34
(W15)	3,025	3,085	1.98

V. CONCLUSIONS

The study utilized three primary data sets. However, it did not account for key factors such as harvesting, irrigation, residue placement, or tillage. Furthermore, relying on climate data from a single year may not fully capture long-term environmental variability, as climate conditions can fluctuate or exhibit unique characteristics specific to that year. Therefore, further research is necessary to identify the factors that significantly impact yields during different planting periods. This should include detailed analyses of climate, water management, and chemical usage to establish a clearer understanding of the relationships between these factors and yield outcomes.

While the DSSAT model provided useful insights for rice agriculture during the 2023 season, incorporating multiple years of weather data would strengthen its predictive power. This expanded analysis would better capture year-to-year climate variations, enabling more reliable long-term agricultural planning.

ACKNOWLEDGMENT

This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant Number B13F670082-8].

The authors thank the Hydro-Informatics Institute (Public Organization) and the University of Phayao for supporting the tools and software to conduct this study . I would like to thanks Dr.Isared Kakarndee, Department of Farm Mechanics, Faculty of Agriculture, Kasetsart University, Thailand., who teaches to operate DSSAT software.

REFERENCES

- [1] S. Srisopa, P. Luamka, S. Rattanawan, K. Somtrakoon, and P. Busababodhin, "Analyzing Spatial Dependence of Rice Production in Northeast Thailand for Sustainable Agriculture: An Optimal Copula Function Approach," *Sustainability*, vol. 15, no. 20, p. 14774, Oct. 2023, doi: 10.3390/su152014774.
- [2] *Infrastructure for a Climate-Resilient Future*. OECD, 2024. doi: 10.1787/a74a45b0-en.
- [3] J. R. B. Tercini, R. F. Perez, A. Schardong, and J. I. G. Bonnecarrère, "Potential Impact of Climate Change Analysis on the Management of Water Resources under Stressed Quantity and Quality Scenarios," *Water (Basel)*, vol. 13, no. 21, p. 2984, Oct. 2021, doi: 10.3390/w13212984.
- [4] M. Ahmadi, B. Motamedvaziri, H. Ahmadi, A. Moeini, and G. R. Zehtabiyani, "Assessment of climate change impact on surface runoff, statistical downscaling and hydrological modeling," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 114, p. 102800, Dec. 2019, doi: 10.1016/j.pce.2019.09.002.
- [5] E. Rozhkova and R. Guo, "Impact of Climate Change on Water Resource Management. Case study on Kura basin," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 3, p. p8791, Mar. 2019, doi: 10.29322/IJSRP.9.03.2019.p8791.
- [6] A. Mulagha-Maganga, L. S. Chiwaula, P. Kambewa, and M. E. Ngaiwi, "A Ricardian analysis of the economic impacts of climate change on agricultural production in the low-income agrarian economy: Estimates from Malawi's 2010–2019 LSMS longitudinal data," *J Agric Food Res*, vol. 15, p. 100995, Mar. 2024, doi: 10.1016/j.jafr.2024.100995.
- [7] B. Prommawin *et al.*, "Impacts of climate change and agricultural diversification on agricultural production value of Thai farm households," *Clim Change*, vol. 177, no. 7, p. 112, Jul. 2024, doi: 10.1007/s10584-024-03732-3.
- [8] T. Amnuaylojaroen, P. Chanvichit, R. Janta, and V. Surapipith, "Projection of Rice and Maize Productions in Northern Thailand under Climate Change Scenario RCP8.5," *Agriculture*, vol. 11, no. 1, p. 23, Jan. 2021, doi: 10.3390/agriculture11010023.
- [9] M. Annie, R. kumar Pal, A. S. Gawai, and A. Sharma, "Assessing the Impact of Climate Change on Agricultural Production Using Crop Simulation Model," *International Journal of Environment and Climate Change*, vol. 13, no. 7, pp. 538–550, May 2023, doi: 10.9734/ijecc/2023/v13i71906.
- [10] D. Bhanusree *et al.*, "Estimation of potential yield of rice using DSSAT-CERES rice model," *International Journal of Research in Agronomy*, vol. 7, no. 2S, pp. 44–52, Feb. 2024, doi: 10.33545/2618060X.2024.v7.i2Sa.334.
- [11] R. Kumar *et al.*, "Impact analysis of moisture stress on growth and yield of cotton using DSSAT-CROPGRO-cotton model under semi-arid climate," *PeerJ*, vol. 11, p. e16329, Nov. 2023, doi: 10.7717/peerj.16329.
- [12] J. W. Jones *et al.*, "The DSSAT cropping system model," *European Journal of Agronomy*, vol. 18, no. 3–4, pp. 235–265, Jan. 2003, doi: 10.1016/S1161-0301(02)00107-7.
- [13] O. Pilahome, C. Homchampa, and W. Kumharn, "Trends of Climate Variables and Aerosol optical Depth in Thailand," *IOP Conf Ser Earth Environ Sci*, vol. 489, no. 1, p. 012029, Apr. 2020, doi: 10.1088/1755-1315/489/1/012029.
- [14] P. Aumjira and P. Trivej, "Rainfall estimation from radar in different seasons over Northern Thailand," *J Phys Conf Ser*, vol. 1144, p. 012122, Dec. 2018, doi: 10.1088/1742-6596/1144/1/012122.
- [15] G. Hoogenboom *et al.*, "Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.8.2," DSSAT

- Foundation. Accessed: Sep. 12, 2024. [Online]. Available: www.DSSAT.net
- [16] G. Hoogenboom *et al.*, “The DSSAT crop modeling ecosystem,” 2019, pp. 173–216, doi: 10.19103/AS.2019.0061.10.
- [17] J. W. Jones *et al.*, “The DSSAT cropping system model,” *European Journal of Agronomy*, vol. 18, no. 3–4, pp. 235–265, Jan. 2003, doi: 10.1016/S1161-0301(02)00107-7.
- [18] H. V., G. K. Das, S. K. Chandniha, H. V. Puranik, and J. L. Chaudhary, “Evaluation of CERES –Rice Model for Simulating Rice Yield and Phenophases,” *Int J Plant Soil Sci*, vol. 36, no. 7, pp. 371–376, Jun. 2024, doi: 10.9734/ijpps/2024/v36i74742.
- [19] K. R. B. Kumar, M. A. Reddy, G. B. Gohain, K. B. Shafeer, S. Sarathkumar, and S. Sundaranayagi, “Assessment of Future Climate and Kharif Paddy Yields using Ceres-Rice in the State of Andhra Pradesh,” *Current World Environment*, vol. 18, no. 1, pp. 176–196, Apr. 2023, doi: 10.12944/CWE.18.1.16.
- [20] J. R. Pinto, J. D. Marshall, R. K. Dumroese, A. S. Davis, and D. R. Cobos, “Seedling establishment and physiological responses to temporal and spatial soil moisture changes,” *New For (Dordr)*, vol. 47, no. 2, pp. 223–241, Mar. 2016, doi: 10.1007/s11056-015-9511-7.
- [21] D. Lal and R. Niwas, “Yield Predication by DSSAT Model of Wheat Crop: A Review,” *International Journal of Environment and Climate Change*, vol. 14, no. 2, pp. 519–524, Feb. 2024, doi: 10.9734/ijecc/2024/v14i23965.
- [22] U. Akumaga *et al.*, “Integration of Remote Sensing and Field Observations in Evaluating DSSAT Model for Estimating Maize and Soybean Growth and Yield in Maryland, USA,” *Agronomy*, vol. 13, no. 6, p. 1540, Jun. 2023, doi: 10.3390/agronomy13061540.
- [23] K. Liu *et al.*, “Optimizing Rice Sowing Dates for High Yield and Climate Adaptation in Central China,” *Agronomy*, vol. 13, no. 5, p. 1339, May 2023, doi: 10.3390/agronomy13051339.

Application of HEC-RAS and Satellite Imagery in Flood Risk Estimation for Rural Area in Thailand

1st Nattaphon Rangsaritvorakarn
*School of Information and
 Communication Technology
 University of Phayao
 Phayao, Thailand
 65025121@up.ac.th*

2nd Niti Iamchuen
*School of Information and
 Communication Technology
 University of Phayao
 Phayao, Thailand
 niti.ia@up.ac.th*

Abstract— The world is facing unavoidable disaster challenges, especially floods in Thailand, these inundations have been becoming increasingly difficult to manage each year. Therefore, a preliminary approach to addressing this issue is to identify areas that may experience future flooding. This study used HEC-RAS model, was calibrated by Manning's N to compare with Sentinel-1 satellite imagery and then were evaluated by Accuracy, F1 and F2 metrics to simulate and assess the affected areas for rural area in Thailand. As the results, no evidence of data overfitting was observed after the calibration between flood maps generated by HEC-RAS and satellite imagery. Moreover, the F metrics and accuracy were not exceeded by 60%, demonstrating a reasonable level of model performance. Subsequently, such model was employed to simulate flooding scenarios with high water levels 1 and 2, was found to be the identified agricultural lands as the most significantly affected area with a high likelihood of impact further. Urban areas were the second most affected. Consequently, these findings benefit the substantial impact of flooding on the livelihoods of agricultural communities and emphasize the consequent damage to household assets.

Keywords—flood, HEC-RAS, satellite imagery, disaster, risk map

I. INTRODUCTION

The world is currently facing increasing challenges from natural disasters, both in frequency and intensity due to climate change.[1]. These events impact human safety and the integrity of the global environment such as rainfall variability, leading to floods, landslides, and droughts [2]. Thailand is one of the countries heavily impacted by climate change and experiences various types of disasters. A study on the effects of climate change from 1999 to 2019 revealed that Thailand experienced 146 major disasters, with damages totaling 7.86 billion US dollars ranking it 9th globally in disaster impact [3].

One widespread way to estimate floods is through in-situ observations. However, these observations are not always available. Some researchers have turned to remote sensing data for flood analysis [4], but floods often happen during cloudy weather, limiting the effectiveness of remote sensing. Additionally, observation-based flood studies are usually specific to certain flood events [5], making them unsuitable for analyzing future floods or their consequences.

Therefore, widely used tools include various hydrological models such as MIKE 11 HEC-HMS or SWAT which focus on rainfall-runoff process [6], [7]. Upon analysing the flooded areas of the study site, it is evident that most of the flooding is caused by water flowing down from the mountains from the Chan River and Kham River, overflowing into the lowland

areas and causing floods. This makes flood analysis tools that focus on river flooding, such as Hydrologic Engineering Centre's: River Analysis System (HEC-RAS).

Several studies have utilized the HEC-RAS hydraulic model for flood mapping. Several studies have utilized HEC-RAS to generate flood hazard maps [8], [9], [10]. Various studies have indicated that tools like SWAT or HEC-RAS are sufficiently capable of predicting floods, as the primary input is the water flow measured at stations [11]. The integration with GIS systems further enhances the accuracy and enables effective application [9], [10], [12]. However, these modelling could not indicate accuracy without validation against flood area data. There were two approaches for validation, first was field surveys, which were time-consuming, or a second, high-resolution satellite imagery such as Sentinel-2 in the visible spectrum, near-infrared (NIR), and shortwave infrared (SWIR) bands, but the cloud cover was the limitation, especially during the rainy season when cloud density was high.

Therefore, it was necessary to use Sentinel-1, an active satellite capable of penetrating clouds, to calibrate against the flood model. Several studies have explored the use of satellite imagery for calibration [13], [14], particularly leveraging Sentinel-1 satellite data to enhance the capability and accuracy of HEC-RAS [15], [16], [17], [18]. Since each area requires a different approach, the challenge in research lies in the fact that there has yet to be a study on using HEC-RAS calibrated with Sentinel-1 satellite maps in rural areas of Thailand. Therefore, this study aims to apply the HEC-RAS hydrological model by calibrated it with satellite maps to estimate flood-prone areas and predict flood-prone areas under extreme climate change scenarios in Mae Chan District, Chiang Rai Province, Thailand.

II. METHODOLOGY

The research method was begun by inputting relevant data into the HEC-RAS model for analysis. Following this, Manning's n (N) coefficient will be calibrated using satellite imagery that has been processed to isolate flooded areas. The model's performance was then validated, ensuring that overfitting was avoided. Subsequently, the predicted water discharge was incorporated into the model to simulate flood risk estimation, along with land use data, under two distinct scenarios. The methodological study has more details as follows:

A. Study Area

the research area was selected within Mae Chan District, Chiang Rai due to its terrain of high hills and two rivers,

namely The Chan River and The Kham River. These rivers have flowed from high mountain sources and traverse the floodplains of Mae Chan District. The study focused on the area to the right side of the region because satellite imagery showed evidence for frequent flooding. Additionally, due to data limitation on water flow through from the stations, this only analysis was the specific measured stations such as Kh.72 of Kham River and Kh.89. for Chan River. Therefore, the study area was limited to only one region as shown in Figure 1.

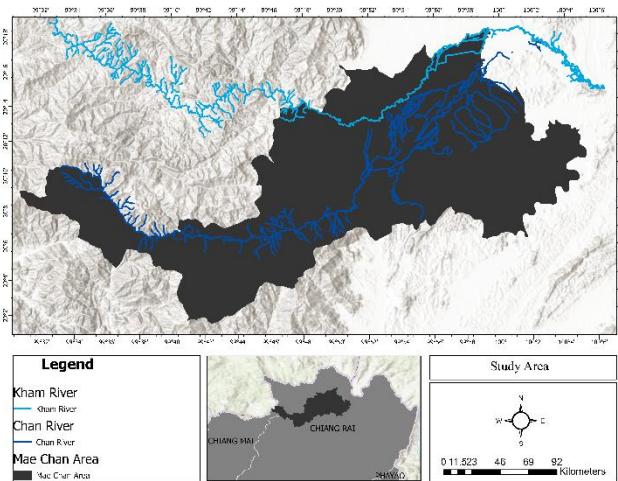


Fig. 1. Study Area and HEC-RAS Modeling Area

B. Satellite-Based Flood Extent Mapping

The study aims to use satellite imagery to evaluate the accuracy of hydrological models, specifically employing Sentinel-1 satellite data-based an active radar satellite from European Union [19], well-suited for distinguishing flooding events and also not impeded by cloud cover. The study used two data sets: 1) Training data from Sentinel-1 VH dated August 17, 2021, and 2) Testing data from Sentinel-1 VH dated August 12, 2022.

Subsequently, the process of delineating flood-prone areas were analyzed by image processing techniques, including color separation and thresholding. A color threshold ranging from 0 to 70 were used to identify flooded areas while values greater than this range were classified as non-flooded by ArcGIS Pro software.

C. Pre-processing Data

The process of data preparation revealed that the hydrological model, 'HEC-RAS' required several key datasets. These key datasets included a digital elevation model (DEM) with a 1:4,000 scale and a land use map from the Land Development Department (LDD) of Thailand [20].and water discharge data from the Upper Northern Regional Hydrology Centre (UNRHC), Royal Irrigation Department (RID) of Thailand.

D. Actual and Predicted Hydrological Data

The study used hourly water discharge data from two measured stations, Kh.72 and Kh.89, with 100 data points from each station. The period used for train data analysis spans from August 13, 2021, at 9:00 PM to August 17, 2021, at 12:00 AM, which corresponds to the availability of Sentinel-1 satellite data for train data. Next, the period was

from August 8, 2022, at 9:00 PM to August 12, 2022, at 12:00 AM, which also corresponded to the availability of Sentinel-1 satellite data for test data. This study also experimented with increasing future water discharge levels by using forecasts from Polynomial Regression to estimate the potential water flow under abnormally high-water levels risk. This approach allows for the simulation of increased water discharge for further analysis. The formula used in the Polynomial Regression analysis is as follows [21]:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n \quad (1)$$

Where y = water discharge, x = water level and $b_0, b_1, b_2 \dots b_n$ = coefficients determined by the regression analysis which n means degree of the polynomial. The result is a forecast of water discharge to be used as input for predicting future water conditions in the event of more severe scenarios.

E. Simulation and Calibration in the HEC-RAS Model

The Hydrologic Engineering Center's River Analysis System (HEC-RAS) version 5.0.7, developed by the US Army Corps of Engineers, is widely used for hydraulic and hydrologic analyses. The model discretizes the area into mesh cells, integrating terrain data with a sub-grid approach for efficiency [22]. Water movement is influenced by topography, flow resistance, land use, and Manning's coefficient [23]. HEC-RAS supports one-dimensional (1D) steady and unsteady flow and two-dimensional (2D) unsteady flow [22].

In this study, only two-dimensional flow analysis was conducted due to the high likelihood of overflow during flooding events and the study utilized unsteady flow simulation because flooding conditions are not constant. The HEC-RAS two-dimensional model used Shallow Water Equations (SWE) or the conservation of mass and momentum calculator as following equation [24].

Mass conservation:

$$\frac{\partial h}{\partial t} + \frac{\partial \partial(hu)}{\partial x} + \frac{\partial \partial(hv)}{\partial y} = q \quad (2)$$

Where t = time, h = water depth, u = velocity in x and v = velocity for y , the result will be the expansion of q = source flux term.

Momentum conservation in x and y direction:

$$\frac{\partial v}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + g \frac{\partial H}{\partial x} + \frac{gn^2|u|}{R^{4/3}} u = 0 \quad (3)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + g \frac{\partial H}{\partial y} + \frac{gn^2|v|}{R^{4/3}} v = 0 \quad (4)$$

In this context, g represents gravitational acceleration measured, n is Manning's coefficient, and R denotes the wetted perimeter in units of length. In the model, the initial 2D computational mesh model was based on DEM of the study. The basic parameters of generated 2D mesh model are number of cells = 3,769 cells, average face length = 189 cells, average Cell Size = 35,910 cells, maximum cell size = 97,481 cells and minimum cell size = 428 cells.

The calibration which approach involves utilizing Manning's roughness coefficient (Manning's n (or N)) which

represented surface roughness, directly influencing water flow resistance in rivers or floodplains. Different values were used based on surface characteristics—smooth surfaces like concrete had low N, while rough, vegetated areas had higher N. The study aims to enhance the model's predictive accuracy. This study involved comparing simulated flood extents with real-world flood occurrences and satellite images to ensure the model's performance is optimized.

F. Validation of HEC-RAS Model

The accuracy assessment involved integrating observed and simulated flood maps within ArcGIS Pro 3.0.2 software. The observed maps, derived from satellite imagery, and the simulated maps, produced by the HEC-RAS model, were overlaid to facilitate a comparative analysis. The accuracy points were generated as a sampling method to test the precision across the study area.

The sample size for accuracy testing was determined using the Binomial Probability Theory, as outlined by Fitzpatrick-Lins [25]. The formula for calculating the sample size was:

$$n = \frac{Z^2(p)(q)}{E^2} \quad (5)$$

where:

n = Sample size

Z = Z-score corresponding to the expected confidence level (1.96 for 95% confidence and round-up to 2)

p = Estimated proportion of the attribute of interest

q = the perfect of 100 – p

E = Margin of error (5% in this case)

These points are then used to assess the model's accuracy through a confusion matrix [26], [27], [28]. The confusion matrix provides values for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Accuracy is quantified using the following metrics:

$$\text{Accuracy} = \frac{(TP+TN)}{TP+FP+FN+TN} \quad (6)$$

$$F1 = \frac{(2 \times TP)}{(2 \times TP) + FP + FN} \quad (7)$$

$$F2 = \frac{(5 \times TP)}{(5 \times TP) + (4 \times FP) + FN} \quad (8)$$

The analysis indicates that a model accuracy of greater than 40% is acceptable, with an optimal target of over 50% for practical applications. This ensures that the flood model provides reliable predictions, which are crucial for effective flood management and mitigation strategies [29], [30], [31].

Finally, it is essential to compare the performance of the model by testing it against both training data and test data to assess the degree of discrepancy. If the model shows an accuracy deviation greater than 10% it indicates potential overfitting. In such cases, the model needs to be revised and refined to improve its accuracy and generalizability. This process ensures that the model performs reliably across different datasets and is not overly tuned to the training data alone.

III. RESULT

This study aims to apply flood modeling based on validation of HEC-RAS model and extracted satellite maps for flood periods. The process involves calibrating parameters to closely match real-world scenarios and assessing accuracy through comparison with satellite maps. Consequently, the results are presented as flood maps and using water discharge forecasting to examine how the impact area might expand under higher water levels. The analysis was conducted by using HEC-RAS 6.5 in conjunction with ArcGIS Pro 3.0.2.

A. HEC-RAS Model Calibration

This study was found that calibration of flood-affected areas should use adjusting Manning's N values in Mae Chan District, Chiang Rai Province, Thailand.. The land use types were converted to manning roughness categories, including such as open water, developed medium intensity, pasture, cultivated crops, forest, and developed open space. Moreover, this analysis revealed that most of land use type as cultivated crops and open water, respectively. Consequently key N value was identified as Cultivated Crops.

TABLE I. MANNING'S N VALUE CALIBRATION

Classification	Minimum	Recommended	Maximum	Calibrate
Open Water	0.01	0.02	0.03	0.05
Developed Med Intensity	0.06	0.18	0.2	1
Pasture	0.15	0.22	0.3	1
Cultivated Crops	0.1	0.17	0.3	0.8
Forest	0.18	0.25	0.3	0.12
Developed Open Space	0.04	0.06	0.1	0.04

From table 1, the HEC-RAS user manual was recommended values, but this study adjusted them based on the characteristics of 'developed' areas, such as houses and 'pasture'. In this region, vegetation was very dense, with some areas overgrown and lacking proper urban planning, which hindered smooth water flow. This led to a significant increase in flow resistance. Furthermore, it was noted that local farms and fields had various types of crops and unusual soil changes, which justified the need for higher resistance values as indicated in the data. When combined with elevation contours and water discharge data, these adjustments resulted in a flood map that closely approximates real-world conditions for both training and test data.

The analysis using Sentinel-1 satellite imagery, with reflectance values below 70, produced a flood map that was compared with flood maps generated by the HEC-RAS software. The comparison revealed that the flood map from Sentinel-1 imagery dated August 17, 2021 (Observed Data) identified an affected area of approximately 61 km², which is close to the flood map from the HEC-RAS model using flow data from August 14-17, 2021, which indicated an affected area of approximately 64 km². This result was in an error of about 6%, a value acceptable for calibration purposes. Therefore, this model can be used for further accuracy analysis.

B. Simulation performance

For calibration, accuracy was assessed over two years: training data (2021) and test data (2022). The process involved measuring accuracy using binomial probability theory through two sampling procedures: one with an expected accuracy of 85% and an acceptable error margin of 5%, and another with an expected accuracy of 95% and an acceptable error margin of 5%. The sample sizes were 203 and 76, respectively. Spatial sampling was then conducted using ArcGIS Pro 3.0.2, and the analysis was performed using the Confusion Matrix tool. The metrics evaluated were Accuracy, F1, and F2, with the requirement that these values exceed 50%. The results of the experiments are presented in table II

TABLE II. MODEL PERFORMANCE RESULT

Sampling	Expected Accuracy 85% (n = 203)		Expected Accuracy 95% (n = 76)		% Diff
	Train data 2021	Test Data 2022	Train data 2021	Test Data 2022	
TP	83	77	34	30	
FP	60	40	21	14	
FN	19	37	7	8	
TN	41	48	13	24	
Precision	58%	66%	-8%	62%	68%
Recall	81%	68%	14%	83%	79%
Accuracy	61%	62%	-1%	63%	71%
F1 Score:	68%	67%	1%	71%	73%
F2 Score:	75%	67%	8%	78%	77%
					1%

The study evaluated the model's performance with an expected accuracy of 85% using a sample size of 203. For the 2021 training data, the model achieved 61% accuracy, a 68% F1 score, and a 75% F2 score. For the 2022 test data, it reached 62% accuracy, a 67% F1 score, and a 67% F2 score, showing a slight improvement in accuracy and F1 score, and a notable increase in the F2 score.

In the second sampling, with an expected accuracy of 95%, the model had 63% accuracy, a 71% F1 score, and a 78% F2 score for the 2021 training data. For the 2022 test data, accuracy improved to 71%, the F1 score was 73%, and the F2 score was 77%. This resulted in an 8% increase in accuracy, a 2% decrease in the F1 score, and a 1% increase in the F2 score.

The results indicate that the model performs better with higher expected accuracy. For the 85% expected accuracy, there were minor changes in performance metrics between training and test data, with a notable increase in the F2 score. For the 95% expected accuracy, the model showed significant improvement in accuracy and slight variations in F1 and F2 scores between training and test datasets. Overall, while there are some fluctuations in individual metrics, the model's performance generally improves with higher expected accuracy, reflecting better consistency and reliability in the test data. Nonetheless, it is observed that both sets of results have values greater than 40%, which are considered acceptable, and values greater than 50% indicate that the model is suitable for use in practical applications.

C. Simulation Result

The application of the HEC-RAS model for flood area assessment in this study used water discharge data for simulation at two levels 1) Higher than the train data included maximum water level at station Kh. 89 is 2.65 meters with a maximum water flow of 106 m³/s, while the maximum water level at station Kh. 72 is 4.80 meters with a maximum water flow of 173 m³/s. 2) The highest recorded level consisted of maximum water level at station Kh. 89 is 2.95 meters with a maximum water flow of 124 m³/s, and the maximum water level at station Kh. 72 is 5.80 meters with a maximum water flow of 221 m³/s. Polynomial Regression was used to evaluate the model, the results show an R² value of 0.9891 and an RMSE value of 0.7516064, indicating minimal error and high reliability. Therefore, it can be confidently used to predict water flow at the specified station levels. The results of the study are shown in the following:

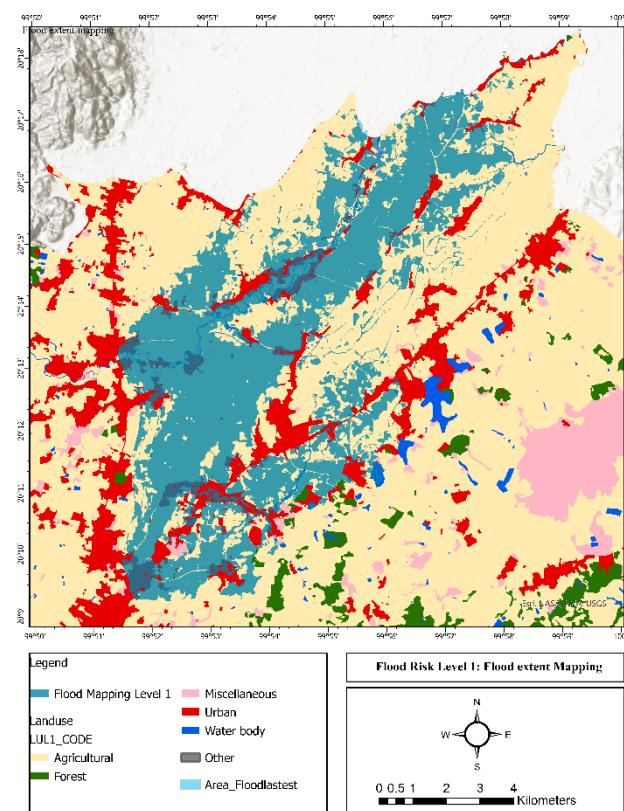


Fig. 2. Flood Risk Level 1: Flood extent Mapping

The results generated two levels of flood maps. When these maps were overlaid with the land use data to estimate the affected areas as figure 2 The flood map at Level 1 revealed that the total affected area is approximately 56.48 km², with most of the impact on agricultural land, covering 49.90 km². Other affected areas included urban land, which accounted for 4.93 km², and water bodies with 1.09 km². Miscellaneous land is also impacted, spanning 0.538 km², while forest land shows minimal effect at 0.005 km².

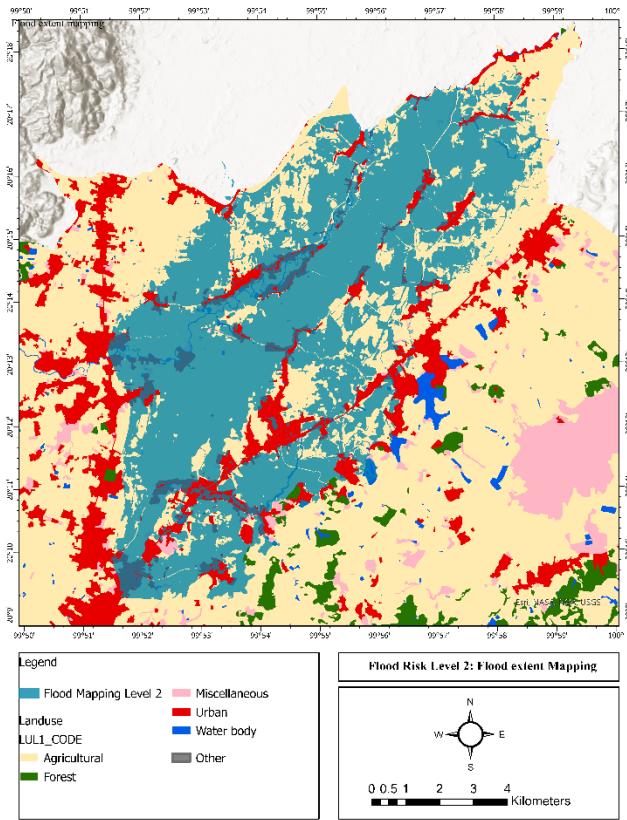


Fig. 3. Flood Risk Level 2: Flood extent Mapping

The next figure 3 shows the extended flood map as the water discharge increases based on the forecast. The total flood extent area now covered 77.1242 km², with the largest impact still on agricultural land, which now spanned 68.9265 km². The urban area affected has expanded to 6.2899 km², while forest land impacts have slightly increased to 0.0064 km². Miscellaneous land covers 0.6329 km², and the affected water body area has decreased slightly to 1.1479 km².

TABLE III. AFFECTED AREAS

Land use and area	Risk lv 1 (km ²)	Risk lv 2 (km ²)	Differences
Flood extent map	56.48	77.1242	37%
Agricultural Land	49.9063	68.9265	38%
Forest Land	0.005	0.0064	28%
Miscellaneous Land	0.5386	0.6329	18%
Urban	4.9392	6.2899	27%
Water body	1.0915	1.1479	5%

From table III, the total flood extent area increased by 37% with the impact on agricultural land growing by 38%, indicating a slightly larger area affected as floodwaters expanded. Urban areas saw a 27% increase, signifying more built-up areas under threat in higher flood risk scenarios.

In addition, forest land experienced an increase of 28%, though the absolute affected area remains small. Miscellaneous land affected at 18% increased. Interestingly, the water body area increased by 5%, possibly due to changes in the distribution of floodwater or shifts in the flood patterns.

IV. CONCLUSION AND DISCUSSION

This study concludes that the calibrated HEC-RAS flood maps can be validated for accuracy using satellite imagery from Sentinel-1. Two datasets and two sampling trials were conducted, revealing that the overall accuracy exceeded 0.5, indicating that the model is suitable for simulating flood scenarios. When assessing overfitting, it was found that the difference in accuracy between the training and test data for sampling with expected accuracies of 85% and 95% was approximately 1% and 8%, respectively. The difference in F1 scores for expected accuracies of 85% and 95% was around 1% and 2%, respectively, while the difference in F2 scores was approximately 8% and 1%, respectively. These findings suggest that the model is not overfitted.

The analysis shows that floods severely impact the agricultural sector, which aligns with Mae Chan district's primary occupation of agriculture. Urban areas are also affected, though to a lesser extent. This highlights that floods heavily damage agricultural lands and residential areas, demonstrating that extreme climate change increases flooding and the proportion of affected areas.

In satellite imagery, several studies have reported consistent results, with model accuracy ranging from 60-70% [17], [15]. Polarized VH and VV produce similar results, although some research indicates that VV provides higher accuracy, ranging from 80-90% [18]. Further investigation into the quality of calibrated satellite imagery is needed. Some studies suggest Sentinel-1 imagery may over-predict by over 60% due to low pixel resolution and reflections from surfaces like rooftops [13]. This contrasts with HEC-RAS, which focuses on terrain elevation or DEM [34]. Higher-resolution satellites like Sentinel-2 may improve HEC-RAS accuracy, as suggested by some studies [32]. However, this research is more appropriate for areas lacking Sentinel-2 imagery.

Another aspect is the difference in calibration and validation processes, particularly the use of Manning's n, which has produced accuracy levels similar to several studies [17], [15]. These studies mainly focus on agricultural areas [15]. For improved accuracy, validation and calibration using Main Channels (Rivers) n based on upstream and downstream water discharge may be necessary [18], [29], [32], [33]. However, this research is particularly suitable for areas where it is not feasible to calibrate using upstream to downstream discharge stations. Such areas, often located in rural regions both in Thailand and worldwide, lack comprehensive water level monitoring along entire river systems. Therefore, this study serves as a contribution by demonstrating that, even in areas with limitations on calibration from water stations, satellite imagery can be effectively used for calibration with reasonable accuracy.

Future research should focus on refining the calibration process using satellite data, improving the adjustment of various parameters, and placing greater emphasis on rural areas that face measurement constraints yet are significantly impacted by flooding.

ACKNOWLEDGMENT

This study is a part of a doctoral dissertation research in Applied Geoinformatics at the School of Information and Communication Technology, University of Phayao. We would like to express our gratitude to the European Space Agency for generously providing satellite imagery, Upper Northern Region Irrigation Hydrology Center of Thailand for

generously providing water discharge and level and to the Land Development Department of Thailand for supplying essential land use maps. Their contributions were crucial to the success of this research and the future research.

REFERENCES

- [1] Krzysztof Goniewicz, N. Islam, and M. Schoch-Spana, "Reimagining natural hazards and disaster preparedness: charting a new course for the future," *BMC Public Health*, vol. 23, no. 1, Mar. 2023, doi: <https://doi.org/10.1186/s12889-023-15497-y>.
- [2] D. Eckstein, V. Künzel, and L. Schäfer, "GLOBAL CLIMATE RISK INDEX 2021," 2021.
- [3] S. H. Erena, H. Worku, and F. De Paola, "Flood hazard mapping using FLO-2D and local management strategies of Dire Dawa city, Ethiopia," *Journal of Hydrology: Regional Studies*, vol. 19, pp. 224–239, Oct. 2018, doi: <https://doi.org/10.1016/j.ejrh.2018.09.005>.
- [4] A. G. Nigusse and O. G. Adhanom, "Flood Hazard and Flood Risk Vulnerability Mapping Using Geo-Spatial and MCDA around Adigrat, Tigray Region, Northern Ethiopia," *Momona Ethiopian Journal of Science*, vol. 11, no. 1, p. 90, May 2019, doi: <https://doi.org/10.4314/mejs.v11i1.6>.
- [5] Mayrén Alavez-Vargas, G. González-Celada, C. Birkel, R. Fonseca, and José Agustín Breña-Naranjo, "Estimated potential erosion in the Papaloapan River: Modeling efficiency and uncertainty" *Tecnología y ciencias del agua*, vol. 14, no. 4, pp. 349–414, Jul. 2023, doi: <https://doi.org/10.24850/j-tyc-14-04-08>.
- [6] Noor Suraya Romali, Zulkifli Yusop, and Abdul Razak Ismail, "Hydrological Modelling using HEC-HMS for Flood Risk Assessment of Segamat Town, Malaysia," IOP conference series, vol. 318, pp. 012029–012029, Mar. 2018, doi: <https://doi.org/10.1088/1757-899x/318/1/012029>.
- [7] I. Sufiyan, M. J.I, and A. T. Ogah, "HYDROLOGIC ASSESSMENT OF FOOD USING SWAT AS GEOSPATIAL TECHNIQUES IN THE CATCHMENT AREA OF TERENGGANU MALAYSIA," *Malaysian Journal of Geosciences*, vol. 4, no. 2, pp. 90–95, Aug. 2020, doi: <https://doi.org/10.26480/mjg.02.2020.90.95>.
- [8] L. A. Mohammed, S. I. Khassaf, and K. R. Al-Murshidi, "Application HEC-RAS Model to Simulate the Flood Wave Due to Dam Failure," Apr. 2019, doi: <https://doi.org/10.1109/sicn47020.2019.9019369>.
- [9] M. Podhoranyi, L. Vojacek, and D. Vojtek, "Flood Risk Monitoring by Using 2D Hydrodynamic Modeling: A Case Study of Frýdek-Místek City," Aug. 2018, doi: <https://doi.org/10.1109/ic3.2018.00-13>.
- [10] A. E. H. Tan, U. Ratnayake and E. K. A. Rahman, "GIS-based techniques to identify rainfall-induced flood risk zones (blue spots): A case study of Brunei River catchment," 7th Brunei International Conference on Engineering and Technology 2018 (BICET 2018), Bandar Seri Begawan, Brunei, 2018, pp. 1–4, doi: [10.1049/cp.2018.1555](https://doi.org/10.1049/cp.2018.1555).
- [11] Prema Malali and G. C. Bellad, "Flood Plain Analysis by using ArcSWAT and HEC-RAS -A Case study," Mar. 2024, doi: <https://doi.org/10.1109/b-htc60740.2024.10564045>.
- [12] Nur Aishah Sulaiman, Nur, Nazirah Md. Tarmizi, Abd Manan Samad, and W. Zurina, "Integration of Geographic Information System (GIS) and hydraulic modelling to simulate floodplain inundation level for Bandar Segamat," Aug. 2014, doi: <https://doi.org/10.1109/icsgrc.2014.6908706>.
- [13] Lahsen Wahib Kebir, Omar Hadj Sahraoui, A. Bennia, and Fouzi Berrichi, "Flood Risk Assessment Using Hydraulic Modeling and Remote Sensing Data: A Case Study of the M'zab Valley, (Algeria)," Apr. 2024, doi: <https://doi.org/10.1109/m2gars57310.2024.10537443>.
- [14] I. Akar, Kaan Kalkan, Derya Maktav, and Y. Ozdemir, "Determination of land use effects on flood risk by using integration of GIS and remote sensing," Jun. 2009, doi: <https://doi.org/10.1109/rast.2009.5158202>.
- [15] A. Ezzine, S. Saidi, T. Hermassi, I. Kammessi, F. Darragi, and H. Rajhi, "Flood mapping using hydraulic modeling and Sentinel-1 image: Case study of Medjerda Basin, northern Tunisia," *The Egyptian Journal of Remote Sensing and Space Science*, Mar. 2020, doi: <https://doi.org/10.1016/j.ejrs.2020.03.001>.
- [16] I. Zotou, V. Bellos, A. Gkouma, V. Karathanassi, and V. A. Tsirhntzis, "Using Sentinel-1 Imagery to Assess Predictive Performance of a Hydraulic Model," *Water Resources Management*, vol. 34, no. 14, pp. 4415–4430, Jul. 2020, doi: <https://doi.org/10.1007/s11269-020-02592-7>.
- [17] I. Elkhrachy et al., "Sentinel-1 remote sensing data and Hydrologic Engineering Centres River Analysis System two-dimensional integration for flash flood detection and modelling in New Cairo City, Egypt," *Journal of Flood Risk Management*, vol. 14, no. 2, Feb. 2021, doi: <https://doi.org/10.1111/jfr3.12692>.
- [18] T. Melkamu, M. Bagyaraj, M. Adimaw, A. Ngusie, and S. Karuppantan, "Detecting and mapping flood inundation areas in Fogera-Dera Floodplain, Ethiopia during an extreme wet season using Sentinel-1 data," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 127, p. 103189, Oct. 2022, doi: <https://doi.org/10.1016/j.pce.2022.103189>.
- [19] E. Union., Key Publications of the European Union. 2016.
- [20] Land Development Department, "Land Development Department Database", Thailand.
- [21] E. Ostertagová, "Modelling using Polynomial Regression," *Procedia Engineering*, vol. 48, pp. 500–506, 2012, doi: <https://doi.org/10.1016/j.proeng.2012.09.545>.
- [22] L. Dasallas, Y. Kim, and H. An, "Case Study of HEC-RAS 1D–2D Coupling Simulation: 2002 Baeksan Flood Event in Korea," *Water*, vol. 11, no. 10, p. 2048, Sep. 2019, doi: <https://doi.org/10.3390/w11102048>.
- [23] A. Mihu-Pintilie, C. I. Cimpianu, C. C. Stoleriu, M. N. Pérez, and L. E. Paveluc, "Using High-Density LiDAR Data and 2D Streamflow Hydraulic Modeling to Improve Urban Flood Hazard Maps: A HEC-RAS Multi-Scenario Approach," *Water*, vol. 11, no. 9, p. 1832, Sep. 2019, doi: <https://doi.org/10.3390/w11091832>.
- [24] D. T. Bulti and B. G. Abebe, "A review of flood modeling methods for urban pluvial flood application," *Modeling Earth Systems and Environment*, vol. 6, no. 3, pp. 1293–1302, May 2020, doi: <https://doi.org/10.1007/s40808-020-00803-z>.
- [25] K. Fitzpatrick-Lins, "Comparison of sampling procedures and data analysis for a land-use and land-cover map," 1981.
- [26] A. M. Al-Abadi, "Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study," *Arabian Journal of Geosciences*, vol. 11, no. 9, May 2018, doi: <https://doi.org/10.1007/s12517-018-3584-5>.
- [27] Moe Moe Myin and L. Lai, "Flood Mapping System of Flooding Prone Area in Myanmar," Mar. 2024, doi: <https://doi.org/10.1109/icca62361.2024.10532869>.
- [28] C.-H. Chang, M.-K. Chung, S.-Y. Yang, C.-T. Hsu, and S.-J. Wu, "A Case Study for the Application of an Operational Two-Dimensional Real-Time Flooding Forecasting System and Smart Water Level Gauges on Roads in Tainan City, Taiwan," *Water*, vol. 10, no. 5, p. 574, Apr. 2018, doi: <https://doi.org/10.3390/w10050574>.
- [29] M. A. Afzal et al., "Flood Inundation Modeling by Integrating HEC-RAS and Satellite Imagery: A Case Study of the Indus River Basin," *Water*, vol. 14, no. 19, p. 2984, Sep. 2022, doi: <https://doi.org/10.3390/w14192984>.
- [30] N. Ongdas, F. Akiyanova, Y. Karakulov, A. Muratbayeva, and N. Zinabdin, "Application of HEC-RAS (2D) for Flood Hazard Maps Generation for Yesil (Ishim) River in Kazakhstan," *Water*, vol. 12, no. 10, p. 2672, Sep. 2020, doi: <https://doi.org/10.3390/w12102672>.
- [31] E. Stephens and P. Bates, "Assessing the reliability of probabilistic flood inundation model predictions," *Hydrological Processes*, vol. 29, no. 19, pp. 4264–4283, May 2015, doi: <https://doi.org/10.1002/hyp.10451>.
- [32] F. Saeed and A. Zaidi, "Application of Geohecras for 2D Flood Modeling of Downstream of Guddu Barrage," *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2022, doi: <https://doi.org/10.1109/igarss46834.2022.9883618>.
- [33] P.G.T.N. Geeshan, T.M.N. Wijayarathne, and J.M.A.S.S. Jayasundara, "Vulnerability of Colombo Suburbs for Kelani River Floods," Nov. 2023, doi: <https://doi.org/10.1109/mercon60487.2023.10355474>.
- [34] M. Farooq, M. Shafique, and M. S. Khattak, "Flood hazard assessment and mapping of River Swat using HEC-RAS 2D model and high-resolution 12-m TanDEM-X DEM (WorldDEM)," *Natural Hazards*, vol. 97, no. 2, pp. 477–492, Jun. 2019, doi: <https://doi.org/10.1007/s11069-019-03638-9>.

Evaluating Models for Urban LULC Classification by Integrating Earth Engine and Python Ecosystem.

Sitthisak Moukomla
Geography Department
Faculty of Liberal Arts
Thammasat University
Pathum Thani 12120 THAILAND.
moukomla@tu.ac.th

Krudsakron Auynirundronkool
Geo-Informatics and Space Technology
Development Agency (Public
Organization)
Bangkok 10210, THAILAND.
krudsakron@gistda.or.th

Phurith Meeprom,
Faculty of Geo-Informatics
Burapha University,
Chonburi 20131, THAILAND.
phurith@buu.ac.th

Supaporn Manajitprasert
Geography department,..
Faculty of Education
Ramkhamhaeng University,
Bangkok 10240, THAILAND
smanajitprasert@gmail.com

Abstract— Urban climate change studies necessitate long-term monitoring to manage urban environments. The Bangkok Metropolitan Region (BMR) faces challenges like increased urban heat island effects, flooding, and air pollution. Long-term land use and land cover (LULC) change monitoring is crucial for analyzing urban development. Landsat series data is vital for tracing LULC changes over 40 years, but classification accuracy often suffers due to spectral overlap. This study introduces deep learning techniques to enhance classification accuracy. We compare three CNN architectures: Conv1D, Conv2D, and VGG16. Conv1D achieved the highest accuracy of 94.86% with a Kappa score of 0.890, showing its effectiveness in handling spectral information. Conv2D achieved an accuracy of 94.25% and a Kappa score of 0.876 by leveraging spatial context. VGG16 showed slightly lower performance with an accuracy of 87.58% and a Kappa score of 0.803. Our study evaluates these models for urban LULC classification using Landsat data specific to BMR. By understanding each model's strengths and limitations, we aim to recommend the most suitable CNN architecture for accurate, scalable urban LULC mapping, contributing to better urban environmental management.

Keywords— Urban LULC, Earth Engine, Models Evaluation

I. INTRODUCTION

Research on urban climate change necessitates comprehensive long-term monitoring to effectively comprehend and manage the shifting dynamics within urban ecosystems [1]. The Bangkok Metropolitan Region (BMR) is encountering considerable challenges because of urban climate change, manifesting in phenomena such as intensified urban heat island effects, increased flooding, and heightened air pollution. The systematic monitoring of land use and land cover (LULC) changes over extended periods is essential for documenting and analyzing urban development and environmental transformations within the BMR. To effectively trace LULC changes over the past four decades, access to long-term data sources, such as those provided by the Landsat satellite series, is critical. Nevertheless, the classification of urban LULC utilizing Landsat data is frequently hindered by low accuracy levels, primarily attributed to spectral overlap, wherein distinct land cover types exhibit analogous spectral signatures [2]. This challenge complicates the classification process, rendering it difficult to accurately differentiate between various urban features [3].

To tackle the challenge of low classification accuracy caused by spectral overlap, we propose the utilization of advanced deep learning methodologies. Notably, Convolutional Neural Networks (CNNs) have exhibited considerable potential in enhancing classification performance within intricate urban settings [4]. These networks have achieved notable success across various image classification applications, including Land Use and Land Cover (LULC) mapping, owing to their capacity to autonomously extract hierarchical features from unprocessed data [5]. This inherent ability renders them particularly adept at managing the complexity and diversity characteristic of urban environments [6]. However, identifying the most appropriate deep learning model for our specific research context remains an unresolved issue. Therefore, this study seeks to evaluate three leading CNN architectures to ascertain the optimal approach for urban LULC classification utilizing Landsat data, with a concentrated focus on the BMR.

Here, we focus on comparing three prominent CNN architectures: one-dimensional convolutional networks (Conv1D), two-dimensional convolutional networks (Conv2D), and the VGG16 model.

One-dimensional convolutional networks, or Conv1D, are built to handle data that flows in a line [7]. They shine when the order of data matters. In LULC classification, Conv1D proves useful for examining the spectral details from various bands of satellite images. By viewing each pixel's spectral signature as a sequence, these models can uncover the complex patterns and connections between spectral bands, which may enhance classification accuracy in regions where bands intertwine closely. While two-dimensional convolutional networks, or Conv2D, have made a name for themselves in processing images [8]. They are adept at recognizing shapes and structures, which allows them to understand the layout of what they see. In classifying land use and land cover in cities, these networks use the arrangement of pixels to tell one type of land from another. When paired with high-resolution satellite images, they become even more powerful. This ability to grasp spatial details aids in pinpointing and categorizing the intricate features of urban landscapes.

Also, we have chosen the VGG16 model. It has a solid structure and has proven itself in many tasks of image

classification. VGG16 is simple yet deep, with 16 layers and small receptive fields [9], [10]. This design helps it capture fine details and layered features well. Its strength lies in its robustness and its ability to perform across various datasets, making it a good choice for classifying land use and land cover in urban areas. We will look at different methods to tackle challenges like spectral overlap and mixed pixels while also finding scalable solutions for urban LULC classification. By understanding what each model does well and where it falls short, we will find the best deep learning framework for using Landsat data in urban land classification.

This study, therefore, aims to measure how well the Conv1D, Conv2D, and VGG16 models perform in classifying land use and land cover in the urban sprawl of Bangkok. We will examine how each model picks up the distinct features and patterns of the city's landscape. We will also look at the strengths and weaknesses of each framework. Furthermore, we will assess how efficient and scalable these methods are for large-scale mapping of urban areas. In the end, we hope to offer guidance on which CNN architecture best fits the land use classification needs and the specific data of the Bangkok region. The results will guide future research and practical uses of deep learning in remote sensing for managing urban environments. In the end, this study confronts the tough challenges of urban landscapes, helping us monitor and manage these ever-changing places better.

II. LITERATURE REVIEW

Land Use and Land Cover (LULC) classification using remote sensing imagery is crucial for environmental monitoring, urban planning, and resource management. The advent of deep learning techniques, particularly Convolutional Neural Networks (CNNs), has significantly improved the accuracy and efficiency of LULC classification. Developing a deep learning CNN model for this purpose involves several key steps, from data preparation to model deployment.

The process begins with data collection and preparation. High-resolution satellite imagery from sources such as Landsat, Sentinel-2, or commercial satellites provides the foundation for analysis [11], [12]. These images are preprocessed through resizing, normalization, and data augmentation techniques to create a robust dataset for training. Annotated datasets like the UC Merced Land Use Dataset or EuroSAT can be valuable resources, though custom-labeled datasets may also be necessary for specific applications [13].

The heart of the LULC classification system lies in the CNN architecture design. CNNs are particularly well-suited for this task due to their ability to capture spatial hierarchies in images [14]. A typical architecture includes an input layer corresponding to the image dimensions, followed by multiple convolutional layers with filters to detect features. Pooling layers are interspersed to reduce spatial dimensions, while fully connected layers and a softmax output layer complete the classification process. The specific architecture can be tailored to the complexity of the classification task and the available computational resources [15].

Training the model is a critical phase that involves careful selection of loss functions, optimizers, and evaluation metrics. Categorical cross-entropy is commonly used as the loss function for multi-class classification, while optimizers

like Adam or SGD guide the learning process. The model's performance is typically assessed using metrics such as accuracy, precision, recall, and F1-score. The training process involves splitting the data into training, validation, and test sets to ensure the model generalizes well to unseen data[16].

After initial training, the model undergoes evaluation and fine-tuning. This iterative process involves assessing the model's performance on a test dataset and adjusting improve its accuracy. Fine-tuning may include tweaking hyperparameters, adding regularization techniques like dropout or batch normalization, or experimenting with different network architectures such as ResNet or VGG[9], [10]. The goal is to achieve the highest possible classification accuracy while avoiding overfitting.

The trained model is prepared for deployment and application. This may involve converting the model to a deployable format like TensorFlow and integrating it with GIS systems or web applications for real-time classification. The deployed model can then be used for various applications such as monitoring changes in land use over time or supporting decision-making in urban planning and environmental conservation.

III. METHODOLOGY

A. Data Preparation

In this study, LULC for the BMR (Fig 1) was classified based on Landsat imagery. Preprocessing was done in Google Earth Engine to ensure the normalization of images through the processes of atmospheric correction and cloud masking. Data were split into training, validation, and test sets. The model is first trained on the training set; then hyperparameter tuning is done using the validation set; and finally, performance is gauged on the test set. Each pixel is labeled using ground truth data collected from high-resolution satellite images and Google Street View.

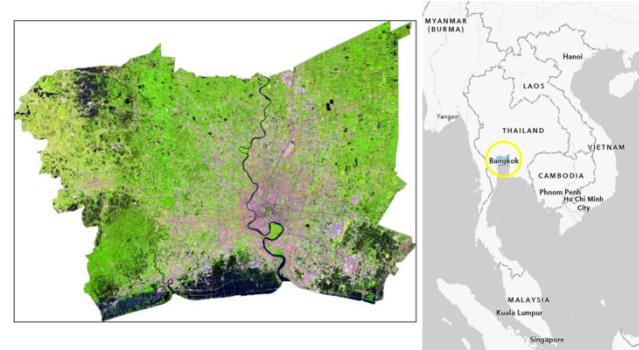


Fig 1 Study area for Bangkok Metropolitan Region

B. CNN Models

1D Convolutional Neural Network (1D CNN) Model uses sequential spectral data arising from the Landsat bands. The model uses three Conv1D blocks with two Conv1D layers each, followed by a MaxPooling1D layer and a Dropout layer with a dropout rate of 0.2 in each block. The Conv1D layers use 64, 128, and 256 filters with a kernel size of 2 and an activation function of ReLU. This is followed by a dimensionality reduction using GlobalMaxPooling1D and two dense layers. The last layer is the softmax layer for classification. This model is beneficial for detecting various

land cover types as it captures significant spectral-temporal characteristics in the data. The optimization techniques used are the same as the other models [17].

2D Convolutional Neural Network (2D CNN) Model used 2D-CNN spatial pattern recognition model. The input to the model is formed by reshaped image patches centered around each pixel. The architecture of the CNN consists of three blocks of convolution, each having two Conv2D layers followed by a MaxPooling2D layer and a Dropout layer of 0.3. The convolutional layers use 64 filters of size 3x3, 128 filters of size 3x3, and 256 filters of size 3x3; the activation function used is ReLU. The feature maps are flattened after all the convolutional layers and then further compressed through two dense layers before the softmax output layer. Since the model captures spatial dependencies and textures in the imagery for correct LULC classification, it is wise to employ the same training strategy as the FCNN but with adjustments for the more complex CNN [18].

The VGG16 model is a pre-trained deep convolutional neural network. It is fine-tuned for LULC classification by replacing the final layers to adapt to the specific number of classes in this study. The initial layers of VGG16, which are pre-trained on ImageNet, are used to extract features from the input images. The final layers, which are fully connected, are replaced with a new dense layer followed by a SoftMax activation layer for classification [10], [19]. This model leverages transfer learning to achieve high accuracy and efficiency.

C. Model Evaluation and Comparison

The performance of each model is evaluated using overall accuracy, kappa coefficient, precision, recall, and F1-score. Confusion matrices are generated to understand the classification performance for each LULC category. Robustness and generalization are checked through cross-validation. A comparative analysis of the strengths and weaknesses in spatial and spectral pattern recognition is conducted among the three models.

IV. RESULTS AND DISCUSSION

A. Model Performance Overview

The comparative analysis of three deep learning models—Conv1D, Conv2D, and VGG16—evaluated for their performance on a classification task. The key metrics (Table 1) considered are accuracy, Kappa, and F1-score, each providing a distinct perspective on how well the models perform. The results reveal important insights into the suitability of each model for this task, highlighting the trade-offs between complexity and performance.

TABLE I. MODEL CAMPARISON

Model	Key Metrics		
	Accuracy	Kappa	F1-Score
Conv1D	0.95	0.89	0.95
Conv2D	0.94	0.88	0.94
VGG16	0.88	0.80	0.88

The evaluation shows that Conv1D performs the best overall, with an accuracy of 0.95, Kappa score of 0.89, and F1-score of 0.95. This indicates not only the highest proportion of correct classifications but also strong agreement in predictions compared to random chance. The high F1-score

reflects a balance between precision and recall, meaning that Conv1D minimizes both false positives and false negatives effectively.

On the other hand, Conv2D closely follows with an accuracy of 0.94, Kappa of 0.88, and F1-score of 0.94. Although slightly behind Conv1D, its performance suggests that it can also handle the classification task well. Conv2D is typically effective for tasks that benefit from spatial relationships in 2D data, such as images, and its strong performance implies that some level of spatial structure might be present in the dataset. However, the minor difference in accuracy suggests that Conv1D's more straightforward, sequential processing of the data may be sufficient and more efficient for this particular task.

In contrast, VGG16, a well-known, pre-trained model, performs relatively lower across all metrics, with an accuracy of 0.88, Kappa of 0.80, and F1-score of 0.88. VGG16 is designed for complex image classification tasks and comes with significant computational overhead, which may not be necessary for this task. Its lower performance could stem from the fact that it is pre-trained on a large-scale dataset (ImageNet) and may not align well with the specific features of this dataset. This underperformance, compared to the simpler Conv1D and Conv2D models, suggests that VGG16's complexity does not provide a clear advantage here [20].

B. Model Training and Inference Time Comparison

The table II presents a comparison of the training and inference times for three models: Conv1D, Conv2D, and VGG16.

TABLE II. MODEL TRANING AND INFERENCE TIME CAMPARISON

Model	Training Time (seconds)	Inference Time (seconds)
Conv1D	26.4	0.68
Conv2D	3941.3	20.47
VGG16	40799.8	211.80

From the results, it is evident that Conv1D is the most efficient model in terms of both training and inference time. With a training time of only 26.4 seconds, Conv1D is exceptionally fast to train. This model also has the shortest inference time, at 0.68 seconds, making it a highly suitable option for applications that demand both quick training and real-time predictions. The minimal computational resources required by Conv1D make it ideal for tasks where speed is critical without sacrificing accuracy [7].

On the other hand, Conv2D requires significantly more time to train, taking 3941.3 seconds (approximately 1.1 hours) to complete the training process. This is over 149 times longer than Conv1D, reflecting the additional complexity that Conv2D introduces with its two-dimensional convolutions. Additionally, Conv2D's inference time is also much higher than Conv1D's, taking 20.47 seconds. Although this is considerably slower than Conv1D, Conv2D's more complex architecture may be justified in cases where 2D spatial information is essential to achieving better performance [21].

VGG16, as expected for a deep and highly complex model, exhibits the longest training and inference times by far. With a training time of 40799.8 seconds (about 11.33 hours), VGG16 is over 1000 times slower than Conv1D and requires considerable computational resources. Its inference time, at 211.80 seconds, also demonstrates the high computational

demand of the model. While VGG16 is known for its powerful performance in large-scale image classification tasks, its resource intensity makes it less practical for time-sensitive applications or projects that require fast training and inference cycles[22], [23].

V. CONCLUSION

Conv1D stands out as the most efficient model in terms of both training and inference time. Its simplicity enables rapid training and fast predictions, making it the optimal choice for tasks where computational efficiency is a priority. Conv2D, while more complex, offers a reasonable balance between performance and computational demand. It is well-suited for tasks that benefit from the extraction of 2D spatial features, but this added complexity comes at the cost of longer training and inference times. On the other hand, VGG16—despite being a highly effective model for large-scale image classification—demands a considerable amount of time for both training and inference. This makes it less suitable for applications that require quick processing or real-time predictions. The results demonstrate that in this case, the complexity of VGG16 does not yield a significant improvement in performance over simpler alternatives, particularly when considering the computational overhead involved. Ultimately, Conv1D's simplicity, speed, and effectiveness make it the most reliable choice for this specific classification task, while Conv2D offers a more moderate option for tasks requiring additional spatial analysis. VGG16, though powerful, proves to be less practical for time-sensitive scenarios in this context.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers. Their keen insights and helpful suggestions have made this manuscript better.

REFERENCES

- [1] Y. Su, C. Lu, Y. Su, Z. Wang, Y. Huang, and N. Yang, “Spatio-temporal Evolution of Urban Expansion Based on a Novel Adjusted Index and GEE: A Case Study of Central Plains Urban Agglomeration, China,” *Chin Geogr Sci*, vol. 33, no. 4, pp. 736–750, Aug. 2023, doi: 10.1007/s11769-023-1369-6.
- [2] S. Amini, M. Saber, H. Rabiei-Dastjerdi, and S. Homayouni, “Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series,” *Remote Sens (Basel)*, vol. 14, no. 11, Jun. 2022, doi: 10.3390/rs14112654.
- [3] M. Zhou et al., “Urban sprawl and changes in land-use efficiency in the beijing-tianjin-hebei region, china from 2000 to 2020: A spatiotemporal analysis using earth observation data,” *Remote Sens (Basel)*, vol. 13, no. 15, Aug. 2021, doi: 10.3390/rs13152850.
- [4] H. Yassine, K. Tout, and M. Jaber, “Improving lulc classification from satellite imagery using deep learning—eurosat dataset,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 369–376, 2021.
- [5] M. Thanga Manickam, M. Karthik Rao, K. Barath, S. Shree Vijay, and R. Karthi, “Convolutional neural network for land cover classification and mapping using landsat images,” in *Innovations in Computer Science and Engineering: Proceedings of the Ninth ICICSE*, 2021, Springer, 2022, pp. 221–232.
- [6] A. T. Balarabe and I. Jordanov, “LULC image classification with convolutional neural network,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 2021, pp. 5985–5988.
- [7] M. I. Habibie et al., “The development land utilization and cover of the Jambi district are examined and forecasted using Google Earth Engine and CNNID,” *Remote Sens Appl*, vol. 34, p. 101175, 2024.
- [8] N. Naik, K. Chandrasekaran, V. Meenakshi Sundaram, and P. Panneer, “Assessment of land use and land cover change detection and prediction using deep learning techniques for the southwestern coastal region, Goa, India,” *Environ Monit Assess*, vol. 196, no. 6, pp. 1–34, 2024.
- [9] S. Chen, F. Lei, Z. Zang, and M. Zhang, “Forest mapping using a VGG16-UNet++ & stacking model based on Google earth engine in the urban area,” *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
- [10] M. Gulhane and S. Kumar, “Oriental Method to Predict Land Cover and Land Usage Using Keras with VGG16 for Image Recognition,” *Advances in Aerial Sensing and Imaging*, pp. 33–46, 2024.
- [11] D. A. Arrechea-Castillo, Y. T. Solano-Correa, J. F. Muñoz-Ordóñez, E. L. Pencue-Fierro, and A. Figueroa-Casas, “Multiclass land use and land cover classification of Andean Sub-Basins in Colombia with Sentinel-2 and Deep Learning,” *Remote Sens (Basel)*, vol. 15, no. 10, p. 2521, 2023.
- [12] C. Liao et al., “Synergistic use of multi-temporal RADARSAT-2 and VENµS data for crop classification based on 1D convolutional neural network,” *Remote Sens (Basel)*, vol. 12, no. 5, p. 832, 2020.
- [13] M. A. Günen, “Performance comparison of deep learning and machine learning methods in determining wetland water areas using EuroSAT dataset,” *Environmental Science and Pollution Research*, vol. 29, no. 14, pp. 21092–21106, 2022.
- [14] J. Yao, T. Mitran, X. Kong, R. Lal, Q. Chu, and M. Shakut, “Landuse and land cover identification and disaggregating socio-economic data with convolutional neural network,” *Geocarto Int*, vol. 35, no. 10, pp. 1109–1123, 2020.
- [15] S. O. Atik and C. Ipbuker, “Integrating convolutional neural network and multiresolution segmentation for land cover and land use mapping using satellite imagery,” *Applied Sciences*, vol. 11, no. 12, p. 5551, 2021.
- [16] R. Naushad, T. Kaur, and E. Ghaderpour, “Deep transfer learning for land use and land cover classification: A comparative study,” *Sensors*, vol. 21, no. 23, p. 8083, 2021.
- [17] W. R. Moskolai, W. Abdou, and A. Dipanda, “A comparative study of deep learning and CA-Markov methods for land use/land cover change prediction,” in *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE, 2022, pp. 190–197.
- [18] J. Jagannathan and C. Divya, “Deep learning for the prediction and classification of land use and land cover changes using deep convolutional neural network,” *Ecol Inform*, vol. 65, p. 101412, 2021.
- [19] A. Bhatt and V. T. Bhatt, “Dcrff-Lhrf: an improvised methodology for efficient land-cover classification on eurosat dataset,” *Multimed Tools Appl*, vol. 83, no. 18, pp. 54001–54025, 2024.
- [20] C. Acuña-Alonso, M. García-Ontiyuelo, D. Barbera-Barragáns, and X. Álvarez, “Development of a convolutional neural network to accurately detect land use and land cover,” *MethodsX*, vol. 12, p. 102719, 2024.
- [21] N. Zaabar, S. Niculescu, and M. M. Kamel, “Application of convolutional neural networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in Ain Témouchent, Algeria,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 5177–5189, 2022.
- [22] J. Jagannathan and C. Divya, “Deep learning for the prediction and classification of land use and land cover changes using deep convolutional neural network,” *Ecol Inform*, vol. 65, p. 101412, 2021.
- [23] M. Mohammadi and A. Sharifi, “Evaluation of convolutional neural networks for urban mapping using satellite images,” *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 9, pp. 2125–2131, 2021.

Assessing the Performance of Machine Learning Models using Satellite Dataset for Classification of Various Crop Types

Montree Pinkaeo
*Geospatial Data Science,
School Of Information And
Communication Technology,
University of Phayao,
Phayao, Thailand.
67025189@up.ac.th
Email: mp.dox69@gmail.com*

Phaisarn Jeefoo
*Geographic Information Science,
School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand.
Email: phaisarn.je@up.ac.th /
p.jeefoo@gmail.com
<https://orcid.org/0000-0001-8504-4026>*

Surachai Chantee
*Research assistant,
Technology and Digital Division,
Hydro-Informatics Institute
(Public Organization),
Bangkok, Thailand
Email: surachai@hiit.or.th*

Pornthep Rojanavasu
*School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand.
Email: pornthep.ro@up.ac.th*

Jirabhorn Chaiwongsai
*School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand.
Email: jirabhorn.ch@up.ac.th*

Prattana Deeprasertkul
*Researcher in Technology and
Information Research Section,
Digital and Technology Development
Division, Hydro-Informatics Institute
(Public Organization)
Bangkok, Thailand.
Email: Prattana@hiit.or.th*

Abstract— This paper aims to evaluate machine learning (ML) models' performance, named Random Forest (RF), Gradient Boosting (GB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) for classifying different crop types using images from the Sentinel 2 satellite. The result shows that RF demonstrated the highest values across all criteria, outperforming the other models with an overall accuracy of 0.7365 (+ 0.010). With an accuracy of 0.7183 (+0.016) overall, GB trailed closely behind. The two models demonstrated robust performance, as evidenced by their respective F1-scores of 0.7344 and 0.7161, which suggest a well-balanced compromise between recall and precision. With total accuracies of 0.5429 (+ 0.0024) and 0.5169 (+ 0.017), respectively, the ANN and SVM algorithms performed comparably worse. In comparison to ensemble techniques like RF and GB, this shows that these models might not be as effective for this specific classification problem. The findings of this study can inform decision-making in the selection of ML algorithms for various applications, such as spatial water management, specifically in agriculture and irrigation, agricultural land management, and planning to mitigate the impact of natural disasters in the agricultural sector.

Keywords— Machine Learning, Sentinel-2A satellite images, Geo-informatics, Classification, Mae Na Rua Sub-District

I. INTRODUCTION

António Guterres [1], the United Nations Secretary-General, stated that "The era of global warming has ended; the era of global boiling has arrived." This statement underscores the significant challenges in current water management. This increasing uncertainty affects the entire water industry[2], [3], [4], particularly the agricultural sector [5], which plays a crucial role in driving society by providing a source of food, generating household income, and contributing to national revenue through agricultural exports [6], [7], [8], [9].

Ineffective water management in the agriculture sector impacts many people directly in both the short and long term, including the country's economic system. Therefore, employing technology to assist in risk management, planning, and decision-making in water management is essential.

Current Geographic Information Systems (GIS) and remote sensing technologies are of high quality and accessible at no cost [10]. Notably, satellite imagery can be utilized for various applications such as monitoring changes in built structures, land cover, crop planning, and even water management [11], [12]. The classification of satellite images is a widely adopted process because the data can be used for analysis, planning, and decision-making in resource management. Many researchers have studied and developed efficient classification methods, using Machine Learning (ML) techniques [13]. Presently, there are numerous algorithms available for both supervised and unsupervised learning [14], each varying in complexity and effectiveness. Therefore, selecting appropriate ML tools requires careful consideration of their efficiency. Consequently, this paper aims to evaluate ML Models' Performance, named Random Forest (RF), Gradient Boosting (GB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) for classifying different crop types using satellite dataset for application in spatial water management, specifically in agricultural and irrigation purposes.

II. MATTERIAL AND METHOD

A. Study area: Mae Na Rua Sub-District, Mueang Phayao District, Phayao Province, Thailand.

The study area is located in the Mae Na Rua sub-district (MNR) of Phayao Province, situated geographically between 19°00' N to 19°08' N and 99°45' E to 99°54' E. This area encompasses approximately 28,953 acres (about 117.17 sq. km). Roughly one-third of the area is forested, while the

remainder consists of agricultural and residential zones, comprising 18 villages [15].

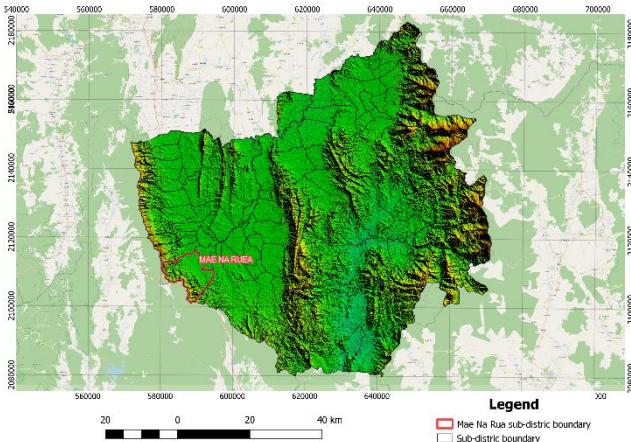


Fig. 1 Geographical location of the study area.

B. Data used

This study utilizes satellite images from Sentinel-2A. The criteria for selecting satellite images are that scenes must have clouds and haze no more than 10 percent of the image area or none at all, only the monthly images from 2020 were selected, with a resolution of 20 meters, it is composed of 10 bands: B02, B03, B04, B05, B06, B07, B08, B08A, B11, and B12. Finally, the monthly photos used include photos from January, February, March, and December.

Land use refers to various types of land utilization, such as forests, buildings, and agricultural areas, which can differentiate the types of crops cultivated in those areas. The land use data for the study area was provided by the Department of Land Development, which is generally accepted and used in Thailand, and was produced in 2020 in the form of shapefiles.

Finally, The MNR boundary was used to define the study area and the extent of the collected data, which was provided by the Hydro Informatics Institute in the form of a shapefile.

C. Methods

Analyzing machine learning's effectiveness requires multiple steps, including data gathering, data preparation, model development, model validation, and outcome comparison. (Figure 2)

Due to the broad scope of the collected satellite images and land use data, it was necessary to exclude unrelated areas, retaining only the study area to expedite the algorithm's inference and reduce the time needed to train the model [13], [16]. ArcMap software was used to crop all satellite images and land use data, using the MNR boundary as a mask to ensure all layers had the same number of pixels and could align correctly. Six classes were identified: (1) etc, (2) longan, (3) corn, (4) rice, (5) rubber, and (6) tobacco. For each class, 5,000 points were randomly selected as input data for machine learning, resulting in a total of 30,000 training points across all six classes. The cell values at locations specified in a training point from each raster were extracted and recorded in the attribute table of the training point, stored separately by band.

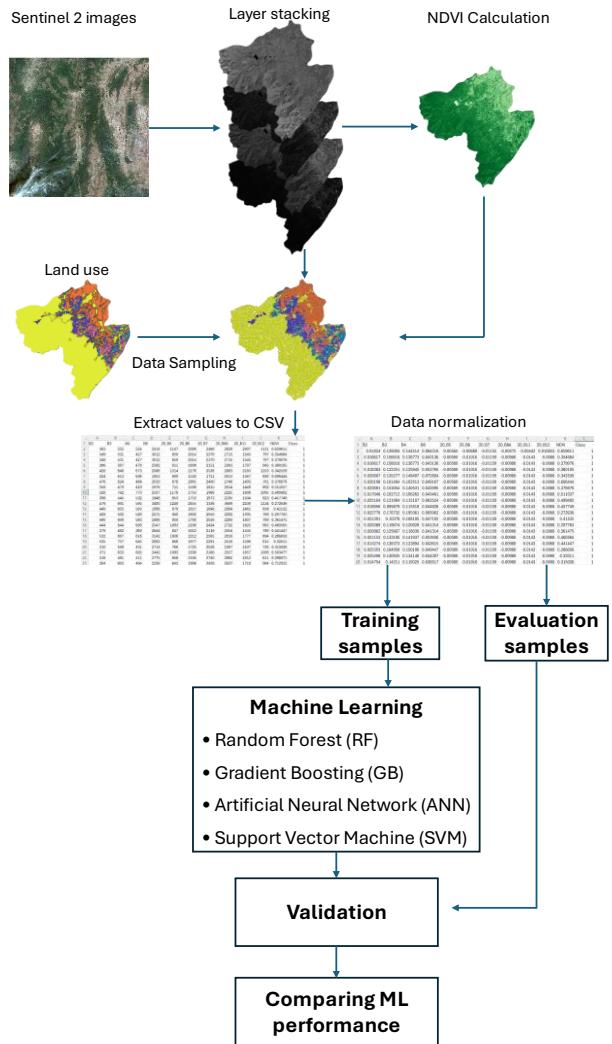


Fig. 2 Research methodology.

To increase the variety of input data, different indices such as the normalized difference vegetation index (NDVI) were added to the table. NDVI was calculated using bands 04 (Red) and 08 (NIR) with the following formula (1) [13], [16], [17], [18]

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

The cell values of NDVI were extracted and added to the attribute table of the training point. The attribute table was then exported as a CSV file for use with machine learning.

Many classification algorithms require the provided data set to be standardized to reduce bias in computation and improve processing speed. Therefore, the data in the attribute table of the training points were standardized using the Min-max technique with the following formula (2) [19], [20]:

$$v'_i = \frac{v_i - MIN_v}{MAX_v - MIN_v} \quad (2)$$

Where v'_i is the new value of the i feature, v_i is the original value of the i feature, $MIN(v)$ is the minimum value of that feature, and $MAX(v)$ is the maximum value of that feature.

Lastly, the training datasets were divided into data for training and testing in an 80:20 ratio from the entire datasets.

D. ML training

This study evaluates the performance of four popular algorithms: Random Forest (RF), Gradient Boosting (GB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) [12]. The models were trained using a High-Performance Computing (HPC) system equipped with an 8-core Intel Xeon 5218 processor, 128 GB of RAM, running Linux operating system, supported by the Hydro Informatics Institute (HII). During model training using the datasets, each model's hyperparameters were optimized using grid search and a 5-fold cross-validation [16] was employed to ensure model robustness and reasonable computational time. Finally, the model's performance is assessed using overall accuracy (OA), precision (PR), recall (RE), and F1-score (F1) values [21] to compare the models' performances and determine which model has the best classification efficiency.

III. RESULTS

A satellite dataset was used to assess the effectiveness of four machine learning models: RF, GB, ANN, and SVM for the classification of different crop varieties. As shown in Table 1, the performance metrics for these models, including OA, PR, RE, and F1, are summarized.

Table 1 Performance of each model.

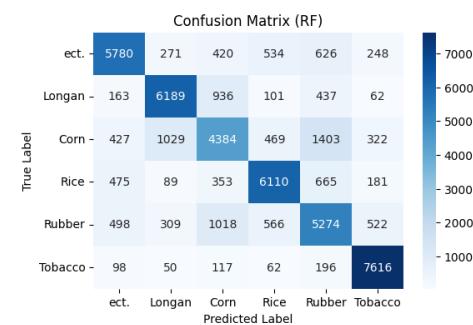
ML	OA	PR	RE	F1
RF	0.7365 (± 0.010)	0.7338	0.7365	0.7344
GB	0.7183 (± 0.016)	0.7153	0.7183	0.7161
ANN	0.5429 (± 0.0024)	0.5438	0.5429	0.5424
SVM	0.5169 (± 0.017)	0.5147	0.5169	0.5145

Table 1 shows that after using the grid search to find the best parameters (Table 2) and the highest accuracy of each model, RF demonstrated the highest values across all criteria, outperforming the other models with an overall accuracy of 0.7365 (± 0.010). With an accuracy of 0.7183 (± 0.016) overall, GB trailed closely behind. The two models demonstrated robust performance, as evidenced by their respective F1-scores of 0.7344 and 0.7161, which suggest a well-balanced compromise between recall and precision. With total accuracies of 0.5429 (± 0.0024) and 0.5170 (± 0.017), respectively, the ANN and SVM algorithms performed comparably worse.

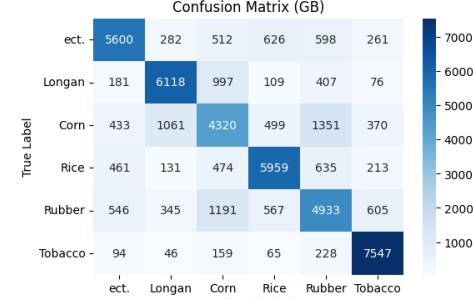
As table 2 demonstrates, RF model worked best with 1200 estimators, using the entropy criterion and a maximum depth of 50. GB was scored best with 1000 estimators, a learning rate of 0.3, and a maximum depth of 5. ANN model achieved optimal performance with three hidden layers of sizes 500, 250, and 500, using the Adam solver and a learning rate invscaling. Finally, SVM was best configured with a C value of 100, Gamma of 500, and the RBF kernel, while also balancing class weights.

Table 2 Best parameter

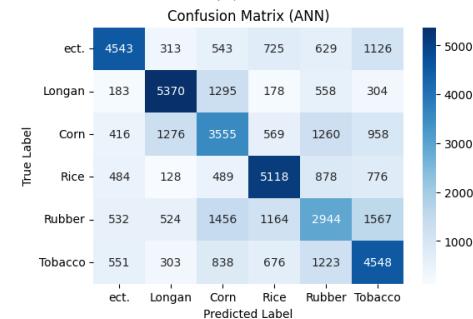
Algorithms	Best parameters
RF	N_estimators: 1200, Bootstrap: False, Criterion: entropy, Maximum depth: 50, Maximum features: sqrt, Minimum samples leaf: 4, Minimum samples split: 3
GB	N_estimators: 1000, Criterion: squared error, Learning rate: 0.3, Loss: log loss, Maximum depth: 5, Minimum samples split: 4, Subsample: 1.0
ANN	Hidden layer sizes: 500, 250, 500, Max iterations: 1200, Solver: adam, Learning rate: invscaling, Alpha: 0.001, Activation: tanh, Momentum: 0.1
SVM	C: 100, Gamma: 500, Kernel: RBF, Class weight: Balance



(a)



(b)

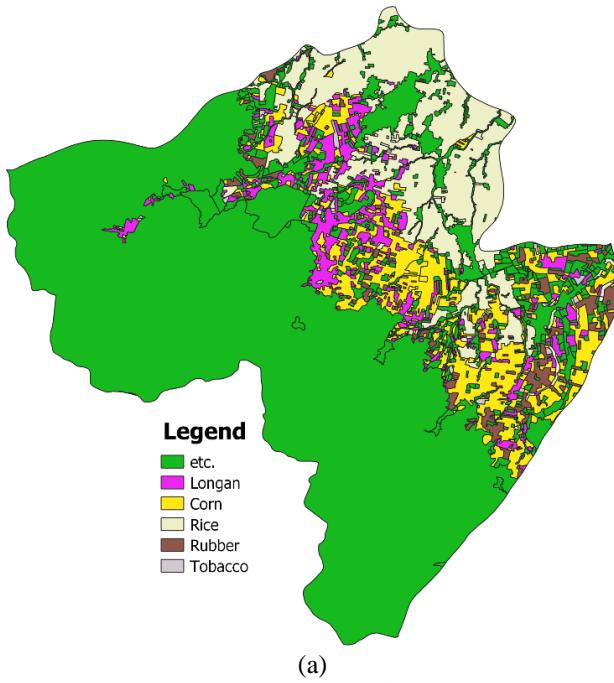


(c)

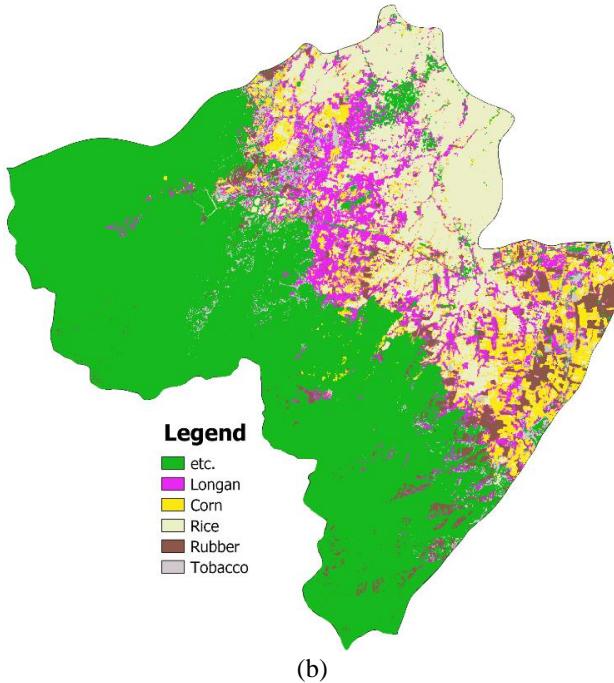
Confusion Matrix (SVM)						
True Label	ect.	Longan	Corn	Rice	Rubber	Tobacco
	ect.	Longan	Corn	Rice	Rubber	Tobacco
ect.	6950	538	633	1073	1037	1692
Longan	348	8228	1705	195	1016	437
Corn	855	2395	4317	792	2199	1481
Rice	1090	308	524	7114	1543	1272
Rubber	1091	1036	1494	1567	4494	2535
Tobacco	1356	683	1016	895	1973	6118

(d)

Fig. 3 Confusion matrix of each model: (a) RF, (b) GB, (c) ANN, (d) SVM



(a)



(b)

Fig. 4 Comparison between (a) Land use and (b) RF classification results.

Confusion matrices for each algorithm were examined to assess the models' performance more thoroughly. Figure 3 reveals that the model distinguishes between classes

effectively, especially the case for classes 2, 4, and 6, where significant numbers of the classes have accurate classifications.

It also shows that all models were confused in classifying some data classes, especially Class 3 and 5, which were the least correctly classified by all models.

The confusion matrices validate that the greatest performance for crop type classification using the satellite dataset is provided by ensemble algorithms, specifically RF, which yields more accurate forecasts with fewer misclassifications. As seen by their confusion matrices and overall performance metrics, the ANN and SVM models, while still functioning satisfactorily, are less successful in this particular environment.

Additionally, Figure 4 presents a comparison between the land use data, which was used to validate the model, and the predictions of the FR model, which achieved the highest accuracy among all models. It is evident that the algorithm was able to classify most crop types closely matching the land use data. However, an area for improvement is the classification of corn and rubber, which may require the incorporation of additional variables to enhance classification accuracy.

IV. DISCUSSION AND CONCLUSIONS

A. Discussion

The study results indicate that ensemble techniques like RF and GB achieved high OA, exceeding 0.7, while the ANN and SVM models had an OA of only up to 0.55. This suggests that ANN and SVM models may not be as effective for this specific classification problem. To further evaluate the performance of RF compared to other models, a T-test was conducted. The T-test results reinforce the superior effectiveness of the RF model. When comparing RF to GB, the T-statistic was 10.322 with a P-value of 6.697, indicating a statistically significant difference in favor of RF. The comparison between RF and ANN yielded a T-statistic of 33.342 and a P-value of 7.146, showing that RF performs considerably better than ANN. Finally, the comparison between RF and SVM resulted in a T-statistic of 52.144 and a P-value of 2.028, demonstrating that RF significantly outperforms SVM.

It is observed that the algorithm's accuracy is not particularly high, which could be due to several factors. For instance, the land use data used for validation lacks sufficient detail. This is evident from the absence of a road classification in the land use layer, despite the study area containing many roads. These roads were categorized based on the adjacent agricultural land. Additionally, the resolution limitations of Sentinel-2 satellite imagery may have contributed to this issue. Most of the bands used for training the model have a resolution of 20 meters, with only four bands: 2, 3, 4, and 8 offering the highest resolution of 10 meters. Furthermore, the variation in crop characteristics over time could be another influencing factor. For example, the satellite imagery used for model training was captured during a period when rubber trees (Class 5) were shedding leaves [22] and corn (Class 5) was in its harvesting phase, during which farmers typically reduce watering, causing the corn leaves to turn brown [23]. All these

factors likely contributed to the increase in model prediction errors.

Considering the model's performance, the results show that RF and GB exhibited high accuracy, while SVM had the lowest accuracy. These findings are consistent with the study by Ouchra, Belangour, and Erraissi (2023), which compared the performance of various ML models in classifying satellite imagery. Their results indicated that Minimum Distance, RF, and GB achieved the highest overall accuracy at 95.76%, 94.91%, and 93.22%, respectively, while the SVM model produced the lowest overall accuracy at 81.35%.

Furthermore, the study found that ANN's performance was not as effective compared to ensemble methods, which aligns with Abdi's (2020) findings. This may be due to the large number of classes and the lack of detail in the validation data, which hindered the model's ability to make accurate predictions.

B. Conclusions

This paper aims to evaluate ML models' performance, named Random Forest (RF), Gradient Boosting (GB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) for classifying different crop types using images from the Sentinel 2 satellite. The result shows that RF demonstrated the highest values across all criteria, outperforming the other models with an overall accuracy of 0.7365 (± 0.010). With an accuracy of 0.7183 (± 0.016) overall, GB trailed closely behind. The two models demonstrated robust performance, as evidenced by their respective F1-scores of 0.7344 and 0.7161, which suggest a well-balanced compromise between recall and precision. With total accuracies of 0.5429 (± 0.0024) and 0.5169 (± 0.017), respectively, the ANN and SVM algorithms performed comparably worse. In comparison to ensemble techniques like RF and GB, this shows that these models might not be as effective for this specific classification problem.

The findings of this study can inform decision-making in the selection of ML algorithms for various applications, such as spatial water management, specifically in agriculture and irrigation, agricultural land management, and planning to mitigate the impact of natural disasters in the agricultural sector.

However, this study faced some limitations: the land use data used for validation lacks sufficient detail, which may affect the accuracy of the model's performance. This suggests that the models may not have been trained or tested with comprehensive data. The satellite imagery used for training has a resolution of 20 meters for most bands, with only four bands (2, 3, 4, and 8) having a higher resolution of 10 meters. This limitation in resolution may reduce the models' ability to accurately classify crop types, as some details might be lost. Additionally, changes in crop characteristics over time can impact model performance. For example, satellite imagery was captured during specific growth phases of the crops, which may not represent their typical appearance throughout the growing season, potentially leading to misclassification.

For future studies, it is recommended to enhance the accuracy of validation data by incorporating ground truth data. This would improve the alignment between land use information and satellite imagery, thereby increasing the model's overall accuracy. Additionally, incorporating further machine learning techniques such as Classification and

Regression Trees and Convolutional Neural Networks etc., although they may require a trade-off in terms of increased processing time.

ACKNOWLEDGMENT

This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant Number B13F670082-8].

The authors thank the Hydro-Informatics Institute (Public Organization) and the University of Phayao for supporting the tools and software to conduct this study.

REFERENCES

- [1] United Nations, "Hottest July ever signals 'era of global boiling has arrived' says UN chief | UN News," UN News. Accessed: Jul. 10, 2024. [Online]. Available: <https://news.un.org/en/story/2023/07/1139162>
- [2] S. Chantee and T. Mayakul, "Modeling Flood Susceptible Areas Using Deep Learning Techniques with Random Subspace: A Case Study of the Mae Chan Basin in Thailand," *International Journal on Advanced science Engineering Information Technology*, vol. 14, no. 1, pp. 344–357, Feb. 2024.
- [3] Ministry of Agriculture and Cooperatives, "The government action plan of the Ministry of Agriculture and Cooperatives for a period of 5 years (2023-2027)," Bangkok, 2024.
- [4] S. Mansour, "Geospatial modelling of drought patterns in Oman: GIS-based and machine learning approach," *Model Earth Syst Environ*, vol. 10, no. 3, pp. 3411–3431, Jun. 2024, doi: 10.1007/s40808-024-01958-9.
- [5] A. A. Ahmed, S. Sayed, A. Abdoulhalik, S. Moutari, and L. Oyedele, "Applications of machine learning to water resources management: A review of present status and future opportunities," *J Clean Prod*, vol. 441, p. 140715, Feb. 2024, doi: 10.1016/j.jclepro.2024.140715.
- [6] C. Khaekswad *et al.*, "Potential for Drought Tolerance and Yield of a Landrace and Recommended Thai Upland Rice Cultivars," *King Mongkut's Agricultural Journal*, vol. 40, no. 2, pp. 187–195, 2022.
- [7] OFFICE OF AGRICULTURAL AFFAIRS, "Situation Report for October 2565," Jakarta, Oct. 2023.
- [8] Office of Agricultural Economics, "Agricultural products, rice." Accessed: Jul 01, 2024. [Online]. Available: <https://mis-app.oae.go.th/product/%E0%B8%82%E0%B9%89%E0%B8%B2%E0%B8%A7%E0%B8%99%E0%B8%B2%E0%B8%9B%E0%B8%B5>
- [9] K. Vikas, K. V. Goud, S. Ponaganti, A. N. Reddy, K. Nagasai, and S. N. Mohanty, "Agricultural Land Classification Based on Machine Learning Algorithms," in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, Oct. 2022, pp. 1–4. doi: 10.1109/ICCCNT54827.2022.9984254.
- [10] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *Glaci Remote Sens*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.
- [11] A. Bianglae and S. Lerk-u-suks, "A Study on Support Vector Machine Image Classification A Case Study on Rice Plantation Area in Chun District Phayao Province," *The Journal of Spatial Innovation Development*, vol. 1, no. 3, pp. 51–62, 2020.
- [12] A. Rahman *et al.*, "Performance of different machine learning algorithms on satellite image classification in rural and urban setup," *Remote Sens Appl*, vol. 20, Nov. 2020, doi: 10.1016/j.rsase.2020.100410.
- [13] H. Ouchra, A. Belangour, and A. Erraissi, "Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study," *IEEE Access*, vol. 11, pp. 71127–71142, 2023, doi: 10.1109/ACCESS.2023.3293828.
- [14] Y. Liu, D. Gui, X. Chen, Q. Liu, and F. Zeng, "Sap flow characteristics and water demand prediction of cash crop in hyper-arid areas," *Agric Water Manag*, vol. 295, Apr. 2024, doi: 10.1016/j.agwat.2024.108767.

- [15] Mae Na Rua Subdistrict Administrative Organization, “Local Development Plan 2023-2027 of the Mae Na Rua Subdistrict Administrative Organization, Mueang District, Phayao Province,” Phayao, Jun. 2024.
- [16] F. J. López-Andreu, M. Erena, J. A. Dominguez-Gómez, and J. A. López-Morales, “Sentinel-2 images and machine learning as tool for monitoring of the common agricultural policy: Calasparra rice as a case study,” *Agronomy*, vol. 11, no. 4, Apr. 2021, doi: 10.3390/agronomy11040621.
- [17] C. Pachanaparn, P. Jeefoo, and P. Rojanavasu, “Application of Remote Sensing for Drought Monitoring with NDVI-based Standardized Vegetation Index in Nan Province, Thailand,” in *7th International Conference on Digital Arts, Media and Technology, DAMT 2022 and 5th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, NCON 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 323–328. doi: 10.1109/ECTIDAMTNCON53731.2022.9720317.
- [18] P. Panjala, M. K. Gumma, and P. Teluguntla, “Machine Learning Approaches and Sentinel-2 Data in Crop Type Mapping,” in *Studies in Big Data*, vol. 96, Springer Science and Business
- [19] Media Deutschland GmbH, 2022, pp. 161–180. doi: 10.1007/978-981-16-5847-1_8.
- [20] C. Munpiensuk, S. Hengphraphrom, K. Hengphraphrom, and D. Thumsiri, “The Impact of Data Normalization to the Learning Algorithm Efficiency of Data Classifiers,” in *Joint Conference on ACTIS & NCOBA*, Nakhon Phanom, 2015, pp. 73–79.
- [21] H. Benhar, A. Idri, and J. L Fernández-Alemán, “Data preprocessing for heart disease classification: A systematic literature review,” Oct. 01, 2020, *Elsevier Ireland Ltd.* doi: 10.1016/j.cmpb.2020.105635.
- [22] N. Chinnapan, “EMPLOYEE PROMOTION PREDICTION USING MACHINE LEARNING,” Srinakharinwirot University, Bangkok, 2022.
- [23] N. Tankeaw, “A study of rubber trees cultivation water footprint in Surin Province. Case Study: Cultivation area Moo.3 and 4 Amphoe. Sang Kha,” Kasetsart university, Surin, 2019.
- Office of the Permanent Secretary for Ministry of Agriculture and Cooperatives, “Basic information of Phayao Province, January 2023,” Bangkok, 2023.

Spatio-Temporal Analysis of Road Accident on Highway 2 between Nakhon Ratchasima Intersection and Joho Intersection

Athiwat Phinyoyang

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology Isan
Nakhon Ratchasima, Thailand
athiwat.pn@rmuti.ac.th*

Kankaew Kanya

*Department of Survey Engineering,
Faculty of Engineering and
Technology, Rajamangala University of
Technology Isan
Nakhon Ratchasima, Thailand
kankaew.ka@rmuti.ac.th*

Tinn Thirakultomorn*

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology Isan
Nakhon Ratchasima, Thailand
tinn.th@rmuti.ac.th*

Wilawan Prasomsup

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology Isan
Nakhon Ratchasima, Thailand
wilawan.pa@rmuti.ac.th*

Bantita Jaemklang

*Department of Survey Engineering,
Faculty of Engineering and
Technology, Rajamangala University of
Technology Isan
Nakhon Ratchasima, Thailand
bantita.ja@rmuti.ac.th*

Buachomphu Phaensanthia

*Department of Survey Engineering,
Faculty of Engineering and
Technology, Rajamangala University of
Technology Isan
Nakhon Ratchasima, Thailand
buachomphu.ph@rmuti.ac.th*

Abstract—This study analyzes the spatio-temporal distribution of road accidents along Highway No. 2 between the Nakhon Ratchasima and Joho Intersections by integrating Geographic Information Systems (GIS) with spatial and temporal analysis techniques. This study used road accident data from 2019 to 2021 to evaluate accident frequency through mean center analysis and kernel density estimation. Additionally, spatial cluster analyses using Moran's I and Getis-Ord Gi* were employed to identify statistically significant accident clusters. Results showed that accidents were concentrated primarily at intersections and road curves. Clustering analysis revealed a random distribution pattern over the three years, while hotspot analysis at 99%, 95%, and 90% confidence levels identified significant hotspots at intersections and curves, with cold spots distributed along other road sections. The findings provide essential insights for developing preventive measures to reduce road accidents in Nakhon Ratchasima Province.

Keywords—Road Accident, Geographic Information System, Spatio-temporal Analysis, Highway No.2, Nakhon Ratchasima Province

I. INTRODUCTION

Economic growth, industrialization, and population increases have led to a rapid rise in the number of vehicles on the roads and the expansion of road networks. Consequently, this has resulted in an increase in road accidents [1]. Road accidents are a significant cause of unnecessary loss of life and property worldwide [2]. According to the World Health Organization (WHO), approximately 1.3 million people die from road accidents each year, with an additional 20 to 50 million injured or disabled. Road accidents are the leading cause of death among individuals aged 15-29, and over 50% of road traffic fatalities involve pedestrians, cyclists, and motorcyclists [3].

In Thailand, road accidents represent a major public health issue, particularly affecting the youth and working-age population, vital national resources. This issue significantly impacts the country's economy. Statistics from the Royal Thai Police indicate that, on average, 7,912 people died from traffic

accidents in Thailand between 2011 and 2021, yielding a fatality rate of 14.81 per 100,000 population. The estimated economic loss from these accidents is around 3.0% of the country's gross domestic product annually [4]. Therefore, road accidents are undeniably a critical issue for Thailand.

Nakhon Ratchasima, the largest province in Thailand, serves as a central hub for business, transportation, education, and tourism, making it a gateway to the northeastern region. This high volume of transportation activity has resulted in a significant increase in road accidents. According to the National Statistical Office, the province averaged 4,549 road accidents annually between 2011 and 2021, with an average of 391 fatalities and 647 injuries per year [4]. Without effective preventive measures and planning, the frequency and severity of road accidents are expected to rise further.

Geographic Information Systems (GIS) are powerful spatial data analysis tools widely used to analyze and visualize road accident locations [5]. GIS helps identify spatial patterns and accident hotspots [6], and assess the density and distribution of accidents [7][8]. Furthermore, GIS enables a deeper understanding of the relationship between accidents and contributing factors, such as socio-economic data, land use, and travel patterns [9].

Recently, GIS has been integrated with spatial statistical analysis techniques, notably Kernel Density Estimation (KDE), a popular method for point pattern and hotspot analysis of accidents [6][10][11]. Additionally, techniques such as Moran's I and Getis-Ord GI* have been introduced for analyzing spatial distributions [12][13], along with temporal analysis to examine the timing of accidents. While there are broader studies on road safety and accident distribution in Nakhon Ratchasima Province, few have focused specifically on high-risk, high-traffic segments like Highway 2, particularly the stretch between the Nakhon Ratchasima and Joho intersections. This section is especially significant due to its heavy transportation flows, high accident rates, and critical intersections and curves that elevate accident risks. This study employs advanced GIS techniques and spatio-temporal analysis to map accident hotspots, analyze frequency over

time, and identify statistically significant clusters along this targeted section of Highway 2. Concentrating on this segment, the research offers actionable insights into accident-prone areas, assisting authorities in developing tailored safety interventions that address the specific risks associated with this key transportation corridor. The findings can serve as foundational information for planning and implementing strategies to prevent and reduce road accidents in Nakhon Ratchasima Province.

II. STUDY AREA

Highway No. 2 (Mitraphap Road) from the Nakhon Ratchasima Intersection to the Johor Intersection (Fig. 1), covers parts of four sub-districts: Nai Mueang, Ban Kho, Mueang Wai, and Johor, within the Mueang Nakhon Ratchasima District of Nakhon Ratchasima Province. The distance is approximately 7.7 kilometers and has been selected as the study area for this research due to its significance as a major transportation route in Nakhon Ratchasima Province, characterized by heavy traffic and frequent accidents.

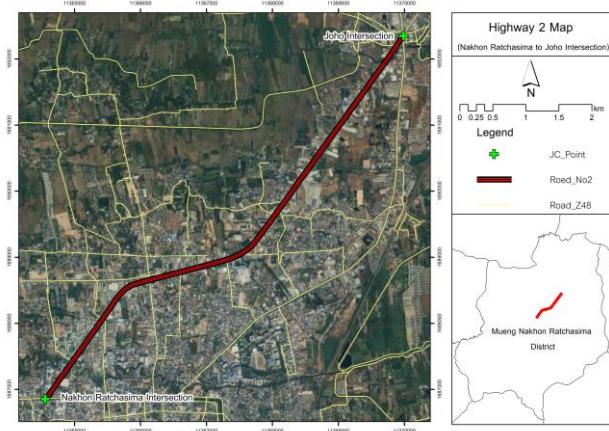


Fig. 1. Location map of the study area.

III. MATERIAL AND METHOD

The research methodology consisted of data collection and preparation and four components (Fig. 2). In this study, the causes of road accidents from the records of The Sawang Metta Dhammasathan Foundation (Sawang Metta) and the Phuttha Dhamma 31 Foundation (Huk 31) were summarized using maps, charts, and tables. To validate these accident records and identify hotspots along the study route, four types of analyses were conducted in ArcGIS: mean center analysis, Kernel Density Estimation (KDE), cluster analysis, and hotspot analysis. Details of each stage are separately described in the following sections.

A. Data Collection and Preparation

In this study, various necessary data were collected for analysis, including GIS data and secondary data, as follows:

Road Network Data: The road network data from 2013, sourced from the Department of Environmental Quality Promotion, was updated using satellite imagery from Google Earth to ensure the dataset's accuracy and currency.

Accident Statistics (2019–2021): Road accident statistics from 2019 to 2021 were provided by the Sawang Metta Dhamma Sathan Foundation, and the Buddha Dhamma 31 Foundation provided road accident statistics from 2019 to 2021. This data was filtered to include only accidents

occurring within the study area. The exact coordinates of accident locations were determined using GPS receivers during field surveys and corroborated with high-resolution satellite imagery from Google Earth. The data was then converted into a spatial format using GIS software for further analysis.

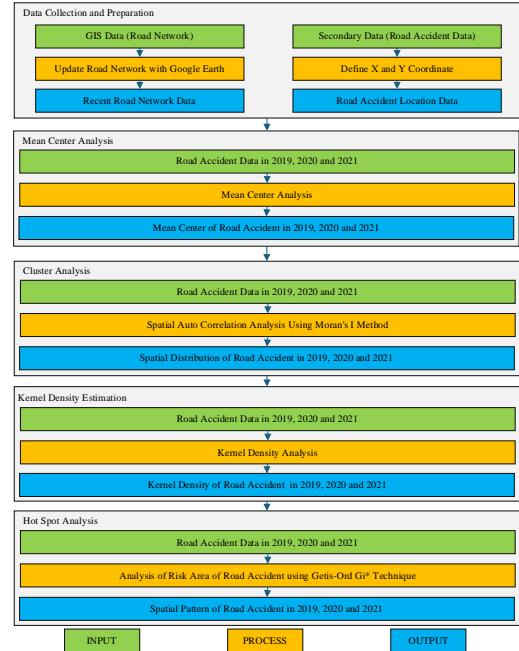


Fig. 2. Overview framework of research procedures.

B. Mean Center Analysis

This method involved measuring the geographic mean of accident locations along the highway network, using accident frequency at each site as a weight. The weighted mean center algorithm adjusts the geographic center or frequency value toward locations with higher accident frequencies. The output of this computation helps analysts identify areas where accidents are more concentrated within the study area. In this study, the road accident location data from the years 2019, 2020, and 2021 were analyzed to determine the mean center by using the average coordinates on the x and y of all the road accident data within the study area. The formulas for calculating the mean center and weighted mean center are given by [14]:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (1)$$

where x_i and y_i are the coordinates of the feature i , and n is the total number of features.

C. Cluster Analysis

The location of road accident data from 2019, 2020, and 2021 were analyzed using spatial autocorrelation, specifically Moran's I method. The spatial autocorrelation (Moran's I) algorithm simultaneously measures both features' locations and values and returns the pattern expressed by the data regarding whether they are clustered, dispersed, or random. Moran's I is an inferential statistical method, which means the analysis results are interpreted within the null hypothesis. This analysis was done for the individual years to observe whether there were changes in cluster intensity. Accident locations with a very low or very high Z-score fall outside the normal distribution and indicate a statistically significant area for analysis. The Moran's I statistic, I , is given by [15][16]:

$$I = \left(\frac{n}{W} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

where N is the total number of observations (locations), w_{ij} are spatial weights between locations i and j , x_i is the value of the variable at the location i , \bar{x} is the mean value of the variable across all locations, W is the total sum of the spatial weights.

D. Kernel Density Estimation

This study used Kernel Density Estimation (KDE) to create a heat map of road accident variability, ranging from high to low values. This method evaluates the proportion of road accidents expected to occur at any given location within a specified area. Road accident location data from 2019, 2020, and 2021 were used as input for the KDE analysis. Additionally, a 50-meter radius for each point was defined, and the density was calculated every 150 meters (Search Radius = 150). A 50-meter radius was used for the KDE analysis to provide a focused assessment of accident density along Highway 2, capturing localized clusters while avoiding excessive smoothing that might obscure significant patterns. This radius aligns with the scale of the highway and typical accident-prone zones, allowing the analysis to highlight high-density areas like intersections and curves without merging distinct hotspots. The result is a map showing the density and distribution of road accidents at each point, along with minimum, maximum, and average density values. The Kernel Density Estimation can be calculated using the following Equation [17]:

$$f(x, y) = \frac{1}{nh^2} \sum_{i=1}^n k\left(\frac{d_i}{h}\right) \quad (3)$$

where f is the density estimate at the location (x, y) , h is the search radius (bandwidth or kernel size), n is a number of observations (total number of accidents), k is the kernel function, and d_i is the distance between the location (event point) (x, y) and the location of the i th observation.

The search radius, R , is given by

$$R = 0.9 \min \left(SD, \sqrt{\frac{1}{ln(2)} D_m} \right) * n^{-0.2} \quad (4)$$

where SD is the standard distance, D_m is the median distance, and n is the number of points (if no population field is used) or the sum of the population field values (if a population field is supplied)

E. Hot Spot Analysis

The hotspot analysis method employs road accident location data from 2019, 2020, and 2021 to identify high-risk areas for road accidents using the Getis-Ord Gi* technique. This approach categorizes the data into two primary groups: high-accident occurrence locations (Hot Spots) and low-accident occurrence locations (Cold Spots). The calculation for this analysis is performed using Equation 5, which quantifies the spatial concentration of accidents. By applying the Getis-Ord Gi* method, this analysis effectively highlights areas with significant concentrations of road accidents, enabling researchers and policymakers to focus on specific locations that require targeted interventions. Identifying these hotspots is crucial for improving road safety, as it allows authorities to allocate resources efficiently and implement tailored safety measures in critical areas [18][19].

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} w_j - \bar{x} \sum_{j=1}^n w_{ij}}{s \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2}{n-1}}} \quad (5)$$

where

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n} \quad (6)$$

$$s = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{x})^2} \quad (7)$$

where x_j is the attribute value for feature j , w_{ij} , j is the spatial weight between feature i and j , and n is equal to the total number of features.

F. Analysis of Road Accident Occurrence Times

The study of the relationship between road accident occurrence times utilized road traffic accident data and hotspot points at confidence levels of 90%, 95%, and 99% from hotspot analysis. This was done to analyze trends in the peak and low times of road accidents and identify the months with the highest and lowest accident occurrences. The analysis was based on overlapping data within a 24-hour period, starting from 00h – 01h to 23h – 00h, and examining the repetition of data across all 12 months, from January to December, for the years 2019, 2020, and 2021.

IV. RESULT AND DISCUSSION

In this study, GIS tools, including Mean Center analysis, Kernel Density Estimation (KDE), Cluster Analysis, and Hotspot Analysis, were utilized to illustrate the spatial characteristics of accident locations. The location of road accident data (2019–2021) was used for the Mean Center and Kernel Density Estimation analyses, while both the location of road accident data and annual accident location data were employed for the Cluster and Hotspot Analyses. The results of the data analysis are detailed as follows:

A. Spatial Distribution of Mean Center

The geographic mean center of the cumulative road accident frequency (2019–2021) and for each year is shown in Fig. 3. Meanwhile, the results of the mean center analysis of cumulative road accidents based on time periods, divided into five-time ranges—late night (00:00–06:00), morning (06:00–12:00), afternoon (12:00–16:00), evening (16:00–19:00), and nighttime (19:00–24:00)—are illustrated in Fig. 4. The study found that the cumulative accidents (2019–2021) were concentrated around the Pradok Intersection and the curve opposite Makro Nakhon Ratchasima. In 2019 and 2021, accident clusters were located around the Pradok Intersection, while in 2020, the cluster was found at the curve opposite Makro Nakhon Ratchasima. Additionally, the mean centers of accidents across all five time periods during 2019–2021 were situated around the Pradok Intersection and the curve opposite Makro Nakhon Ratchasima.

B. Spatial Distribution of Cluster Analysis

The results of the spatial autocorrelation statistic (Moran's I) for the accident data from 2019 to 2021 are shown in Fig. 5. It was found that the distribution pattern of accidents over the past three years was random, with Moran's I value of 0.014, -0.017, and -0.120, and z-scores of 0.2961, -0.1275, and -1.2901 for 2019, 2020, and 2021, respectively (as shown in Table I.). This indicates that, with a confidence level of 90% and a probability of 10%, the clustering of the accident data may be due to random processes. This suggests that the accident patterns over the three-year period (2019–2021) do not significantly differ from a random distribution. In other words, there is no significant clustering of accident data, and the overall distribution is random during these three years.



Fig. 3. Geographic mean center of road accident from 2019 and 2021.

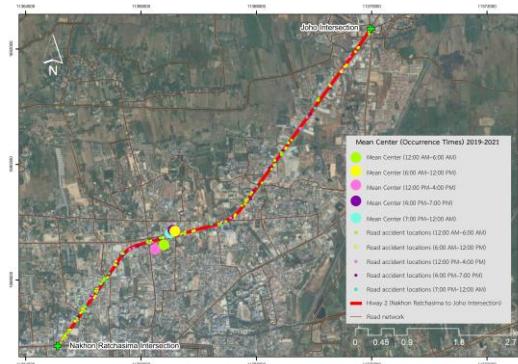


Fig. 4. Geographic mean center of road accident in 5 different times.

In addition, The results of the spatial correlation statistics (Moran's I) for accident occurrences across five time periods—late-night (00:00-06:00), morning (06:00-12:00), afternoon (12:00-16:00), evening (16:00-19:00), and night (19:00-24:00)—from the years 2019 to 2021 are shown in Fig. 6. It was found that the occurrences of accidents during all time periods over the three years displayed a random distribution, with Moran's I values presented in Table II. The study results indicate a pattern of clustering in the accident data at a 90% confidence level, with a 10% probability that the clustering may be due to random processes. This suggests that the pattern of morning accidents from 2019 to 2021 does not significantly differ from a random distribution, indicating that there is no clustering of accident data overall during these three years, and the distribution is random.

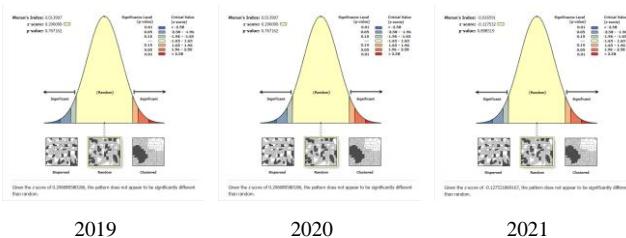


Fig. 5. Spatial correlation statistics for road accident occurrences from the years 2019 to 2021.

TABLE I. MORAN'S INDEX, Z-SCORE, AND P-VALUE FOR ROAD ACCIDENTS FROM THE YEARS 2019 TO 2021.

Year	Moran's Index	z-score	p-value
2019	0.0139	0.2961	0.7672
2020	-0.0166	-0.1275	0.8985
2021	-0.1204	-1.2901	0.1970

TABLE II. MORAN'S INDEX, Z-SCORE, AND P-VALUE FOR ROAD OF 5 OCCURRENCE TIMES FROM THE YEARS 2019 TO 2021.

Year	Moran's Index	z-score	p-value
Late-night time (00:00-06:00)			
2019	0.1244	0.4035	0.6866
2020	0.2315	1.2055	0.2280
2021	0.1312	1.2779	0.2013
Morning time (06:00-12:00)			
2019	-0.0004	0.1127	0.9103
2020	-0.0236	0.0595	0.9525
2021	-0.0378	-0.0022	0.9982
Afternoon time (12:00-16:00)			
2019	-0.0146	0.1081	0.9139
2020	-0.1081	-0.1603	0.8726
2021	-0.2541	-0.7801	0.4353
Evening time (16:00-19:00)			
2019	0.1876	0.4430	0.6578
2020	-0.2023	-0.4837	0.6286
2021	-0.4170	-1.1941	0.2324
Night time (19:00-24:00)			
2019	-0.2103	-0.6277	0.5302
2020	0.0491	0.3730	0.7092
2021	-0.1844	-0.5217	0.6019

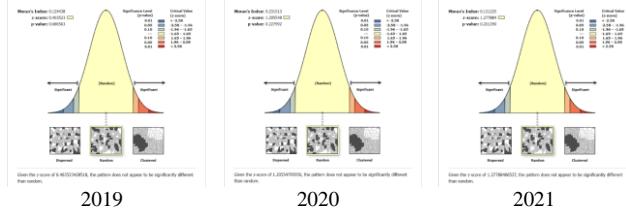


Fig. 6. Spatial correlation statistics for road accident of 5 occurrence times from the years 2019 to 2021.

C. Spatial Distribution of Kernel Density Estimation

The kernel density estimation (KDE) data for the cumulative accident frequency along the study area's routes includes the annual accident density surfaces and the KDE surfaces for accidents categorized by time periods (late night, morning, afternoon, evening, and nighttime), as shown in Fig. 7. These density surfaces depict the spatial variation in road accident frequency, ranging from high to low. However, it is important to note that these KDE maps are subjective, as they do not indicate the statistical significance of accident frequency at various locations throughout the study area. In other words, the variations may result from random processes and are not necessarily linked to the causes of the accidents. Nonetheless, when the KDE data is analyzed in relation to the road network, it suggests that areas with higher overall accident frequency are associated with road intersections (such as the Pradok Intersection) and curves near shopping malls (Makro and Central Plaza).

D. Spatial Distribution of Hot Spot Analysis

The results of the hotspot analysis using the Getis-Ord Gi* statistic, with parameters defined based on the characteristics of the input data—road accident data from 2019 to 2021—are shown in Fig. 8. The analysis reveals that hotspots with a 99% confidence level are located around the Pradok Intersection. Hotspots with 95% and 90% confidence levels are dispersed across various intersections and curves along the roads within the study area. In contrast, cold spots for cumulative accident data over the three-year study period are scattered throughout the roads in the study area.

The results of the hotspot analysis using the Getis-Ord Gi* statistic for road accident data categorized by 5 time periods

(late night, morning, afternoon, evening, and night) during 2019–2021 (as shown in Fig. 12) reveal the following:

- Late Night: No hotspots were found at the 99% or 95% confidence levels, with only 90% confidence level hotspots detected. These were scattered mainly at intersections and road curves from the Nakhon Ratchasima Intersection to the Pradok Intersection. Cold spots were rare during this period.

- Morning: Similar to the late-night period, no hotspots were found at the 99% confidence level, but hotspots at the 95% and 90% confidence levels were present, mainly at intersections and curves along the roads in the study area. Cold spots were widely distributed across the roads.

- Afternoon: Hotspots were detected at the 99% confidence level, but no hotspots were found at the 95% or 90% confidence levels. Cold spots were widely spread along the roads in the study area.

- Evening: No hotspots were found at the 99% confidence level, but hotspots were identified at the 95% and 90% confidence levels, mainly at intersections and curves throughout the study area. Cold spots were also dispersed along the roads.

- Nighttime: Only 90% confidence level hotspots were found, primarily around the Nakhon Ratchasima Intersection. Cold spots during this period were scattered along the roads, similar to other time periods.

This analysis indicates that accident hotspots and cold spots vary depending on the time of day, with intersections and curves being the most common areas for higher accident frequencies.

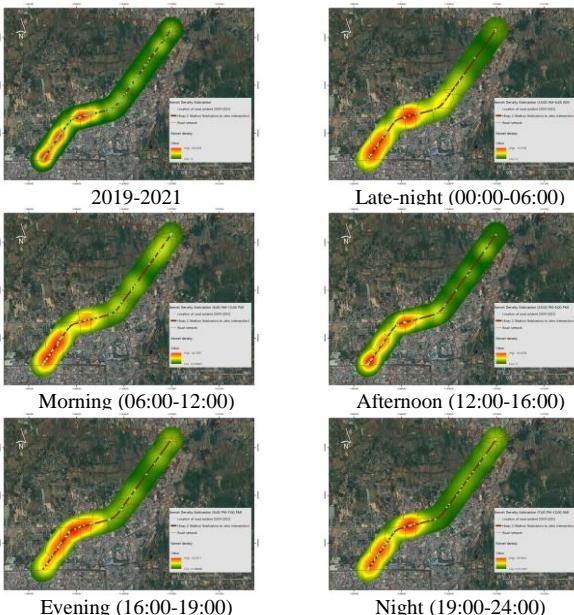


Fig. 7. Spatial distribution of kernel density of road accidents in the years of 2019-2021.

The analysis in this study revealed that intersections and curves are generally high-risk areas for road accidents in Thailand, consistent with findings from [1] and [21]. This research on Highway 2 confirms that these geometric features significantly contribute to accident hotspots, particularly near critical intersections like Pradok and Joho. By utilizing Kernel Density Estimation (KDE) to identify these hotspots and further verify them with Getis-Ord Gi*, this study enhances

the precision of previous findings, highlighting specific spatial patterns across multiple years and times of day.

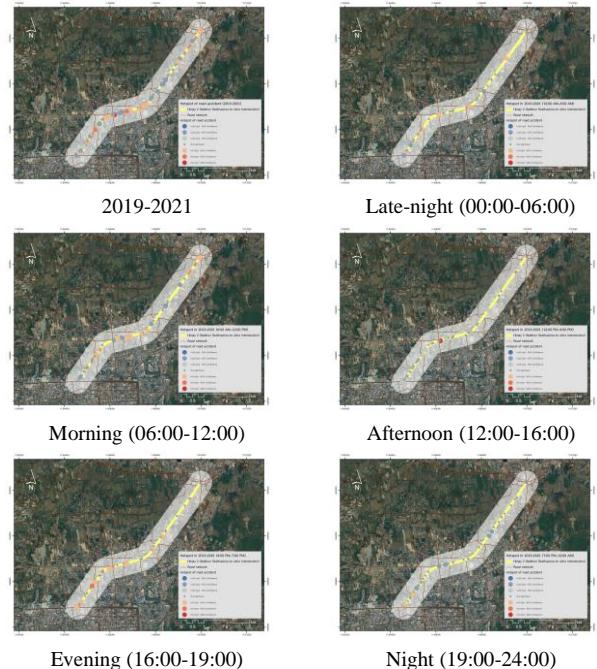


Fig. 8. Spatial distribution of hot spots of road accidents in the years of 2019-2021.

Internationally, GIS-based analyses, including those from Romania [6] and Nigeria [20], have proven the utility of KDE and Moran's I in identifying accident hotspots. This study not only applies these techniques to locate clusters but also examines temporal trends, such as peak accident hours, to guide safety strategies. The recurrence of consistent patterns across various regions underscores the universal relevance of these methods and the critical need for tailored, localized interventions.

E. Road Accident Occurrence Times

The analysis of hourly accident occurrences can be summarized as follows: the time period from 15:00 to 17:00 had the highest frequency of road accidents in 2019. In 2020 and 2021, however, the highest frequencies were observed between 08:00 to 09:00 and 11:00 to 12:00, respectively, as shown in Fig. 9.

When examining the hourly accident occurrences across the three seasons, it was found that during the summer and rainy seasons, 08:00 to 09:00 and 16:00 to 17:00 had the highest frequency of road accidents. In contrast, during the winter, the highest frequency occurred between 00:00 to 01:00 and 14:00 to 15:00 (Fig. 10).

Additionally, the analysis of hourly accident occurrences across all five time periods indicates that from 00:00 to 01:00, there was the highest frequency of road accidents during the late-night hours (00:00-06:00). This is likely due to the closing hours of entertainment venues. The time period from 08:00 to 09:00 recorded the highest frequency during the morning (06:00-12:00), which may correspond to commuting times. Meanwhile, the highest frequency during the afternoon (12:00-16:00) was found between 14:00 to 15:00. For the evening period (16:00-19:00), the highest frequency occurred from 16:00 to 17:00. Lastly, the time period from 19:00 to 20:00 showed the highest frequency of road accidents during

the night (19:00-24:00), likely due to post-work travel or shift changes, which can increase the risk of traffic accidents, as illustrated in Fig. 11.

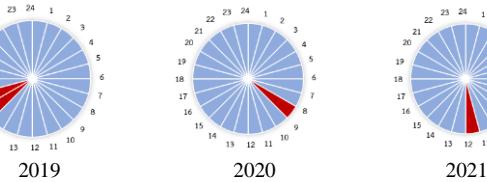


Fig. 9. Hourly statistics on road accident occurrences from the years 2019 to 2021.

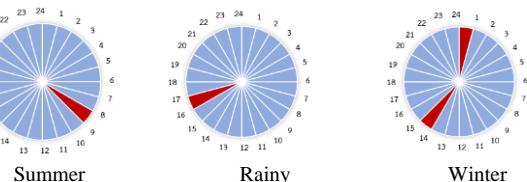


Fig. 10. Road accident statistics based on the season from the years 2019 to 2021.

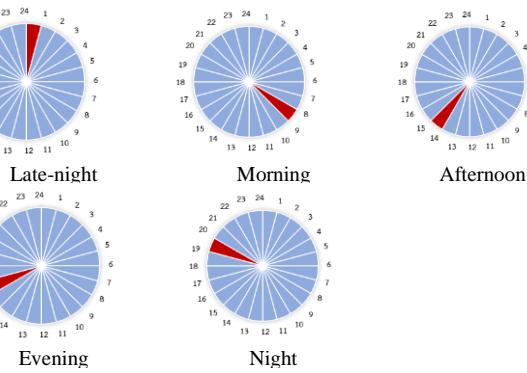


Fig. 11. Road accident statistics based on five time periods from the years 2019 to 2021.

V. CONCLUSIONS

The analysis of mean centers and kernel density for accident frequency (2019–2021) identified intersections and curves, such as the Nakhon Ratchasima intersection and the curve near the Makro supercenter, as primary hotspots. These areas experience high traffic density and complex routing, consistent with [20]'s findings that accidents often concentrate at intersections and curves. Kernel Density Estimation (KDE) further highlighted high-risk zones linked to traffic density and inadequate road design, supporting similar results from [21] in Chonburi Province.

Moran's I analysis revealed no significant clustering of accidents, with values ranging from 0.014 to -0.120 and z-scores between 0.2961 and -1.2901, indicating a random distribution over the three years. In contrast, Getis-Ord Gi* hotspot analysis identified significant hotspots at intersections and curves, with 99%, 95%, and 90% confidence levels, while cold spots were found in lower-traffic areas. Temporal patterns showed peak accident occurrences during morning and evening rush hours, correlating with increased vehicle movement.

These findings align with [22]'s study in Pattani municipality, reinforcing the need for targeted interventions to enhance road safety. By addressing high-risk areas through improved road design, optimized traffic management, and tailored safety measures, these insights can significantly

reduce accident risks and promote safer transportation systems.

REFERENCES

- [1] N. Suesat, The development of hazardous route map in Thailand. M.Eng. Thesis, Suranaree University of Technology, 2015.
- [2] World Bank, Traffic Crash Injuries and Disabilities: The Burden on Indian Society. The World Bank, Washington, DC, 2021.
- [3] World Health Organization. "Road Traffic injuries." 2023. [Online]. Available: https://www.who.int/health-topics/road-safety#tab=tab_1.
- [4] Royal Thai Police, "Accident Information Center for Strengthening Road Safety Culture." August 2023. [Online]. Available: <https://www.thairsc.com/>.
- [5] J. K. Deepthi and B. Ganeshkumar, "Identification of accident hot spots: a GIS based implementation for Kannur district, Kerala," International Journal of Geomatics and Geoscience, Vol. 1(1), pp. 51-59, 2010.
- [6] J. Benedek, S. M. Ciobanub and T. C. Man, "Hotspotsn and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania)." Accident Analysis and Prevention, Vol. 87, pp. 117-126, 2016.
- [7] A. S. Mohaymany, M. Shahri, and B. Mirbagheri, "GIS-based method for detecting high-crash-risk road segments using network kernel density estimation,". Geo-Spatial Information Science, Vol. 16(2), pp. 113-119, 2013.
- [8] R. Satria and M. Castro, "GIS tools for analyzing accidents and road design: a review". Transportation Research Procedia, Vol. 18, pp. 242-247, 2016.
- [9] L. Li, L. Zhu, and Z. D. Sui, "A GIS-Based Bayesian Approach for Analyzing Spatial-Temporal Patterns of Intra-City Motor Vehicle Crashes," Journal of Transport Geography, Vol. 15, pp. 274-285, 2007.
- [10] M. Bil, R. Andrášik, and Z. Janoska, "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation," Accident Analysis and Prevention, Vol. 55, pp. 265-273, 2013.
- [11] S. Hashimoto, S. Yoshiki, R. Saeki, Y. Mimura, R. Ando, and S. Nanba, "Development and application of traffic accident density estimation models using kernel density estimation,". Journal of Traffic and Transportation Engineering, Vol. 3(3), pp. 262-270, 2016.
- [12] J. K. Ord, and A. Getis, "Local spatial autocorrelation statistics: Distributional Issues and an Application,". Geographical Analysis, Vol. 27(4), pp. 286-306, 1995.
- [13] A. Soltani, and S. Askari, "Exploring spatial autocorrelation of traffic crashes based on severity". Injury, Vol. 48, pp. 637-647, 2017.
- [14] ESRI, "Mean Center (Spatial Statistics)" 2024. [Online]. Available: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-mean-center-spatial-statistics-works.htm>.
- [15] T. C. Bailey, and A.C. Gatrell, "Interactive spatial data analysis," Addison Wesley Longman Limited, Harlow, England, 1995.
- [16] D.A. Griffith, "Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization," Springer, Berlin, Germany, 2003.
- [17] J. K. Deepthi, and B. Ganeshkumar, "Identification of accident hot spots: A GIS-based implementation for Kannur District, Kerala," Int. J. Geomat. Geosci, Vol. 1, pp. 51–59, 2010.
- [18] J. P. DeGroote, R. Sugumaran, S. M. Brend, B. J. Tucker, and L. C. Bartholomay, "Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA," Inter J Health Geog Vol. 7(19), 2008.
- [19] X. Zhixiao, and Y. Jun, "Kernel density estimation of traffic accidents in a network space," Geography/Geology Faculty Publications.
- [20] A. Afolayan, S. M. Easa, O. S. Abiola, F. M. Alayaki, O. Foluronso, "GIS-Based Spatial Analysis of Accident Hotspots: A Nigerian Case Study," Infrastructures, Vol. 7, 103, 2022.
- [21] J. Waiwatthana, N. Pleerux, K. Nualchawee, and N. Intarawichian, "Spatio-temporal analysis of road accident in chon buri province," Phranakorn Rajabhat Research Journal (Science and Technology), Vol.14(2), pp. 31-42, 2019.
- [22] S. Arthan, "The study of road accident risk areas in Pattani Municipality, Pattani Province," Department of Geography, Faculty of Humanities and Social Sciences, Prince of Songkla University, 2019.

Geoinformatics-based Flood Extent Mapping and Analysis Using Sentinel-1 Imagery: A Case Study of Phayao Province

Pannatorn Woottipriyatorn
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 67025156@up.ac.th

Thidapath Anucharn
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 thidapath.an@up.ac.th

Niti Iamchuen
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 niti.ia@up.ac.th

Thitisorn Sriprom
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 67025213@up.ac.th

Nakarin Chaikaew
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 nakarin.ch@up.ac.th

Phongsakorn Hongpradit
School of Information and Communication Technology
University of Phayao
 Phayao, Thailand
 67025224@up.ac.th

Abstract— Climate change will worsen global flooding. This study used geoinformatics and Sentinel-1 satellite data to map and analyze flood extent in Phayao Province, Thailand. The research defines the area of interest, preprocessing satellite imagery, identifying flood extent, and validating results. Phayao Province was inundated for 19,295.87 hectares on August 25, 2024. Further analysis showed that the flood extent changed over the study period, down to 10,620.52 hectares on September 6 and up to 11,990.81 on September 18. Dok Kham Tai, Chiang Kham, and Chun were most affected by the flooding. The researchers identified places with repeated flooding episodes using satellite imagery from multiple time, estimating the inundated area at 390,964.16 hectares. The results showed that geoinformatics and remote sensing can quickly and accurately flood extent mapping for evidence-based disaster management and mitigation. This study can be applied in similar flood areas to enhance awareness of flooding patterns and design flood risk management measures.

Keywords— *Flood mapping, geoinformatics, Phayao, sentinel-1*

I. INTRODUCTION

Floods are among the most terrible natural disasters globally, impacting millions of lives and causing significant economic losses annually [1]. The Intergovernmental Panel on Climate Change defines floods as "the overflowing of the normal confines of a stream or other body of water, or the accumulation of water over areas that are not normally submerged." There are various types of floods, including river floods, flash floods, urban inundations, and coastal flooding. Flash floods, totaling over 85% of flooding cases worldwide, have the highest mortality rate. The increasing frequency of heavy rainfall, changes in upstream land use, and a rising concentration of population and assets in flood-prone areas contribute to the rising flood risks [2]. These factors, combined with insufficient flood planning and management strategies, worsen the impact of flooding on human survival and economic development [3].

Thailand, particularly the northern region including Phayao Province, has significant flooding issues. The severity of inundation fluctuates annually, influenced by factors such as rainfall patterns, geographical features, and climate change [4]. In Phayao, two primary types of flooding occur: overflow of main rivers like the Yom and Ing due to heavy rainfall

during severe storms, and flash floods. These events impact both urban and agricultural areas, affecting property and human lives. The Phayao Provincial Disaster Prevention and Mitigation Office reports that main rivers are impacted to inadequate water monitoring systems, limiting flood notifications to specific areas [5]. This lack of comprehensive monitoring and early warning systems poses a significant challenge in flood management [6] and mitigation efforts.

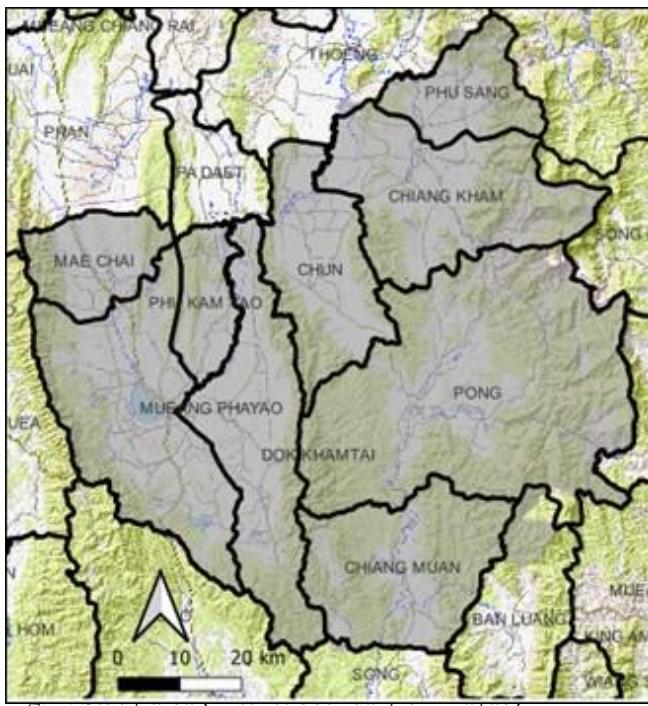
To address these challenges, integrating geoinformatics technologies offers promising solutions. Geographic Information Systems (GIS) can be utilized for various functions, including water management planning, land use assessment, and flood extent mapping. Satellite imagery, particularly from Sentinel-1, provides valuable data for flood extent mapping, situational assessment, and continuous monitoring [7] [8]. The study "Geoinformatics-based Flood Extent Mapping and Analysis Using Sentinel-1 Imagery: A Case Study of Phayao Province" aims to investigate and analyze flooding in Phayao using these advanced techniques. By utilizing geoinformatics, the research seeks to improve flood disaster assessment, enhance rescue operations, and optimize resource allocation. This approach has the potential to lead to more effective flood mitigation strategies, improved early warning systems, and better overall flood management in the region [9].

II. METHODOLOGY

A. Study Area

This research area was chosen in Phayao Province, Northern Thailand, where mountain ranges cover the eastern, western, southern, and central sections of the province. The mountain ranges extend from north to south. Agricultural plains are situated between the rivers. The province is characterized by high terrain to varied extents. Plains and hills constitute around 35% of the region, whereas lowlands account for the remaining 18%. The Phi Pan Nam Mountain ranges, regarded as natural barriers, encompass the province. Doi Phu Langka, Doi San Pan Nam, and Doi Mae Suk are prominent mountain ranges that serve as the sources of various streams. The Ing River, Lao River, and Yom River are

the major rivers [10], that the study area was confined, to the region shown in Figure 1.



B. Data Collection

This research utilizes Sentinel-1 imagery to map and analyze flood extent in Phayao Province from July 8 to September 18, 2024. Sentinel-1 is an active radar satellite that is well-suited for flood monitoring because it can penetrate cloud cover, allowing for consistent data collection even in cloudy conditions [11]. The study has steps to survey and identify flood-affected areas using Sentinel-1 imagery, monitor changes in flood extent over the study period, and identify specific areas suffering from flooding. By using Sentinel-1 imagery [12], the research resolves limitations of optical satellites that are hindered by cloud cover, enhancing accurate and timely flood mapping and analysis in Phayao Province.

C. Data Processing

To monitor pre- and post-flood conditions, the following methodology was implemented:

1. Define the area of interest (AOI) for the study.
2. Acquire Sentinel-1 satellite imagery for the AOI.
3. Preprocess the Sentinel-1 data by using the Sentinel Application Platform (SNAP) software:

3.1 Apply orbit file to improve geometric precision and overall data quality [13].

3.2 Perform thermal noise removal to enhance the images quality and backscatter measurements [14] [15].

3.3 Apply Land-Sea Mask to distinguish between land and water areas.

3.4 Calibrate image data using logarithmic scaling for consistency and accuracy [16].

3.5 Reduce speckle noise using a speckle filter.

3.6 Apply terrain correction to adjust geometric distortions.

4. Conduct flood extent mapping: Perform thresholding to separate flood extent using the criterion Sigma0_VV_db < -20.

5. Export preprocessed images to geographic information system processing program for further analysis:

5.1 Use raster calculator to combine pre-flood and post-flood images, highlighting areas of change indicative of flood extent.

5.2 Identify and delineate flooded areas using flooded SAR techniques.

5.3 Convert raster data of inundated areas to vector polygons.

5.4 Calculate flood extent using raster-to-polygon conversion.

5.5 Assess accuracy using random point sampling method and Kappa coefficient.

6. Generate final flood extent map showing inundated areas derived from processed Sentinel-1 imagery (Figure 2).

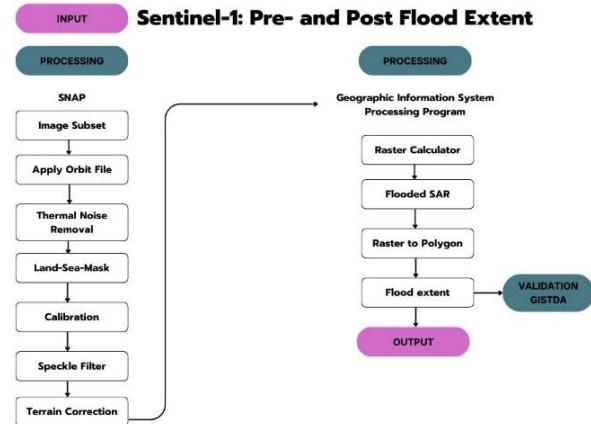


Fig. 2. Methodology workflow of the study

D. Validation

To determine the appropriate sample size for accuracy testing, this study employs Binomial Probability Theory [17], as described by Fitzpatrick-Lins. The sample size is calculated using the following equation:

$$n = \frac{Z^2(p)(q)}{E^2} \quad (1)$$

where:

n = Sample size

Z = Z-score corresponding to the desired confidence level (1.96 for 95% confidence and round-up to 2)

p = Estimated proportion of the attribute of interest

q = the perfect of 100 – p

E = Margin of error (5% in this case)

Accuracy assessment is crucial in remote sensing data processing to determine its usefulness for consumers. Effective utilization of geospatial data relies on understanding data quality. The validation process involves several key components:

Overall Accuracy: This metric assesses how well each pixel's classification aligns with actual ground conditions derived from ground truth data.

Producer's Accuracy: This measure evaluates omission errors, indicating how effectively the classification identifies actual land cover types.

User's Accuracy: This metric quantifies commission errors, showing the likelihood that a classified pixel matches the real-world land cover type.

The overall accuracy and Kappa coefficient are widely used techniques for assessing classification accuracy [18]. The Kappa coefficient is calculated using the following equation:

$$K = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n + i + \dots}{n^2 - \sum_{i=1}^k n + i + \dots} \quad (2)$$

where:

K = Kappa coefficient

n = Total number of observations

k = Number of categories

n_{ii} = Number of observations in row i and column i of the error matrix

n_{i+} = Total for row i

n_{+i} = Total for column i

III. RESULT

This article explores the results of flood mapping in Phayao Province, employing geoinformatics techniques: The study utilized Sentinel-1 imagery to analyze flood extent between July 8, 2024, and September 18, 2024, and does not include water bodies. Sentinel-1 was selected for flood detection because of its ability to penetrate cloud cover. On Aug 25, 2024, the inundated area was 19,295.87 hectares; on Sep 6, 2024 it was 10,620.52 hectares; and on Sep 18, 2024 it reached 11,990.81 hectares as shown in Table I.

Table I. Flood extent sort by date

Date	Flood extent (ha)	Average (%)
25 Aug 2024	19,295.87	46.04
6 Sep 2024	10,620.52	25.34
18 Sep 2024	11,990.81	28.61

The flood extent does not include water bodies with the most significant impact on individuals categorized by the date of incident:

August 25, 2024, affecting an area of 19,295.87 hectares. Additionally, it was determined that the districts that were most significantly impacted during this period were Dok Kham Tai (6,208.51), Chiang Kham (3,953.72), and Chun (3,101.66) as illustrated in Figure 3.

September 6, 2024, covers 10,620.52 hectares. Furthermore, it was found that the districts that were most seriously impacted during this period were Dok Kham Tai

(3,797.56 ha), Chiang Kham (3,376.81 ha), and Phu Kam Yao (1,827.42 ha), as illustrated in Figure 4.

September 18, 2024, with an area of 11,990.81 hectares. Additionally, it was determined that the districts that were most significantly impacted during this period were Dok Kham Tai (5,003.76 ha), Chiang Kham (2,383.67 ha), and Phu Kam Yao (2,357.79 ha) as illustrated in Figure 5.

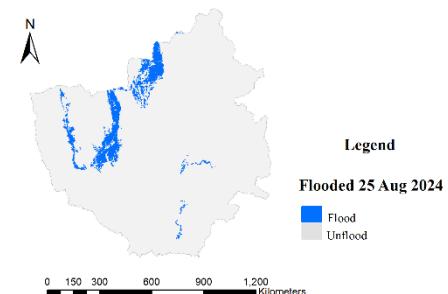


Fig. 3. Flood extent mapping of 25 Aug 2024

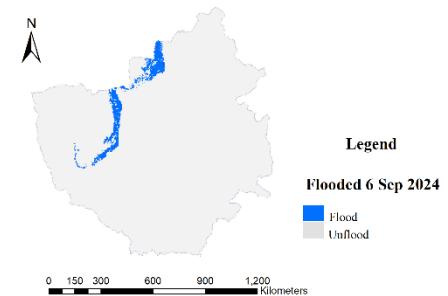


Fig. 4. Flood extent mapping of 6 Sep 2024

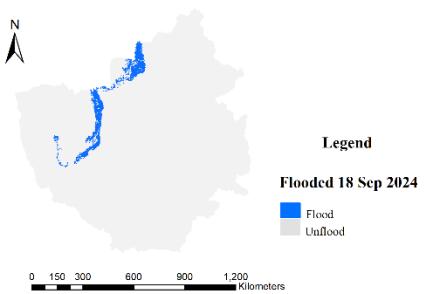


Fig. 5. Flood extent mapping of 18 Sep 2024

By combining satellite imagery from three various time periods, which provide data on flood extent, we can pinpoint locations that have experienced flooding multiple times. The total inundated area is estimated to be approximately 390,964.16 hectares, as shown in Figures 6 and 7.

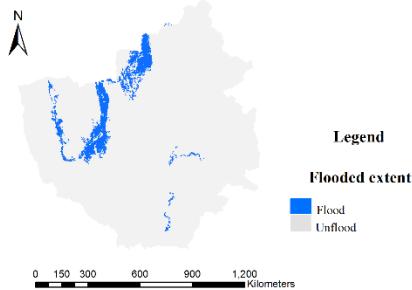


Fig. 6. Flood combine mapping

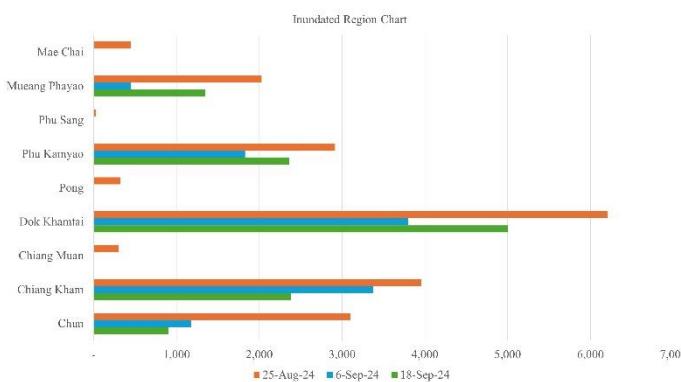


Fig. 7. Inundated region chart

An accuracy assessment for flood mapping using four datasets: July 8, 2024 (pre-flood), August 25, September 6, and September 18, 2024. These dates served for training data, while testing data from GISTDA, acquired on August 23, 2024, was used for validation. The accuracy assessment employed binomial probability theory with simple random sampling to ensure adequate representation across various flood conditions. The expected accuracy was set at 85%, with an acceptable error margin of 5%. Based on these totals of 203 sample points were identified as necessary for the accuracy assessment.

Table 2 details the flood mapping accuracy assessment for August 25, 2024, highlighting its high reliability. The classification achieves an overall accuracy of 84% with a Kappa coefficient of 0.64, indicating match with ground truth data. For flooded areas, the user's accuracy is exceptionally high at 89%, while the producer's accuracy is 87%. In contrast, non-flooded areas show a slightly higher producer's accuracy of 78% but a higher user's accuracy of 75%, suggesting some commission errors. Despite these errors, the classification effectively delineates flood extent, especially in identifying flooded areas.

Table 3 details the flood mapping accuracy assessment for September 6, 2024, highlighting its slightly high reliability. The classification achieves an overall accuracy of 79% with a Kappa coefficient of 0.57, indicating match with ground truth data. For flooded areas, the user's accuracy is exceptionally at 78%, while the producer's accuracy is 83%. In contrast, non-flooded areas show a slightly lower producer's accuracy of 74% but a lower user's accuracy of 80%, suggesting some commission errors. Despite these

errors, the classification effectively delineates flood extent, especially in identifying flooded areas.

Table 4 details the flood mapping accuracy assessment for September 18, 2024, highlighting its high reliability. The classification achieves an overall accuracy of 80% with a Kappa coefficient of 0.59, indicating match with ground truth data. For flooded areas, the user's accuracy is exceptionally high at 88%, while the producer's accuracy is 80%. In contrast, non-flooded areas show a higher producer's accuracy of 81% and higher user's accuracy of 70%, suggesting some commission errors. Despite these errors, the classification effectively delineates flood extent, especially in identifying flooded areas.

Table II. Accuracy assessment: 25 Aug 2024

Classify 25 Aug 2024	GISTDA			User's Accuracy
	Flood	Unflood	Total	
Flood	117	15	132	0.89
Unflood	18	53	71	0.75
Total	135	68	203	
Producer's Accuracy	0.87	0.78		
Overall Accuracy	0.84			
Kappa Coefficient	0.64			

Table III. Accuracy assessment: 6 Sep 2024

Classify 6 Sep 2024	GISTDA			User's Accuracy
	Flood	Unflood	Total	
Flood	89	25	114	0.78
Unflood	18	71	89	0.80
Total	107	96	203	
Producer's Accuracy	0.83	0.74		
Overall Accuracy	0.79			
Kappa Coefficient	0.57			

Table IV. Accuracy assessment: 18 Sep 2024

Classify 18 Sep 2024	GISTDA			User's Accuracy
	Flood	Unflood	Total	
Flood	102	14	116	0.88
Unflood	26	61	87	0.70
Total	128	75	203	
Producer's Accuracy	0.80	0.81		
Overall Accuracy	0.80			
Kappa Coefficient	0.59			

Table V presents the spatial rainfall patterns in Phayao province from 2022 to 2024. The data reveals significant temporal variations across the three-year period. The most

notable observation is the exceptional peak in August 2024, reaching 480.1 mm, which stands as the highest recorded rainfall during the study period. This was followed by sustained high rainfall in September 2024 at 375.6 mm.

The year 2022 exhibited two distinct rainfall peaks, with April showing 250.6 mm and September recording the year's highest at 368 mm. In contrast, 2023 displayed more moderate rainfall patterns throughout the year, with peak measurements occurring in September-October (195-256.6 mm). The 2024 data shows the most dramatic variation, characterized by relatively low rainfall in the early months, followed by the unprecedented August peak and high September rainfall.

Consistently across all three years, the period from November through March shows minimal rainfall, typically below 100 mm, indicating a clear dry season. The data demonstrates a recurring pattern where the rainy season primarily occurs from July to October, with occasional elevated rainfall events in April-May. This seasonal pattern is particularly evident in the graphical representation, where the lines for all three years follow similar trajectories despite varying intensities.

The exceptional rainfall events in late 2024, being significantly higher than the previous two years, suggest unusual precipitation patterns that likely contributed to flooding conditions in the province. This comprehensive dataset provides valuable insights into the temporal distribution of rainfall and its potential implications for flood risk management in Phayao province.

Table V. Spatial Rainfall in Phayao Province (2022-2024)

Month	Spatial Rainfall (mm)		
	2022	2023	2024
Jan	29.5	0.00	1.2
Feb	26.6	57.6	0.00
Mar	87.3	1.4	15.9
Apr	250.6	30.2	32
May	222	96.2	144.6
Jun	69.4	0.00	89.7
Jul	165.6	80.9	147.2
Aug	154.4	159.9	480.1
Sep	368	195	375.6
Oct	196.1	256.6	153.9
Nov	11	20.4	16.9
Dec	16.8	25.3	N/A

Source: Hydro-Informatics Institute, 2024

Note: N/A indicates data not available.

Figure 8 shows the spatial rainfall patterns in Phayao province from 2022 to 2024, represented by three distinct lines: blue (2022), orange (2023), and green (2024). The graph reveals significant temporal variations, with the most striking feature being the sharp peak in August 2024 at 480.1 mm.

The rainfall patterns demonstrate clear seasonal characteristics, with minimal precipitation during November-March and higher rainfall during July-October. The year 2024 exhibits the most dramatic variation, particularly with its unprecedented August peak, while 2023 shows more moderate

fluctuations throughout the year. The graph effectively illustrates how the exceptional rainfall in late 2024 likely contributed to flooding conditions in the province.

Spatial rainfall of Phayao Province: 2022-2024

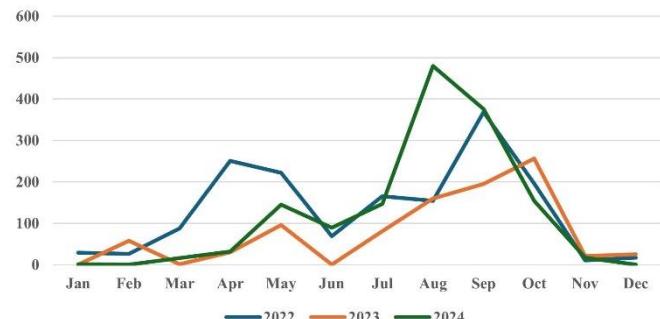


Fig. 8. Spatial rainfall of Phayao province chart

IV. CONCLUSION AND DISCUSSION

This study concludes that the flood extent mapping in Phayao province was calibrated and validated. Utilizing images from the Sentinel-1 used with geoinformatics data, most of the regions that are impacted by inundation are in close contact with the major river, including areas that are also susceptible to flooding due to their low altitude. Furthermore, the effects of continuous rainfall during periods of flooding issues. Whatever are the benefits of the apply geoinformatics has been demonstrated to be effective in delivering real-time and dependable access to flood extent information for disaster management planning and decision-making, as well as regional disaster mitigation, in floodplain surveys [19]. Moreover, prompt and precise flood mapping [20] is essential because of the extensive flood and the expanse of the inundated region [21].

In conclusion, geoinformatics is instrumental for flood mapping and analysis in Phayao Province [22]. The findings can facilitate the development of evidence-based disaster management and mitigation strategies, including early warning systems, emergency response funding, and long-term infrastructure and development of land uses, thereby encouraging local communities, such as evaluating and finding suitable areas for water retention. Defined region must be elevation and sufficient far from the primary river. Significantly improving the drainage of water into secondary streams [23][24]. to effectively manage flooding events. Another aspect of this study just considers flood-affected regions since it is a retrospective examination of data from Sentinel-1. It lacks to define damaged areas, particularly agricultural areas and living areas.

The study's findings have significant implications for flood management in Phayao Province:

Early Warning Systems: The ability to rapidly map flood extent can enhance early warning capabilities, allowing for more timely evacuations and emergency responses.

Resource Allocation: Identifying the most flood-prone districts can help authorities prioritize resource allocation for flood mitigation and response efforts.

Land Use Planning: The recurrent flooding patterns observed can inform long-term land use planning and zoning decisions to reduce flood risk in vulnerable areas.

Infrastructure Development: Understanding flood extent and frequency can guide the development of flood-resistant infrastructure and improve drainage systems in high-risk areas.

While this study provides valuable insights, it primarily focuses on mapping flood extent without detailed analysis of damage to specific land use types, such as agricultural or residential areas. Future research could incorporate land use data to assess the socio-economic impacts of flooding more comprehensively. Additionally, integrating this flood mapping approach with hydrological models and real-time weather data could further enhance flood prediction capabilities and support more proactive flood management strategies.

ACKNOWLEDGMENT

This study is a part of a Remote Sensing Application and Geographic Information System Application courses program in Applied Geoinformatics at the School of Information and Communication Technology, University of Phayao, Phayao, Thailand. Thank you Geo-Informatics and Space Technology Development Agency (GISTDA) and Hydro-Informatics Institute for the information regarding the research process.

REFERENCES

- [1] J. Li, J. Wang, and H. Ye, "Rapid Flood Mapping based on Remote Sensing Cloud Computing and Sentinel-1," *Journal of Physics: Conference Series*, vol. 1952, no. 2, p. 022051, Jun. 2021, doi: <https://doi.org/10.1088/1742-6596/1952/2/022051>.
- [2] "Flood" World Health Organization. Accessed: Sep. 25, 2024. [Online]. Available: https://www.who.int/europe/health-topics/floods#tab=tab_1
- [3] "Summarize of disaster statistics 2023" Department of Disaster Prevention and Mitigation, Bangkok. Accessed: Sep. 25, 2024. [Online]. Available: <https://shorturl.at/DaKoa>.
- [4] "Flood Disasters Prevention and Mitigation plan (Rainy Season 2024)" Royal Irrigation Department. <https://shorturl.at/hLShj>.
- [5] "PY flood" Center of excellence in Natural Disaster Management, Accessed: Sep. 23, 2024. [Online]. Available: <https://shorturl.at/vB1LC>
- [6] K. Kuntiyawichai, *Hydrologic system*, 1st ed. Khon Kaen: Department of Civil Engineering, Faculty of Engineering, Khon Kaen University, 2020.
- [7] "The essential of geospatial information in disaster management," GISTDA – Thailand's future living in space. Accessed: Sep. 25, 2024. [Online]. Available: <https://go.up.ac.th/AHYLTo>
- [8] G. P. Petropoulos and Tanvir Islam, *Remote sensing of hydrometeorological hazards*. Boca Raton, Fl: Crc Press, Taylor & Francis Group, 2018.
- [9] "Flood Situation," Geo-Informatics and Space Technology Development Agency (Public Organization). Accessed: Aug. 24, 2024. [Online]. Available: <https://go.up.ac.th/SCiA3s>
- [10] "Phayao Weather." Accessed: Sep. 25, 2024. [Online]. Available: <https://go.up.ac.th/8M2IKG>.
- [11] D. Amitrano *et al.*, "Sentinel-1 for Monitoring Reservoirs: A Performance Analysis," *Remote Sensing*, vol. 6, no. 11, pp. 10676–10693, Nov. 2014, doi: <https://doi.org/10.3390/rs61110676>.
- [12] D. Marino, M. Palmieri, A. Marucci, M. Soraci, A. Barone, and S. Pili, "Linking Flood Risk Mitigation and Food Security: An Analysis of Land-Use Change in the Metropolitan Area of Rome," vol. 12, no. 2, pp. 366–366, Jan. 2023, doi: <https://doi.org/10.3390/land12020366>.
- [13] M. Li, H. Luo, and R. Zhang, "Spatial-Temporal Distribution of Air Quality and the Influencing Factors in Complex Mountainous Cities," Jul. 2021, doi: <https://doi.org/10.1109/igarss47720.2021.9554824>.
- [14] P. Potin *et al.*, "Sentinel-1 mission status," Jul. 2015, doi: <https://doi.org/10.1109/igarss.2015.7326401>.
- [15] L. Mascolo, J. M. Lopez-Sanchez, and S. R. Cloude, "Thermal Noise Removal From Polarimetric Sentinel-1 Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: <https://doi.org/10.1109/lgrs.2021.3050921>.
- [16] E. Chuvieco, *Fundamentals of Satellite Remote Sensing*. CRC Press, 2020.
- [17] K. Fitzpatrick-Lins, "Comparison of sampling procedures and data analysis for a land-use and land-cover map," *Photogrammetric Engineering and Remote Sensing*, vol. 47, no. 3, pp. 343–351, Mar. 1981.
- [18] S. S. Rwanga and J. M. Ndambuki, "Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS," *International Journal of Geosciences*, vol. 08, no. 04, pp. 611–622, 2017, doi: <https://doi.org/10.4236/ijg.2017.84033>.
- [19] P. F. Hudson and H. Middelkoop, *Geomorphic Approaches to Integrated Floodplain Management of Lowland Fluvial Systems in North America and Europe*. Springer, 2015.
- [20] T. Sarmah, S. Das, A. Narendr, and B. H. Aithal, "Assessing human vulnerability to urban flood hazard using the analytic hierarchy process and geographic information system," *International Journal of Disaster Risk Reduction*, vol. 50, p. 101659, Aug. 2020, doi: [10.1016/j.ijdrr.2020.101659](https://doi.org/10.1016/j.ijdrr.2020.101659).
- [21] P. Matgen, R. Hostache, G. Schumann, L. Pfister, L. Hoffmann, and H. H. G. Savenije, "Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies," *Physics and Chemistry of the Earth Parts a/B/C*, vol. 36, no. 7–8, pp. 241–252, Dec. 2010, doi: [10.1016/j.pce.2010.12.009](https://doi.org/10.1016/j.pce.2010.12.009).
- [22] K. Nan, "Geomorphological change of the upper Yom river in Phrae province," 2022. doi: [10.58837/chula.the.2022.166](https://doi.org/10.58837/chula.the.2022.166).
- [23] K. Meeching, S. Thongkungand, and T. Suteerasak, "Application of Geographic Information System Data for Flood Susceptibility Area and Catchment Area Assessment: A Case Study the Southern Area of SuphanBuri Province," *The Journal of King Mongkut S University of Technology North Bangkok*, vol. 29, no. 3, Apr. 2019, doi: [10.14416/j.kmutnb.2019.04.001](https://doi.org/10.14416/j.kmutnb.2019.04.001).
- [24] L. J. Bracken *et al.*, "Flood risk management, an approach to managing cross-border hazards," *Natural Hazards*, vol. 82, no. S2, pp. 217–240, Apr. 2016, doi: [10.1007/s11069-016-2284-2](https://doi.org/10.1007/s11069-016-2284-2).

Characteristics of Infrasound Wave Background Noise on the Coast of Jakarta, Indonesia

1st Mario Batubara
 Research Center for Space
 National Research and
 Innovation Agency of
 Indonesia (BRIN)
 Bandung, Indonesia
 mari009@brin.go.id or
 0000-0001-8330-2124

2nd Ibnu Fathrio
 Research Center for
 Atmospheric and Climate
 National Research and
 Innovation Agency of
 Indonesia (BRIN)
 Bandung, Indonesia
 ibnu003@brin.go.id

3rd Musthofa Lathif
 Research Center for Space
 National Research and
 Innovation Agency of
 Indonesia (BRIN)
 Bandung, Indonesia
 must009@brin.go.id

4th Poki Agung Budiantoro
 Research Center for Satellite
 Technology
 National Research and
 Innovation Agency of
 Indonesia (BRIN)
 Bogor, Indonesia
 poki001@brin.go.id

5th Asep Sulaeman
 Research Center for Space
 National Research and Innovation
 Agency of Indonesia (BRIN)
 Bandung, Indonesia
 asep024@brin.go.id

6th Heru Subarkah
 Research Center for Space
 National Research and Innovation
 Agency of Indonesia (BRIN)
 Bandung, Indonesia
 heru019@brin.go.id

7th Yopie Kristiana
 Research Center for Space
 National Research and Innovation
 Agency of Indonesia (BRIN)
 Bandung, Indonesia
 yopi002@brin.go.id

Abstract— This article reviews the observation data of infrasound campaign for 8 months in the North and South West Java region. We use statistical approach in Power Spectral Density scheme to obtain a profile representative to the long-term infrasonic noise near the observation site. We see the microbaroms wave is dominant during the observation period. In addition, the wind direction at the observation site affects the observed infrasound amplitude level. Finally, our approach is useful to obtain the background noise profile in various observation targets.

Keywords—*infrasound, Indonesian Infrasound Network, microbaroms, Statistic PSD*

I. INTRODUCTION

Infrasonic waves are atmospheric longitudinal waves with a frequency band between 0.01 Hz and 20 Hz, which is below the human hearing threshold. Infrasonic signals have the advantage of propagating over long distances due to their low attenuation rate compared to audible sound, since absorption increases with the square of the frequency [1]. These characteristics have made infrasonics an alternative technology for monitoring the early occurrence of natural disasters and studying the dynamics of a planet's atmosphere. [2,3]. Infrasonic networks have been built in some countries in the world for decades. Infrasound has been routinely recorded by the Infrasonic Monitoring System in Indonesia since early 2024, and its purpose is to detect and locate shock infrasonic waves. Historically, Indonesia has experienced many natural events that emit infrasonic from volcanoes, earthquakes, severe weather events, lightning, sea waves [4]. On the one hand, the inventory of infrasonic sources needs to be understood to help indicate the characterization of each source. Infrasonic propagation can also provide information about the mechanisms that are worked on, as well as their strength at the

source location which can allow the classification of the type of source generator such as Power Spectral Density (PSD) [5,6]. They use the PSD model to describe the ambient noise spectra of the seismic source.

The ambient infrasound noise is dominated by long-range pressure fluctuations generated over the ocean as a microbarom and by short-range pressure fluctuations due to local eddies and winds. This noise at infrasound stations varies greatly over time. Location-dependent factors that contribute to the noise include climate, station location relative to the ocean and other artificial sources that interfere with the recorded infrasound signal.

Microbaromes are observed worldwide with frequencies of 0.1–0.4 Hz and amplitudes of several tenths of a Pascal. Due to their pervasiveness, they routinely regulate the noise level and thereby determine the detection threshold in that band. Their generation depends on the amplitude of ocean waves generated by storms [7], and its propagation is highly dependent on stratospheric winds [8,9].

Ambient infrasound noise has been examined on a limited scale in recent years. Infrasound noise has been surveyed when selecting sites for new infrasound stations [10,11]. The relationship between surface wind speed and noise has also been explored [12,11].

In this paper, we present a spectral analysis combined with probability statistics to investigate the background noise conditions at selected locations and show examples of the influence of wind speed direction on infrasound observations recorded by the network since early 2024.

II. DATA

A. Infrasonik Data

Figure 1 and Table 1 determine the location of infrasound observation stations that have been built in 2024. We use infrasound data from the I2S2 station whose data has been available since early 2024 and is located relatively close to ocean waters. The volume of data availability from each observation location is shown in Figure 2, where observation data from March to July 2024 is fully covered by observation data from all three locations. In addition, complete data is obtained from observations of I2S3_BRIN Pameungpeuk followed by IS2_BRIN Ancol and I2S4_BRIN Cibodas. This paper uses I2S2_BRIN Ancol data because of the availability of AWS data at the Jakarta location, while in the southern region of West Java there are no AWS observations yet.

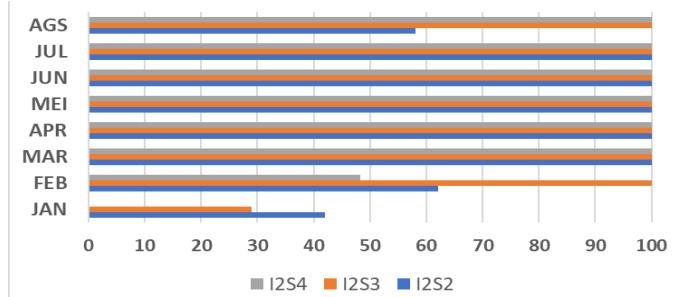


Fig. 1. Statistics on the availability of infrasound observation data in 2024 in West Java.

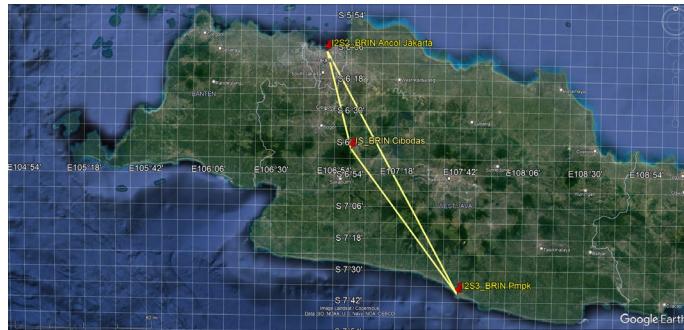


Fig. 2. Map showing the locations of infrasound observations in West Java.

TABLE I. LOCATION OF INFRASONIC STATIONS

No .	Stati on	Location	Longitude	Latitude
1.	I2S2	BRIN North Jakarta	106°50'45.97"E	6°7'31.62"S
2.	I2S3	BRIN Pameungpeuk, West Java	107°41'36.82"E	7°38'59.11"S
3.	I2S4	BRIN Cibodas West Java	107° 0'20.04"E	6°44'26.39"S

Each location consists of an infrasound sensor, accelerometer, temperature sensor, digitizer and single board computer as its main components. All components are integrated into one box and placed indoors to reduce and suppress uncorrelated high-frequency noise. All locations use the same infrasound sensor INF04.

B. Weather Conditions on the Ground Surface

In this study, we use surface meteorological measurement data to see the surface conditions that are factors that influence the propagation of infrasonic waves. Sound propagation in the atmospheric wave guides deals with the atmospheric layers having vertical profiles of the effective sound speed $C_{eff}(z)$ as follow:

$$C_{eff}(z) = \vec{V} \cdot \vec{n} + c \approx V \cos(\theta) + c + O(V^2/c^2)$$

where $c = \sqrt{\gamma RT}$ is the sound speed, R is gas constant, γ is the ratio of heat capacities of air at constant volume and pressure, $\gamma = C_p/C_v$, \vec{V} is the wind velocity vector, T is the absolute temperature, \vec{n} is the normal to the wave front of acoustic wave and θ is the azimuth of the reception point with respect to the direction of the wind velocity vector.

The data used in this paper is weather condition data around the infrasonic observation location obtained from AWS measurements obtained from the AWS network in the Jakarta area. The AWS network location map recorded on the ecowitt website is shown in figure 2.

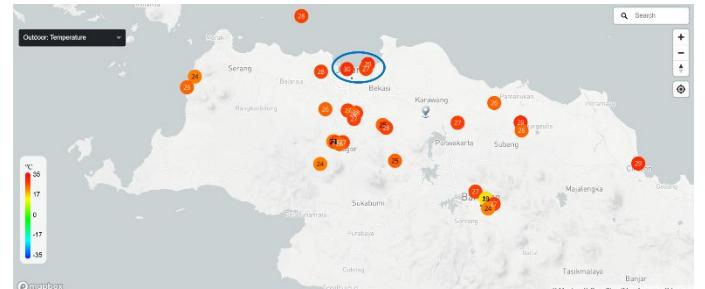


Fig. 3. Map showing the location of surface meteorological condition observations via AWS in West Java. (source: ecowitt.net). The AWS measurement area used in this paper is Jakarta. (blue circle).



Fig. 4. Locations of weather condition observations using AWS around the infrasound observation station in Jakarta. There are two AWS locations, AWS-1 (red balloon A) and AWS-2 (red balloon B).

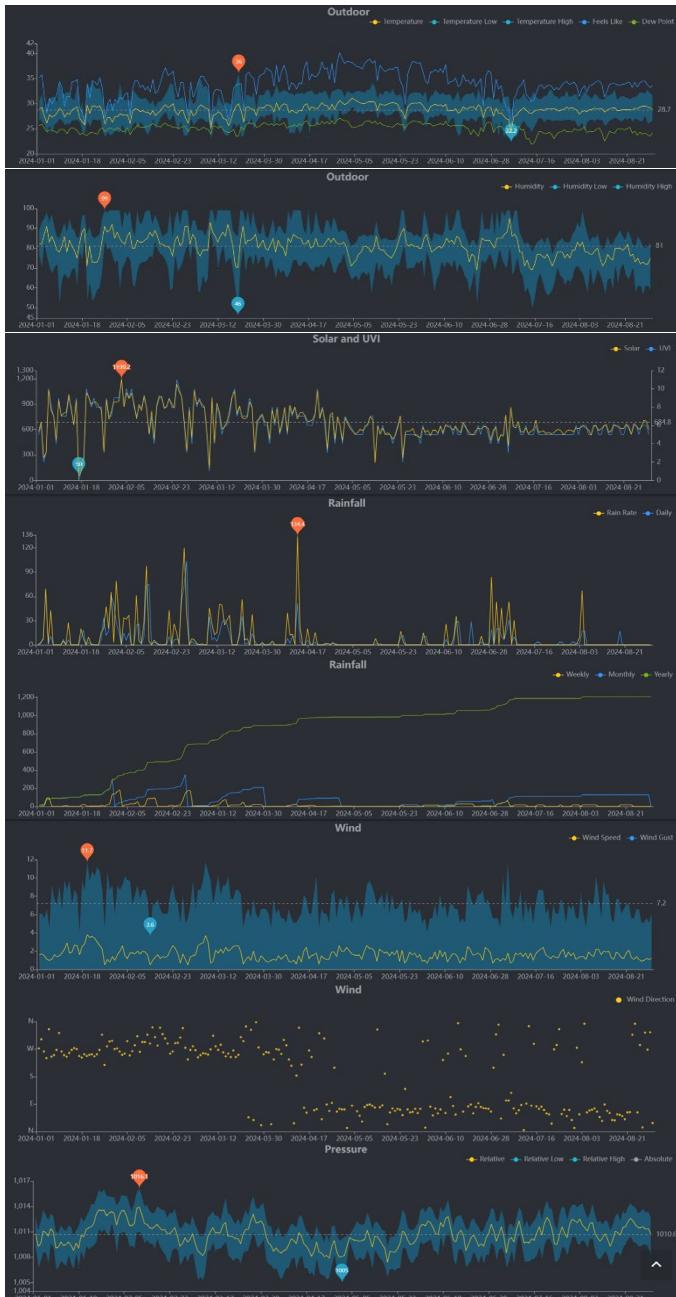


Fig. 5. Surface weather condition observation data profile in the Jakarta area in 2024 based on AWS-1 measurements in figure-4 (data source: www.ecowitt.net).

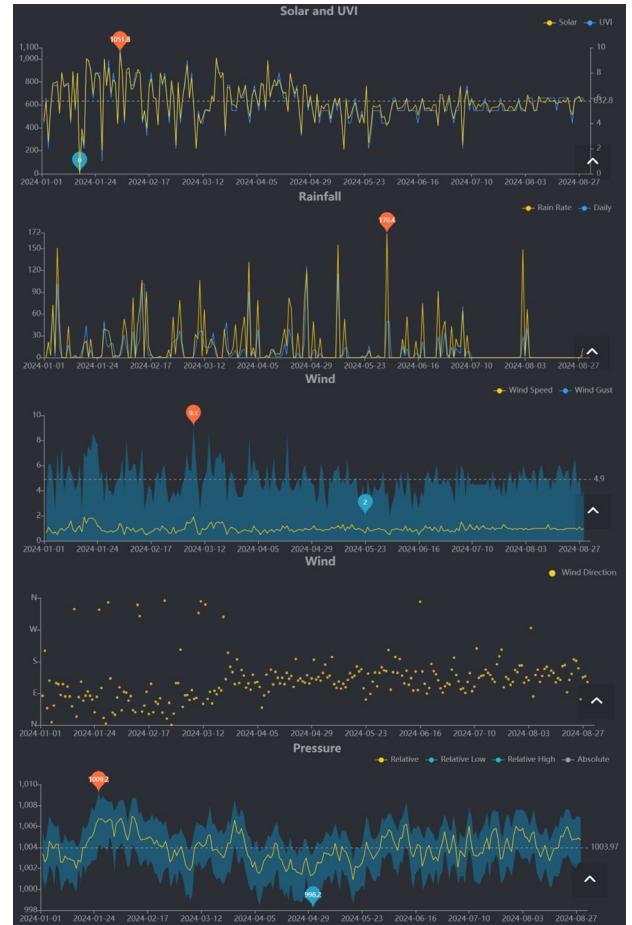


Fig. 6. Surface weather condition observation data profile in the Jakarta area in 2024 based on AWS-2 measurements in figure-4 (data source: www.ecowitt.net).

III. DATA PROCESSING

The method used in this paper is the 2024 infrasonic data in Jakarta. Infrasonic data was obtained from a single station that operates continuously in North Jakarta. In addition, to see the effect of infrasonic propagation received by the system, surface weather observation data in 2024 in Jakarta was also used. Surface weather data in Jakarta in 2024 was obtained from two Automatic Weather Stations (AWS) in North Jakarta.

Data analysis was carried out in the following manner:

1. Infrasonic data that has a stochastic time series $u(t)$, and its discrete form is represented as $u(n)$. With the periodic data $u(t)$ obtained, for analysis purposes, the data is divided into 1-hour finite-length time series of M segments, overlapping 50 percent to reduce the variance in the Power Spectral Density (PSD) estimation [13]. Each PSD can be directly estimated as:

$$p(\omega) = \frac{1}{T} |U(\omega)|^2$$

Where $U(\omega)$ is the Fourier spectrum of $u(t)$ that is

$$U(\omega) = \int_{-T/2}^{T/2} u(t) e^{-i\omega t} dt$$

While the discrete form of $p(\omega)$ is

$$p(k) = \frac{1}{N} |U(k)|^2$$

Where N is the total number of data points from $u(n)$.

We use Power Spectral Density calculation to characterize the signals received by our sensors that are distributed across various frequencies. The infrasound signal data obtained from our sensors is a very long time series data so that it can be done in characterizing the infrasound data in the frequency domain. This requires two calculations of Fast Fourier Transform (FFT) and Power Spectral Density (PSD). We do not just estimate the FFT value of the infrasound signal, but from the data we see the energy distribution across the frequency spectrum. Therefore, in the PSD calculation, the selection of the window function, its overlap, and the number of FFT points are the main parameters in the PSD calculation. Basically, When the window size is smaller it will increase the total number of data windows to be divided. This will help get a smooth PSD estimate. However, the disadvantage of a smaller window size is that the frequency resolution is disturbed because the distance between two frequency points increases and this will result in lower frequency resolution. On the other hand, when the total number of windows is reduced due to the increase in window length, the PSD estimates show only slightly sharper frequency resolution but become noisier because the noise effect is not cancelled out due to the reduced number of windows. We also compared different percentages of window overlap. The results show that with 0% overlap (for a given window size), the invertibility is lost, the total number of windows is less than with 50% overlap and the PSD results are relatively noisier and bumpier compared to 50% overlap. Likewise, when increasing the overlap percentage to higher values, 90% and 99%. This does not help much because of the high correlation between window samples and thus the average does not cancel out the noise effect followed by its variance. Finally, when the overlap percentage is 75%, the results are not much different from 50% overlap.

As mentioned above that PSD estimates the distribution of a signal's strength across a frequency spectrum. Therefore, the PSD units for a signal with an acceleration measurement in unit m/s^2 are $m^2/s^4/Hz$. We convert the PSD estimate into decibels (dB) with respect to acceleration $m^2/s^4/Hz$, by $10 * \log_{10}(PSD)$.

2. After a set of estimates of M segments and the length of each segment is L (1-hour), the average PSD can be calculated by:

$$\bar{p}(k) = \frac{1}{M} \sum_{m=1}^M \left[\frac{1}{L} |U_m(k)|^2 \right]$$

At this stage, the average value of $\bar{p}(k)$ is corrected for the sine taper at the beginning of processing and then deconvolves the seismometer instrument response by dividing the PSD estimate, $p(k)$, by the instrument transfer function with respect to acceleration, in the frequency domain. Finally, we convert the smoothed PSD estimate to

decibels (dB) with respect to acceleration, for direct comparison with the Low Noise Model (LNM), based on:

$$\bar{p}(k) = 10 * \log_{10}(p^-(k))$$

3. Finally, the goal of this paper is to gain insight into the actual noise variation at a given station. We do this by generating an infrasonic noise PDF from the processed PSD using the method discussed in the previous section. To adequately sample the PSD, full-octave averages are taken at 1/8 octave intervals. This procedure reduces the number of frequencies by a factor of 169 from nfft=16,385 to 97. Thus, the power is averaged between the short period (high frequency), T_s , and the long period (low frequency) $T_l=2*T_s$, with the centre period, T_c , so that $T_c=\sqrt{T_s*T_l}$ is the geometric mean period within the octave. The geometric means are evenly spaced in logarithmic space. The average power for that octave, spanning the period from T_s to T_l , is stored with the octave centre period, T_c , for the upcoming 0.125 analysis. T_s is increased by one 1/8 octave so that $T_s = T_s * 2$, to calculate the average power for the next bin period. T_l and T_c are recalculated, the power is averaged over the next period range T_s to T_l , and the process continues until we reach the longest period that can be resolved by the time series window length of the original data, $T_r/10$. Basically, the octave band in octave analysis is a frequency band where the highest frequency is twice the lowest frequency (i.e., $f_u = 2^{1/2}f_c$ and $f_l = \frac{f_c}{2^{1/2}}$). In addition, the eighth octave band is defined as a frequency band whose upper band frequency f_u is the lower band frequency f_l times the eighth root of two; The relationship between the upper, middle, and lower band frequencies is as follows:

$$f_u = 2^{1/8}f_l, f_u = 2^{1/16}f_c = 1.044f_c,$$

$$f_l = \frac{f_c}{2^{1/16}} = 0.958f_c$$

where f_c is the centre frequency. The fractional bandwidth percentage per eighth octave band at each centre frequency is $100 \left(\frac{f_u - f_l}{f_c} \right) = 8.6\%$. So the bandwidth of each band is the centre frequency f_c times 0.086.

4. This process is repeated for each 1-hour PSD estimate, producing thousands of refined PSD estimates for each station component. Power is accumulated in 1 dB intervals to produce a frequency distribution plot for each period.
5. The next step is to plot the distribution of powers per period, as mentioned above, using a probability density function (PDF). Then in a plot of the probability of the distribution PDF, the minimum, mode, and maximum powers for each period are plotted.

The analysis of the infrasound propagation test was carried out by interpreting the significance of the daily magnitude of the received infrasound signal to see the significance of the influence of surface weather conditions on infrasound reception at the infrasound observation station.

IV. RESULTS AND DISCUSSION

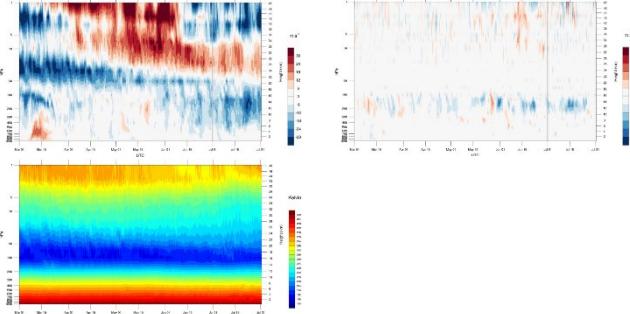


Fig. 7. The wind velocity (u, v) components of the wind field measured at the geographical coordinate of infrasound station. The top, middle and bottom plots present the u and v components and temperature in consecutive 25-min periods predicted based on ECMWF data.

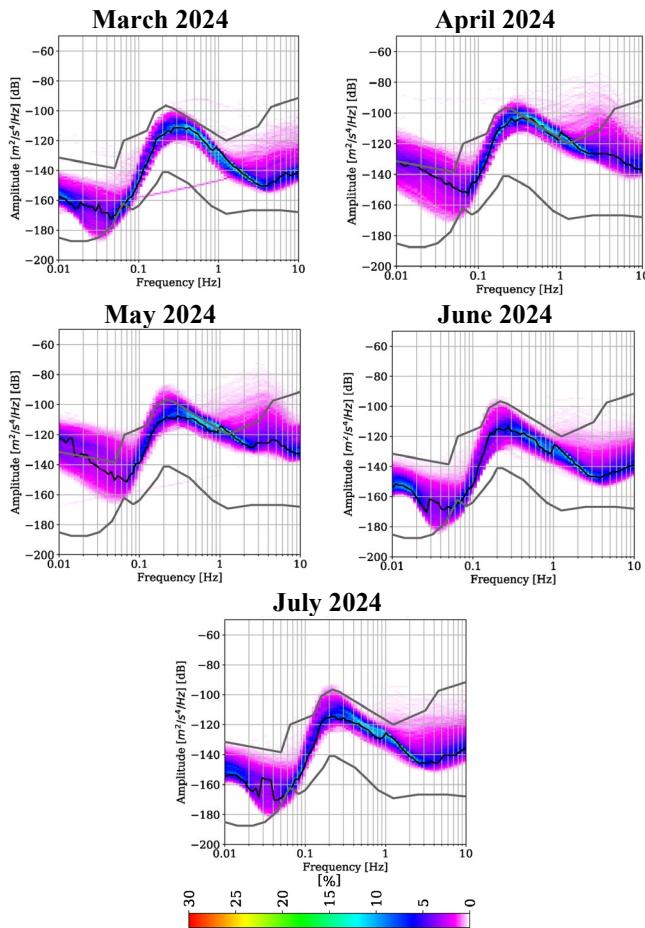


Fig. 8. Probabilistic Power Spectral Density (PPSD) of infrasonic station at North Jakarta (I2S2) calculated over one month of operations. The thick grey lines at the top and bottom represent the HNM and LNM, respectively from Peterson (1993). The solid black line represent the energy mode of the ambient variation levels (McNamara and Buland, 2004).

The characterization of noise energy compared to HNM and LNM of Peterson (1993), especially at lower frequencies, can verify the cause of the stable frequency range and provide important noise level information for the analysis of its dominance at the infrasonic level. Here, the noise power levels of each month are discussed. As above, the data used in

our analysis is the infrasonic component in the low frequency range between March and July 2024. By combining the interpretation of the results of the European Centre for Medium-Range Weather Forecasts (ECMWF) [15] atmospheric vertical model with the parameters of the horizontal wind components (u, v) for one year of data in 2024 (figure-7) with a plot of weather condition observation data using AWS at two locations (figures-5 and 6) and the results of our model calculations using infrasonic data are as follows. The easterly wind starts to blow strongly in the period from March to June and gradually continues thereafter until August 2024. During this period, it is possible for the formation of a wavefront deflector layer in the propagation of infrasonic waves in the mesosphere layer ($\sim 30-50$ Km) and returns to the Earth's surface. On the other hand, the wind dominates blowing towards the West from March to May. At the beginning of this period, the stratospheric wind blows very strongly towards the West during the period of March and April which has the potential to form a ducting layer in the stratosphere region. Meanwhile, the meridional wind component is seen to dominate in the troposphere layer where from April to August 2024, the wind direction tends to blow towards the South. This condition strengthens the infrasonic wave propagation mechanism towards the infrasonic station. Complementing these results, the PPSD profile shown in Figure 8 shows that from May to July 2024, the spectrum peak tends to narrow at a frequency of ~ 0.2 Hz. Sutherland L.C and, Bass H.E (2004) measured microbaroms at frequency range between $0.1 - 0.5$ Hz can potentially propagate over long distance due to low absorption. The probability value in this observation period, especially in the 0.2 Hz spectrum, the probability distribution is located in the estimated limit region of the High Noise Model (HNM) and Low Noise Model (LNM) values. In contrast, observations in March and April 2024, the probability distribution of the infrasonic data PSD is generally seen outside and exceeding the estimated limits of the HNM and LNM values. The peak around the 0.2 Hz spectrum looks wider compared to observations after May-July 2024.

Noise Levels in the Coast of Jakarta

The characterization of noise energy compared with the HNM and LLNM o Peterson (1993), especially at lower frequency (< 0.1 Hz), could verify the cause of the stable frequency range. Here, we elaborate noise power levels according to four cascade frequency bands.

a) Very low frequency band (< 0.02 Hz)

In this frequency band, the noise energy on June and July 2024 are well within the HNM and LNM. In contrast, the noise energy levels for the March-May 2024 fluctuated dominantly (~ 25 dB), with several PSDs and the noise mode exceeding the HNM (Fig. 8). Moreover, an ~ 30 dB difference between the noise levels of the March-May 2024 and the June-July 2024 was evident from the noise mode. Tilt noise induced by seafloor currents give rise to such unstable and elevated ambient variations levels in this frequency band.

b) *Middle Low Frequency band (0.02 – 0.1 Hz)*

In this frequency band, the noise can be generated by conversion of far-field energy induced by very large oceanic storms, low frequency ocean wave energy generated by vertical pressure variation of ocean currents at the seafloor or by breaking of surf on a sloping beach (McNamara & Buland 2004). Meanwhile, noise power in this band is completely dependent on climate conditions (Greeu et al, 2018).

c) *Middle Up Frequency Band (0.1 – 1 Hz)*

This band was consistently the most energetic part of the PPSDD plots within 0.01 – 10 Hz (Fig. 8). The PSDs with high probability were mostly below the HNM for all components, including the noise energy mode. The plots of March-May 2024 in this band were ~10 dB higher than the June-July 2024 plots. The origin of the secondary peak is related to standing gravity waves, the amplitude of which could be markedly increased by local storms (McNamara and Buland 2004).

d) *High Frequency Noise Band (> 1 Hz)*

In the high frequency range (>1 Hz), peaks were registered on all of the March-May 2024 results. The peak in this range is generally related on the wavefield of short-wavelength shear modes on the ocean floor. Low probability smeared signals exceeding the HNM at > 1 Hz were probably related to body waves from local small events (Vassallo 2012).

V. CONCLUSION

The infrasonic wave observation system has been built since early 2024. Infrasonic observation data from a station in North Jakarta (I2S2) is used to implement the PSD combination method with probability statistics. The probability of infrasonic data is seen with a peak spectrum around 0.2 Hz with a narrowing spectrum width in May-July 2024 and the probability area is located in the region between the estimated limits of the High Noise Model and Low Noise Model values. The horizontal wind speed vector around the location also determines the quality of the infrasonic signal reception data at the station.

ACKNOWLEDGMENT

This activity was funded by the Satellite Technology Research Center Program House-Aeronautics and Space Organization of BRIN and the Disaster Research Center Program House-Disaster and Maritime Research Organization of BRIN. The author also expresses gratitude for all technical assistance in the field to all parties involved in the installation and monitoring process of the system at the observation location.

REFERENCES

- [1] Louis C. Sutherland, Henry E. Bass; Atmospheric absorption in the atmosphere up to 160 km. *J. Acoust. Soc. Am.* 1 March 2004; 115 (3): 1012–1032. <https://doi.org/10.1121/1.1631937>
- [2] Brachet, N., Brown, D., Le Bras, R., Cansi, Y., Mialle, P., Coyne, J. (2010). Monitoring the Earth's Atmosphere with the Global IMS Infrasound Network. In: Le Pichon, A., Blanc, E., Hauchecorne, A. (eds) Infrasound Monitoring for Atmospheric Studies. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9508-5_3
- [3] Pichon, A.L. and Blanc, E. and Hauchecorne, A., Infrasound Monitoring for Atmospheric Studies: Challenges in Middle Atmosphere Dynamics and Societal Benefits. Springer International Publishing, 2018
- [4] Christie, D.R., Campus, P. (2010). The IMS Infrasound Network: Design and Establishment of Infrasound Stations. In: Le Pichon, A., Blanc, E., Hauchecorne, A. (eds) Infrasound Monitoring for Atmospheric Studies. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9508-5_2
- [5] Peterson, J., 1993. Observations and modeling of seismic background noise, Open File Rep, Albuquerque Seismological Laboratory - USGS, 93-322, <https://doi.org/10.3133/ofr93322>.
- [6] Jana, Niptika & Singh, Chandrani & Biswas, Rahul & Grewal, Nitin & Arun, Singh. (2017). Seismic noise analysis of broadband stations in the Eastern Ghats Mobile Belt of India using power spectral density. *Geomatics, Natural Hazards and Risk.* 7. <https://doi.org/10.1080/19475705.2017.136577>.
- [7] Arendt, Steve & Fritts, David. (2000). Acoustic radiation by ocean surface waves. *Journal of Fluid Mechanics.* 415. 1 - 21. [10.1017/S0022112000008636](https://doi.org/10.1017/S0022112000008636).
- [8] Garcés, M., M. Willis, C. Hetzer, A. Le Pichon, and D. Drob (2004), On using ocean swells for continuous infrasonic measurements of winds and temperature in the lower, middle, and upper atmosphere, *Geophys. Res. Lett.*, 31, L19304, <https://doi.org/10.1029/2004GL020696>.
- [9] Le Pichon, A., V. Maurer, D. Raymond, and O. Hyvernaud (2004), Infrasound from ocean waves observed in Tahiti, *Geophys. Res. Lett.*, 31, L19103, <https://doi.org/10.1029/2004GL020676>.
- [10] Garcés, Milton & Potter, Colin & Drob, Douglas & Picone, Michael & Hansen, Roger & Lindquist, Kent. (2000). Characterization of Predicted Infrasonic Phases For IS99, Hawaii. Proceedings of the 22nd Annual DoD/DOE Seismic Research Symposium: Planning for Verification of and compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT) held in New Orleans, Louisiana on September 13-15, 2000, U.S. Government or Federal Rights.
- [11] Hedlin, M.A.H., Berger, J., Vernon, F.L. (2002). Surveying Infrasonic Noise on Oceanic Islands. In: Der, Z.A., Shumway, R.H., Herrin, E.T. (eds) Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Data Processing and Infrasound. Pageoph Topical Volumes. Birkhäuser, Basel. https://doi.org/10.1007/978-3-0348-8144-9_11
- [12] Le Pichon, A., Ceranna, L., Vergoz, J., Tailpied, D. (2019). Modeling the Detection Capability of the Global IMS Infrasound Network. In: Le Pichon, A., Blanc, E., Hauchecorne, A. (eds) Infrasound Monitoring for Atmospheric Studies. Springer, Cham. https://doi.org/10.1007/978-3-319-75140-5_17.
- [13] Cooley, J.W., & Tukey, J.W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19, 297-301.
- [14] Sutherland L.C and, Bass H.E (2004) measured microbaroms at frequency range between 0.1 – 0.5 Hz can potentially propagate over long distance due to low absorption.
- [15] Landès M. , Le Pichon A. , Shapiro N.M. , Hillers G. , Campillo M. , 2014. Explaining global patterns of microbarom observations with wave action models, *Geophys. J. Int.*, 199, 1328–1337.. [10.1093/gji/ggu324](https://doi.org/10.1093/gji/ggu324)
- [16] Vassallo, M., Festa, G. & Bobbio, A., 2012. Seismic ambient noise analysis in southern Italy, *Bull. seism. Soc. Am.*, 102, 574–586.
- [17] McNamara, D.E., & Buland, R.P., 2004. Ambient Noise Levels in the Continental United States, *Bull. Seism. Soc. Am.*, 94, 1517-1527.

The Systematic Effectiveness Comparison of Spatiotemporal Carbon Dioxide Emission with Spatial Interpolation Methods: Case Study in Thailand

Suriyawate Boonthalarath
Applied Geoinformatics
School of Information and
Communication Technology
Phayao Province, Thailand
suriyawate.b@gmail.com

Nakarin Chaikaew
Applied Geoinformatics
School of Information and
Communication Technology
Phayao Province, Thailand
nakarin.ch@up.ac.th

Phaisarn Jeefoo
Applied Geoinformatics
School of Information and
Communication Technology
Phayao Province, Thailand
phaisarn.je@up.ac.th

Abstract— The rapid increase in greenhouse gases, especially the high concentration of carbon dioxide in the atmosphere, has aggravated the effects of the worsening global warming crisis. There is a growing focus on managing and tracking spatio-temporal carbon emissions to achieve carbon neutrality in academic research. However, it is challenging to measure and assess the dynamic release of human-made and natural carbon dioxide across broader spatial resolutions due to numerous missing data points. The interpolation method is an essential tool for predicting missing data in datasets that monitor variations in carbon dioxide from sampling points, satellite measurements, or ground-based observations. Various interpolation techniques have recently been used to analyze and evaluate the continuous increase in carbon dioxide pollution at different spatial scales across various disciplines. These techniques have proven to be valuable in decision-making and problem-solving. Unfortunately, there is a lack of research explaining the specific use of these systematic interpolation techniques. This study aims to compare the performance of four interpolation techniques in evaluating carbon dioxide emissions. The objective is to analyze the spatial and temporal dynamics in atmospheric carbon dioxide concentration by comparing the geo-statistics of the satellite OCO-2 dataset and the ground-based TCCON dataset within Thailand's boundaries from 2015 to 2020. Furthermore, each interpolation technique was assessed for accuracy using Root Mean Square Error (RMSE) in cross-validation studies as the criteria for selecting the most appropriate spatial interpolation techniques for specific scenarios. It was found that the four interpolation techniques demonstrated strong predictive performance in visualizing and analyzing Thailand's boundaries, as well as revealing the spatiotemporal patterns of atmospheric carbon dioxide concentrations. Although the results indicated that IDW and Spline interpolation outperformed the Natural Neighbor and Kriging methods, achieving a total RMSE of less than 0.01 in cross-validation. When comparing the IDW and Spline interpolation methods, it was discovered that the Spline algorithm showed overfitting for both datasets in comparison to IDW. As a result, IDW demonstrates a more suitable sensitivity variation for improving atmospheric carbon dioxide concentration in response to spatiotemporal human-made emissions.

Keywords—*Interpolation, OCO-2, TCCON, Atmospheric Transport Models, Spatiotemporal Model Simulations*

I. INTRODUCTION

In this day and age, global warming is producing significant impacts, resulting in dramatic environmental changes and driving global climate change. This presents the most pressing environmental challenge facing humanity today [1], [2], [3], largely driven by the surge in greenhouse gases, particularly anthropogenic carbon dioxide emissions since the Industrial Revolution [4]. Consequently, the resultant increase in atmospheric carbon dioxide levels significantly raises Earth's average land surface temperature [5], [6], [7]. Hence, the adverse effects of global warming on the environment are increasingly gaining attention as a critical challenge. Monitoring atmospheric carbon dioxide emissions is pivotal for comprehending the sources and sinks of carbon dioxide, which substantially contribute to global warming. This entails monitoring atmospheric carbon dioxide emissions to track changes in land cover [8] arising from rapid urbanization and industrialization [9], [10], [11], utilizing geospatial data, ground-based observations, and satellite-tracking datasets. Statistical measurements reveal that the emission data were influenced by geographical factors and constrained by various limitations. For instance, when examining the ground-based observation dataset, it was found that there were missing and unsampled data points across the continuous domain [12], [13], [14]. Furthermore, air pollution monitoring datasets only contain a few sampling points with limited spatial resolution, resulting in incomplete data in specific locations and offering limited information about the spatio-temporal variation of carbon dioxide emissions. Alternatively, analyzing data from remote sensing offers an effective approach to consistently monitor and capture extensive areas, providing valuable insights into key variables, even under challenging conditions such as cloud cover and aerosol-contaminated conditions or equipment failure [15], [16], [17].

The passage above underscores the challenges of staying abreast of the latest research and innovations in accurately predicting and tracking human-induced atmospheric carbon dioxide emissions at specific locations. This task necessitates the use of geospatial statistical models in conjunction with observational datasets, particularly when there is ample spatial gaseous compound data. At present, the estimation of carbon dioxide emissions is often improved through the use of

advanced techniques such as artificial neural networks (ANN) [18] and convolutional neural networks (CNN) [19], including generalized regression neural networks (GRNN) [20]. These approaches aim to predict carbon dioxide fluxes across various regions by leveraging published data that includes factors such as latitude, age, and potential net primary productivity. The findings demonstrate a strong correlation between the estimates generated and the existing carbon dioxide emission datasets [21]. Furthermore, these methodologies enable the development of super-resolution mapping, also referred to as sub-pixel mapping, which addresses the mixed-pixel issue. This strategy is designed to provide more accurate sub-pixel information by relying on the assumption of spatial dependence [22]. While these methods can yield promising results and effectively integrate data from diverse spatial, spectral, and temporal sources, they are primarily applied in isolation and designed for specific tasks under limited conditions [23]. Their inability to accurately estimate carbon dioxide levels across extensive areas undermines their effectiveness as indicators for wide-ranging carbon reduction policies [24] and as tools for mitigating carbon dioxide emissions at the national level [25]. Selective high-performance interpolation methods are employed for exploratory model analysis and geo-visualization to validate estimated values. These techniques aid in depicting spatial and temporal relationships between air quality and land-use patterns with precision. Overall, interpolation methods have been widely utilized to analyze and extrapolate data in fields such as environmental science [26], [27], [28], [29] and engineering [30], [31], [32], [33], thereby enhancing geospatial statistics, improving image quality, and facilitating the restoration of missing or distorted images from interpreted datasets [34], [35].

When examining the release of atmospheric carbon dioxide, it is crucial to take into account various factors related to sources and sinks. This involves evaluating the characteristics of the area [36], [37], [38], selecting the appropriate tracking dataset [39], [40], [41], and employing a suitable interpolation method [42]. All studies strive to assess the diffusion levels and human-caused carbon dioxide sources consistently. Therefore, it is essential to scrutinize the spatial variation of intricate factors and their impact on the escalating levels of carbon dioxide in the atmosphere from an academic standpoint. However, the justifications for the selection of methods are frequently disregarded. It is vital to consider interpolation methods and the nature of the dataset used for analysis, while also providing rationale within the context of the mounting trend of carbon dioxide levels in the atmosphere. In this research, we aim to decipher the intricate spatial elements associated with the increase in carbon dioxide levels in academic settings. Our primary objective is to elucidate the use of interpolation methods and decision criteria, in conjunction with data sets acquired from various collection techniques, in order to achieve a comprehensive comprehension. The investigation will be carried out within the geographical confines of Thailand as a case study, spanning the period from 2017 to 2022.

II. MATERIALS AND METHODS

In the past century, the growth of cities and industries has resulted in increased burning of fossil fuels, which has led to a significant rise in atmospheric carbon dioxide, a major greenhouse gas [43] that affects surface temperature. Understanding the details of cyclical carbon dioxide

emissions, including sources, sinks, and global distribution [44], [45], [46], is important for identifying the processes driving changes in its concentration. Satellite observations from instruments such as SCIAMACHY [47], GOSAT [48], and OCO-2 [49] have been used to gather global monitoring data. These instruments measure specific wavelengths related to reflected solar radiation to quantify atmospheric carbon dioxide concentration in the spectral region. Ground-based measurements have been widely used to assess high-precision spatiotemporal carbon dioxide emissions [50] in challenging areas more effectively than satellite observations. The monitoring ground-based network, known as the TCCON station, tracks dynamic anthropogenic emission sources by interpreting solar spectra to accurately retrieve carbon dioxide emissions [21]. In this article, we will compare the characteristics of both datasets to evaluate the performance of measuring atmospheric carbon dioxide emissions using various interpolation techniques in Thailand's study area.

A. OCO-2 Satellite Dataset

In this research paper, a dataset from the OCO-2 satellite, which was launched by NASA in July 2014, was used to study the monitoring of atmospheric carbon dioxide. The OCO-2 satellite is equipped with a three-channel high-resolution grating spectrometer that measures sunlight [51]. The retrieval of carbon dioxide concentrations from the satellite's high-spectral-resolution measurements is based on nonlinear Bayesian optimization theory [52], which involves analyzing scattered and reflected radiation from the land surface or atmosphere across near-infrared to short-wave infrared bands [53] using inversion algorithms [54], [55], [56]. For this study, the OCO-2 XCO₂ (OCO2_L2_Lite_FP v10r) observations spanning from 1 Jan 15 to 1 Jan 20 were selected, and the dataset utilizes the ACOS algorithm developed by NASA. Product specifications are detailed in TABLE I.

TABLE I. CHARACTERISTICS SUMMARY OF OCO-2 SATELLITE DATASET FOR MONITORING ATMOSPHERIC CARBON DIOXIDE EMISSION

Satellite	Characteristics of Satellite Dataset Used [57]	
OCO-2	<i>Version</i>	OCO2_L2_Lite_FP v10r
	<i>Algorithm</i>	ACOS
	<i>Revisit</i>	16 day
	<i>Application</i>	Greenhouse Gas Monitoring
	<i>Spatial Resolution</i>	2.25 km × 1.29 km
	<i>Swath Width</i>	10 km with 8 Across-Track Measurements
	<i>Temporal resolution</i>	Monthly

B. TCCON Groud-base Dataset

The Total Carbon Column Observing Network (TCCON) is a global ground-based network that monitors greenhouse gases in the atmosphere. Measurements are taken using a high-precision Fourier transform spectrometer, which captures the absorption of atmospheric emission gases related to the direct solar spectra in the near-infrared band. An advanced nonlinear least squares spectral fitting algorithm [58] is then used to determine the total concentration of the gases accurately. All TCCON measurement datasets in this

article used version GGG2020 as the investigated dataset and specific data is explained in TABLE II below.

TABLE II. CHARACTERISTICS SUMMARY OF TCCON GROUND-BASE DATASET FOR MONITORING ATMOSPHERIC CARBON DIOXIDE EMISSION

TCCON Station	General Information of the TCCON Site Stations and the Temporal Range of Selected Data [59].	
Burgos	<i>Country</i>	Philippines
	<i>Longitude (°)</i>	120.6496 E
	<i>Latitude (°)</i>	18.5325 N
	<i>Abbreviation</i>	BU
	<i>Temporal coverage</i>	20170303-20230623

TCCON observation data is currently the optimal choice for validating OCO-2 satellite observations and is widely utilized in accuracy assessments of satellite estimations of atmospheric carbon dioxide concentrations and corrections of relevant parameters [60]. When observing space and time, we cannot directly compare the carbon emission data collected by the OCO-2 satellite with the ground-based TCCON data. This is due to the different ways in which the two datasets retrieve atmospheric carbon emission data in terms of spatial and temporal resolution. However, the TCCON dataset provides a standard static sampling point, which we used as a reference for comparing with satellite observations within specific spatial boundaries [61], [62].

C. The Specifications Study Area

Thailand, situated in Southeast Asia, boasts a varied landscape featuring towering mountains in the north and west, a central plain, and a sizable plateau in the east. As of the end of 2023, the population stood at approximately 66 million [63], with around 33% residing in the northeastern region and roughly 30% in the central region. This demographic distribution has led to unequal economic and social development throughout the country. The rapid pace of urban development, outpacing population growth, has the potential to result in increased carbon dioxide emissions [64]. The southern and western regions, renowned for their plains and mountains, primarily focus on resource utilization and offer tourism services. These areas are home to cities of varying sizes and economic significance. The unique natural and socioeconomic characteristics of each region make them intriguing subjects for the study of atmospheric carbon dioxide emissions [65]. Thailand's geographical map and population distribution are depicted in Fig.1 for a visual representation.

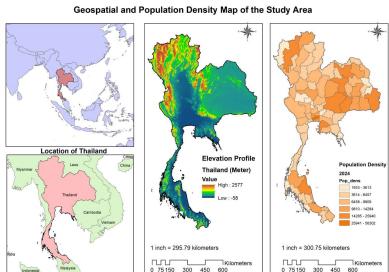


Figure 1 Thailand's study site includes the location of the boundary (left) different morphological features (middle) and a population density map (right)

D. Interpolation Techniques and Cross-Validation

Interpolation techniques and model simulation are important for analyzing the spatio-temporal characteristics of carbon dioxide emissions. While satellite observations and ground-based monitoring have limitations, model simulation can provide comprehensive coverage of atmospheric carbon dioxide concentrations on both a global and local scale. This research aims to estimate carbon dioxide concentrations in areas where measurements are not available using four methods: Inverse Distance Weighting (IDW), Spline, Ordinary Kriging, and Natural Neighbors method [66], [67], [68], [69]. Here's a summary of each method:

a) Inverse Distance Weighting (IDW) Interpolation: is a straightforward and automated technique used to estimate missing neighborhood parameters with minimal data input [70]. It effectively smooths geo-visual data and is well-suited for small datasets. Additionally, it is highly adaptable for estimating datasets with trends in neighborhood shaping. In contrast, other fitting techniques may be susceptible to errors [71]. IDW directly assumes that missing data from an unsampled location is a weighted average of known data points within a local neighborhood surrounding the unsampled location [72], as shown in the following formula:

$$u(x, y) = \frac{\sum_{n=1}^N \frac{u_n(x_n, y_n)}{d_n}}{\sum_{n=1}^N \frac{1}{d_n}} \quad (1)$$

This method simulates the pollutant measurements at N locations surrounding the unknown location (x_n, y_n) where d_n is a commonly squared distance

b) Spline Interpolation: Spline interpolation can be performed using two methods: regularized and tension. The regularized method results in a more flexible, smoothly varying surface, whereas the tension spline typically produces smoother surfaces based on the same sample points [73]. The tension curve is flatter than the regularized curve, exerting greater influence on the estimates to remain closer to the sample data. While the tension spline incorporates only the first (slope) and second derivatives (rate of change in slope), the regularized method involves the first, second, and third derivatives (rate of change in the second derivative) in the spline computation. The equation for a spline with a tension function is given by (2) [74].

$$\phi(r) = \ln\left(\frac{\sigma \cdot r}{2}\right) + K_0(\sigma \cdot r)^2 + C_E \quad (2)$$

Where r —distance between the prediction point and the sample; σ —tension parameter; E_1 —exponential integral function; C_E —constant of Eulero (0.577215); K_0 —modified Bessel function.

c) Ordinary Kriging Interpolation: Kriging is a method that involves combining measurements from nearby monitoring stations using weighted averages. Instead of assuming a function based on inverse distance, kriging calculates weights for each concentration based on the spatial correlation among the observed measurements [75], [76]. One of the main benefits of kriging is its ability to produce estimates and standard errors at locations where no measurements were taken. These standard errors indicate the uncertainty associated with the forecast [77]. Two common

forms of kriging are used in geostatistics: ordinary and universal. Ordinary kriging assumes a constant unknown mean in the local neighborhood, while universal kriging assumes a general polynomial trend model. Ordinary kriging is the more widely used method and is preferred for its flexibility [78], [79]. Kriging estimate $Z(x_0)$ and error estimation variance $\sigma_k^2(x_0)$ at any point x_0 were, respectively, calculated as follows (3) and (4)

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (3)$$

$$\sigma_k^2(x_0) = \mu + \sum_{i=1}^n \lambda_i \gamma(x_0 - x_i) \quad (4)$$

Where λ_i are the weights, μ is the Lagrange constant, and $\gamma(x_0 - x_i)$ is the semivariogram value corresponding to the distance between x_0 and x_i [80]

d) Natural Neighbor Interpolation: The natural neighbor (NN) interpolation method uses a weighted average of local data points. It is based on the concept of natural neighbor coordinates derived from Thiessen polygons for bivariate data and Thiessen polyhedra for trivariate data [81]. This method finds the closest subset of input points to an unknown point and applies weights to them based on proportionate areas in order to interpolate a value [82]. The natural neighbors of any point are those associated with the neighboring Voronoi polygons. Initially, a Voronoi diagram is constructed from all given points, and a new Voronoi polygon is then created around the interpolation point. The proportion of overlap between this new polygon and the initial polygons is then used as weights. The value in an unsampled location is computed as a weighted average of the nearest neighbor values, where the weights are dependent on areas or volumes rather than distances. The number of given points used for the computation at each unsampled point is variable, dependent on the spatial configuration of data points. The natural neighbor interpolation is defined as (5) and (6) respectively [83].

$$\hat{z} = \frac{1}{\sum_{k=1}^n Area_i} \sum_{k=1}^n Area_k \cdot Z_k \quad (5)$$

$$Area = \frac{1}{2} \sum_{i=1}^n x_i y_{i+1} - x_{i+1} y_i \quad (6)$$

Where the value Z_k for each neighbor is combined with its associated weight to compute an estimated value for \hat{z} at the interpolation coordinates and $Area_i$ is the overlap area corresponding to the known data site i

Natural neighbor linear interpolation results in a rubber-sheet character of the resulting surface. When blended gradient information (derived from data points by local ‘pre-interpolation’) is added, the resulting surface can be made smooth everywhere with tautness, analogous to tension, tuned according to the character of the modelled phenomenon. The value of tautness is controlled by two empirically selected parameters that modify the shape of the blending function. The result is a surface with smoothly changing gradients that passes through data points, blended from natural neighbor local trends, with local tunable tautness, and with the capability to calculate derivatives and integrals. The surface passes through the input samples and is smoothed everywhere except at the locations of the input samples. It adapts locally

to the structure of the input data, requiring no input from the user pertaining to search radius, sample count, or shape. It works equally well with regularly and irregularly distributed data.

Cross-validation is a widely used method for assessing the predictive accuracy of various interpolation techniques. This technique entails systematically leaving out individual data points and utilizing the remaining data [73], [84], [85] to perform interpolation. Subsequently, the disparities between the actual value of the omitted point and its estimated value are computed. This process is iterated until each observation is estimated using the remaining observations in the sample, one by one [86]. During cross-validation, the accuracy of each interpolation is assessed using statistical methods such as mean error (ME) and root mean square error (RMSE). RMSE is especially useful for determining the most accurate interpolator for estimates at unsampled locations, as well as for uncovering prominent patterns in the data series [87]. It can be mathematically represented as [88]:

$$ME = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (8)$$

Where y_i is the actual value, \hat{y}_i is the predicted value and n is the number of observations.

E. Preparing the Datasets in terms of Spatial and Temporal Resolutions

Before comparing the ground-based and satellite datasets, it's important to note that different CO₂ products can't be directly compared because of variations in their data sources and sampling methods. To account for these differences, an adjustment needs to be made before comparison. In this study, the model-simulated CO₂ profiles were converted to XCO₂ using an averaging kernel and a priori profile of the OCO-2 dataset referred to as Rodgers and Connor [89]. The adjustment equation is expressed as follows:

$$XCO_2^m = XCO_2^a + \sum_j h_j a_j (y_m - y_a)_j \quad (9)$$

The notation used in the equation is as follows: XCO₂^m represents the integrated XCO₂ from OCO-2; XCO₂^a denotes the TCCON a priori XCO₂; h stands for the pressure weighting function; a is the TCCON column averaging kernel, which depends on both pressure and the solar zenith angle; y_m is the simulated CO₂ profile; and y_a is the TCCON a priori CO₂ profile. For a more detailed explanation of this method, please refer to Wunch et al [58].

Additionally, the CO₂ data from both the OCO-2 dataset and TCCON are represented in CO₂ profiles at varying vertical levels. To compare them on a global scale, the CO₂ profiles are uniformly converted to XCO₂ using the weighting pressure-averaged method outlined by O'Dell et al. [54]:

$$XCO_2 = h^T \hat{u} \quad (10)$$

Where \hat{u} is the CO₂ profiles from the OCO-2 dataset on discrete pressure levels and h is the pressure weighting function. The specific conducted method of h is described in O'Dell et al [54].

III. RESULTS AND DISCUSSION

It is widely accepted that ground TCCON XCO₂ data are highly accurate for monitoring CO₂ emissions and are frequently used as a reference [90], [91]. However, Rodgers and Connor have noted that it is not ideal to directly compare measurements made by different remote sounders due to variations in their a priori profiles and averaging kernels [89]. Nonetheless, several studies have demonstrated that direct comparisons can still be valuable for validating satellite measurements [92], as demonstrated in the latest research on satellite measurement validation [93]. Additionally, there is compelling evidence that specific underestimations of XCO₂ over the tropical region in the winter season may be linked to satellite retrieval errors arising from tropical cloud cover during these periods. These findings are consistent with the research of Cogan et al [94]. In the present study, we relied on the TCCON ground-based data version GGG2020 from the Burgos site in the Philippines as our reference, spanning the time period between 20170303 to 20230623. All data collected are depicted in Figure 2A. However, it's worth noting that the XCO₂ data exhibit uneven distribution, characterized by multiple gaps in both spatial and temporal coverage due to observation factors such as cloud cover and instrument data screening limitations. Therefore, we create a trendline for the investigation using spatio-temporal geostatistics as shown in Figure 2B [21].

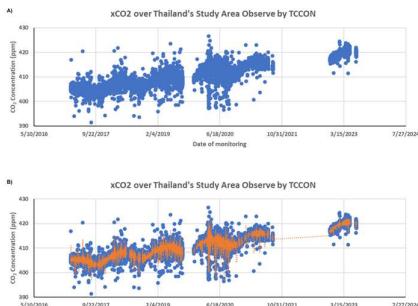


Figure 2 Number of available retrievals and temporal changes of collected XCO₂ data by TCCON
 (a) the number of XCO₂ data points
 (b) the temporal variation of XCO₂ and trendline

In this particular case, the TCCON observation station provided the hourly XCO₂ dataset, accessible at <https://tccondata.org/>. The dataset encompasses CO₂ emissions measured from 3 March 2017 to 23 June 2023. However, for the purposes of this study, the data used pertains to the period from 3 March 2017 to 3 October 2021 as there were notable data gaps, as depicted in Figure 2. We transformed the measured hourly XCO₂ dataset into a monthly dataset with a mean value [95], ensuring it reflects the geostatistical representation of the TCCON data. These monthly TCCON datasets are consistent over time and can be compared to satellite data for the development of an interpolation model in the next section.

A. Spatial Variability of Atmospheric CO₂ Concentrations Analysis and Interpolation Method

The research investigates the relationship between seasonal CO₂ concentration patterns and the efficacy of interpolation techniques at specific locations within the study area. Each interpolation method's predicted CO₂ distribution is integrated with a linear regression model of spatial and temporal CO₂ trends to evaluate the impact on variance

between time periods. This allows for a better understanding of how terrain attributes and climatic variables in the area influence CO₂ levels. The resulting geographical representation of each interpolation technique was assessed by comparing the CO₂ distribution with the applied estimates, as illustrated in Figure 3. The interpolation method includes IDW, Spline, Ordinary Kriging, and Natural Neighbor respectively.

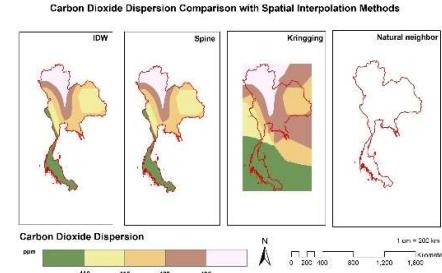


Figure 3 the performance of geo-visual in each interpolation method applied to estimate the distribution of CO₂ on the sample site

It is crucial to carefully choose the most suitable interpolation method, considering various factors including the characteristics of the observed variable, the timeframe of representation, the surface types for modeling, and the geographic area features [96]. Different interpolation methods operate subtly and distinctly to analyze pollutant characteristics across the entire area of interest, akin to utilizing a mathematical function's shape. The analysis presented demonstrated the XCO₂ OCO-2 dataset in 2017 validated with TCCON in Figure 3, we assessed the quality of predicted surfaces for the variability of atmospheric CO₂ concentrations using various interpolation methods. Our findings indicated that the IDW, Spline, and Ordinary Kriging approaches yielded similar surfaces. However, the natural neighbor (NN) method did not provide an accurate geographical visualization and produced inaccurate values, particularly in close proximity to large sample point locations. IDW and Spline interpolation consistently performed well in capturing CO₂ distribution, especially considering edge effects (unsampled locations) and clustered sample points. The prediction had a much greater morphology of the map (high differentiation between color ranges). Notably, IDW techniques produced values within the range of the original observations, while kriging and spline methods were capable of generating interpolated values both below and above the observed range [97]. When utilizing the Natural Neighbour method, it may not precisely convey the geographic features over the large area, as this interpolation technique is more effective when working with closely grouped scatter points [98]. However, a limitation of this method is that it only generates an interpolated surface within the confines of the data points and does not extend beyond them. Moreover, the Natural Neighbour method yields inferior results when the data points are unevenly distributed. Additionally, the composition of the sampling set plays a crucial role in determining the most suitable interpolation method. It's apparent that kriging interpolation is a more sophisticated approach compared to other methods. However, the presence of a restricted number of sampling points, irregular datasets, and clustered point locations could potentially affect the accuracy of the interpolated results [99], leading to display errors beyond the study's boundaries.

TABLE III. SUMMARY STATISTICS OF THE INTERPOLATION METHODS

Interpolation Method	Atmospheric CO ₂ Concentration (ppm)			
	Min	Max	Mean	RMSE
Observed Values	399.816	401.375	400.597	-
IDW	399.837	400.871	400.372	0.0021
Spline	399.835	401.879	400.357	0.0014
Kriging	399.818	401.238	400.469	0.4148
Natural Neighbor	399.920	400.873	400.329	1.2832

The accuracy of various interpolation methods is often assessed using cross-validation. In cross-validation, a single data point is removed, and interpolation is carried out using the data from the remaining locations [73], [100], [101]. The disparity between the measured value of the removed point and its estimation is then computed. This process is repeated until each observation in turn has been estimated from the remaining observations [86]. The model error for each interpolation in cross-validation is appraised using statistical measures such as the mean error (ME) and root mean square error (RMSE) [86]. RMSE identifies an interpolator that offers the most reliable estimates at an unsampled location and demonstrates a dominant pattern in the data series [87]. Hence, we select RMSE as a statistical measure to signify the effectiveness of the interpolation method and depict it in Table III.

The methods of IDW and spline are suitable for examining CO₂ variability in Thailand over a specific time frame when compared to natural neighbor and kriging interpolation techniques. The use of spline and IDW techniques in predicting spatiotemporal changes in atmospheric CO₂ concentration resulted in relatively accurate outcomes. According to the average prediction RMSE values (displayed in Table III), the spline technique produced the lowest RMSE values, signifying lower interpolation errors compared to the other methods. Both the IDW and spline interpolation results demonstrated minimal error variances (as indicated in Table 2). However, the IDW approach involved heavy averaging of the data and did not generate values greater than the maximum or smaller than the minimum values of the observations. In general, while the differences in RMSE between IDW and spline were generally small, the IDW method tended to exhibit higher errors, possibly as a result of the uneven distribution of input points across the study area, leading to edge effects and clusterization. The predictive capability of the IDW method was generally lower compared to spline predictions (resulting in higher RMSE values), possibly due to a slight underestimation of the observed concentrations. Both IDW interpolation and spline are deterministic methods that generate surfaces from samples based on the extent of cell similarity or degree of smoothing [102]. However, a spline mathematical function passes exactly through each sample point, while IDW does not pass through any of the points [103]. It's important to note that IDW is effective in identifying pollution hotspots from interpolated surfaces. On the other hand, spline tends to have a 'smoothing' or 'overfitting' effect, which could be caused by the insertion process of regional data sets into mathematical algorithms.

However, the natural neighbor and kriging estimation methods exhibited higher errors, as the predicted values closely aligned with the observed ones. Nonetheless, the

natural neighbor method did not consider extrapolation, which could result in significant errors in areas beyond the sampling locations. The statistical model was estimated by calculating the mean prediction and root-mean-square prediction errors (RMSE). The kriging experimental results revealed that the accuracy of CO₂ variation interpolation across different locations is primarily influenced by sampling density, sampling mode, and sampling location. When specific sampling datasets fail to sufficiently capture spatial characteristics, interpolation errors can quickly escalate, potentially leading to under- or overestimation of values. In typical Kriging methods, value variation is expected to increase with distance according to the semivariogram. However, in this instance, the visual representations of geographical data do not align with the anticipated results. The disparities in geographical relationships can be primarily attributed to the nature of data collection and organization, and the sampling mode significantly influences future predictions.

IV. CONCLUSIONS

The significant increase in atmospheric CO₂ levels primarily stems from human activities, notably the combustion of fossil fuels and biomass. Tackling this issue requires the implementation of ambitious measures to curtail carbon emissions. Advanced remote sensing technology, particularly satellite measurements, plays a pivotal role in gathering global and regional CO₂ data. However, it's important to acknowledge that estimates of XCO₂ from satellites may contain errors and uncertainties influenced by various factors, including the inversion model, measurement instruments, spatial resolution, atmospheric conditions, and surface parameters. Enhancing the accuracy of the data could involve the use of spatial interpolation methods, leveraging model assimilation data as a reference to accommodate the variability of CO₂ emissions in the atmosphere.

In our research, we conducted a comparative analysis of four distinct spatial interpolation methods for generating interpolated maps of atmospheric CO₂ concentrations. After evaluating these methods within the context of Thailand's boundaries, we found that the inverse distance weighting (IDW) and spline interpolation techniques yielded the most precise results compared to natural neighbour interpolation and Ordinary Kriging. We also determined that spline with tension interpolation was best suited for both spatial and temporal modeling of pollution in the area under study. Nonetheless, our findings have prompted further inquiries into the most suitable interpolation technique for specific geospatial features, the distribution of local emission sources, and the influence of atmospheric transport on observed seasonal variations in CO₂ levels.

ACKNOWLEDGMENT (Heading 5)

The authors acknowledge the School of Information and Communication Technology, Geographic Information Science at the University of Phayao for providing access to the use of server computer and this research was funded by the program in Geographic Information Science at the University of Phayao.

REFERENCES

- [1] B. Abrahms *et al.*, "Climate change as a global amplifier of human–wildlife conflict," *Nat. Clim.*

- Chang.*, vol. 13, no. 3, pp. 224–234, Mar. 2023, doi: 10.1038/s41558-023-01608-5.
- [2] T. M. Lenton *et al.*, “Quantifying the human cost of global warming,” *Nat Sustain*, vol. 6, no. 10, pp. 1237–1247, Oct. 2023, doi: 10.1038/s41893-023-01132-6.
- [3] Y.-J. Zhang, Z. Liu, H. Zhang, and T.-D. Tan, “The impact of economic growth, industrial structure and urbanization on carbon emission intensity in China,” *Natural Hazards*, vol. 73, no. 2, pp. 579–595, 2014, doi: 10.1007/s11069-014-1091-x.
- [4] P. Friedlingstein *et al.*, “Global Carbon Budget 2022,” *Earth System Science Data*, vol. 14, no. 11, pp. 4811–4900, Nov. 2022, doi: 10.5194/essd-14-4811-2022.
- [5] M. Zhang and Y. Liu, “Influence of digital finance and green technology innovation on China’s carbon emission efficiency: Empirical analysis based on spatial metrology,” *Science of The Total Environment*, vol. 838, p. 156463, Sep. 2022, doi: 10.1016/j.scitotenv.2022.156463.
- [6] Y. Wang, L. Li, J. Kubota, R. Han, X. Zhu, and G. Lu, “Does urbanization lead to more carbon emission? Evidence from a panel of BRICS countries,” *Applied Energy*, vol. 168, pp. 375–380, Apr. 2016, doi: 10.1016/j.apenergy.2016.01.105.
- [7] Q. Wang, L. Li, and R. Li, “Uncovering the impact of income inequality and population aging on carbon emission efficiency: An empirical analysis of 139 countries,” *Science of The Total Environment*, vol. 857, p. 159508, Jan. 2023, doi: 10.1016/j.scitotenv.2022.159508.
- [8] Md. A. Fattah, S. R. Morshed, and S. Y. Morshed, “Impacts of land use-based carbon emission pattern on surface temperature dynamics: Experience from the urban and suburban areas of Khulna, Bangladesh,” *Remote Sensing Applications: Society and Environment*, vol. 22, p. 100508, Apr. 2021, doi: 10.1016/j.rsase.2021.100508.
- [9] R. Nassar *et al.*, “Advances in quantifying power plant CO₂ emissions with OCO-2,” *Remote Sensing of Environment*, vol. 264, p. 112579, Oct. 2021, doi: 10.1016/j.rse.2021.112579.
- [10] F. Raymand, P. Ahmadi, and S. Mashayekhi, “Evaluating a light duty vehicle fleet against climate change mitigation targets under different scenarios up to 2050 on a national level,” *Energy Policy*, vol. 149, p. 111942, Feb. 2021, doi: 10.1016/j.enpol.2020.111942.
- [11] Y.-F. Sun, Y.-J. Zhang, and B. Su, “How does global transport sector improve the emissions reduction performance? A demand-side analysis,” *Applied Energy*, vol. 311, p. 118648, Apr. 2022, doi: 10.1016/j.apenergy.2022.118648.
- [12] T. Mammo, “Crustal Structure of the Flood Basalt Province of Ethiopia from Constrained 3-D Gravity Inversion,” *Pure and Applied Geophysics*, vol. 170, no. 12, pp. 2185–2206, Dec. 2013, doi: 10.1007/s00024-013-0663-0.
- [13] B. Kebede, A. Alemu, H. Kebede, E. Melaku, D. Feyissa, and G. Legesse, “Earth & Environmental Science Research & Reviews,” 2023.
- [14] K. Mickus, K. Tadesse, G. R. Keller, and B. Oluma, “Gravity analysis of the main Ethiopian rift,” *Journal of African Earth Sciences*, vol. 48, no. 2, pp. 59–69, Jun. 2007, doi: 10.1016/j.jafrearsci.2007.02.008.
- [15] A. Masria, “Bridging coastal challenges: The role of remote sensing and future research,” *Regional Studies in Marine Science*, vol. 73, p. 103502, Jul. 2024, doi: 10.1016/j.rsma.2024.103502.
- [16] V. S. Manivasagam, “Remote sensing of irrigation: Research trends and the direction to next-generation agriculture through data-driven scientometric analysis,” *Water Security*, vol. 21, p. 100161, Apr. 2024, doi: 10.1016/j.wasec.2023.100161.
- [17] F. F. da Costa, I. A. A. Rufino, R. de Aragão, and R. da S. R. Filho, “Performance evaluation of four remote-sensing products throughout precipitation estimation in the State of Paraíba, Northeast Brazil,” *Remote Sensing Applications: Society and Environment*, vol. 35, p. 101256, Aug. 2024, doi: 10.1016/j.rsase.2024.101256.
- [18] M. Aqil, I. Kita, A. Yano, and N. Soichi, “Decision Support System for Flood Crisis Management using Artificial Neural Network,” Jan. 2006.
- [19] Y. Le Cun *et al.*, “Handwritten digit recognition with a back-propagation network,” in *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, in NIPS’89. Cambridge, MA, USA: MIT Press, 1989, pp. 396–404.
- [20] L. Li, T. Xu, and Y. Chen, “Improved Urban Flooding Mapping from Remote Sensing Images Using Generalized Regression Neural Network-Based Super-Resolution Algorithm,” *Remote Sensing*, vol. 8, no. 8, 2016, doi: 10.3390/rs8080625.
- [21] S. Yang, L. Lei, Z. Zeng, Z. He, and H. Zhong, “An Assessment of Anthropogenic CO₂ Emissions by Satellite-Based Observations in China,” *Sensors*, vol. 19, no. 5, 2019, doi: 10.3390/s19051118.
- [22] P. M. Atkinson, “Sub-pixel Target Mapping from Soft-classified, Remotely Sensed Imagery,” *Photogrammetric Engineering and Remote Sensing*, vol. 71, pp. 839–846, 2005.
- [23] M. K. -P. Ng, Q. Yuan, L. Yan, and J. Sun, “An Adaptive Weighted Tensor Completion Method for the Recovery of Remote Sensing Images With Missing Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3367–3381, Jun. 2017, doi: 10.1109/TGRS.2017.2670021.
- [24] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, “Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018, doi: 10.1109/TGRS.2018.2810208.
- [25] tdri, “Transforming Thailand into a Low-Carbon Economy and Society,” TDRI: Thailand Development Research Institute. Accessed: Nov. 06, 2024. [Online]. Available: <https://tdri.or.th/en/2023/10/transforming-thailand-into-a-low-carbon-economy-and-society/>
- [26] L. Zappa, S. Schlaffer, B. Bauer-Marschallinger, C. Nendel, B. Zimmerman, and W. Dorigo, “Detection and Quantification of Irrigation Water Amounts at 500 m Using Sentinel-1 Surface Soil Moisture,” *Remote Sensing*, vol. 13, no. 9, 2021, doi: 10.3390/rs13091727.

- [27] J. L. Peña-Arancibia, M. Mainuddin, J. M. Kirby, F. H. S. Chiew, T. R. McVicar, and J. Vaze, "Assessing irrigated agriculture's surface water and groundwater consumption by combining satellite remote sensing and hydrologic modelling," *Science of The Total Environment*, vol. 542, pp. 372–382, Jan. 2016, doi: 10.1016/j.scitotenv.2015.10.086.
- [28] M. Romaguera, M. Krol, Mhd. Salama, Z. Su, and A. Hoekstra, "Application of a Remote Sensing Method for Estimating Monthly Blue Water Evapotranspiration in Irrigated Agriculture," *Remote Sensing*, vol. 6, no. 10, pp. 10033–10050, Oct. 2014, doi: 10.3390/rs61010033.
- [29] E. Jalilvand, M. Tajrishy, S. A. Ghazi Zadeh Hashemi, and L. Brocca, "Quantification of irrigation water using remote sensing of soil moisture in a semi-arid region," *Remote Sensing of Environment*, vol. 231, p. 111226, Sep. 2019, doi: 10.1016/j.rse.2019.111226.
- [30] Y. Chen *et al.*, "Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, Landsat imagery, and ancillary data," *Remote Sensing of Environment*, vol. 204, pp. 197–211, Jan. 2018, doi: 10.1016/j.rse.2017.10.030.
- [31] G. Kaplan *et al.*, "Estimating Processing Tomato Water Consumption, Leaf Area Index, and Height Using Sentinel-2 and VENµS Imagery," *Remote Sensing*, vol. 13, no. 6, 2021, doi: 10.3390/rs13061046.
- [32] J. M. Deines, A. D. Kendall, M. A. Crowley, J. Rapp, J. A. Cardille, and D. W. Hyndman, "Mapping three decades of annual irrigation across the US High Plains Aquifer using Landsat and Google Earth Engine," *Remote Sensing of Environment*, vol. 233, p. 111400, Nov. 2019, doi: 10.1016/j.rse.2019.111400.
- [33] G. Kaplan, L. Fine, V. Lukyanov, V. S. Manivasagam, J. Tanny, and O. Rozenstein, "Normalizing the Local Incidence Angle in Sentinel-1 Imagery to Improve Leaf Area Index, Vegetation Height, and Crop Coefficient Estimations," *Land*, vol. 10, no. 7, 2021, doi: 10.3390/land10070680.
- [34] D. Deligiorgi and K. Philippopoulos, "Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece," 2011. doi: 10.5772/17734.
- [35] L. Li, X. Zhou, M. Kalo, and R. Piltner, "Spatiotemporal Interpolation Methods for the Application of Estimating Population Exposure to Fine Particulate Matter in the Contiguous U.S. and a Real-Time Web Application," *International Journal of Environmental Research and Public Health*, vol. 13, no. 8, 2016, doi: 10.3390/ijerph13080749.
- [36] "Characterizing Spatial Processes: The Covariance and Variogram," in *Geostatistics for Environmental Scientists*, 2007, pp. 47–76. doi: 10.1002/9780470517277.ch4.
- [37] D. Were, F. Kansiime, T. Fetahi, A. Cooper, and C. Jjuuko, "Carbon Sequestration by Wetlands: A Critical Review of Enhancement Measures for Climate Change Mitigation," *Earth Systems and Environment*, vol. 3, no. 2, pp. 327–340, Aug. 2019, doi: 10.1007/s41748-019-00094-0.
- [38] H. Hou, S. Zhang, Z. Ding, Y. Wang, Y. Yang, and S. Guo, "Temporal variation of near-surface CO₂ concentrations over different land uses in Suzhou City," *Environmental Earth Sciences*, vol. 75, no. 16, p. 1197, Aug. 2016, doi: 10.1007/s12665-016-5988-4.
- [39] K. R. Verhulst *et al.*, "Carbon dioxide and methane measurements from the Los Angeles Megacity Carbon Project – Part 1: calibration, urban enhancements, and uncertainty estimates," *Atmos. Chem. Phys.*, vol. 17, no. 13, pp. 8313–8341, Jul. 2017, doi: 10.5194/acp-17-8313-2017.
- [40] L. E. Mitchell *et al.*, "Long-term urban carbon dioxide observations reveal spatial and temporal dynamics related to urban characteristics and growth," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, no. 12, pp. 2912–2917, Mar. 2018, doi: 10.1073/pnas.1702393115.
- [41] M. Krishnapriya, A. B. Chandra, R. K. Nayak, N. R. Patel, P. V. N. Rao, and V. K. Dadhwal, "Seasonal and inter-annual variability of atmosphere CO₂ based on NOAA Carbon Tracker analysis and satellite observations," *J Indian Soc Remote Sens.*, vol. 46, no. 2, pp. 309–320, Feb. 2018, doi: 10.1007/s12524-017-0688-4.
- [42] D. De Bortoli Teixeira, A. R. Panosso, C. E. P. Cerri, G. T. Pereira, and N. La Scala, "Soil CO₂ emission estimated by different interpolation techniques," *Plant and Soil*, vol. 345, no. 1, pp. 187–194, Aug. 2011, doi: 10.1007/s11104-011-0770-6.
- [43] K. Calvin *et al.*, "IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland," Intergovernmental Panel on Climate Change (IPCC), Jul. 2023. doi: 10.59327/IPCC/AR6-9789291691647.
- [44] T. Yue, L. Zhang, M. Zhao, Y. Wang, and J. Wilson, "Space- and ground-based CO₂ measurements: A review," *Science China Earth Sciences*, vol. 59, no. 11, pp. 2089–2097, Nov. 2016, doi: 10.1007/s11430-015-0239-7.
- [45] F. Chevallier *et al.*, "On the accuracy of the CO₂ surface fluxes to be estimated from the GOSAT observations," *Geophysical Research Letters*, vol. 36, no. 19, Oct. 2009, doi: 10.1029/2009GL040108.
- [46] F. Chevallier, F.-M. Bréon, and P. J. Rayner, "Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks: Theoretical study in a variational data assimilation framework," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D9, May 2007, doi: 10.1029/2006JD007375.
- [47] H. Bovensmann *et al.*, "SCIAMACHY: Mission Objectives and Measurement Modes," *J. Atmos. Sci.*, vol. 56, no. 2, pp. 127–150, Jan. 1999, doi: 10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.
- [48] Y. Yoshida *et al.*, "Retrieval algorithm for CO₂ and CH₄ column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite," *Atmos. Meas. Tech.*, vol. 4, no. 4, pp. 717–734, Apr. 2011, doi: 10.5194/amt-4-717-2011.

- [49] H. Boesch, D. Baker, B. Connor, D. Crisp, and C. Miller, "Global Characterization of CO₂ Column Retrievals from Shortwave-Infrared Satellite Observations of the Orbiting Carbon Observatory-2 Mission," *Remote Sensing*, vol. 3, no. 2, pp. 270–304, 2011, doi: 10.3390/rs3020270.
- [50] D. Wunch *et al.*, "The Total Carbon Column Observing Network," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1943, pp. 2087–2112, May 2011, doi: 10.1098/rsta.2010.0240.
- [51] G. Meng *et al.*, "The status and development proposal of carbon sources and sinks monitoring satellite system," *Carbon Neutrality*, vol. 1, no. 1, p. 32, Sep. 2022, doi: 10.1007/s43979-022-00033-5.
- [52] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding*, vol. Volume 2. in Series on Atmospheric, Oceanic and Planetary Physics, no. Volume 2, vol. Volume 2. WORLD SCIENTIFIC, 2000. doi: 10.1142/3171.
- [53] Y. Bi, Q. Wang, Z. Yang, J. Chen, and W. Bai, "Validation of Column-Averaged Dry-Air Mole Fraction of CO₂ Retrieved from OCO-2 Using Ground-Based FTS Measurements," *Journal of Meteorological Research*, vol. 32, no. 3, pp. 433–443, Jun. 2018, doi: 10.1007/s13351-018-7118-6.
- [54] C. W. O'Dell *et al.*, "The ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic observations," *Atmos. Meas. Tech.*, vol. 5, no. 1, pp. 99–121, Jan. 2012, doi: 10.5194/amt-5-99-2012.
- [55] D. Crisp *et al.*, "The ACOS CO₂ retrieval algorithm – Part II: Global XCO₂ data characterization," *Atmos. Meas. Tech.*, vol. 5, no. 4, pp. 687–707, Apr. 2012, doi: 10.5194/amt-5-687-2012.
- [56] Y. Yoshida *et al.*, "Improvement of the retrieval algorithm for GOSAT SWIR XCO₂ and XCH₄ and their validation using TCCON data," *Atmos. Meas. Tech.*, vol. 6, no. 6, pp. 1533–1547, Jun. 2013, doi: 10.5194/amt-6-1533-2013.
- [57] C. Das, R. K. Kunchala, N. Chandra, A. Chhabra, and M. R. Pandya, "Characterizing the regional XCO₂ variability and its association with ENSO over India inferred from GOSAT and OCO-2 satellite observations," *Science of The Total Environment*, vol. 902, p. 166176, Dec. 2023, doi: 10.1016/j.scitotenv.2023.166176.
- [58] D. Wunch *et al.*, "Calibration of the Total Carbon Column Observing Network using aircraft profile data," *Atmos. Meas. Tech.*, vol. 3, no. 5, pp. 1351–1362, Oct. 2010, doi: 10.5194/amt-3-1351-2010.
- [59] "TCCON - Total Carbon Column Observing Network." Accessed: Jul. 29, 2024. [Online]. Available: <http://www.tccon.caltech.edu/index.html>
- [60] J. Li *et al.*, "High-spatiotemporal resolution mapping of spatiotemporally continuous atmospheric CO₂ concentrations over the global continent," *International Journal of Applied Earth Observation and Geoinformation*, vol. 108, p. 102743, Apr. 2022, doi: 10.1016/j.jag.2022.102743.
- [61] B. J. Connor, H. Boesch, G. Toon, B. Sen, C. Miller, and D. Crisp, "Orbiting Carbon Observatory: Inverse method and prospective error analysis," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D5, Mar. 2008, doi: 10.1029/2006JD008336.
- [62] Y. Jing, T. Wang, P. Zhang, L. Chen, N. Xu, and Y. Ma, "Global Atmospheric CO₂ Concentrations Simulated by GEOS-Chem: Comparison with GOSAT, Carbon Tracker and Ground-Based Measurements," *Atmosphere*, vol. 9, no. 5, 2018, doi: 10.3390/atmos9050175.
- [63] Official statistics registration systems, "สถิติประชากรทางการเมืองรายจังหวัด." Accessed: Aug. 03, 2024. [Online]. Available: <https://stat.bora.dopa.go.th/stat/statnew/statMONTH/stamonth/#/mainpage>
- [64] B. Wu, H. Huang, Y. Wang, S. Shi, J. Wu, and B. Yu, "Global spatial patterns between nighttime light intensity and urban building morphology," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, p. 103495, Nov. 2023, doi: 10.1016/j.jag.2023.103495.
- [65] K. Shi, Z. Chang, Z. Chen, J. Wu, and B. Yu, "Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China," *Journal of Cleaner Production*, vol. 255, p. 120245, May 2020, doi: 10.1016/j.jclepro.2020.120245.
- [66] M. Johnson, V. Isakov, J. S. Touma, S. Mukerjee, and H. Özkaynak, "Evaluation of land-use regression models used to predict air quality concentrations in an urban area," *Atmospheric Environment*, vol. 44, no. 30, pp. 3660–3668, Sep. 2010, doi: 10.1016/j.atmosenv.2010.06.041.
- [67] J. E. Clougherty, E. A. Houseman, and J. I. Levy, "Source apportionment of indoor residential fine particulate matter using land use regression and constrained factor analysis," *Indoor Air*, vol. 21, no. 1, pp. 53–66, Feb. 2011, doi: 10.1111/j.1600-0668.2010.00682.x.
- [68] M. Brauer, C. Lencar, L. Tamburic, M. Koehoorn, P. Demers, and C. Karr, "A cohort study of traffic-related air pollution impacts on birth outcomes," *Environ Health Perspect*, vol. 116, no. 5, pp. 680–686, May 2008, doi: 10.1289/ehp.10952.
- [69] S. Henderson, B. Beckerman, M. Jerrett, and M. Brauer, "Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter," *Environmental science & technology*, vol. 41, pp. 2422–8, May 2007, doi: 10.1021/es0606780.
- [70] A. Hessl, J. Miller, J. Kernan, D. Keenum, and D. McKenzie, "Mapping Paleo-Fire Boundaries from Binary Point Data: Comparing Interpolation Methods," *The Professional Geographer*, vol. 59, no. 1, pp. 87–104, Feb. 2007, doi: 10.1111/j.1467-9272.2007.00593.x.
- [71] M. Tomczak, "Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach," *Journal of Geographic Information and Decision Analysis*, vol. 2, pp. 18–30, Jan. 1998.
- [72] J. Hofierka, J. Parajka, H. Mitasova, and L. Mitas, "Multivariate Interpolation of Precipitation Using

- Regularized Spline with Tension," *T. GIS*, vol. 6, pp. 135–150, Mar. 2002, doi: 10.1111/1467-9671.00101.
- [73] Stanislaw Grundas and Andrzej Stepniewski, Eds., "Advances in Agrophysical Research," Jul. 2013, doi: 10.5772/3341.
- [74] L. R. Hutrya *et al.*, "Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective," *Earth's Future*, vol. 2, no. 10, pp. 473–495, Oct. 2014, doi: 10.1002/2014EF000255.
- [75] M. A. OLIVER and R. WEBSTER, "Kriging: a method of interpolation for geographical information systems," *International journal of geographical information systems*, vol. 4, no. 3, pp. 313–332, Jul. 1990, doi: 10.1080/02693799008941549.
- [76] R. Woodard, "Interpolation of Spatial Data: Some Theory for Kriging," *Technometrics*, vol. 42, no. 4, pp. 436–437, Nov. 2000, doi: 10.1080/00401706.2000.10485731.
- [77] J. Mulholland, A. Butler, J. Wilkinson, A. Russell, and P. Tolbert, "Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications," *J Air Waste Manag Assoc*, vol. 48, no. 5, pp. 418–426, May 1998, doi: 10.1080/10473289.1998.10463695.
- [78] F. Ferreira, H. Tente, P. Torres, S. Cardoso, and J. M. Palma-Oliveira, "Air Quality Monitoring and Management in Lisbon," *Environmental Monitoring and Assessment*, vol. 65, no. 1, pp. 443–450, Nov. 2000, doi: 10.1023/A:1006433313316.
- [79] S. Janssen, G. Dumont, F. Fierens, and C. Mensink, "Spatial interpolation of air pollution measurements using CORINE land cover data," *Atmospheric Environment*, vol. 42, no. 20, pp. 4884–4903, Jun. 2008, doi: 10.1016/j.atmosenv.2008.02.043.
- [80] G. S. Bhunia, P. K. Shit, and R. Maiti, "Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)," *Journal of the Saudi Society of Agricultural Sciences*, vol. 17, no. 2, pp. 114–126, Apr. 2018, doi: 10.1016/j.jssas.2016.02.001.
- [81] H. Ledoux and C. Gold, "An Efficient Natural Neighbour Interpolation Algorithm for Geoscientific Modelling," in *Developments in Spatial Data Handling*, P. F. Fisher, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 97–108.
- [82] T. Etherington, "Discrete natural neighbour interpolation with uncertainty using cross-validation error-distance fields," *PeerJ Computer Science*, vol. 6, p. e282, Jul. 2020, doi: 10.7717/peerj.cs.282.
- [83] R. Sibson, "A brief description of natural neighbor interpolation," 1981. Accessed: Sep. 05, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/A-brief-description-of-natural-neighbor-Sibson/1ca6926bf459db0a4d89a8c88b9fd64c32760bc2>
- [84] R. R. Rodriguez, "A Case Study of the Klamath River Estuary".
- [85] X. Ruan, M. Guo, and Z. Zhan, "A regional digital bathymetric model fusion method based on topographic slope: A case study of the south China sea and surrounding waters," *Heliyon*, vol. 10, no. 4, p. e26644, Feb. 2024, doi: 10.1016/j.heliyon.2024.e26644.
- [86] R. J. Peckham and G. Jordan, "Digital terrain modelling: Development and applications in a policy support environment," *Lecture Notes in Geoinformation and Cartography*, Jan. 2007.
- [87] W. Luo, M. C. Taylor, and S. R. Parker, "A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales," *International Journal of Climatology*, vol. 28, no. 7, pp. 947–959, Jun. 2008, doi: 10.1002/joc.1583.
- [88] B. A., P. Fernández, and O. Duval, "Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves," *Soil Use and Management*, vol. 19, pp. 232–242, Jan. 2006, doi: 10.1111/j.1475-2743.2003.tb00309.x.
- [89] C. D. Rodgers and B. J. Connor, "Intercomparison of remote sounding instruments," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D3, Feb. 2003, doi: 10.1029/2002JD002299.
- [90] A. Liang, W. Gong, G. Han, and C. Xiang, "Comparison of Satellite-Observed XCO₂ from GOSAT, OCO-2, and Ground-Based TCCON," *Remote Sensing*, vol. 9, no. 10, 2017, doi: 10.3390/rs9101033.
- [91] F. Chevallier, R. Engelen, and P. Peylin, "The contribution of AIRS data to the estimation of CO₂ sources and sinks," *Geophysical Research Letters - GEOPHYS RES LETT*, vol. 322, Dec. 2005, doi: 10.1029/2005GL024229.
- [92] M. Zhou *et al.*, "Validation of TANSO-FTS/GOSAT XCO₂ and XCH₄ glint mode retrievals using TCCON data from near-ocean sites," *Atmos. Meas. Tech.*, vol. 9, no. 3, pp. 1415–1430, Apr. 2016, doi: 10.5194/amt-9-1415-2016.
- [93] M. Inoue *et al.*, "Validation of XCO₂ derived from SWIR spectra of GOSAT TANSO-FTS with aircraft measurement data," *Atmos. Chem. Phys.*, vol. 13, no. 19, pp. 9771–9788, Oct. 2013, doi: 10.5194/acp-13-9771-2013.
- [94] A. J. Cogan *et al.*, "Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D21, Nov. 2012, doi: 10.1029/2012JD018087.
- [95] D. Wunch *et al.*, "A method for evaluating bias in global measurements of CO₂ total columns from space," *Atmos. Chem. Phys.*, vol. 11, no. 23, pp. 12317–12337, Dec. 2011, doi: 10.5194/acp-11-12317-2011.
- [96] C.-Y. Wu, J. Mossa, L. Mao, and M. Almulla, "Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River," *Annals of GIS*, vol. 25, pp. 1–19, Mar. 2019, doi: 10.1080/19475683.2019.1588781.
- [97] L. Anselin and J. Le Gallo, "Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects," *Spatial Economic Analysis*, vol. 1, pp. 31–52, Jun. 2006, doi: 10.1080/17421770600661337.
- [98] J. Knijff, J. Younis, and A. Roo, "LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation," *International Journal of*

- Geographical Information Science*, vol. 24, pp. 189–212, Feb. 2010, doi: 10.1080/13658810802549154.
- [99] Y. Meng, M. Cave, and C. Zhang, “Comparison of methods for addressing the point-to-area data transformation to make data suitable for environmental, health and socio-economic studies,” *Science of The Total Environment*, vol. 689, pp. 797–807, Nov. 2019, doi: 10.1016/j.scitotenv.2019.06.452.
- [100] Z. Li *et al.*, “Exploring modern bathymetry: A comprehensive review of data acquisition devices, model accuracy, and interpolation techniques for enhanced underwater mapping,” *Frontiers in Marine Science*, vol. 10, May 2023, doi: 10.3389/fmars.2023.1178845.
- [101] C. J. Okolie and J. L. Smit, “A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 1–29, Jun. 2022, doi: 10.1016/j.isprsjprs.2022.03.016.
- [102] M. Arseni, M. Voiculescu, L. P. Georgescu, C. Itecescu, and A. Rosu, “Testing Different Interpolation Methods Based on Single Beam Echosounder River Surveying. Case Study: Siret River,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 11, 2019, doi: 10.3390/ijgi8110507.
- [103] I. Maduako, E. Ebinne, U. Idorenyin, and R. Ndukwu, “Accuracy Assessment and Comparative Analysis of IDW, Spline and Kriging in Spatial Interpolation of Landform (Topography): An Experimental Study,” *Journal of Geographic Information System*, vol. 09, pp. 354–371, Jan. 2017, doi: 10.4236/jgis.2017.93022.
- .

Research on developing an automated tool for evaluating spatial data quality according to the national geographic data standard technical model

Yen Phan Quoc *, Nga Nguyen Thi Thu, Hanh Tong Thi

Le Quy Don Technical University, Hanoi, Vietnam

Abstract - Currently, the inspection and control of spatial data quality according to national geographic data standards are often performed manually, which is typically time-consuming and costly. This process requires inspectors to have deep expertise in the fields of GIS and geographic database foundations. The inspection results can be influenced by various factors such as the inspector's expertise level, with a high possibility of missing data and subjective errors. Therefore, this study focuses on the application of GIS programming in developing an automated tool to evaluate the quality of spatial data for the Population group according to the structural model of national geographic data standard technical specifications. The tool is developed using the C# language and ArcObjects library, operating as an Add-In for ArcMap to perform the 7 steps of the spatial data quality assessment process. Testing the tool with the Population group of geographic data in Thanh Hoa province according to QCVN 73:2023/BTNMT, the results indicate errors in area-shaped houses, point-shaped houses, and house blocks overlapping with road objects, and some data acquisition criteria below permissible levels in terms of area, length, and height. The software accurately and quickly detects the exact number of errors.

Keywords - *GIS, spatial data quality, Add-Ins, ArcObjects, Geodatabase*

INTRODUCTION

Currently, the explosion of geographic information systems (GIS) and artificial intelligence has created a turning point in spatial support and decision-making systems. GIS data plays a crucial role. If the data is not standardized and contains errors, spatial analysis algorithms will produce incorrect results. Therefore, computer-assisted design and adherence to standards are vital in data preparation (Rasheed et al., 2015). Geographic spatial data is defined as data on the geographic location of geographic objects, and geographic base data is the foundational spatial data used to build other geographic spatial data (National Assembly, 2018). The national geographic database (NGDB) is a collection of geographic base data built in the national coordinate and elevation systems according to national standards, technical regulations, and specifications to be used uniformly nationwide (Government, 2019).

Based on inheriting and developing international standards, Vietnam has issued technical regulations on the structural model and content of the NGDB at different scale groups, such as the 2,000-10,000, 10,000-25,000, and 50,000-100,000 scale groups. Each technical regulation includes an annex specifying the quality of spatial data, known as the Annex on the Quality of the National Geographic Database (Ministry of Natural Resources and Environment, 2018, 2019, 2020a, 2020b, 2022, 2023). For example, Annex C of QCVN 73:2023/BTNMT, Annex 4 for the 1:10,000, 1:25,000 scale group. According to these annexes, five groups of data quality factors are mentioned:

Data completeness, Data conformity with the structural data model, Positional accuracy of geographic objects, Temporal accuracy of geographic objects, and Thematic attribute accuracy (Ministry of Natural Resources and Environment, 2020a, 2020b, 2022, 2023). All these factors are important for an NGDB, influencing the final outcome of spatial decision-support problems. However, the "Compliance with spatial relationships" factor within the "Data conformity with the structural data model" group has the most impact on spatial analysis.

According to regulations in the standards, NGDB quality can be checked manually or automatically using software (Ministry of Natural Resources and Environment, 2020a, 2020b, 2022, 2023). Manual inspection can review and compare datasets against standards visually, supported by specialized mapping software. Checking data quality with hundreds of data layers and dozens of quality parameters is time-consuming and costly. Additionally, it requires the inspector to have deep expertise in GIS, with potential discrepancies and data omissions due to human subjective factors (Quoc Yen Phan et al., 2023). Currently, there are several specialized tools for checking data quality such as ArcGIS Data Reviewer, SQUAD for QGIS, Check for Topological Errors, Check Validity, DSG Tool (Kukulska et al., 2018; Sehra et al., 2017; Li et al., 2020; Hendricks et al., 2022). However, these tools require deep expertise in GIS and geographic base data. Building hundreds of rules for spatial relationships to conform to Vietnamese standards is challenging. Particularly, integrating reference sample datasets with these software tools is difficult.

There is a need to develop software where rules related to spatial data are pre-integrated according to Vietnamese standards. Currently, NGDB construction is mainly done in the geodatabase (*.gdb) format of the ArcGIS software suite. It is an advanced software suite capable of expanding to create, manage, integrate, analyze, and present geographic data, including a suite of integrated applications: ArcMap, ArcCatalog, and ArcToolbox (Booth and Mitchell, 2021). ArcGIS is built on a technological framework called ArcObjects SDK (Eldrandaly, 2013; Li and Li, 2020; Amirian, 2013). Using ArcObjects to build auxiliary tools for the ArcGIS software suite, especially ArcMap, to support map data editing and standardization is very common. Therefore, this study applies the ArcObjects library to develop a tool that functions as a plugin for ArcMap to identify spatial relationship errors according to the NGDB structural model at 1:2,000, 1:5,000, 1:10,000 scales according to QCVN 73:2023/BTNMT.

RESEARCH METHODOLOGY

Spatial Data Quality Criteria According to Technical Regulations

Group 2 involves the relationships between objects in different layers, such as overlay errors, gaps between surfaces, point-to-line relationship errors (e.g., a power station must be at the end of a power line), line-to-line relationship errors (e.g., a bridge must align with the centerline of a road), arc-to-surface boundary errors, and line-to-region relationship errors (Fig. 1).

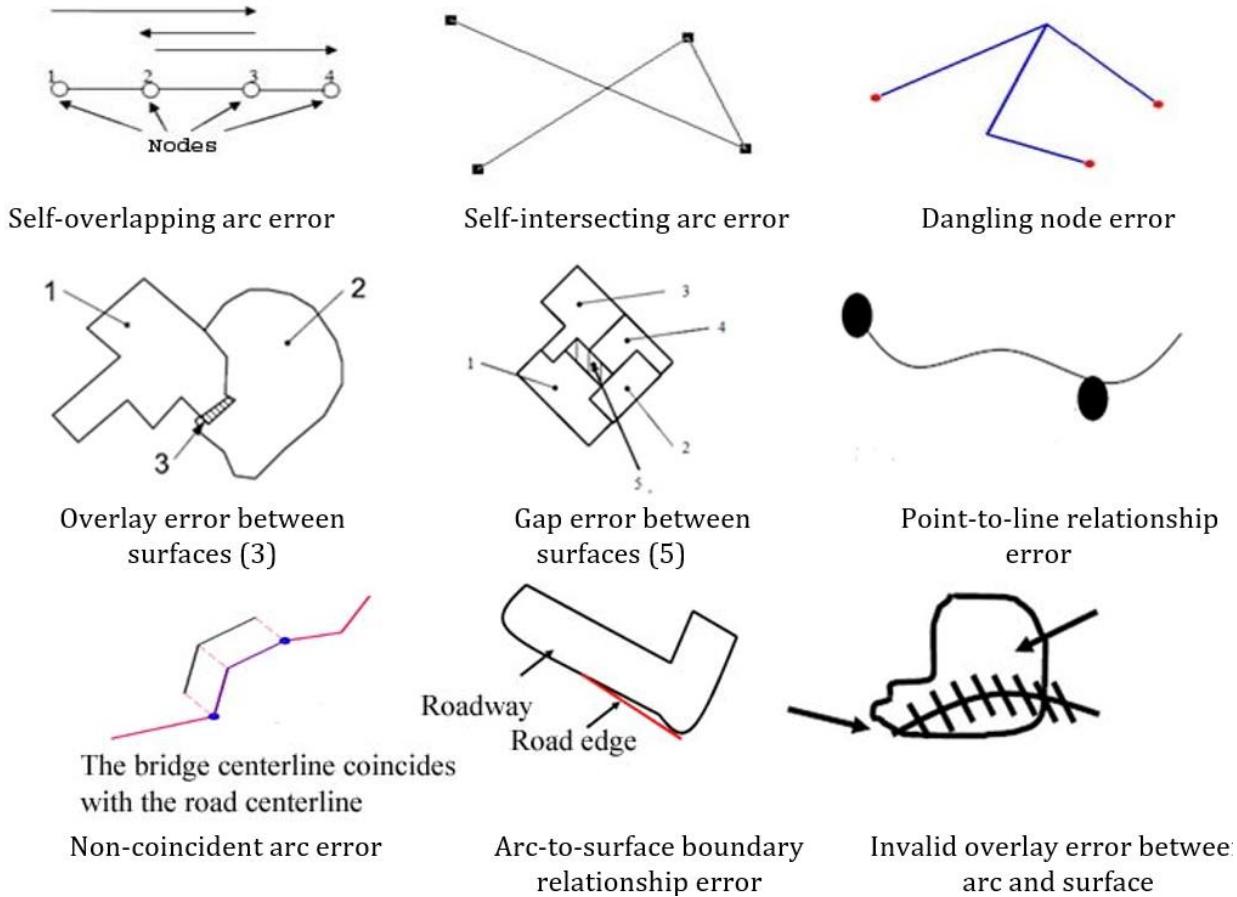


Figure 1. Spatial relationship errors according to QCVN 73:2023/BTNMT [9].

To check the errors mentioned above, it is necessary to develop specific spatial data validation rules for all layers in the geodatabase. The rules in Group 1 are simple rules for each individual layer (line, point, area) and are commonly referred to as general rules. The rules in Group 2 are known as specific rules and are defined based on the prescribed data structure model.

Analyze the application schema describing the residential package to identify spatial relationship rules between object layers (Fig. 3). Some basic rules for spatial relationships between object layers in the residential group include: "Urban residential areas" must be within "Ward, town boundaries"; "Rural residential areas" must be within "Commune boundaries"; "Building S"

must not overlap or coincide with "Traffic bridge S"; "Building S" must not overlap or coincide with "Traffic tunnel S"; "Building S" must not overlap or coincide with "Car passable ford S"; "Building S" must not overlap or coincide with "Roadway surface objects S"; "Building P" must not be located on "Traffic bridge S"; "Building P" must not be located on "Traffic tunnel S"; "Building P" must not be located on "Car passable ford S"; "Building P" must not be located on "Roadway surface objects S"; "Building block S" must not overlap or coincide with "Traffic S"; "Transformer station P" must be at the endpoint of "Power line C"; "Electric pole P" must be on "Power line C". Here, the symbols P, C, and S respectively represent point, line, and area geographic objects.

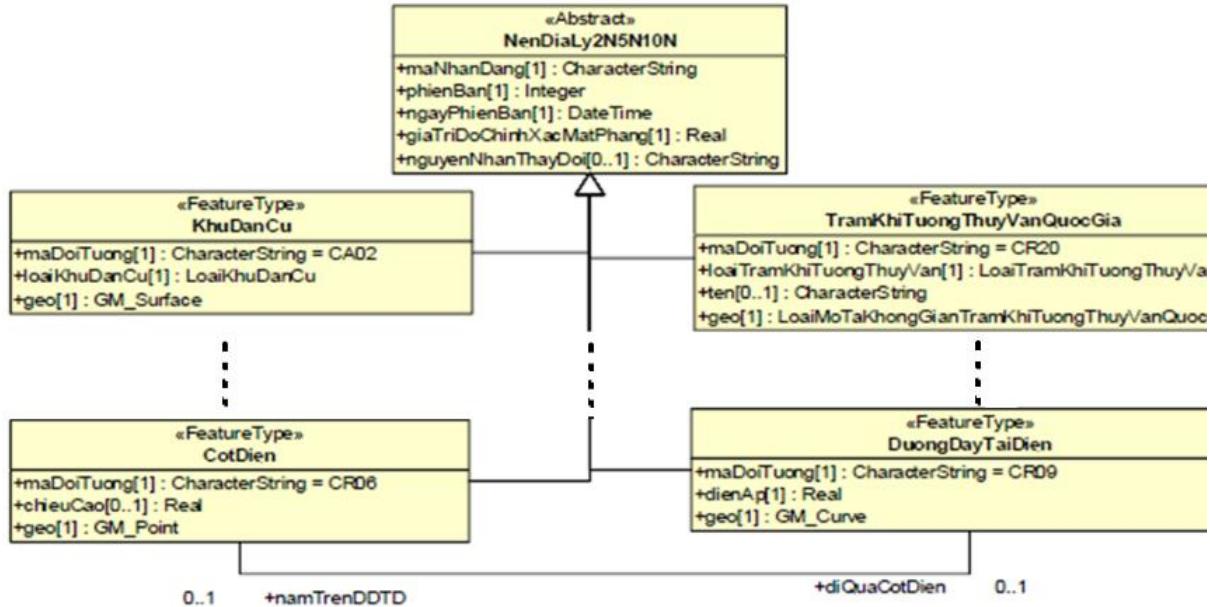


Figure 2. Excerpt from the application schema describing geographic object layers in the residential data package (9. Ministry of Natural Resources and Environment, 2023)

GIS Programming Technology in Software Development for Quality Control

Data Quality Assessment Process

According to the QCVN 42:2020/BTNMT standard, the geographic base data quality assessment process involves 5 steps (left panel of Fig. 3) (Ministry of Natural Resources and Environment, 2020). These 5 steps comprehensively cover the overall data quality assessment package. Steps 1 and 2 are performed based on the requirements, data content characteristics, and established standards. In Step 3, the method for data quality assessment is

identified and applied. Manual visual inspection for spatial relationships is very difficult, time-consuming, and costly. Utilizing the power of GIS techniques for automated identification will yield high efficiency, especially with pre-built rules in the software, and only the first 5 steps (right panel of Fig. 3) need to be executed for the software to automatically detect errors based on the selected rules. Step 4 in the regulations involves analyzing the inspection results, particularly the results of erroneous objects and writing reports, corresponding to steps 6 and 7 in the software. This will significantly support the entire data quality assessment process.

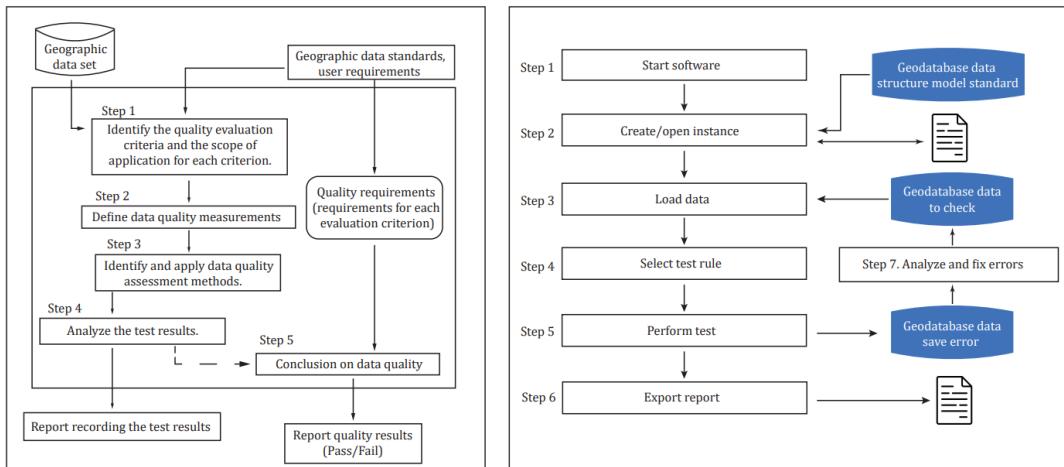


Figure 3. Geographic base data quality inspection process

GIS Programming Technology in Development of Quality Inspection Software

Currently, there are seven operators in GIS commonly used to determine spatial relationships between objects and between geographic layers (Table 1). The Equals operator indicates that the interiors and geometries of both objects are exactly the same. Intersects involves two objects having a relationship where they contain, are contained, cross, touch, or overlap each other. Contains means

one object is a subset of the other, with their interiors intersecting. Within is the inverse of Contains. Crosses is a subset of Intersects, specifically for relationships between lines and areas. Touches describes two objects that intersect at vertices but do not overlap. Overlaps applies to two geometrically similar objects whose intersection is of the same geometry type but differs from both original geometries (ESRI, 2023).

Table 1. Spatial operators supporting relationships between two geometric objects (ESRI, 2023)

	Equals	Intersects	Contains	Within	Crosses	Touches	Overlaps
P-P	✓	✓	✓	✓			
P-C		✓		✓		✓	
P-S		✓		✓			
C-P		✓	✓			✓	
C-C	✓	✓	✓	✓	✓	✓	✓
C-S		✓		✓	✓	✓	
S-P	✓		✓			✓	
S-C		✓	✓		✓	✓	
S-S	✓	✓	✓	✓		✓	✓

Due to the predominance of geographic base data stored in the *.gdb format of ArcGIS, the ArcGIS software suite is widely used for standardizing and editing national geographic base data. Within the ArcMap data editing software, users can incorporate tools to support various functions. Therefore, Add-ins tools in ArcMap are implemented to automate the verification of model structure compliance, which is the optimal solution developed using Objects SDK under the C#.Net language (Amirian, 2023).

The spatial comparison operators between two geometric objects in the ArcObject SDK return boolean values indicating whether the desired relationship exists or not. The IRelationalOperator interface is available in the Geometry library and provides several methods to evaluate whether two input geometries satisfy a spatial relationship with each other. To execute queries, spatial constraints are applied using spatial filters, which serve as parameters for IFeatureClass.search, IFeatureClass.select, or similar methods across selected object classes (Fig. 4) (Amirian, 2023).

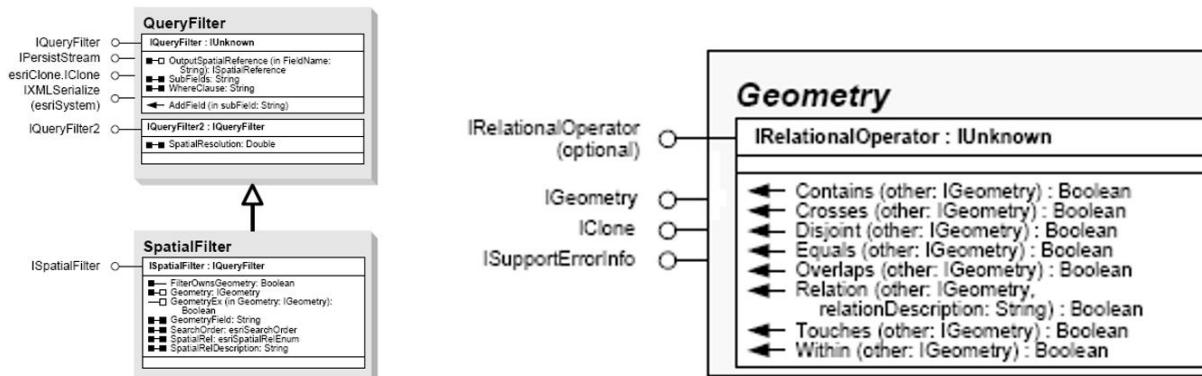


Figure 4. Some basic and popular libraries of ArcObject (Amirian, 2013)

RESEARCH RESULTS AND DISCUSSION

Test Data

The test data is the "Population" dataset of Thanh Hoa province, standardized and edited according to QCVN 73:2023/BTNMT for national geographic database scales of 1:2,000, 1:5,000, 1:10,000, including various layers as shown in Fig. 5. The population group data layers include: Point-type houses with 549,394 objects, Other technical infrastructure points with 681, National meteorological stations as point features 12, Environmental monitoring stations as point features 3, Electric poles 28,485, Health facilities as point features 367, Educational facilities as point features 610, Sports facilities as point features 10, Cultural facilities as point features 857, Commercial service facilities as point features 1,181, Office headquarters as point features 274, Religious buildings as point features 465, State agency headquarters as point features 464, Industrial facilities as point features 1,587, Agricultural production facilities as point features 7, Waste treatment facilities as point features 1, Security facilities as point features 51, National defense

facilities as point features 36, Power transmission lines 4,948, Boundaries (C) 10,543, Residential areas (S) 18,137, Building blocks (S) 1,944, Other technical infrastructure as polygon features 2,174, Sports facilities as polygon features 356, Cultural facilities as polygon features 224, Commercial service facilities as polygon features 580, Industrial facilities as polygon features 751, Agricultural production facilities as polygon features 419, Special function zones (S) 9, Residential areas as polygon features 172,285, National meteorological stations as polygon features 3, Health facilities as polygon features 399, Educational facilities as polygon features 2,150, Office headquarters as polygon features 595, Religious buildings as polygon features 235, State agency headquarters as polygon features 849, Waste treatment facilities as polygon features 19, Security facilities as polygon features 93, National defense facilities as polygon features 113. In addition to the test data layers, there will be other related data layers such as transportation thematic group data as polygon features.

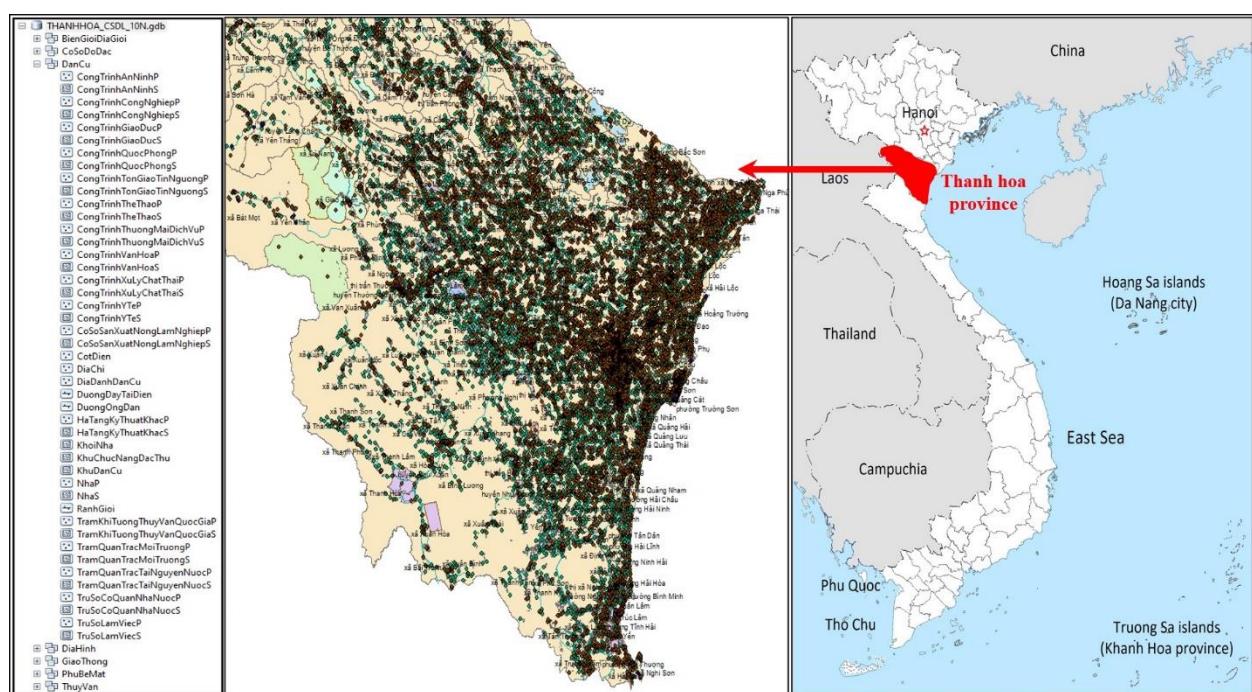


Figure 5. Data and experimental area

Results of tool development, testing, and evaluation

Software interface and data quality testing process

The software is developed as a Plug-In for ArcMap. After installation, it integrates

seamlessly into ArcMap with a main interface as shown in Fig. 6. Quality testing rules based on technical criteria for content structure models are programmed and embedded within the software. Users only need to follow the 7 steps as illustrated in the right panel of Fig. 3 above to automatically evaluate data quality.

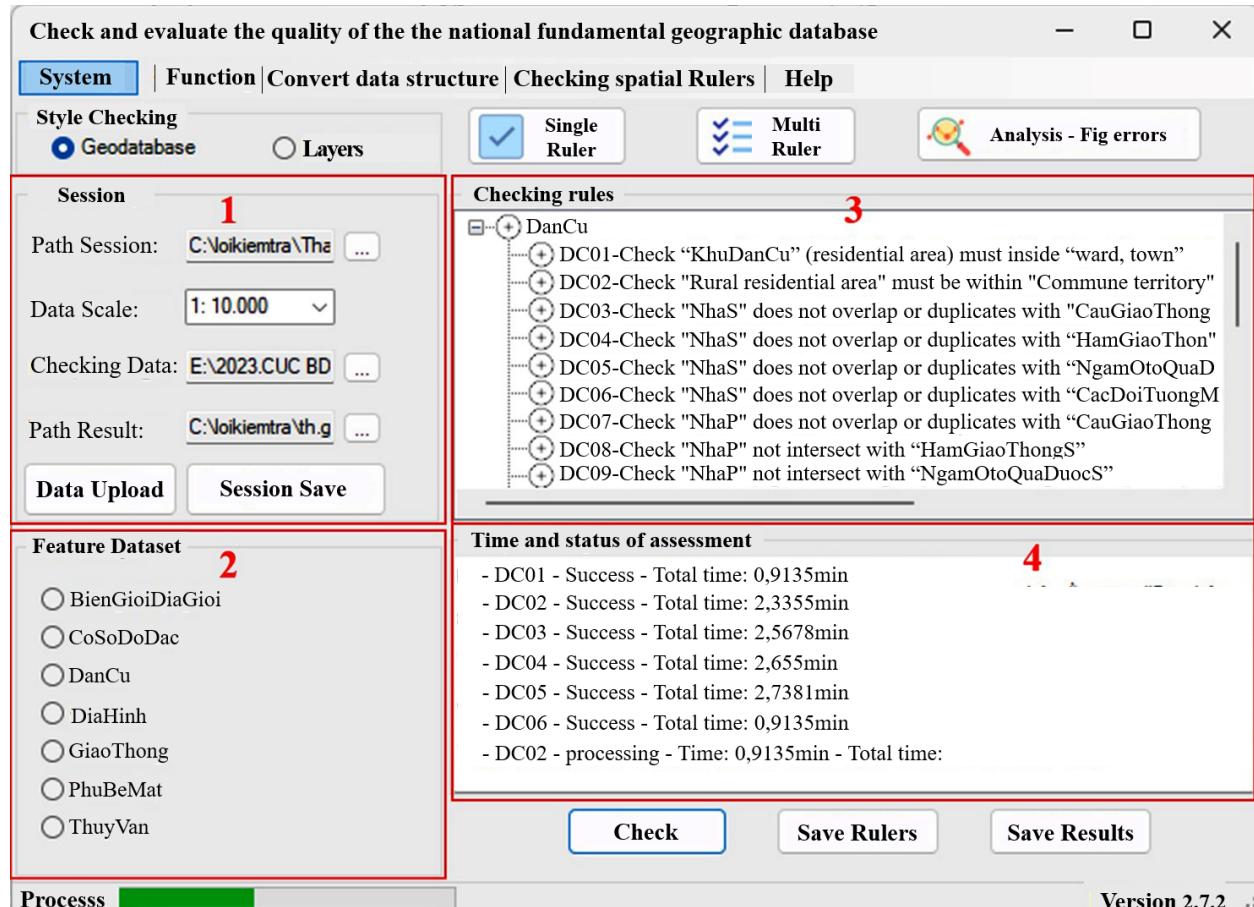


Figure 6. Main interface of the software tool

Step 1: Start ArcMap software and activate the tool by clicking on the icon on the toolbar. Step 2: Create a new session or open an existing one: go to the "System/create session" menu. This is the first step to initialize a Job for work. The created session is stored in a folder and reopened when needed to perform data quality checks. Creating a session includes information on the evaluation scale of the data, loading geodatabase data according to standards, selecting the location to save the evaluation result, and saving the session (Area 1 of Fig. 6). After creating the session, proceed to Step 3 by

clicking on the "Load data" button to load the data to be checked into the software. After successful loading, the groups of data layers are displayed in Area 2, allowing users to select them for testing. Step 4 involves selecting accuracy checking rules for spatial data by accessing the "Assessment/Rules 10N ratio" menu, which contains a list of general and specific evaluation rules; simply click on the corresponding rule to be tested (Fig. 7). The selected rule lists will be added to the evaluation content in Area 3 of Fig. 6.

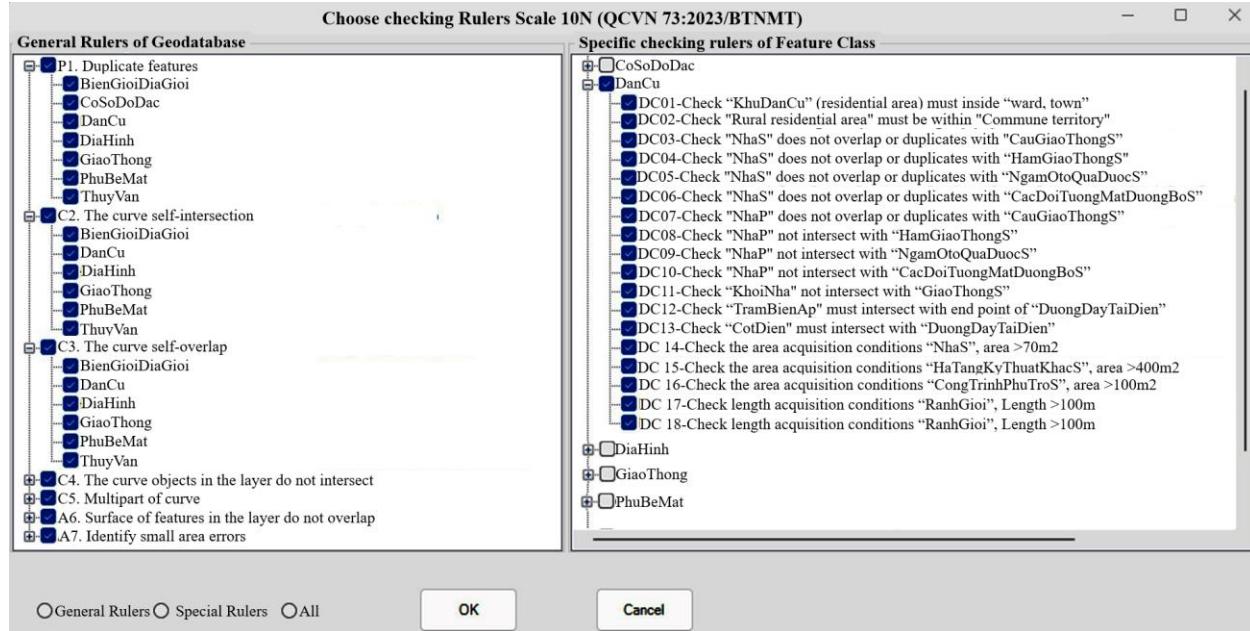


Figure 7. Interface for selecting spatial data validation rules

After selecting the rules, proceed with Step 5 by clicking on the "Check" button in Fig. 6 to automatically validate the data quality. The progress information for each data layer's validation process is displayed in Area 4 of Fig. 6. Once the validation is complete, proceed with steps 6 and 7 to generate reports and analyze errors. The interface for error analysis and correction, as shown in Fig. 8, includes: A list of datasets with errors and the corresponding error-prone layers within each dataset. Information table detailing the validation rules and the total number of errors for each selected layer in Area (1). Information table listing

specific objects that violate the selected erroneous rule in Area (2). 4.1. Allows viewing and displaying error objects according to the list in Area (3). 4.2. A table within ArcMap containing error-marked layers and related layers for easy editing and verification. 4.3. The map area of ArcMap showing error objects and related information, where users can use object editing tools in ArcMap software combined with options in Area (5). This area displays statuses that allow users to make repairs and mark objects as edited or not. Information about edited objects is stored in the database for subsequent editing and verification phases.

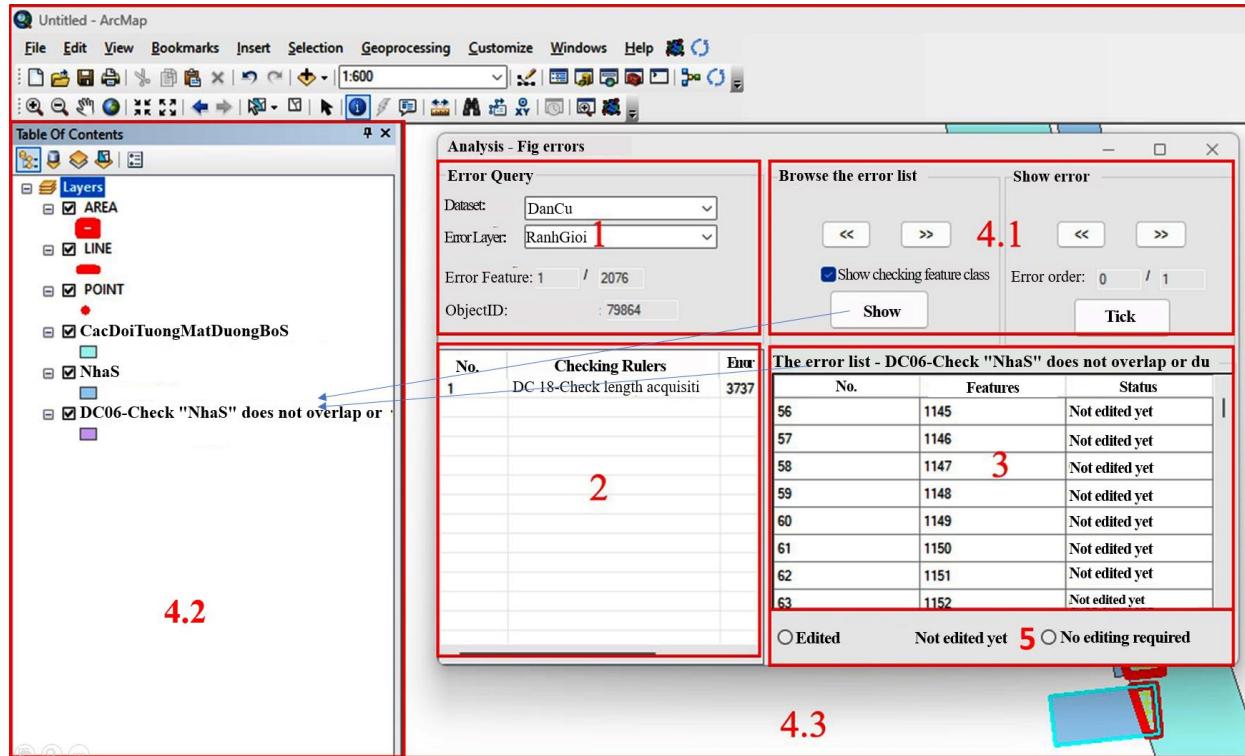


Figure 8. Interface for error analysis and correction

Test results with the "Population" dataset of Thanh Hoa province

Conducted testing with predefined rules in the software for the "Population" dataset.

Exported results after verification, listing layers with erroneous data as in statistical table 2.

Table 2. List of spatial errors in the Population group

No.	Name Ruler	Checked class	error number
1	DC01-Check "KhuDanCu" (residential area) must inside "ward, town"	KhuDanCu	131
2	DC02-Check "Rural residential area" must be within "Commune territory"	KhuDanCu	107
3	DC03-Check "NhaS" does not overlap or duplicates with "CauGiaoThongS"	NhaS	1
4	DC06-Check "NhaS" does not overlap or duplicates with "CacDoiTuongMatDuongBoS"	NhaS	2076
5	DC 14-Check the area acquisition conditions "NhaS", area >70m ²	NhaS	7
6	DC11-Check "KhoiNha" not intersect with "GiaoThongS"	KhoiNha	176
7	DC12-Check "TramBienAp" must intersect with end point of "DuongDayTaiDien"	CongTrinhCongNghiepP	101
8	DC 18-Check length acquisition conditions "RanhGioi" (TuongVay), Length >100m	RanhGioi	3737

Using the "Analysis-Error Editing" function to view, analyze, and edit errors for each layer on the map. Some specific errors are shown and analyzed as follows:

The "Population" layer has 02 error groups corresponding to rules DC01 and DC02. According to rule DC01 in the table above, there are a total of 131 erroneous objects. Population

objects with the object code "CA02," which represents "Urban," should be within administrative areas on land with the object code "Administrative boundaries of wards, townships." However, these erroneous areas in the data show administrative boundaries as communes, for example, object 390 in Fig. 9.

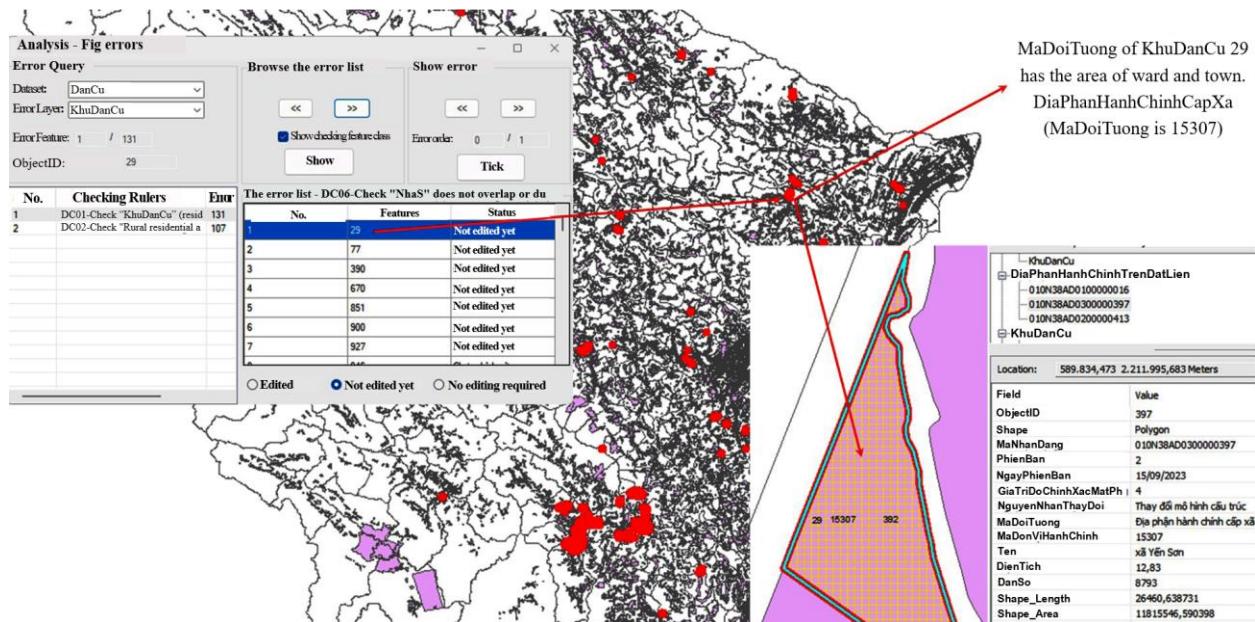


Figure 9. Urban residential area is categorized as urban but located within the administrative boundaries of a commune.

Similarly, rule DC02 - "rural residential area" must be within "Commune boundaries," with 107 objects in error. Rural residential areas found within the administrative boundaries categorized as towns or townships are flagged

as errors by the software, as shown in Fig. 10. The error area may cover the entire residential area or overlap partially with inappropriate administrative boundaries.

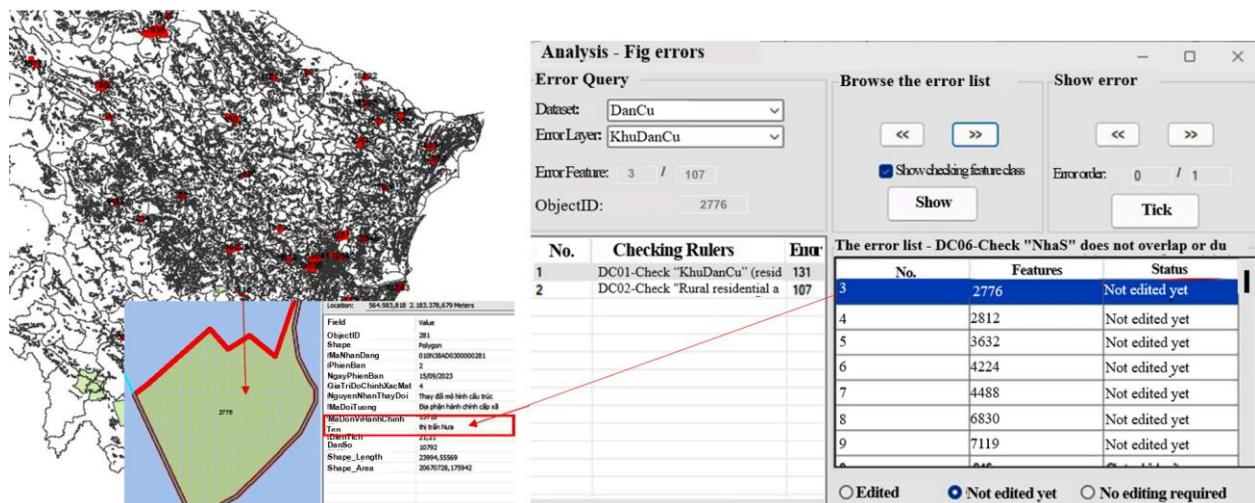


Figure 10. Rural residential area in error

The "Area type houses" layer has 3 error groups equivalent to rules DC03, DC06, and DC15. Rule DC03 - Checking "Area type houses" should not overlap or coincide with "Area type

bridge" has 1 object in error. The standalone house object with ID 79444 has its boundary overlapping with a portion of the boundary of the Kiều Đại bridge area (Fig. 11).

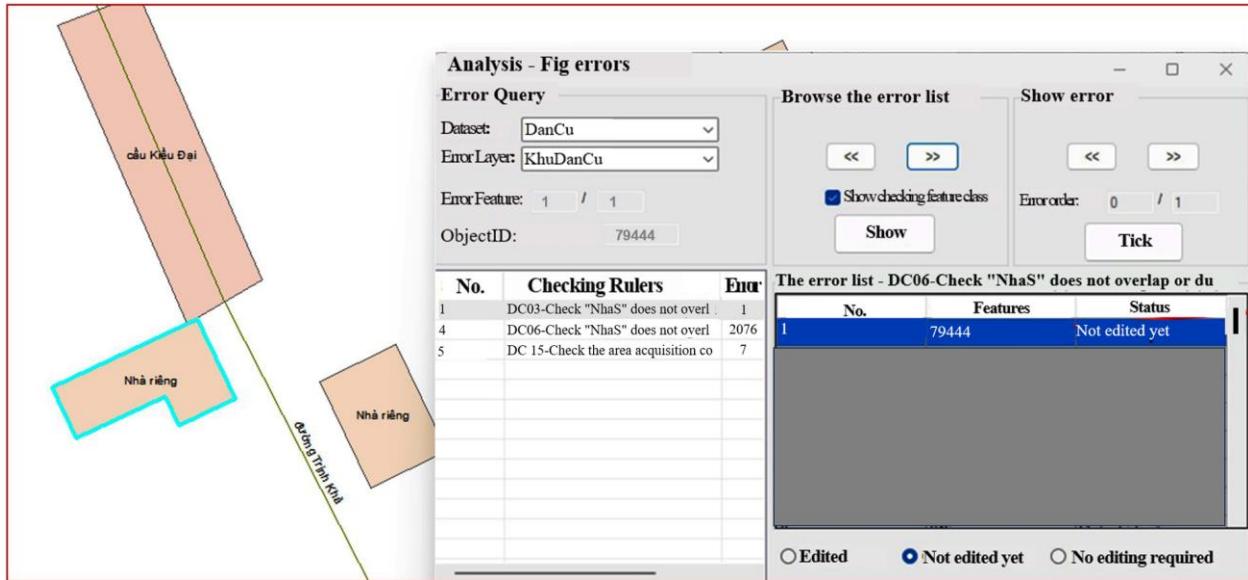


Figure 11. Error S houses should not overlap or coincide with Road S

Rule DC06 - Check "House (S)" should not overlap or coincide with "Road polygons". There are 2076 spatial error objects. The error locations are marked in red on the map (Fig. 12).

For example, error object numbered 2076 in the displayed list is highlighted in blue, with the overlapping part with the road polygon circled in red, as zoomed in on in Fig. 12.

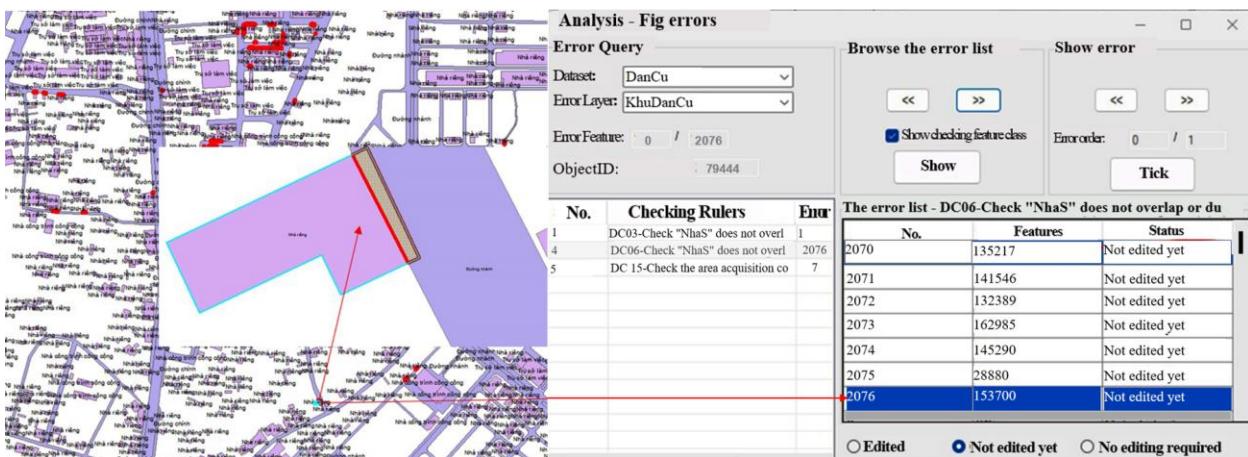


Figure 12. Error S houses overlap or coincide with Road S

Rule DC15 - Check "House (S)" with an area < 70m² has a total of 7 error objects, which violate the specified small area requirement (left panel of Fig. 13). This error is similar to Rule DC18 regarding boundary length, with 3737 error objects failing to meet the length

requirement of less than 100m (right panel of Fig. 13). The software checks the area and length of spatial objects in the geodatabase. Cases that do not meet the standards are flagged as errors, allowing users to delete or merge objects.

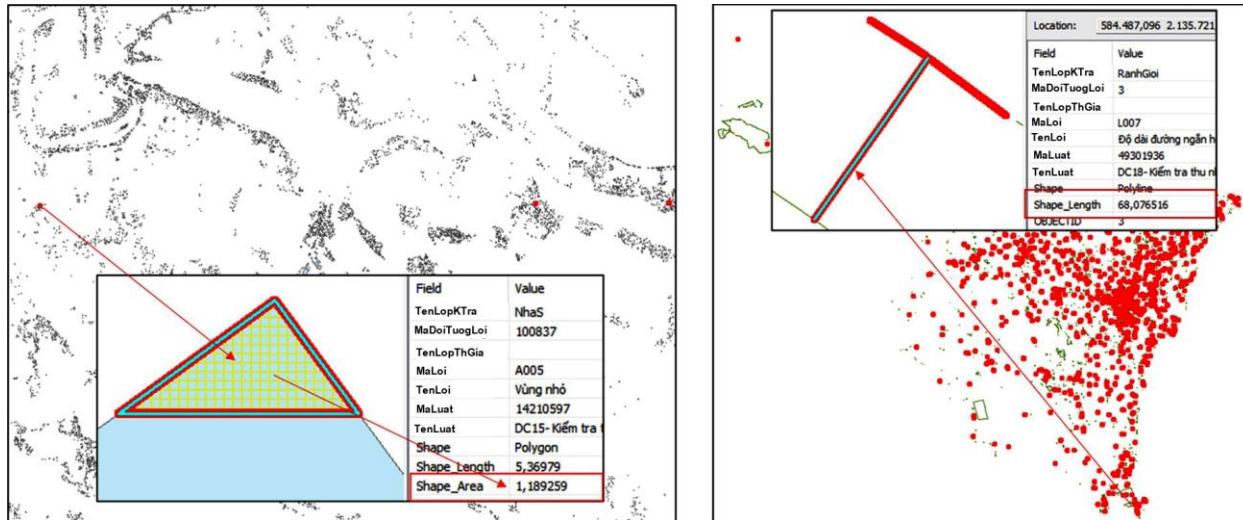


Figure 13. Error in data capture of "Block of Houses (S)" and "Boundary"

The "Block of Houses (S)" layer has Rule DC11 concerning the spatial relationship between "Block of Houses (S)" and the "Transport Infrastructure (Area)" layer. According to this rule, a block of houses should not overlap with or be on top of any transport infrastructure area objects. This rule identifies 176 error objects in the "Block of Houses (S)"

layer. The transport infrastructure area objects encompass various layers in the transportation specialization group, such as tunnels, train bridges, road tunnels, road bridges, roadways (Area), etc. (Fig. 14). Many errors indicate very small overlapping areas that are difficult to detect with the naked eye.

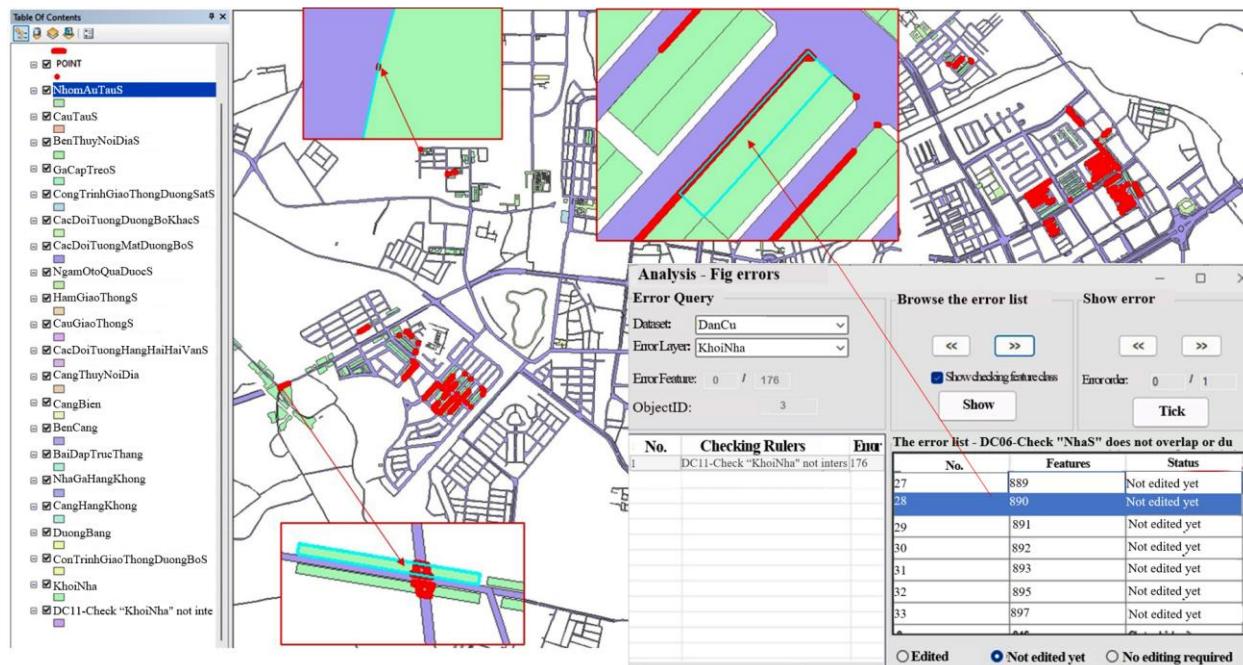


Figure 14. Spatial relationship error between "Block of Houses" and "Transport Infrastructure (Area)"

The "Block of Houses" layer has Rule DC11 regarding the spatial relationship between "Block of Houses" and the "Transport Infrastructure (Area)" layer. According to this rule, blocks of houses should not overlap or be on top of any transport infrastructure area objects. This rule identifies 176 error objects in the "Block of Houses" layer. The transport

infrastructure area objects encompass various layers within the transportation specialization group, such as tunnels, train bridges, road tunnels, road bridges, roadways (Area), etc. (Fig. 14). Many errors indicate very small overlapping areas that are difficult to detect with the naked eye.

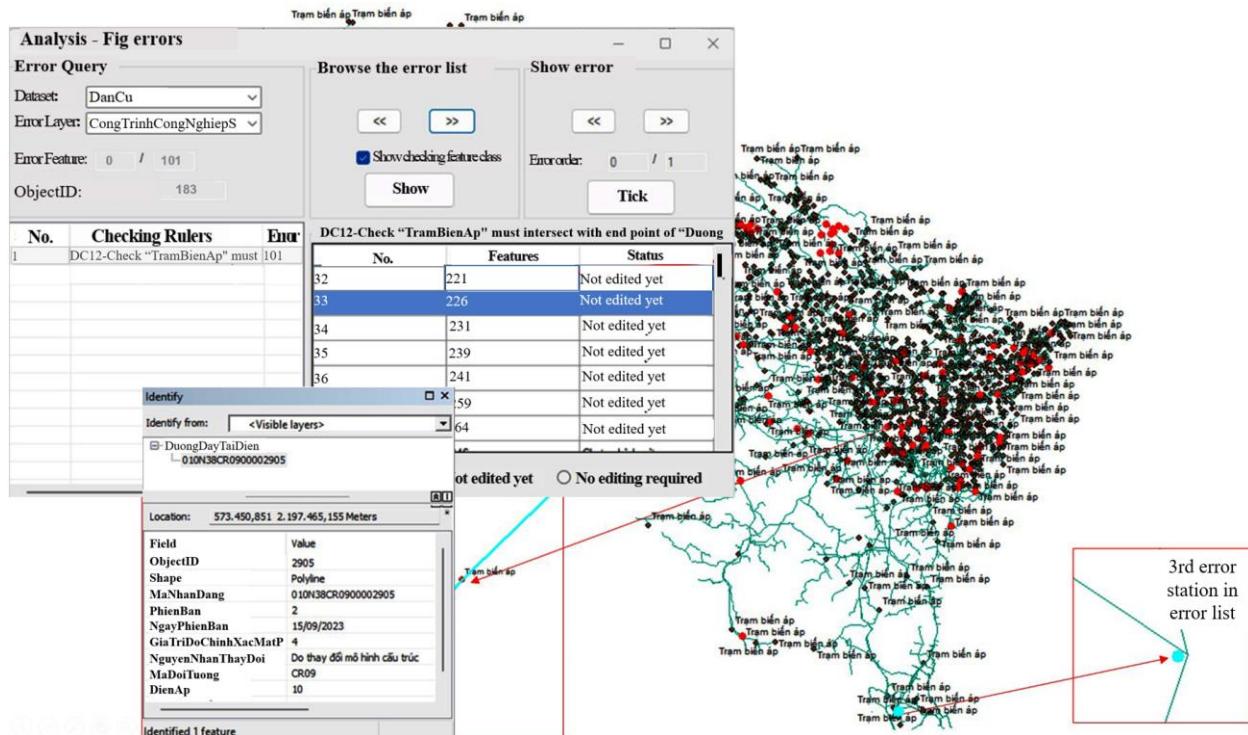


Figure 15. Error of spatial relationship between transformer stations and power transmission lines

In general, the spatial analysis rules for the Population group have been pre-built into the software. Users simply need to select the rules, click the check button, and after a very short time, the software can check the entire spatial relationship according to the national geographic data standards.

CONCLUSIONS

a/ The paper has conducted research and specified spatial data consistency rules for the Population group of national geographic data standards at scales of 1:2,000, 1:5,000, 1:10,000, QCVN 73:2023/BTNMT.

b/ The paper also studied GIS programming libraries, especially ArcObject libraries, to develop automated tools for spatial data quality checks in the Population group.

c/ The tool results were applied to check and evaluate the spatial data consistency of the Population group of the national geographic dataset scale 10,000 in Thanh Hoa province according to QCVN 73:2023/BTNMT.

d/ The results show the effectiveness of the tool clearly through processing time compared to reality, and the objective transparency of data checking work.

REFERENCES

1. Rasheed, M.U., Ahmad S.R., and Saleem R.J.J., 2015. Building a Geodatabase for Parcel and Cadaster Mapping and Add-Ins Development: A Case Study for Defense Housing Authority (DHA). Lahore, Pakistan, 7(6): p. 588-597.

2. National Assembly, 2018. Law on Surveying and Mapping. Law number: 27/2018/QH14 dated June 14, 2018.
3. Government, 2019. Decree detailing a number of provisions of the Law on Surveying and Mapping. Number: 27/2019/NĐ-CP, dated March 13, 2019.
4. Ministry of Natural Resources and Environment, 2018. Regulations on inspection, appraisal, and acceptance of quality of measurement and mapping products. Circular No. 24/2018/TT-BTNMT dated November 15, 2018.
5. Ministry of Natural Resources and Environment, 2019. Technical regulations on the model structure, content of national geographic database at scales 1:2,000 and 5,000. Circular No. 23/2019/TT-BTNMT dated December 25, 2019.
6. Ministry of Natural Resources and Environment, 2020. QCVN 42:2020/BTNMT, National technical regulations on geographic information base standards. Circular No. 06/2020/TT-BTNMT dated August 31, 2020.
7. Ministry of Natural Resources and Environment, 2020. Technical regulations on the model structure, content of national geographic database at scales 1:10,000, 1:25,000. Circular No. 15/2020/TT-BTNMT dated November 30, 2020.
8. Ministry of Natural Resources and Environment, 2022. QCVN 71:2022/BTNMT, National technical standards on geographic database base at scales 1:50,000, 1:100,000. Circular No. 07/2022/TT-BTNMT dated June 30, 2022.
9. Ministry of Natural Resources and Environment, 2023. QCVN 73:2023/BTNMT, National technical standards on geographic database base at scales 1:2,000, 1:5,000, 1:10,000. Circular No. 08/2023/TT-BTNMT dated July 31, 2023.
10. Quoc Yen Phan, Anh Tuan Tran, and Bui T.X.H., 2023. Research on spatial consistency checking according to National Technical Regulation on basic Geographic Information Section on Special Construction Engineering 6(2).
11. Kukulska, A., Salata, T., Cegielska, K. and Szylar, M., 2018. Methodology of evaluation and correction of geometric data topology in QGIS software. *Acta Scientiarum Polonorum. Formatio Circumiectus* 17(1), 137-150.
12. Sehra, S.S., Singh J., and Rai H.S., 2017. Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet* 9(2), pp. 15.
13. Li, B., Huang, X., Song, N., Wang, Q., Zhang, G. and Nie, L., 2020. Development of an aquaculture suitability assessment system of sea areas based on plug-in technology. In 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, pp. 250-254.
14. Hendricks, M., Athey J., and Macpherson A., 2022. QA/QC of GeMS Data Case Study: The Alaska DGGS Geologic Mapping System..
15. Booth, B. and Mitchell A., 2001. Getting started with ArcGIS. Esri Redlands, CA, USA.
16. Eldrandaly, K.J.I.A.J.I.T., 2013. Developing a GIS-based MCE site selection tool in ArcGIS using COM technology, 10(3), pp. 276-282.
17. Li, Z. and Li, Z., 2020. Component GIS, ArcObjects and ArcGIS Server. *Pipeline Spatial Data Modeling and Pipeline WebGIS, Digital Oil and Gas Pipeline. Research and Practice*, pp.103-117.
18. Amirian, P., 2013. Beginning ArcGIS for desktop development using .NET. John Wiley & Sons.
19. ESRI, 2023. Spatial relationships in ArcGIS GeoAnalytics Server. Available from: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/big-data-analytics/spatial-relationships-with-big-data.htm#>.

Integrating Sentinel-2 Imagery and Machine Learning for Crop Damage Evaluation in Thai Agriculture

1st Viphada Bonnlerd
Geo-Informatics and Space Technology Development Agency (Public Organization), Bangkok 10210, Thailand
 Bangkok 10210, Thailand
 viphada.boo@gistda.or.th

5th Sansita Rattanasupa
Office of the Permanent Secretary, Ministry of Agriculture and Cooperatives
 Bangkok 10900, Thailand
 disas.plan@gmail.com

2nd Woranut Chansury
Geo-Informatics and Space Technology Development Agency (Public Organization), Bangkok 10210, Thailand
 Bangkok 10210, Thailand
 woranut@gistda.or.th

6th Sirilux Noikeaing
Department of Agricultural Extension
 Bangkok 10900, Thailand
 siriluxr@gmail.com

3rd Panu Nueangjumnong
 2: *Geo-Informatics and Space Technology Development Agency (Public Organization), Bangkok 10210, Thailand*
 Bangkok 10210, Thailand
 panu.n@gistda.or.th

7th Apantri Yutthapan
Thailand Meteorological Department
 Bangkok 10200, Thailand
 aphantreey@yahoo.com

4th Pakorn Petchprayoon
Geo-Informatics and Space Technology Development Agency (Public Organization), Bangkok 10210, Thailand
 Bangkok 10210, Thailand
 pakorn@gistda.or.th

8th Pheeraphong Rattanaburi
Department of Water Resources Bangkok 10900, Thailand
 popirre60@gmail.com

Abstract— This research enhances food security and sustainable agriculture by improving crop damage assessments, aligning with SDG 2: Zero Hunger and SDG 13: Climate Action. Accurate damage evaluations can promote growth in Thailand's agribusiness sector, where over 30% of the population works in agriculture, frequently affected by drought. The goal is to create a Crop Damage Assessment Index (CDAI) model for drought-affected areas with the assumption underlying this research is that vegetation indices derived from satellite imagery can effectively monitor and predict crop damage. We are analyzing over 5 million damaged and non-damaged plots across nine provinces to train the model. Using 9 vegetation indices and the Random Forest machine learning technique, the model distinguished between damaged and undamaged plots with Sentinel-2 satellite imagery at 20 meters resolution. MNDWI stands out as the most influential index, especially for maize and rice. NDVI is notable for sugarcane, while REIP is crucial for cassava. Other indices, such as NDWI, S2REP, and SAVI, show varying contributions. These tailored indices support the model's strong performance, with ROC values of 86% for rice, 91% for maize, 89% for cassava, and 86% for sugarcane. These results highlight the reliability of the CDAI model in distinguishing drought-affected plots. The CDAI model supports farmers and agricultural officers with accurate assessments, essential for drought management and fair compensation. Accessible via the GISTDA platform (cropsdrought.gistda.or.th), it aids decision-making without requiring GIS knowledge, though ongoing validation and data updates are necessary for improved efficiency.

Keywords— *crop damage assessment, machine learning, remote sensing, vegetation indices, sustainable agriculture, drought management*

I. INTRODUCTION

Agriculture is a critical sector in Thailand, with over 30% of the population engaged in farming activities. However, frequent droughts pose significant challenges to crop productivity, leading to economic losses and food insecurity. Addressing these issues is crucial for sustainable agricultural development and aligns with global initiatives such as

Sustainable Development Goals (SDGs) 2: Zero Hunger and 13: Climate Action. Current crop damage assessment methods rely heavily on local committees following the Ministry of Agriculture and Cooperatives' guidelines, which can be time-consuming and inconsistent due to limited resources and subjective evaluations. There is a pressing need for more accurate, efficient, and scalable methods to assess crop damage, especially in drought-prone areas. This model supports government officers and local committees by decreasing the number of workers needed, reducing the time required, and lowering budget expenditures.

Recent advances in remote sensing and machine learning offer promising solutions for agricultural monitoring. Satellite imagery, particularly from the Sentinel-2 satellite, provides high-resolution data that can be analyzed using various vegetation indices to assess crop health. Machine learning techniques, such as the Random Forest algorithm, can effectively process large datasets to distinguish between damaged and undamaged crops, providing precise and objective assessments.

This study aims to develop a Crop Damage Assessment Index (CDAI) model for individual agricultural plots in nine drought-affected provinces of Thailand. The research analyzes over 5 million plots using 13 vegetation indices derived from Sentinel-2 imagery. However, after model optimization, the number of significant indices was reduced to 9, which included the Modified Normalized Difference Water Index (MNDWI), Normalized Difference Vegetation Index (NDVI), and Normalized Difference Water Index (NDWI), among others. These indices were selected based on their importance in accurately classifying crop damage, with MNDWI being the most significant.

However, challenges remain in ensuring the accuracy and reliability of these models, particularly when applied to diverse regions and crop types.

In this context, our research addresses these challenges by developing a robust CDAI model tailored to the specific

conditions of drought-prone areas in Thailand. The principal conclusions of this work underscore the importance of integrating advanced technologies in agricultural management to support sustainable development and resilience against climate variability.

II. MATERIALS AND METHODS

A. Study Area and Data Collection

The research focuses on forecasting drought-risk areas for Thailand by using data from nine provinces: Kamphaeng Phet, Uthai Thani, Nakhon Ratchasima, Surin, Roi Et, Sakon Nakhon, Nakhon Sawan, Phichit, and Suphan Buri. The nine provinces selected represent Thailand's major agricultural areas with varying drought frequencies and severities. These provinces were chosen based on historical drought data. Rice, maize, cassava, and sugarcane were selected as they are contributing significantly to the national economy. Data was collected on 97,444 damaged and 5,615,607 non-damaged plots from 2015 - 2021, with crop data obtained from the

Department of Agriculture Extension (DOAE). These plots from the nine provinces are used for training the model, which will then be applied to monitor and predict drought risk across the entire country.



Fig. 1. Study Area: Selected Provinces for Crop Damage Assessment in Thailand

Aggregated data to assist farmers affected by drought (KORSOR02) from the DOAE from 2015 to 2021, covering four types of crops: rice, maize, cassava, and sugarcane, in nine provinces, with a total of 97,444 plots.

TABLE I. NUMBER OF AGRICULTURAL PLOTS AFFECTED BY DROUGHT, CATEGORIZED BY CROP TYPE, IN THE NINE PROVINCES FROM 2015 – 2021

Province	Crops			
	Rice	Maize	Cassava	Sugarcane
Nakhon Sawan	7,240	791	1	22
Phichit	3,708	509	26	9
Kamphaeng Phet	126	32	0	0
Suphan Buri	7,767	190	0	0
Uthai Thani	6,961	376	84	58
Nakhon Ratchasima	35,265	4,279	0	0
Surin	2,032	0	0	0
Roi Et	27,221	0	0	0

a. Source: Department of Agricultural Extension (2022)

B. Satellite Imagery

We utilized Sentinel-2 satellite imagery, providing 20-meter resolution data. The images were pre-processed to correct atmospheric and radiometric distortions, ensuring accurate vegetation index calculations.

The Sentinel-2 satellite is a high-resolution, wide-area multispectral imaging mission supporting the Copernicus Land Monitoring studies. The resource survey satellites consist of two satellites, Sentinel-2A and Sentinel-2B, launched into orbit in 2015 and 2017, respectively, by the European Space Agency. Sentinel-2 satellites provide optical images with spatial resolutions of 10, 20, and 60 meters, varying depending on the wavelength. The Sentinel-2 satellites support three wavelength ranges: Short Wave Infrared (SWIR), Near-Infrared (NIR), and Visible, with a total of 13 spectral bands. The coverage area spans the entire globe, both land and ocean, with data recorded every 5 days, enabling us to update the results every week. They were developed to support various applications such as agriculture, resource exploration, disaster management, water resource management, and more.

C. Model Preparation

The data preparation process began with downloading Sentinel-2 vegetation indices from Google Earth Engine (GEE) that matched the corresponding years of the KORSOR02 data. The KORSOR02 data provided ground-truth records of drought-affected and normal plots, allowing for precise alignment with the satellite imagery. For each crop field, the mean values of the Sentinel-2 indices were calculated by matching the satellite pixel values to the corresponding crop field boundaries. These aggregated vegetation index values were then used as input features for the Random Forest model, ensuring that the satellite data accurately reflected the conditions within each crop field. Finally, the Random Forest Classifier was applied to analyze the prepared datasets and predict drought-affected and normal plots for each crop type.

Sampling points were determined by first separating the data by crop type (rice, maize, sugarcane, cassava) to ensure that each crop was modeled independently. Within each crop-specific dataset, plots were labeled as drought-affected (1) or normal (0) based on historical drought records. This crop-wise stratification ensured that the Random Forest Classifier could learn and predict patterns specific to each crop type effectively.

III. RESULTS

A. Model Performance

TABLE II. SUMMARIZE OF ALGORITHM PERFORMANCE

Model	Rice	Maize	Cassava	Sugarcane
	Accuracy			
Logistic Regression	0.58	0.57	0.62	0.62
Decision Trees	0.87	1.00	1.00	0.91
XG Boost	0.87	0.95	0.95	0.85
Random Forest Classifier	0.86	0.86	0.91	0.89
ROC				
Linear Regression	0.65	0.68	0.71	0.67
Logistic Regression	0.55	0.57	0.62	0.62
Decision Trees	0.94	1.00	1.00	0.91

b. Source: Department of Agricultural Extension (2022)

After comparing the models, a used metric for assessing classification performance is the Receiver Operating Characteristic (ROC) curve. This metric evaluates the model across all potential classification thresholds by plotting the True Positive Rate (TPR), representing the percentage of correctly predicted positive cases, against the False Positive Rate (FPR), the percentage of actual negative cases mistakenly classified as positive. The overall performance of the model is summarized by the Area Under the ROC Curve (AUC-ROC), with values closer to 1 indicating better discrimination capability between positive and negative classes, making it a reliable measure for model evaluation.

We selected the Random Forest Classifier as our final model. We chose the Random Forest classifier because of its general robustness and consistency across all crop types, even though the Decision Tree model performed better in accuracy and ROC for crops. Random Forest is less prone to overfitting than Decision Trees, particularly when working with large datasets and multiple features. It was also more efficient and practical because it needed less computing time than more complicated models like XGBoost.

The development of a damage assessment model revealed that using Machine Learning with the Random Forest method selected 9 indices out of a total of 13 indices that are important for indicating crop damage and non-damage. The top three indices important for distinguishing between damage and non-damage in crop are as follows: Modified Normalized Difference Water Index (MNDWI) is the most important followed by Normalized Difference Vegetation Index (NDVI), and the Normalized Difference Water Index (NDWI). The important details of each index are shown in Table III.

To summarize the performance of models used to assess crop damage for rice, maize, cassava, and sugarcane. The table lists the most important features for each crop, their feature values, and the corresponding model performance metrics, including accuracy and ROC Curve values. The accuracy for each crop model is shown in the Table III.

TABLE III. FEATURE IMPORTANCE

Feature	Crops			
	Rice	Maize	Cassava	Sugarcane
MNDWI	0.357	0.403	0.166	0.195
NDVI	0.147	0.158	0.156	0.221
NDWI	0.103	0.096	0.103	0.110
S2REP	0.087	0.059	0.075	0.075
REIP	0.074	0.065	0.181	0.062
IPVI	0.0708	0.057	0.047	0.082
SAVI	0.065	0.059	0.120	0.109
NDRE	0.048	0.066	0.097	0.083
MCARI	0.045	0.032	0.052	0.061

c. Source: Authors (2023)

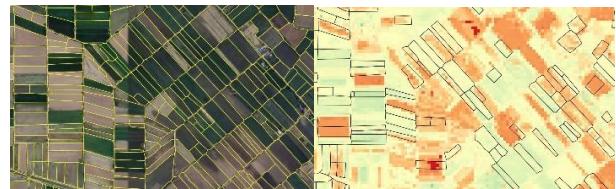


Fig. 2. Example of CDAI map in the rice field

IV. DISCUSSION

A. Interpretation of Results

The results of this study highlight the effectiveness of the Crop Damage Assessment Index (CDAI) model in accurately identifying drought-affected agricultural plots using remote sensing and machine learning techniques. The high accuracy rates and robust ROC values across different crops demonstrate the model's reliability and applicability in real-world scenarios. The key vegetation indices identified, such as MNDWI, NDVI, and NDWI, effectively address the assumption that vegetation indices derived from satellite imagery can monitor and predict crop damage with good results.

B. Limitations and Future Improvements

The CDAI model primarily relies on vegetation indices derived from Sentinel-2 satellite imagery, which effectively reflect crop health and damage, making it robust for assessing drought impacts. Machine-learning models, such as CDAI, provide scalability, objectivity, and the ability to process large datasets efficiently, offering a superior alternative to traditional crop damage assessment methods. In contrast, manual surveys are time-intensive, resource-demanding, and prone to human error.

However, machine-learning models are still undergoing continuous learning and refinement. As the CDAI model relies on satellite-derived data, periodic validation using ground-truth data is critical to maintaining its accuracy and reliability. This validation will help bridge potential gaps between satellite observations and on-ground realities, further enhancing the model's practical applicability in agricultural decision-making.

Moreover, to ensure the identified damage is specifically caused by drought and no other factors, our next step is to incorporate the Drought Risk Index (DRI) and Drought Severity Index (DSI) as input conditions. These indices will help confirm that crop damage is drought-induced. Additionally, the CDAI model is designed to integrate real-time Sentinel-2 data for frequent updates, and we propose establishing a periodic validation protocol using ground-truth data collected by agricultural agencies. This will ensure the model remains accurate and responsive to evolving climatic conditions.

C. Implications for Agricultural Management

The CDAI model has significant implications for agricultural management in drought-prone regions. By providing timely and accurate damage assessments, it can help in the equitable distribution of compensation and resources, thereby supporting farmers' livelihoods and promoting food security. The model's accessibility via the GISTDA platform (cropsdrought.gistda.or.th) ensures that decision-makers can easily use the tool without needing extensive GIS knowledge, thus democratizing access to advanced agricultural monitoring technologies.

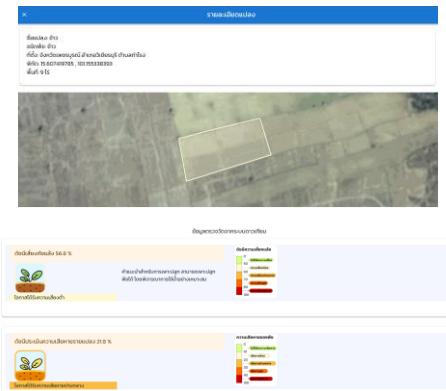


Fig. 3. Example Output from the CropDrought Platform

V. CONCLUSIONS

The Crop Damage Assessment Index (CDAI) model developed in this study presents a significant step forward in the field of agricultural damage assessment, particularly for drought-prone regions in Thailand. By integrating Sentinel-2 satellite imagery with machine learning techniques, we achieved high accuracy rates in distinguishing between damaged and undamaged plots across various crops, including rice, maize, cassava, and sugarcane. The key vegetation indices identified, such as MNDWI, NDVI, and NDWI, proved essential in assessing crop health and damage accurately.

Our findings demonstrate that the CDAI model not only enhances the precision of damage assessments but also supports timely and fair compensation for farmers, thereby promoting food security and sustainable agricultural practices. The model's accessibility via the GISTDA platform allows for easy adoption by decision-makers without requiring extensive technical expertise.

In the broader context, this research contributes to the global efforts to achieve Sustainable Development Goals (SDGs) 2 (Zero Hunger) and 13 (Climate Action) by improving the resilience of agricultural systems to climate variability. Future research should focus on incorporating

additional data sources, expanding the study to more diverse regions and crop types, and conducting continuous field validations to further refine and validate the model.

In conclusion, the CDAI model provides a robust, scalable, and accessible tool that can significantly aid in managing drought impacts and supporting sustainable agricultural development in Thailand and potentially other regions facing similar challenges.

VI. FUNDING:

This research was funded by the National Research Council of Thailand for the year 2022 under the project "Geo-Informatics Approach for Assessing the Risk of Crops Loss and Damage Associated with Drought at Farm Level"

ACKNOWLEDGMENT

We acknowledge the Ministry of Agriculture and Cooperatives, Department of Agriculture Extension, Royal Irrigation Department, Thai Meteorological Department, Office of the National Water Resources for providing data and GISTDA for access to geospatial data platforms.

REFERENCES

- [1] Berger, K.; Atzberger, C.; Danner, M.; D'Urso, G.; Mauser, W.; Vuolo, F.; Hank, T. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sens.* 2018, 10, 85.
- [2] Budagovsky, A.I. Meteorological factor and soil moisture influence on evaporation. In Irrigation Problems in Lower Amu-Darya River; USSR RAS Publication: Moscow, 1956.
- [3] Chang, S.; Chen, H.; Wu, B.; Nasanbat, E.; Yan, N.; Davdai, B. A Practical Satellite-Derived Vegetation Drought Index for Arid and Semi-Arid Grassland Drought Monitoring. *Remote Sens.* 2021, 13, 414. <https://doi.org/10.3390/rs13030414>.
- [4] Demarez, V.; Dutchoit, S.; Baret, F.; Weiss, M.; Dedieu, G. Estimation of leaf area and clumping indexes of crops with hemispherical photographs. *Agric. For. Meteorol.* 2008, 148, 644–655.
- [5] Faridatul, M. I.; Ahmed, B. Assessing Agricultural Vulnerability to Drought in a Heterogeneous Environment: A Remote Sensing-Based Approach. *Remote Sens.* 2020, 12, 3363. <https://doi.org/10.3390/rs12203363>.
- [6] Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; Michaelsen, J. Climate Hazards Group. Available online: <http://dx.doi.org/10.15780/G2RP4Q> (accessed on 7 June 2023).
- [7] Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 1996, 257–266.
- [8] Kamonpat, T. Development of Monthly Grid Rainfall Data for Thailand Using Station and Remote Sensing Data. Master's Thesis, Kasetsart University, Bangkok, Thailand, 2019. Available online: https://senswat.eng.ku.ac.th/website/wp-content/uploads/2021/09/Kamonpat_5714501845.pdf (accessed on 7 June 2023).
- [9] Polina, L. Hyperspectral Vegetation Indices Calculated by QGIS Using Landsat TM Image: A Case Study of Northern Iceland. *Adv. Res. Life Sci.* 2020, 4, 70–78.
- [10] Hargreaves, G. Moisture Adequacies for Agriculture in the Southeast. Department of Agricultural and Irrigation Engineering, Utah State University, 1974.
- [11] Hu, X.; Ren, H.; Tansey, K.; Zheng, Y.; Ghent, D.; Liu, X.; Yan, L. Agricultural drought monitoring using European Space Agency Sentinel 3A land surface temperature and normalized difference vegetation index imageries. *Agric. For. Meteorol.* 2019, 279, 107707.
- [12] Hydrology. SCS National Engineering Handbook Section 4. Soil Conservation Service (SCS), U.S. Department of Agriculture, 1972.
- [13] Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 1988, 25, 295–309.

- [14] Jasinski, M. F.; Eagleson, P. S. Estimation of Subpixel Vegetation Cover Using Red-Infrared Scattergrams. *IEEE Trans. Geosci. Remote Sens.* 1990, 28, 253–267.
- [15] McKee, T. B.; Doesken, N. J.; Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993; pp. 179–184.
- [16] Mishra, S. K.; Singh, V. P. Soil Conservation Service Curve Number (SCS-CN) Methodology. Springer Science+Business Media: Dordrecht, The Netherlands, 2003.
- [17] Nicholson, C.; Sorlien, C. C.; Atwater, T.; Crowell, J. C.; Luyendyk, B. P. Microplate capture, rotation of the Western Transverse Ranges, and initiation of the San Andreas transform as a low-angle fault system. *Geology* 1994, 22, 491–495.
- [18] Palmer, W. C. Meteorological Drought. Research Paper No.45, U.S. Department of Commerce, Washington, D.C., 1965.
- [19] Palmer, W. C. Keeping Track of Crop Moisture Conditions, Nationwide: The New Crop Moisture Index. *Weatherwise* 1968, 21, 156–161.
- [20] Wang, J.; Price, K. P.; Rich, P. M. Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *Int. J. Remote Sens.* 2001, 22, 3827–3844.
- [21] W3schools. Node.js Introduction. Available online: https://www.w3schools.com/nodejs/nodejs_intro.asp (accessed on 7 June 2023).
- [22] Yoon, D.-H.; Nam, W.-H.; Lee, H.-J.; Hong, E.-M.; Feng, S.; Wardlow, B. D.; ... Kim, D.-E. Agricultural Drought Assessment in East Asia Using Satellite-Based Indices. *Remote Sens.* 2020, 12, 444. <https://doi.org/10.3390/rs12030444>.
- [23] Zargar, A.; Sadiq, R.; Khan, F. I. Uncertainty-Driven Characterization of Climate Change Effects on Drought Frequency Using Enhanced SPI. *Water Resour. Manag.* 2014, 15–40.
- [24] Nga Nguyen, N., et al. (2016). Exploring the potential of Sentinel-2 data for drought monitoring: Demonstrating the effectiveness of vegetation indices derived from satellite imagery in capturing drought conditions. https://www.researchgate.net/publication/373890243_Exploring_the_Potential_Use_of_Sentinel-1_and_2_Satellite_Imagery_for_Monitoring_Winter_Wheat_Growth_under_Agricultural_Drought_Conditions_in_North-Western_Poland
- [25] Rakapatcharawong, M., et al. (2020). Utilizing satellite-based assessments to evaluate the impact of drought on rice yields in Thailand: Showcasing the practical application of remote sensing for agricultural decision-making. <https://www.mdpi.com/2072-4292/12/13/2099>
- [26] Yoon, J., et al. (2020). Conducting a broader assessment of agricultural drought in East Asia using satellite-based indices: Emphasizing the value of remote sensing in regional drought monitoring. <https://www.mdpi.com/2072-4292/12/3/444>

GIS-based Hybrid Machine Learning for Landslide Susceptibility Assessment in Thai Nguyen Province, Vietnam

Thuong Tran
International School
Thai Nguyen University, Vietnam
 tranthanhtuong@tnu.edu.vn

Hoa Trieu
University of Agriculture and Forestry
Thai Nguyen University, Vietnam
 xuanhoatrieu@tuaf.edu.vn

Nathaniel Bantayan
University of the Philippines Los Baños
Philippines
 ncbantayan@up.edu.ph

Abstract— Landslides are a natural hazard frequently occurring in mountainous areas and have been responsible for numerous fatalities and injuries. They were considered the primary cause of natural disasters leading to loss of life in Asia. In Vietnam, landslides have become increasingly common in recent years, particularly in mountainous regions, posing a direct and significant threat to human life and property. To mitigate this risk, improve land-use planning, and establish early warning systems, landslide susceptibility assessment is a valuable tool, but it requires continuous updating and further development with practical applications to address the complexity of landslide events. Given the importance of landslide susceptibility assessment, this study aimed to assess landslide susceptibility using novel hybrid machine learning (ML) techniques. Four ML models have been utilized for landslide susceptibility prediction, including SVM, RF, SVMMMLP, and RFMLP. These models have been constructed based on support vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) algorithms. To develop the models, the study has collected nine conditioning factors (slope, elevation, relief degree of land surface, drainage density, lithology, soil, land use land cover, rainfall, normalized difference vegetation index) relied on the correlation between historical landslide occurrences and specific geo-environmental conditions in the study area. The relative importance of these factors has been determined by excluding that factor and then calculating the overall accuracy of the model. The difference in overall accuracy between the models with and without the conditioning factor indicates the quantitative importance of the factor. The performance of the models has been assessed using different statistical measures, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC). Typically, higher values of these metrics indicate better performance of the models. The results have shown that the RFMLP model has been the most appropriate model with the best performance, with an accuracy of 85.33%, sensitivity of 85.71%, and specificity of 83.33%. Additionally, the RFMLP model has achieved the highest AUC value (86%), indicating that it has been the most effective model for predicting landslide susceptibility. Based on the results, the RFMLP model has been used to produce a landslide susceptibility map for the study area. This study has provided a feasible approach for assessing landslide susceptibility using hybrid ML models. Its significance lies in its practical applications for disaster risk reduction. The generated susceptibility map, particularly tailored to Thai Nguyen Province, will aid local authorities in effective land-use planning and disaster management. By utilizing advanced ML

techniques and continuously updating these maps, authorities can proactively mitigate landslide risks, safeguarding lives and properties in regions prone to this natural hazard.

Keywords— *Landslide susceptibility, ML, RF, MLP, SVM, RFMLP, SVMMMLP*

I. INTRODUCTION

Landslides pose significant geological hazards globally, endangering lives and property [1]. About 14% of all victims of natural disasters, especially in hilly regions, were due to landslides [2], [3]. Landslides not only cause physical devastation but also lead to short- and long-term psychological impacts due to losses of family members, property, livestock, and crops. Moreover, critical services like water, electricity, and communication infrastructure are often disrupted by landslides, affecting the healthcare system and community well-being.

Landslide susceptibility maps are vital in preventing disaster losses and developing sustainable land use because they can help identify areas susceptible to landslides [4]. Based on these maps, landslide mitigation and prevention can be achieved by building landslide defense structures or moving local communities and industrial facilities to safer areas. Therefore, susceptibility mapping is considered the crucial step in landslide hazard management.

Vietnam is one of the most disaster-prone countries in the world, suffers from many different types of natural hazards including both hydrometeorological (e.g. typhoon, floods, heavy rainfalls and droughts) and geophysical (e.g. landslides) [5]. It was indicated that 22% of the country's landslide occurs during monsoon season and was triggered mainly by rainfall [6].

In recent years, landslides have occurred frequently in the country, especially in the Northern mountainous regions [7]. According to the survey reports from Vietnamese Data Center for Hydrometeorology and Vietnam Disaster Management Authority, more than 800 landslides have been recorded in 14 Northern mountainous provinces over the past two decades, causing 748 fatalities, the destruction of 52,544 houses, and the displacement of 3,910 households.

Thai Nguyen province is situated in the center of the Northern Midland and Mountainous Region, one of the most landslide-prone regions in Vietnam [8], [9]. In the past two decades, Thai Nguyen Province has faced significant devastation from landslides, resulting in 110 fatalities and extensive damage to numerous houses [10].

Various approaches have been developed for GIS-based landslide susceptibility mapping, categorized into inventory-based, index-based, statistically based, and deterministic approaches. Deterministic methods offer high precision but are limited to areas with uniform geology. Index-based methods are less precise and rely on expert judgment. Statistically based models, favored for medium-scale assessments, analyze relationships between instability factors and landslides objectively but require extensive data collection [11]. Hence, new methods such as machine learning (ML) have been proposed and gained popularity for their ability to handle multiple factors, offer statistical objectivity, and allow for continuous updates in landslide modeling [11], [12].

Many models based on ML techniques have been applied for landslides susceptibility mapping such as support vector machine [13], [14]; random forest [15], [16]; fuzzy logic [17]; artificial neural networks [18]; neuro-fuzzy and adaptive neuro-fuzzy inference system [19]; decision tree [20]; generalized additive model and quadratic discriminant analysis [13]; adaBoost [21]; naïve bayes [22]; kernel logistic regression [23]; boosted regression tree; classification and regression tree; multivariate adaptive regression splines; maximum entropy [24]; etc.

In recent years, a shift towards the integration of multiple ML models has been observed alongside the utilization of individual models. research conducted by authors referenced in [25], [12], [26], etc., exemplifies hybrid approaches. However, determining the most effective method for all regions remains inconclusive [19]. A model effective in one area may not perform well in another due to the intricate nature of landslides. Therefore, ongoing research into novel models or enhancements of existing ones remains crucial.

This work aims to assess landslide susceptibility in Cau River Watershed in Thai Nguyen province (CRWTN) using GIS-based hybrid machine learning models. By integrating ML techniques and GIS, the study seeks to: (i) Propose hybrid ML models for landslide susceptibility assessment; (ii) Evaluate the effectiveness of ML models used; (iii) Identify landslide-prone areas within CRWTN.

II. STUDY AREA

Cau River Basin in the Middle Mountains Areas of Northern Vietnam covers almost Thai Nguyen province and the southern region of Bac Kan province. The basin covers a total area of about 3257 km², of which about 2808 km² belongs to Thai Nguyen province. This study focuses on the area of Cau River Watershed in Thai Nguyen province.

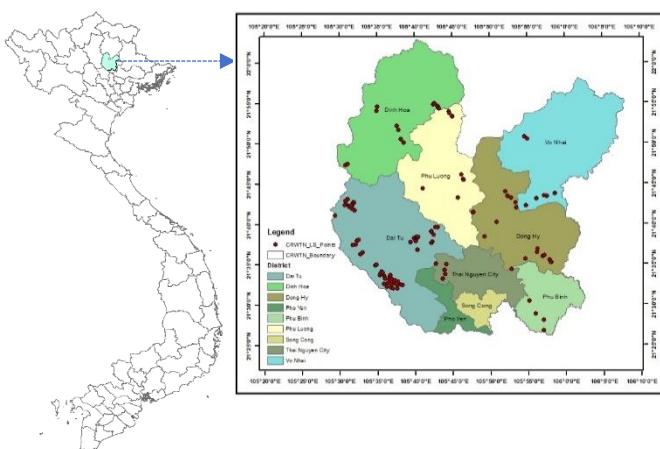


Fig. 1. Landslide inventory in CRWTN

The study area is located between $21^{\circ}26'8''N$ and $22^{\circ}2'54''N$ latitudes, $105^{\circ}28'36''E$ and $106^{\circ}7'41''E$ longitudes and covers 79.6 % area of Thai Nguyen province (3527 km²). CRWTN covers more than two-thirds of Thai Nguyen province (fig.1). The study area is characterized by all natural characteristics of a mountainous province in tropical monsoon climate zone in Northern Vietnam with the topography of mainly hills and mountains. The majority of the study site exhibits elevations below 300 meters, with approximately two-thirds of this area comprising hills and mountains exceeding 100 meters in elevation [9]. The typical climate characteristic of this region is a tropical monsoon climate with a very cold winter. Each year, the area experiences high levels of rainfall, with the rainy season lasting for 6 months, beginning in May and extending through October annually. During this period, heavy and prolonged rainfall is concentrated primarily in the months of June, July, and August, with maximum daily precipitation ranging from approximately 208 to 496 mm. The annual rainfall in this region ranges from 1,360 mm to 2,572 mm. Landslides have frequently occurred in this area, as illustrated by the inventory map in fig. 1.

III. METHODOLOGY

A. Methodology flowchart

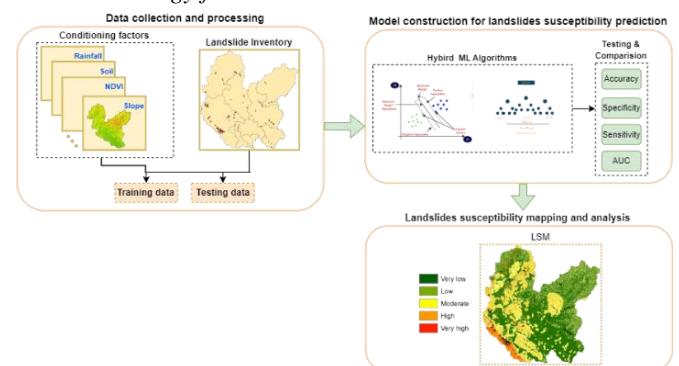


Fig. 2. Methodology flowchart

The workflow of the study contains three main phases (fig. 2), including (1) Data collection and processing; (2). Model construction for landslide susceptibility prediction. The landslide models have been constructed based on ML algorithms and then trained over the training dataset, which has been subsequently evaluated for performance using a separate dataset that was not used in the training phase. (3) Landslide susceptibility mapping and analysis. The resulting model has been then applied to generate the landslide susceptibility index (LSI) for the entire study area. LSI obtained from the ML model has then been imported into a GIS tool to produce a landslide susceptibility map.

B. Inventory map

An inventory map for the CRWTN has been constructed using collected landslide data, primarily from the flash flood and landslide inventory data provided in a report by the Thai Nguyen Department of Science and Technology [27] and the Global Landslide Catalog. Other sources of data have included previous projects, remote sensing data, and field trips. The investigated data has been used to create a landslide inventory map for the study area (fig. 1).

C. Landslide conditioning factors

The landslide susceptibility has affected by many factors, depending on specific conditions, the influencing levels of these factors are different. In this work, slope, elevation, relief

degree of land surface (RDLS), drainage density, lithology, soil type, normalized difference vegetation index (NDVI), rainfall, and land use land cover (LULC) are the conditional factors that are considered in which the rainfall factor was taken according to the average rainfall value of the rainy season months (May to October) from 2010 to 2020 (fig. 3).

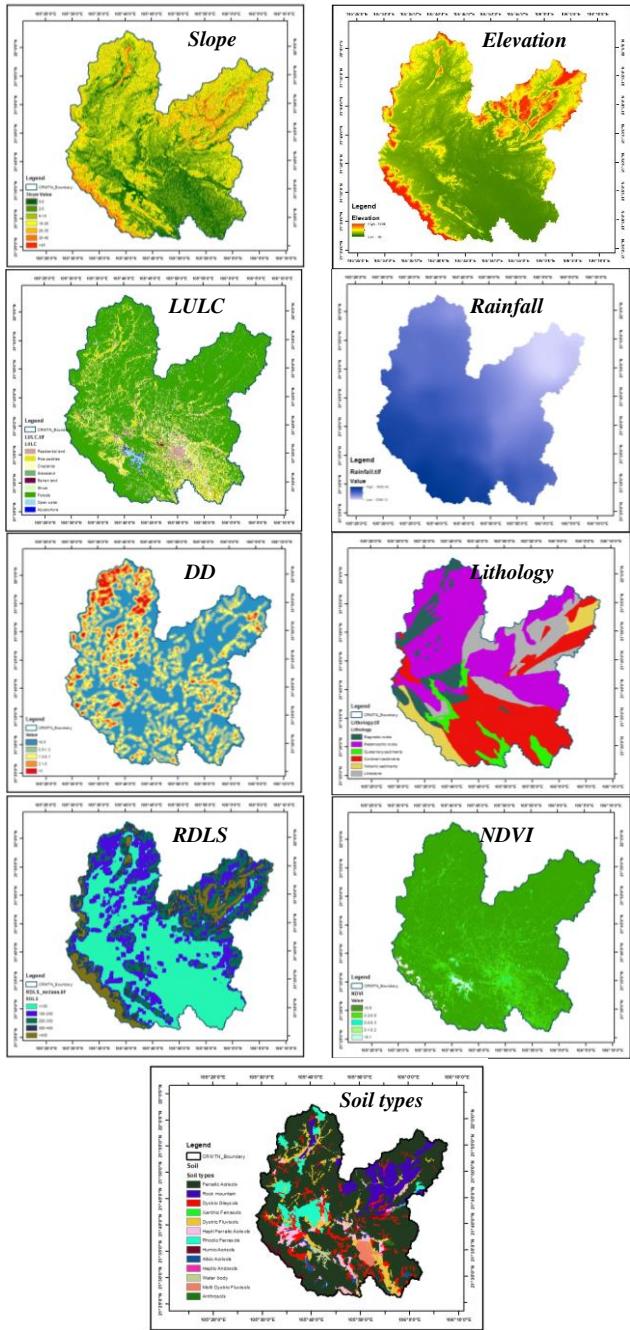


Fig. 3. Landslide conditioning factors

Evaluation and analysis of the relationship of factor classes to landslide susceptibility have been implemented based on the relative importance of each conditioning factor that has been determined by excluding that factor and then calculating the overall accuracy of the model [28]. The difference in overall accuracy between the models with and without the conditioning factor indicates the quantitative importance of the factor.

D. Landslide susceptibility modeling

In addition to RF and SVM, two hybrid ML models, namely RFMLP and SVMMLP have been proposed to predict landslide susceptibility. These models have been developed using three algorithms, including RF, SVM, and MLP. The input data of these models has been created using the landslide inventory and LCFs. The historical landslide dataset employed in this work consisted of 125 landslide events. This dataset has been divided into two subsets for training and testing landslide models in which the first subset has been employed for constructing landslide models whilst the second one has been used for evaluating landslide models and estimating their accuracy. All grid cells in the dataset referring to landslides have been assigned the value of 1. An equal number of grid cells referring to non-landslide has been randomly sampled from the landslide-free area and assigned a value of 0. Values of the LCFs have been then extracted to create the training and testing datasets. These datasets have contained one label variable referring to the landslide presence/absence and values of nine LCFs. Moreover, to ensure a desirable classification model, this research employs fivefold cross-validation. Since the performance of the model in the training process alone may not be sufficient to indicate a good model due to the problem of over-fitting. It is noted that an over-fitted model simply memorizes the input–output mapping provided in the training dataset instead of generalizing it. As a consequence, the model may classify the training set very well, but poorly predict new input patterns which are not included in the training set. The fivefold cross-validation divides the dataset into five parts, training on four and testing on the fifth, repeated five times for reliable performance estimation.

1) Random forest (RF)

To improve the performance of the random forest algorithm, the grid search technique can be used to find the optimal hyperparameters for the model. Hyperparameters are settings that are not learned during the training process but rather are set by the user before training. Hyperparameter tuning can be achieved through random search or grid search. In random search, hyperparameter values are chosen independently based on a probability distribution. Conversely, grid search is a process that searches exhaustively through a manually specified subset of the hyperparameter space of the targeted algorithm. Grid search involves testing a range of hyperparameters to identify the combination that produces the best performance on a validation set. Some hyper-parameters need to be tuned in RF models, including (i) the number of trees to be combined; (ii) the maximum depth of the trees; and (iii) the maximum number of features considered at each split.

2) Support vector machine (SVM)

The SVM algorithm involves training a model on a training set using a kernel function and finding the hyperplane that best separates the data into different classes. In addition to kernel function, the performance of a SVM model also depend upon on other hyper-parameters, such as C-value and Gamma [29].

(i) Kernel functions: In machine learning, a ‘kernel’ is used as a linear classifier to solve a non-linear problem. It transforms the input data into a high-dimensional space, allowing the model to achieve the desired output. After

training in the high-dimensional space, the results are brought back into the original 2-dimensional space. Mathematically, a kernel function is expressed as: $K(x, y) = \{f(x), f(y)\}$ where, K is the kernel function, x and y are n -dimensional inputs, f is a map from n -dimension to m -dimension space, and $\{x, y\}$ denotes the dot product.

(ii) C-value (penalty function or regularization parameter): It controls the tradeoff between model complexity and the amount of error allowed in the model. A larger value of C will result in a more complex model that fits the training data more closely, while a smaller value of C will result in a simpler model with more error [30]. A reasonable approach is to assign a C -value equal to the range of output values [31].

(iii) Gamma (γ): This is the kernel coefficient (kernel parameter). The higher the value of gamma, the closer the fit to the training data and vice versa.

3) Multilayer perceptron (MLP)

A fully connected multi-layer neural network is called a Multilayer Perceptron (MLP). It is a supervised learning algorithm that learns a function $f : R^m \rightarrow R^o$ by training on a dataset, where m is the number of dimensions for input and o is the number of dimensions for output. The MLP network in this landslide susceptibility model has three main parts, including one input layer, three hidden layers, and one output layer. The input layer uses nine variables, each of which corresponds to one landslide conditioning factor. In the network, each neuron receives input from the neurons in the previous layer and computes a weighted sum of the inputs (fig. 4). During the training process, the weights and biases of the neurons in the model are adjusted using the backpropagation algorithm to minimize the difference between the predicted output and the actual output.

4) Hybrid ML models

In the proposed hybrid models in this work, RFMLP and SVVMLP, the first machine learning algorithm is served to augment training data for the model, while the second algorithm is employed to predict landslide susceptibility. In these hybrid models, either Random Forest or Support Vector Machine is used as the first classifier, while a Multi-Layer Perceptron acts as the second classifier. The ML model has been first trained using the input data from the training dataset. The resulting model outputs have been processed and combined with the original training dataset to create the input for the model in the next phase. Then, the model has been retrained using the updated input data, and its performance has been evaluated. This approach has been applied to enhance the dataset and improve the accuracy of the model. The outputs of these models are utilized to generate the final prediction of landslide susceptibility.

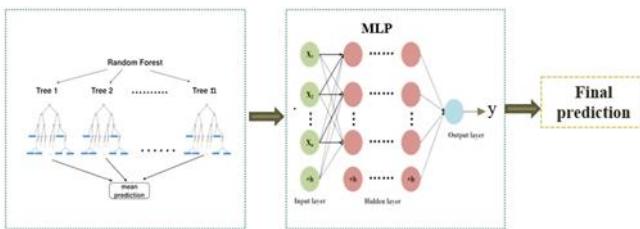


Fig. 4. Hybrid ML model architecture for landslide susceptibility prediction using RF and MLP

E. Model evaluation

In this study, some statistical criteria, including Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and Area under the ROC curve (AUC), have been used to evaluate the applied prediction models. Accuracy (ACC) is the overall performance of the model. This quality criterion refers to how close a measurement is to the true or accepted value [32]. Sensitivity (SEN) refers to the ability of a model to correctly identify the positive class [32]. Specificity (SPE) refers to the ability of a model to correctly identify the negative class [33]. AUC, which is the area under the ROC curve, is used to judge the performance of a model in which a higher AUC value indicates better performance of the model in distinguishing between positive and negative classes with the ROC curve being plotted based on the sensitivity and (100-specificity) on the x- and y-axis, respectively. Specifically, a value of 1 indicates an accurate model, while an AUC equal to 0.5 is an inaccurate model [33].

Typically, higher values of ACC, SEN, SPE, and AUC indicate better performance of the models [34]. Additionally, these models' results are compared to each other.

The evaluation metrics can be calculated using the following formulas.

$$\text{ACC} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{SEN} = \frac{TP}{TP + FN}$$

$$\text{SPE} = \frac{TN}{TN + FP}$$

where,

- TP (True Positive) refers to the proportion of positive cases that were correctly predicted (landslides correctly classified as landslides).
- FP (False Positive) refers to the proportion of negative cases incorrectly predicted as positive (non-landslides incorrectly classified as landslides).
- TN (True Negative) refers to the proportion of correctly predicted negative cases (non-landslides correctly classified as non-landslides).
- FN (False Negative) refers to the proportion of positive cases incorrectly predicted as negative (landslide incorrectly classified as non-landslide).

P and N are defined as the total number of landslide and non-landslide samples, respectively.

IV. RESULTS AND DISCUSSION

A. Relative importance of landslide conditioning factors

The relative importance of the conditioning factors in a landslide model can vary depending on the methods and criteria used for evaluation. Therefore, a factor with a high contribution in one model may not be useful in another model, and vice versa. The importance of a conditioning factor may have considerable variability. In this study, the relative importance and contribution of each conditioning factor were determined by excluding that factor and then calculating the overall accuracy of the model. The difference in overall accuracy between the models with and without the conditioning factor indicates the quantitative importance of the factor. The relative importance of conditioning factors to

landslide occurrences determined by the ML models has been presented in fig. 5. It was suggested all selected factors had an impact on landslides, of which occurrences, slope, rainfall, and NDVI had a more significant impact compared to other factors. Landslides in Thai Nguyen can be classified into three main categories based on major causing types: rainfall on construction sites, rainfall on natural slopes, and rainfall and water drawdown on banks of reservoirs or streams [9]. Additionally, according to the forest inventory data of the Thai Nguyen Forest Protection Department, the forest covers an area of more than 1575 hectares of CRWTN, which accounts for 56.1% of the study area. This information may help explain the significance of slope, rainfall, and NDVI factors in landslide occurrences in the area.

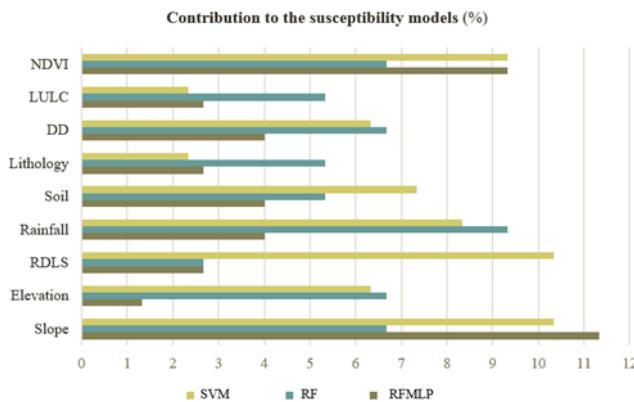


Fig. 5. Relative importance of conditioning factors to landslide occurrences

B. Landslide susceptibility mapping

Two hybrid models (SVMMLP and RFMLP) have been constructed for landslide susceptibility prediction and the cross-validation technique has been utilized to verify the performance of these models. Additionally, the Random Forest and SVM model have been implemented to test their performance compared to the hybrid method. Table 1 provides an overview of the performance of these models in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE).

TABLE I. THE PERFORMANCE OF RF, SVM, SVMMLP, AND RFMLP

Model	RF	SVM	SVMMLP	RFMLP
ACC	78.66%	76%	80%	85.33%
SEN	79%	76%	80%	85.71%
SPE	79%	76%	81%	83.33%

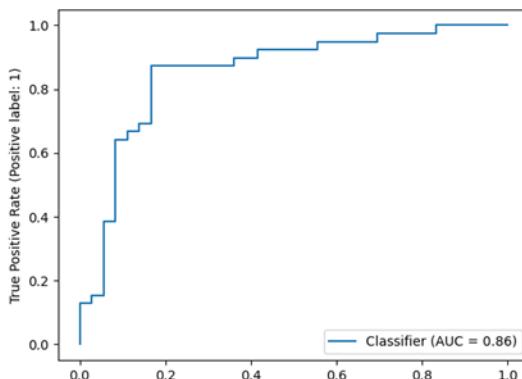


Fig. 6. ROC curves and AUC values of RFMLP model

Accordingly, the hybrid models outperformed the individual model in terms of accuracy, sensitivity, and specificity, providing evidence for the effectiveness of the hybrid models developed in this study. Among the models used, the RFMLP model showed the highest accuracy, and therefore it was selected to produce an appropriate landslide susceptibility map for CRWTN. The ROC curves and AUC values of the RFMLP model have been illustrated in fig 6 with its AUC value of 86%.

Based on the statistical indicators analyzed above, the RFMLP model provided the best prediction among models used. As a result, this model was selected to produce the most appropriate susceptibility map for the study area. The resulting map was classified into five categories: very low, low, moderate, high, and very high susceptibility (fig 7).

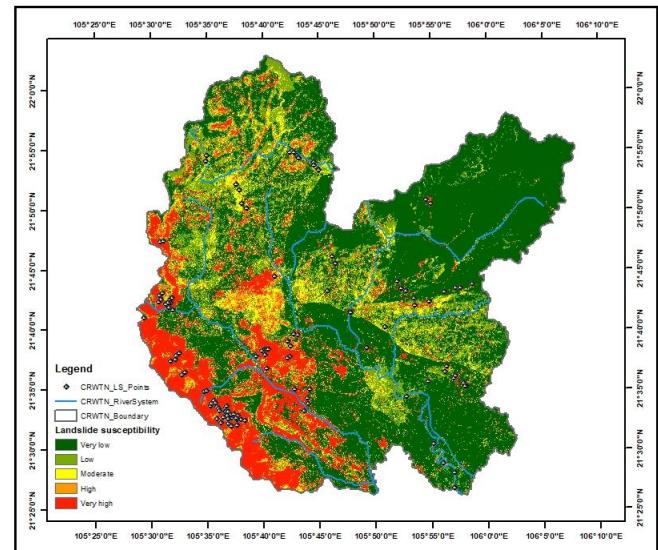


Fig. 7. Landslide susceptibility map for CRWTN produced by the RFMLP model

The final results represented 41,973 ha (14.94%) in very high-susceptibility areas, 14,258 ha (5.08%) in high-susceptibility areas, 22,159 ha (7.89%) in moderate-susceptibility areas, 48,334 ha (17.21%) in low-susceptibility areas and 154,129 ha (54.88%) in very low-susceptibility areas.

C. Discussion

Landslides are hazardous natural phenomena that often occur in mountainous areas [34]. They resulted in a high number of fatalities and injuries each year and were considered the primary cause of natural disasters leading to loss of life in Asia [35]. As a result, it is crucial to continually update information on areas susceptible to landslides due to their complex geological processes [36].

Machine learning techniques are regarded as a promising and efficient approach for evaluating natural hazards due to their adaptability and high level of predictive accuracy [37]. Previously, single models, such as the MLP model, were often applied to build landslide susceptibility maps [38], [39]. Several research studies combined MLP with other ML models, demonstrating improved predictive accuracy. Particle Swarm Optimization and MLP were used to generate a landslide susceptibility map for Shicheng County in China with ACC of 88.1% [40]. In another investigation, an ensemble of Stochastic Gradient Descent and MLP to forecast landslide-prone areas for Yanshan County in Jiangxi province in China, obtained an ACC of 82.2% [41]. The RFMLP model with Rotation Forest and

MLP achieved the predictive accuracy of 81.8% in landslide susceptibility prediction mapping for Son La, Vietnam [42].

This study presented a practical method for assessing landslide susceptibility using machine learning techniques. ML models were proposed to generate predictive landslide susceptibility maps for the CRWTN case study. Among used models, RFMLP was identified as the most suitable option, demonstrating a higher accuracy (85.33%) compared to previous research. This result shown the effectiveness of the RFMLP model and suggested its practical implications for evaluating landslide risk in other regions. However, depending on the specific conditions of the target site, the landslide conditioning factors will be studied to select appropriate factors. This finding suggested a novel and efficient hybrid model for evaluating landslide susceptibility in mountainous regions and its results can assist local authorities in effective land-use planning and disaster management. Additionally, this work focused on utilizing static data. For future work, real-time data (e.g. rainfall) can be considered to integrate into the model to develop a dynamic early warning system for landslides.

V. CONCLUSION

Machine learning techniques were employed in this study for landslide susceptibility assessment at CRWTN. In addition to Random Forest and Support Vector Machine, two hybrid models, namely SVMMLP and RFMLP, were developed for predicting landslide susceptibility. Among the tested models, the RFMLP model exhibited the highest performance in predicting landslide susceptibility with an accuracy of 85.33%. Consequently, the RFMLP model was chosen for generating a comprehensive landslide susceptibility map. Besides, the SVMMLP hybrid model also outperformed the individual RF and SVM models in terms of accuracy, sensitivity, and specificity, underscoring the effectiveness of the hybrid proposed models. The predictive map established from the RFMLP model output has provided valuable insights for analyzing and evaluating landslide susceptibility at a local scale.

REFERENCES

- [1] S. Lee, "Current and future status of gis-based landslide susceptibility mapping: a literature review," *Korean Journal of Remote Sensing*, vol. 35, no. 1, pp. 179–193, 2019.
- [2] A. B. Shirzadi, "Shallow landslide susceptibility assessment using a novel hybrid intelligence approach," *Environmental Earth Sciences*, vol. 76, no. 2, p. 60, 2017.
- [3] M. J. Froude, "Global fatal landslide occurrence from 2004 to 2016," *Natural Hazards and Earth System Sciences*, vol. 18, no. 8, 2018.
- [4] W. P. Chen, "Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility," *Catena*, vol. 172, pp. 212–231, 2019.
- [5] WHO, "Disaster in vietnam," <https://www.who.int/vietnam/health-topics/disasters#>, 2021, retrieved October 11, 2021.
- [6] T. V. Phong, "Landslide susceptibility modeling using different artificial intelligence methods: A case study at muong lay district, vietnam," *Geocarto International*, vol. 36, no. 15, pp. 1683–1708, 2021.
- [7] D. T. Bui, "A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the lang son city area (vietnam)," *Geomatics, Natural Hazards and Risk*, vol. 6, no. 3, pp. 243–271, 2015.
- [8] D. T. Tien Bui, "Spatial prediction of rainfall-induced landslides for the lao cai area (vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization," *Landslides*, pp. 447–458, 2017.
- [9] T. T. Le, "Landslide detection analysis in north vietnam based on satellite images and digital geographical information-landsat 8 satellite and historical data approaches," *Japanese Journal of JSCE*, vol. 73, no. 5, pp. I 239–I 249, 2017.
- [10] VNDMA, *The Vietnam 20-year floods and landslides handbook*. Vietnam Disaster Management Authority, 2020.
- [11] D. T. Bui, "Landslide susceptibility assessment in the hoa binh province of vietnam: a comparison of the levenberg-marquardt and bayesian regularized neural networks," *Geomorphology*, vol. 171, pp. 12–29, 2012.
- [12] A. M. Youssef and H. R. Pourghasemi, "Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at abha basin, asir region, saudi arabia," *Geoscience Frontiers*, vol. 12, no. 2, pp. 639–655, 2021.
- [13] H. R. Pourghasemi, "Prediction of the landslide susceptibility: Which algorithm, which precision?" *Catena*, vol. 62, pp. 177–192, 2018.
- [14] V. K. Pandey, "Landslide susceptibility mapping using maximum entropy and support vector machine models along the highway corridor, garhwal himalaya," *Geocarto International*, vol. 35, no. 2, pp. 168–187, 2020.
- [15] H. R. Pourghasemi, "Land-subsidence spatial modeling using the random forest data-mining technique," in *Spatial modeling in GIS and R for Earth and environmental sciences*. Elsevier, 2019, pp. 147–159.
- [16] O. K. Rahmati, "Spatial prediction of landslide susceptibility using random forest algorithm," in *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation*. Springer, Singapore, 2021.
- [17] A. B. Abdi, "Landslide susceptibility mapping using gis-based fuzzy logic and the analytical hierarchical processes approach: A case study in constantine (north-east algeria)," *Geotechnical and Geological Engineering*, vol. 39, no. 8, pp. 5675–5691, 2021.
- [18] D. J.-G. Van Dao, "A spatially explicit deep learning neural network model for the prediction of landslide susceptibility," *Catena*, vol. 188, 2020.
- [19] B. T. Pham, "A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area," *Geotechnical and Geological Engineering*, vol. 34, no. 6, pp. 1807–1824, 2016.
- [20] N. F. Micheletti, "Machine learning feature selection methods for landslide susceptibility mapping," *Mathematical Geosciences*, vol. 46, 2014.
- [21] W. Z. Chen, "Performance evaluation of the gis-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling," *Science of the Total Environment*, vol. 644, pp. 1006–1018, 2018.
- [22] W. Chen, H. Shahabi, S. Zhang, K. Khosravi, A. Shirzadi, K. Chapi, B. T. Pham, T. Zhang, L. Zhang, H. Chai et al., "Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression," *Applied Sciences*, vol. 8, no. 12, p. 2540, 2018.
- [23] A. M.-K. Youssef, "Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at wadi tayyah basin, asir region, saudi arabia," *Landslides*, vol. 13, no. 5, 2016.
- [24] A M. Felic'ísimo, A. Cuartero, J. Remondo, and E. Quiro's, "Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study," *Landslides*, vol. 10, pp. 175–189, 2013.
- [25] D. T. Bui, "Comparing the prediction performance of a deep learning neural network model with conventional machine learning models in landslide susceptibility assessment," *Catena*, vol. 188, p. 104426, 2020.
- [26] X. Chen, "Gis-based landslide susceptibility assessment using optimized hybrid machine learning methods," *Catena*, vol. 196, 2021.
- [27] T. V. Khanh, "Application of a weighted model for landslide mapping," *TNU Journal of Science and Technology*, vol. 226, no. 07, 2021.
- [28] D. T. Bui, "Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree," *Landslides*, vol. 13, no. 2, pp. 361–378, 2016.
- [29] D. P. Basak, "Support vector regression," *Neural Information Processing-Letters and Reviews*, pp. 203–224, 2007.
- [30] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
- [31] V. Cherkassky, "Selection of meta-parameters for support vector regression," in *Artificial Neural Networks—ICANN 2002: International Conference Madrid, Spain, August 28–30, 2002 Proceedings*. Springer Berlin Heidelberg, 2002, pp. 687–693.
- [32] G. W. James, *An Introduction to Statistical Learning*. New York: Springer, 2013, vol. 112.
- [33] P. T. Nguyen, "Development of a novel hybrid intelligence approach for landslide spatial prediction," *Applied Sciences*, vol. 9, no. 14, 2019.
- [34] K. S ilha'n, "Dendrogeomorphology of landslides: principles, results and perspectives," *Landslides*, vol. 17, no. 10, pp. 2421–2441, 2020.
- [35] L.Highland, *The Landslide Handbook - A guide to understanding landslides*. Reston, VA, USA: US Geological Survey, 2008.
- [36] W. L. Y. Chen, "Gis-based evaluation of landslide susceptibility using hybrid computational intelligence models," *Catena*, 2020.
- [37] D. S. Tien Bui, "Shallow landslide prediction using a novel hybrid functional machine learning algorithm," *Remote Sensing*, vol. 11, no. 8, p. 931, 2019.
- [38] M. S. Adnan, "Improving spatial agreement in machine learning-based landslide susceptibility mapping," *Remote Sensing*, vol. 12, no. 20, p. 3347, 2020.
- [39] D. M. Meghanad, "A multi-criteria landslide susceptibility mapping using deep multi-layer perceptron network: A case study of srinagar-rudraprayag region (india)," *Advances in Space Research*, vol. 69, no. 4, pp. 1883–1893, 2022.
- [40] D. Li, F. Huang, L. Yan, Z. Cao, J. Chen, and Z. Ye, "Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, bp neural network, and information value models," *Applied Sciences*, vol. 9, no. 18, p. 3664, 2019.
- [41] H. Hong, "Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping," *Science of the total environment*, vol. 742, p. 140549, 2020.
- [42] Q. D. Bui, "Landslide susceptibility prediction mapping with advanced ensemble models: Son la province, vietnam," *Natural Hazards*, pp. 1, 2022.



รายละเอียดแนบท้าย 1

ประกาศมหาวิทยาลัยพะเยา เรื่อง กำหนดปฏิทินการศึกษา ปีการศึกษา 2567

สำหรับนิสิตระดับปริญญาตรี ภาคปกติ

กิจกรรม	ภาคการศึกษาต้น	ภาคการศึกษาปลาย	ภาคฤดูร้อน
คณะกรรมการรายวิชาตามแผนการศึกษาที่เปิดสอน ถึง กองบริการการศึกษา	27 ก.พ. 67	12 ก.ค. 67	27 ธ.ค. 67
วันสุดท้ายของการรับเอกสารขอโอนย้ายสถานศึกษา	10 พ.ค. 67	20 ก.ย. 67	-
นิสิตยื่นคำร้องขอเทียบโอนรายวิชาและผลการเรียนจากสถานศึกษาเดิม (UP09)	10 พ.ค.-10 มิ.ย. 67	20 ก.ย.- 21 ต.ค. 67	27 ม.ค.-10 มี.ค. 68
คณะกรรมการรับคำร้องขอเทียบโอนรายวิชาและผลการเรียนจากสถานศึกษาเดิม (UP09)	10 พ.ค.-11 มิ.ย. 67	20 ก.ย.- 18 ต.ค. 67	27 ม.ค.-11 มี.ค. 68
นิสิตยื่นคำร้องขอย้ายคณะ หลักสูตร และแผนการศึกษา ถึง คณะต้นสังกัดของนิสิต (UP10)	22 เม.ย.-3 พ.ค. 67	2-13 ก.ย. 67	20-31 ม.ค. 68
คณะกรรมการตั้งสังกัดส่งคำร้องขอย้ายคณะ หลักสูตร และแผนการศึกษา ไปยัง คณะที่รับย้าย	7-10 พ.ค. 67	16-20 ก.ย. 67	3-7 ก.พ. 68
คณะกรรมการตั้งสังกัดส่งคำร้องขอย้ายคณะ หลักสูตร และแผนการศึกษา ไปยัง กองบริการการศึกษา	13-17 พ.ค. 67	23 - 27 ก.ย. 67	10-14 ก.พ. 68
คณะกรรมการตรวจสอบรายวิชาตามแผนการศึกษา	20-24 พ.ค. 67	30 ก.ย.-4 ต.ค. 67	17-21 ก.พ. 68
วันประกาศตารางเรียนตารางสอน	27 พ.ค. 67	7 ต.ค. 67	24 ก.พ. 68
วันพบอาจารย์ที่ปรึกษา	27 พ.ค. 67	7 ต.ค. 67	24 ก.พ. 68
นิสิตยื่นความประสงค์ลงทะเบียนรายวิชาตกลงแผนการศึกษาผ่านระบบ (UP20.1)	27 พ.ค.- 19 มิ.ย. 67	15-19 ต.ค. 67	24 ก.พ.-12 มี.ค. 68
นิสิตเตรียมลงทะเบียนเรียน	27-31 พ.ค. 67	7-11 ต.ค. 67	24-28 ก.พ. 68
วันลงทะเบียนเรียน ลงทะเบียนรักษาสภาพการเป็นนิสิต (วันแรกเริ่มลงทะเบียนเวลา 09.00 น.)			
นิสิตรหัส 67	4-7 มิ.ย. 67	15-18 ต.ค. 67	3-7 มี.ค. 68
นิสิตรหัส 66	4-7 มิ.ย. 67	16-18 ต.ค. 67	4-7 มี.ค. 68
นิสิตรหัส 65	5-7 มิ.ย. 67	17-18 ต.ค. 67	5-7 มี.ค. 68
นิสิตรหัส 64 และน้อยกว่า	6-7 มิ.ย. 67	17-18 ต.ค. 67	6-7 มี.ค. 68
ชำระเงินค่าลงทะเบียน (เริ่มชำระได้ตั้งแต่นิสิตยังไม่ได้รับอักษร W)	4-7 มิ.ย. 67	15-18 ต.ค. 67	3-7 มี.ค. 68
นิสิตยื่นคำร้องขอผ่อนผันการลงทะเบียนเรียน (UP 18)	4-7 มิ.ย. 67	15-18 ต.ค. 67	3-7 มี.ค. 68
คณะกรรมการรับคำร้องขอผ่อนผันการลงทะเบียนเรียน (UP 18) ถึง กองบริการการศึกษา	4-10 มิ.ย. 67	15-21 ต.ค. 67	3-10 มี.ค. 68
คณะกรรมการรับแบบฟอร์มการขอเปลี่ยนแปลงตารางเรียนตารางสอน ถึง กองบริการการศึกษา (UP20)	5-20 มิ.ย. 67	17 - 31 ต.ค. 67	5-13 มี.ค. 68
คณะกรรมการรับแบบฟอร์มการขอเปิดรายวิชา ถึง กองบริการการศึกษา (UP21)	5-20 มิ.ย. 67	17 - 31 ต.ค. 67	5-13 มี.ค. 68
วันเปิดภาคการศึกษา	10 มิ.ย. 67	21 ต.ค. 67	10 มี.ค. 68
ลงทะเบียนเรียนล่าช้าและลงทะเบียนรักษาสภาพการเป็นนิสิต,เพิ่ม-ถอนรายวิชาไม่ได้รับอักษร W			
นิสิตรหัส 65-67 (วันแรกเริ่มลงทะเบียนเวลา 13.00 น.)	10-21 มิ.ย. 67	21 ต.ค.-1 พ.ย. 67	10-14 มี.ค. 68
นิสิตรหัส 64 และน้อยกว่า (วันแรกเริ่มลงทะเบียนเวลา 09.00 น.)	11-21 มิ.ย. 67	22 ต.ค.-1 พ.ย. 67	11-14 มี.ค. 68
ชำระเงินค่าลงทะเบียนเรียนทุกชั้นปี (สำหรับนิสิตที่ลงทะเบียนล่าช้า)	11-21 มิ.ย. 67	22 ต.ค.-1 พ.ย. 67	11-14 มี.ค. 68
นิสิตยื่นคำร้องขอผ่อนผันการลงทะเบียนเรียน (UP 18) (สำหรับนิสิตที่ลงทะเบียนล่าช้า)	11-21 มิ.ย. 67	21 ต.ค.-1 พ.ย. 67	11-14 มี.ค. 68
คณะกรรมการรับคำร้องขอผ่อนผันการลงทะเบียนเรียน (UP 18) (สำหรับนิสิตที่ลงทะเบียนล่าช้า)	11-24 มิ.ย. 67	21 ต.ค.-4 พ.ย. 67	11-17 มี.ค. 68
นิสิตยื่นคำร้องขอสำเร็จการศึกษา (UP 24) และชำระเงินค่าขึ้นทะเบียนปริญญาบัตร	10 มิ.ย.-9 ก.ค. 67	21 ต.ค.-19 พ.ย. 67	10 มี.ค.-8 เม.ย. 68
วันสุดท้ายคณะรับคำร้องขอสำเร็จการศึกษา (UP 24)	16 ก.ค. 67	29 พ.ย. 67	23 เม.ย. 68
วันสุดท้ายคณะรวมรายชื่อขอสำเร็จการศึกษา (UP 24) ถึง กองบริการการศึกษา	19 ก.ค. 67	4 ธ.ค. 67	25 เม.ย. 68



รายละเอียดแนบท้าย 1

ประกาศมหาวิทยาลัยพะเยา เรื่อง กำหนดปฏิทินการศึกษา ปีการศึกษา 2567

สำหรับนิสิตระดับปริญญาตรี ภาคปกติ

กิจกรรม	ภาคการศึกษาต้น	ภาคการศึกษาปลาย	ภาคฤดูร้อน
วันไหว้ครู (งดการเรียนการสอนทุกชั้นปี)*			
วันสุดท้ายของการส่งรายงานผลขอแก้ไขอักษร ถึง คณะเจ้าของรายวิชา	5 ก.ค. 67	15 พ.ย. 67	11 เม.ย. 68
วันสุดท้ายของการส่งรายงานผลขอแก้ไขอักษร ถึง กองบริการการศึกษา	9 ก.ค. 67	19 พ.ย. 67	8 เม.ย. 68
นิสิตยื่นคำร้องขอถอนรายวิชาโดยได้รับอักษร W (UP 08)	13 ก.ย. 67	17 ม.ค. 68	25 เม.ย. 68
วันสุดท้ายคณะส่งคำร้องขอถอนรายวิชาโดยได้รับอักษร W (UP 08) ถึง กองบริการการศึกษา	16 ก.ย. 67	20 ม.ค. 68	28 เม.ย. 68
นิสิตยื่นคำร้องขอสำเร็จการศึกษาล่าช้า (UP 25)	10 ก.ค.- 26 ก.ย. 67	19 พ.ย. 67 - 11 ก.พ. 68	9 เม.ย.-7 พ.ค. 68
คณะรับคำร้องขอสำเร็จการศึกษาล่าช้า (UP 25)	10 ก.ค.- 27 ก.ย. 67	19 พ.ย. 67 - 13 ก.พ. 68	9 เม.ย.-8 พ.ค. 68
วันครบรอบวันสถาปนามหาวิทยาลัย (งดการเรียนการสอนทุกชั้นปี)*			
วันสุดท้ายของการยื่นคำร้องขอคืนสภาพการเป็นนิสิตเพื่อลังเหลบเปลี่ยนเรียน (UP 4) (นิสิต)	26 ก.ค. 67	6 ธ.ค. 67	4 เม.ย. 68
วันสุดท้ายของการยื่นคำร้องขอเพิ่มรายวิชาหลังกำหนด (UP 5) (นิสิต)	26 ก.ค. 67	6 ธ.ค. 67	11 เม.ย. 68
วันสุดท้ายคณะส่งคำร้องขอคืนสภาพการเป็นนิสิตเพื่อลังเหลบเปลี่ยนเรียน (UP 4) ถึง กองบริการการศึกษา	30 ก.ค. 67	9 ธ.ค. 67	17 เม.ย. 68
วันสุดท้ายคณะส่งคำร้องขอเพิ่มรายวิชาหลังกำหนด (UP 5) ถึง กองบริการการศึกษา	30 ก.ค. 67	9 ธ.ค. 67	17 เม.ย. 68
วันสุดท้ายของการชำระค่าลงทะเบียนเบียน กรณีนิสิตที่ได้รับอนุมัติให้ผ่อนผัน (UP 18)	2 ส.ค. 67	13 ธ.ค. 67	11 เม.ย. 68
วันสอบกลางภาค (งดการเรียนการสอนทุกชั้นปี)	3-11 ส.ค. 67	14-22 ธ.ค. 67	-
วันประเมินการสอนผ่านระบบ簿เบียนออนไลน์	16-27 ก.ย. 67	3-14 ก.พ. 68	28 เม.ย.-9 พ.ค. 68
พิธีพระราชทานปริญญาบัตร (งดการเรียนการสอนทุกชั้นปี)*			
วันสุดท้ายของการยื่นคำร้องขอลาพักการศึกษา (UP 13)	27 ก.ย. 67	14 ก.พ. 68	9 พ.ค. 68
วันสุดท้ายของการยื่นคำร้องขอลาออกโดยไม่ได้รับผลการเรียน (UP 14)	27 ก.ย. 67	14 ก.พ. 68	9 พ.ค. 68
วันสอบปลายภาค (งดการเรียนการสอนทุกชั้นปี)	30 ก.ย.-11 ต.ค. 67	17-28 ก.พ. 68	13-16 พ.ค. 68
วันสุดท้ายของการยื่นคำร้องขอแก้ไขข้อ - ลักษณะ (UP 11) ของนิสิตที่คาดว่าจะสำเร็จการศึกษา	11 ต.ค. 67	28 ก.พ. 68	16 พ.ค. 68
วันปิดภาคการศึกษา	15 ต.ค. 67	3 มี.ค. 68	19 พ.ค. 68
วันสุดท้ายของการส่งรายงานผลการเรียนรายวิชา ถึง คณะเจ้าของรายวิชา	25 ต.ค. 67	14 มี.ค. 68	30 พ.ค. 68
วันสุดท้ายของการส่งรายงานผลการเรียนรายวิชา ถึง กองบริการการศึกษา	29 ต.ค. 67	18 มี.ค. 68	4 มิ.ย. 68
วันสำเร็จการศึกษา	15 ต.ค. 67	3 มี.ค. 68	19 พ.ค. 68

หมายเหตุ: * มหาวิทยาลัยจะประกาศให้ทราบในภายหลัง

update 27 มี.ค.67

Identification of Spatial Patterns of Dengue Hemorrhagic Fever Using Spatial Statistics

Danh-tuyen Vu

*Faculty of Surveying, Mapping and Geographic Information
Hanoi University of Natural Resources and Environment
Hanoi, Vietnam
vdtuyen@hunre.edu.vn*

Anh-huy Hoang

*Faculty of Environment
Hanoi University of Natural Resources and Environment
Hanoi, Vietnam
hahuy@hunre.edu.vn*

Tien-thanh Nguyen

*Faculty of Surveying, Mapping and Geographic Information
Hanoi University of Natural Resources and Environment
Hanoi, Vietnam
ntthanh@hunre.edu.vn*

Abstract—Background: Spatial data pertaining to the dengue hemorrhagic fever (DHF) distribution pattern can be obtained through spatial statistics. This study attempts to determine the spatial patterns of dengue hemorrhagic fever in Ho Chi Minh City, Vietnam using spatial statistics. **Methods:** the spatiotemporal patterns of dengue hemorrhagic fever at the local and global levels were examined using the global Moran's I statistic, the Moran's I scatterplot, and the local Moran's I statistic. More specifically, using the local Moran's I statistic, geographic clusters (high-high and low-low) and spatial outliers (low-high and high-low) of dengue hemorrhagic fever were found. **Results:** there were still high rates of DHF infections in some districts in August, no spatial clusters (high-high and low-low) were found. Instead, low-high and high-low spatial outliers of DHF cases were primarily found in the city's western and southern districts. Districts 4 and Thu Duc were found to have a low-high spatial outlier and a high-high spatial cluster, respectively. **Conclusions:** The study's conclusions showed that Ho Chi Minh City had statistically significant spatial clusters and outliers of DHF prevalence. The findings of this investigation also show how well local, global, and Moran's I statistics, as well as scatterplots, may be used to analyze spatial patterns in DHF. The results of this investigation shed light on how dengue hemorrhagic fever is spread.

Keywords—spatial patterns, spatial statistics, dengue hemorrhagic fever, local indicators of spatial autocorrelation

I. INTRODUCTION

Dengue is a mosquito-borne acute systemic viral infection caused by any one of four serotypes of dengue virus (DENV) (1). The WHO estimates that between 100 and 400 million cases of dengue virus infections have occurred annually over the past 20 years, with an eight-fold rise in cases. (2). One of the biggest obstacles facing developing countries' health systems is the continuous dengue pandemic (1). According to recent estimates of the disease burden, transmission may occur in as many as 124 countries, resulting in at least 35 million symptomatic cases and 20,000 fatalities annually (3). Since dengue is primarily an urban disease, the current rise in cases is probably related to the city's rapid urbanization (4). More than 70% of outbreaks between 2000 and 2015 happened in urban regions or in a mix of urban and rural settings (5). Dengue virus (DENV) spreads more effectively in urban settings due to the density of human population there, which also increases vector-human contact (4). But according to current seroprevalence research, rural and urban populations have both been exposed to DENV to the same extent (6,7). The high seropositive prevalence in rural regions is probably caused by a shift from rural to urban areas and an increase in the number of mobile DENV-infected individuals from urban areas (8).

Geographic Information System (GIS) technologies have long been widely used in public health studies and related challenges, such as disease outbreak during natural disasters, at the regional or national level (9,10). GIS have shown to have significant potential in solving epidemic issues (11). A GIS is a valuable tool for investigating whether seropositivity for dengue is clustered (12). GIS spatial analysis techniques are frequently used to represent disease spatial patterns and determine whether there is a significant clustering of disease incidences in a specific area (13). To be more precise, dead crow data was subjected to kernel density functions in order to record the geographic density of the cases. Using virus-positive mosquito sample sites in the same areas, the grouping of cases that were validated was displayed using maps of human cases and cluster analysis (14,15). The tendency of observations to assume values that are not independent of those of nearby observations is known as spatial dependency, or spatial autocorrelation (16). Locational geography has effectively used spatial autocorrelation, a potent method for analyzing spatial patterning in variate values (17). With spatial autocorrelation analysis, the pattern of distribution of infectious diseases, such dengue instances, can be found.

In Vietnam, DHF is still a dangerously developing arboviral illness. DHF was initially discovered in Vietnam in 1963 in the southern Mekong Delta region (18). Vietnam recorded 1,518,808 cases of dengue hemorrhagic fever and 14,133 fatalities between 1963 and 1995 (19). According to data from the dengue monitoring program in southern Vietnam, between 1975 and 1987, there were around five-year intervals between epidemic peaks of increasing severity. An 11-year interval preceded a major outbreak of dengue hemorrhagic fever (DHF) in 1998, which resulted in 119,429 cases and 342 deaths (20). DHF is endemic in Vietnam, with an estimated 1.6 million cases occurring annually in urban and periurban areas (21). Previous studies have shown that all four DENV serotypes had circulated in Vietnam at some point, with DENV-1 and DENV-2 being the most often detected serotypes, as is normal in hyperendemic nations (22). In 2022, the Ho Chi Minh Centre for Disease Control (HCDC) reported 78,561 cases of dengue, indicating a significant DENV outbreak in the city. Thus, the purpose of this study is to determine the spatiotemporal patterns of dengue hemorrhagic fever in Vietnam's Ho Chi Minh City.

II. STUDY AREA, DATA USED AND METHODS

A. Study area and data used

Nearly 9 million people (more than 8% of the nation's total population) lived in Ho Chi Minh City with an area of 2,061 km² in 2019, and nearly 7 million lived in the 19 inner districts (urban area: 283 km²) (8). In 2014, the city had about 210

parks (8). A. albopictus and A. aegypti in the major parks and the adjacent residential areas in Ho Chi Minh City (8). High dengue virus transmission generally happens in the second half of the year during the rainy season (23). Dengue morbidity and mortality increased in recent years (8). Ho Chi Minh metropolis, the largest metropolis in Vietnam, saw its largest DENV outbreak in over a decade. Significant morbidity and mortality result from it in hyperendemic countries like Vietnam (24). According to reports, DENV-1 was the most common serotype in Ho Chi Minh city and its surrounding areas until 2018 (24). The Ho Chi Minh Centre for Disease Control (HCDC) reported 78,561 dengue cases in 2022, indicating a significant DENV outbreak in the city. Therefore, this study focuses on the identification of spatio-temporal patterns of DHF in Ho Chi Minh city, Vietnam using a dataset of DHF incidence that was gathered in August and September of 2023.

B. Methods

Global Moran's I statistic

The overall presence or absence of spatial autocorrelation is shown by the global Moran's I. This study used the global Moran's I statistic to ascertain the worldwide geographical clustering of dengue hemorrhagic fever incidence (25,26). The definition of the global Moran's I statistic is given in equation (4):

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^n \sum_{j=i}^n W_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Where: x_i and x_j are the dengue hemorrhagic fever incidence for district i and district j; \bar{x} is the mean of the dengue hemorrhagic fever incidence and be given by $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$; n is the total number of districts in the whole study area; and W_{ij} is a $(n \times n)$ spatial weight matrix (27).

The global Moran's I coefficient values are in the interval $[-1, +1]$ (27). When there is positive spatial autocorrelation in the data, global Moran's I values are positive; conversely, when there is negative spatial autocorrelation, global Moran's I values are negative (28). The DHF distribution shows no signs of randomness or regional autocorrelation when global Moran's I coefficient values are near zero.

The local Moran's I statistic

One way to condense a whole spatial distribution into a single number is to utilize the Moran's I statistic. When working with big spatial datasets, a major drawback of the global index is that it might obscure local patterns of spatial clustering or spatial heterogeneity (29). The local Moran's I statistic can be used to overcome this restriction. Since the local Moran's I statistic is one of the LISA statistics that is most frequently employed in research, this study used it to measure the spatial clustering of low and high dengue hemorrhagic fever incidence in each district (27). The local Moran's I statistic (I_i) of dengue hemorrhagic fever incidence at district i is given by the following equation (30):

$$I_i = \frac{(x_i - \bar{x})}{\sigma^2} \sum_{j \# i, j \in J_i}^N W_{ij}(x_j - \bar{x}) \quad (2)$$

Where: x_i , x_j , \bar{x} , and W_{ij} are defined in equation (1); N is the total number of neighborhood districts (27); J_i denotes the neighborhood set of dengue hemorrhagic fever incidence at district i; $j \# i$ implies that the sum of all $(x_j - \bar{x})$ of nearby

neighbourhood districts of district i but not including x_j ; and σ^2 is the variance of x , given in equation (3). W_{ij} defines neighbor connectivity and can be constructed using first or second order of contiguity.

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2 \quad (3)$$

The statistical significance of Moran's I statistic can be tested if I_i follows the normal distribution, the Z-scores can be determined as follows:

$$Z_{I_i} = -\frac{I_i - E(I_i)}{\sqrt{Var(I_i)}} \quad (4)$$

where: $E(I_i)$ and $Var(I_i)$ are the arithmetic expectation and variance of the Moran statistic at district i, respectively.

It is evident that the local Moran's I statistic indicates the degree of spatial clustering of dengue hemorrhagic fever incidence at each district. The local Moran's I value at district i (I_i) lies between -1 and +1, just like the global Moran's I statistic. If the local Moran's I coefficient in district I is 0, then there is no spatial autocorrelation of dengue hemorrhagic fever incidence ($I_i = 0$). If $I_i > 0$ then there will be a positive spatial autocorrelation of dengue hemorrhagic fever incidence (27). If $I_i < 0$ then there will be a negative spatial autocorrelation of dengue hemorrhagic fever incidence.

The Moran's I scatterplot

The Moran scatterplot was utilized to visualize and determine the degree of spatial instability (30). The original variable's deviations from the global mean on the x-axis and the average neighborhood weight deviations from the global mean on the y-axis make up the four quadrants centered on the global mean (29). The upper right quadrant of the Moran's I scatterplot shows high values above the mean surrounded by high values; the lower left quadrant shows low values surrounded by low values; the upper left quadrant shows low values surrounded by high values; and the lower right quadrant shows high values surrounded by low values (28). The local Moran's I statistic in this instance shows two different types of geographic clusters for dengue hemorrhagic fever incidence cases: high-high spatial clusters and low-low spatial clusters. When there is a negative local spatial autocorrelation, the local Moran's I statistic can also be used to identify two more types of spatial outliers: low-high and high-low clusters. The Moran scatterplot better illustrates the various patterns, outliers, and uncertainty/transition zones the stronger the Moran's I (31).

III. RESULTS AND DISCUSSION

A. Spatial distribution of DHF incidence

The statistics on infection rates/100,000 people were summarized in Table I. The distribution of DHF incidence in August and September was shown in the boxplots in Figure 1 and maps in Figure 2. Data in Table I shows the number of DHF cases (DHF infection rate) has increased gradually each month with the mean value increasing in the range of 16.1 and 21.5, respectively. The dispersion of the DHF incidence also increased month by month with the corresponding values of IQR and SDEV of 8 and 8, and of 6.1 and 7.2, respectively. Data from the boxplots in Figure 2 illustrate that, in August and September, a high rate of DHF infections (median is

greater than the average value) were discovered in the majority of districts.

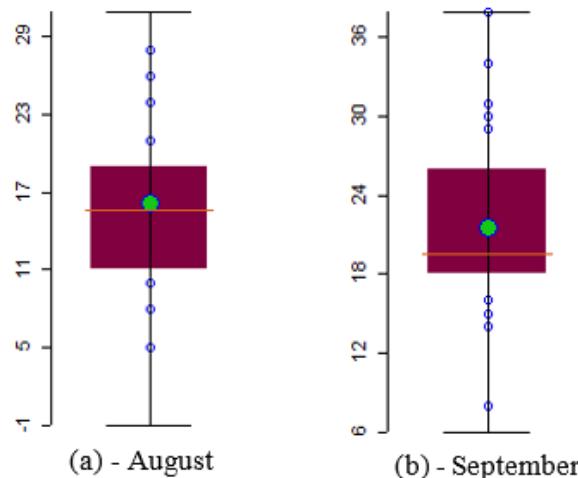


Figure 1. Boxplots of the incidence of DHF.

Data from Figure 2 shows that high DHF infection rates mainly occurred in districts in the center of the city, whereas low DHF infection rates were detected in suburban districts in the North and South of the city. In particular, there was a tendency for the infection rate in urban districts to increase over time. In August, high infection rates were mainly concentrated in the west of the city. The high infection rate shifted to the western districts of the city in August and moved to the east of the city in September. Thus, it can be seen that there was a continuous change in the locations of high DHF infection rates.

B. Moran's I scatterplot

Data from Moran scatter plots in Figure 3 illustrates the degree of globally spatial autocorrelation of dengue hemorrhagic fever in August and September, 2023. The values of global Moran's I statistic identified in these months were -0.29 and -0.04, respectively. The negative values of global Moran show that, overall, there existed negative autocorrelation between DHF infection rates in these districts in these two months.

TABLE I. Summary table of statistical descriptives for DHF.

Time	Statistical descriptives							
	Min	Mean	Median	Max	Q1	Q3	IQR	SDEV
August	5	16.1	15.5	28	11	19	8	6.1
September	8	21.5	19.5	38	18	26	8	7.2

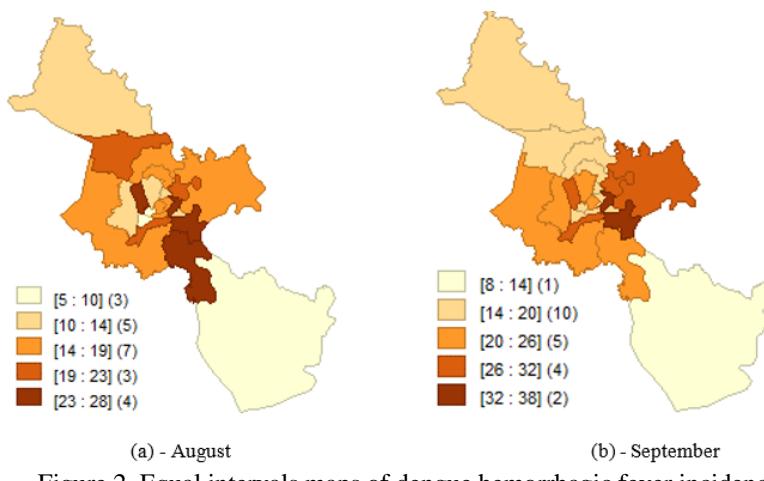


Figure 2. Equal intervals maps of dengue hemorrhagic fever incidence.

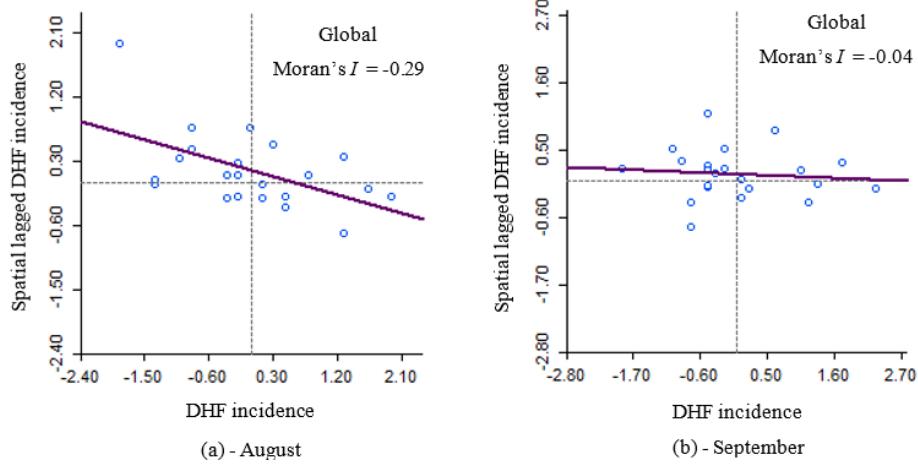


Figure 3. Moran's I scatterplots of dengue hemorrhagic fever incidence.

C. Spatio-temporal clustering analysis of DHF

The spatial distribution of local Moran' I coefficients obtained in August and September were shown in boxplots in Figure 4. The corresponding descriptive statistics were summarized in Table II. Data in the Table II illustrate that the minimum values of local Moran's I statistic in these two months were negative, indicating negative spatial autocorrelation for DHF infection rates. The median values of local Moran's I were almost concentrated around zero (ranging of from -0.05 to -0.04).

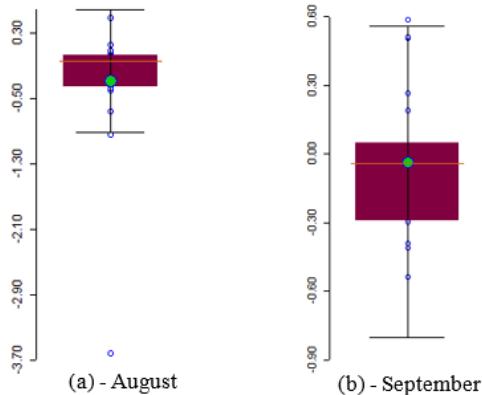


Figure 4. LISA: (a) boxmap and (b) boxplot.

The maximum value of local Moran's I statistic was detected in September with a value of 0.59. Meanwhile, measures of central tendency shows that values of IQR and SDEV were in the range of [0.34; 0.38] and [0.31; 0.80], respectively. Data in boxplots in Figure 4-a shows Moran's I coefficients in August were skewed below the median, showing that many values of the Moan's I statistic were below mean values. In September, data from the boxplot in Figure 4-b demonstrates that local Moran's I statistic was skewed to the top, showing that many high values of local Moran's I statistic were greater than the mean.

Data from the Boxmap (hinge = 1.5) in Figure 5-a shows that, in August, low values of local Moran's I coefficients were found in the northern ans southern districts, whereas moderate and high values of local Moran's I statistic were detected in the eastern districts. Similar to those obtained in August, low Moran's I values were also mainly detected in the northern and southern districts of the city, where moderate and high Moran's I values were mostly detected in the western and eastern districts, respectively. However, in September, high values of local Moran's I index were mainly concentrated in the northern and eastern districts, while low Moran's I coefficient appeared in the western and southern districts of the city.

TABLE II. Summary table of statistical descriptives for LISA.

Time	Statistical descriptives							
	Min	Mean	Median	Max	Q1	Q3	IQR	SDEV
August	-3.6	-0.28	-0.05	0.49	-0.35	0.03	0.38	0.80
September	-0.54	-0.04	-0.04	0.59	-0.29	0.05	0.34	0.31

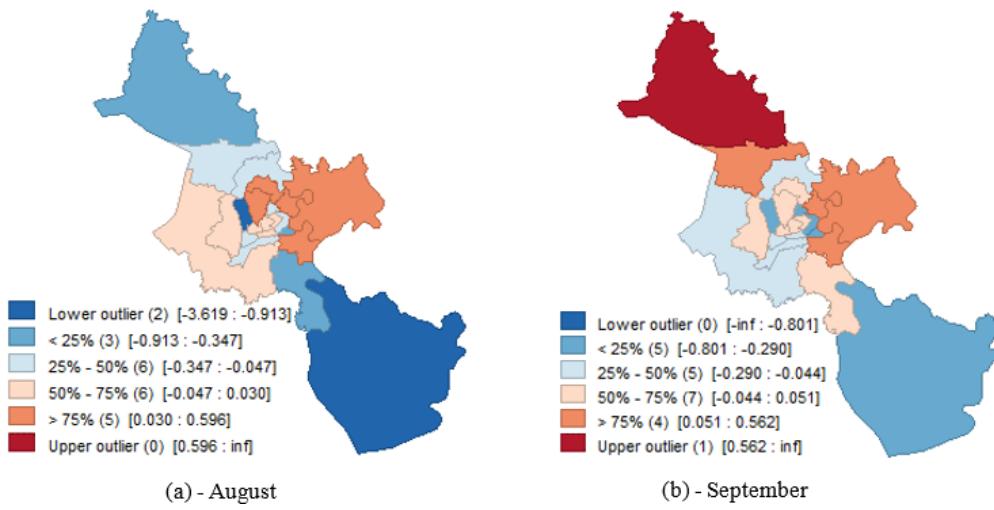


Figure 5. Boxmap of DHF incidence.

Spatial clusters of DHF incidence obtained from the local Moran's I statistic were shown in Figure 6. Overall, data from the LISA cluster map shown in Figure 4 shows that spatial clusters were mainly detected in the city center. The dynamics of spatial clusters from August to September specifically has changed as follows: in August, the Moran's I statistic successfully detected 4 spatial outliers clusters with three low-high and one high-low outliers. Low-high and high-low spatial outliers appeared in the western and southern districts of the city, respectively including District 4 (11 cases/100,000 people), Binh Chanh (16 cases/100,000 people), Can Gio (5 cases/100,000 people), and Tan Phu (24 cases/100,000 people). No spatial clusters (high-high and low-low) were detected in August although there were still high rates of DHF infections in some districts such as Nha Be (28 cases/100,000 people), District 1 (26 cases/100,000 people), and District 7

(24 cases/100,000 people). It can be seen that the spatial clusters detected in August moved to the west and south of the city. In September, the Moran's I index detected one spatial cluster and one spatial outliers. High-high and low-high spatial clusters were found districts of Thu Duc (26 cases/100,000 people) and District 4 (18 cases/100,000 people), respectively. No spatial clusters, especially high-high clusters, were detected in this month although some districts were found with relatively high number cases of DHF such as District 1 (38 cases/100,000 people), District 7 (34 cases/100,000 people), and District 8 (31 cases/100,000 people). In addition, the remaining 20 districts were also identified with statistically unsignificance at the 0.05 level. Therefore, compared with those found in August, the spatial cluster of DHF infection rates in August moved to the east of the city in September 2023.

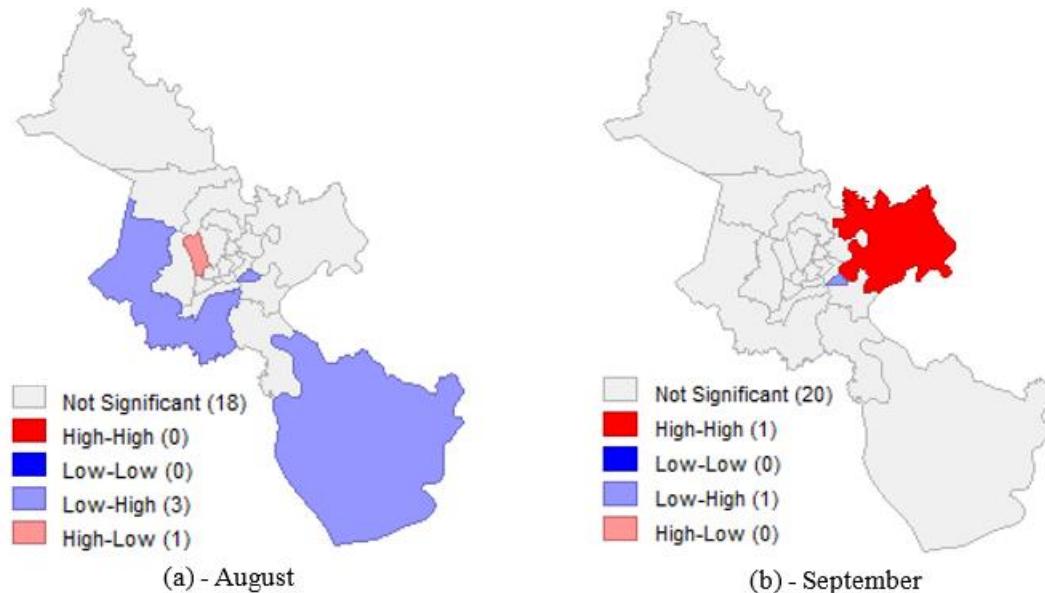


Figure 6. LISA cluster maps of dengue hemorrhagic fever incidence.

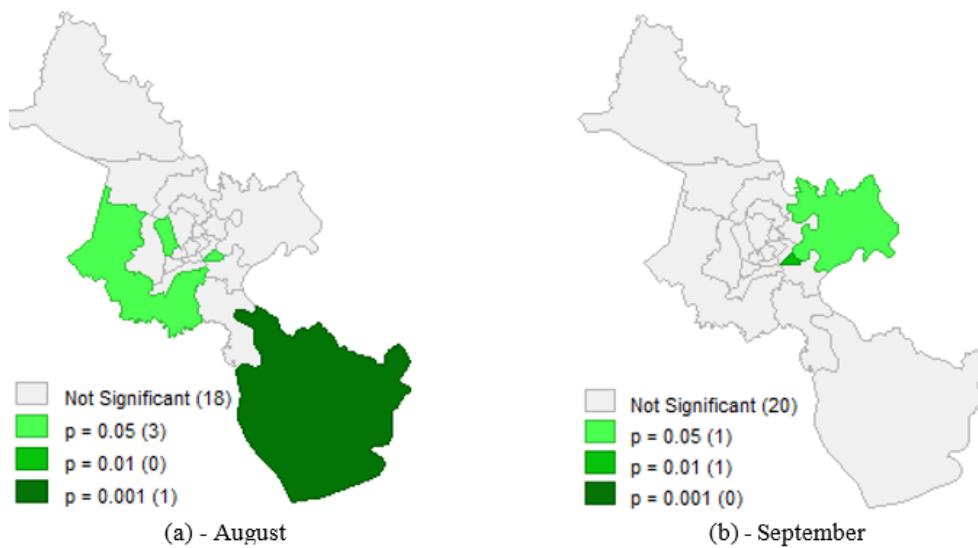


Figure 7. LISA cluster maps of dengue hemorrhagic fever incidence.

Data from significant maps in Figure 7 shows statistical significance levels for DHF infection rates for each district of Ho Chi Minh city in August and September, 2023. Statistical levels were presented in four levels including statistically unsignificance (> 0.05) and statistically significance at the 0.05, 0.01 and 0.001 levels. Data from Figure 7-a illustrate that, three districts were found with statistically significance at the 0.05 level and one district with statistical significance at

a very high level of 0.001 in August (Figure 7-a). In September, two districts were identified with statistically significance at the levels of 0.05 and 0.01 (Figure 7-b). The results in Figures 7-a and 7-b show that more high clusters of DHF were detected in August than in September 2023. This indicates that the DHF infection rate in August was higher than that in September 2023.

IV. CONCLUSIONS

The purpose of this study was to determine the geographic distribution of dengue hemorrhagic fever in Vietnam's Ho Chi Minh City. To look at the spatial distribution of dengue hemorrhagic fever on a local and worldwide scale, the global Moran's I statistic, the Moran's I scatterplot, and the local Moran's I statistic were used. More precisely, using the local Moran's I statistic, geographic clusters (high-high and low-low) and spatial outliers (low-high and high-low) of dengue hemorrhagic fever were found. While there were still high rates of DHF infections in some districts in August, it was discovered that low-high and high-low spatial outliers of DHF cases were primarily found in the city's western and southern districts. In contrast, no high-high or low-low spatial clusters were found in August. A high-high spatial cluster and a low-high spatial outlier were found districts of Thu Duc and District 4, respectively. In summary, the results of this investigation showed that Ho Chi Minh City had statistically significant spatial clusters and outliers of DHF prevalence. The findings of this investigation also show how well local, global, and Moran's I statistics, as well as scatterplots, may be used to analyze spatial patterns in DHF. The results of this investigation shed light on how dengue hemorrhagic fever is spread.

ACKNOWLEDGMENT

The authors would like to thank HCDC for providing the data. The authors acknowledge the editors and anonymous reviewers for their careful reading of our paper and their many insightful comments and suggestions.

REFERENCES

- [1] Anders KL, Nguyet NM, Chau NNV, Hung NT, Thuy TT, Farrar J, et al. Epidemiological factors associated with dengue shock syndrome

- [13] Ward MP, Levy M, Thacker HL, Ash M, Norman SKL, Moore GE, et al. Investigation of an outbreak of encephalomyelitis caused by West Nile virus in 136 horses. *J Am Vet Med Assoc.* 2004;225(1):84–9.
- [14] Wittich CA. Spatial analysis of West Nile Virus and predictors of hyperendemicity in the Texas equine industry. Texas A & M University; 2008.
- [15] Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, et al. Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne Zoonotic Dis.* 2002;2(3):157–64.
- [16] Getis A. Cliff, ad and ord, jk 1973: Spatial autocorrelation. london: Pion. *Prog Hum Geogr.* 1995;19(2):245–9.
- [17] Premo LS. Local spatial autocorrelation statistics quantify multi-scale patterns in distributional data: an example from the Maya Lowlands. *J Archaeol Sci.* 2004;31(7):855–66.
- [18] Halstead SB, Voulgarpoulas E, Tien NH, Udomsakdi S. Dengue hemorrhagic fever in South Vietnam: report of the 1963 outbreak. *Am J Trop Med Hyg.* 1965;14(5):819–30.
- [19] Halstead SB. Epidemiology of dengue and dengue hemorrhagic fever. *Dengue and dengue hemorrhagic fever.* 1997;23–44.
- [20] Ha DQ, Tien NT, Huong VT, Loan HT, Thang CM. Dengue epidemic in southern Vietnam, 1998. *Emerg Infect Dis.* 2000;6(4):422.
- [21] Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, Ferguson NM. Mapping global variation in dengue transmission intensity. *Sci Transl Med.* 2020;12(528):eaax4144.
- [22] Ty Hang VT, Holmes EC, Veasna D, Quy NT, Tinh Hien T, Quail M, et al. Emergence of the Asian 1 genotype of dengue virus serotype 2 in viet nam: in vivo fitness advantage and lineage replacement in South-East Asia. *PLoS Negl Trop Dis.* 2010;4(7):e757.
- [23] Mogi M, Khamboonruang C, Choochote W, Suwanpanit P. Ovitrap surveys of dengue vector mosquitoes in Chiang Mai, northern Thailand: seasonal shifts in relative abundance of *Aedes albopictus* and *Ae. aegypti*. *Med Vet Entomol.* 1988;2(4):319–24.
- [24] Tran VT, Inward RPD, Gutierrez B, Nguyen NM, Rajendiran I, Thanh PN, et al. Cryptic transmission and re-emergence of Cosmopolitan genotype of Dengue Virus Serotype 2 within Ho Chi Minh City and Southern Vietnam. *medRxiv.* 2023;2004–23.
- [25] Cliff AD, Ord JK. Spatial processes: models & applications. (No Title). 1981;
- [26] Getis A, Ord JK. Local spatial statistics: An overview. *Spatial analysis: Modeling in a GIS environment.* Longley, P., and M. Batty. Wiley, New York; 1996.
- [27] Vu D-T, Nguyen T-T, Hoang A-H. Spatial clustering analysis of the COVID-19 pandemic: A case study of the fourth wave in Vietnam. *Geogr Environ Sustain.* 2021;14(4).
- [28] Nguyen TT, Vu TD. Identification of multivariate geochemical anomalies using spatial autocorrelation analysis and robust statistics. *Ore Geol Rev.* 2019;111.
- [29] Negreiros JG, Painho MT, Aguilar FJ, Aguilar MA. A comprehensive framework for exploratory spatial data analysis: Moran location and variance scatterplots. *Int J Digit Earth.* 2010;3(2):157–86.
- [30] Anselin L. Local indicators of spatial association—LISA. *Geogr Anal.* 1995;27(2):93–115.
- [31] Negreiros J. Spatial neighborhood matrix computation: inverse distance weighted versus binary contiguity. 2009;

Acid Precipitation of Lignin from Alkali Black Liquor: A Study on the Optimization of Lignin Recovery Using H₂SO₄ and HCl

1st Khatiya Weerasai

The Joint Graduate School for
Energy and Environment
King Mongkut's University of
Technology Thonburi
Prachauthit Road, Bangmod,
Bangkok 10140, Thailand
k.weerasai@gmail.com

Saksit Imman
School of Energy and
Environment
University of Phayao
Tambon Maeka, Amphur Muang
Phayao 56000, Thailand
saksit.im@up.ac.th

Torpong Kreetachat
School of Energy and
Environment
University of Phayao
Tambon Maeka, Amphur Muang
Phayao 56000, Thailand
torpong.kr@up.ac.th

Nopparat Suriyachai*
School of Energy and
Environment
University of Phayao
Tambon Maeka, Amphur Muang
Phayao 56000, Thailand
nopparat.su@up.ac.th

Abstract— In this study, black liquor derived from an alkali-catalyzed organosolv process was used as the source for lignin extraction, with the aim of maximizing lignin yield and exploring its utilization in sustainable applications. To reduce the pH of the alkali black liquor from 13 to below 2, sulfuric acid (H₂SO₄) and hydrochloric acid (HCl) were employed for acidification. The results showed that H₂SO₄ produced the highest solid precipitate yield, with a lignin yield of 86.09% and a corresponding solid yield of 26.19% at a reaction time of 12 hours. In contrast, HCl required larger volumes but achieved higher lignin purity, reaching 92.57% compared to 86.63% for H₂SO₄. The process efficiency was examined in terms of acid volume consumption and lignin precipitation yield. While H₂SO₄ offered better efficiency in terms of lower acid volumes, HCl demonstrated superior lignin and solid yields, making it more effective in maximizing product recovery. Moreover, the high initial ash content (72.1%) in the black liquor, which was linked to the solid yield, was significantly reduced after acid precipitation, improving the purity of the extracted lignin. The Klason lignin content was enhanced through the process, demonstrating the effectiveness of the acid precipitation in isolating lignin with reduced inorganic contaminants. These findings contribute to the development of more efficient lignin valorization processes, supporting the broader goal of sustainable biomass utilization by converting black liquor into high-purity lignin suitable for further applications in bioplastics, adhesives, and biofuels.

Keywords— Black liquor, Acid precipitation, Derived lignin, Lignin recovery

I. INTRODUCTION

As the world faces the pressing challenge of climate change, the need for sustainable alternatives to fossil-based resources has become more critical than ever. Biorefineries, which utilize renewable biomass for the production of fuels, chemicals, and materials, offer a promising solution to reducing greenhouse gas emissions and promoting circular economies [1]. By maximizing the use of all components of biomass, biorefineries can contribute significantly to mitigating environmental impacts while providing a sustainable source of energy and materials [2]. Among the various strategies employed, lignocellulosic biomass

valorization has gained attention due to its abundant availability and potential to replace conventional petrochemicals.

Black liquor, a byproduct of the industrial pulping process, is a lignin-rich waste stream generated during the chemical separation of lignin from cellulose fibers. In large-scale kraft and organosolv pulping processes, black liquor is typically burned as a low-value fuel to recover energy, while the remaining inorganic chemicals are recycled back into the pulping process [3]. However, this practice underutilizes the potential of lignin as a valuable feedstock for high-value chemicals and materials. Advances in biorefinery technologies have paved the way for the valorization of lignin from black liquor, transforming it from a waste product into a source of biobased products, such as adhesives, carbon fibers, and biofuels. In this study, the black liquor derived from an alkali-catalyzed organosolv process was used as the source for lignin extraction, aiming to maximize lignin yield and explore its utilization in more sustainable applications [4].

One promising method for lignin recovery is acid precipitation [5], which facilitates the separation of lignin from black liquor. In this study, lignin was extracted from black liquor obtained from a lab-scale alkali-catalyzed organosolv process. Strong acids, including sulfuric acid (H₂SO₄) and hydrochloric acid (HCl), were used to determine the optimal acid dosages for lignin precipitation. Following extraction, the lignin was characterized for key physical and chemical properties, providing valuable insights into its potential applications in various industries. This research aims to contribute to the development of more efficient lignin valorization processes, ultimately supporting the broader goal of sustainable biomass utilization.

II. MATERIALS AND METHODS

A. Raw material

The starting material is black liquor fraction that provided from alkaline-catalyzed organosolv pretreatment developed in a

laboratory. Separation of liquid and solid fractions will be carried out by acidification using strong acids (i.e. H_2SO_4 and HCl) under various conditions. Then, the result of black liquor especially solid part will be characterized.

B. Lignin precipitation using strong acids

To obtain first lignin fraction, 100 ml of alkaline black liquor (pH around 9-11) will be acidified by strong acids i.e. H_2SO_4 and HCl at atmospheric pressure. In order to study the optimum condition, the concentration of acid, pH range, time for precipitation, and temperature will be varied. The amount of strong acid concentration will be varied including 1 M, 2 M, 3 M, and 6 M. The pH is an important factor which has an effect of precipitation. The temperature ranges for this research will be altered from 25-75 °C; time for precipitation will be trialed in range of 2, 4, 6, and 8 h. After that, the precipitated lignin will be centrifuged at 4000 rpm for 10 minutes.

C. Lignin separation process

After lignin precipitation, the black liquor will be separated into solid and liquid fractions by the filtration process. The black liquor will be filtrated with Glass fiber filter 1.2 µm (Whatman). Then, the solid cake will be dried at 60-70 °C and characterized chemical and physical properties.

III. RESULTS AND DISCUSSION

The acid precipitation process was employed to separate lignin from black liquor, a lignin-rich byproduct of the alkali-catalyzed organosolv process developed in the laboratory. Different strong acids, specifically sulfuric acid (H_2SO_4) and hydrochloric acid (HCl), were used to determine the optimal acid dosages for lignin extraction, aiming to maximize lignin yields. Following the precipitation process, the extracted lignin will be characterized for its key physical and chemical properties.

A. Effect of reaction time in acid precipitation process

To decrease the pH value from 13 to below 2 in alkali black liquor, both H_2SO_4 and HCl were used for acidification, with different volumes required for each acid. For 200 ml of black liquor, approximately 3 ml of sulfuric acid (98% w/w) and 6 ml of hydrochloric acid (37% w/w) were used, as shown in Table I. The effects of varying the reaction time on the precipitation yield of lignin from alkali black liquor were examined for both acids. In the case of H_2SO_4 and HCl, sulfuric acid produced the highest solid precipitate compared to hydrochloric acid [6]. At a reaction time of 12 hours, the percentage of lignin yield was 86.09%, corresponding to a 26.19% solid yield (Fig. 1). Moreover, the increase in reaction time was more beneficial to lignin precipitation when using H_2SO_4 compared to HCl. However, a larger volume of HCl was required compared to H_2SO_4 in this process [5].

Table I. Compare the used of volume to acidify alkali black liquor to pH 2.

Acid	Concentration (w/w)	Acid volume (ml)
------	------------------------	------------------------

H_2SO_4	98% (w/w)	3
HCl	37% (w/w)	6

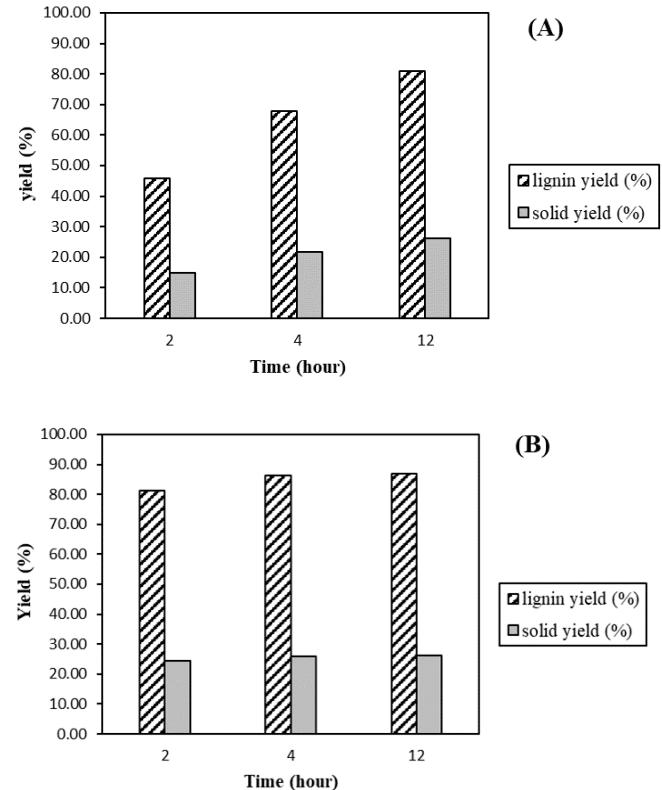


Fig. 1 Effect of time on acid precipitation in the presence of H_2SO_4 (A) and HCl (B).

B. Effect of concentration in acid precipitation process

The reduction of pH from 13 to below 2 in 200 ml of alkali black liquor was achieved using sulfuric acid (H_2SO_4) and hydrochloric acid (HCl) for acidification. For sulfuric acid concentrations of 1, 2, and 3 molar, the required volumes ranged from approximately 1.5 to 4 ml, while for hydrochloric acid at the same concentrations, the volumes used were between 2 and 7 ml, as shown in Figures (A) and (B).

In the H_2SO_4 system, increasing the concentration did not significantly affect the lignin and solid yields. However, a higher acid concentration led to a reduction in the acid volume required for the process, suggesting improved efficiency. On the other hand, the use of HCl showed a more pronounced effect on both lignin and solid yields, with higher concentrations resulting in a substantial increase in yield. The comparison of the two acids reveals that while H_2SO_4 offers better efficiency in terms of volume consumption, HCl is more effective in achieving higher yields of lignin and solids [7].

Additionally, the results further highlight that, at equivalent concentrations, H_2SO_4 required a smaller volume compared to HCl for the same degree of acidification. This volume efficiency makes H_2SO_4 a more practical option for processes where

minimizing acid usage is critical. However, the higher lignin and solid yields observed in the HCl system suggest its potential for applications where maximizing product recovery is the primary goal.

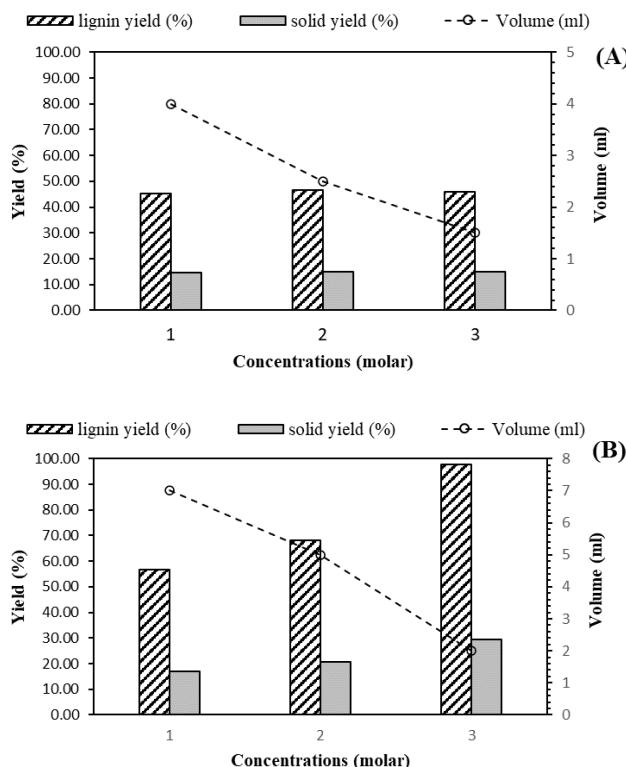


Fig. 2 Effect of concentration on acid precipitation in the presence of H_2SO_4 (A) and HCl (B).

C. The purity of lignin precipitation

The composition of the black liquor prior to the acid precipitation process reveals that it contains 27.9% lignin, with the remaining majority, 72.1%, being ash. The high ash content indicates a significant presence of inorganic materials, which could impact the efficiency and purity of lignin extraction. The solid yield observed in different precipitation methods is directly linked to this high ash content, as the solids contain both organic (lignin) and inorganic (ash) components.

After the precipitation process, which involved the use of acids (H_2SO_4 and HCl), the purity of the lignin significantly improved. The purity of lignin after precipitation with H_2SO_4 increased to 86.63%, and with HCl, it rose even further to 92.57%. This demonstrates the effectiveness of the acid precipitation process in removing the inorganic contaminants, specifically the ash, from the black liquor, thereby enhancing the overall purity of the recovered lignin.

The decrease in ash content following the acid treatment suggests that the acids were effective in separating and precipitating out the lignin, while the ash and other contaminants

were largely removed. This purification step is crucial for valorizing lignin, as high-purity lignin is more suitable for further applications in bio-based and chemical products.

The Klason lignin content, a measure of the lignin purity, was also enhanced after the acid precipitation process. This improvement in Klason lignin content indicates that the process is successful in isolating lignin in a form that is more concentrated and less contaminated with non-lignin materials. This purified lignin can now be further valorized in various applications, such as the production of bioplastics, adhesives, or as a precursor for biofuels.

IV. CONCLUSIONS

This study successfully demonstrated that acid precipitation using sulfuric acid (H_2SO_4) and hydrochloric acid (HCl) is an effective method for extracting and purifying lignin from alkali black liquor derived from an alkali-catalyzed organosolv process. While H_2SO_4 was more efficient in terms of acid volume consumption, HCl achieved higher lignin purity and yield, making it advantageous for applications where maximizing product recovery is essential. The process also significantly reduced the high initial ash content in black liquor, enhancing the overall purity of the lignin. These findings support the development of more efficient lignin valorization processes, contributing to sustainable biomass utilization and providing high-purity lignin for further use in bio-based products such as bioplastics, adhesives, and biofuels.

ACKNOWLEDGMENT

Nopparat Suriyachai was financially supported by Unit of Excellence (UOE220/2567) from the University of Phaya.

REFERENCES

- [1] T. Giwa, M. Akbari, and A. Kumar, "Assessment of the greenhouse gas emission footprint of a biorefinery over its life cycle," *Energy Conversion and Management*, vol. 271, pp. 116327, 2022.
- [2] H. Y. Leong, C.K. Chang, K.S. Khoo, K.W. Chew, S.R. Chia, J.W. Lim, J.S. Chang, and P.L. Show, "Waste biorefinery towards a sustainable circular bioeconomy: a solution to global issues," *Biotechnology for Biofuels*, vol. 14(1), PP. 87, 2021.
- [3] L. Pola, S. Collado, P. Oulego, and M. Díaz, "Kraft black liquor as a renewable source of value-added chemicals," *Chemical Engineering Journal*, vol. 448, pp. 137728, 2022.
- [4] R. Morya, M. Kumar, I. Tyagi, A. Kumar Pandey, J. Park, T. Raj, R. Sirohi, V. Kumar, and S.H. Kim, "Recent advances in black liquor valorization," *Bioresource Technology*, vol. 350, pp. 126916, 2022.
- [5] R.C. Andeme Ela, L. Spahn, N. Safaie, R.C.Jr. Ferrier, and R.G. Ong, "Understanding the Effect of Precipitation Process Variables on Hardwood Lignin Characteristics and Recovery from Black Liquor," *ACS Sustainable Chemistry & Engineering*, vol. 8(37), pp. 13997-14005, 2020.
- [6] E.C. Watumlawar, and B.D. Park, "Effects of Precipitation pH of Black Liquor on Characteristics of Precipitated and Acetone-Fractionated Kraft Lignin," *Journal of the Korean Wood Science and Technology*, vol. 51(1), pp. 38-48, 2023.
- [7] S. Priyanto, A.P. Intan, B. Rianto, T.D. Kusworo, B. Pramudono, E. Untoro, and P. Ratn, "The Effect of Acid Concentration (H_2SO_4) on the Yield and Functional Group during Lignin Isolation of Biomass Waste Pulp and Paper Industry." *Reaktor*, vol. 19, no. 4, pp. 162-167, Dec. 2019.

Satellite-based Nitrogen Uptake Evaluation in Rice Cultivation

Porutai Thianthai

*School of Information and Communication Technology,
University of Phayao
Phayao, Thailand.
67025291@up.ac.th*

Sukij Skawsang

*Geo-informatics and Space Technology
Development Agency
(Public Organization)
Bangkok, Thailand.
sukij@gistda.or.th*

Phaisarn Jeefoo

*School of Information and Communication Technology,
University of Phayao
Phayao, Thailand.
Phaisarn.je@up.ac.th*

Sawarin Lerk-U-Suke

*School of Information and Communication Technology,
University of Phayao
Phayao, Thailand
sawarin.le@up.ac.th*

Panu Nuangjamnong

*Geo-informatics and Space Technology
Development Agency
(Public Organization)
Bangkok, Thailand.
panu.n@gistda.or.th*

Nuntikorn Kitratporn

*Geo-informatics and Space Technology
Development Agency
(Public Organization)
Bangkok, Thailand.
nuntikorn.kit@gistda.or.th*

Abstract— This research illustrates the efficacy of utilizing Sentinel-2 and Sentinel-1 satellite data to evaluate nitrogen uptake. In 2023, nine rice plots in Suphan Buri Province, Thailand were chosen as study areas. The ground-based nitrogen uptake from plant samples at the panicle initiation stage was used to develop model with several satellite indices. The result indicated that the Modified Soil Adjusted Vegetation Index (MCARI) from Sentinel-2 showed good performance with the R^2 of 0.748 and RMSE of 16.614 kg/ha. On the other hand, Sentinel-1's Dual Polarization SAR Vegetation Index (DPSVI) showed more accuracy with R^2 of 0.814 and RMSE of 13.678 kg/ha. The research confirms that satellite data can efficiently evaluate nitrogen levels in rice fields and enable the delineation of nitrogen content-rich regions, reducing the time and resources required for field surveys.

Keywords—*Sentinel-1, Sentinel-2, Nitrogen Uptake, Panicle Initiation Stage*

I. INTRODUCTION

Rice, which serves as a crucial source of sustenance for more than fifty percent of the global populace, thereby underscoring the vital significance of ensuring food security for individuals across the world. Projections indicate that the production of rice may reach an impressive total of 567 million tons by the year 2030, reflecting a substantial increase in agricultural output [1]. Nitrogen plays an indispensable role in promoting both the growth and yield of rice crops, particularly when it is utilized in a manner that is both correct and appropriate. The presence of nitrogen has a direct impact on the efficiency of photosynthesis within rice plants, leading to the manifestation of darker green foliage and an expansion of leaf area [2],[3],[4] both of which are critical determinants for the absorption of light and overall plant development. Nonetheless, the misuse of nitrogen fertilizer poses a considerable challenge, as it has detrimental effects on the environment [5],[6], resulting in the contamination and accumulation of nitrogen within both soil and aquatic systems, which in turn contributes to the deterioration of soil quality [7].

When nitrogen fertilizer is applied correctly, it has the potential to significantly enhance rice yield by fortifying the plants themselves, encouraging robust growth, increasing

the quantity of grains produced per panicle initiation stage, and elevating the overall quality of the grains harvested [8]. Conventional techniques for assessing nitrogen levels typically involve direct measurements taken from rice plants, which are not only time-intensive but also financially burdensome [9]. Consequently, the meticulous management and application of nitrogen fertilizer emerge as pivotal elements in the endeavor to improve both the yield and the quality of rice, while simultaneously maintaining the delicate ecological balance within rice paddies, and mitigating the environmental challenges that can arise as a result of the excessive application of nitrogen fertilizers.

With the rapid advancements of remote sensing technologies and satellite imaging, the process of nitrogen assessment has now become markedly more accessible, expeditious, and capable of effectively covering extensive geographical areas [10]. In this research, remote sensing data derived from both Optical and Synthetic Aperture Radar (SAR) systems are employed to evaluate nitrogen uptake within rice cultivation. This data is obtained from the Sentinel-2 and Sentinel-1 satellites, which are integral components of the European Union's Copernicus program. Sentinel-2 is comprised of two satellites, specifically Sentinel-2A and Sentinel-2B, which were launched into orbit in the years 2015 and 2017, respectively. These satellites acquire multispectral imagery with spatial resolutions of 10, 20, and 60 m., encompassing 13 bands across the Visible (VIS), Near-Infrared (NIR), and Shortwave Infrared (SWIR) spectral ranges. They have a revisit interval of every 5 days for the same geographic location.

Conversely, Sentinel-1, which likewise consists of two satellites, Sentinel-1A and Sentinel-1B, was launched into orbit in 2014 and 2016. This system utilizes SAR technology, which possesses the capability to gather data irrespective of weather conditions and at any hour of the day or night [11]. Sentinel-1 provides a variety of operational modes to facilitate diverse applications on Earth, including the Interferometric Wide Swath (IW) mode, featuring a resolution of 5 x 20 meters and covering an extent of 250 kilometers; the Extra-Wide Swath (EW) mode, possessing a resolution of 20 x 40 meters and covering 400 kilometers;

and the Stripmap (SM) mode, which attains a maximum resolution of 5 x 5 meters.

The primary aim of research is to analyze the relationship between satellite-derived indices and empirical field data to develop a satellite-based model for assessing nitrogen levels in rice paddies, as well as to generate a nitrogen concentration map utilizing remote sensing approach. Should a definitive relationship be ascertained, these indices could be employed to effectively monitor and manage nitrogen levels across extensive areas, thereby diminishing the necessity for labor-intensive and time-consuming fieldwork. The developed model serves to enhance the efficacy of nitrogen fertilizer application, augment crop yields, and mitigate environmental repercussions.

II. MATERIAL AND METHODS

A. Study Area

This study was conducted in Suphan Buri Province, situated in central Thailand (about $14^{\circ} 4'$ - $15^{\circ} 5'$ N, $99^{\circ} 17'$ - $100^{\circ} 16'$ E), approximately 5,358.01 km² in size [12]. The climate is classified as tropical humid, with an average annual temperature of around 28.4°C and yearly precipitation between 900 and 1,200 mm [13]. Suphan Buri ranks as one of the nation's largest and most significant rice-producing regions due to its effective irrigation infrastructure. This research area highlights the cultivation of the RD85 rice variety recognized for its excellent yield and resilience to diseases and pests [12],[13] [14]. The research chose the transplanting paddy field, growing seedlings in small plots until they attain sufficient vigor for transplantation into the primary flooded areas. This technique regulates the spacing of rice plants, facilitating uniform growth, reducing competition among them, and preventing plant development.

B. Method

The workflow is summarized in Fig. 1, where ground data is collected and analyzed in the laboratory to calculate nitrogen. Then, Sentinel-2 and Sentinel-1 images close to the sample collection date are searched for and obtained. The images are processed to reduce noise and correct for reflection effects. Subsequently, vegetation indices (VI) are extracted to analyze with laboratory data. The data is divided into training and testing sets to find the most suitable index for assessing nitrogen absorption. Finally, a map showing nitrogen distribution is created using color gradients in each region.

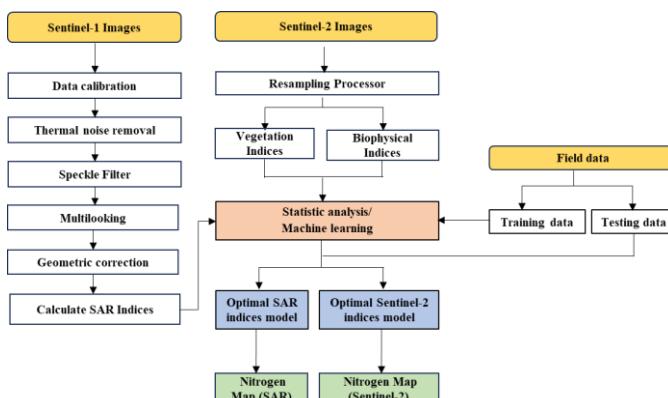


Fig. 1. Methodological

C. Field Data

The data collected on RD85 rice variety includes the information from a total of 9 fields and 72 designated sampling points. These samples were collected during the panicle initiation stage phase (PI) before the second round of fertilization. In each field, samples were taken from 8-9 specific points, with each sampling point covering an area of 1x1 m. Systematic sampling methods were utilized for ensuring an even distribution of sampling sites throughout the plot and to minimize collecting data errors. Through the collection of rice leaves and stems, the leaves, with particular emphasis on the flag leaf (the uppermost leaf), serve as the most accurate indicators of the nitrogen status within the rice plant. The flag leaf reflects the nitrogen absorption utilized for grain production, whereas the stem conveys information regarding nitrogen storage and its distribution to various sections of the plant. Subsequent to the collection of samples, they are transported to the laboratory for the quantification of total nitrogen content in the rice plant (%) and the assessment of its dry weight (kg/ha). The nitrogen uptake can subsequently be computed utilizing Equation 1.

$$N \text{ uptake (Kg N / ha)} = \frac{\text{Total N} (\%)}{100} \times \text{Dry weight (kg N/ha)} \quad (1)$$

The overall nitrogen concentration, represented as N (%), indicates the amount of nitrogen present within the plant tissue, measured as a percentage of the plant's dry weight. Dry Weight signifies the mass of rice straw that endures after harvesting and desiccation, subsequent to the comprehensive extraction of moisture content.

Field data was collected on two separate occasions, first on January 13, 2023 for plot 5-7 where rice was transplanted earlier, and another on February 1, 2023 for the rest of the plot (Table 1).

TABLE.1 FIELD DATA ACQUISITION DATE AND IMAGE ANALYSIS DATE

P	Area (ha)	Transplanted Date (2022)	Field data (2023)	Satellite Observation Acquisition (2023)	
				Sentinel-2	Sentinel-1
1	2.864	1 December	1 February	January 31	31 January
2	1.424	4 December	1 February	January 31	31 January
3	2.4	2 December	1 February	January 31	31 January
4	2.432	1 December	1 February	January 31	31 January
5	3.776	18 November	13 January	January 6	7 January
6	1.52	17 November	13 January	January 6	7 January
7	1.824	17 November	13 January	January 6	7 January
8	2.16	1 December	1 February	January 31	31 January
9	1.92	1 December	1 February	January 31	31 January

D. Satellite Data

In order to ensure the acquisition of data that is both accurate in its representation and precise in its measurements, we meticulously chose Sentinel-2 and Sentinel-1 images which were captured on a date that was in close temporal proximity to the date on which the data was collected in the field. (Table.1) The data that has been obtained through this rigorous process will specifically pertain to the quantification of nitrogen uptake by plants during the aforementioned period of interest.

E. Spectral VI from Sentinel-2

The pre-processing of Sentinel-2 image data includes Geometric Resampling for all bands to a spatial resolution of 20 meters. The SNAP software is used to calculate relevant VI associated with estimating nitrogen levels in plant as specified in Table. 2

The selection of indices from Sentinel-2 for nitrogen analysis considers multiple spectral bands that are related to plant status and nitrogen absorption. Key indices include LAI for leaf area, Chl for chlorophyll related to nitrogen, and MCARI for nitrogen-related chlorophyll absorption. Other indices like NDVI and GNDVI assess plant density and health, while MSAVI and SAVI reduce soil impact. These indices provide efficient tools for nitrogen management and plant analysis.

TABLE. 2 THE INDICES UTILIZED OF ANALYSIS AND PROCESSING

Indices	Formula	Ref
LAI	Using Parameters of the Satellite Imaging System by SNAP (Sentinel Application Platform)	[15]
Chl		
FAPAR		
GNDVI	(B7 - B3) / (B7 + B3)	[16]
MCARI	((B5 - B4) - 1 × (B5 - B3)) × (B5 / B4)	[17]
MSAVI	(2 × B8 + 1 - √((2 × B8 + 1)² - 8 × (B8 - B4))) / 2	[18]
NDVI	(B8 - B4) / (B8 + B4)	[19]
RVI	B8 / B4	[20]
SAVI	(B8 - B4) / (B8 + B4 + 0.5)	[18]

*Leaf Area Index (LAI), Chlorophyll content in the Leaf (Chl), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Green Normalized Difference Vegetation Index (GNDVI), Modified Chlorophyll Absorption in Reflectance Index (MCARI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalized Difference Vegetation Index (NDVI), Ratio Vegetation Index (RVI), Soil-Adjusted Vegetation Index (SAVI)

F. Spectral VI from Sentinel-1

Sentinel-1 data is processed by calibrating the image with both VV and VH polarizations. Firstly, Noise Removal and Speckle Filtering are conducted by applying the Lee Sigma filter to reduce speckle noise with a window size of 5x5 and a target window size of 3x3. Afterward, multilook processing is used to further reduce speckle noise in radar images, resulting in clearer data that is easier to analyze. Lastly, terrain correction is done to minimize terrain distortion using the SRTM 3-Second Digital Elevation Model (DEM) with a resampling method of nearest neighbor with a pixel spacing of 10 m and converted into WGS84 coordinate reference. Various indices, as listed on Table 3, are then calculated.

Indices from Sentinel-1 use backscatter values to assess plant canopy structure, leaf characteristics, and density, which indicate plant health. CR measures wave reflection between polarizations to assess structure, while DPSVI and DpRVIC analyze plant density and nitrogen status. DPSVIm and RVIm enhance accuracy by reducing environmental impact, and H_c and thetac analyze plant distribution. These indices enable accurate plant analysis in all weather conditions.

TABLE.3 THE INDICES APPLIED FOR THE ASSESSMENT USING SENTINEL-1 SATELLITE DATA

Indices	Formula	Ref
CR	σ^0vv / σ^0vh	[21]
DPSVI	$DPSVI = IDPDD \times VDDPI \times \sigma^0vh$ Where: $IDPDD = [(\sigma^0vv(max) - \sigma^0vv) + \sigma^0vh] / \sqrt{2}$ $VDDPI = (\sigma^0vv + \sigma^0vh) / \sigma^0vv$	[22]
DpRVIC	$[\{(\sigma^0vv / \sigma^0vh) \times (\sigma^0vh / \sigma^0vv + 3)\} / (\sigma^0vh / \sigma^0vv + 1)^2]$	[23]
DPSVIm	$(\sigma^0vv^2 + \sigma^0vv \times \sigma^0vh + \sigma^0vh^2) / \sqrt{2}$	[24]
RVIm	$\sigma^0vh / 4 \times \sigma^0vv$	[25]
H_c	$H_c = -1 \times [(p1 \times \log_2(p1)) + (p2 \times \log_2(p2))]$ Where: $p1 = 1 / [1 + (\sigma^0vh / \sigma^0vv)]$ $p2 = 1 / [1 + (\sigma^0vv / \sigma^0vh)]$	[26]
thetac	$\tan^{-1} [(\sigma^0vh / \sigma^0vv)^2 / \{1 - (\sigma^0vh / \sigma^0vv)^2\}]$	[23]

*Cross Ratio (CR), Dual Polarization SAR Vegetation Index (DPSVI), Dual-polarization Radar Vegetation Index (DpRVIC), Modified Dual Polarization SAR Vegetation Index (DPSVIm), Modified Radar Vegetation Index (RVIm), Dual-polarimetric descriptors (H_c), Dual-polarimetric descriptors (thetac)

G. Regression analysis

Regression analysis is a quantitative method used to examine the relationship between one or more independent variables and a dependent variable.

$$y = mx + c \quad (2)$$

where y is the dependent variable or the outcome we are predicting, x is the independent variable or the predictor utilized to clarify variations in y, m is the slope of the line and c is y-intercept. The performance of the regression models developed using those listed VI from Sentinel-2 and Sentinel-1 is assessed. Multiple performance indicators are chosen.

The Root Mean Squared Error (RMSE) is defined as the square root of the Mean Squared Error (MSE) and serves as a widely employed metric for evaluating the precision of predictive models.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

where n is the total amount of data, y_i is the observe values, and \hat{y}_i is the predict values.

The Pearson Correlation Coefficient (r) formula is utilized to evaluate the linear relationship between two variables, and its value ranges from -1 to 1

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (4)$$

where n is the total amount of data, x_i is the value of the variable x at locus i, \bar{x}_i is the mean of the variable x, y_i is the value of the variable y at locus i and \bar{y}_i is the mean of the variable x.

III. RESULTS AND DISCUSSION

A. Nitrogen Uptake Estimation Model from Sentinel-2

According to the data analysis, the MCARI showed the strongest correlation in estimating nitrogen status in rice. It showed R^2 value of 0.8709, indicating a high level of accuracy, and RMSE of 10.822 kg/ha, reflecting a low margin of error. Additionally, the correlation coefficient ($r = 0.933$) indicates a strong positive relationship between the MCARI and nitrogen content in rice. When compared to other indices from Sentinel-2, such as RVI, GNDVI, and Chl, MCARI proved to be the most consistent and reliable in assessing nitrogen status (Table 4).

The MCARI indices quantify the wavelengths absorbed by chlorophyll, namely in the blue and red spectra, where chlorophyll in plant foliage captures light. The reflection of the green wavelength results in the green coloration of leaves. By integrating the Red Edge band into its computation, MCARI improves its precision in identifying variations in chlorophyll content. The index's efficacy in measuring chlorophyll correlates with nitrogen detection in rice fields, given that nitrogen is an essential constituent of chlorophyll and is vital for photosynthesis and plant development.

The linear regression analysis indicated that the regression equation possesses a slope of 582.22 and y-intercept of -22.075. This unequivocally demonstrates the variation in nitrogen estimate in plants relative to the MCARI index. The MCARI index, which corresponds with chlorophyll analysis results, exhibits the strongest association in evaluating the nitrogen condition of rice fields (Fig.2), rendering it an effective instrument for detecting nitrogen levels in agricultural regions.

The analysis results using MCARI and EVI indices in this research align (Fabbri, 2020) study, which showed these indices are effective tools for assessing nitrogen content and biomass in wheat. This study explores the use of RapidEye satellite data for nitrogen management in Durum wheat by assessing nitrogen status through the Nitrogen Nutrition Index (NNI), used to evaluate plant nitrogen levels. [27]

TABLE.4 RELATIONSHIPS BETWEEN FIELD DATA AND SENTINEL-2 INDICES

Index	Estimate Model	R^2	RMSE	r
SAVI	$y = 203x - 31.975$	0.7527	12.805	0.867
MSAVI	$y = 204.07x - 22.395$	0.7548	14.918	0.868
FAPAR	$y = 186.01x - 55.141$	0.7813	14.087	0.883
LAI	$y = 55.662x - 35.726$	0.7892	13.832	0.888
NDVI	$y = 180.62x - 78.521$	0.8167	12.897	0.903
Chl	$y = 0.7065x - 1.8986$	0.8185	12.832	0.904
GNDVI	$y = 202.46x - 68.62$	0.8204	12.765	0.905
RVI	$y = 4.5556x + 9.3548$	0.8219	12.715	0.906
MCARI	$y = 582.22x - 22.075$	0.8709	10.822	0.933

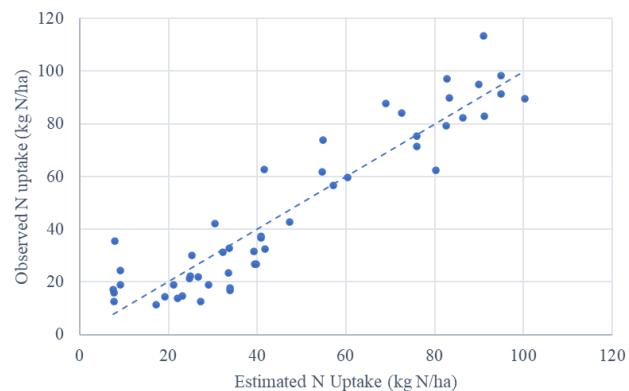


Fig. 2. Estimate model of rice nitrogen uptake from MCARI

B. Nitrogen Uptake Estimation Model from Sentinel-1

The data analysis indicated that the DPSVI index from Sentinel-1 exhibits the highest connection with nitrogen levels in rice ($R^2 = 0.8345$, RMSE = 12.254 kg/ha). The r value of 0.913 demonstrates a robust positive association between the DPSVI and nitrogen estimation in rice.

Table 5 shows that DPSVI exhibits superior accuracy in evaluating nitrogen status relative to other indices, including RVIm, DpRVIC, and CR. However, Fig. 3 depicts the categorization of the outcomes into two distinct regions. The green circle shows the rice farming area in Mueang Suphan Buri District, a region characterized by high fertility. The soil comprises a blend of loamy and clayey components, enriched with a substantial quantity of organic matter. Organic matter is a crucial element that augments soil fertility, strengthens soil structure, boosts water and nutrient retention, and facilitates optimal plant growth. The blue circle shows samples from the region in Don Chedi District, Sam Chuk District, and Si Prachan District, characterized by clay and loamy soil intermixed with clay, exhibiting a low organic matter content relative to Mueang Suphan Buri District.

The linear regression of DPSVI analysis revealed a strong correlation in predicting nitrogen estimates, with a slope of 0.1961 and a y-intercept of 159.08. The slope signifies that nitrogen status rises with an increase in the DPSVI value, while the y-intercept denotes the projected nitrogen value when the DPSVI is zero. The DPSVI may be used to accurately measure nitrogen levels, but further study is necessary to ensure robustness of its application in area with different levels of soil fertility.

TABLE.5 RELATIONSHIPS BETWEEN FIELD DATA AND SENTINEL-1 INDICES

Index	Estimate Model	R^2	RMSE	r
DPSVIm	$y = -0.1689x + 18.195$	0.0492	29.376	0.221
thetaC	$y = -60.54x + 64.832$	0.1913	27.092	0.437
H_c	$y = -734x + 135.06$	0.4022	23.293	0.634
CR	$y = 228.55x - 80.738$	0.403	23.276	0.634
DpRVIC	$y = 672.3x - 491.07$	0.4162	23.018	0.645
RVIm	$y = -0.1532x + 6.2371$	0.7102	49.939	0.842
DPSVI	$y = 0.1961x + 159.08$	0.8345	12.254	0.913

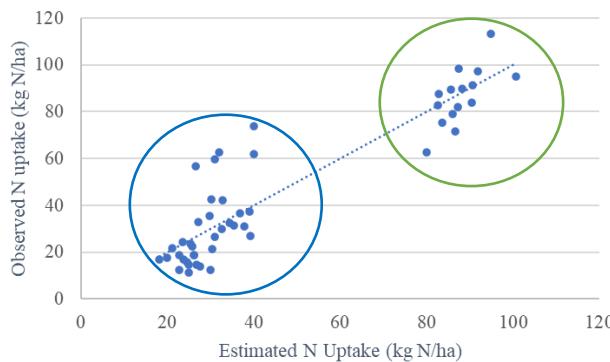


Fig. 3. Estimate model of rice nitrogen uptake form DPSVI

C. Accuracy assessment

To validate the precision of our research, we separated the data into two segments: 70% for training and 30% for testing, to ascertain that the generated model can accurately forecast nitrogen levels in rice fields.

The developed model was evaluated using the provided test dataset, yielding a R^2 of 0.748, RMSE of 16.614 kg/ha, and r of 0.865 for the MCARI from Sentinel-2, indicating its efficacy in predicting nitrogen levels. However, based on Sentinel-1's DPSVI outcomes, which exhibited $R^2 = 0.814$, RMSE = 13.678 kg/ha, and $r = 0.902$, it may be inferred that DPSVI surpasses MCARI in predicting efficacy. DPSVI is more effective than MCARI because it can be assessed from the plant's canopy, which reflects growth. A strong canopy often indicates the health of the plant.

D. Nitrogen Uptake Map

The models developed using relationship between nitrogen levels in rice fields and the satellite derive data from Sentinel-2 and Sentinel-1 can be used to spatially visualize nitrogen uptake. Fig. 4 illustrates nitrogen uptake maps based on MCARI and DPSVI models from Sentinel-2 and Sentinel-1, respectively. Both approaches, utilized an assessment range of 0 to 150 kg/ha, revealed that plots 5, 6, and 7 are regions of elevated fertility, consistent with soil analysis conducted prior to rice cultivation which indicated high organic matter content resulting from the soil preparation phase. Soil organic matter is essential for nutrient and water retention, enabling plants to effectively utilize nitrogen. However, the map from DPSVI showed more homogeneous results within fields compared to MCARI, especially for plot 5-7.

These satellite-based maps identify regions with elevated or diminished nitrogen concentrations and offer information regarding soil fertility and plant growth conditions in each plot. Moreover, the development of such maps facilitates more efficient planning of fertilizer management and maintenance of rice fields, enabling the identification of locations necessitating greater nitrogen or those potentially receiving excessive amounts. This allows farmers to more effectively evaluate and administer resources in their fields, leading to enhanced yields and less superfluous fertilizer application.

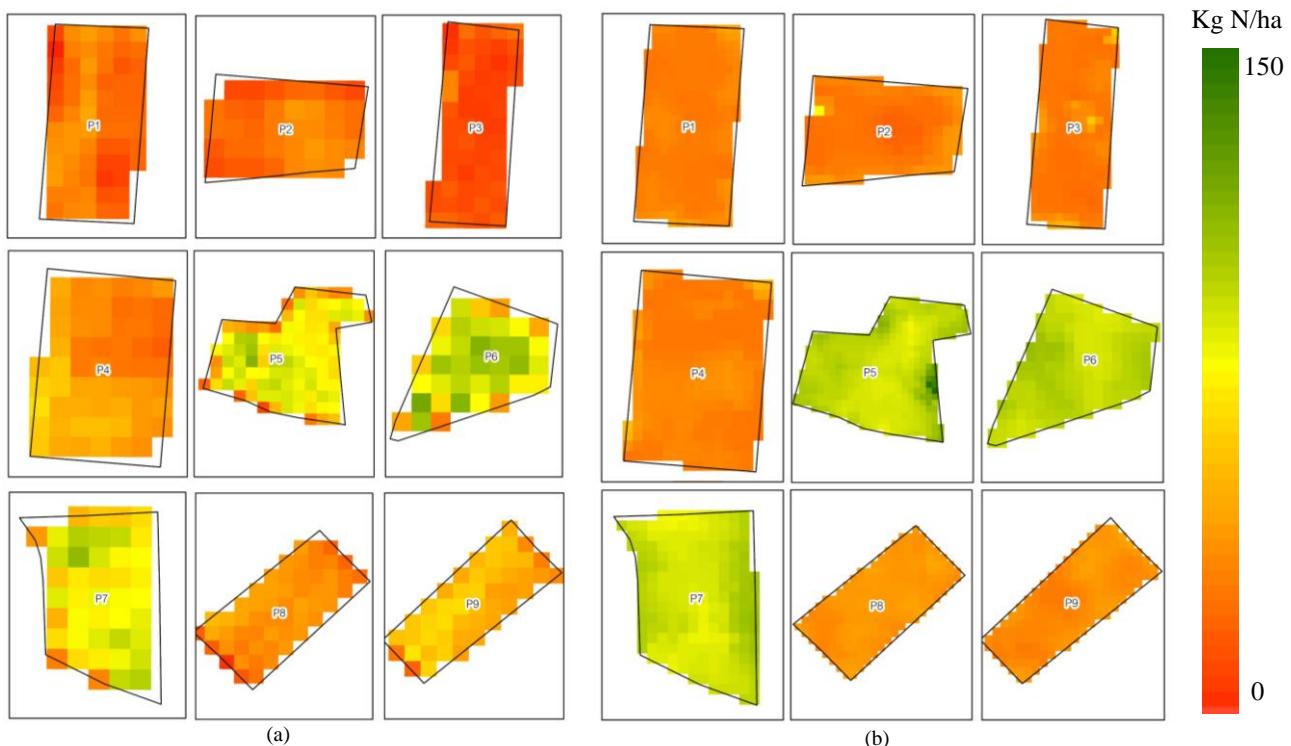


Fig.4 Sentinel-2 MCARI (a) and Sentinel-1 DPSVI (b) derived Rice nitrogen Uptake and Field

IV CONCLUSION

This study evaluates satellite indices to create maps of nitrogen absorption that are crucial for precision agriculture, particularly nitrogen recommendations in rice fields. The main tools used in this work are Sentinel-2 and Sentinel-1. Although MCARI from Sentinel-2 can estimate nitrogen concentration through plant reflectivity, it is limited by cloud cover. DPSVI from Sentinel-1 collects data in all weather, day or night, using SAR radar, but requires complex processing and has lower imaging frequency. Future studies should expand the study area to improve data accuracy. Incorporating multivariate algorithms like Random Forest (RF), Support Vector Machines (SVM), and others can enhance nitrogen absorption. The developed model was evaluated using the provided test dataset, yielding good performance from MCARI and greater performance from DPSVI. Therefore, Sentinel-1's DPSVI-based model is promising to be further developed as an precisely and effectively method to evaluate nitrogen levels in rice fields.

ACKNOWLEDGMENT

This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development Research and Innovation [grant number B13F670082-8].

The authors thank Geo-Informatics and Space Technology Development Agency (Public Organization) and the University of Phayao for supporting the tools and software to conduct this study.

REFERENCES

- [1] Mohidem, N. A., Hashim, N., Shamsudin, R., & Che Man, H. (2022). Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content. *Agriculture*, 12(6), 741. <https://doi.org/10.3390/agriculture12060741>
- [2] Zhao, C., Liu, G., Chen, Y., Jiang, Y., Shi, Y., Zhao, L., Liao, P., Wang, W., Xu, K., Dai, Q., & Huo, Z. (2022). Excessive nitrogen application leads to lower rice yield and grain quality by inhibiting the grain filling of inferior grains. *Agriculture*, 12(7), 962. <https://doi.org/10.3390/agriculture12070962>
- [3] Zhang, Z., Wang, Y., Wang, Z., Ashraf, U., Mo, Z., Tian, H., Duan, M., Li, Y., Tang, X., & Pan, S. (2022). Precise delivery of nitrogen at tillering stage enhances grain yield and nitrogen use efficiency in double rice cropping systems of South China. *Field Crops Research*, 289(108736), 108736. <https://doi.org/10.1016/j.fcr.2022.108736>
- [4] She, Y., Li, P., Du, Z., Qi, X., Zhao, S., Li, T., & Guo, W. (2022). Nitrogen fertilization effects on soil nitrate, water use, growth attributes and yield of winter wheat under shallow groundwater table condition. *Agronomy* (Basel, Switzerland), 12(12), 3048. <https://doi.org/10.3390/agronomy12123048>
- [5] Lu, B., Li, Z., Yang, F., Xu, C., Tang, X., Zhang, J., Ding, C., Xu, D., Zhu, G., Zhang, J., Li, W., Qian, H., Jiang, Y., Liu, Z., Ding, Y., & Li, G. (2024). Nitrogen application increased yield sensitivity of indica hybrid rice to climate resource. *European Journal of Agronomy: The Journal of the European Society for Agronomy*, 159(127257), 127257. <https://doi.org/10.1016/j.eja.2024.127257>
- [6] Lee, S. (2021). Recent advances on nitrogen use efficiency in rice. *Agronomy* (Basel, Switzerland), 11(4), 753. <https://doi.org/10.3390/agronomy11040753>
- [7] Ramalingam, S., Panneerselvam, B., & Kalaiappan, S. P. (2022). Effect of high nitrate contamination of groundwater on human health and water quality index in semi-arid region, South India. *Arabian Journal of Geosciences*, 15(3). <https://doi.org/10.1007/s12517-022-09553-x>
- [8] Wang, D., Struik, P. C., Liang, L., & Yin, X. (2024). Developing remote sensing- and crop model-based methods to optimize nitrogen management in rice fields. *Computers and Electronics in Agriculture*, 220(108899), 108899. <https://doi.org/10.1016/j.compag.2024.108899>
- [9] Khon Kaen Agriculture Journal. (n.d.). Tci-thaijo.org. Retrieved August 3, 2024, from <https://li01.tci-thaijo.org/index.php/agkasetkaj>
- [10] Li, D., Zhang, P., Chen, T., & Qin, W. (2020). Recent development and challenges in spectroscopy and machine vision technologies for crop nitrogen diagnosis: A review. *Remote Sensing*, 12(16), 2578. <https://doi.org/10.3390/rs12162578>
- [11] Sentinel NASA, "Sentinel-1 and Sentinel-2" Accessed: Aug. 3, 2024. [Online]. Available <https://sentinel.esa.int/web/sentinel/home>
- [12] Suphanburi Province Office, "General Information about the province" Accessed: Aug. 3, 2024. [Online]. Available: <https://www.suphanburi.go.th/content/general>
- [13] Thai Meteorological Department, "Climate Suphanburi" Accessed: Aug. 3, 2024. [Online]. Available: <http://climate.tmd.go.th/content/article/7>
- [14] Ministry of Agriculture and Cooperatives, "Rice" Accessed: Aug. 10, 2024. [Online]. Available: <https://www.moac.go.th/news-preview-451291791121>
- [15] S2ToolBox Level 2 products, "LAI, FAPAR, FCOVER" Accessed: 5 Aug 2024. [Online]. Available https://step.esa.int/docs/extras/ATBD_S2ToolBox_L2B_V1.1.pdf
- [16] Soltanikazemi, M., Minaei, S., Shafizadeh-Moghadam, H., & Mahdavian, A. (2022). Field-scale estimation of sugarcane leaf nitrogen content using vegetation indices and spectral bands of Sentinel-2: Application of random forest and support vector regression. *Computers and Electronics in Agriculture*, 200, 107130. <https://doi.org/10.1016/j.compag.2022.107130>
- [17] Fang, G., He, X., Weng, Y., & Fang, L. (2023). Texture features derived from Sentinel-2 vegetation indices for estimating and mapping forest growing stock volume. *Remote Sensing*, 15(11), 2821. <https://doi.org/10.3390/rs15112821>
- [18] Muchsin, F., Supriyatna, A., Harmoko, I., & Prasasti, I. (2022). Evaluation of atmospheric correction methods of Sentinel-2 for monitoring paddy rice growth in Cianjur and Klaten Regency. *Journal of Physics: Conference Series*, 2243, 012023. <https://doi.org/10.1088/1742-6596/2243/1/012023>
- [19] Nadzirah, R., Rizqon, M. K., & Indarto. (2024). Application of Sentinel-2A images for land cover classification using NDVI in Jember Regency. *Geosfera Indonesia*, 9(1), 41-55. <https://doi.org/10.19184/geosi.v9i1.28846>
- [20] Ma, T., Hu, Y., Wang, J., Beckline, M., Pang, D., Chen, L., Ni, X., & Li, X. (2023). A novel vegetation index approach using Sentinel-2 data and random forest algorithm for estimating forest stock volume in the Helan Mountains, Ningxia, China. *Remote Sensing*, 15(7), 1853. <https://doi.org/10.3390/rs15071853>
- [21] Frison, P.-L., Fruneau, B., Kmicha, S., Soudani, K., Dufrêne, E., Le Toan, T., Koleck, T., Villard, L., Mougin, E., & Rudant, J.-P. (2018). Potential of Sentinel-1 data for monitoring temperate mixed forest phenology. *Remote Sensing*, 10(12), 2049. <https://doi.org/10.3390/rs10122049>
- [22] Dos Santos, E. P., Da Silva, D. D., & do Amaral, C. H. (2021). Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. *International Journal of Remote Sensing*, 42(19), 7581–7609. <https://doi.org/10.1080/01431161.2021.1959955>
- [23] Bhogapurapu, N., Dey, S., Mandal, D., Bhattacharya, A., Karthikeyan, L., McNairn, H., & Rao, Y. S. (2022). Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data. *Remote Sensing of Environment*, 271(112900), 112900. <https://doi.org/10.1016/j.rse.2022.112900>
- [24] Dos Santos, E. P., Da Silva, D. D., & do Amaral, C. H. (2021). Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. *International Journal of Remote Sensing*, 42(19), 7581–7609. <https://doi.org/10.1080/01431161.2021.1959955>
- [25] Nasirzadehdzaji, R., Balik Sanli, F., Abdikan, S., Cakir, Z., Sekertekin, A., & Ustuner, M. (2019). Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. *Applied Sciences* (Basel, Switzerland), 9(4), 655. <https://doi.org/10.3390/app9040655>
- [26] Bhogapurapu, N., Dey, S., Bhattacharya, A., Mandal, D., Lopez-Sanchez, J. M., McNairn, H., López-Martínez, C., & Rao, Y. S. (2021). Dual-polarimetric descriptors from Sentinel-1 GRD SAR data for crop growth assessment. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, 178, 20–35. <https://doi.org/10.1016/j.isprsjprs.2021.05.013>
- [27] Fabbri, C. et al. (2020). Integrating satellite data with a Nitrogen Nutrition Curve for durum wheat. *Eur. J. Agron.*, 120, 126148. <https://doi.org/10.1016/j.eja.2020.126148>

Accuracy Assessment of UAV-Derived Data from Multi-Flight Plans with RTK GNSS Surveys

Tinn Thirakultomorn

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology
Isan*

Nakhon Ratchasima, Thailand
tinn.th@rmuti.ac.th

Sumitta Khosungnoen

*Department of Survey Engineering,
Faculty of Engineering and
Technology, Rajamangala University of
Technology Isan*

Nakhon Ratchasima, Thailand
sumitta.kh@rmuti.ac.th

Athiwat Phinyoyang*

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology
Isan*

Nakhon Ratchasima, Thailand
athiwat.pn@rmuti.ac.th

Wilawan Prasomsup

*Research Unit of Geoinformatics
Innovation for Spatial Management,
Department of Transportation, Faculty
of Railway systems and Transportation,
Rajamangala University of Technology
Isan*

Nakhon Ratchasima, Thailand
wilawan.pa@rmuti.ac.th

Kodchakon Phonak

*Department of Survey Engineering,
Faculty of Engineering and
Technology, Rajamangala University of
Technology Isan*

Nakhon Ratchasima, Thailand
kodchakorn.pu@rmuti.ac.th

Abstract—This study aimed to create an orthophotograph and digital surface model (DSM) using an unmanned aerial vehicle (UAV) with a multi-flight plan and to assess their accuracy by comparing the results to real-time kinematic (RTK) survey data. This study also focuses on the limited comparisons of various flight plans and their effects on the accuracy of UAV-derived data. The study area covered 4,000 m² and employed various flight plans with speed and trigger mode adjustments while maintaining a consistent flying height for a ground sample distance (GSD) of 5 cm and 70% overlap in both front and side directions. Aerial imagery was processed using Pix4D Mapper software, and RTK GNSS was utilized to obtain coordinates and elevations at predefined survey points. The methodology consisted of three main components: UAV surveying, RTK GNSS surveys, and accuracy assessments of both horizontal and vertical dimensions. Survey points were arranged in a grid pattern, totaling 30 survey points. The DJI Phantom 4 Multispectral RTK was employed for the UAV survey, while the V90 Plus GNSS RTK System was used for the RTK surveys. The accuracy assessment was evaluated using the Root Mean Square Error (RMSE) for the X and Y coordinates and overall horizontal accuracy, while vertical accuracy focused on Z values. The overall assessment provided a comprehensive view of data precision. Results showed that the FE flight plan achieved the highest accuracy for horizontal coordinates (RMSE of 0.19), whereas the NE flight plan demonstrated the highest vertical accuracy (RMSE of 0.15). Ultimately, the overall accuracy assessment (RMSE_T) indicated that the NE flight plan, utilizing Normal Flight Speed and Equal Distance Trigger modes, provided the best accuracy with an RMSE of 0.38. These results highlight the effectiveness of UAVs in generating accurate geospatial data for various applications.

Keywords—UAV flight plans, RTK survey, Orthophotograph, DSM, Accuracy Assessment.

I. INTRODUCTION

A. Problem statement

Currently, unmanned aerial vehicles (UAVs) are being applied in various fields, especially in surveying, including construction, agriculture, architecture, urban planning, high-resolution aerial photography and terrain mapping, and geographic information systems (GIS). UAVs have led to

significant advancements in remote sensing technologies, emerging as powerful tools for capturing aerial imagery and generating accurate digital surface models (DSMs). This is due to the rapid and cost-effective processes associated with UAV surveying [1][2][3]. UAVs are readily accessible, can be operated using smartphones and free flight control applications, and have access to free cloud-based software for processing aerial imagery.

However, data obtained from UAV surveys, including aerial photographs and terrain maps, lack standardized accuracy for positional and elevation information. This contrasts with ground surveys using total stations, which adhere to established surveying standards. Nevertheless, ground surveys are subject to procedural limitations and challenges in areas with highly variable terrain. Additionally, UAV surveys differ from satellite-based surveys (Global Navigation Satellite System, GNSS), which have accuracy defined by the surveying process, such as RTK surveys that provide precision of 1-5 cm. Nonetheless, GNSS surveys have limitations related to satellite signal transmission [4]. Furthermore, UAV surveys lack established flight patterns or operational methods that ensure high accuracy and precision in data collection [5].

Therefore, this study aims to assess the accuracy of UAV-derived data (orthophotos and DSMs) from multi-flight plans by comparing the results to RTK GNSS survey data. The multi-flight plan integrates different flight settings, including flight speed and image recording modes, while other settings remain the same. The findings are expected to identify suitable flight plans and their advantages in enhancing the accuracy and precision of orthophotos and DSMs.

B. Objectives

The objectives aim to create an orthophotograph and digital surface model (DSM) using a UAV with a multi-flight plan and to assess their accuracy by comparing them to data obtained from real-time kinematic (RTK) surveys.

C. Scope of the study

The study area is a square, with a total coverage of 4,000 m², and utilizes various flight plans, including speed and mode

adjustments, while maintaining a consistent flying height for a ground sample distance (GSD) of 5 cm and overlaps (front and side overlap) of 70%. Aerial imagery will be processed with Pix4D Mapper software. RTK GNSS will obtain coordinates and elevation of predefined survey points. The accuracy of the aerial imagery will be assessed by the Root Mean Square Error (RMSE) for the coordinates and elevation compared to RTK GNSS data.

II. METHODOLOGY

The methodology comprises three main components: (1) surveying using UAVs, (2) surveys with RTK GNSS, and (3) assessing horizontal and vertical accuracy (See Fig. 1).

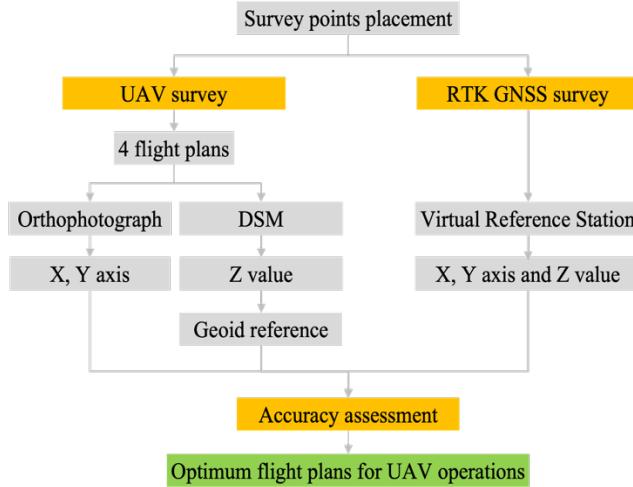


Fig. 1. The research methodology

A. Survey points placement

The survey points will be distributed in a grid pattern [4] (See Fig. 2). This method provides the best accuracy in both horizontal and vertical dimensions. The grid size is 50 m x 50 m, resulting in 16 grids, with at least one survey point placed in each grid, totaling 30 survey points. Additionally, each point will be spaced no less than one-tenth of the diagonal length of the grid from one another [6].



Fig. 2. The distribution of survey points

B. UAV survey

The DJI Phantom 4 Multispectral RTK with the D-RTK 2 Mobile Base Station is used for surveying, while the DJI GS PRO application is employed for flight control.

The flight plan settings for the UAV survey consist of four options: NH, NE, FH, and FE. In this context, N represents Normal Flight speed (7.7 m/s), F denotes Fast Flight speed (14.8 m/s), H indicates Hovering Trigger mode, and E stands for Equal Distance Trigger mode. For the other setting, the GSD, front and side overlaps, and the flight pattern of the UAV follows a grid formation for 2D maps, as shown in Figure 3. Overall, selecting these flight plans aims to optimize data quality by adjusting speed and image capture modes specifically for aerial photo mapping tasks.

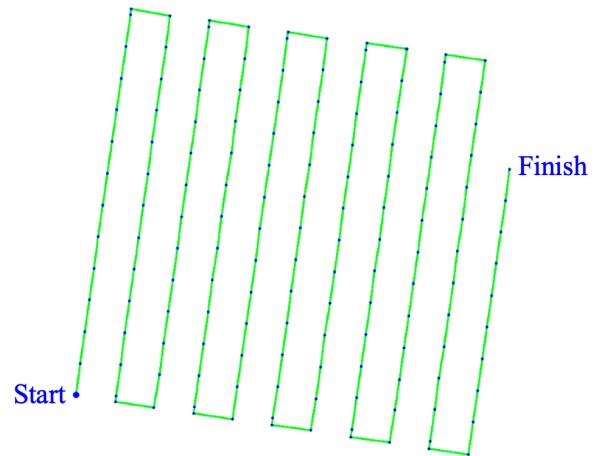


Fig. 3. A grid formation for 2D maps

The height obtained from the UAV will be calibrated from the ellipsoid reference to the geoid reference, similar to the data from RTK GNSS. This calibration will be performed using the geoid height calculator available on the UNAVCO website (See Fig. 4). The relationship is defined as:

$$H = h + N \quad (1)$$

Where: H is the Orthometric height, h is the Ellipsoidal height from GPS, and N is the Geoid height.

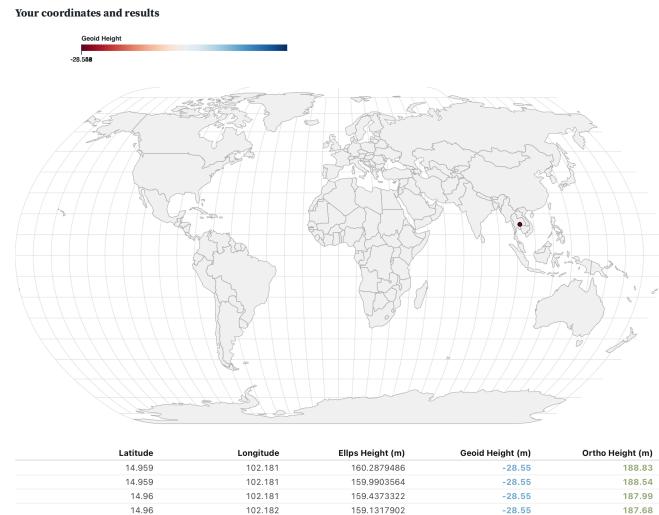


Fig. 4. Geoid Height Calculator

C. Orthomosaic and DSM processing

Orthomosaic and DSM processing involves several vital steps [7][8][9]. First, image matching is performed to calculate positions and match images, creating tie points to ensure accuracy. Next, georeferencing adjusts these tie points using control points through bundle block adjustment, followed by accuracy verification. The process then generates a dense point cloud to create detailed 3D coordinates, allowing for the removal of any anomalies. Afterward, a mesh model is created to produce surface data for DSM, focusing on resolution. Finally, high-quality aerial imagery is made as a true orthophoto, ensuring appropriate scale and modeling accuracy.

D. RTK GNSS survey

The GNSS equipment used for relative positioning surveys using the RTK method [10] was the V90 Plus GNSS RTK System. This system measures coordinates and elevation, utilizing the Virtual Reference Station (VRS) provided by the Department of Lands [11].

The operational principle of the VRS system allows users to send their approximate position coordinates to a control center and request position corrections via GSM/GPRS/EDGE/3G communication. The control station then generates a VRS model from all reference stations within the network. This VRS station is situated close to the user's position, ensuring high accuracy in processing by utilizing a mathematical model to adjust for errors caused by atmospheric conditions and environmental factors in the surrounding area.

E. Accuracy assessment

The accuracy assessment follows the National Standard for Spatial Data Accuracy (NSSDA) outlined by Thailand's Federal Geographic Data Committee (FGDC). FGDC adheres to the ASPRS accuracy standards for large-scale maps based on statistical principles to report spatial data accuracy. These standards express the standard error at a specific confidence level, measured in ground distance [12].

The horizontal positional accuracy evaluates the accuracy of the eastern and northern coordinates in the WGS 1984 UTM Zone 48N reference. The results are presented as the Root Mean Square Error (RMSE) for the X and Y coordinates and the overall horizontal accuracy. Vertical positional accuracy focuses on the accuracy of elevation values. The results are expressed as the RMSE for elevation. Overall accuracy assessment evaluates horizontal and vertical accuracy together, providing a comprehensive view of the data's precision. The RMSE equations are defined as follows [13]:

$$RMSE_X = \sqrt{\frac{\sum(X_{data_i} - X_{check_i})^2}{n}} \quad (2)$$

$$RMSE_Y = \sqrt{\frac{\sum(Y_{data_i} - Y_{check_i})^2}{n}} \quad (3)$$

$$RMSE_Z = \sqrt{\frac{\sum(Z_{data_i} - Z_{check_i})^2}{n}} \quad (4)$$

$$RMSE_R = \sqrt{RMSE_X^2 + RMSE_Y^2} \quad (5)$$

$$RMSE_T = \sqrt{RMSE_R^2 + RMSE_Z^2} \quad (6)$$

Where: $RMSE_X$, $RMSE_Y$, and $RMSE_Z$ are root mean square errors in the X, Y, and Z axis, respectively. $RMSE_R$ is the horizontal positional error, while $RMSE_T$ is the overall positional error. The coordinate of test point i in the test set for X, Y, and Z data are represented by X_{data_i} , Y_{data_i} , and Z_{data_i} , respectively. In contrast, X_{check_i} , Y_{check_i} , Z_{check_i} refer to the coordinates of the reference measurement for test point i for X, Y, and Z data, respectively.

III. RESULTS

A. An orthophograph and DSM

Based on the NH, NE, FH, and FE flight plans, processing transformed UAV imagery into high-quality outputs, including an orthophoto and DSM. In the initial processing phase (See Fig. 5), key points were computed, images were calibrated and oriented, standard features were matched, and a 3D point cloud was generated. The point cloud and mesh generation step improved the density and accuracy of the 3D model by cleaning and classifying the point cloud, followed by creating the DSM and mesh. The final outputs were produced by blending the images.

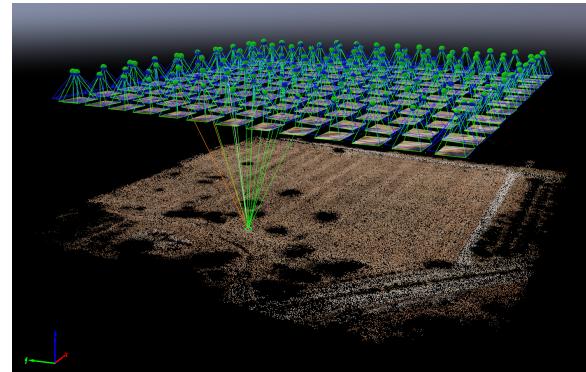


Fig. 5. Tie points matching process

One of the results from the NH flight plan featuring the orthophotograph (See Fig. 6) and DSM is shown in Figure 7. The DSM corresponds to this study's digital terrain model (DTM) since the area consists of bare soil. The spatial resolution of the orthophotograph achieved after processing is characterized by a cell size of 5.20 cm, allowing for a detailed and accurate representation of the ground control point on the orthophotograph, shown in Figure 8)



Fig. 6. Orthophograph

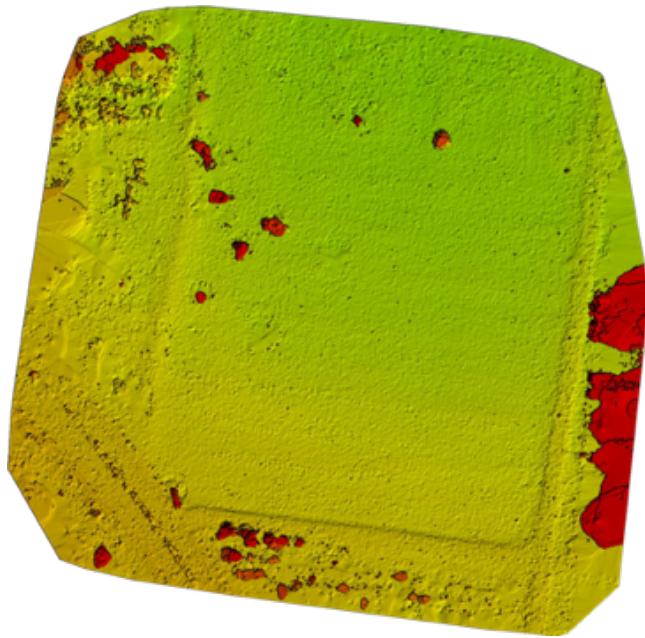


Fig. 7. Digital Surface Model (DSM)



Fig. 8. Ground control point on orthophotograph

The results of the UAV survey for each flight plan include the number of images captured and the duration of the survey flights, as shown in Table 1. Notably, the FE flight plan had the shortest flight duration, taking only 4 minutes and 13 seconds.

TABLE I. SUMMARY OF UAV FLIGHT SURVEY

Flight plan	Flight speed	Trigger mode	Number of images	Flight duration
NH	7.7 m/s	Hovering	155	14 m 54 s
NE	7.7 m/s	Equal distance	155	7 m 43 s
FH	14.8 m/s	Hovering	155	9 m 4 s
FE	14.8 m/s	Equal distance	155	4 m 13 s

B. An accuracy assessment

The accuracy assessment results for each flight plan are presented in Tables II to V for the NH, NE, FH, and FE flight plans, respectively. These tables indicate horizontal and vertical positional accuracy, as well as overall accuracy, as detailed in Table VI. The horizontal positional accuracy was evaluated based on the WGS 1984 UTM Zone 48N reference coordinates. These results were presented as the RMSE for the X and Y coordinates (RMSE_X and RMSE_Y) and the overall horizontal accuracy (RMSE_H). For vertical positional accuracy, the assessment focused on the accuracy of elevation values, which were expressed as the RMSE for elevation (RMSE_Z). The overall accuracy assessment integrated

horizontal and vertical evaluations, offering a comprehensive view of the data's precision (RMSE_T).

TABLE II. THE ACCURACY ASSESSMENT BETWEEN RTK GNSS DATA AND THE UAV DATA WITH THE NH FLIGHT PLAN

Point	diff X	diff Y	diff Z	Sq (diff X)	Sq (diff Y)	Sq (diff Z)
1	0.23	0.33	1.81	0.05	0.11	3.28
2	0.30	0.15	1.90	0.09	0.02	3.61
3	0.21	-0.13	2.01	0.04	0.02	4.04
4	0.15	-0.25	2.11	0.02	0.06	4.45
5	0.31	-0.42	1.90	0.10	0.18	3.61
6	0.06	-0.69	2.01	0.00	0.48	4.04
7	0.25	-0.91	1.69	0.06	0.83	2.86
8	-0.17	-0.88	1.94	0.03	0.77	3.76
9	0.01	-0.63	2.05	0.00	0.40	4.20
10	-0.24	-0.44	2.11	0.06	0.19	4.45
11	0.01	-0.27	2.12	0.00	0.07	4.49
12	-0.17	-0.16	2.09	0.03	0.03	4.37
13	-0.05	-0.06	2.26	0.00	0.00	5.11
14	0.06	0.02	2.06	0.00	0.00	4.24
15	0.05	0.28	1.98	0.00	0.08	3.92
16	-0.21	0.40	1.79	0.04	0.16	3.20
17	-0.38	0.22	1.94	0.14	0.05	3.76
18	-0.20	0.13	2.15	0.04	0.02	4.62
19	-0.34	-0.05	2.14	0.12	0.00	4.58
20	-0.37	-0.24	2.14	0.14	0.06	4.58
21	-0.53	-0.42	2.03	0.28	0.18	4.12
22	-0.42	-0.66	1.96	0.18	0.44	3.84
23	-0.89	-0.71	1.58	0.79	0.50	2.50
24	-0.67	-0.59	1.90	0.45	0.35	3.61
25	-0.76	-0.39	1.89	0.58	0.15	3.57
26	-0.63	-0.15	2.03	0.40	0.02	4.12
27	-0.65	0.08	1.83	0.42	0.01	3.35
28	-0.48	0.10	1.96	0.23	0.01	3.84
29	-0.69	0.24	1.94	0.48	0.06	3.76
30	-0.55	0.32	1.83	0.30	0.10	3.35
Sum of Sq (Diff)					5.07	5.35
X and Y coordinate accuracy					0.41	0.42
Horizontal and Vertical accuracy					0.59	1.98
Total accuracy					2.06	

TABLE III. THE ACCURACY ASSESSMENT BETWEEN RTK GNSS DATA AND THE UAV DATA WITH THE NE FLIGHT PLAN

Point	diff X	diff Y	diff Z	Sq (diff X)	Sq (diff Y)	Sq (diff Z)
1	0.03	0.04	0.05	0.00	0.00	0.00
2	0.00	-0.05	0.08	0.00	0.00	0.01
3	-0.03	-0.15	0.13	0.00	0.02	0.02
4	-0.06	-0.23	0.21	0.00	0.05	0.04
5	0.03	-0.31	0.15	0.00	0.10	0.02
6	-0.08	-0.45	0.03	0.01	0.20	0.00
7	0.04	-0.53	0.05	0.00	0.28	0.00
8	-0.10	-0.59	0.20	0.01	0.35	0.04
9	-0.11	-0.39	0.23	0.01	0.15	0.05
10	-0.22	-0.35	0.21	0.05	0.12	0.04
11	-0.08	-0.21	0.19	0.01	0.04	0.04
12	-0.14	-0.18	0.21	0.02	0.03	0.04
13	-0.07	-0.15	0.21	0.00	0.02	0.04
14	-0.05	-0.12	0.23	0.00	0.01	0.05
15	-0.05	0.01	0.14	0.00	0.00	0.02
16	-0.17	0.09	0.08	0.03	0.01	0.01
17	-0.25	-0.04	0.07	0.06	0.00	0.00
18	-0.14	-0.05	0.11	0.02	0.00	0.01
19	-0.21	-0.15	0.27	0.04	0.02	0.07
20	-0.20	-0.25	0.25	0.04	0.06	0.06
21	-0.25	-0.30	0.18	0.06	0.09	0.03
22	-0.22	-0.46	0.18	0.05	0.21	0.03
23	-0.46	-0.47	0.00	0.21	0.22	0.00
24	-0.34	-0.40	0.09	0.12	0.16	0.01
25	-0.37	-0.29	0.06	0.14	0.08	0.00
26	-0.32	-0.13	0.13	0.10	0.02	0.02
27	-0.29	-0.09	0.05	0.08	0.01	0.00
28	-0.26	-0.08	0.12	0.07	0.01	0.01
29	-0.34	0.01	0.04	0.12	0.00	0.00
30	-0.29	-0.03	0.08	0.08	0.00	0.01
Sum of Sq (Diff)					1.33	2.26
X and Y coordinate accuracy					0.21	0.27
Horizontal and Vertical accuracy					0.35	0.15
Total accuracy					0.38	

TABLE IV. THE ACCURACY ASSESSMENT BETWEEN RTK GNSS DATA AND THE UAV DATA WITH THE FH FLIGHT PLAN

Point	diff X	diff Y	diff Z	Sq (diff X)	Sq (diff Y)	Sq (diff Z)
1	0.24	0.23	2.85	0.06	0.05	8.12
2	0.25	0.11	3.04	0.06	0.01	9.24
3	0.19	-0.10	3.09	0.04	0.01	9.55
4	0.14	-0.23	3.26	0.02	0.05	10.63
5	0.24	-0.37	3.12	0.06	0.14	9.73
6	0.04	-0.64	2.98	0.00	0.41	8.88
7	0.23	-0.77	2.76	0.05	0.59	7.62
8	-0.13	-0.77	3.17	0.02	0.59	10.05
9	-0.01	-0.56	3.16	0.00	0.31	9.99
10	-0.21	-0.38	3.23	0.04	0.14	10.43
11	0.01	-0.26	3.25	0.00	0.07	10.56
12	-0.12	-0.12	3.24	0.01	0.01	10.50
13	-0.02	-0.04	3.25	0.00	0.00	10.56
14	0.04	0.00	3.24	0.00	0.00	10.50
15	0.02	0.19	3.23	0.00	0.04	10.43
16	-0.16	0.32	2.98	0.03	0.10	8.88
17	-0.31	0.19	3.14	0.10	0.04	9.86
18	-0.12	0.11	3.21	0.01	0.01	10.30
19	-0.29	-0.05	3.28	0.08	0.00	10.76
20	-0.33	-0.23	3.27	0.11	0.05	10.69
21	-0.42	-0.37	3.27	0.18	0.14	10.69
22	-0.36	-0.59	3.09	0.13	0.35	9.55
23	-0.77	-0.59	2.86	0.59	0.35	8.18
24	-0.55	-0.52	3.07	0.30	0.27	9.42
25	-0.63	-0.32	3.04	0.40	0.10	9.24
26	-0.51	-0.17	3.23	0.26	0.03	10.43
27	-0.51	0.06	2.94	0.26	0.00	8.64
28	-0.40	0.07	3.11	0.16	0.00	9.67
29	-0.56	0.17	2.98	0.31	0.03	8.88
30	-0.49	0.25	2.98	0.24	0.06	8.88
Sum of Sq (Diff)			3.52	3.95	290.86	
X and Y coordinate accuracy			0.34	0.36	-	
Horizontal and Vertical accuracy			0.50	3.11		
Total accuracy				3.15		

TABLE V. THE ACCURACY ASSESSMENT BETWEEN RTK GNSS DATA AND THE UAV DATA WITH THE FE FLIGHT PLAN

Point	diff X	diff Y	diff Z	Sq (diff X)	Sq (diff Y)	Sq (diff Z)
1	-0.21	-0.15	4.97	0.04	0.02	24.70
2	-0.24	-0.11	5.00	0.06	0.01	25.00
3	-0.25	-0.12	5.09	0.06	0.01	25.91
4	-0.19	-0.14	5.18	0.04	0.02	26.83
5	-0.19	-0.19	5.13	0.04	0.04	26.32
6	-0.22	-0.21	4.99	0.05	0.04	24.90
7	-0.15	-0.22	5.12	0.02	0.05	26.21
8	-0.10	-0.32	5.19	0.01	0.10	26.94
9	-0.18	-0.21	5.06	0.03	0.04	25.60
10	-0.21	-0.20	5.10	0.04	0.04	26.01
11	-0.17	-0.14	5.08	0.03	0.02	25.81
12	-0.15	-0.15	5.11	0.02	0.02	26.11
13	-0.14	-0.17	5.12	0.02	0.03	26.21
14	-0.19	-0.17	5.16	0.04	0.03	26.63
15	-0.18	-0.13	5.02	0.03	0.02	25.20
16	-0.20	-0.11	4.94	0.04	0.01	24.40
17	-0.21	-0.17	5.09	0.04	0.03	25.91
18	-0.21	-0.19	5.19	0.04	0.04	26.94
19	-0.14	-0.20	5.11	0.02	0.04	26.11
20	-0.17	-0.19	5.17	0.03	0.04	26.73
21	-0.18	-0.23	5.12	0.03	0.05	26.21
22	-0.21	-0.27	5.15	0.04	0.07	26.52
23	-0.24	-0.26	5.19	0.06	0.07	26.94
24	-0.18	-0.27	5.15	0.03	0.07	26.52
25	-0.17	-0.21	5.08	0.03	0.04	25.81
26	-0.15	-0.16	5.08	0.02	0.03	25.81
27	-0.17	-0.17	4.98	0.03	0.03	24.80
28	-0.16	-0.13	5.07	0.03	0.02	25.70
29	-0.17	-0.14	5.14	0.03	0.02	26.42
30	-0.15	-0.16	5.12	0.02	0.03	26.21
Sum of Sq (Diff)			1.02	1.08	779.41	
X and Y coordinate accuracy			0.18	0.19	-	
Horizontal and Vertical accuracy			0.26	5.10		
Total accuracy				5.10		

TABLE VI. SUMMARY OF THE ACCURACY ASSESSMENT RESULTS

Flight plan	RMSE _X	RMSE _Y	RMSE _R	RMSE _Z	RMSE _T
NH	0.41	0.42	0.59	1.98	2.06
NE	0.21	0.27	0.35	0.15	0.38
FH	0.34	0.36	0.50	3.11	3.15
FE	0.18	0.19	0.26	5.10	5.10

Based on Table II, the RMSE_X and RMSE_Y values are relatively low, indicating that the X and Y coordinates of all points from the orthophotograph differ only slightly from the results obtained through RTK GNSS (See Fig. 9 and 10). The FE flight plan provided the highest accuracy for X and Y coordinates, with an RMSE of 0.18 and 0.19, respectively, while the NH flight plan had the lowest X and Y accuracy, with an RMSE of 0.41 and 0.42, respectively. For RMSE_R, the FE flight plan demonstrated the highest accuracy, with an RMSE of 0.26. This indicates that the horizontal accuracy of the data obtained from this flight plan is superior to the other plans.

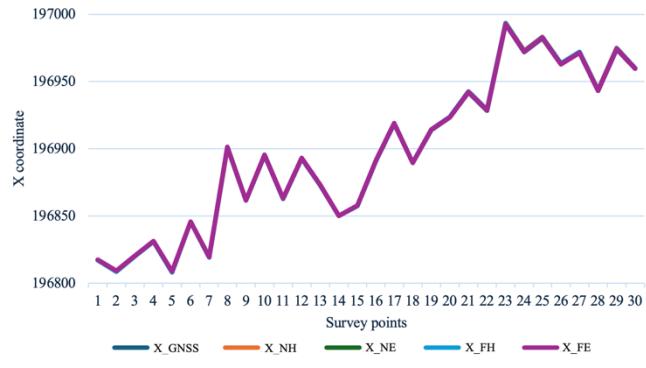


Fig. 9. Comparison of X coordinates between RTK GNSS survey data and UAV-derived data across the four flight plans.

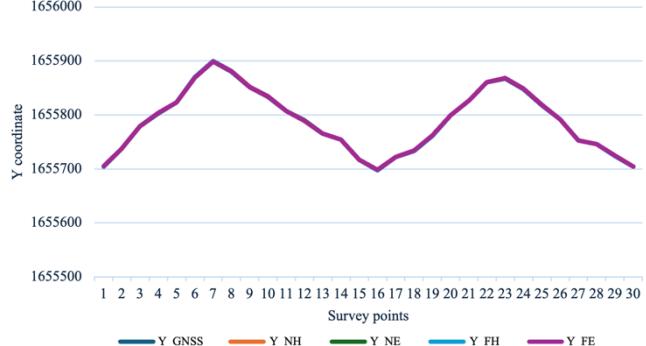


Fig. 10. Comparison of Y coordinates between RTK GNSS survey data and UAV-derived data across the four flight plans.

The NE flight plan demonstrated the highest vertical accuracy (RMSE_Z) assessment. The graph for each survey point shows values close to those obtained from the RTK GNSS survey (See Fig. 11), with an RMSE of 0.15. In comparison, the NH, FH, and FE flight plans had RMSE values of 1.98, 3.11, and 5.10, respectively.

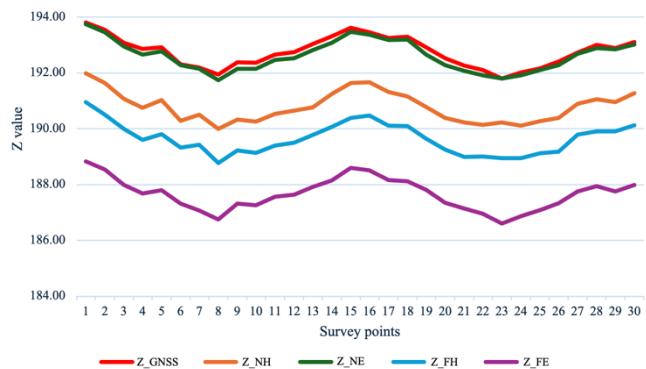


Fig. 11. Comparison of Z value between RTK GNSS survey data and UAV-derived data across the four flight plans.

The overall accuracy assessment (RMSE_T) results for horizontal and vertical accuracy indicate that the NE flight plan has the highest accuracy. This plan utilizes the Normal Flight Speed mode and Equal Distance Trigger mode, achieving an RMSE of 0.38.

IV. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

This study successfully achieved its objectives of creating an orthophotograph and a DSM using a UAV (DJI Phantom 4 multispectral RTK with the D-RTK 2 mobile base station) with a multi-flight plan approach while assessing the accuracy of these outputs against data obtained from RTK surveys (V90 Plus GNSS RTK system). The findings demonstrate the effectiveness of UAV technology in capturing high-resolution geospatial data across a specified area of 4,000 m² with 30 survey points.

The methodology included UAV surveys, RTK GNSS surveys, and a rigorous accuracy assessment. The different flight plans (NH, NE, FH, and FE) provided valuable insights into the impacts of flight speed and image capture modes on data quality. The results revealed that the NE flight plan, which utilized normal flight speed and equal distance triggering, had the highest overall accuracy ($\text{RMSE}_T = 0.38$ m) for horizontal and vertical measurements.

Regarding horizontal accuracy, the FE flight plan performed best for X and Y coordinates ($\text{RMSE}_R = 0.27$ m), indicating that faster flight speeds can enhance data collection efficiency without significantly compromising accuracy. In another way, when using UAVs for surveying to obtain horizontal coordinates, the FE flight plan can be selected automatically since it requires the shortest flight time of only 4 minutes and 13 seconds, thereby saving both time and battery life. Conversely, the NE flight plan excelled in vertical accuracy, making it suitable for applications requiring precise elevation measurements ($\text{RMSE}_Z = 0.15$ m).

Overall, this research contributes to the growing knowledge of UAV applications in geospatial surveying and highlights the importance of selecting appropriate flight plans based on specific project requirements. The findings of this study could significantly influence the standardization of UAV survey methods in several ways. The results underscore the potential of UAVs as reliable tools for generating accurate geospatial information, meeting the demands of various stakeholders in diverse fields such as urban planning, agriculture, and environmental monitoring.

B. Recommendations

1) Reference coordinates and elevation from UAVs: Applications using UAV data, including aerial imagery and digital elevation models, should carefully consider reference coordinates and elevation values, especially the studies that compare these values with other data sources.

2) Reference coordinates and elevation from RTK GNSS Network: For operations utilizing the RTK GNSS Network and Virtual Reference Stations (VRS), it should be crucial to consider reference coordinates and elevation for accurate comparisons.

3) Incorporation of Ground Control Points: This research highlights the need to include Ground Control Points (GCPs) to better calibrate datasets with different foundational references.

4) Flight Characteristics: The study focused solely on changes in flight speed and image recording modes while keeping other settings constant. Future studies should explore adjustments to other flight plans, such as flight height, overlap ratios, or software processing parameters, to potentially different results.

5) Design of Ground Survey Markers: Ground survey markers should be sized to match the specified spatial resolution, ensuring they exceed the Ground Sample Distance (GSD) of 5 centimeters for visibility in aerial imagery.

REFERENCES

- [1] T. Kanpleemjit, C. Nimmad, and S. Sangkthong. "Assessment of the accuracy of numerical terrain models from 3D point classification techniques in UAV photogrammetry". In Proceedings of the 26th National Civil Engineering Conference (pp. SGI-21-1 - SGI-21-9). June 23-25, 2021, Online Conference.
- [2] M. H. Chaudhry, A. Ahmad, and Q. Gulzar. "A comparative study of modern UAV platforms for topographic mapping." IOP Conference Series: Earth and Environmental Science, vol. 540, no. 1, p. 012019. doi: 10.1088/1755-1315/540/1/012019, 2020.
- [3] Department of Land. "Manual for Aerial Photography Mapping for UAV Surveying". National Land Parcel Center, Department of Land, Bangkok, 2020
- [4] R. Tangphumjit, S. Songsaeng, A. Pasda, K. Wongyao, and T. Lapitichayangkool. "Study of the control point patterns affecting the accuracy of UAV-based mapping." In Proceedings of the 25th National Civil Engineering Conference (Chonburi). July 15-17, 2020.
- [5] P. Suknaka, N. Phetthong, A. Pasda, K. Wongyao, and T. Lapitichayangkool. "Study of flight altitude and speed during imaging that affects accuracy in UAV mapping". In Proceedings of the 25th National Civil Engineering Conference (Chonburi). July 15-17, 2020.
- [6] American Society for Photogrammetry and Remote Sensing (ASPRS). "Positional Accuracy Standards for Digital Geospatial Data (Edition 1, Version 1.0)". Photogrammetric Engineering & Remote Sensing, 81(3), A1–A26. 2014.
- [7] C. Waisurasong. "Surveying with Photogrammetry". Sustainable Infrastructure Research and Development Center, Department of Civil Engineering, Khon Kaen University, 2020
- [8] Department of Public Works and Town & Country Planning. "Surveying and Mapping with Unmanned Aerial Vehicles". Urban Engineering Department and Institute for Urban Planning Personnel Development, Department of Public Works and Town & Country Planning, Bangkok, 2020
- [9] B. Paudel. "Drone Image Processing Using Pix4D". March 2023. [Online]. Available: <https://www.researchgate.net/publication/369362467>
- [10] A. Hakhrisuek. "A comparison of positional accuracy assessment of orthophotos obtained from real-time kinematic and static satellite-based surveys". Master's thesis in Geoinformatics, Faculty of Social Sciences, Srinakharinwirot University, 2018
- [11] Department of Land. "Operational Manual for Real-Time Kinematic (RTK) Network Surveying". Mapping Technology Division, Land Department, Bangkok, 2015
- [12] Federal Geographic Data Committee Secretariat. "Geospatial Positioning Accuracy Standards: Part 4 - Standards for Architecture, Engineering, Construction (A/E/C) and Facility Management". U.S. Geological Survey, 2002.
- [13] K. Chuchip. "Accuracy Assessment". Remote Sensing Technical Note NO.3. Faculty of Forestry, Kasetsart University, 2018.

Service area of Elderly care facility in the Perimeter Region of Thailand, using GIS

Kulapramote Prathumchai

Department of Geography,

Faculty of Social Science

Kasetsart University

Bangkok, Thailand

ORCID 0000-0002-5197-7298

Chitsanu Anurakjaturong

Department of Geography,

Faculty of Social Science

Kasetsart University

Bangkok, Thailand

csn.chitsanu.a@gmail.com

Nitis Boonanun

Department of Geography,

Faculty of Social Science

Kasetsart University

Bangkok, Thailand

boonanan13@gmail.com

Kitthiya Pooriwitayawattana

Department of Geography,

Faculty of Social Science

Bangkok, Thailand

kitthiya2520@gmail.com

Supattra Tookhokkruad

and Sopita Ountuang

Department of Geography

Faculty of Social Science

Bangkok, Thailand

phaknham2559@gmail.com

sopita.ou21@gmail.com

Wilawan Khrutsuwan

Department of Land

Ministry of Interior

Bangkok, Thailand

wilawan.khrut@gmail.com

Abstract—Thailand is experiencing a significant demographic shift, transitioning into an aging society. The elderly population in Nonthaburi, Pathum Thani, and Samut Prakan provinces is growing rapidly. This demographic trend and urban expansion raises critical concerns about the adequacy and accessibility of elderly care facilities in these regions. In alignment with Sustainable Development Goal (SDG) 3, which aims to ensure healthy lives and promote well-being for all ages, this research explores the service areas of facilities for the elderly in these provinces. This study applies service area and network analysis to examine the spatial distribution and service areas of existing elderly care facilities. By analyzing the coverage and accessibility of these facilities, the research aims to identify potential gaps and assess whether the current infrastructure can meet the growing demand. The findings will provide valuable insights for policymakers and urban planners, helping them to make informed decisions about the future development and allocation of elderly care resources. This research contributes to the broader understanding of effectively managing and planning for an aging population in rapidly urbanizing regions.

Keywords— *Elderly care facility, Network analysis, Service area*

I. INTRODUCTION

Today's technology plays a significant role in human life, particularly medical technology, which helps people live longer. This, in turn, has placed greater importance on education. As people gain knowledge, they recognize the importance of building a complete family and preparing physically and financially, and government policies may indirectly contribute to lower birth rates. These factors have resulted in a declining birth rate, while technological advancements have also contributed to lower death rates. Thailand has entered an aging society, which, by definition, is a society where more than 10% of the population is aged 60 and above, or more than 7% is aged 65 and above. Thailand has been an aging society since 2005 when 10.4% of the population was 60 and above. It is also projected that in the next 20 years, Thailand will fully transition into a complete aging society [1].

In 2013, statistics from Thailand's leading real estate website [2] revealed that the five provinces where Thai people searched for homes the most during the first quarter were Bangkok, Chiang Mai, Nonthaburi, Pathum Thani, and Samut

Prakan. This indicates that more people are beginning to look for housing in suburban areas, though major cities like Bangkok and Chiang Mai remain prominent. The growing population has led to concentrated urban areas, causing resource shortages and environmental impacts, such as increased pollution. There are also broader social and economic effects, such as continuously rising urban housing prices. As a result, urban residents, especially those starting new families, are increasingly relocating to suburban areas [3]. A point of interest is how many elderly care facilities exist in the suburban areas surrounding major cities like Bangkok.

A survey of the elderly population in suburban areas, including Nakhon Pathom, Nonthaburi, Pathum Thani, Samut Prakan, and Samut Sakhon, identified the three provinces with the largest elderly populations. These are: 1) Nonthaburi, with a total population of 1,295,916, of which 242,640 are elderly (18.72% of the total population); 2) Samut Prakan, with a total population of 1,360,227, of which 215,640 are elderly (15.85% of the total population); and 3) Pathum Thani, with a total population of 1,201,532, of which 179,753 are elderly (14.96% of the total population) [4]. Given the projected increase in the elderly population in the future, this study focuses on analyzing the service areas of elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan provinces. The research aims to provide valuable information and guidance to relevant authorities for making informed decisions about future planning and addressing upcoming demographic changes.

II. OBJECTIVE

This study aims to analyze the service areas of elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan provinces. The research seeks to answer two main questions: Do the elderly care facilities provide adequate coverage within the service radius, and is the capacity of these facilities aligned with the demand from the elderly population in the service areas in a present ?

The research scope is defined in four aspects: geographically, it covers Nonthaburi, Pathum Thani, and Samut Prakan provinces, with Nonthaburi divided into six districts, Pathum Thani into seven districts, and Samut Prakan into six districts. The content scope focuses on analyzing the service areas of elderly care facilities in these provinces,

considering facility locations, capacity, and service radius based on distance and village coordinates. This data was collected on elderly care facilities in September 2023, with direct field surveys conducted in November 2023. The target population includes licensed elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan.

III. LITERATURE REVIEW

A. Network analysis

Network analysis encompasses basic tasks like determining the shortest path and connectivity. More advanced analyses are commonly performed on network data by transport, electricity, gas, and telecommunications industries. These complex tasks include simulating flows in intricate networks, managing load distribution in electrical grids, analyzing traffic flow, and calculating pressure drops in gas pipelines. Additionally, methods are used to assess the presence of a generalized spatial surface. [5]

B. Service Area

A service area refers to managing a service location, which involves the operation and execution of tasks by relevant agencies or organizations related to managing the service area. It includes managing the service, location, equipment, and facilities provided and organizing various activities in that area. Critical components of service area management include the scope of the service area, management of the space, and the equipment and facilities used within that area [6] [7]. In this research, a service area refers to the area that facilitates users who visit elderly care facilities in the study area. Key components include the proximity (distance), and the capacity of the facilities.

C. Elderly People

Elderly people are defined as those 60 years of age or older. They are entitled to protection, promotion, and support in various areas, such as healthcare, education, employment, and personal development [8]. In this research, the term "elderly" refers to Thai citizens aged 60 and above, who were considered the population in the study of service areas of elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan provinces.

D. Applications GIS for elderly health care facilities

In the study "Healthcare Logistics System Planning for Facility Location of Aging Society," Geographic Information Systems (GIS) were utilized in several significant ways to enhance the planning and management of healthcare services for the aging population in Bangkok. The research utilized GIS to analyze service coverage for elderly care facilities, revealing that existing primary healthcare centers in Bangkok are insufficient. The Dijkstra Algorithm was employed to assess accessibility based on walking times for the elderly [9].

IV. METHODOLOGY

This research is based on a conceptual framework that illustrates the service area of elderly care facilities in three regions: Nonthaburi, Pathum Thani, and Samut Prakan

provinces. The data analysis is divided into two main aspects. The first aspect is determining the service radius of elderly care facilities, which involves analyzing data such as the list of elderly care facilities, road networks, distances, facilities, and village locations. The second aspect focuses on the capacity of elderly care facilities based on the collected data, which includes the capacity of the facilities, the demand for elderly care, and the population in the study areas. This forms the conceptual framework for the research. The overall research framework is illustrated in Fig.1.

A. Data and Sources of Data

This research is descriptive and involves surveying the locations of elderly care facilities and analyzing their service areas. The methodology includes two types of data: primary data, which consists of geographic coordinates and the capacities of elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan provinces, and secondary data, which includes the list of elderly care facilities, village coordinates, elderly population data, road network data, provincial boundaries, and demand for elderly care facilities. The data sources include the Department of Health Service Support (list of elderly care facilities), the Department of Provincial Administration (elderly population and provincial boundaries), the Office of the Permanent Secretary, the Ministry of Transport (road network), the Office of Transport and Traffic Policy and Planning (distance information), internet searches from the Geo-Informatics and Space Technology Development Agency (GISTDA) website (village coordinates), and the Real Estate Information Center website (demand for elderly care facilities).

B. Analysis Methods

This study utilized network analysis in GIS to evaluate the accessibility and coverage of elderly care facilities within designated service areas. By applying the service area function, the study achieved a comprehensive view of how well these facilities meet the needs of the elderly population, highlighting both strengths and areas for improvement in accessibility and capacity. Unlike straight-line distances, network analysis takes actual travel routes into account, such as roads and pathways, providing a more accurate representation of service areas by considering the real-world paths people would use to access elderly care facilities. Additionally, network analysis allows for the calculation of travel distances and times from various locations to these facilities, which is essential for understanding their accessibility, especially in terms of practical travel time within specific distance thresholds (e.g., 5, 10, or 15 kilometers). Identification of coverage gaps: By mapping service areas, network analysis identifies regions that fall outside the coverage of existing elderly care facilities. These uncovered areas can point to locations where additional facilities may be needed or where existing infrastructure could be enhanced to better meet demand.

The analysis included assessing the road network, geographic coordinates of elderly care facilities, and village coordinates to determine service area boundaries. Network analysis with the service area technique was performed in GIS

software to evaluate whether the service areas of elderly care facilities provide adequate coverage for the elderly population in the study region.

The service area mapping process involved setting parameters for driving distances from facilities at 5, 10, 15, 20, 25, and 30 kilometers. Overlay analysis was then applied to identify villages both within and outside the coverage areas, allowing for a clear assessment of service adequacy across the region.

Data on the capacity of elderly care facilities were collected through field surveys at each location. This capacity data was compared to the demand from the elderly population using information provided by the Real Estate Information Center [10].

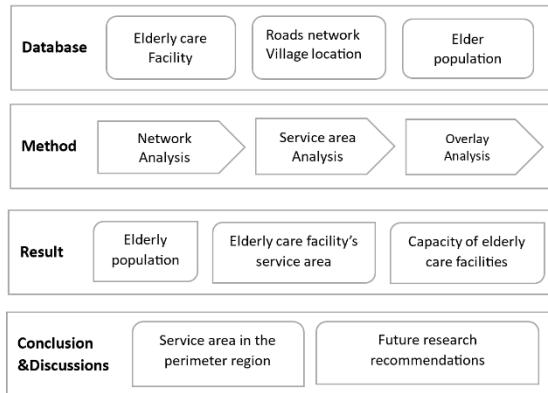


Fig.1 Research Framework

V. RESEARCH RESULTS

The study on the service areas of elderly care facilities in Nonthaburi, Pathum Thani, and Samut Prakan provinces analyzes three aspects: the elderly population density, the service area boundaries of elderly care facilities (focusing within 10 kilometers of service area coverage), and the capacity of elderly care facilities with the demand from the elderly population within the service areas. The findings are as follows:

A. Elderly Population

The survey of the elderly population in Nonthaburi, Pathum Thani, and Samut Prakan provinces revealed varying population densities. In Nonthaburi Province, the total elderly population is 118,770, with the highest density in Bang Bua Thong District, housing over 43,000 elderly residents, while Sai Noi, Pak Kret, and Mueang Nonthaburi districts have between 10,000 and 16,000 elderly residents. In Pathum Thani Province, the elderly population totals 126,186. The densest areas are the Mueang and Khlong Luang districts, each with over 28,800 elderly residents, and the least dense areas are Sam Khok, Lat Lum Kaeo, and Nong Suea districts, with 7,800 to 10,000 elderly residents. In Samut Prakan Province, the elderly population is 118,712, with the highest densities in Bang Phli and Mueang districts, each exceeding 35,000 elderly residents, while Bang Sao Thong District has fewer than 7,000 elderly residents.

B. Service Area of Elderly Care Facilities

1) Nonthaburi Province

The analysis of the service areas of elderly care facilities in Nonthaburi Province was conducted using the geographic coordinates of the facilities, with data requested from the Department of Health Service Support, the Ministry of Public Health, and additional data collected through direct field surveys. The analysis revealed 74 elderly care facilities in the province, 416 village coordinates, and road network data. The service area was analyzed using network analysis with the service area technique under a 10-kilometer service area. The results showed that the 74 elderly care facilities provide services to 348 out of the 416 villages. Two districts, Pak Kret, and Sai Noi, remain partially uncovered. Pak Kret District covers 104 villages, while Sai Noi District covers 46 villages. This indicates that the 74 elderly care facilities are insufficient to fully cover the service area across the entire province (Fig. 2).

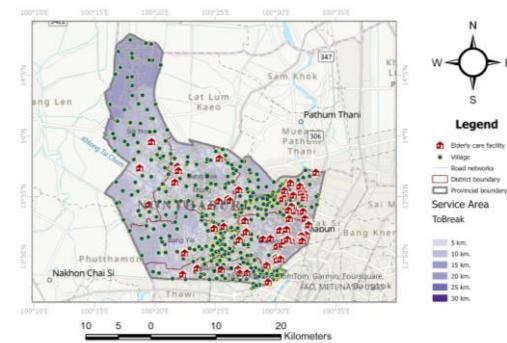


Fig.2. Service area of elderly care facilities in Nonthaburi province

2) Pathum Thani Province

The analysis of the service areas of elderly care facilities in Pathum Thani Province was conducted using the geographic coordinates of the elderly care facilities, with data requested from the Department of Health Service Support, the Ministry of Public Health, and additional data collected through direct field surveys. The analysis revealed 28 elderly care facilities in the province, 708 village coordinates, and road network data. The service area was analyzed using network analysis with the service area technique under a 10-kilometer radius. The results showed that the 28 elderly care facilities provide services to 536 out of 708 villages, with the highest coverage in Mueang District, covering 104 villages. However, Nong Suea District is completely uncovered. This indicates that the 28 existing elderly care facilities are insufficient to fully cover the service area across the entire province. (Fig. 3)

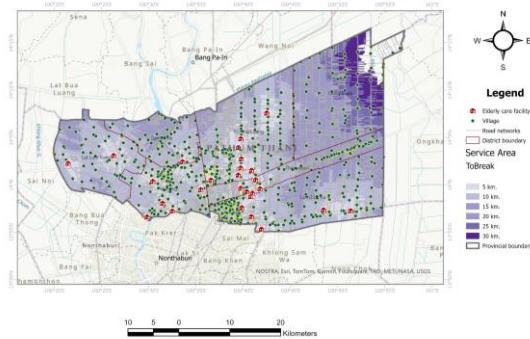


Fig.3. Service area of elderly care facilities in Pathum Thani province

3) Samut Prakan Province

The analysis of the service areas of elderly care facilities in Samut Prakan Province was conducted using the geographic coordinates of the elderly care facilities, with data requested from the Department of Health Service Support, Ministry of Public Health, and additional data collected through direct field surveys. The analysis revealed 19 elderly care facilities in the province, 402 village coordinates and road network data. The service area was analyzed using network analysis with the service area technique under a 10 kilometer coverage. The results showed that the 19 elderly care facilities provide services to 219 out of 402 villages. Two districts, Phra Pradaeng, and Phra Samut Chedi, remain partially uncovered. Phra Pradaeng District covers only one village, while Phra Samut Chedi District has no coverage. This indicates that the 19 elderly care facilities are insufficient to fully cover the service area in the study area. (Fig.4)

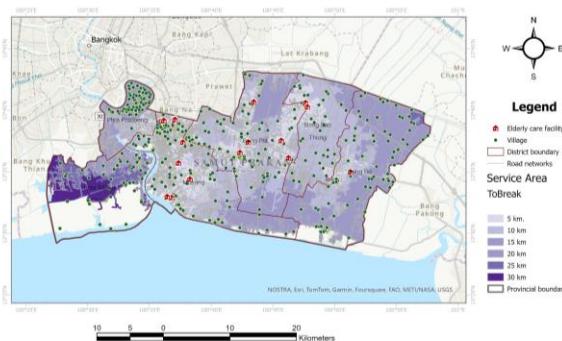


Fig.4. Service area of elderly care facilities in Samut Prakan Pathum Thani province

C. Capacity of Elderly Care Facilities concerning the Demand from the Elderly Population in the Service Area

Analysis of elderly care facility capacity in the studied provinces reveals a significant shortfall in meeting the aging population's needs. In Nonthaburi, Pathum Thani, and Samut Prakan, the estimated demand for elderly care far exceeds the current capacity of available facilities. This shortage is particularly acute in most districts within these provinces,

with only Pak Kret District in Nonthaburi demonstrating sufficient capacity to meet local needs.

Specifically, in Nonthaburi, while 113,668 elderly individuals reside within the service areas of existing facilities, the current capacity of 1,538 beds falls short of the estimated demand of 5,683 individuals. In Pathum Thani, the gap is even more pronounced, with a total facility capacity of 318 beds compared to an estimated demand of 5,301. Samut Prakan also faces a similar challenge, with a capacity of 545 beds and an estimated demand of 5,936.

These findings could be significant for future plans to increase the elderly care infrastructure in these regions to ensure adequate support for the growing elderly population.

VI. CONCLUSIONS AND DISCUSSIONS

A. Service area of Elderly Care Facilities in the perimeter region

1) Nonthaburi Province

Analysis of service areas in Nonthaburi revealed that while 74 elderly care facilities cover 95.70% of the province's elderly population, there remains a significant coverage gap, particularly in Pak Kret and Sai Noi districts. This is due to geographical constraints in Pak Kret and a lack of healthcare infrastructure in Sai Noi. Despite 5% of the elderly population within the service area requiring care, the total capacity of facilities falls far short of the estimated demand, underscoring a critical shortage in all districts.

2) Pathum Thani Province

Pathum Thani's elderly care facilities cover 84.03% of the province, with Mueang District demonstrating the highest coverage due to accessibility and infrastructure. However, Nong Suea District remains completely uncovered due to lacking facilities. The estimated demand for care far exceeds the total capacity of facilities, highlighting a significant shortage across all districts.

3) Samut Prakan Province

In Samut Prakan, 19 facilities cover 67.35% of the elderly population. While Mueang Samut Prakan has high coverage, Phra Samut Chedi, a coastal district with dispersed communities and limited healthcare, remains entirely uncovered. Similar to the other provinces, the estimated demand for care significantly outpaces the available capacity, resulting in a province-wide shortage.

The analysis reveals a consistent pattern of inadequate elderly care facility coverage and capacity across all three provinces. This significance could imply the need to expand and improve infrastructure to meet the growing needs of the aging population.

In comparison to previous studies, such as the research on healthcare logistics for the elderly in Bangkok, aligns closely in both methodology and findings. The Bangkok study utilized Geographic Information Systems (GIS) to assess healthcare accessibility for the elderly and applied network analysis to optimize facility locations based on

accessibility[9]. Both studies highlight the need for improved accessibility and reveal significant gaps in current healthcare coverage, underscoring a shared conclusion: strategically expanding and placing elderly care facilities is essential to meet the growing demand. The use of GIS and network analysis in both studies emphasizes the effectiveness of these methods for spatial analysis and planning, showcasing the advantages of integrating real travel paths and population forecasts into service area planning.

This research effectively addresses its main objectives by assessing whether elderly care facilities provide sufficient coverage within their service radius and determining if facility capacity aligns with the demand from the elderly population. The findings indicate that, although service areas cover substantial portions of the elderly population, facility capacity is inadequate across all districts, falling short of meeting the needs of an aging demographic. This highlights the pressing need for infrastructure enhancements and aligns with the research's goal to provide actionable insights for bridging these gaps.

This study was conducted within a limited timeframe for data collection, with a focus on private healthcare facilities and a straightforward methodology. While these results offer a valuable preliminary assessment of elderly healthcare facility coverage, future studies could benefit from including existing Primary Healthcare Centers (PHCs) and projected data on the elderly population. This addition would provide a more comprehensive view and further enhance the accuracy of planning for elderly care infrastructure.

B. Future Research Recommendations

This study relies on data from the Department of Health Service Support and field surveys conducted over a limited period (September and November 2023). Consequently, the findings may not capture seasonal variations in service demand or shifts in the elderly population over time, which could affect the study's applicability and generalizability regarding facility capacity and coverage. Additionally, while this study evaluates whether elderly care facilities meet the needs of the elderly population, certain complex demand factors such as socio-economic status, health conditions, and individual preferences were not fully addressed, potentially leading to an oversimplified analysis.

For future research, we recommend longitudinal studies to monitor changes in the elderly population and their evolving care needs over time. Such an approach would enable researchers to understand the effects of demographic shifts, urbanization, and policy changes on demand for elderly care facilities, thereby supporting more informed planning and resource allocation. Future studies should also conduct a comprehensive analysis of factors influencing demand for elderly care, including socio-economic conditions, health status, and specific preferences of the elderly. This more detailed approach would improve service alignment with the diverse needs of elderly groups, enhancing service delivery and better addressing the unique needs of Thailand's aging population.

The findings of this research were generated using GIS software, which estimates travel time based on preset distances and speeds. In reality, however, multiple factors affect travel, including traffic conditions and available transportation options. Integrating these additional factors into future analyses would increase the accuracy of results.

It is also important to recognize that not all elderly individuals in Nonthaburi, Pathum Thani, and Samut Prakan will utilize elderly care facilities, as these are just one of several options available to them. Additionally, elderly residents from neighboring or other provinces may also seek services in these areas, influenced by factors such as available amenities, environmental preferences, and personal choice. To improve accuracy and reliability, future research should consider including the elderly population from other provinces in the analysis.

REFERENCES

- [1] C. Promphak. 2013. Thailand's Transition into an Aging Society (Online). https://library.senate.go.th/document/Ext6078/6078440_0002.PDF, September 23, 2023.
- [2] DDproperty Editorial Team. 2020. The 5 Most Searched Provinces for Homes in Q1 2020 (Online). <https://www.ddproperty.com/คุณอยู่ที่ไหน/5-จังหวัดที่คนไทยหานักบ้านมากที่สุด-24696>, September 25, 2023.
- [3] P. Vutichatiwanich. 2015. The Impact of Global Urban Sprawl (Online). <https://www.ftpi.or.th/2015/2791>, September 24, 2023.
- [4] Department of Provincial Administration, Ministry of Interior. 2022. Population Statistics (Online). https://stat.bora.dopa.go.th/new_stat/webPage/statByProvince, September 20, 2023.
- [5] M. Kannan and M. Singh, 2021. Geographical Information System and Crime Mapping, CRC Press, Taylor&Francis Group, p.13
- [6] T. Kulthawatchawichai. 2015. Service Areas, Perceived Ease of Travel, and Perceived Benefits Affecting the Decision to Visit Trade Shows. Independent Study, MBA, Bangkok University.
- [7] W. Puyawattana. 2013. Managing Bank Branch Spaces in Shopping Centers in Bangkok's Central Business District: Case Study of Kasikornbank. Independent Study, MSc, Thammasat University.
- [8] Department of Older Persons. 2021. Rights and Welfare for the Elderly (Online). <https://www.dop.go.th/th/know/15/646>, September 13, 2023.
- [9] C. Amchang and J. Suraraksa, (2023). Healthcare Logistics System Planning for Facility Location of Aging Society.*International Journal of Membrane Science and Technology*. <https://doi.org/10.15379/ijmst.v10i2.3119>
- [10] Real Estate Information Center (online) <https://reic.or.th/>, September 13, 2023.

Enhancing Thailand's Cultural Diplomacy: A GIS-Based Analysis of Global Sentiments Toward Thai Food and Tourism

Kednipa Prasongsri

*School of Information, Computer
and Communication Technology*

*Sirindhorn international institute of technology
Thammasat university
Pathum Thani, Thailand
6322774439@g.siit.ac.th*

Teerayut Horanont

*School of Information, Computer
and Communication Technology*

*Sirindhorn international institute of technology
Thammasat university
Pathum Thani, Thailand
teerayut@siit.tu.ac.th*

Abstract—This study leverages social media data from the X platform, analyzing over 30,000 posts and discussions about Thai cuisine, tourism and festivals across three key countries. By utilizing Geographic Information Systems (GIS) to map regional sentiment patterns, the research provides a comprehensive analysis of global attitudes toward these aspects of Thai culture. The study incorporates the concept of soft power, emphasizing how Thailand's cuisine, tourism and festivals can serve as cultural assets to influence international perceptions and relationships. By evaluating the effectiveness of Thailand's cultural diplomacy and tourism initiatives, this research aims to advance Thailand's cultural influence on the global stage. The findings offer valuable insights to enhance Thailand's soft power, boost its global appeal, and strengthen its role in international cultural relations.

Keywords—Cultural diplomacy, Geographic Information Systems (GIS), Sentiment analysis, Thai food and tourism, Soft power strategy

I. INTRODUCTION

Soft power is defined as the ability to shape the preferences and behaviors of others through attraction and appeal rather than coercion or monetary incentives. This influence stems from a nation's cultural assets, political values, and foreign policies, which, when perceived as legitimate or admirable, inspire others to align with its goals willingly [1], [2]. It plays a significant role in shaping a country's economic and international relationships by leveraging long-term strategies such as encouraging tourism and short-term tactics like promoting its cuisine or products [3]. Thailand's international reputation is profoundly influenced by its gastronomy, tourism, and cultural events, all of which are integral components of its soft power. While these sectors contribute significantly to positive global perceptions, the effectiveness of Thailand's cultural and tourism campaigns may vary across regions due to differing socio-cultural contexts and market dynamics [4]. This study employs Geographic Information Systems (GIS) and sentiment analysis to systematically analyze global perceptions of Thai cuisine, tourism, and cultural initiatives. By mapping

sentiment trends and engagement metrics within key markets such as Japan, India, and the United States, the research aims to critically assess the impact and effectiveness of Thailand's cultural diplomacy strategies. The findings will provide data-driven insights to refine strategic marketing and promotional efforts, ultimately strengthening Thailand's cultural diplomacy and enhancing its global influence. This approach underscores the importance of leveraging digital tools and analytics to inform soft power strategies in an increasingly interconnected world. This paper is organized as follows: Section II outlines the methodology, including data processing and GIS visualization. Section III presents results with regional insights. Section IV discusses findings and recommendations, while Section V concludes with implications and future directions for enhancing Thailand's soft power.

II. METHODS

A. Data Collection

This study collected data from X (formerly known as Twitter), focusing on three key countries: Japan, India, and the United States. The data collection process involved filtering tweets related to Thai cuisine, tourism, and cultural content. Selection criteria included language relevance, geographical location, and content type. Engagement metrics such as likes and reach were also captured to evaluate the visibility and interaction levels of the posts. Additionally, user attributes and geo-location information were gathered to provide a detailed understanding of the spatial distribution and demographic characteristics of the contributors.

B. Data Cleaning

The collected dataset underwent a comprehensive cleaning process to ensure its suitability for sentiment analysis using a pre-trained model. Since this study focused on English-language comments, all non-English content was excluded to maintain linguistic consistency. Initially, the dataset was organized and preprocessed in Excel, where extraneous elements

such as special characters, URLs, and irrelevant hashtags were removed. Common stopwords were filtered out to improve input data quality. Subsequently, the text was tokenized and lemmatized to standardize words into their base forms, ensuring consistency for analysis. Records with incomplete or irrelevant information were excluded to preserve the dataset's integrity. Finally, the cleaned dataset was structured and formatted for sentiment analysis, incorporating sentiment labels and engagement metrics for use with the pre-trained model.

C. Sentiment Analysis

The sentiment analysis for this study was conducted using TextBlob, a pre-trained natural language processing (NLP) tool widely recognized for its efficiency and simplicity in processing textual data. The primary goal of this analysis was to compute the polarity score for each post and classify sentiments into three categories: positive, negative, and neutral. These categories provide a comprehensive understanding of public perceptions reflected in the dataset. Each post underwent individual processing to extract its polarity score, which quantifies the emotional tone of the text on a scale ranging from -1 to +1. Specifically:

$$\text{Polarity} = \frac{\sum_{i=1}^n p_i}{n}$$

Fig. 1. Polarity Equation

- A score of -1 represents a highly negative sentiment, indicating strong unfavorable emotions.
- A score of +1 indicates a highly positive sentiment, reflecting strong favorable emotions.
- A score of 0 reflects a neutral sentiment, suggesting an absence of strong emotional tone.

TextBlob computes polarity scores by analyzing the syntactic structure and lexical features of the text, such as word choice and contextual relationships. Based on the computed polarity scores, posts were classified into three distinct sentiment categories: Positive Sentiment: Posts with polarity scores greater than 0, indicating a favorable or optimistic tone. Negative Sentiment: Posts with polarity scores less than 0, suggesting an unfavorable or pessimistic tone. Neutral Sentiment: Posts with a polarity score of 0, reflecting the absence of strong positive or negative emotions. The analysis relies on TextBlob's Pattern Analyzer, which uses a predefined lexicon of polarity and subjectivity scores for individual words. These scores include: Polarity: Ranges from -1.0 (very negative) to +1.0 (very positive). Subjectivity: Ranges from 0.0 (completely objective) to 1.0 (completely subjective).

For example, the word "happy" has a polarity of +0.8 and a subjectivity of 0.9, indicating a positive and highly subjective expression. In contrast, the word "sad" has a polarity of -0.8 and a subjectivity of 0.7, reflecting a negative sentiment

with moderate subjectivity. TextBlob calculates the overall polarity and subjectivity of a post by averaging the polarity and subjectivity scores of all recognized words within the text: Polarity: The total polarity scores are summed and divided by the number of words with sentiment scores. Subjectivity: Similarly, subjectivity scores are averaged to determine the overall subjectivity of the text. These calculations yield two key metrics: Polarity Score: Indicates the direction and intensity of sentiment, with negative values denoting negative sentiment, positive values denoting positive sentiment, and values near zero indicating neutrality. Subjectivity Score: Reflects the degree of subjectivity, with higher values indicating a stronger subjective tone in the text.

To complement the sentiment analysis, additional metadata for each post was collected and incorporated into the results. This metadata includes: Sentiment Score: A numerical value indicating the intensity and direction of sentiment. Sentiment Label: The categorical sentiment classification (Positive, Negative, Neutral) derived from the polarity score. Reach: The estimated number of users exposed to the post, providing insight into the post's potential influence. Geographical Information: Data such as longitude, latitude, and country, which enables the spatial analysis of sentiment distribution across regions.

The inclusion of geographical metadata enhances the study by facilitating the identification of regional sentiment trends and patterns. For instance, positive sentiment may dominate in certain regions due to cultural affinities, while neutral or negative sentiment could prevail elsewhere. This detailed analysis provides actionable insights into the spatial dynamics of public sentiment. Overall, this comprehensive approach to sentiment analysis, supported by metadata and spatial information, ensures a robust framework for understanding public perceptions and trends. By leveraging these insights, the study provides a valuable basis for informing strategies in cultural and tourism campaigns. This methodology aligns with the study's objective of analyzing global sentiments and identifying actionable patterns across different regions.

D. Visualization

The final phase of the study involved visualizing the analyzed data using Geographic Information Systems (GIS). This process mapped the sentiment data geographically to provide a visual representation of regional perceptions of Thai cuisine, tourism, and festivals. GIS visualization enabled the identification of spatial patterns and variations, offering a clear and insightful depiction of how different regions perceive Thailand's cultural and tourism offerings. Such visual representations are crucial for informing strategic decisions and enhancing Thailand's soft power by addressing specific regional perceptions and trends.

III. RESULTS

The sentiment analysis results focus on three key regions: Japan, the United States, and India. These regions were chosen to represent a diverse range of global perspectives on Thai cuisine, tourism, and cultural festivals. The findings reveal

notable differences in sentiment, indicating varying degrees of interest and engagement with Thailand's cultural offerings. The following sections will provide a detailed discussion of these results, emphasizing their implications for enhancing Thailand's cultural diplomacy and marketing efforts.

A. Japan

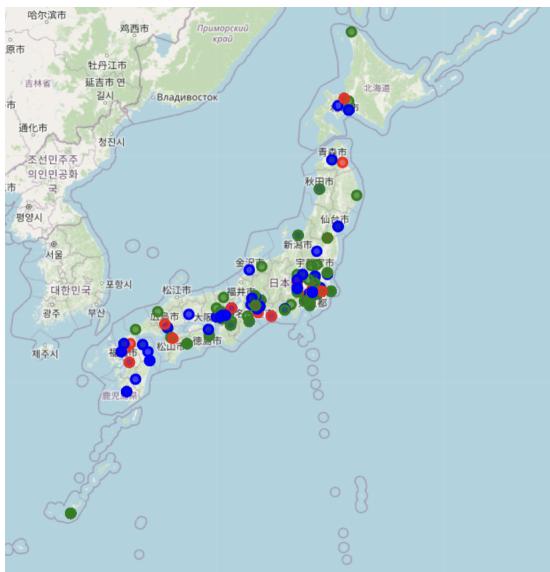


Fig. 2. The Sentiment Distribution of Japanese

The sentiment map of Japan highlights strong positive sentiment (green dots) in major urban areas like Tokyo, Osaka, and Fukuoka, reflecting a favorable reception of Thai culture, cuisine, and festivals in these regions. Neutral sentiment (blue dots) is dispersed across both urban and rural areas, indicating opportunities to engage audiences with more emotionally resonant campaigns. Negative sentiment (red dots) is minimal and scattered, suggesting rare unfavorable perceptions that may still warrant attention. Overall, the map underscores the success of Thai cultural efforts in metropolitan hubs while highlighting potential for further engagement in neutral sentiment areas.

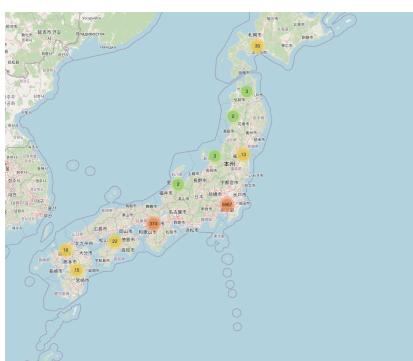


Fig. 3. The Post Distribution of Japanese

The map also shows a clear concentration of posts in major urban areas like Tokyo, Osaka, and Fukuoka, which could

be influenced by two main factors. One possibility is the presence of Thai communities in these cities, as areas like Shinjuku in Tokyo and Namba in Osaka are known hubs for Thai expatriates and businesses. Another factor could be the local interest in Thai culture, driven by the popularity of Thai restaurants, festivals, and cultural events that are often held in these metropolitan areas. These cities are also frequently targeted by Thai tourism campaigns, which may explain the high density of posts and the predominantly positive sentiment reflected on the map. In contrast, rural areas show much lower activity, indicating that Thai cultural influence is primarily concentrated in urban centers.

B. The United States



Fig. 4. The Sentiment Distribution of The United States of America

The sentiment distribution map of the United States reveals a strong positive reception of Thai culture, cuisine, and festivals, particularly in regions like the West Coast, South, and East Coast. Cities such as Los Angeles, San Francisco, Miami, and Austin display clusters of green points, signifying widespread admiration and appreciation. These urban centers, known for their cultural diversity and openness to international cuisine, are natural hubs for Thai cultural engagement. The presence of established Thai communities, restaurants, and cultural events in these areas likely contributes to the positive sentiment observed. Additionally, the concentration of Thai tourism campaigns and the popularity of Thai cuisine, particularly dishes like Pad Thai and Tom Yum, resonate strongly with local populations, further enhancing engagement.

On the East Coast, cities like New York, Washington D.C., and Boston exhibit a mix of positive and neutral sentiment. These areas are major metropolitan hubs with a significant interest in international cultures, and the prominence of Thai restaurants, cultural events, and tourism-related mentions underscores this trend. Neutral sentiment in these regions often reflects informational mentions, indicating potential for more emotionally engaging campaigns to shift perceptions toward positivity.

The Midwest, while showing pockets of positive sentiment in cities like Chicago, reveals a higher prevalence of neutral discussions. This suggests that while there is awareness of Thai culture, it may lack the strong emotional connection seen in other regions. The Midwest's demographic composition and less frequent exposure to international cultural campaigns may

account for this trend. Targeted outreach, such as Thai food festivals or tourism promotions, could help bridge this gap.

Negative sentiment, represented by red points, is minimal but present in areas like Sacramento and Saint Paul. These isolated instances may stem from specific issues, such as cultural misunderstandings or dissatisfaction with service quality. Addressing these localized concerns can help mitigate negative perceptions and reinforce Thailand's cultural image.

Geographically, the clustering of positive sentiment aligns closely with regions known for cultural openness and a high density of Thai cultural establishments. This suggests that the accessibility of Thai cuisine, events, and tourism-related activities is a significant driver of positive engagement. Conversely, areas with fewer Thai establishments or cultural campaigns, such as parts of the Midwest, tend to show more neutral sentiment, emphasizing the importance of strategic expansion to these regions.

The findings underscore the importance of leveraging regional characteristics to enhance Thailand's cultural influence. By focusing on urban centers with existing positive sentiment and addressing neutral or negative perceptions in other areas, Thailand can further strengthen its cultural and tourism footprint in the United States.



Fig. 5. The Post Distribution of The United States of America

The engagement with Thai culture, cuisine, and festivals in Washington D.C., New York, and Los Angeles reflects the unique roles these cities play in shaping cultural perceptions in the United States. New York, as a global metropolis and cultural hub, demonstrates the highest concentration of activity, driven by its diverse population and vibrant culinary scene. The city's extensive network of Thai restaurants, cultural events, and promotional campaigns significantly contributes to the widespread positive sentiment toward Thai culture. The multicultural nature of New York provides an ideal platform for fostering appreciation and dialogue around Thai festivals and tourism.

In Washington D.C., engagement with Thai culture is influenced by its position as the nation's capital and a center for international diplomacy. The presence of embassies, including the Royal Thai Embassy, along with global organizations, creates opportunities for cultural exchanges that resonate with both the local and international communities. Events promoting Thai cuisine and festivals in Washington are likely to attract attention from policymakers and diplomats, strengthening Thailand's cultural diplomacy efforts in this influential city.

Los Angeles, with its strong ties to the Asian diaspora and the presence of Thai Town, serves as a critical hub for Thai cultural promotion on the West Coast. The city's established reputation as an entertainment and media capital allows Thai festivals and cuisine to gain visibility across a wide audience. The high engagement levels in Los Angeles reflect the city's openness to Asian influences and its role as a gateway for introducing Thai culture to both local and global audiences.

C. India

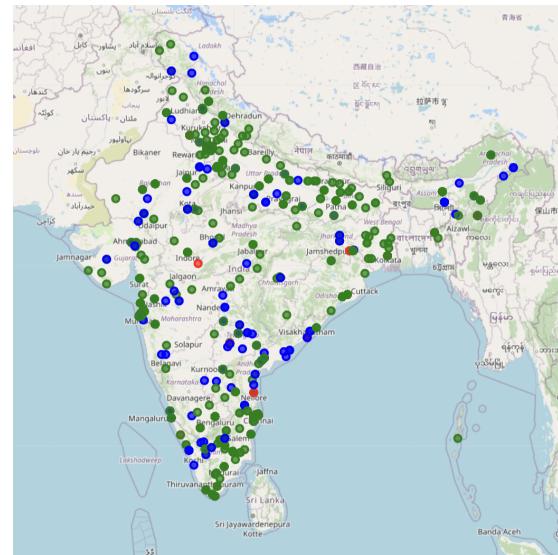


Fig. 6. The Sentiment Distribution of India

The map illustrates the distribution of sentiments across India, with green points (positive sentiment) heavily concentrated in urban areas like Delhi, Mumbai, Bengaluru, Chennai, and Hyderabad. This prevalence of positive sentiment underscores the successful reception of Thai cuisine and cultural initiatives in these cities. These regions are characterized by their multicultural demographics, exposure to global culinary trends, and frequent hosting of cultural festivals, all of which contribute to the favorable sentiment.

Blue points (neutral sentiment) are distributed more evenly across less urbanized areas, indicating informational mentions or moderate engagement. These areas represent an opportunity for targeted campaigns to create deeper emotional connections with Thai culture. Red points (negative sentiment) are minimal and scattered, appearing in cities like Indore and Kolkata. These isolated cases warrant further investigation to address potential cultural misunderstandings or issues with Thai-related experiences in these areas.

The first map illustrates the density of posts about Thai culture, cuisine, and tourism in India. Northern India, particularly Delhi, shows the highest activity due to its role in diplomacy and the presence of the Royal Thai Embassy. Western cities like Mumbai and Ahmedabad exhibit significant engagement, driven by Mumbai's cosmopolitan population and culinary diversity. Southern cities, including Bengaluru,

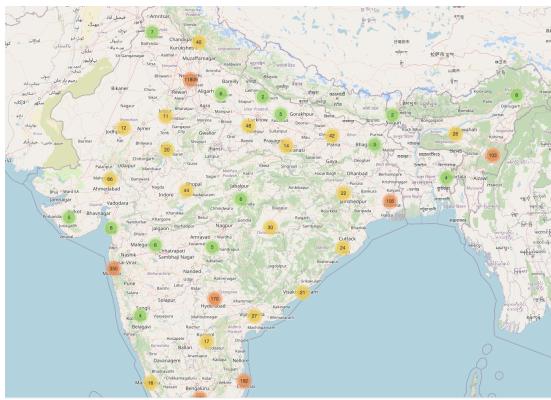


Fig. 7. The Post Distribution of India

Chennai, and Hyderabad, show moderate post densities, reflecting openness to global influences. Eastern regions, such as Kolkata and Guwahati, reveal smaller clusters, indicating growing awareness. Overall, major metropolitan areas dominate, while smaller cities present opportunities for further cultural outreach.

IV. RECOMMENDATIONS

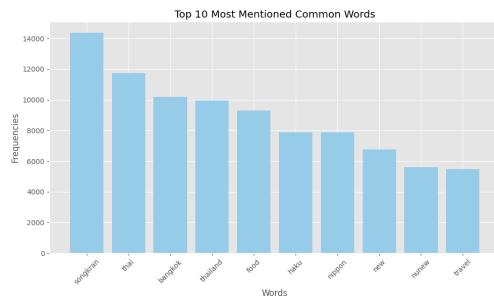


Fig. 8. Top 10 mentioned words

The chart illustrates the top 10 most mentioned words in social media discussions related to Thai culture, cuisine, and tourism. The term "Songkran" leads with the highest frequency, reflecting the global interest in Thailand's famous water festival and its prominence in cultural discourse. Following closely are terms like "Thai," "Bangkok," and "Thailand," indicating a strong focus on the nation's identity, capital city, and overall cultural appeal. Words such as "food" highlight the pivotal role of Thai cuisine in Thailand's cultural diplomacy, while mentions of "travel" underscore the global audience's interest in exploring the country.

- Japan Japan exhibits a strong cultural alignment with Thailand, particularly through shared interests in cuisine and historical tourism. The challenge of finding authentic Thai cuisine, despite its popularity in major cities like Tokyo and Osaka, presents an opportunity for the Thai Ministry of Commerce's "Thai Select" certification program. Promoting this certification can improve the authenticity and visibility of Thai restaurants, appealing

to Japanese diners seeking genuine culinary experiences. Additionally, leveraging platforms like TikTok to share detailed cooking demonstrations aligns with Japanese consumers' interest in culinary experimentation, fostering a deeper appreciation for Thai cuisine. In tourism, Japanese visitors display a preference for historical attractions, such as Wat Arun and the Grand Palace, and enjoy shopping for souvenirs. These trends suggest that Thailand should promote easily accessible markets and provide practical travel tips to enhance the experience for Japanese tourists. Tailoring marketing efforts to highlight cultural heritage, paired with practical advice on navigating Thailand's tropical climate, supports long-term engagement and strengthens cultural diplomacy with Japan.

- The United States The United States presents significant potential for enhancing Thailand's soft power through culinary and cultural tourism. Thai cuisine is highly popular but requires careful management of spice levels to cater to varying tolerances among American diners. Addressing inaccuracies in how some influencers portray Thai food by providing authentic recipes and adaptable guidelines can enhance the perception of Thai cuisine. This strategy ensures Americans associate Thai food with authenticity and quality. In tourism, Thailand's affordability, combined with its immersive cultural experiences, positions the country as an attractive destination for budget-conscious travelers. Promoting destinations like Chiang Mai and Krabi with an emphasis on sustainable tourism and cultural immersion aligns with American travelers' growing interest in eco-friendly travel. Additionally, developing mobile applications and digital guides offering travel tips, insights into Thai customs, and etiquette will enhance cross-cultural understanding and provide a seamless experience for American visitors. Culinary tourism, including food tours and cooking workshops, remains a critical area to highlight Thailand's gastronomic diversity.
- India India demonstrates a high level of interest in Thai culture, particularly in cuisine and adventure tourism. The detailed discussions among Indian audiences about Thai cooking techniques and street food suggest significant opportunities for engagement. Organizing food festivals, live cooking demonstrations, and immersive culinary experiences such as cooking classes can cater to this interest. Collaborations with Indian food influencers further amplify Thailand's culinary appeal by reaching wider audiences. Showcasing Thailand's rich culinary heritage through digital platforms strengthens its soft power and positions Thailand as a top gastronomic destination for Indian tourists.

Additionally, India's interest in adventure and nature-based tourism, reflected in discussions about destinations like Pattaya Tiger Park, Chiang Mai, and Pai, suggests an opportunity to focus marketing efforts on wildlife encounters and cultural immersion. Tailored travel packages, adventure festivals, and

collaborations with Indian travel influencers specializing in adventure tourism can attract Indian tourists seeking unique experiences. Emphasizing Thailand's commitment to wildlife conservation and cultural heritage in global media enhances its image as a responsible and attractive destination, aligning promotional efforts with Indian travelers' interests.

ACKNOWLEDGMENT

Gratitude is extended to all individuals and organizations who contributed to the successful completion of this research project. Particular acknowledgment is given to Assoc. Prof. Dr. Teerayut Horanont and the academic advisors for their invaluable guidance, constructive feedback, and continuous support throughout the research process. Appreciation is also directed to the cultural and tourism organizations for providing essential data and insightful perspectives that significantly enhanced the study. Special thanks are given to the participants in the data collection phase, whose contributions were pivotal in shaping the findings. Finally, recognition is due to the support and encouragement provided by peers and family, whose understanding and motivation played a critical role in overcoming the challenges of this academic endeavor.

REFERENCES

- [1] J. Nye, "Soft Power: The Origins and Political Progress of a Concept," *Journal of Political Science & International Relations*, vol. 1, no. 3, pp. 15–20, 2008.
- [2] "Soft power," Wikipedia, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Soft_power
- [3] "Tools of Foreign Policy - What Is Soft Power?" CFR Education, 2022. [Online]. Available: <https://education.cfr.org/learn/learning-journey/tools-foreign-policy/what-is-soft-power>
- [4] A. Anholt, "Nation-brands and the value of provenance," *Economic Development Quarterly*, vol. 17, no. 4, pp. 334–343, 2003. [Online]. Available: <https://doi.org/10.1177/0891242403250714>
- [5] M. Fan, "Branding the Nation: Towards A Better Understanding," *Place Branding and Public Diplomacy*, vol. 6, no. 2, pp. 97–103, 2010. [Online]. Available: <https://doi.org/10.1057/pb.2010.16>
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, Aug. 1987. [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [8] S. Anholt, *Competitive Identity: The New Brand Management for Nations, Cities and Regions*. New York, NY: Palgrave Macmillan, 2007.
- [9] P. Rockower, "Recipes for Gastrodiplomacy," *Place Branding and Public Diplomacy*, vol. 8, no. 3, pp. 235–246, 2012. [Online]. Available: <https://doi.org/10.1057/pb.2012.17>
- [10] G. Richards, *Cultural Tourism: Global and Local Perspectives*. Routledge Studies in Cultural History. Routledge, 2007.
- [11] D. Buhalis and R. Law, "Progress in Information Technology and Tourism Management: 20 Years On and 10 Years After the Internet," *Tourism Management*, vol. 29, no. 4, pp. 609–623, 2008. [Online]. Available: <https://doi.org/10.1016/j.tourman.2008.01.005>
- [12] R. Buckley, *Adventure Tourism Management*. Elsevier, 2011.
- [13] L. M. Long, *Culinary Tourism*. Lexington, KY: University Press of Kentucky, 2004.

Development of IoT-Based Compact Mushroom Cultivation Monitoring System

1st Noor Ropidah Bujal

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
noorropidah@gmail.com*

4th Syazana Md Dali

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
syazana@polisas.edu.my*

2nd Shaizan Jusoh

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
shaizeen@polisas.edu.my*

3rd Mokhtar Hashim

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
mokhtar1@polisas.edu.my*

5th Nor Sukor Ali

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
norsukorali@polisas.edu.my*

6th Mohamed Noor Azman Bidin

*Dept. of Electrical Engineering
Sultan Haji Ahmad Shah Polytechnic
Kuantan, Pahang, Malaysia
noorazman@polisas.edu.my*

Abstract— Development of an IoT-Based Compact Mushroom Cultivation System (CMCS) project integrates advanced Internet of Things (IoT) technology with traditional mushroom cultivation practices to address challenges enthusiasts, small-scale farmers, and researchers face. CMCS aims to create an intelligent cultivation environment that accurately controls important environmental factors like temperature and humidity to optimize the conditions for mushroom growth. This project aligns with Sustainable Development Goal (SDG) 2, which seeks to eliminate hunger, achieve food security, improve nutrition, and adopt sustainable agriculture. The compact system will be outfitted with an Internet of Things (IoT) based automatic control system that can regulate temperature and humidity fluctuations in the environment. This system will include the Raspberry Pi 4, Arduino UNO, DHT22 sensor, LDR sensor, relays, Node-Red, and MQTT Dashboard. Together, these elements create a comprehensive structure that allows for the real-time monitoring and control of the cultivation environment. The data gathered from the system undergoes processing and analysis using a cloud-based platform, which offers real-time insights and control through a user-friendly interface. This study investigates the creation and execution of the system, emphasizing the difficulties faced and the strategies developed to overcome them. The implementation of CMCS promises to transform conventional mushroom farming through its ability to offer accurate and immediate monitoring and control, resulting in improved efficiency and long-term viability. In addition, it also supports sustainable agricultural practices, which are in line with the broader goals of SDG 2.

Keywords— *Compact Mushroom Cultivation System, IoT monitoring system, smart farming, smart agriculture, real monitoring system*

I. INTRODUCTION

Mushroom cultivation, which involves growing a highly nutritious and commercially valuable food product, poses various challenges due to the specific environmental conditions needed. Conventional techniques require significant manual labor and could be more effective, resulting in unpredictable crop production. [1]. The advent of IoT technology offers a promising solution by enabling automated, precise control of growing conditions[2]. The author in [3] designed a real-time control and monitoring system for mushroom cultivation fields based on a Wireless

Sensor Network (WSN) and the Internet of Things (IoT) to control and monitor temperature and humidity parameters in each field. The main components of the node board consist of Arduino Uno, which is connected to XBee, a DHT22 sensor, an actuator in the form of a sprayer pump, a blower, and a lamp together with Raspberry Pi and XBee in the master board. Another system uses a DHT-11 sensor, Arduino Uno microcontroller, MCU ESP8266, and the internet-of-things (IoT) platforms called Cayenne [1]. In [4], researchers developed an IoT-based Oyster Mushroom Farming Monitoring System that incorporates sensor networks, data analytics, and automation to enhance the efficiency of the cultivation process by offering meticulous regulation of environmental factors such as temperature, humidity, light, and ventilation.

Therefore, the conceptual framework for an IoT-based Compact Mushroom Cultivation System is presented in this paper. Specific goals include outlining the system design, discussing its implementation, evaluating its benefits and challenges, and proposing future enhancements. By doing so, the paper seeks to demonstrate the potential of IoT technology in revolutionizing mushroom farming practices. This paper is organized into several sections, which include an introduction and previous research, which provide context and identify gaps in current practices. The methodology section includes the system design section, which details the components and architecture of the proposed system. Project concept and implementation describe the prototype development and initial testing results. Then, benefits and challenges are discussed, followed by a discussion of the results and their implications. The paper concludes with future work, acknowledgment, and references.

II. LITERATURE REVIEW

Traditional mushroom farming methods involve manually monitoring and adjusting environmental conditions, which can be labor-intensive and prone to human error [5]. Techniques such as substrate preparation, spawn inoculation, and environmental control are crucial but often lack precision, leading to variable yields. Recent advancements have introduced more automated and controlled environments, but many systems remain costly

and complex [6]. IoT technology has been increasingly applied in agriculture to create intelligent farming solutions that enhance efficiency and productivity. Examples include precision farming, automated irrigation systems, and real-time soil monitoring [7]-[8]. These applications have significantly improved crop management, resource utilization, and yield prediction [9]. However, there is limited research specifically focused on IoT applications in mushroom cultivation. Traditional methods face several challenges, including inconsistent environmental conditions, high labor requirements, and limited scalability [10]. Manual monitoring can delay responses to environmental changes, negatively impacting yield and quality. Additionally, traditional systems often lack the ability to collect and analyze data, hindering efforts to optimize and improve cultivation practices.

III. METHODOLOGY

A. Project concept

Mushroom cultivation is an ancient practice that has gained modern innovation through the integration of technology. This research explores the creation of an Arduino-based automated mushroom cultivation system using simple components like an Arduino Uno, a DHT22 temperature and humidity sensor, and a relay module to control a water pump. The system aims to provide a controlled environment for optimal mushroom growth, ensuring that the temperature and humidity levels are maintained at ideal levels. Fig. 1 shows the block diagram of the developed IoT-based Compact Mushroom Cultivation System.

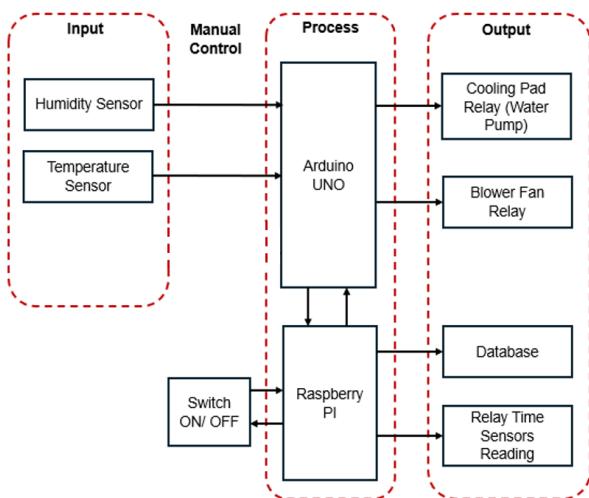


Fig. 1. Block diagram of developed IoT-based Compact Mushroom Cultivation System

The Arduino Uno serves as the brain of the automated cultivation system, acting as the central processing unit to control and coordinate the various components. Its open-source nature and ease of programming make it an excellent choice for this project. Programmed with the Arduino IDE, the Uno reads data from the DHT11 sensor and sends signals to the relay module, allowing precise control over the environmental conditions essential for successful mushroom cultivation.

The DHT22 sensor is a crucial component in this project, as it provides real-time data on temperature and humidity levels within the cultivation environment. Placed strategically within the mushroom growing chamber, DHT22 constantly monitors the conditions and relays this information to the Arduino Uno. The accuracy of the DHT22 ensures that the cultivation environment remains within the optimal range for the chosen mushroom species.

To maintain the necessary humidity levels, an automated watering system is implemented using a relay module. The relay acts as a switch, controlled by the Arduino Uno, to turn the water pump on and off. When the humidity levels drop below the preset threshold, the Arduino signals the relay to activate the water pump, ensuring that the growing medium remains adequately moistened for optimal mushroom development.

Calibrating the system is a crucial step to ensuring accurate and reliable operation. Users can input specific parameters, such as the ideal temperature and humidity range for the targeted mushroom species, into the Arduino program. Through trial and error, adjustments can be made to achieve the optimal settings, creating a tailored environment that maximizes mushroom yield. Regular monitoring of the cultivation system is essential for success. Users should regularly check the DHT22 sensor readings, ensuring that the temperature and humidity remain within the specified range. The water pump and relay connections should also be inspected to guarantee proper functionality. Any deviations from the desired conditions should prompt immediate adjustments to prevent adverse effects on mushroom growth.

The Arduino-based automated mushroom cultivation system presented here offers an efficient and cost-effective solution for small-scale farmers. By integrating the Arduino UNO, DHT22 sensor, and relay module, enthusiasts can create a controlled environment that promotes optimal mushroom growth. This project serves as an example of how accessible technology can enhance traditional agricultural practices, opening new possibilities for sustainable and automated farming techniques.

The Compact Mushroom Cultivation System (CMCS) is based on combining advanced, especially the Internet of Things (IoT), with traditional mushroom-growing methods. Recognizing the challenges enthusiasts, small-scale farmers, and researchers face in maintaining optimal conditions for mushroom growth, this CMCS aims to offer a thorough and flexible solution. Central to this idea is the creation of an intelligent growing environment. It uses IoT connections to control essential environmental factors, like temperature and humidity, precisely.

The main idea of CMCS is to bring together advanced parts, including the Raspberry Pi 4, Arduino UNO, D22 sensor, LDR sensor, relays, Node-Red, & MQTT Dashboard. Together, these elements create a complex system that enables real-time monitoring & control of the cultivation environment. By employing state-of-the-art technology, the project seeks to automate and enhance the cultivation process, addressing the limitations of traditional manual methods. The development of CMCS involves a systematic and integrated approach to leveraging Internet of Things (IoT) technology to enhance mushroom cultivation. The methodology comprises several key steps and components

that collectively contribute to creating an efficient and automated cultivation environment, as shown in Figure 1.

B. System Design Architecture

CMCS is a combination of hardware and software where Arduino Uno R3 and Raspberry Pi 4 are the heart of the project. For the software part, Arduino IDE and Proteus Professional are used to program Arduino. The proposed IoT-based system consists of interconnected components designed to create an optimal growing environment for mushrooms. The scope of the work extends to providing a user-friendly interface for monitoring and control. The integration of Node-Red and MQTT Dashboard ensures accessibility for users to view real-time data, receive alerts, and remotely manage the cultivation environment. This user-friendly interface addresses the need for a practical and efficient tool that allows farmers, researchers, and enthusiasts to interact with the system regardless of their physical proximity to the cultivation site. The Message Queuing Telemetry Transport (MQTT) Dashboard plays a pivotal role, serving as a powerful tool for managing and monitoring IoT devices. It could facilitate real-time communication between devices. Using the publish-subscribe architecture, MQTT enables seamless data exchange between connected devices. The architecture includes sensors, actuators, a microcontroller unit (MCU), and cloud connectivity. [11] These components work together to monitor and control environmental conditions in real-time, ensuring consistent and optimal growth parameters.

Node-RED is an open-source, flow-based development tool that provides a visual programming environment for wiring together devices, APIs, and online services. One of Node-RED's distinctive features is its intuitive, web-based interface, which allows users to create visual representations of their workflows. It supports a wide range of nodes out of the box, enabling users to interact with various devices, protocols, and data sources. The platform allows users to respond to real-time events by triggering specific actions, making it particularly well-suited for applications that require rapid and dynamic responses.

A network of sensors is deployed to continuously monitor critical parameters such as temperature, humidity, and light intensity. [12]. These sensors provide accurate, real-time data that is essential for maintaining the optimal conditions required for mushroom growth at various stages of development. Actuators are integrated into the system to adjust the growing environment based on sensor data [13]. Automated control mechanisms enable precise adjustments to be made quickly and efficiently, ensuring the environment remains within the desired range. The MCU serves as the system's brain, processing sensor data and sending commands to the actuators [14]. It is programmed with algorithms that use historical data and real-time feedback to adjust the growing conditions, optimizing the environment for maximum yield.

The system is connected to a cloud platform, enabling remote monitoring and control via a user-friendly interface [15]. Farmers can access real-time data, receive alerts, and adjust from any location using a smartphone or computer. Cloud connectivity also facilitates data storage and analysis, providing valuable insights for continuous improvement.

Advanced data analytics tools on the cloud platform help farmers analyze trends, predict outcomes, and optimize growing conditions [16], [17].

A vital concept is developing an intelligent watering system. This system, powered by a Raspberry Pi 4 and an Arduino UNO, monitors and regulates the water supply to the cooling pad based on predefined parameters and real-time environmental conditions. This innovation eliminates manual intervention, ensuring consistent and controlled moisture levels conducive to optimal mushroom growth. The concept emphasizes the precision and efficiency of integrating these technologies into the cultivation process. Fig. 2 and 3 show a flowchart of the monitoring system process and hardware design of CMCS. The system incorporates IoT-based temperature and humidity real-monitoring controls., which can be monitored and managed via a smartphone application or computer.

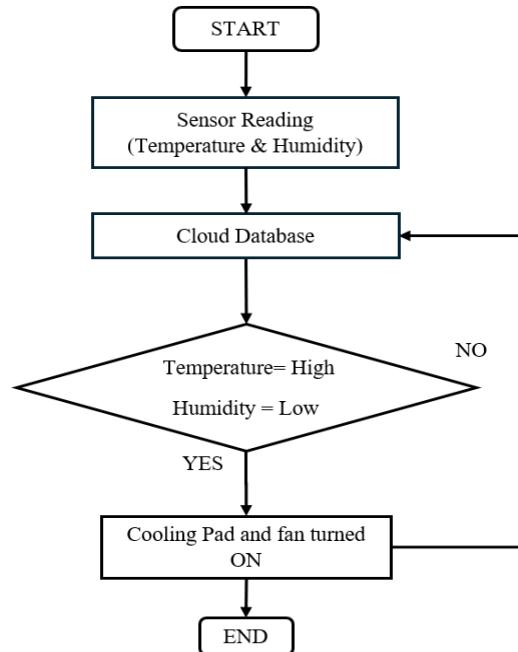


Fig. 2. Overall process of automated cultivation system

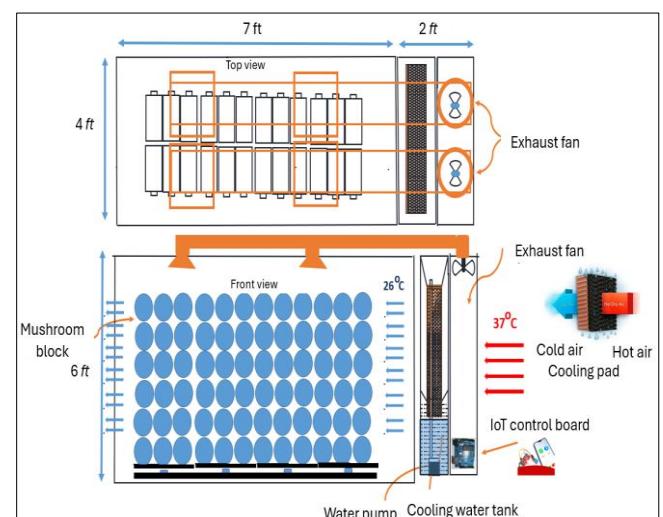


Fig. 3. Hardware design of CMCS

IV. RESULTS AND DISCUSSION

A. Implementation of project

The Compact Mushroom Cultivation System (CMCS) prototype, which is based on the Internet of Things (IoT), was developed to verify the potential of an intelligent system for mushroom farming. Various technological components, including sensors, actuators, and a microcontroller unit (MCU), are integrated into the prototype's small grow chamber. This system is linked to a cloud platform, which facilitates the remote monitoring and management of the cultivation environment.

The hardware components of the prototype were carefully designed to satisfy the requirements of mushroom cultivation. This encompasses actuators that can modify conditions as required and sensors that monitor environmental parameters. In terms of MCU programming, the MCU, which could be based on platforms such as the Raspberry Pi 4 and Arduino UNO, was programmed to regulate the system's operations, which included the processing of sensor data and the management of actuator responses. The grow chamber can be remotely monitored and controlled by users through the integration of the system with a cloud platform. In order to facilitate data communication, this connection implements protocols such as MQTT (Message Queuing Telemetry Transport). Data was gathered during the testing phase to assess the extent to which the CMCS maintained optimal conditions for mushroom growth.

To ensure that the grow chamber remains within the optimal temperature range for mushroom cultivation maintains appropriate humidity levels to support mushroom development and assesses the system's efficiency in utilizing resources such as water and energy, key parameters monitored include temperature, humidity, and resource utilization. The effectiveness of environmental sensors and actuators will be examined in the context of analysis of system components. The objective was to ascertain the degree of precision and dependability with which these components could regulate the conditions of the grow chamber.

The system employs node-Red and the MQTT Dashboard to create a user-friendly interface for monitoring and control. The analysis will concentrate on the degree of the interface's intuitiveness and effectiveness for users, as well as its integration with the rest of the system. The research evaluated the system's overall performance in real-world scenarios. This encompasses an assessment of the system's adaptability to varying conditions or user requirements, as well as the precision of environmental control. The comprehensive analysis was designed to offer a thorough understanding of the system's overall impact on mushroom cultivation, as well as its functionality and reliability. The CMCS endeavors to improve the efficiency and sustainability of agricultural practices by optimizing the cultivation process. This can result in enhanced yield and resource management, making mushroom farming more environmentally friendly and economically viable.

B. Performance of CMCS

A comparison between the traditional mushroom cultivation method and the developed IoT-based Compact

Mushroom Cultivation Monitoring System (CMCS) is presented in Table I.

TABLE I. COMPARISON OF TRADITIONAL MUSHROOM CULTIVATION METHOD AND CMCS

Aspect	Traditional Method	IoT-based CMCS
Temperature Condition	Daytime temperature of 30-35°C	Controlled temperature of 25-28°C
Humidity	Uncontrolled, dependent on weather	Automatically controlled for optimal humidity
Mushroom Yield	80 grams per block	250 grams per block
Yield Reduction	Decreases by 50% due to hot weather	No significant reduction in yield
Quality Consistency	Inconsistent, affected by weather factors	More consistent quality with IoT control

The pilot test data on mushroom production and income were collected for 300 blocks of mushrooms using the development of IoT-based CMCS as in Fig. 4 (a) and 4 (b). The traditional method yields 80 grams per block, while the IoT system yields 250 grams per block. The yield increases by 51 kg (from 24 kg to 75 kg), which is over three times higher. Fig. 4 (b) indicates that the IoT-based system generates a much higher income. The income rises from RM 360 to RM 1125 with IoT implementation, showing significant potential for increased productivity and profitability. A brief comparison of CMCS with the previous system is shown in Table II.

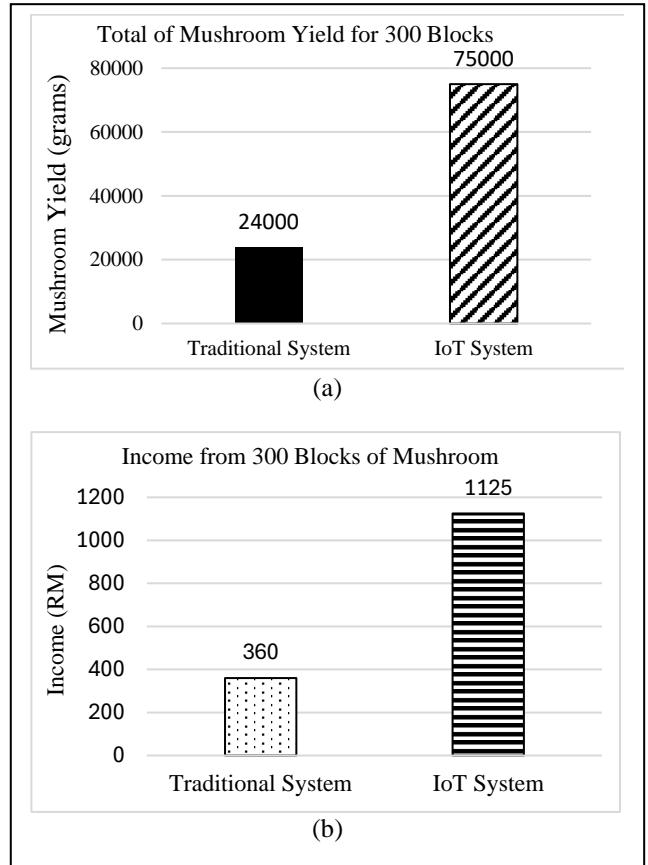


Fig. 4. Mushroom production and income for 300 blocks of mushrooms

TABLE II. COMPARISON OF CMCS WITH THE PREVIOUS SYSTEM

Aspect	Previous System	IoT-based CMCS
Cost	- Incorporates XBee modules and advanced architecture, potentially increasing costs [3]. - Focuses on advanced analytics, raising costs [4].	- It uses affordable components like Raspberry Pi 4, Arduino UNO, and DHT22 sensors. - Ensuring cost-effectiveness for small-scale farmers.
Scalability	A centralized low-cost system may limit scalability across different operational scales [5].	Modular design allows easy scalability for both small-scale and commercial operations and is adaptable to various mushroom types.
Efficiency	-Integrates solar energy but focuses more on sustainability than automation precision [1]. -Broader agriculture focus, less specific[7]-[8].	Provides precise environmental control and automation, increasing yield from 80 grams to 250 grams per block.
User-friendliness	Systems like Cayenne provide interfaces but may require more setup [4].	It includes a Node-RED interface and an MQTT Dashboard for intuitive monitoring and control.
Sustainability	It emphasizes renewable energy integration (solar) but lacks detailed optimization for specific parameters [1].	Optimize water, energy, and other resources, contributing to eco-friendly farming practices.

C. Benefits and Challenges

The automated monitoring and control capabilities of the IoT-based system led to more consistent and optimal growing conditions, resulting in higher yields and better-quality mushrooms. The system reduces the need for manual intervention, freeing up labor for other tasks and improving overall efficiency [18]. The system's modular design allows for easy scalability, making it suitable for both small-scale urban farms and large commercial operations. The flexibility of the system enables it to be adapted to different mushroom varieties and growing conditions, enhancing its applicability across diverse contexts [19]. Precise control of environmental factors leads to better resource utilization, reducing waste, and lowering costs. The system optimizes the use of water, energy, and other inputs, contributing to more sustainable farming practices [20].

The IoT-based system simplifies the cultivation process, making it more accessible to small-scale and urban farmers who may lack the expertise or space for traditional methods. The compact design and automated features enable efficient mushroom farming in limited spaces [11]. The initial investment in IoT technology and equipment may be a barrier for some farmers. However, a cost-benefit analysis suggests that the long-term savings from increased yields, reduced labor, and optimized resource use can offset the initial costs, making the system economically viable. Farmers may require training to use and maintain the IoT-based system effectively. Providing support and educational resources is essential to

ensure successful adoption and operation. Developing user-friendly interfaces and comprehensive manuals can help mitigate this.

V. CONCLUSION

The Development of the IoT-Based Compact Mushroom Cultivation System (CMCS) project has successfully integrated advanced IoT technologies, including the Raspberry Pi 4, Arduino UNO, and various sensors, to create an efficient and automated framework for mushroom cultivation. The system's precision will maintain optimal environmental conditions and enhance yield and quality while reducing labor intensity, making it highly practical for small-scale and urban farmers. It promotes sustainable agricultural practices by optimizing resource use and reducing waste. However, initial costs and the need for reliable internet and power supply present challenges. Future research should focus on cost reduction, alternative power sources, extensive field trials, and advanced data analytics. The CMCS project demonstrates the transformative potential of IoT in agriculture, paving the way for more efficient, sustainable, and productive farming practices and supporting Malaysia's Sustainable Development Goal.

ACKNOWLEDGMENT

The authors are highly grateful to Sultan Haji Ahmad Shah Polytechnic (POLISAS) for providing a research platform and endless support from the beginning of this research. We also would like to show our gratitude to the Ministry of Higher Education as this study is sponsored under the TVET Applied Research Grant Scheme (T-ARGS), which made this research possible. Their commitment to advancing technical and vocational education has contributed significantly to the success of this work.

REFERENCES

- [1] M. Rukhiran, C. Sutanthavibul, S. Boonsong, and P. Netinant, "IoT-Based Mushroom Cultivation System with Solar Renewable Energy Integration: Assessing the Sustainable Impact of the Yield and Quality," *Sustainability*, vol. 15, no. 18, p. 13968, Sep. 2023, doi: 10.3390/su151813968.
- [2] M. Ataei Kachouei, A. Kaushik, and M. A. Ali, "Internet of Things-Enabled Food and Plant Sensors to Empower Sustainability," *Adv. Intell. Syst.*, vol. 5, no. 12, 2023, doi: 10.1002/aisy.202300321.
- [3] Wajiran, S. D. Riskiono, P. Prasetyawan, A. Mulyanto, M. Iqbal, and R. Prabowo, "Control and Realtime Monitoring System for Mushroom Cultivation Fields based on WSN and IoT," *J. Phys. Conf. Ser.*, vol. 1655, no. 1, 2020, doi: 10.1088/1742-6596/1655/1/012003.
- [4] N. A. Zainuddin, K. M. Nordin, A. Ishak, M. S. Nuriman, and Tajul Irfan Danial Tajul Azhar, "IoT-Based Oyster Mushroom Farming Monitoring System," *Int. J. Multidiscip. Res.*, vol. 5, no. 3, pp. 1–7, Jun. 2023, doi: 10.36948/ijfmr.2023.v05i03.4051.
- [5] D. P. Guragain, B. Shrestha, and I. Bajracharya, "A low-cost centralized IoT ecosystem for enhancing oyster mushroom cultivation," *J. Agric. Food Res.*, vol. 15, p. 100952, Mar. 2024, doi: 10.1016/j.jafr.2023.100952.
- [6] Y. Edan, G. Adamides, and R. Oberti, "Agriculture Automation," Springer, Cham, 2023, pp. 1055–1078.
- [7] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Enhancing smart farming through the applications of Agriculture 4.0 technologies," *Int. J. Intell. Networks*, vol. 3, pp. 150–164, Jan. 2022, doi: 10.1016/j.ijn.2022.09.004.

- [8] E. Bwambale, F. K. Abagale, and G. K. Anornu, "Smart irrigation monitoring and control strategies for improving water use efficiency in precision agriculture: A review," *Agric. Water Manag.*, vol. 260, p. 107324, Feb. 2022, doi: 10.1016/J.AGWAT.2021.107324.
- [9] A. Ali, T. Hussain, N. Tantashutikun, N. Hussain, and G. Cocetta, "Application of Smart Techniques, Internet of Things and Data Mining for Resource Use Efficient and Sustainable Crop Production," *Agric.*, vol. 13, no. 2, 2023, doi: 10.3390/agriculture13020397.
- [10] M. Shirur, A. Barh, and S. K. Annepu, "Sustainable production of edible and medicinal mushrooms: Implications on mushroom consumption," in *Climate Change and Resilient Food Systems*, Singapore: Springer Singapore, 2021, pp. 315–346.
- [11] S. Rudrakar and P. Rughani, "IoT based agriculture (Ag-IoT): A detailed study on architecture, security and forensics," *Inf. Process. Agric.*, Sep. 2023, doi: 10.1016/j.inpa.2023.09.002.
- [12] T. L. Narayana et al., "Advances in real-time smart monitoring of environmental parameters using IoT and sensors," *Heliyon*, vol. 10, no. 7, p. e28195, Apr. 2024, doi: 10.1016/j.heliyon.2024.e28195.
- [13] M. N. Halgamuge, A. Bojovschi, P. M. J. Fisher, T. C. Le, S. Adelolu, and S. Murphy, "Internet of Things and autonomous control for vertical cultivation walls towards smart food growing: A review," *Urban For. Urban Green.*, vol. 61, p. 127094, Jun. 2021, doi: 10.1016/J.UFUG.2021.127094.
- [14] A. L. Pedida, "Integrating AI and IoT in Mushroom Growing Chamber Integrating AI and IoT in Mushroom Growing Chamber Arthur Liwanan Pedida Senior Lecturer Institute of Technical Education (ITE College East) School of Electronics and Info-Comm Technology," no. June, pp. 0–12, 2024.
- [15] A. Elewi, A. Hajhamed, R. Khankan, S. Duman, A. Souag, and A. Ahmed, "Design and implementation of a cost-aware and smart oyster mushroom cultivation system," *Smart Agric. Technol.*, vol. 8, no. March, p. 100439, 2024, doi: 10.1016/j.atech.2024.100439.
- [16] Punsala S.M.P., Bentara B.W.H.T, Ferdinando D.V.P., Chathura W.B.D., Silva S., and Liyanage R., "IoT based Plant Growth and Health Monitoring System for Mushrooms," *Int. J. Eng. Manag. Res.*, vol. 13, no. 4, pp. 33–40, 2023, doi: 10.31033/ijemr.13.4.4.
- [17] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5602–5618, Sep. 2022, doi: 10.1016/j.jksuci.2021.05.013.
- [18] L. Klerkx, E. Jakku, and P. Labarthe, "A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda," *NJAS - Wageningen J. Life Sci.*, vol. 90–91, p. 100315, 2019, doi: 10.1016/j.njas.2019.100315.
- [19] M. Hafizuddin, M. Hishamuddin, and A. Ubin, "Mushroom House Monitoring System Using Internet of Things (IoT)," *Evol. Electr. Electron. Eng.*, vol. 3, no. 1, pp. 545–553, 2022, [Online]. Available: <http://publisher.uthm.edu.my/periodicals/index.php/eeee>.
- [20] J. de A. Dornelles, N. F. Ayala, and A. G. Frank, "Smart Working in Industry 4.0: How digital technologies enhance manufacturing workers' activities," *Comput. Ind. Eng.*, vol. 163, 2022, doi: 10.1016/j.cie.2021.107804.

Development of a Time Series Visualization Application for PM2.5 Level Changes Using Google Earth Engine

1st Natima Udon

Department of Natural Resources and Environment, Faculty of Agriculture Nature Resources and Environment, Naresuan University. Phisanulok, Thailand.

natimau66@nu.ac.th

2nd Kampanart Piyathamrongchai

Department of Natural Resources and Environment, Faculty of Agriculture Nature Resources and Environment, Naresuan University. Phisanulok, Thailand.

3rd Sittichai Choosumrong*

Department of Natural Resources and Environment, Faculty of Agriculture Nature Resources and Environment, Naresuan University. Phisanulok, Thailand.

Center of Excellence in Nonlinear Analysis and Optimization, Faculty of Science, Naresuan University, Phitsanulok, Thailand

*Coresponding Author:
sittichaic@nu.ac.th

Abstract — Over the years, Thailand has been facing a severe air pollution problem that has had a significant impact on people's health. One of the most important and dangerous pollution problems is fine particulate matter (PM2.5), which is a particulate matter smaller than 2.5 microns that can enter the body through the respiratory and circulatory systems, causing various health problems. This is mainly due to the burning of agricultural and forestry land. Burning in open areas causes a lot of smoke. If people are affected by PM2.5 accumulated for a long time, there will be a health risk. Respiratory and vascular diseases, which will affect the working process of various organs in the body.

Therefore, managing the PM2.5 problem in Thailand requires collaboration between the government and the government. This research focuses on the development of an application that can visualize time series data to analyze changes in PM2.5 levels by applying Google Earth Engine (GEE), a platform that helps to process large-scale geospatial data. This developed application will allow users to effectively track and analyze changes in PM2.5 levels through the visualization of data in the form of interactive graphs and maps. Using data from the Sentinel-5P satellite and air quality monitoring stations to process and analyze PM2.5 data at different times, and present it through an application platform developed through JavaScript coding on GEE tools and technologies, this research will help local executives and stakeholders to make effective decisions and plan for air quality management and can be deployed in other areas to support future air quality management and related research.

Keywords — Air pollution, Fine particulate matter, Satellite imagery, Spatial analysis.

I. INTRODUCTION

Thailand is facing a severe air pollution problem that has a significant impact on people's health. One of the most important and extremely dangerous pollutants is fine particulate matter (PM2.5), which is a dust smaller than 2.5

microns, making it easy to enter the respiratory and circulatory systems. It causes health problems such as respiratory diseases. The PM2.5 problem in Thailand is caused by many factors, including fuel burning in vehicles, the manufacturing industry, and the production industry. This problem is likely to worsen during the winter months as calm winds and high air pressure cause dust to accumulate in the low atmosphere for a long time. Many provinces in Thailand, especially in the North, such as Chiang Mai Province, Lamphun and Phitsanulok. Faced with this problem hard. This is mainly due to the burning of agricultural and forest areas that produce a large amount of smoke. In addition, in Bangkok and its vicinity, high PM2.5 levels have been found due to heavy traffic and various constructions [2].

Problem management PM2.5 in Thailand requires collaboration between the government, private sector and citizens. By taking various measures such as controlling the burning of biomass, promoting the use of clean energy, increasing green spaces, and continuously monitoring air quality to reduce the impact on public health and improve air quality in the long term [3].

This research focuses on developing a platform that provides an overview of PM2.5 level changes, specifically targeting nine provinces in the lower northern region of Thailand. The platform utilizes Google Earth Engine (GEE), a tool for processing large-scale geospatial data, to create a web application through JavaScript. This platform will enable users to efficiently track and analyze PM2.5 level changes through interactive graphs and map displays, leveraging data from Sentinel-5P satellite imagery for analysis and processing on the GEE platform. This research aims to develop an application capable of visualizing time-series data to analyze changes in PM2.5 levels, providing a tool to support decision-making and air quality management planning for local authorities and stakeholders. The application is expected to serve as a valuable tool for monitoring and analyzing PM2.5 variations, with potential applicability in other regions to support air quality management and related research in the future [4].

II. METHODOLOGY AND METHODS

This research employs a quantitative approach to develop and validate an application for visualizing time-series data to analyze changes in PM_{2.5} levels. Leveraging the capabilities of GEE for processing and analyzing large geospatial datasets, the application utilizes Sentinel-5P satellite imagery.

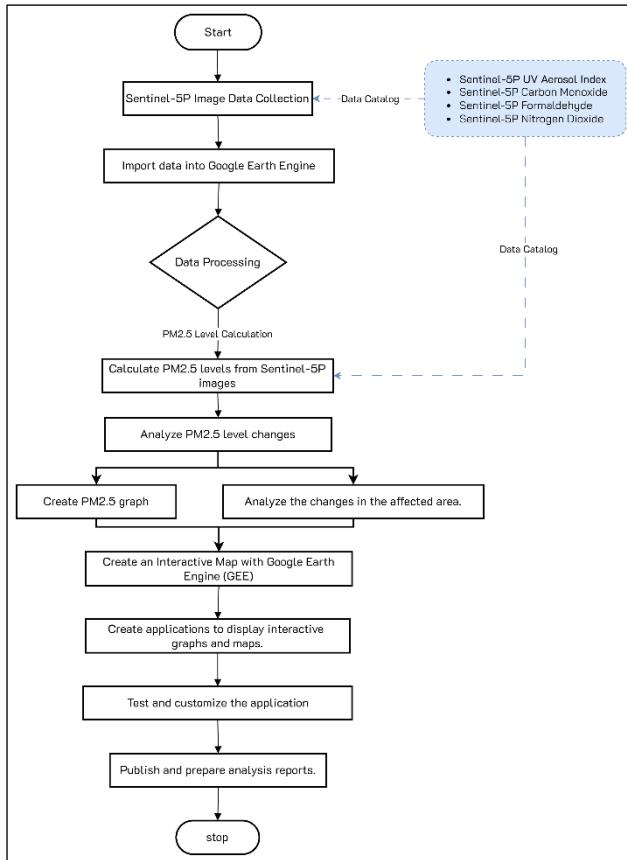


Figure 1 Flowchart Operation

Based on the above conceptual framework, (Figure1) this research will analyze PM_{2.5} levels using Sentinel-5P satellite data, including the Sentinel-5P UV Aerosol Index, Carbon Monoxide, Formaldehyde, and Nitrogen Dioxide datasets, to investigate factors influencing PM_{2.5} formation. The PM_{2.5} data will then be processed and prepared as follows: import data, filter by time period and area, fill in missing values with the average or median, and calculate daily or monthly averages, with display parameters set according to analysis needs. Pre-processing in GEE prepares the data for effective analysis and visualization, facilitating more accurate PM_{2.5} level assessments [5][6][7]. In the processing phase, daily, weekly, or monthly averages are calculated from raw data to provide a representative view of each period, with calculations of the mean, standard deviation, or maximum and minimum values as needed. Time-series visualizations will display PM_{2.5} trends over time. The GEE application development process comprises several key stages, including planning, design, development, and testing. Earth Engine's UI components will be used to create interactive maps and graphs, enabling users to engage with the data.

Once completed, the code and application can be saved and shared via a link for broader access [8][9][10].

III. METERIALS

For research focused on PM_{2.5} analysis using satellite imagery and application development GEE, the tools used include.

A. Study Area

For this research, we selected the study area for PM_{2.5} analysis in 9 provinces in the lower northern part of Thailand, namely Phitsanulok, Kamphaeng Phet, Sukhothai, Uttaradit, Uthai Thani, Phichit, Phetchabun, Tak and Nakhon Sawan.

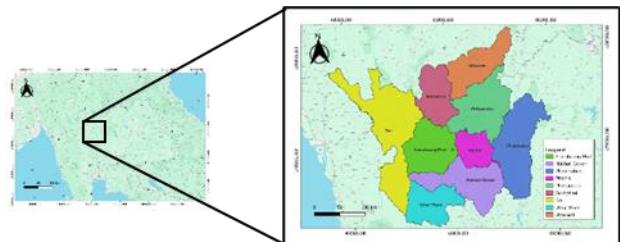


Figure 2 Study area with particulate matter (PM) of 9 Lower Northern Provinces

B. Google Earth Engine (GEE)

GEE is a cloud-based geospatial analysis tool that allows users to view and analyze satellite images of Earth. Researchers in fields such as remote sensing, disease outbreak forecasting, natural resource management, and more, including academics and nonprofit organizations, use Earth Engine extensively. With this knowledge, your students can engage in these discussions and potentially pursue careers as data scientists [11][12].

C. JavaScript

In GEE, JavaScript is a dynamic programming language used for developing websites, web apps, games, and more. It allows you to add interactive elements to web pages that cannot be achieved with HTML and CSS alone.

D. Data

This study processes and analyzes PM_{2.5} is The Sentinel-5 Precursor mission instrument gathers data to evaluate air quality. With a spatial resolution of 0.01 arc degrees, the TROPOMI instrument is a multispectral sensor that measures the reflectance of wavelengths crucial for determining the concentrations of ozone, methane, formaldehyde, aerosol, carbon monoxide, nitrogen oxide, and sulfur dioxide in the atmosphere. It also records cloud characteristics [13][14].

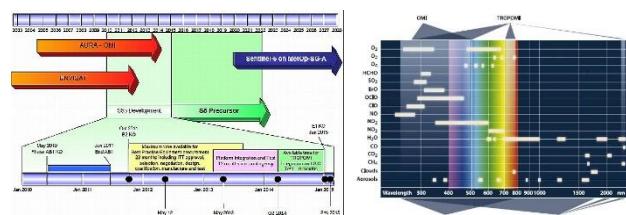


Figure 3 Tropospheric Monitoring Instrument (TROPOMI)

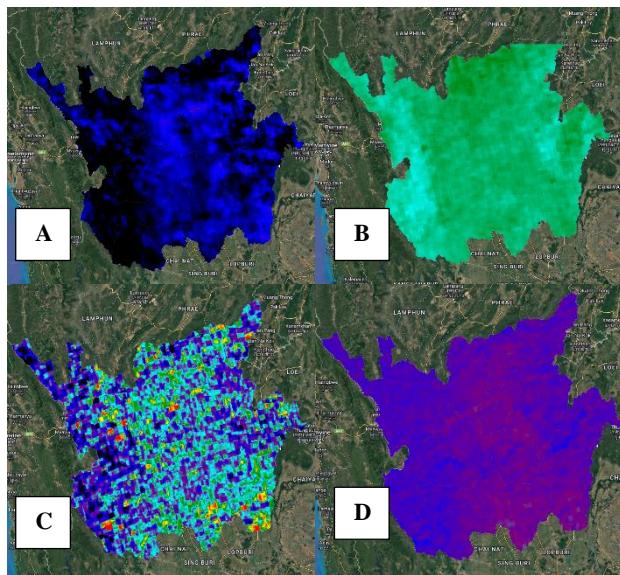


Figure 4 Sentinel-5P UV Aerosol Index (A), Sentinel-5P Carbon Monoxide (B), Sentinel-5P Formaldehyde(C) and Sentinel-5P Nitrogen Dioxide (D)

From *Figure 4*, These data can be used to assess the distribution of particulate matter and airborne pollutants that correlate with PM2.5 values, especially in areas of burning or industrial activity. which is the source of fine dust. The use of NO₂, CO, HCHO, and AER AI data on GEE will allow for efficient estimation of PM2.5 at the local level [15][16][17].

IV. RESULT AND DISCUSSION

In this section, the results of PM2.5 analysis using data from the Sentinel-5P satellite on Google Earth Engine (GEE) will be presented. Additionally, an application for visualizing PM2.5 air pollution time-series data will be developed, and the analysis results will be discussed in the context of air quality assessment and interpretation in the 9 lower northern provinces of Thailand. The analysis selected historical data from 2021 to 2024, which is the period during which changes in PM2.5 dust levels in Thailand were observed.

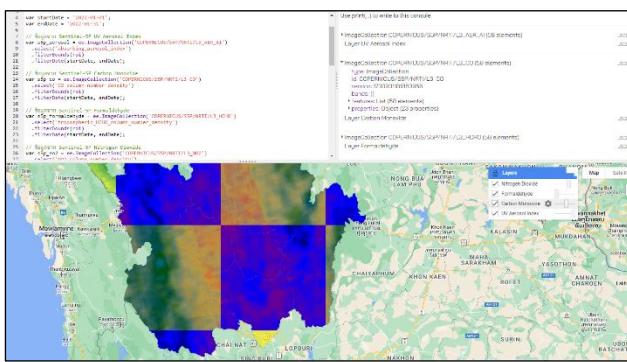


Figure 5 Retrieve data from Sentinel-5P, filter data before deployment.

In data analysis, the average of all data is selected for analysis. One advantage of GEE is that it allows you to code the calculation of averages without the need to create your own formula [18].

A line plot of the daily or monthly average of each indicator can help you see the trend of increasing or decreasing pollution during 2022 and compare higher NO₂ values from burning more fuel.

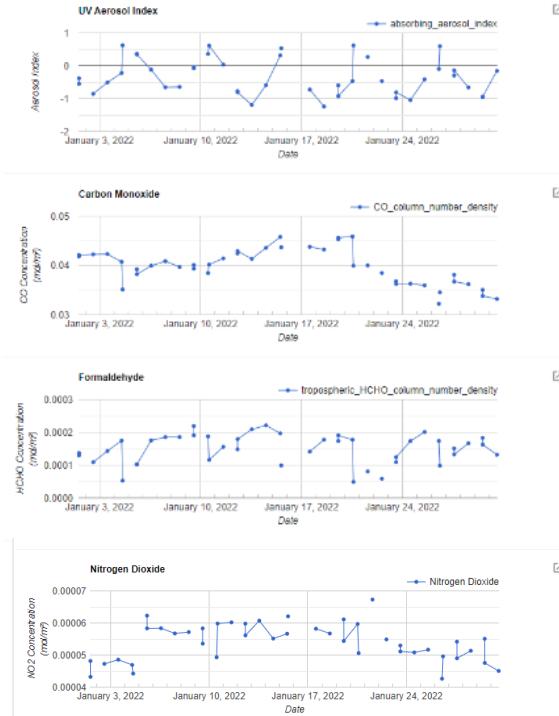


Figure 6 Example of a graph from coding to show the results of the analysis of the four variables.

From the analysis, it was found that the Southeast province, Kamphaeng Phet provinces with higher levels of PM2.5 than other provinces in the lower north. This study indicates that these provinces are more at risk of air pollution than others

Table 1 THE AGRICULTURAL AREA OF THE NINE NORTHERN PROVINCES BELOW.

order	Table Column Head	
	Province	Area / rai
1	Nakhon Sawan	3,070,993
2	Phitsanulok	2,781,826
3	Phetchabun	2,669,207
4	Kamphaeng Phet	2,310,975
5	Uttaradit	1,812,832
6	Sukhothai	1,707,703
7	Phichit	1,451,63
8	Tak	1,327,413
9	Uthai Thani	935,238

. Geographical characteristics: Areas of the province with high accumulation of PM2.5, usually have geographical features that make it easier to accumulate dust, such as being located in a valley or low area, which

reduces the dispersion of dust. Agricultural burns: These provinces may have extensive agricultural areas, which causes burning of foam or other agricultural materials to occur more frequently.

V. CONCLUSION

This research focuses on developing applications that can analyze and visualize time-series data on PM2.5 changes using the Google Earth Engine (GEE) platform and data from the Sentinel-5 Precursor satellite to support air quality management decisions. The developed application will allow users to monitor PM2.5 level changes effectively and be useful in planning air pollution management. The purpose of this research is to develop an application that can display time series data on PM2.5 levels using Sentinel-5P UV Aerosol Index, Sentinel-5P Carbon Monoxide, Sentinel-5P Formaldehyde, and Sentinel-5P Nitrogen Dioxide data. To process PM2.5 data in 9 Lower Northern Provinces, namely Phitsanulok, Kamphaeng Phet, Sukhothai, Uttaradit, Uthai Thani, Phichit, Phetchabun, Tak and Nakhon Sawan. and create application tools: designing and developing easy-to-use and interactive interfaces. the results of the PM2.5 analysis using Sentinel-5P data on Google Earth Engine (GEE) are presented and discussed. This includes the evaluation of the different satellite data layers, their application in assessing PM2.5 levels.

Analysis of Sentinel-5P Data Layers The analysis was conducted using various Sentinel-5P data layers, including,

UV Aerosol Index (AER AI): The UV Aerosol Index provides insights into the presence of aerosols in the atmosphere, which can be associated with the dispersion of fine particulate matter like PM2.5. The AER AI index was used to identify areas with high aerosol concentrations, which correlate with high PM2.5 levels.

Nitrogen Dioxide (NO_2): This data layer is particularly useful for detecting sources of air pollution, particularly from vehicles and industrial activities. High concentrations of NO_2 are often linked to increased levels of PM2.5 in urban areas, making it an important indicator for estimating PM2.5.

Carbon Monoxide (CO): CO is a secondary pollutant that is often produced alongside PM2.5 from combustion processes. This layer was used to detect areas where high combustion activities, such as traffic and industrial operations, may contribute to higher levels of PM2.5.

Formaldehyde (HCHO): Formaldehyde, a byproduct of biomass burning and other industrial processes, was analyzed to identify potential hotspots for PM2.5 in areas affected by open burning, such as agricultural practices.

In conclusion, Sentinel-5P satellite data, particularly the NO_2 , CO, and AER AI layers, proved to be effective tools for assessing PM2.5 levels in the lower northern region of Thailand. This study highlights the potential of using remote sensing data in conjunction with ground-based measurements for better air quality monitoring and management. Further research and data validation are necessary to refine the models and improve the accuracy of PM2.5 estimations.

Analysis of the quantity of PM2.5 or small particle dust with a diameter of less than 2.5 microns in the nine lower northern provinces from 2019 to 2024 revealed important data on the trend of change in air pollution in this region, which is a problem with widespread impact on public health and the environment. Data analysis reveals that 2022 is the year with higher PM2.5 levels than any other year during the study period. Especially in the lower northerly provinces. The study indicates that this year's concentration of PM2.5 is likely to increase significantly compared to the previous or subsequent years. The factors that may influence the increase are Agricultural burns 2022 may have increased burning activity in agricultural areas [19][20][21].

In terms of the design of the application for displaying the results of the analysis on the GEE, a web page consisting of four types of map functions can be created, namely the map showing the UV Aerosol Index, Carbon Monoxide, Formaldehyde, and Nitrogen Dioxide, and displaying the values from the graph to see the effect of changes in the amount of air pollution [22][23], which is a major factor in PM2.5 generation.

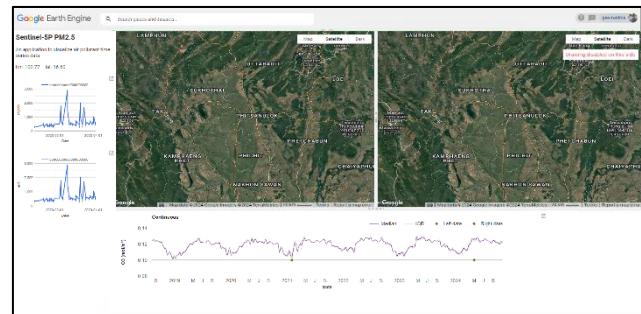


Figure 7 An application to visualize air pollutant and PM2.5 time series data.

Challenges and Limitations Several challenges were encountered during the analysis, including, Spatial Resolution, the spatial resolution of Sentinel-5P data (7x3.5 km) may limit the precision of localized PM2.5 estimates, particularly in highly urbanized areas. Data Gaps, Cloud cover and atmospheric conditions occasionally resulted in data gaps, which can affect the accuracy of the analysis in some areas [24][25]. Despite these challenges, the Sentinel-5P data provided valuable insights into PM2.5 trends and served as an effective tool for large-scale air quality monitoring.

REFERENCES

- [1] World Health Organization (WHO). (2021). "Air Quality Guidelines: Global Update 2021."
- [2] Yang, L., Li, L., & Song, X. (2020). "Satellite-based PM2.5 estimation in Southeast Asia using machine learning approaches." *Remote Sensing of Environment*, 239, 111662.
- [3] Department of Pollution Control, Thailand. (2023). "Annual Report on Air Quality Monitoring."
- [4] Jin, X., & Holloway, T. (2015). "Spatial and temporal variability of PM2.5 sources over Southeast Asia." *Atmospheric Environment*, 120, 90-101.

- [5] Gorelick, N., et al. (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sensing of Environment*, 202, 18-27.
- [6] Aspen Morgan, Jeremy Crowley and Raja M. Nagisetty. (2024). Montana Statewide Google Earth Engine-Based Wildfire Hazardous Particulate (PM2.5) Concentration Estimation. *Air* 2024, 2(2), 142-161; <https://doi.org/10.3390/air2020009>.
- [7] Morgan, A., Crowley, J., & Nagisetty, R. M. (2024). Montana statewide Google Earth Engine-based wildfire hazardous particulate (PM2.5) concentration estimation. *Air*, 2(2), 142-161. <https://doi.org/10.3390/air2020009>.
- [8] Vladimir Tabunschik, Roman Gorbunov and Tatiana Gorbunova. (2022). Unveiling Air Pollution in Crimean Mountain Rivers: Analysis of Sentinel-5 Satellite Images Using Google Earth Engine (GEE). *Remote Sens.* 2023, 15(13), 3364; <https://doi.org/10.3390/rs15133364>.
- [9] Patrick Reddy. (2024). Using TROPOMI Satellite CO and AAI with MODIS Satellite AOD to Screen High Ozone and PM2.5 Concentrations Associated with Wildfire Smoke. LinkedIn article: <https://www.linkedin.com/pulse/using-tropomi-satellite-co-aai-modis-aod-screen-high-ozone-reddy-1cxjc/>.
- [10] Thanh Thi Nhat Nguyen, Hung Quang Bui, Ha Pham Van and Hung Luu. (2015). Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study. *IOP Publishing Environmental Research Letters* 10(9):095016. DOI:10.1088/1748-9326/10/9/095016
- [11] Rabiei-Dastjerdi, H., Mohammadi, S., Saber, M., Amini, S., & McArdle, G. (2022). Spatiotemporal analysis of NO₂ production using TROPOMI time-series images and Google Earth Engine in a Middle Eastern country. *Remote Sensing*, 14(7), 1725. <https://doi.org/10.3390/rs14071725>.
- [12] Bijay Halder, Iman Ahmadianfar, Salim Heddam, Zainab Haider Mussa, Leonardo Goliatto, Mou Leong Tan, Zulfaqar Sa'adi, Zainab Al-Khafaji, Nadhir Al-Ansari, Ali H. Jawad & Zaher Mundher Yaseen. (2023). Machine learning-based country-level annual air pollutants exploration using Sentinel-5P and Google Earth Engine. *Scientific Reports* volume 13, Article number: 7968.
- [13] Chudnovsky, C. Tang, A. Lyapustin, Y. Wang, J. Schwartz, and P. Koutrakis. (2013). A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions. *Atmos. Chem. Phys.*, 13, 10907–10917, <https://doi.org/10.5194/acp-13-10907-2013>, 2013.
- [14] Lin, L., Liang, Y., Liu, L., Zhang, Y., Xie, D., Yin, F., & Ashraf, T. (2022). Estimating PM2.5 concentrations using the machine learning RF-XGBoost model in Guanzhong urban agglomeration, China. *Remote Sens.* 14(20), 5239. <https://doi.org/10.3390/rs14205239>
- [15] Xia Wang, Xiaojie Yao, Changzheng Jiang & Wei Duan. (2022). Dynamic monitoring and analysis of factors influencing ecological environment quality in northern Anhui, China, based on the Google Earth Engine. *Scientific Reports* volume 12, Article number: 20307 (2022).
- [16] de Sherbinin, A., Levy, M. A., Zell, E., Weber, S., & Jaiteh, M. (2014). Using satellite data to develop environmental indicators. *Environmental Research Letters*, 9(8), 084013. <https://doi.org/10.1088/1748-9326/9/8/084013>.
- [17] Feng, Y., Fan, S., Xia, K., & Wang, L. (2022). Estimation of regional ground-level PM2.5 concentrations directly from satellite top-of-atmosphere reflectance using a hybrid learning model. *Remote Sens.* 14(11), 2714. <https://doi.org/10.3390/rs14112714>.
- [18] Md. Nazmul Haque, Md. Shahriar Sharif, Rhyme Rubayet Rudra, Mahdi Mansur Mahi, Md. Jahir Uddin, Radwan G. Abd Ellah. (2022). Analyzing the spatio-temporal directions of air pollutants for the initial wave of Covid-19 epidemic over Bangladesh: Application of satellite imageries and Google Earth Engine. *Remote Sensing Applications: Society and Environment*. Volume 28, November 2022, 100862.
- [19] Sarath Guttikunda and Nishadh KA. (2022). Evolution of India's PM2.5 pollution between 1998 and 2020 using global reanalysis fields coupled with satellite observations and fuel consumption patterns. Received 28th March 2022 , Accepted 22nd September 2022.
- [20] Gupta, P., Zhan, S., Mishra, V., Aekakkararungroj, A., Markert, A., Paibong, S., Chishtie, F. (2021). Machine Learning Algorithm for Estimating Surface PM2.5 in Thailand. *Aerosol Air Qual. Res.* 21, 210105. <https://doi.org/10.4209/aaqr.210105>.
- [21] Hemant Bherwani, Suman Kumar, Kavya Musugu and Moorthy Nair. (2021). Assessment and valuation of health impacts of fine particulate matter during COVID-19 lockdown: a comprehensive study of tropical and sub tropical countries. *Environmental Science and Pollution Research* 28(7). DOI:10.1007/s11356-021-13813-w.
- [22] Jaffe, D. A., Schnieder, B., & Inouye, D. (2022). Technical note: Use of PM2.5 to CO ratio as an indicator of wildfire smoke in urban areas. *Atmospheric Chemistry and Physics*, 22(18), 12695–12704. <https://doi.org/10.5194/acp-22-12695-2022>.
- [23] Sagnik Dey, Bhavesh Purohit, Palak Balyan, Kuldeep Dixit, Kunal Bali, Alok Kumar, Fahad Imam, Sourangsu Chowdhury, Dilip Ganguly, Prashant Gargava andV. K. Shukla. (2020). A Satellite-Based High-Resolution (1-km) Ambient PM2.5 Database for India over Two Decades (2000–2019): Applications for Air Quality Management. *Remote Sens.* 2020, 12(23), 3872; <https://doi.org/10.3390/rs12233872>.
- [24] Park, S., Shin, M., Im, J., Song, C.-K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J., Lee, D.-W., & Kim, S.-K. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. *Atmospheric Chemistry and Physics*, 19(3), 1097-1113. <https://doi.org/10.5194/acp-19-1097-2019>.
- [25] Seohui Park, Minso Shin, Jungho Im and Chang-Keun Song. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. *Atmospheric Chemistry and Physics* 19(2):1097-1113. *Atmospheric Chemistry and Physics* 19(2):1097-1113.

Sex determination in leave of date palm (*Phoenix doactylifera*) on the juvenile vegetative phase between tissue culture and seeding by using LC-QTOF-MS technique

Pranorm Khruawean
*School of Information and Communication Technology
 University of Phayao ao Phayao, Thailand
 pranorm.kh@up.ac.th*

Nakarin Chaikaew
*Research Unit of Spatial Innovation Development (RUSID)
 School of Information and Communication Technology
 University of Phayao Phayao, Thailand
 nakarin.ch@up.ac.th*

Siriwat Boonchaisri
*School of Sciences
 University of Phayao Phayao, Thailand
 siriwat.bo@up.ac.th*

Winai Wiriyalangkorn
*Faculty of Agricultural Production
 Maejo University Chiang Mai, Thailand
 naimiat56@gmail.com*

Achara Kleawkla
*Faculty of Science
 Maejo University Chiang Mai, Thailand
 achara_kleawkla@yahoo.co.uk*

Watcharin Jantawan
*The Office of Agricultural Research and Extension
 Maejo University Chiang Mai, Thailand
 Jantawan16@gmail.com*

Chattakan Jantaput
*Faculty of Science
 Maejo University Chiang Mai, Thailand
 mju6104103309@mju.ac.th*

Ekawit Threenet*
*Faculty of Science
 Maejo University Chiang Mai, Thailand
 Ekawit73@hotmail.com*

*Corresponding author

Abstract The knowledge of identify key metabolites for the purpose of separating female date palm from the male counterparts in the juvenile vegetative phase will help farmers make accurate investment decisions for date palm cultivation to generate the maximum profit. The research aims to develop metabolomics technique to identify both genders by using the high resolution liquid chromatography-quadrupole time-of-flight-mass spectrometer (LC-QTOF-MS) together with multivariate statistical analysis. The study was conducted by extraction of leaves from 2-year old date palms with known sex which were propagated in both tissue culture and conventional seeding in 50%(v/v) methanol. The result of LC-QTOF-MS demonstrated that the date-palm leaves obtained by tissue culture technique could be clearly distinguished from the seed-grown date-palm leaves because the analysis of heatmap patterns clearly showed a reduction of 11 unique metabolites (e.g., metabolites 593.1359, 2370552, 366.9829 Da). This data confirmed that female date palms grown from tissue culture were metabolically different from male date palms. On the other hand, it is very difficult to distinguish the gender of date palms grown from seeds. Only when intensive analysis of metabolite fingerprints was performed, it revealed that dates grown from female seeds differed from males due to the differences in gibberellin and artocarpin as well as an unknown metabolite of 439.0743867 Da.

Keywords—metabolomics, date palm, sex, tissue culture, seeding

I. INTRODUCTION

The date palm (*Phoenix doactylifera*), as classified in the palm family, produces a fruit that is gaining popularity in Thailand. They are alternative fruits for farmers, and are an economically important plant that people all over the world want to consume. In particular, Thailand has a demand no less than other countries. It is originated in the Middle East and North Africa by growing in hot and dry regions similar to deserts (most suitable temperature is 32°C - 40°C) [1]. It is a monocotyledonous plant with imperfect flowers and both

male and female flowers are separated on different trees. However, only female date palms can produce fruits and when grown from seeds, there is an equal chance of getting either male or female date palms. However, farmers can only know the gender of date palms grown from seeds by observing the characteristics of the flowers, which takes 2-3 years or more [2]. Therefore, it is a huge risk to plan an efficient date palm planting. Metabolomics studies are the collection of data and quantities all metabolites synthesized within cells and substances secreted outside cells, which are the products of metabolic pathways [3]. The researchers hypothesized that metabolomics techniques have the potential sex determination of date palm in non-flowering young trees, thus facilitating farmers to plan production that determines the sex ratio of date palms quickly and accurately.

II. METHODOLOGY

A. Sample collection

Date palms were planted in completely randomized design (CRD) from seeding in an agricultural garden at Maejo University, Nong Harn Subdistrict, Amphur San Sai, Chiang Mai Province and from tissue culture in a local orchard at Talat Khwan Subdistrict, Doi Saket, Chiang Mai Province (Fig. 1). The care was given by fertilizing the plants through leaves and watering them every day during the period of weak sunlight from 09.00 to 11.00 hrs. until the date palms were 2 years old. The samples were collected from the youngest mature leaves from both sexes of date palms in equal quantities by cutting the leaves to the desired size until the fresh weight is 100-120 mg/sample. The metabolic process were instantly quenched by placing the samples in 15 ml centrifuge tubes that were frozen by floating in liquid nitrogen before sending them to the laboratory under the frozen condition in liquid nitrogen at all times for storage in a refrigerator at -80 °C awaiting for analysis.

B. Study of metabolite compounds

B 1. Sample preparation

A 30-mg sample of date leaves stored at -80 °C was weighed and 1 mL of 500% methanol solution (including 2% v/v of four internal standards: $^{13}\text{C}_6$ sorbitol, $^{13}\text{C}_{15}\text{N}_5$ valine, aminoanthracene and pentafluorobenzoic acid) was added. Vortexing was performed for 10 min at room temperature and centrifuged at 15,000 rpm for 15 min. The clear solution was collected in 1.5 ml Eppendorf and 100 μL aliquot of this crude extract was placed in a glass insert in a vial for immediate analysis.

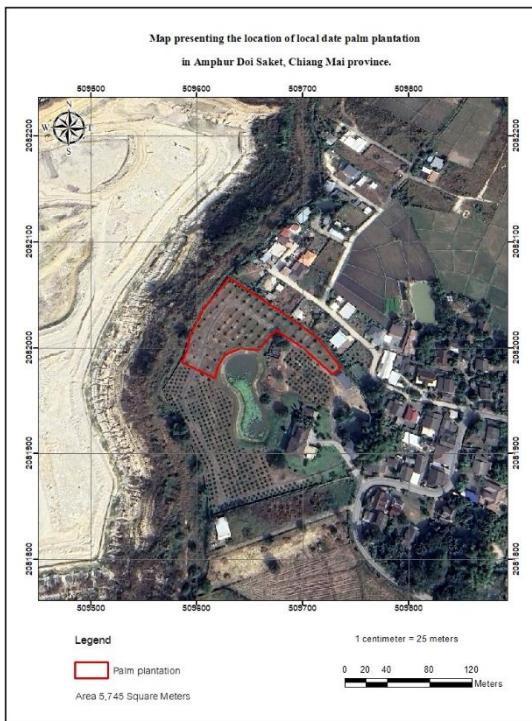


Fig. 1. Location of a planting site of date palms propagated from tissue culture in a local orchard at Talat Khwan Subdistrict, Doi Saket, Chiang Mai Province.

B.2 Analysis of metabolite profile

The samples were analyzed for a global metabolome using LC-QTOF-MS to identify the type and amounts of metabolites. The method followed the previous references [4] where the analysis can be divided into two steps.

Firstly, 5 μL the date palm-leaf crude extract was injected into an Agilent 1290 UHPLC system (Agilent, Santa Clara, CA, USA). The stationary phase conditions was set on a Hypersil GOLD™ C18 Selectivity Column (150 x 2.1 mm, 1.9 μm , Thermo Fisher Scientific Inc.) when a gradient mobile phase were consisted of 0.1% (v:v) formic acid in Milli-Q water (phase A) and 0.1% (v:v) formic acid in acetonitrile (phase B). The initial gradient conditions were started from 2% phase B and the concentration was gradually increased to 100% by linear gradient elution over 15 min. The final concentration (100% phase B) was maintained for 3 min to remove the carry-over metabolites from the column, and then reduced to 2% over 2 min to prepare for the next sample analysis. The sample flow rate was set to 0.3 mL/min and the column temperature was maintained constant at 30 °C

throughout the analysis (total 20 min). The analysis was controlled using Agilent Open Lab software (v2.1.0.433).

The second step was to analyze the mass of the metabolites by high-resolution mass spectrometry. The mass of the metabolites that pass through the LC column were either positively or negatively charged (ionized) by Electrospray ionization (ESI) technique. Then, the mass of all the charged metabolites was analyzed by Quadrupole (Q) mass spectrometry, scanning the m/z from 100-1700 Da to generate the total ion chromatograms. In order to identify the type of metabolites, the high-intensity charged metabolites in each chromatographic peak were further analyzed by fragmentation of the metabolites, using Collision-induced dissociation (CID) technique at 35% energy and 0.25 activation Q with a collision time of 10 msec. in the second Quadrupole spectrometry. The mass of individual fragments was measured again by Time of Flight (TOF) mass spectrometry to generate the second level of mass spectrogram (MS/MS).

B.3 Pre-processing in data

The quality of the mass data was enhanced using the XCMS online platform (<https://xcmsonline.scripps.edu>). Several parameters were optimized to reduce noise, align peaks across samples, group peaks representing the same metabolite, and annotate putative compounds. Additionally, data visualization tools were employed to generate simplified images, aiding in the interpretation of the results

B.4 Data analysis

A comparison of the mass spectrogram of male and female date samples was conducted to select metabolites of interest. Mass peaks that showed clear differences between the two sexes were considered potential biomarkers for sex discrimination. The area under the corresponding peaks was integrated to estimate the quantity of these metabolites, and their masses were searched against public databases such as Metabolomic Workbench, ChemSpider, and PubChem, with a mass accuracy of 0.0002-0.002 m/z to aid in their identification. Manual inspection of differential peaks was conducted using MassHunter Qualitative Analysis software. A comprehensive analysis of metabolite fingerprints was performed using multivariate Partial Least Squares-Discriminant Analysis (PLS-DA), with the intensities of 112 mass peaks as input. The results were visualized in 2D score plots, with further visualization in the form of heatmaps based on Euclidean distance and Ward's linkage.

Due to the high similarity in metabolite fingerprints of seed-grown date palm samples, a pairwise analysis using Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) was conducted. Potential biomarkers were identified based on Variable Importance in Projection (VIP) scores and further validated using Receiver Operating Characteristic (ROC) analysis. Youden's J Statistic was applied to select a threshold of 0.8 for the ROC analysis that minimized false positives and negatives. The optimal cutoff was determined near the top-left corner of the ROC curve, indicating maximum sensitivity and specificity.

III. RESULTS AND DISCUSSION

The total ion chromatograms (TICs) from LC-QTOF-MS analysis of male and female date palms were largely similar in both positive and negative ionization modes. However, distinct differences between the sexes were still observable in the palms propagated from tissue culture technique. For

example, the peak at 9.112 minutes was significantly higher in male date palms compared to their female counterparts (Fig. 2).

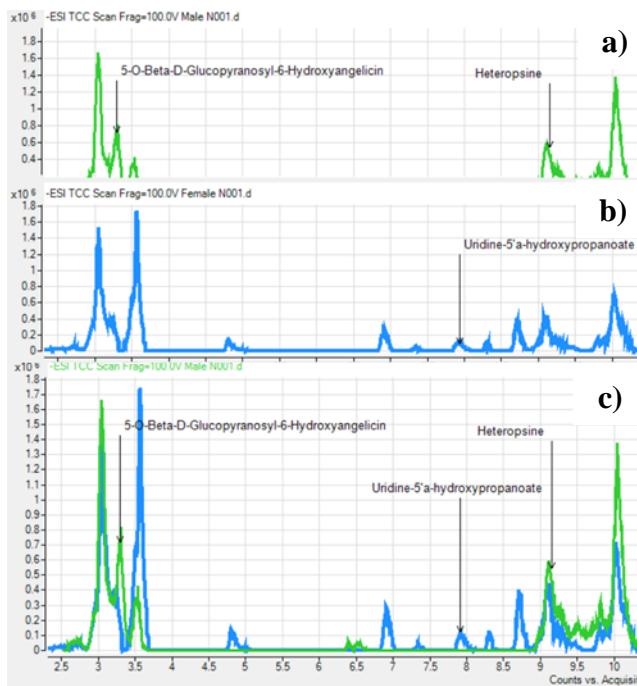


Fig. 2. Negative mode chromatograms of leaf extracts from tissue culture-grown a) male, b) female and c) both male and female date palms.

A manual search identified this compound as heteropsine, naturally occurring alkaloid found *in planta*, known for its inhibitory activity against α -glucosidase, which causes diabetes [5]. The metabolite fingerprints were extracted from the TICs. These fingerprints which consisted of a total 112 metabolites were used as input data to classify male and female date palm samples, by PLS-DA. The analysis revealed that male dates were separated from female dates into three distinct groups: females grown from seed (red), females grown from tissue culture (green), and males grown from tissue culture (light blue). However, male dates grown from seed (purple) exhibited metabolite profiles that overlapped with all other groups (Fig. 3).

The classification of date palm samples was further validated using Hierarchical Clustering Analysis (HCA), visualized in a heatmap (Fig. 4). The comparison of all 112 metabolites (rows) across female date palms grown from seed (red), female date palms grown from tissue culture (green), male date palms grown from seed (blue), and male date palms grown from tissue culture (light blue) successfully grouped the samples into three clusters, consistent with the PLS-DA results.

In the case of date palms grown from tissue culture, analysis of the color scale patterns revealed that a decrease in 11 metabolites (from 593.1359–366.9829 Da) was a key factor in distinguishing female (green) from male (light blue) date palms (Fig. 5).

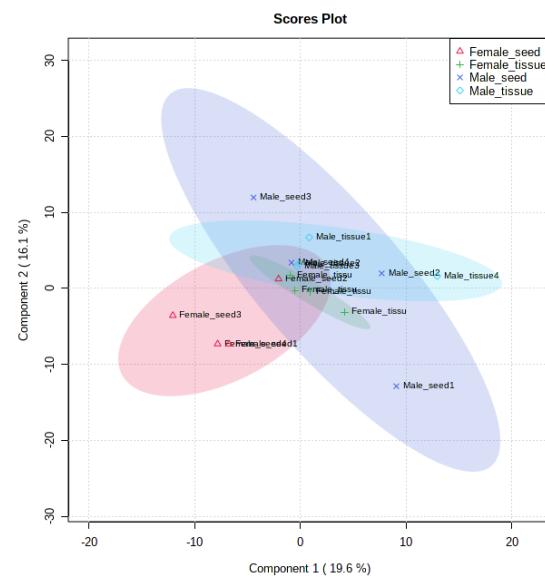


Fig. 3. Score plots of date palm samples clustering into 3 groups using PLS-DA.

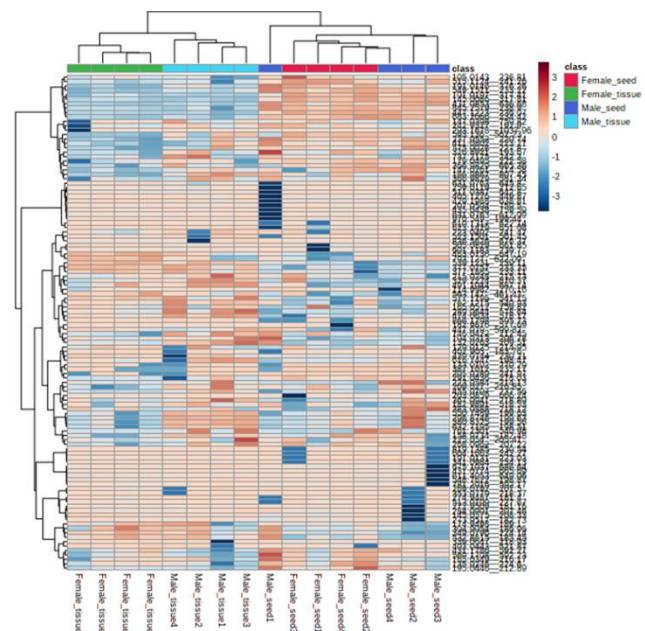


Fig. 4. Heat map showing the clustering of date palm samples using hierarchical clustering analysis (HCA). Note: Each column represents the sample name, each row represents the metabolite type. The color scale indicates the amount of metabolites, where brown = the amount of metabolites in that column sample is higher than the average, blue = the amount of metabolites in that column sample is lower than the average.

On the other hand, the analyses confirmed that sex differentiation in seed-grown date palms was not possible using these methods. This suggests that identifying the sex of date palms grown from tissue culture is easier than for those grown from seeds. One possible explanation is that the tissue-cultured date palms used in this study were derived from several-year-old tissues, having already entered the adult

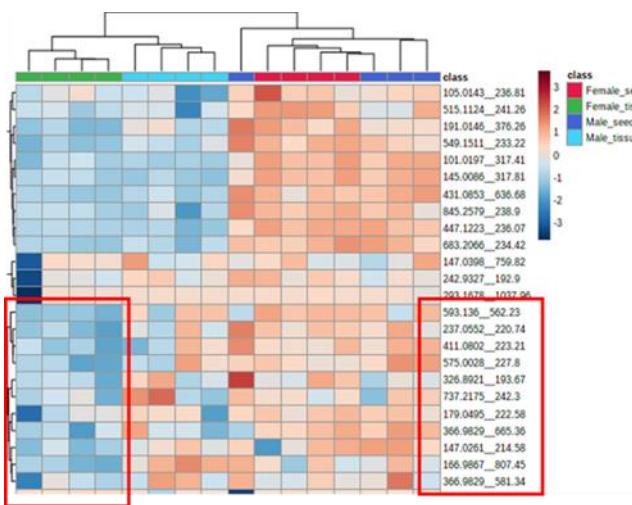


Fig. 5. Enlarged section of the heat map presented in Fig. 4, highlighting the substance groups contributing to the differentiation of male and female date palms.

Note: Includes only samples derived from tissue culture.

vegetative phase and been induced to produce flowers [6, 7, 8]. Consequently, these palms may contain sex-related metabolites, which are downstream products of specific genes, such as those matching a male-specific sequence characterized amplified region (SCAR) marker [9].

In contrast, seed-grown date palms may be too young to express any sex-specific metabolites, resulting in both sexes exhibiting a similar global metabolome. However, when metabolite fingerprints were analyzed using OPLS-DA, the model was able to distinguish the two sex groups (Fig. 6).

The important metabolites that affect the discrimination of male and female date palms grown from seeds were extracted by calculating the Variable Importance Projection (VIP) scores from the OPLS-DA model. The results are the metabolites with molecular weights of 383.0887, 439.0743, and 515.1124 Da, respectively (Fig. 7). Specifically, the metabolite at 383.0887 Da, one of the top three compounds with the highest VIP scores ($VIP = 1.0$), can accurately differentiate male from female date palms grown from seed, with no false positives. This result aligned with the Box plot of raw data, which showed a significant accumulation of the 383.0887 Da metabolite in male date palms (Fig. 8). A comprehensive database search putatively identified the 383.0887 Da and 515.1124 Da metabolites as gibberellin, a phytohormone that promotes plant elongation, and artocarpin, an isoprenyl flavonoid that induced apoptosis to kill lung cancer [10], respectively. However, the metabolite at 439.0743 Da remains unidentified.

IV. CONCLUSION

In conclusion, sex differentiation in date palms grown from tissue culture was easier than in those grown from seeds. The reduction of 11 metabolites (e.g., 593.136 Da, 366.9829 Da) played a key role in distinguishing female date palms grown from tissue culture from their male counterparts. Although sex differentiation in seed-grown date palms was more challenging, intensive metabolite fingerprint analysis revealed that female date palms grown from seeds could be distinguished from males due to the accumulation of gibberellin

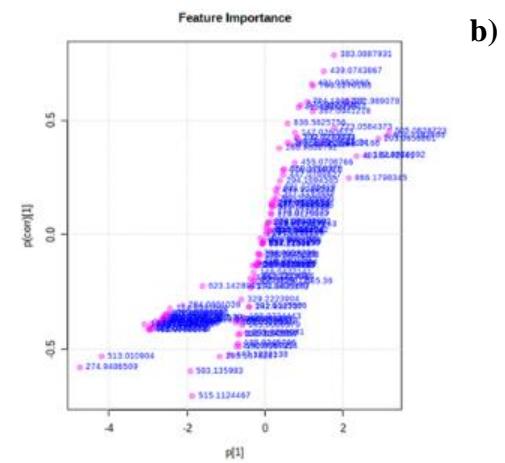
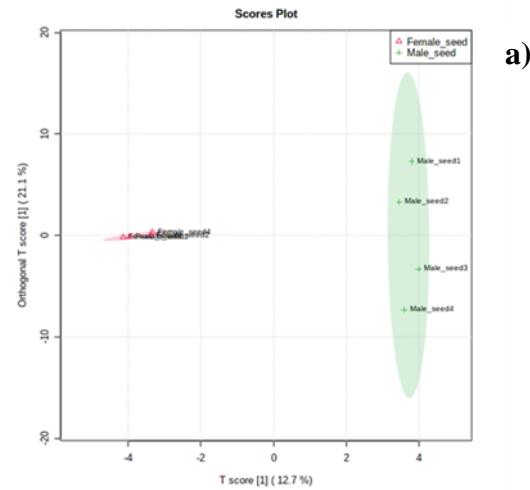


Fig. 6. Orthogonal Partial Least Square-Discriminant Analysis (OPLS-DA) of date palm from seed grown cultivation: a) scores plot and b) loading plot.

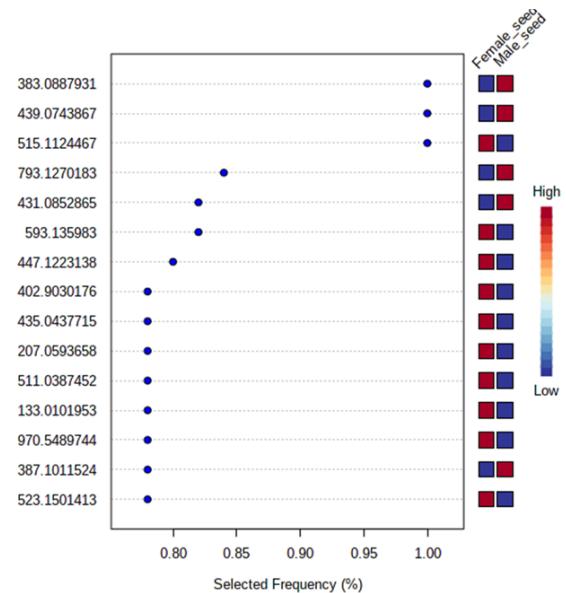


Fig. 7. Calculation of Variable Importance in Projection (VIP) scores from the OPLS-DA model (Fig. 6)

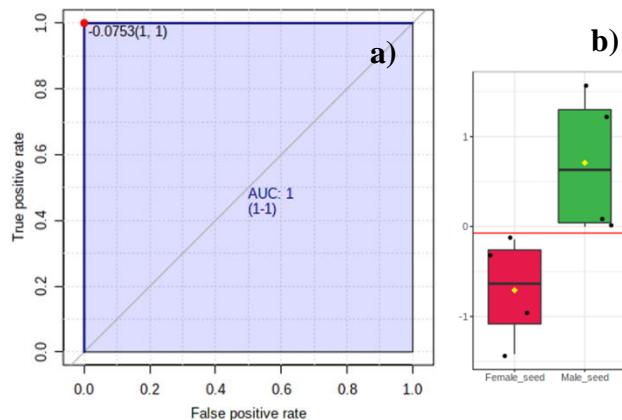


Fig. 8. Analysis of the metabolite at 383.0887 Da using a) a receiver operating characteristic (ROC) curve and b) Box plot of raw data.

artocarpin, and an unidentified metabolite with the mass of 439.0743 Da. These findings highlight the potential of metabolomic profiling as a reliable tool for early and accurate sex determination in date palms, irrespective of their propagation method. Future research could focus on expanding the database of sex-related metabolites and optimizing detection methods for large-scale agricultural applications.

ACKNOWLEDGMENT

This study was financially supported by the Thailand Science Research and Innovation (TSRI), National Science Research and Innovation Fund (NSRF), Fundamental Fund (FF) in 2022 as well as Young Scientist Fund from School of Sciences, University of Phayao. We also thank Mrs. Pimporn Manochai for her kind help in facility at Division of Chemistry, Faculty of Science, Maejo University, Chiang Mai.

REFERENCES

- [1] R.R. Krueger, "Date palm (*Phoenix dactylifera* L.) biology and utilization. In: The Date Palm Genome, Vol. 1: Phylogeny, Biodiversity and Mapping," Cham: Springer International Publishing, 2021. pp. 3-28.
- [2] S.A. Bekheet and M.S. Hanafy, "Towards sex determination of date palm," Date palm biotechnology, pp. 551-566, 2011.
- [3] F. Danzi, R. Pacchiana, A. Mafficini, M. Scupoli, A. Scarpa, M. Donadelli and A. Fiore, "To metabolomics and beyond: a technological portfolio to investigate cancer metabolism," Signal Transduction and Targeted Therapy, 8(1), 137, 2023.
- [4] S. Boonchaisri, S. Rochfort, T. Stevenson and D. Dias, "LC-MS untargeted metabolomics assesses the delayed response of glufosinate treatment of transgenic glufosinate resistant (GR) buffalo grasses (*Stenotaphrum secundatum* L.)," Metabolomics, vol. 17, pp.1-17, 2021.
- [5] P. Teerapongpisan, V. Suthiphasilp, P. Phukhatmuen, N. Rujanapun, B. Chaiyosang, S. Tontapha, T. Maneerat, B. O. Patrick, T. Duangyod, R. Charoensup, R. J. Andersen, and S. Laphookhieo, "Dimeric aporphine alkaloids from the twigs of *Trivalvaria costata* (Hook.f. & Thomson) I.M.Turner," Phytochem., vol. 207, p. 113586, 2023. doi: 10.1016/j.phytochem.2023.113586.
- [6] A.R. Sheldrake, "The ageing, growth and death of cells," Nature, vol. 250(5465), pp. 381-385, 1974.
- [7] H. Thomas, "Ageing in plants. Mechanisms of ageing and development," vol. 123(7), 747-753, 2002.
- [8] H. Thomas, "Senescence, ageing and death of the whole plant. New Phytologist," vol. 197(3), 696-711, 2013.
- [9] S. A. Naqvi, W. Shafqat, M. S. Haider, F. S. Awan, I. A. Khan and M. J. Jaskani, Gender determination of date palm. In The Date Palm Genome, Vol. 1: Phylogeny, Biodiversity and Mapping. Cham: Springer International Publishing, 2021, pp. 161-177.
- [10] M. H. Tsai, J. F. Liu, Y. C. Chiang, S. C. Hu, L. F. Hsu, Y. C. Lin, Z. C. Lin, H. C. Lee, M. C. Chen, C. L. Huang, and C. W. Lee, "Artocarpin, an isoprenyl flavonoid, induces p53-dependent or independent apoptosis via ROS-mediated MAPKs and Akt activation in non-small cell lung cancer cells," Oncotarget, vol. 8, no. 17, pp. 28342-28358, 2017, doi: 10.18632/oncotarget.16058.

Land Use Changes Prediction in the Phra Phimon Irrigation Scheme using Multi-temporal Landsat Imagery and CA-ANN Model

Kasidet Sirisakorn
Department of Irrigation Engineering
Kasetsart University
Thailand
Kasidet.sir@ku.ac.th

Ketvara Sittichok
Department of Irrigation Engineering
Kasetsart University
Thailand
fengkrs@ku.ac.th

Chuphan Chompuchan
Department of Irrigation Engineering
Kasetsart University
Thailand
fengcpcc@ku.ac.th

Abstract—This study investigates land use changes in the Phra Phimon Operation and Maintenance Project, a large irrigation scheme near Bangkok, Thailand where rapid urbanization has directly contributed to significant changes in agricultural land use and cultivation practices. Landsat satellite imagery from 1993, 2006, and 2019 was used to classify land use into four categories: paddy fields, orchards, urban areas, and water bodies. Future land use changes for 2032 and 2045 were predicted using the CA-Markov couple with ANN model (CA-ANN), implemented through QGIS's MOLUSCE plugin. Land use classification achieved 90.95% accuracy (F1-score: 0.91), and prediction accuracy was 67.84% (F1-score: 0.67). Results show significant urban expansion, with urban areas projected to increase from 23.65% in 2019 to 34.87% in 2045, while agricultural areas decrease from 73.28% to 62.00%. This urban growth may result in agricultural land fragmentation, decreased water distribution efficiency, and potential irrigation canal deterioration. This research provides insights for irrigation scheme managers to prepare for future changes in water requirements and infrastructure management.

Keywords—Land use change; Satellite imagery; CA-ANN model

I. INTRODUCTION

Land use and land cover are subject to continuous changes due to both natural processes and human activities. Factors including urban expansion, industrial development, and changes in agricultural practices contribute to rapid shifts in land use patterns [1]. These changes significantly affect ecosystems and agricultural systems, particularly impacting water resources. Irrigated areas, which are essential for global food production, are especially sensitive to these transformations. As urbanization and farming practices change, water demand and distribution patterns also shift, requiring adaptive management approaches [2]. Understanding and predicting land use changes is therefore crucial for effective agricultural water allocation and resource management. Accurate forecasts of future land use patterns can guide policymakers, urban planners, and irrigation managers in developing strategies that balance urban development with agricultural sustainability and efficient water use [3],[4].

The Phra Phimon Operation and Maintenance Project (PP Project), a large irrigation scheme near Bangkok, Thailand, shows similar land use changes to those observed globally. This area has experienced significant land use changes over the years, reflecting trends seen in many developing regions. Recent data from the Land Development Department indicates a substantial expansion of urban areas within the PP Project. This urban growth has directly contributed to changes in agricultural land use and cultivation practices. Consequently, the resulting shifts in agricultural land and farming activities have created challenges for irrigation water

planning and management. Predicting future land use changes is essential in this context. Research by [5] modeled urban expansion and showed that urban areas tend to replace agricultural and forested lands. According to [6] found that an increase of 1% in urbanization was associated with a reduction in agricultural water consumption of 0.639 million cubic meters relative to the baseline year. This emphasizes the importance of sustainable land use planning.

The CA-Markov model is an effective method for predicting future land use changes, integrating cellular automata (CA) and Markov chains. This approach has been successfully applied in various contexts, including crop loss warning [7], sustainable land-use development planning [8], and land and water resource management [9]. This study analyzes historical land use changes from 1993 to 2019 and applies the CA-Markov model to predict future land use for 2032 and 2045. The results aim to help irrigation scheme managers understand and prepare for potential future land use changes in the PP Project area.

II. MATERIAL AND METHOD

A. Study Area

The Phra Phimon Operation and Maintenance Project, encompassing an area of 47,384.82 hectares, is situated in the central region of Thailand, west of the capital, Bangkok (Figure 1). The project area features flat terrain in the lower western plain of the Chao Phraya River. A network of irrigation canals connects the Tha Chin and Chao Phraya rivers, which serve as the major water sources for this inundation-type irrigation scheme. This system plays a crucial role in supporting agricultural production in the region. The majority of the cultivated area is used for rice paddies and orchards.

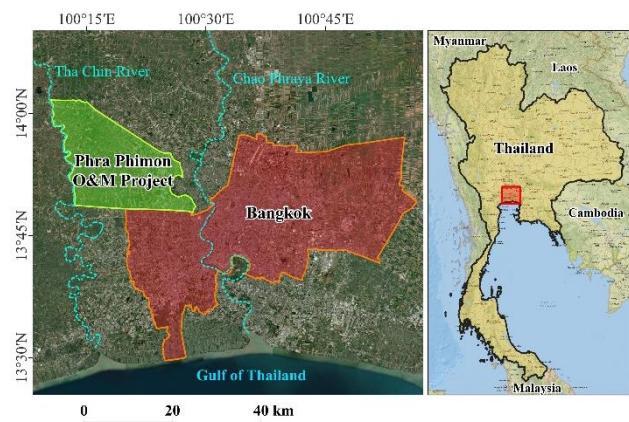


Fig. 1. Map of study area

B. Satellite Imagery Classification

This study uses Landsat Collection 2 Level-2 satellite data from 1993, 2006, and 2019 (Table I). The data covers path 129, row 050, with a 30-meter spatial resolution. Multiple images were selected for each year due to the dynamic nature of crop phenology, particularly for rice cultivation. Rice paddies experience significant changes throughout their growth cycle, transitioning between water-covered, vegetated, and bare land states. Using a single satellite image for each year could lead to misinterpretation of land use patterns [10]. Therefore, multiple images were used to accurately represent the dynamic land cover changes and minimize potential errors in land use classification.

TABLE I. SUMMARY OF SATELLITE IMAGERY USED IN THIS STUDY

Period	Satellite	Date of Acquisition
1993	Landsat-5 TM	07 November 1993
	Landsat-5 TM	25 December 1993
	Landsat-5 TM	26 January 1994
	Landsat-5 TM	31 March 1994
2006	Landsat-5 TM	27 November 2006
	Landsat-5 TM	29 December 2006
	Landsat-5 TM	14 January 2007
	Landsat-5 TM	15 February 2007
2019	Landsat-8 OLI	15 November 2019
	Landsat-8 OLI	17 December 2019
	Landsat-8 OLI	02 January 2020
	Landsat-8 OLI	19 February 2020

This study used three spectral indices: (1) Normalized Difference Vegetation Index (NDVI) to assess vegetation cover [11], (2) Normalized Difference Water Index (NDWI) to detect water bodies [12], and (3) Normalized Difference Built-up Index (NDBI) to quantify urban development [13]. These indices were calculated using the following formulas:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \quad (2)$$

$$NDBI = \frac{\rho_{SWIR1} - \rho_{NIR}}{\rho_{SWIR1} + \rho_{NIR}} \quad (3)$$

where, ρ_{Green} , ρ_{Red} , ρ_{NIR} and ρ_{SWIR1} are surface reflectance for band 2,3,4 and 5 in Landsat-5 TM, respectively and band 3,4,5 and 6 in Landsat-8 OLI, respectively.

The interpretation of Land Use/Land Cover (LULC) utilized spectral signature analysis from multitemporal satellite imagery, based on three indices: NDVI, NDWI, and NDBI [14]. These multitemporal spectral index values were transformed into spectral curves, representing index values across various bands over time. These curves were used to identify representative areas within each land use category— paddy fields, orchards, urban areas, and water bodies. Each of these representative areas exhibited distinct spectral signatures, with clear differences in both the magnitude of spectral index values and their temporal patterns (Fig. 2). These unique characteristics allowed for accurate identification of each land use category. The Spectral Angle Mapping (SAM) technique, which uses an n-dimensional angle to match pixels to reference spectra [15], was applied to classify the entire image into these four LULC categories.

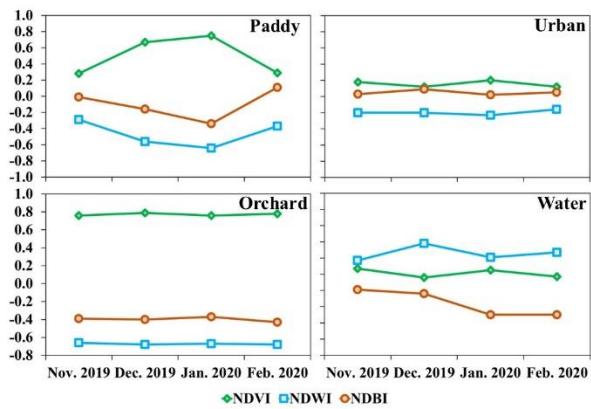


Fig. 2. Spectral curve of multitemporal indices values of representative areas for land use classification.

C. The CA-Markov couple with ANN Model (CA-ANN)

The CA-Markov model combines Cellular Automata (CA) and Markov chain analysis. CA simulates spatial changes based on neighboring cell states, while Markov chain analysis calculates transition probabilities between land use classes. This study used the Modules of Land Use Change Evaluation (MOLUSCE) plugin in QGIS to forecast future LULC changes. MOLUSCE, which uses artificial neural networks (ANN), provides a reliable method for predicting LULC changes and was employed in this study for modeling purposes. It can be an effective tool for land use planning and management [16].

Three driving factors were used to predict LULC changes: distance from built-up areas, distance from irrigation canals, and population density (Fig 3). These socioeconomic factors influence land use change patterns distinctly: distance from built-up areas represents urban expansion patterns, where adjacent areas are likely to be developed due to existing infrastructure [17]; areas near irrigation canals tend to remain agricultural due to reliable water access [18]; and population density indicates potential new urban centers, where high-density areas may develop independently of existing built-up areas [19]. These factors were chosen based on previous research [20],[21] and local relevance. The distance factors (from built-up areas and irrigation canals) were calculated using a Proximity algorithm in QGIS. Data for built-up areas came from the Google Open Building dataset, while irrigation canal data was provided by the Royal Irrigation Department. Population density data was sourced from the Gridded Population of the World (GPW) dataset. All factors were resampled to 30x30 pixel resolution to align with the satellite imagery resolution.

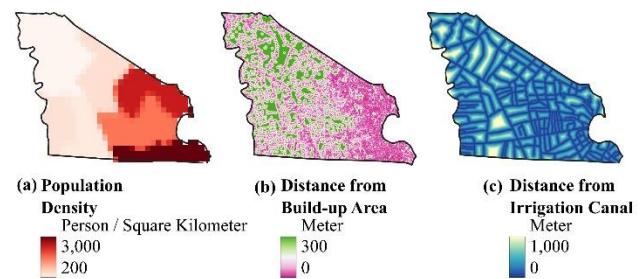


Fig. 3. Geographic and socioeconomic factors (a) Population density, (b) Distance from building and (c) Distance from irrigation canal.

Classified land use data from satellite imagery for 1993, 2006, and 2019 were used in this study. The 1993 and 2006 data were used to predict 2019 land use, which was then validated against the actual 2019 classified data. After validation, the 2006 and 2019 data were used to predict future land use for 2032 and 2045. (Fig. 4).

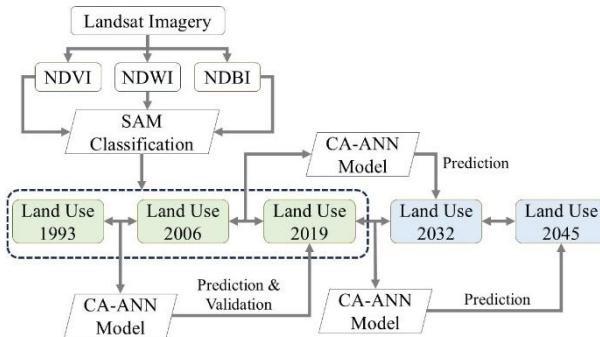


Fig. 4. Flow chart of study

D. Accuracy Assessment

The 2019 land use classification accuracy assessment was conducted using Google Street View data from 199 random sample sites (Fig. 5), which efficiently provides reliable ground reference data covering large areas at low cost [22]. Historical land use classifications (1993 and 2006) could not be validated as historical reference data was limited, which is consistent with other studies [23]. Overall Accuracy (OA) and weighted F1-score were used as evaluation metrics. Both metrics are commonly used to assess the performance of satellite image classification systems [24]. The Overall Accuracy (OA) is calculated as:

$$OA_i = \frac{\sum_{i=1}^k a_{ii}}{n} \quad (4)$$

where k represents the number of rows in the matrix, where each row corresponds to a unique type of land cover a_{ii} represents the number of observations found at the intersection of the i row and the j column. n represents the total sample size.

$$F1 - score_i = \frac{2 \times \frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \quad (5)$$

where True Positives (TP) represent instances correctly classified as positive, while False Positives (FP) denote instances incorrectly classified as positive. False Negatives (FN) refer to instances incorrectly classified as negative. The equation for the weighted F1-score using the number of samples in each class as weights as equation (6)

$$Weighted F1 - score_i = \frac{\sum_{i=1}^N (F1 - score_i \times N_i)}{\sum_{i=1}^N N_i} \quad (6)$$

where N_i represents the sample size of each category i , and N represents the total sample size across all categories.

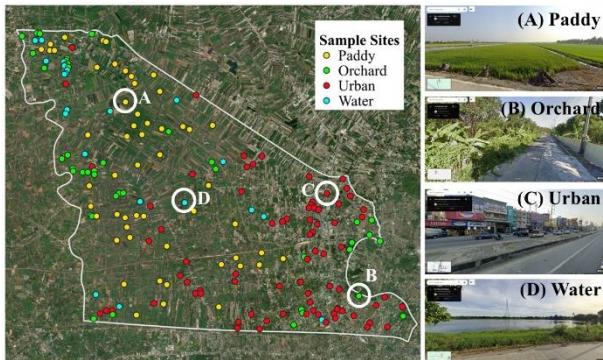


Fig. 5. Distribution of ground truth sample sites in the PP Project area with Google Street View.

III. RESULT AND DISCUSSION

A. LULC Classification and CA-ANN Model Validation

The 2019 Land use classification using the SAM algorithm, validated with Google Street View data, achieved 90.95% overall accuracy and a 0.91 weighted F1-score (Table II). These results demonstrate accurate classification performance, offering a solid basis for further analyses. The acquired results were utilized as reference data to predict future land use changes. This predictive model was trained using historical land use data from 1993 and 2006 to predict 2019 land use. The validated result against classified 2019 land use reached an overall accuracy of 67.84% and a weighted F1-score of 0.67. This accuracy level is consistent with other LULC prediction studies using CA-ANN models: 65.80% [25], 76.28% [16], 70% [26], and 76.23% [27], indicating good consistency between predicted and actual LULC situations. These studies suggest that accuracy levels between 65-76% are typical and acceptable for LULC prediction models, considering the complexity of urban growth patterns and socioeconomic drivers influencing land use changes.

TABLE II. ERROR MATRIX OF THE 2019 LAND USE CLASSIFICATION.

Classified Data	Reference Data					User Acc. (%)
	Paddy	Orchard	Urban	Water	Total	
Paddy	50	8	3	0	61	81.97
Orchard	3	35	1	0	39	89.74
Urban	1	0	75	2	78	96.15
Water	0	0	0	21	21	100.00
Total	54	43	79	23	199	
Producer Acc. (%)	92.59	81.40	94.94	91.30		
Overall Accuracy = 90.95%, Weighted F1-score = 0.91						

B. LULC Change Prediction

Land use changes were projected for 2032 and 2045, as depicted in Fig. 6 and Table III. These spatial data maps, covering the PP Project area from 1993 to 2045, revealed a noticeably increase in urban areas. Urban areas are projected to expand from 23.65% in 2019 to 32.00% in 2032 and 34.87% in 2045. Conversely, agricultural areas are expected to decrease from 73.28% in 2019 to 64.78% in 2032 and 62.00% in 2045. Population density was identified as a key driver of these changes, with residential areas concentrating on the project's eastern side, adjacent to the capital city. Orchard land has declined more significantly than other land use types, due to its initial concentration in the project's eastern region, which has been subjected to urban expansion. These results align with the findings of [28], showing that urban expansion directly results in the conversion of agricultural land into urban areas. This study indicates that agricultural land is projected to be the most affected by urban growth.

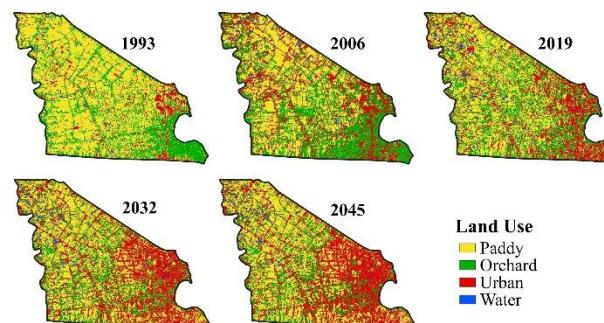


Fig. 6. LULC maps of PP Project for 1993, 2006, 2019, 2032 and 2045.

TABLE III. CHANGES IN LULC CLASSES BETWEEN THE YEARS OF 1993-2045 (AREA UNIT: HA).

Year	LU Classes				
	Paddy	Orchard	Urban	Water	
1993	Area	28,977	13,800	3,872	735
	%	61.15%	29.13%	8.17%	1.55%
2006	Area	18,909	16,502	11,037	936
	%	39.91%	34.83%	23.29%	1.97%
2019	Area	21,026	13,697	11,205	1,456
	%	44.37%	28.91%	23.65%	3.07%
2032	Area	20,304	10,391	15,165	1,526
	%	42.85%	21.93%	32.00%	3.22%
2045	Area	19,480	9,900	16,522	1,483
	%	41.11%	20.89%	34.87%	3.13%

The projected decrease in agricultural land is expected to reduce overall irrigation water demand. However, this reduction may not be proportional to the land use change due to factors such as increased water demand from remaining crops, climate change, and shifts in cropping patterns [29]. Urban expansion, particularly in the eastern project area, may fragment agricultural land and decrease water distribution efficiency [30]. This may require the irrigation scheme to become more flexible to accommodate changing water demand dynamics. Urban growth can also complicate access to and maintenance of irrigation infrastructure, potentially accelerating canal deterioration. As canal efficiency declines, reduced water supply could lead to agricultural abandonment, further accelerating urbanization [31]. Additionally, increased urbanization poses risks to irrigation water quality. Urban runoff may introduce pollutants like oils, heavy metals, and nutrients into the canal system [32]. These potential issues in the irrigation scheme area present significant challenges for policymakers and irrigation scheme managers. To address these challenges, adaptive management strategies are crucial. These could include implementing smart irrigation systems that adjust water delivery based on real-time crop needs [33], upgrading canal infrastructure to improve water distribution efficiency, and creating riparian buffer zones between urban and agricultural areas to minimize pollution [34].

Integration of additional socioeconomic variables such as GDP, land prices, and infrastructure development would enhance future analyses. The eastern region demonstrated accelerated urbanization patterns influenced by multiple growth factors beyond the model parameters. Future studies should consider incorporating these variables to improve model performance in rapidly urbanizing areas.

IV. CONCLUSION AND SUGGESTION

This study analyzed land use changes in the Phra Phimon Operation and Maintenance Project area using multitemporal satellite imagery and the CA-ANN model. The land use classification achieved high accuracy (90.95%), providing a reliable basis for future projections. The CA-ANN model predicted significant urban expansion from 23.65% in 2019 to 34.87% in 2045, primarily at the expense of agricultural land, which is expected to decrease from 73.28% to 62.00% in the same period. These projections highlight potential challenges for irrigation management, including reduced water distribution efficiency, increased maintenance difficulties, and water quality risks due to urban runoff. To address these challenges, adaptive management strategies such as smart irrigation systems, infrastructure upgrades, and riparian buffer zone creation are recommended.

ACKNOWLEDGE

This research was funded by the Faculty of Engineering at Kamphaeng Saen, Kasetsart University (Kamphaeng Saen Campus), Thailand. The authors would like to thank Google for providing the Open Building Dataset, also thank the Royal Irrigation Department, NASA Socioeconomic Data and Applications Center (SEDAC), and USGS Earth Explorer for their valuable support and data contributions to this study.

REFERENCES

- [1] A. Raza et al., "Impact of Land Use and Land Cover Change on Agricultural Production in District Bahawalnagar, Pakistan," *Environmental Sciences Proceedings*, vol. 29, no. 1, p. 46, 2024.
- [2] M. H. Giacomo, L. Kanta, and M. E. Zechman, "Complex Adaptive Systems Approach to Simulate the Sustainability of Water Resources and Urbanization," *Journal of Water Resources Planning and Management*, vol. 139, no. 5, pp. 554–564, 2013.
- [3] A. S. Mahiny and K. C. Clarke, "Simulating hydrologic impacts of urban growth using SLEUTH, multi criteria evaluation and runoff modeling," *Journal of Environmental Informatics*, vol. 22, no. 1, pp. 27–38, 2013.
- [4] R. S. Gragg, A. Anandhi, M. Jiru, and K. M. Usher, "A conceptualization of the urban food-energy-water nexus sustainability paradigm: Modeling from theory to practice," *Frontiers in Environmental Science*, vol. 6, no. 133, pp. 1–14, 2018.
- [5] N. Mozaffaree Pour and T. Oja, "Urban Expansion Simulated by Integrated Cellular Automata and Agent-Based Models; An Example of Tallinn, Estonia," *Urban Science*, vol. 5, no. 4, p. 85, 2021.
- [6] S. Avazdahandeh and S. Khalilian, "The effect of urbanization on agricultural water consumption and production: the extended positive mathematical programming approach," *Environmental Geochemistry and Health*, vol. 43, no. 1, pp. 247–258, 2021.
- [7] M. Beroho et al., "Future Scenarios of Land Use/Land Cover (LULC) Based on a CA-Markov Simulation Model: Case of a Mediterranean Watershed in Morocco," *Remote Sensing*, vol. 15, no. 4, p. 1162, 2023.
- [8] M. R. Rahnama, "Forecasting land-use changes in Mashhad Metropolitan area using Cellular Automata and Markov chain model for 2016–2030," *Sustainable Cities and Society*, vol. 64, p. 102548, 2021.
- [9] G. F. Arfasa, E. Owusu-Sekyere, and D. A. Doke, "Predictions of land use/land cover change, drivers, and their implications on water availability for irrigation in the Vea catchment, Ghana," *Geocarto International*, vol. 38, no. 1, p. 2243093, 2023.
- [10] X.-Z. Pan, S. Uchida, Y. Liang, A. Hirano, and B. Sun, "Discriminating different landuse types by using multitemporal NDVI in a rice planting area," *International Journal of Remote Sensing*, vol. 31, no. 3, pp. 585–596, 2010.
- [11] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS," in *3rd Earth Resources Technology Satellite-1 Symposium*, 1974, pp. 309–317.
- [12] S. K. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features," *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, 1996.
- [13] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583–594, 2003.
- [14] M. Kim et al., "Convolutional Neural Network-Based Land Cover Classification Using 2-D Spectral Reflectance Curve Graphs with Multitemporal Satellite Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4604–4617, 2018.
- [15] T. Jiang, H. van der Werff, and F. van der Meer, "Classification endmember selection with multi-temporal hyperspectral data," *Remote Sensing*, vol. 12, no. 10, p. 1575, 2020.
- [16] M. Kamaraj and S. Rangarajan, "Predicting the future land use and land cover changes for Bhavani basin, Tamil Nadu, India, using QGIS MOLUSCE plugin," *Environmental Science and Pollution Research*, vol. 29, no. 57, pp. 86337–86348, 2022.
- [17] M. Salem et al., "Urban expansion simulation based on various driving factors using a logistic regression model: Delhi as a case study," *Sustainability*, vol. 13, no. 19, pp. 1–17, 2021.
- [18] D. Blakeslee, A. Dar, R. Fishman, S. Malik, H. S. Pellegrina, and K. S. Bagavathinathan, "Irrigation and the spatial pattern of local economic development in India," *Journal of Development Economics*, vol. 161, p. 102997, 2023.
- [19] Z. Li, L. Jiao, B. Zhang, G. Xu, and J. Liu, "Understanding the pattern and mechanism of spatial concentration of urban land use, population and economic activities: a case study in Wuhan, China," *Geo-Spatial Information Science*, vol. 24, no. 4, pp. 678–694, 2021.
- [20] S. Mansour, M. Al-Belushi, and T. Al-Awadhi, "Monitoring land use and land cover changes in the mountainous cities of Oman using GIS

- and CA-Markov modelling techniques,” *Land Use Policy*, vol. 91, p. 104414, 2020.
- [21] B. Matlhodi, P. K. Kenabatho, B. P. Parida, and J. G. Maphanyane, “Analysis of the future land use land cover changes in the gaborone dam catchment using ca-markov model: Implications on water resources,” *Remote Sensing*, vol. 13, no. 13, p. 2427, 2021.
- [22] Y. Yan and Y. Ryu, “Exploring Google Street View with deep learning for crop type mapping,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 278–296, 2021.
- [23] M. T. U. Rahman et al., “Temporal dynamics of land use/land cover change and its prediction using CA-ANN model for southwestern coastal Bangladesh,” *Environmental Monitoring and Assessment*, vol. 189, no. 11, p. 565, 2017.
- [24] N. M. Enwright et al., “Modeling barrier island habitats using landscape position information,” *Remote Sensing*, vol. 11, no. 8, p. 907, 2019.
- [25] R. Muhammad, W. Zhang, Z. Abbas, F. Guo, and L. Gwiazdzinski, “Spatiotemporal Change Analysis and Prediction of Future Land Use and Land Cover Changes Using QGIS MOLUSCE Plugin and Remote Sensing Big Data: A Case Study of Linyi, China,” *Land*, vol. 11, no. 3, p. 419, 2022.
- [26] S. Ullah et al., “Remote sensing-based quantification of the relationships between land use land cover changes and surface temperature over the lower Himalayan region,” *Sustainability*, vol. 11, no. 19, p. 5492, 2019.
- [27] B. Aneesa Satya, M. Shashi, and P. Deva, “Future land use land cover scenario simulation using open source GIS for the city of Warangal, Telangana, India,” *Applied Geomatics*, vol. 12, no. 3, pp. 281–290, 2020.
- [28] F. Martellozzo, F. Amato, B. Murgante, and K. C. Clarke, “Modelling the impact of urban growth on agriculture and natural land in Italy to 2030,” *Applied Geography*, vol. 91, pp. 156–167, 2018.
- [29] S. Sun, T. Zhou, P. Wu, Y. Wang, X. Zhao, and Y. Yin, “Impacts of future climate and agricultural land-use changes on regional agricultural water use in a large irrigation district of northwest China,” *Land Degradation & Development*, vol. 30, no. 10, pp. 1158–1171, 2019.
- [30] G. Bonaiti and G. Fipps, “Methodology for quantifying impacts of urbanization on the water distribution networks of irrigation districts,” *Agricultural Water Management*, vol. 176, pp. 270–279, 2016.
- [31] V. Davivongs, M. Yokohari, and Y. Hara, “Neglected canals: Deterioration of indigenous irrigation system by urbanization in the west peri-urban area of Bangkok Metropolitan Region,” *Water*, vol. 4, no. 1, pp. 12–27, 2012.
- [32] J. Mahesh, P. Amerasinghe, and P. Pavelic, “An integrated approach to assess the dynamics of a peri-urban watershed influenced by wastewater irrigation,” *Journal of Hydrology*, vol. 523, pp. 427–440, 2015.
- [33] A. Goap, D. Sharma, A. K. Shukla, and C. Rama Krishna, “An IoT based smart irrigation management system using Machine learning and open source technologies,” *Computers and Electronics in Agriculture*, vol. 155, pp. 41–49, 2018.
- [34] S. Wu, M. A. Bashir, Q. U. A. Raza, A. Rehim, Y. Geng, and L. Cao, “Application of riparian buffer zone in agricultural non-point source pollution control—A review,” *Frontiers in Sustainable Food Systems*, vol. 7, p. 985870, 2023.

Estimating Monthly PM2.5 Levels using Integrated Satellite and Meteorological Data: A Case Study of Suphan Buri Province, Thailand

Marut Phuphaniat

*Geographic Information Science,
School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand
marut@phcsuphan.ac.th*

Phaisarn Jeefoo

*Geographic Information Science,
School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand
phaisarn.je@up.ac.th*

Wipop Paengwangthong*

*Geographic Information Science,
School of Information and
Communication Technology,
University of Phayao
Phayao, Thailand
wipop.pa@up.ac.th*

Abstract— PM2.5 is an air pollutant that causes respiratory diseases. However, in some areas of Suphan Buri Province, Thailand, there are no monitoring stations for early warning. This study integrated satellite data and meteorological data to develop a linear regression model for predicting monthly PM2.5 levels. The predictions were made for two grid sizes: 5x5 km and 10x10 km. The independent variables were satellite data and meteorological data, including Aerosol Optical Depth (AOD), hotspots, air temperature, and relative humidity. The dependent variable was PM2.5 data from five ground-based monitoring stations. After developing a multiple linear regression model, the results showed that the model had R-squared values of 0.665 and 0.760, respectively. Furthermore, model validation showed that the predicted PM2.5 had root mean square errors (RMSE) of 12.82 mg/m³ and 11.53 mg/m³, respectively.

Keywords— PM2.5, Remote sensing, AOD

I. INTRODUCTION

The issue of fine particulate matter with a diameter of less than 2.5 microns (PM2.5) is a significant air pollution concern. PM 2.5 is generated from combustion processes, including vehicle emissions, agricultural burning, wildfires, and industrial activities. Due to its small size, PM 2.5 is more dangerous than larger particulate matter because it can penetrate deep into the lungs. Studies have shown that every 10 micrograms per cubic meter increase in PM 2.5 raises the risk of death by 1-2% for patients with respiratory diseases [1]. Research indicates that PM 2.5 levels in Thailand typically exceed the standard from January to April each year and are increasing across all regions. One of the areas of interest is agricultural zones [2]. Suphan Buri province, located in the central region of Thailand, covers an area of 5,358 square kilometers. The majority of this area, 72.44%, is agricultural land [3]. Most farmers aim to increase their production and often resort to burning agricultural materials to expedite the process. Studies have shown that the period with the highest air pollution coincides with biomass burning from agricultural activities. Agricultural burning is a major source of PM2.5, contributing up to 54.2% [4]. Therefore, having a large agricultural area may increase the risk of PM2.5 levels exceeding the standard, especially during the harvest season [5]. The Department of Pollution Control in Thailand will conduct measurements of PM2.5 particulate matter through ground-level air quality monitoring stations. These measurements correlate with surrounding areas within a limited radius due to constraints in the number and coverage of monitoring stations. Therefore, a satellite-based technique

has been developed using Aerosol Optical Depth (AOD) data to assess air quality comprehensively across all areas [6]. AOD data products are derived from remote sensing techniques, such as those from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors on Terra and Aqua satellites [7]. Studies have shown a significant correlation ($R^2 = 0.69-0.85$) between AOD and PM2.5 concentrations in Southeast Asia [8]. Integrating AOD data with meteorological information such as temperature, humidity [9], and hotspots enhances the accuracy of PM2.5 estimation [10]. This study applies Geographic Information Systems (GIS) and remote sensing techniques to investigate PM2.5 levels in Suphan Buri province. The aim is to use this model to assess PM2.5 situations in areas lacking monitoring stations, for planning control measures and preventing hazards from high PM2.5 levels.

II. MATERIAL AND METHOD

A. Study area: Suphan Buri province, Thailand

Suphan Buri province covers approximately 5,358 square kilometers with boundaries ranging from latitude 14.4 to 15.5 N and longitude 99.17 to 100.16 E. The terrain is predominantly flat with some parts being gently undulating, with slopes ranging from 0-3% predominantly in the west. The lowest elevation is in the southeast, slightly above sea level at around 3 meters, while the highest elevation is in the northwest, about 10 meters above sea level. The majority of the area is agricultural, accounting for 72.44% of the total, with rice paddies covering 51.32% and upland crops 34.82%. It is the province with the highest agricultural area in the central region [3].

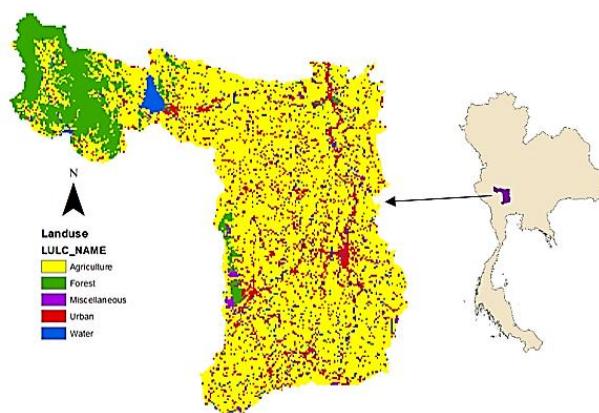


Fig. 1: Suphan Buri Province, Thailand

B. Data Usage

Aerosol Optical Depth (AOD) data from MODIS satellite sensors using the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm with 1-kilometer spatial resolution. The product name is MCD19A2, processed by NASA, available from <https://worldview.earthdata.nasa.gov>. Hotspot data from MODIS satellite sensors, processed by NASA, available from <https://firms.modaps.eosdis.nasa.gov>. Define 5 grid box areas using field sites as the center points. Create buffer zones of 2 sizes: 5 square kilometers and 10 square kilometers. Specify the period for daily data use from January 1, 2022 to December 31, 2023.

Meteorological data and PM2.5 data were obtained from PM2.5 measuring instruments at 5 ground stations (Fig. 2) using daily data from January 1, 2022, to December 31, 2023.

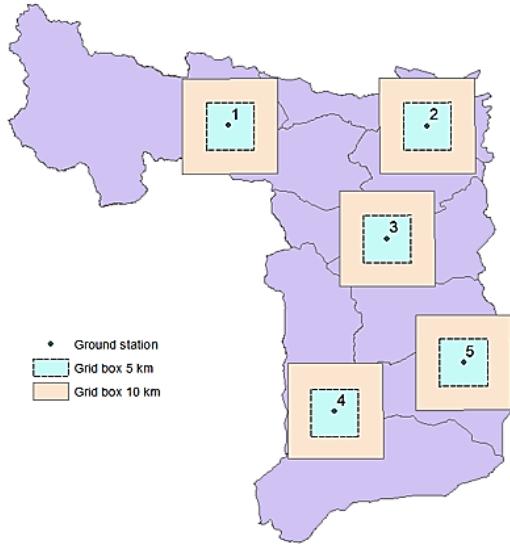


Fig. 2: Location of PM2.5 measuring site.

TABLE I. Coordinates of PM2.5 monitoring devices

ID	Ground stations	Latitude	Longitude
1	Dan Chang	99.703401	14.840323
2	Doem Bang Nang Buat	100.09892	14.848478
3	Don Chedi	100.02014	14.634461
4	Sa Yai Som	99.896471	14.281104
5	Bang Pla Ma	100.16775	14.398392

C. Data Processing

The AOD and Hotspot data were processed using GIS software and converted into CSV format. Subsequently, the AOD and Hotspot data and ground station data, including air temperature, relative humidity, and PM2.5, were aggregated as monthly averages to facilitate joint analysis. All processed data were stored in Excel for further use. A multiple linear regression analysis created a mathematical model for calculating PM2.5 concentrations from AOD, hotspots, relative humidity, and air temperature. An accuracy assessment was conducted by designating one ground station area to compare the model's predicted values with the PM2.5 measurements from the ground station, covering the period

from January 2022 to December 2023, using Root Mean Square Error (RMSE).

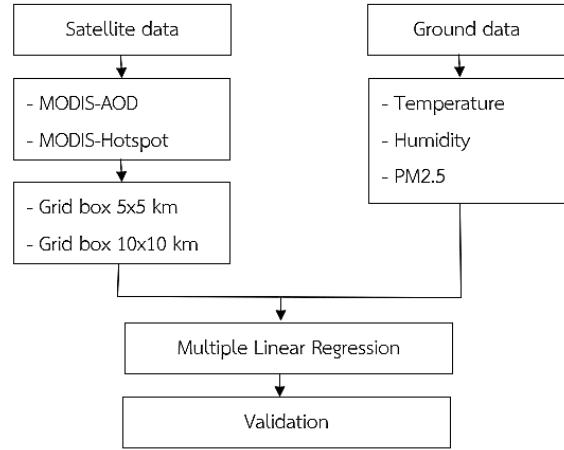


Fig. 3: Conceptual Framework

III. RESULTS

A. Trend of AOD and PM2.5

The AOD values from satellites and PM2.5 concentrations from ground monitoring stations in 2022-2023 showed consistency. AOD values in 5 square kilometer areas, AOD in 10 square kilometer areas, and PM2.5 concentrations were high during January to April, which is the dry season in both years. The average PM2.5 concentrations exceeded the standard at all stations. When considering hotspots and relative humidity, there was also consistency. As observed, during periods with numerous hotspots, both AOD values and PM2.5 concentrations were high. However, when relative humidity was high, AOD values and PM2.5 concentrations decreased. The hotspots were most prevalent during the dry season, which coincided with the harvest period from December to April each year.

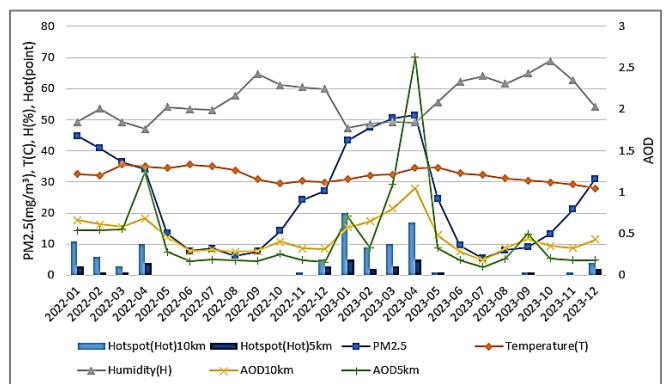


Fig. 4: Trend of AOD and ground PM2.5 from 2022-2023

B. Prediction of PM2.5 concentrations

Predicting PM2.5 concentrations through a mathematical model using multiple linear regression analysis, incorporating data on AOD, hotspots, relative humidity, and air temperature, was conducted over areas of 5 square kilometers and 10 square kilometers. The analysis revealed an R square value of 0.665 for the 5 square kilometer area and 0.760 for the 10 square kilometer area. It was found that AOD has a significantly positive correlation with PM2.5 concentrations (Fig. 5), and

the number of hotspots also has a strong positive correlation with PM2.5 levels. Conversely, relative humidity and air temperature correlate negatively with PM2.5 concentrations.

$$\text{PM2.5}(5\text{x}5\text{km}) = 85.227 + 41.462(\text{AOD}) + 2.393(\text{hotspots}) - 1.292(\text{temperature}) - 0.639(\text{humidity})$$

$$\text{PM2.5}(10\text{x}10\text{km}) = 61.598 + 46.454(\text{AOD}) + 0.943(\text{hotspots}) - 0.995(\text{temperature}) - 0.505(\text{humidity})$$

Based on the assessment of PM2.5 accuracy with the model, the Grid box size of 5 square kilometers showed an error of 12.82 mg/m³, while the Grid box size of 10 square kilometers showed an error of 11.53 mg/m³. However, the PM2.5 trends are consistent (Fig. 6).

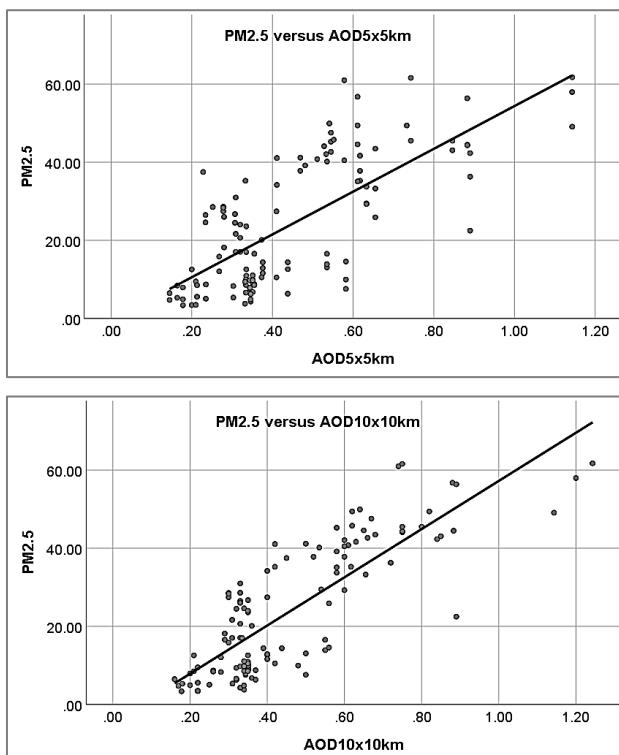


Fig. 5: Scatter plots of AOD and ground PM2.5

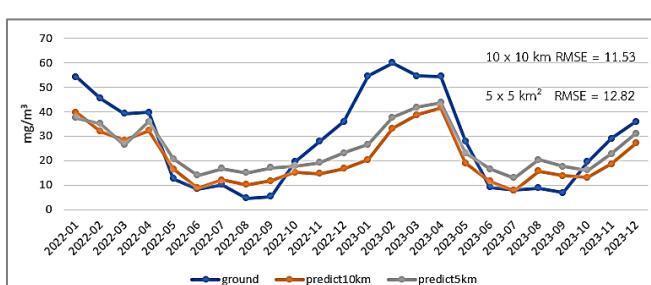


Fig. 6: Differences in PM2.5 from ground stations and predictive models

IV. CONCLUSION

The study found that Aerosol Optical Depth (AOD) has a significantly positive correlation with PM2.5 concentrations.

AOD, derived from remote sensing techniques using satellite data, measures the density of fine particles in the atmosphere. Therefore, an increase in AOD indicates a higher concentration of particles in the atmosphere [7]. Additionally, the number of fire hotspots also shows a strong positive correlation with PM2.5 levels. Hotspots are areas of burning, often found in agricultural regions due to the burning of agricultural residues. Such burning is a primary source of smoke in the study area. Hotspots were assessed on a monthly cumulative basis for suitable analysis since they do not occur daily [5]. Relative humidity has a significantly negative correlation with PM2.5 concentrations, as high humidity helps particulate matter settle to the ground more effectively, reducing airborne dust [11]. Similarly, air temperature exhibits a negative correlation with PM2.5 levels. Higher temperatures can stimulate the movement of airborne particles, causing them to settle more or increase wind movement, which may help disperse the dust particles further away [12].

The variable selection for this study includes variables for predicting, such as AOD data, hotspots, air temperature, and relative humidity. Previous studies have shown that AOD data, air temperature, relative humidity, and wind speed are correlated with PM2.5 levels [9]. Additionally, reports indicate that hotspot data and relative humidity are correlated with PM2.5 levels [13]. Due to spatial limitations, wind speed and atmospheric layer levels cannot be used. However, this PM2.5 estimation model can predict PM2.5 levels in Suphan Buri Province with an accuracy of 76%.

In determining the grid box size for this study, previous research has utilized various sizes for satellite-derived AOD data, such as 30 km², 10 km², 5 km², and 3 km². For this study, a grid box size of 10 km² was selected as a moderate size, and a grid box size of 5 km² as a small size. The 30 km² size was not chosen due to its large size exceeding the study area's boundaries, and the 3 km² size was not selected because it is too small, leading to significant variability in hotspot data. Previous studies have shown that larger grid box sizes tend to have higher R-squared values compared to smaller grid box sizes [14], as they provide more data points which reflect central tendencies better than smaller grid box sizes. However, because PM2.5 values are influenced by multiple factors, combining AOD data, which serves as a representative of the area, with meteorological data improves PM2.5 prediction accuracy.

The accuracy assessment of the model, the grid box size of 5 km² has a deviation of 12.82 mg/m³, while the grid box size of 10 km² has a deviation of 11.53 mg/m³, which is less than the minimum deviation of 14.00 mg/m³ found by Phuengsamran & Lalitaporn [9]. Additionally, data collection revealed significant data incompleteness during the rainy season, including both satellite and ground-based data, making daily data unusable. Therefore, it is necessary to adjust the data to monthly averages.

ACKNOWLEDGMENT

The authors acknowledge the School of Information and Communication Technology (ICT), University of Phayao, Thailand for providing the available facilities to carry out this work. Special thanks would like to deliver to Suphan Buri Provincial Meteorological Station and Hospitals in Suphan Buri Province for supporting data.

REFERENCES

- [1] K. Maciejewska, "Short-term impact of PM2.5, PM10, and PMc on mortality and morbidity in the agglomeration of Warsaw, Poland," *Air Qual Atmos Health*, vol. 13, pp. 659-672, 2020.
- [2] Department of Pollution Control, Guidelines for Prevention and Mitigation of PM2.5 Particulate Matter Issues. Bangkok: Ministry of Natural Resources and Environment, 2021.
- [3] Department of Land Development, "Report on Land Use Survey/Improvement," 2023. [Online]. Available: http://www1.ldd.go.th/web_OLP/report_research_C.html. [Accessed: 11-Jan-2024].
- [4] M. M. Rahman, B. A. Begum, P. K. Hopke, K. Nahar, and G. D. Thurston, "Assessing the PM2.5 impact of biomass in megacity Dhaka, Bangladesh," *Environmental Pollution*, vol. 264, p. 114798, 2020, doi: 10.1016/j.envpol.2020.114798.
- [5] M. Zamanielaei, A. M. Shew, J. J. Fain, A. Borkowski, and J. L. McCarty, "Crop residue burning from high-resolution satellite imagery and PM2.5 dispersion: A case study of Mississippi County, Arkansas, USA," *Sustainable Environment*, vol. 9, no. 1, p. 2274646, 2023, doi: 10.1080/27658511.2023.2274646.
- [6] M. Sorek-Hamer, R. Chatfield, and Y. Liu, "Review: Strategies for using satellite-based products in modeling PM2.5 and short-term pollution episodes," *Environment International*, vol. 144, p. 106057, 2020, doi: 10.1016/j.envint.2020.106057.
- [7] M. Stafoggia, T. Bellander, S. Bucci, M. Davoli, K. de Hoogh, F. De' Donato, C. Gariazzo, A. Lyapustin, P. Michelozzi, M. Renzi, M. Scorticini, A. Shtein, G. Viegi, I. Kloog, and J. Schwartz, "Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model," *Environment International*, vol. 124, pp. 170-179, 2019, doi: 10.1016/j.envint.2019.01.016.
- [8] Y. N. Ma, J. Y. Xin, W. Y. Zhang, Z. R. Liu, Y. J. Ma, L. B. Kong, Y. S. Wang, Y. Deng, S. H. Lin, and Z. M. He, "Long-term variations of the PM2.5 concentration identified by MODIS in the tropical rain forest, Southeast Asia," *Atmospheric Research*, vol. 219, pp. 140-152, 2018.
- [9] P. Phuengsamran and P. Lalitaporn, "Estimating particulate matter concentrations in central Thailand using satellite data," *Thai Environmental Engineering Journal*, vol. 35, no. 3, pp. 1-11, 2021.
- [10] D. Ruidas and S. C. Pal, "Potential hotspot modeling and monitoring of PM2.5 concentration for sustainable environmental health in Maharashtra, India," *Sustainable Water Resources Management*, vol. 8, no. 4, p. 98, 2022, doi: 10.1007/s40899-022-00682-5.
- [11] X. Wang, R. Zhang, and W. Yu, "The effects of PM2.5 concentrations and relative humidity on atmospheric visibility in Beijing," *Journal of Geophysical Research: Atmospheres*, vol. 124, pp. 2235-2259, 2019, doi: 10.1029/2018JD029269.
- [12] Vaishali, G. Verma, and R. M. Das, "Influence of temperature and relative humidity on PM2.5 concentration over Delhi," *MAPAN*, vol. 38, pp. 759-769, 2023, doi: 10.1007/s12647-023-00656-8.
- [13] C. Sritong-aon, J. Thomya, C. Kertpromphan, and A. Phosri, "Estimated effects of meteorological factors and fire hotspots on ambient particulate matter in the northern region of Thailand," *Air Qual. Atmos. Health*, vol. 14, pp. 1857-1868, 2021.
- [14] S. K. Sahu, M. Beig, S. K. Dash, and J. P. Pandey, "Impact of horizontal resolution on simulation of aerosol optical properties in East Asia: A case study using WRF-Chem," *Atmospheric Environment*, vol. 130, pp. 156-169, 2016, doi: 10.1016/j.atmosenv.2015.09.038.

Visualizing Awareness and Behavior in Reducing Single-Use Plastic among High School Students Using the Interactive ArcGIS Storymaps

*Note: Sub-titles are not captured in Xplore and should not be used

Do Thi Viet Huong

Department of Environmental Resources Management and Geoinformatics Faculty of Geography and Geology Hue University of Sciences Hue City, Vietnam dtvhuong@hueuni.edu.vn

Nguyen Phuong Nghi

Hue University of Sciences High School for Gifted Students Hue City, Vietnam

nguyenphuongnghi0405@gmail.com

Nguyen Thi Thanh Ninh

Hue University of Sciences High School for Gifted Students Hue City, Vietnam

mimithanhnin@gmail.com

Abstract - Single-use plastic (SUP) products, currently known for being "super cheap and convenient," pose a significant threat to biodiversity, climate change, and various other repercussions that humans and nature must bear when they end up in the environment, particularly in the oceans. Educating about environmental awareness is a long-term direction aimed at building a better environment for the future, especially for the generation of students sitting in classrooms. The Interactive StoryMap Web application contributes to innovating solutions to increase awareness of reducing single-use plastic waste among students. The product is designed using GIS technology, integrating online survey applications ArcGIS Survey123 on high student awareness and behavior in reducing single-use plastic, interactive web maps ArcGIS Web Application, real-time reporting map applications ArcGIS Dashboard, and packaged within the ArcGIS multimedia StoryMap. The StoryMaps content focuses on (1) raising awareness about the issue of plastic waste pollution, (2) sharing information about solutions and actions to reduce single-use plastic among high school students, and (3) creating interactive digital maps and dashboards about plastic reduction community participation. The interactive StoryMap application product has been applied in practice, spreading and promoting the participation of students in the Hue University of Science High School for Gifted Students and several high schools in the province through the network of students and teachers, open day activity participation in live interactive games, and joining hands in the plastic reduction journey of students and the community. This product can edit, supplement, update, and display instant information quickly and effectively. It can also be used and interacted with on different devices to help communicate powerfully with every student, from inside and outside the school, decision-makers on environmental protection and sustainable development, suitable for the context of technological development 4.0, GIS in schools, and the National Digital Transformation Program contributing to the goal of Thua Thien Hue - the Green City, aiming to become a City directly under the Central Government.

Keywords— Single-use plastic, Thua Thien Hue, High school students, ArcGIS StoryMaps, interactive Web application

I. INTRODUCTION

Plastic has become widely used worldwide but is also a dangerous pollutant in the air, land, and water. According to the Organization for Economic Cooperation and Development (OECD), in 2021, the world generated 353 million tons of plastic waste. Still, less than 10% of plastic waste was

recycled [10]. In Vietnam, rapid economic growth and urbanization, accompanied by rapid changes in lifestyle, have been leading to a plastic pollution crisis. The International Organization for Nature Conservation in Vietnam (WWF) has also identified that the use of single-use plastic products (Single-use Plastic - SUP) is a widespread behavior in Vietnam [12].

Towards a national roadmap to reduce single-use plastic in Vietnam, in: "National Action Plan on Ocean Plastic Waste Management to 2030," Vietnam aims to cut plastic waste by 50% in the sea and ocean in 2025 and reach 75% in 2030. In particular, the measure seeks to ban the production and import of plastic bags for SUP for domestic use (2026) by 2031. The Government is drastically implementing policies and effective control solutions, focusing on minimizing use, training, raising awareness, changing plastic waste disposal behavior, increasing recycling, reusing, and promoting a circular economy and green growth [7].

Along with the general trend of making Thua Thien-Hue a centrally run city by 2025 and a typical heritage city of Vietnam by 2030, the city's socio-economy is increasingly developing and poses the risk of plastic waste problems, significantly when expanding administrative boundaries if not well controlled [7]. Single-use plastic products (SUP) are now "super cheap and convenient," such as foam boxes, plastic bags, cups, and straws to store breakfast food or buy drinks at school. When released into the environment, the ocean threatens biodiversity, climate change, and many other consequences humans and nature must suffer. In Vietnam, single-use plastic is used and disposed of every day, especially among young people and students [11, 12]. Raising awareness and changing the younger generation's behavior is essential in solving this problem. However, traditional methods to educate and raise awareness often face limitations in effectiveness and reach. Around the world, several interactive communication methods on the ArcGIS StoryMaps application system have been deployed by some environmental research organizations and projects to integrate multimedia in raising awareness of the problem of plastic waste in the mountains, ocean, seaside, school, etc... [1, 4, 13].

Campion School in New Zealand has implemented a school-based GIS project for high school students, in which ArcGIS StoryMaps are used to build stories about the waste journey at local beaches. This communication approach

guides students to understand how community citizens contribute to plastic reduction, supporting contributions to achieving sustainable development goals [3]. A StoryMap about: "Tracking Plastic Waste" was created by Shah N. Selbe and a group of students to track the journey of plastic waste into the sea [13]. Spoiled Oceans StoryMap also shows students' natural perspectives on microplastics and food waste in ecosystems, communities, and school food [9].

In the context of the development of 4.0 technology, geographic information systems (GIS) in schools, the National digital transformation program, and the characteristics of the GenZ generation, diversifying communication solutions and enhancing student interaction are essential for applications. Spreading this to the community in joining hands to reduce plastic, thereby promoting behavioral change towards a circular economy, is necessary.

ArcGIS StoryMap is a new communication trend in the Internet age that informs, inspires, and attracts stakeholders about a resource, environmental, or social issue. This powerful media tool converts maps and geographic information systems (GIS) into interactive information visualization content to create interactive cartographic stories, bringing significant potential in vividly and attractively conveying information (ESRI, 2023). By combining visual elements, text, and spatial data, StoryMap can help students become more aware of the impact of single-use plastics and motivate them to take action to reduce the use of plastic.

The article aims to (1) analyze the current status of students' awareness and behavior of plastic reduction and (2) contribute to innovating solutions to increase awareness of reducing single-use plastic waste among students, thereby promoting reducing behavior disposable plastic among students through visualize Interactive ArcGIS StoryMap Web application in the ArcGIS Online platform.

II. DATA AND METHODOLOGY

A. Data collection

The data for this study were collected through a combination of surveys, interactive web maps, and real-time reporting tools in the ArcGIS Online platform.

Data were gathered using the ArcGIS Survey123 application, which enabled the collection of high school students' awareness, attitudes, and behaviors toward reducing SUP waste. The survey first targeted students from the Hue University of Science High School for Gifted Students (HUSC High School), followed the sampling Taro Yamane (1967), and then expanded to other high schools in Thua Thien Hue province [14].

$$n = \frac{N}{1 + N(e)^2} \quad (1)$$

where n is the sample size, N is the population size, and e is the margin of error. To enhance accuracy, apply a 10% margin of error, with a total number of students of 374 and a sample size of approximately 87 students (plus 10%). Students were selected for the survey at their convenience, promoting active participation in filling out the survey link.

GIS data were utilized to create interactive ArcGIS web maps and dashboards. These maps incorporated spatial data on engaged survey students' plastic waste hotspots and community participation in plastic reduction efforts.

Data on student interaction with the ArcGIS StoryMap application were collected in real-time, including metrics such as participation in live interactive games, visits to the web application, and contributions to the plastic reduction journey.

B. Methodology

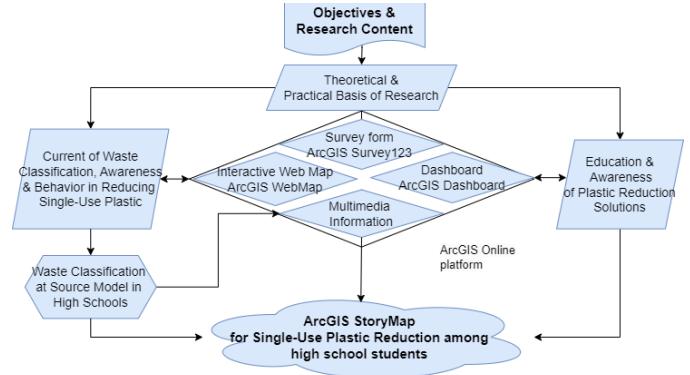


Figure 1. Procedure of research

1) Survey Design and Deployment:

Based on the objectives, research content, theoretical and practical basis related to single-use plastic waste are identified and analyzed. An online survey was designed using the ArcGIS Survey123 application, tailored for smartphone use. This survey was deployed in schools to collect data on students' knowledge and behaviors towards SUP. The study aimed to capture insights from a broad student population, focusing initially on HUSC High School. The survey content concentrates on (1) Awareness of SUP: decomposition time, hazards, knowledge of 3T terminology, waste classification in schools, (2) SUP usage behavior: commonly used SUP types, level of use, reasons for use, post-use habits, factors in maintaining SUP use, recycling methods, willingness to change to reduce SUP, and (3) Solution: Solution to minimize, influencing factors and communication methods to raise awareness. In addition, information about school grades and living areas was also collected during the survey.

2) Pilot Implementation of Waste Classification Model:

A pilot project was initiated at HUSC High School, where a waste classification model at the source was tested. This pilot included designing instructions and implementing suitable bin locations to reduce plastic waste within the school. This pilot's success would inform the model's potential expansion to other high schools and the broader community.

3) Development of the StoryMap Application:

The StoryMap application was designed using ArcGIS technology, integrating various GIS tools to create an interactive and engaging platform. The application combined multimedia elements, including text, images, videos, and interactive maps, to present information on SUP pollution and reduction strategies. Connecting modern GIS components, including ArcGIS Survey123 field surveys, ArcGIS Webmap, ArcGIS Dashboards, multimedia, and interactive maps, can help promote spatial thinking. The integration process is sequential from (1) Design a survey form on ArcGIS Survey123; (2) Create and analyze maps (interactive/multimedia Map) from survey results; (3) Create an ArcGIS Dashboard; (4) Integration on ArcGIS StoryMaps.

4) Data Analysis Visualization and Dissemination

Comprehensive information on solutions to reduce plastic waste and knowledge about SUP was compiled. This

information was then digitized and integrated into interactive designs using the ArcGIS Survey123 application, ArcGIS Webmap for interactive web mapping, and ArcGIS Dashboard for real-time data visualization. Instant and multimedia information (videos, images, etc...) are integrated into the ArcGIS StoryMap application. Through groups of friends and classes, the digital interactive story map application product will gradually be shared with students outside the school and spread on the school's website and the Facebook page of HUSC High School, social networks, and career guidance activities OPEN DAY in Hue University of Sciences.

III. RESULT AND DISCUSSION

A. High school student awareness and behavior about Single-Use Plastics

1) Awareness about Single-Use Plastics

Analysis results show that the total number of students surveyed was 87, with 85 valid survey contents, accounting for 97.70%. Of these, 58.62% were female, and 41.38% were male, who responded to the questionnaire. The distribution map of the living areas surveyed by the students in the Data function is edited and displayed on ArcGIS Online, showing that they are located within Hue City (Figure 2).

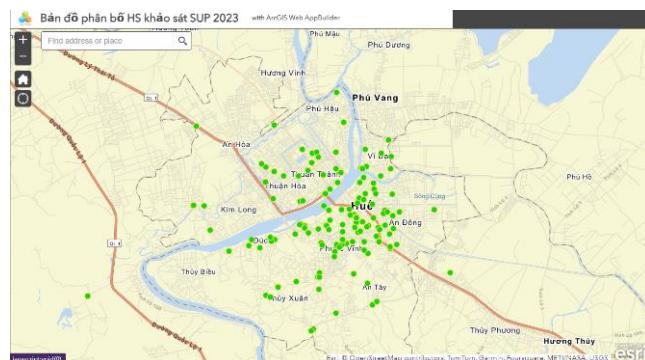


Figure 2. Distribution map of students participating in the research survey

Regarding the time it takes for plastic waste (plastic water bottles, straws, bags..., etc.) to decompose in the environment, 60.3% of surveyed students believe that it takes hundreds to more than 500 years, while up to 23.8% of students think that plastic waste can decompose in the environment in about 10 - 50 years.

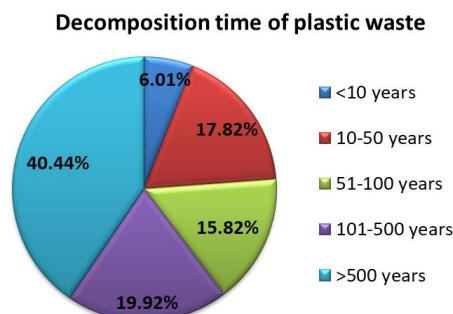


Figure 3. Awareness of plastic waste decomposition time (Source: Survey conducted by the research team)

This shows that in addition to a relatively high proportion of students being aware of the difficulty of decomposing plastic waste in the environment, several students are also

unaware of the long-term existence of plastic, which will negatively impact the environment, creatures, and humans.

Regarding awareness of health and environmental hazards caused by SUP waste, 74% of students are concerned about the risks, while 26% do not know or do not care. Regarding awareness of waste classified into several groups in school, only about 59% of students are interested in classifying waste in school (Figure 4). This shows that it accurately reflects the current reality in the school where there are only trash bins and no classified domestic solid waste containers model like the project "Hue - Plastic Waste Reduction Urban in Central Vietnam" funded by WWF - Vietnam, which is being implemented for Thua Thien Hue Province.

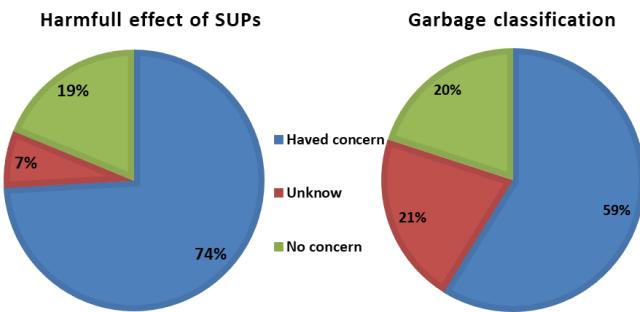


Figure 4. Awareness of (a) Harmfull effect of SUP and (b) Garbage classification (Source: Survey conducted by the research team)

Regarding awareness of the term solution of 3T (in Vietnamese) or 3R (in English) in the circular economy with plastic waste. By analyzing in a word cloud form, statistical results show that 32 students do not know or do not care, 14 students see the term "Recude," 10 students think it is "Reduce - Reuse - Recycle," 3 students know "Reuse," and also understand the terms 3R - 3T - "Reduce - Reuse - Recycle, Reduce - Recycle - Reuse" with about 4 students. This shows that students have somewhat heard of and are aware of some information related to 3R - 3T, but the information is still not clear and specific.



Figure 5. Awareness of the term 3T (Vietnamese) or 3R (English) in the circular economy with plastic waste (Source: Survey conducted by the research team)

2) Behavioral Patterns in Single-Use Plastics Usage

According to students' opinions, they use foam food boxes the most, followed by plastic bags, corresponding to 64.71% and 57.65% of the total 85 survey opinions (Figure 6). This reflects the reality that in the morning, students often buy breakfast in front of the school gate and bring it into the classroom, and a plastic bag also accompanies the outside of the foam box. In addition, plastic bottles and plastic cups for soft drinks are also popular choices for students, accounting for over 40%.

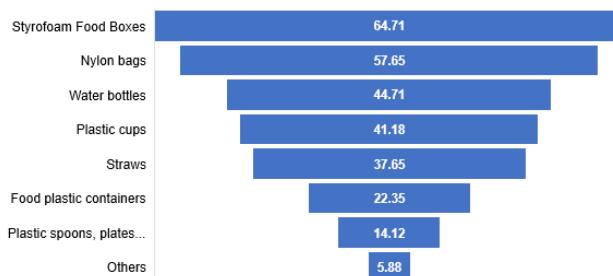


Figure 6. The common type of SUP consumed by high school students

Regarding the location of SUPs, students said that most of them are scattered around trash cans (unsorted bins), in the schoolyard, on the road, and in the classroom.

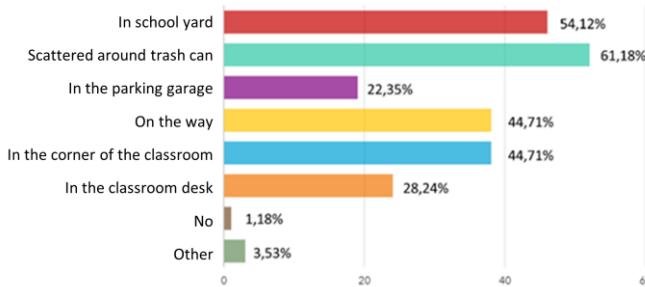


Figure 7. Frequency and Locations of SUP Usage

Regarding how often students use SUPs, 44.7% think that the level of use is average (accounting for the highest percentage), and 25% use it often and very often (Figure 8). This shows that students are familiar with the problem of using disposable plastic items in daily activities. Most students use 1 to 2 disposable plastic products, accounting for nearly 75%; on average, three or more products account for about 23%.

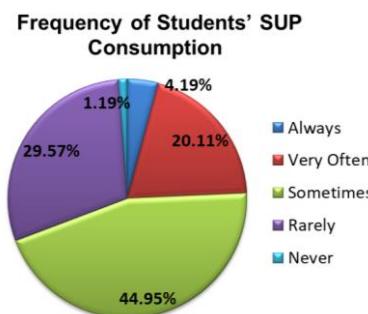


Figure 8. Frequency of Students' SUP consumption

Regarding disposal Habits of used SUPs, 42.35% of students use and throw away the trash. The remaining 30.59% were reused for other purposes, while 20% collected waste (putting it in separate bins). This shows that the need for reuse and collection among students is still significant, but schools need a waste classification and collection system.

In motivations for SUP use, the main reason for using SUP among students was that the majority said it was light (51.76%) and convenient (42.35%), while nearly 30% also said it was because there were no substitute products. (Figure 9). This is also the basis for considering the issue: the "root" of reducing single-use plastic in the community.

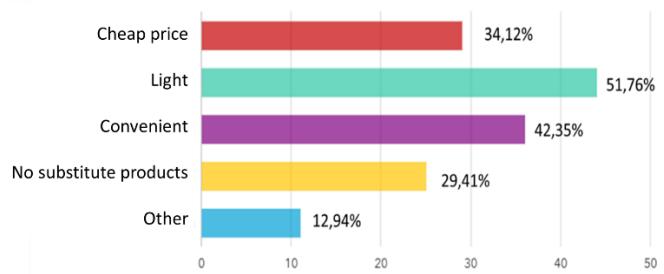


Figure 9. The main reason for using SUBs among students

The factor contributing to maintaining students' single-use plastic behavior is the frequency of regular use (habit), accounting for 42.35%. Next is convenience and compactness (accounting for 30.59%). Besides, up to 22.35% of opinions said that it was due to the proactive provision of the seller (Figure 10).

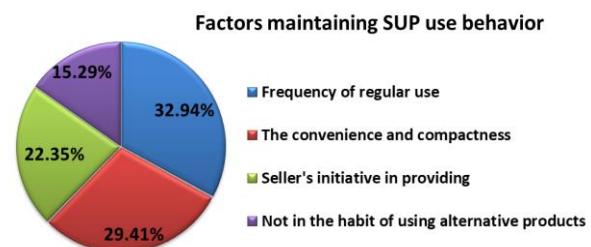


Figure 10. Factor contributes to maintaining students' single-use plastic behavior

Regarding willingness to change habits to reduce SUP use, up to 74.12% of students are willing to change, but there is still a sizeable remaining percentage who are not or are not interested (about 25%) (Figure 11).

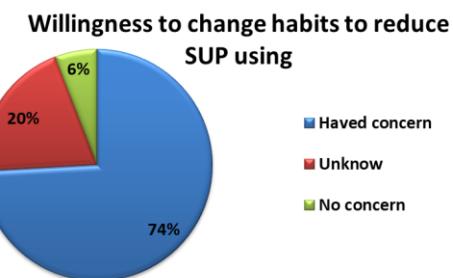


Figure 11. willingness to change habits to reduce SUP use

Regarding answers about solutions/ideas about some ways of recycling for students and solutions to reduce single-use plastic waste, many opinions focused on traditional solutions such as recycling products. Disposable plastic used to make decorations, toys, and plants; reuse plastic bags and straws; Change the habit of using personal water bottles and cloth bags; and refuse to use plastic straws. At the same time, technological solutions appear: enhancing interactive communication with short films, animations, quiz games, etc (Figure 12).

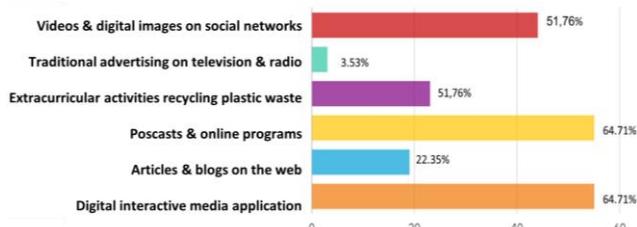


Figure 12. Solutions/ideas for recycling and reducing single-use plastic waste

To have a basis for designing digital interactive media content on ArcGIS StoryMap, some surveys on traditional digital solution ideas suitable for the GenZ generation (through media selection and stimulating media elements) are conducted. ratio) is shown as shown in Figure 13.

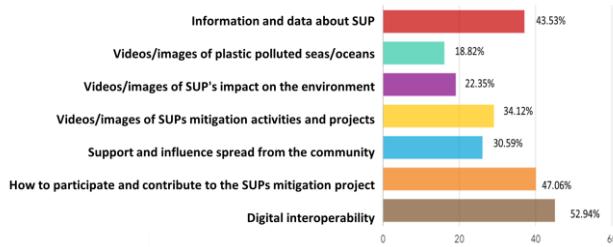


Figure 13. Opinions about media attract GenZ's attention

B. Waste Classification Model in High School

1) Design instructions and determine the location for installing a waste classification model

Based on the survey results on students' awareness and behavior about single-use plastic waste show that raising awareness as well as taking specific action promotes reducing single-use plastic among students. The model of waste classification at source was installed and deployed, and the first test instructions were noted in May 2023, in which the blue bin was recycled and reused waste, the Yellow bin was organic waste, and the Orange bin was hazardous and inorganic waste (Figure 14).



Figure 14. Model of garbage classification in schools

Based on the results of an actual survey of the buildings of HUSC High School, which currently do not have a sortable

trash system, the research team has designed a garbage collection model in the school and arranged it on the floors according to the location shown in diagram figure 15. Trash can locations are in empty spaces (stairs on the 2nd floor) and the middle corridor (floors 1, 3, and 4).

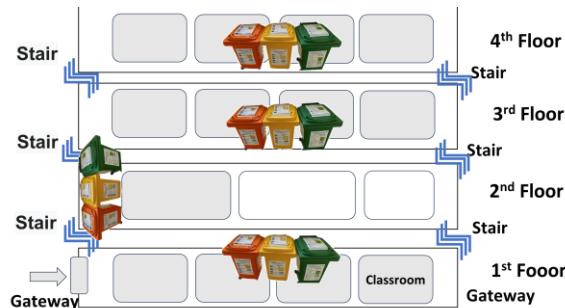


Figure 15. Location of setting the model of garbage classification in schools

2) Initial effectiveness of the implementation model

The process of implementing the model and evaluating the initial effectiveness of the model is assessed by observing the waste sorting behavior at the source of students on all floors and examining changes over time. The initial effectiveness of the model is evaluated semi-quantitatively according to 3 levels: (O) > 80% correct classification, few mistakes, awareness of keeping trash bins clean; (o) 50 - 79% are relatively classified, but mixed with other bins; (x) < 50% correctly classified, thrown away, left out.

The results of observation and general assessment of the effectiveness of the initial installation of the waste classification model in schools show that students are beginning to be aware of waste classification. Still, in reality, the amount of recyclable and recyclable waste is used (blue). Students generate much more waste than organic, hazardous, and inorganic waste. These findings are the basis for considering adjusting the quantity and size or changing the collection frequency in trash bin arrangements to make it convenient for students to classify at source in schools. It is also important to note that the two bins with the highest frequency of littering are the blue bin (recyclable/reusable waste), the yellow bin (organic waste), and the orange bin (hazardous and inorganic waste) almost very little can use small old crates and accompanying instruction notes.

TABLE 1. INITIAL EFFECTIVENESS OF THE WASTE CLASSIFICATION MODEL IN SCHOOLS

Month	Day/ Week	Building floor			
		1 st	2 nd	3 rd	4 th
May (15-20/5/2023)	Mon		x	x	
	Tue		x	x	
	Wend		x	x	
	Thurs		x	x	
	Fri		o	o	
	Sat		o	x	
September 2023	Week 1	x	x	x	x
	Week 2	x	o	x	x
	Week 3	x	o	x	o
	Week 4	o	o	o	o
October 2023	Week 1	o	o	o	o
	Week 2	o	o	o	o

	Week 3	o	o	o	o
	Week 4	o	o	o	o
Decmeber 2023	Week 1	o	O	o	o
	Week 2	o	O	o	o
	Week 3	O	O	O	O
	Week 4	O	O	O	O

(O) > 80% correct classification, few mistakes, awareness of keeping trash bins clean;

(o) 50 - 79% are relatively classified but mixed with other bins;

(x) < 50% correctly classified, thrown away, left out.

C. Interactive StoryMap supports raising students' awareness and behavior in reducing single-use plastic

1) StoryMap application interface

The StoryMap theme is REDUCING SINGLE-USED PLASTIC, with the message "Creating sustainable change - towards a Green – Clean - Bright Hue." This is a digital transformation approach in communication to raise awareness of single-use plastic reduction, integrating plastic reduction efforts in multimedia storytelling from Problem Identification > Planning > Field Research > Data Management > Analysis > Collaborate and Share.



Figure 16. Interface of StoryMap on Desktop screen and smartphone

2) Interactive content and functions on the StoryMap application

* Educational content to raise awareness of SUP:

Coming to the story's beginning, the experience will lead to taking a short survey on the Survey123 link about disposable plastic. Encourage students to answer what they have perceived and their actual behavior, not afraid of being wrong or right, and send information about participating students' living/school locations on the interactive digital map. This is an online survey designed to examine the perceptions and behaviors of students at Hue High School for the Gifted in Science.

Next, students will experience the interactive game QUIZ - SUP, designed and integrated into ArcGIS Survey123 and ArcGIS Dashboard. The implementation process includes designing cognitive question forms in single selection (True/False) and returning results according to ArcGIS Survey123 Quiz - SUP conditions to provide students with more information about single-use plastic. Each experience corresponds to the results returned by analyzing Map Interactive data; the ArcGIS Dashboard is designed to include the survey form and information on the number of current users (updated continuously after refreshing). Finally, this dashboard will be integrated into ArcGIS StoryMap using the Media tool with the shared Embed code. The results are shown in Figure 17, which shows 280 game interactions when the web link is refreshed.



Figure 17. Interactive game QUIZ - SUP experience

In addition, statistical information about plastic waste and single-use plastic worldwide is also presented through self-designed videos on the Canva application, comic storytelling, infographics, and maps of plastic waste distribution worldwide. Statistical data is directly linked from the website <https://ourworldindata.org/>. Information is displayed through maps, data tables, and visual graphs to help viewers interact and access more updated information (Meijer et al., 2021). In addition, information on plastic waste worldwide, including statistics and facts, was also collected and updated (Bruna Alves, 2023). The information is designed to interact with the Media tool through the options: Text, Embedded, Image, Slide car, and Video. Figure 18 illustrates the video media for explaining the truth of SUP.



Figure 18. Multimedia content integrated into educational content to raise awareness about plastics

* Pilot project implemented at HUSC High School sharing

Information on the methods and content of implementing small projects by student groups to raise awareness and behavior on plastic reduction is presented and shared. In particular, the model of waste classification at source in the School is given with the Media tool in the form of an Image Gallery. In addition, the results of comparing the initial changes in students' awareness and behavior on plastic reduction are also presented through the integration of the ArcGIS Dashboard application.

* Interactive map spreads community participation

This StoryMap application product is aimed at students in HUSC High School and spreads to students and the community to join hands in reducing single-use plastic. Notably, the StoryMaps application allows instant data creation on the ArcGIS Dashboard application, integrating ArcGIS Online survey forms and, for example, creating interactive maps with information on plastic alternatives, recycling points, communities developing environmentally friendly products, or research projects and encouraging participation. This information will help users better understand plastic alternatives and promote involvement in reducing plastic use, especially single-use plastic. The community can spread information on plastic reduction

solutions by filling out the form in the Dashboard application according to the required information and selecting Submit; the results will be displayed on the map. This is a community outreach solution to communicate and reduce plastic waste. This process is also based on integrating ArcGIS Survey123, ArcGIS web map, ArcGIS Dashboard, and ArcGIS StoryMap (Figure 19).

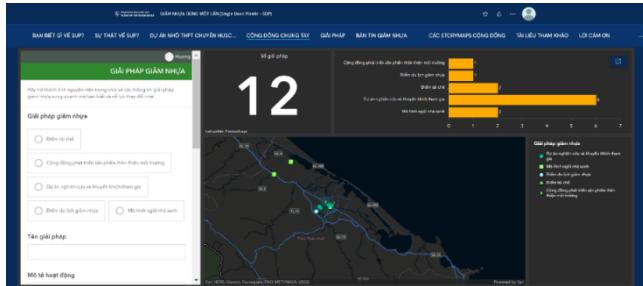


Figure 19. ArcGIS Dashboard application spreads sharing of plastic reduction solutions.

* Solutions, plastic reduction newsletters, and community map stories:

Some solutions recommend that students promote plastic reduction behavior. Updates on plastic reduction programs and projects at local and regional scales are also integrated into the reduction story map application. Minimize plastic. StoryMaps publishing applications with user communities worldwide are integrated to give users multi-dimensional views of solutions and practices for plastic reduction efforts locally, nationally, and worldwide. Information is provided on the menu: "SOLUTIONS," "PLASTIC REDUCTION NEWSLETTER," and "COMMUNITY STORYMAP."

* Awareness and behavior about SUPs:

To provide preliminary assessments of changes before and after intervention in communication solutions to users on awareness and behavior of reducing single-use plastic, the menu "AWARENESS AND BEHAVIOR ABOUT SUPS" is designed, in which The tester will continue to submit the Survey123 survey form - the content is the same as the first survey content before interacting with the SUP Reduction Story Map application to review the changed responses as follows: how. After submitting, the immediate response results will provide a map report of the distribution of survey participants and a Dashboard report comparing before and after the application experience (during the implementation of the waste classification model at school, experience the story map application).

3) Sharing, dissemination, and replication of the StoryMaps application

This StoryMap application allows designers to adjust, add, and update information conveniently and quickly. It can be shared with everyone or within a separate organization group on social networks Facebook, Twitter, and LinkedIn, and it can be opened on computer screen devices (laptop, desktop), smartphones, and tablets conveniently through the sharing link or QR code on the application. The interactive web maps, Survey123 online survey forms, and ArcGIS Dashboard panels designed in map stories are all hosted on ArcGIS Online accounts with cloud computing.

D. Effectiveness of the Digital Tools ArcGIS Storymap

1) Drive engagement with Interactive ArcGIS Storymaps

The interactive StoryMap on plastic reduction has promoted student participation by implementing a survey on the awareness and behavior of students at the HUSC High School regarding plastic reduction before and after implementing the waste classification model. The results of implementing this ArcGIS StoryMaps application were also initially tested to verify the effectiveness of support based on a re-survey of students who had participated in a survey on awareness and behavior of reducing single-use plastic before deployment solution (April 2023). As a result, 81/85 participating students, reaching 95%, responded through the survey form. Data synthesis and online analysis from ArcGIS Survey123 from data and visual design on ArcGIS Dashboard show that student responses are more optimistic.

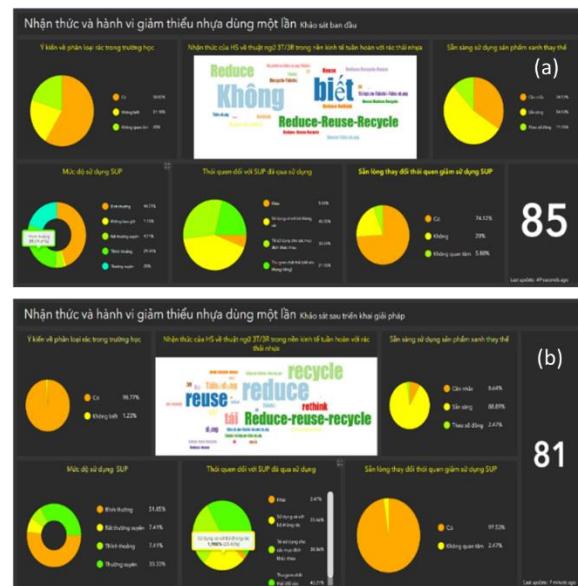


Figure 20. Dashboard reports survey comments and behavior reduce SUP (a) before and (b) after implementing solutions.

Through the model of waste classification at source in schools, digital interactive sharing information on the Story Map application "Reduce Plastic Once" has partly influenced opinions about the level of willingness to use the product. Instead of mainly throwing away the trash after using the SUP area, reducing the use of SUP will aim to increase the option of reusing or collecting waste to put in classified trash cans.

The number of people visiting the StoryMap application is 2619, of which participating in live interactive games is 310, and contributing to the map 21 information points about plastic alternative solutions - plastic reduction initiatives; 35 community networks are not afraid to reduce plastic participating in the Hue Project - Urban plastic reduction in Central Vietnam - Campaign "Modern users, not afraid to reduce plastic." Includes information on which cafes and milk tea shops have vouchers to visit, and do not forget to bring your own cup/water bottle to help reduce plastic; 6 proposed points to detect and clean up plastic waste hot spots.

In particular, the Plastic Reduction Story Map product was displayed within the framework of the TOT Training Course on environmental education and ocean plastic waste reduction for teachers and key staff in the Provinces/Cities of

the Central Coast region of Vietnam on June 4th and 5th, 2024 by the International Organization for Conservation of Nature in Vietnam (WWF - Vietnam) in collaboration with the University of Sciences, Hue University. The results showed feedback from students and teachers about the effectiveness and potential spread of StoryMap's application products, affecting SUP usage awareness and behavior.

Changing awareness has helped students gradually change their behavior toward reducing plastic. However, it is not possible to evaluate the effectiveness immediately in a short time or just by looking at the information shared on the StoryMap application. It is a long journey and requires the coordination of individuals accompanying the community to spread. For the GenZ generation, who are constantly exposed to new and agile technology, integrating interactive applications and engaging with users creates opportunities to experience and encourage the contributions of the younger generation, which is very important and meaningful. In the context of national digital transformation, Thua Thien Hue province is one of the pioneering localities in the digital transformation program; therefore, by applying GIS technology in resource and environmental management, it can be seen that this solution is on the right track.

2) Effectiveness of the Digital Tools

** Conveying environmental messages without using plastic materials:*

Instead of previously commonly used plastic printing materials such as banners, posters, or brochures, with ArcGIS StoryMap - Reduce single-use plastic, you can save costs and limit plastic items. This helps convey the environmental message effectively without using toxic plastic materials and contributes to reducing the amount of plastic waste.

** Creating intuitive and interactive digital experiences: Adaptability and User Interaction*

The content transmitted on the Story Map allows the creation of rich, instant interactive maps and information, combining images, videos, and other multimedia elements. Interactive maps can convey messages vividly and attractively instead of using visual materials such as posters and banners. You can use images and videos to illustrate the impact of using plastic and disposable plastic on the environment, provide information on solutions to reduce and replace plastic, and encourage participation in conservation activities, such as environmental protection.

** Compatible with digital media platforms:*

Story Map application ArcGIS StoryMap's real-time data visualization capabilities can be integrated and shared on digital media platforms such as websites, mobile applications, and social networks. Through widespread media channels, this helps spread the plastic reduction message to many users, including the GenZ generation and enhances user experience and learning outcomes.

** Promote individual and community contributions:*

By creating an interactive map of plastic alternatives and recycling points, the community develops environmentally friendly products and research projects that encourage participation. These insights help users better understand alternative reduction solutions and encourage participation in

actively reducing the use of single-use plastics. They contribute to raising awareness and driving behavior change.

3) Novelty and creativity of Interactive StoryMap

- Spatial data visualization technology:

Applying ArcGIS StoryMap to environmental education is a new method for visualizing spatial and environmental data. Instead of dry theoretical content, students can access interactive maps, images, and videos, helping them better understand environmental issues and climate change. This spatial data visualization technology, in particular, shows modern, pervasive, and updated aspects.

* Modern: We are applying a new way (Redesign) to raising awareness through using ArcGIS StoryMap interactive digital technology on the ArcGIS Online cloud computing platform instead of conventional propaganda methods (posters, banners). This can be shared on all social platforms and digital devices suitable for the 4.0 era and GenZ generation; it meets the Resolution on digital transformation of Thua Thien Hue province.

* Pervasiveness: Promote community participation in joining hands to reduce SUP (Reduce) through interactive actions of sharing plastic reduction solutions such as information about recycling points, community product development, Environmentally friendly products, research projects to encourage participation, plastic-reducing tourist attractions, environmental sanitation/beach clean-up campaigns (eliminating plastic waste hotspots),...

* Updating: Information is continuously updated anytime, anywhere, and vividly visualized on the application; instant data reporting helps people gradually change their awareness about environmental protection and Minimize SUP use (Rethink).

- Integrating information and communication technology (ICT) in environmental education: ArcGIS StoryMap is a typical example of integrating ICT into education. It helps students grasp knowledge in a lively and interactive way, improving educational effectiveness and raising awareness of environmental issues among the younger generation.

- Action - and practice-oriented: ArcGIS StoryMap's focus on reducing single-use plastic does not stop at providing knowledge; this product also aims to change student behavior through lessons and project reality. This helps convert awareness into concrete action, contributing to reducing plastic pollution and protecting the environment.

- Creativity is shown in the aspects of (1) Combining multi-dimensional data and interactive stories, (2) Developing digital skills and creative thinking, (3) Practical applications in community projects, (4) Scalability and reuse, (5) Support for decision and policy-making.

4) Ability to apply socio-economic efficiency, development direction, and recommendations

The ArcGIS StoryMap plastic reduction story map product is a powerful and flexible tool that allows users to expand, update, and create interactive stories with images, text, and data maps. It facilitates application deployment in various educational contexts, from elementary to high school and university. Schools can integrate ArcGIS StoryMap into their Geography, Science, and Civics curriculum to convey knowledge about plastic pollution and environmental protection visually and vividly.

This product can be used in environmental clubs or student projects to encourage them to learn and take action for the environment. The app can also be used in public education events or exhibitions to raise awareness of plastic pollution.

For economic efficiency: Minimizing single-use plastic will save resources, reduce plastic production, and minimize waste management and pollution treatment costs. Furthermore, raising awareness and changing behavior from the student stage will help reduce long-term costs related to environmental pollution, thereby creating sustainable change toward a Green, Clean, and Bright Hue.

For social effectiveness: Educating students about the harmful effects of plastic and how to reduce it will create a generation conscious of protecting the environment. This benefits the environment and promotes sustainable lifestyles, building a more responsible and cohesive community towards sustainable urban development in the future.

Studies in educational settings reveal that GIS platforms like ArcGIS StoryMap can successfully incorporate digital storytelling into place-based learning, strengthening students' connections to the natural environment. This approach encourages a shift from conventional science instruction, fostering students' exploration and appreciation of ecological, cultural, and environmental connections. Similar research conducted at various educational institutions, including a Canadian study utilizing ArcGIS StoryMap to explore hibiscus plants in virtual classrooms, has shown that digital storytelling can cultivate a comprehensive understanding of the interdependence between nature and culture [3, 6].

IV. CONCLUSION

Plastic waste and single-use plastic are considered urgent global environmental problems, and immediate priority solutions are needed to reduce them to be appropriate to each context, age, and region. The research visualized the awareness and behavior of students regarding plastic reduction through interactive communication. It has initially been applied in practice, spreading and promoting the participation of students in schools and universities. Several high schools in the province, through the student network and Open day activities of the University of Sciences, Hue University, with more than 2,500 visits, participation in direct interactive games, and joining hands in the journey to reduce plastic of students and the community.

This product can edit, supplement, update, and display instant information quickly and effectively. It can also be used and interacted with on different devices to help communicate powerfully with everyone—students, from inside and outside the school, and even decision-makers. In the coming time, the research team will continue to update and supplement content, expand the application scale, seek funding cooperation sources, and evaluate and improve the model to meet the transformation era, to respond to climate change, protect the environment, and develop the economy. The application product creates a community factor in increasing awareness of SUPs among students (GenZ generation) first in the school, thereby encouraging and spreading to local high school students and other participants. Participate in taking positive actions to reduce single-use plastic in everyday life (behavioral change). Continue to promote the expansion of the application's scope of application from local schools to other

provinces and nationwide. Promote community participation in building a story map application about reducing single-use plastic.

In the following research phase, we will improve content quality through continued development of single-use plastic reduction ArcGIS StoryMap content to be more prosperous and engaging, including updating the latest data on plastic pollution, interactive questions about plastic awareness, lots of information about stores and organizations in Thua Thien Hue province responding to saying no plastic waste from SUPs, solutions, innovative models to reduce plastic.

A detailed plan for integrating technology-based strategies, such as using mobile apps for tracking recycling efforts, implementing interactive digital platforms like ArcGIS StoryMap for environmental education, and establishing a digital reward system to incentivize sustainable habits will be proposed in a framework for setting up these solutions in school settings, covering essential resources, staff roles, and student involvement. In addition, using artificial intelligence (AI) to build digital maps of plastic waste hotspots along rivers and coastal areas, digital maps of service providers such as "green" restaurants and hotels,... will be considered as contributing part towards the goal of Thua Thien Hue - Green City, becoming a centrally run city in the direction of: "Heritage, cultural, ecological, environmentally friendly and smart urban area" and meeting the needs of the city current digital transformation.

ACKNOWLEDGMENT

This research was implemented thanks to the support of funding from the scientific research project for students at Hue University of Science High School for Gifted Students, code 24/HD-DHKH, 2023.

REFERENCES

- [1] Basel Convention Secretariat & GRID-Arendal (2022), Plastic Waste on the Peak: How plastic waste came to pollute mountains and remote areas - and what we can do about it. GRID-Arendal.
- [2] Bruna Alves, 2023, Plastic waste worldwide - statistics & facts. Truy cập tại: <https://www.statista.com/topics/5401/global-plastic-waste/#topicOverview>, ngày truy cập 20/8/2023.
- [3] ESRI, (2022), ArcGIS StoryMaps Helps Students Show How Far Beach Litter Travels, ArcNEWS. Access on <https://www.esri.com/about/newsroom/arcnews/arcgis-storymaps-helps-students-show-how-far-beach-litter-travels/>, 20/06/2023.
- [4] ESRI (2023), What is ArcGIS StoryMaps? Access is on <https://doc.arcgis.com/en/arcgis-storymaps/get-started/what-is-arcgis-storymaps.htm>, and the date of access is 4/2023.
- [5] Meijer et al. (2021). More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean. Science Advances, <https://ourworldindata.org/plastic-pollution>
- [6] Poh Tan, David Zandvliet (2024), Accessible and Enriched Community-Engaged Learning: A Botanical Virtual Classroom Field Trip, International Conference Proceedings, New perspectives in science education, 13th Edition, Filodiritto Editore, DOI: 10.26352/I314_2384-9509.
- [7] Prime Minister (2019), Decision No. 1746/QD-TTg On promulgating the national action plan on ocean plastic waste management until 2030.
- [8] Prime Minister (2009). Decision No. 445/QD-TTg dated April 7, 2009, of the Prime Minister on Approving the adjustment of the master plan for the development of Vietnam's urban system to 2025 and vision to 2050.
- [9] Tatiana Hlinka, Chris Serrao, and Bani Meghadri (2021), Spoiled Oceans. Truy cập tai: <https://storymaps.arcgis.com/collections/fee26719ccf242b1b7dc9c096019c6df?item=>.

- [10] Tekman, M. B., Walther, B. A., Peter, C., Gutow, L. and Bergmann, M. (2022): Impacts of plastic pollution in the oceans on marine species, biodiversity, and ecosystems, 1–221, WWF Germany, Berlin. Doi: 10.5281/zenodo.5898684.
- [11] World Bank (2022a), Towards a national roadmap for single-use plastics in Vietnam - Strategic options to prioritize single-use plastic reduction, Ocean Plastics Report Series: East Asia and Pacific Region. 126 pages.
- [12] World Bank (2022b), Analysis of plastic pollution in Vietnam, Report series on Ocean Plastic: East Asia and Pacific Region. 137 pages.
- [13] Shah N. Selbe, 2021, Tracking Plastic Waste. Truy cập tại: <https://storymaps.arcgis.com/stories/6012a4f5e0c546bb8536fae9c9c9ed9c>.
- [14] Yamane, Taro. (1967), Statistics: An Introductory Analysis, 2nd Ed., New York: Harper and Row

Landslide Risk Mapping and development of landslide database and mobile app for Bhutan

Thongley

Department of Civil Engineering and Surveying
 Jigme Namgyel Engineering College
 Royal University of Bhutan
thongley21@gmail.com

Younten Tshering

Department of Information Technology
 Jigme Namgyel Engineering College
 Royal University of Bhutan
yountentshering.jnec@rub.edu.bt

Abstract— Bhutan experiences several landslides during the monsoon season. Although there are a few studies, they do not cover the entire country. Additionally, these maps are published in hard copy, making them less accessible for reference. There is also no comprehensive landslide database in Bhutan. Therefore, this study aims to produce a digital landslide risk map of Bhutan and create a database through a mobile app known as Bhutan Landslide Information (BLI). The landslide map was produced using twelve influencing factors and a landslide inventory collected through field visits and from relevant offices. A total of 2,324 landslide locations were documented, with 70% (1,620) used for training and 30% (704) for testing. The analysis was conducted using logistic regression. Before the regression model, tolerance (TOL) and Variance Inflation Factor (VIF) were used to check for collinearity. The analysis showed all factors had a VIF less than 10 and tolerance above 0.1, indicating none of the independent factors were redundant. According to the logistic regression model, Land Use Land Cover (LULC) had the highest coefficient of 0.8027, making it the most influential factor for landslide events. The landslide risk map was validated using sensitivity, accuracy, and Area Under Curve (AUC), with values of 0.8852 and 0.9457, respectively, indicating high accuracy. The landslide map was then exported to QGIS2Web to be made available online and incorporated into the BLI mobile app. Users can locate and view the landslide risk zone. Additionally, a landslide database was developed, allowing users to take photographs of landslide events through the BLI app. The app automatically records the landslide location (latitude and longitude) and the time the photograph was taken. Users can download the landslide database as a CSV file for research and decision-making purposes.

Keywords—Area Under Curve, Database, Landslide, Landslide risk mapping, Logistic regression, Mobile App

I. INTRODUCTION

Landslides are a frequent phenomenon in the mountainous nations of the Himalayan region, particularly during the monsoon season, constituting one of the most prevalent natural disasters. These occurrences lead to substantial losses in terms of property, crops, infrastructure, and human lives [1]. Recent years have seen a rise in the frequency of landslide incidents due to climate change and human activities [2]. According to a report prepared by [3], landslides are ranked as the 7th most dangerous natural disaster worldwide. However, in the fragile Himalayan region, landslides are considered one of the top disasters significantly impacting daily life, especially during the monsoon season [4]. As per the international disaster database [5], landslides accounted for the loss of 13,214 lives, adversely affecting 4,255,459 people between 2000 and 2023. The economic damage in Asian countries alone reached a staggering US\$2,450,085.

Bhutan's challenging topography, characterized by steep mountainous terrain, makes it more particularly vulnerable to landslides, especially during the monsoon season [6]. The landslide poses a significant threat to the country's

infrastructure and the well-being of the people. The escalating impact of climate change, intensified by various human activities, has led to an increase in both the frequency and severity of landslides in Bhutan in recent years [7]. The majority of landslides in the Himalayan region are primarily triggered by intense rainfall. However, limited studies have been conducted at a large scale covering the entire Bhutan [8].

A landslide claimed the lives of 10 people in the northern district of Gasa on June 16, 2021. These individuals were tragically killed while collecting cordyceps [9]. Several landslide incidences killed many people in Bhutan apart from disturbances in the daily schedules due to landslides. The frequent casualties from landslides in Bhutan underscore the alarming nature of this situation, calling for heightened attention and preventive measures in the vulnerable country of Bhutan.

The comprehensive digital landslide risk map for Bhutan plays a crucial role in the mitigation and reduction of landslide disasters in landslide-prone places. The digital landslide map which is stored in a mobile app will help individual and relevant implementing agencies for strategic planning and preparedness. In addition, the landslide database incorporated in the mobile app known as Bhutan Landslide Information (BLI) will help future researchers with the accessibility of landslide data.

The objectives of this study were to develop a digital landslide risk map of Bhutan and a landslide database. Another objective is to develop a mobile app known as Bhutan Landslide Information (BLI) which will display a landslide risk map and store landslide database. The landslide database will be like the open data concept as recommended in [10].

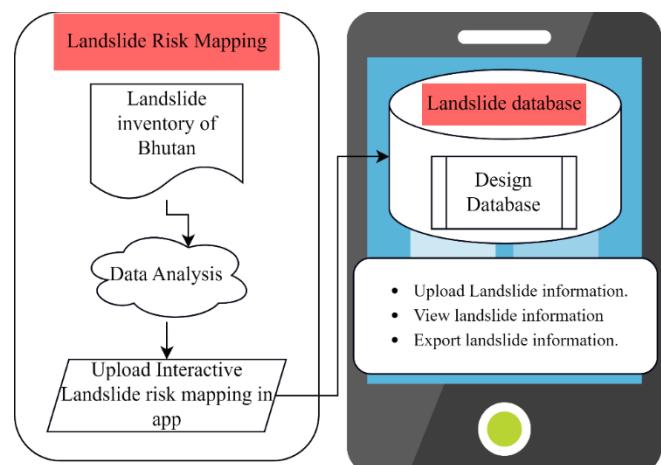


Fig. 1. Conceptual framework of research

II. STUDY AREA

Bhutan (Fig. 2) is a landlocked country located in the eastern Himalayas. It is bordered by China to the north and

India to the south. The country is relatively small, covering an area of 38,394 square kilometers. Bhutan's geographical coordinates are approximately 26.3475°N-28.4086°N latitude and 88.7516°E-92.1332°E longitude. Bhutan constitutes a fragile landscape and experiences heavy rainfall during the monsoon season. Frequent landslide during monsoon affects daily work in some parts of the country.

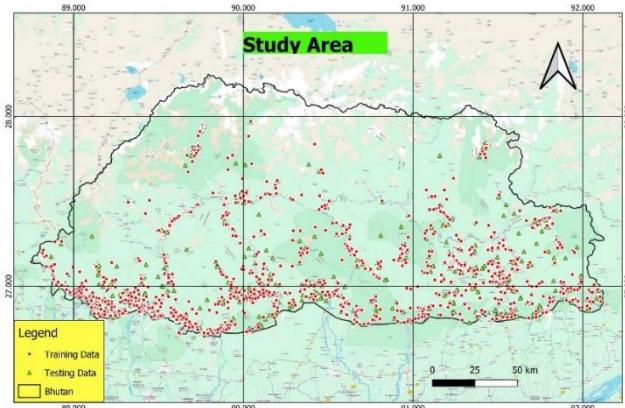


Fig. 2. Existing landslide locations in Bhutan

Landslides pose significant risks to lives, infrastructure, and the environment. However, the absence of a comprehensive landslide risk map and database hinders effective monitoring, early warning, and mitigation efforts. To address this gap, the BLI mobile app was developed. This app aims to provide a centralized platform for collecting, storing, and disseminating landslide information, facilitating timely responses and informed decision-making to reduce landslide risks.

III. METHODOLOGY

The overall methodology of the research is shown in Fig. 3. The workflow consists of landslide inventory, factor preparation, developing a landslide risk map, its validation, designing a database, and mobile app development.

A. Landslide inventory and preparation of factors

Landslide inventory consists of collecting landslide locations using different methods such as collecting its location from high-resolution satellite imageries, and site visits using handheld GPS. A total of 2324 landslide locations (Fig. 2) were collected as a landslide inventory which will be used for data training and validation purposes. 70% (1627 landslide locations) of total landslide and an equal number of non-landslide points were used as a dependent factor. The remaining 30% (697 landslide locations) were used for validation purposes. The study used different independent factors such as aspect map, elevation map, slope map, rainfall map, stream power index map (SPI), soil map, land use land cover (LULC) map, topographic wetness index (TWI), drainage map, geology, road map, and Normalized Difference Vegetation Index (NDVI). The SRTM DEM was used to derive an aspect map, elevation map, slope map, rainfall map, stream power index map (SPI), topographic wetness index (TWI), and drainage map. NDVI map was derived from NIR and Red band of Landsat 8 OLI. Similarly, the LULC map was derived from the classification of Landsat 8 imageries. The geological map was shared by the Department of Geology and the Mine of Bhutan.

B. Multicollinearity diagnosis

The logistic regression model is sensitive to collinearity among the independent factors, and it is important to check the collinearity among the independent factors [11]. Multicollinearity is caused by the high correlation between the independent factors [12]. The tolerance (TOL) and variance inflation factor (VIF) are commonly used to check multicollinearity. The TOL and VIF are calculated using Equations 1 and 2. The factors are said to be non-collinear when the TOL is more than 0.1 and the VIF is less than 5 [13] and redundant factors were discarded.

$$\text{TOL} = 1 - R^2 \quad (1)$$

$$\text{VIF} = \frac{1}{\text{TOL}} \quad (2)$$

C. Logistic Regression Model

The logistic regression model gives good results for landslide mapping [14] and many researchers used the logistic regression model for landslide studies [15, 16]. Logistic regression is commonly used in landslide studies analysis. The logistic regression requires several independent factors on a single dichotomous dependent factor [17]. The single dichotomous factors contain binary points (1 and 0) whereby 1 indicates a landslide point and an equal number of 0 indicates a non-landslide point. The logistic regression evaluates the correlation between the landslide event and the influencing factors [18].

The probability (p) of the landslide in logistic regression is calculated using Equation 3

$$P = \frac{1}{1 + e^{-z}} \quad (3)$$

Where p is the probability of landslide which ranges between 0 to 1 on an S-shaped curve. Z represents a linear combination.

$$Z = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_n x_n \quad (4)$$

where b_0 is the intercept of the model, b_i ($i = 0, 1, 2, \dots, n$) represents the coefficients of the logistic regression model, and x_i ($i = 0, 1, 2, \dots, n$) denotes the independent factors.

D. Accuracy assessment

The accuracy assessment is a critical step in validating the reliability and effectiveness of the generated landslide risk map [19]. Accuracy and Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve was employed for the validation purpose. The AUC and accuracy were calculated using Equation 7 and Equation 8 respectively. The ROC curve illustrates a False Positive Rate (FPR) on the x-axis and a True Positive Rate (TPR) on the y-axis. The accuracy and AUC values are interpreted as excellent (0.9-1.0), very good (0.8-0.9), good (0.7-0.8), moderate (0.6-0.7), and poor (0.5-0.6) [20]. The True Positive Rate (TPR) and False Positive Rate (FPR) are calculated using Equation 5 and 6.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (6)$$

$$\text{AUC} = \frac{\sum \text{TP} + \sum \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

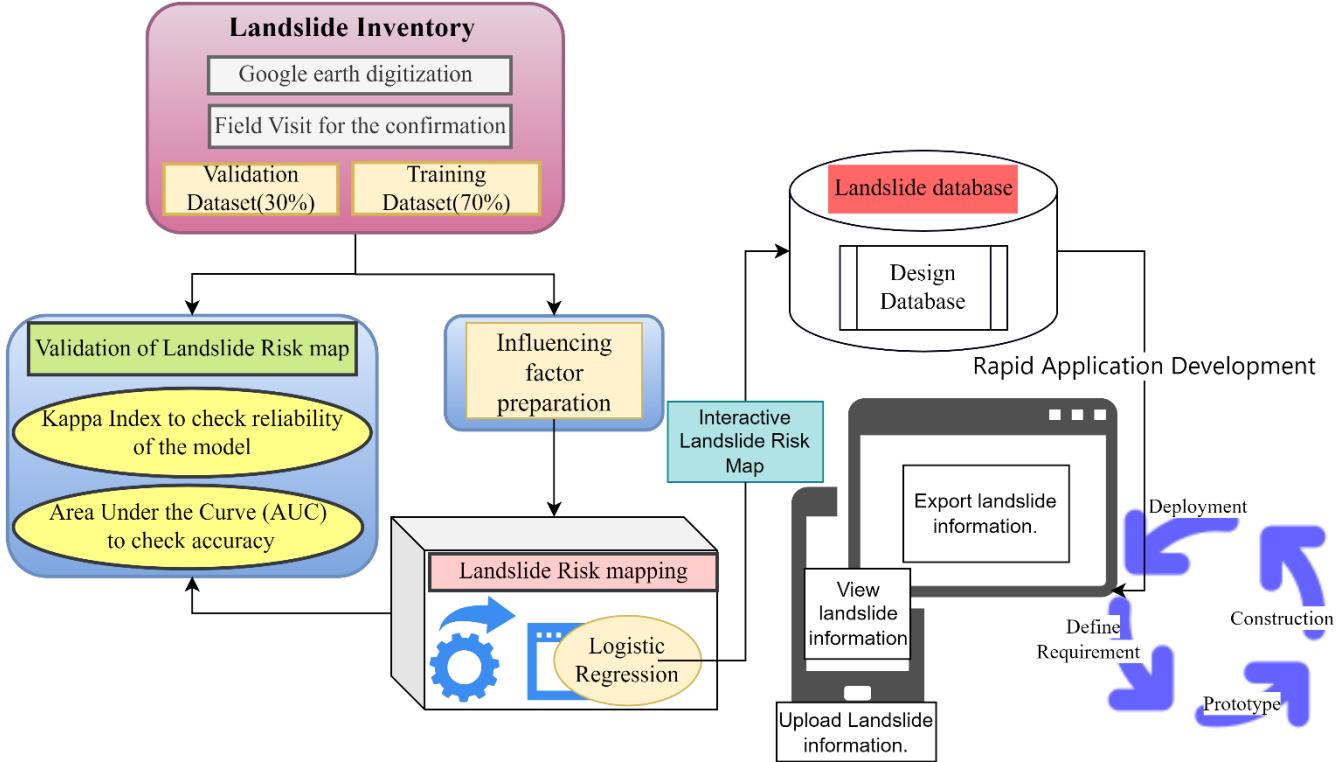


Fig. 3. Overall workflow for the research.

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

E. Database Design and Development Process

An accurately designed database is vital for maintaining up-to-date and correct information. Initially, the landslide database was intended to be stored in MySQL, but Firebase [21] was chosen for its ability to handle large volumes of data and provide real-time updates.

F. Mobile App Development Using the RAD Model

The System Development Life Cycle (SDLC) provides a structured approach to developing information systems. In this project, the SDLC was integrated with the RAD model as shown in Fig. 3 to ensure a comprehensive development process. As per [22], the Rapid Application Development (RAD) model is a popular approach for software development, especially when quick iterations and user feedback are essential.

In [23], the project integrated SDLC with the Code Fix approach and for this project, the SDLC was integrated with the RAD model where the project benefitted from a structured development approach while maintaining the flexibility to incorporate user feedback and make iterative improvements. This method ensured the successful development of the mobile app and its underlying database, providing a strong solution for landslide risk management in Bhutan.

IV. RESULTS AND DISCUSSION

A. Multi-collinearity Diagnosis and Landslide Risk Map

Multicollinearity was assessed using Tolerance (TOL) and Variance Inflation Factor (VIF). TOL and VIF are calculated using Equation 1 and Equation 2 respectively. Factors were deemed non-collinear if TOL exceeded 0.1 and VIF was below 5. Table I displays the coefficient, TOL, and VIF for each independent factor. It is observed that all independent

factors exhibit no signs of collinearity since the TOL is more than 0.1 and the VIF is less than 5. All the factors are within the range for TOL and VIF. Therefore, none of the factors were discarded.

TABLE I. COEFFICIENT OF FACTORS AND THE VALUE OF TOLERANCE (TOL) AND VARIANCE INFLATION FACTOR (VIF)

Factors	Coefficient	TOL	VIF
Constant	0.363		
Aspect	0.001	0.99	1.01
Drainage	-0.082	0.8	1.25
Elevation	-0.862	0.26	3.86
Geology	0.137	0.51	1.95
LULC	0.803	0.77	1.29
NDVI	-0.128	0.68	1.47
Rainfall	0.157	0.58	1.72
Road	-0.231	0.77	1.3
Slope	0.094	0.88	1.14
Soil	-0.007	0.89	1.12
SPI	0.027	0.96	1.05
TWI	-0.01	0.96	1.04

The coefficients presented in Table I were employed in the computation of the linear combination Z using Equation 3. This coefficient value will subsequently serve as the basis for determining the landslide probability P using Equation 3 and, ultimately, for generating a landslide risk map.

The coefficients from Table I were applied in Equation 4 and Equation 3 to produce the landslide risk map as depicted in Fig. 4. Utilizing the Jenk classification method, known for its clear breakpoints between classes, the landslide risk map was categorized into five zones: very low (8.76%), low (21.05%), moderate (30.91%), high (29.85%), and very high (9.44%). It is noted that the majority of the area falls within the moderate risk zone, aligning with the observations made during ground truthing. The result from the landslide map

prepared by [24] generated using the Frequency Ratio Model is similar to the generated map for this study.

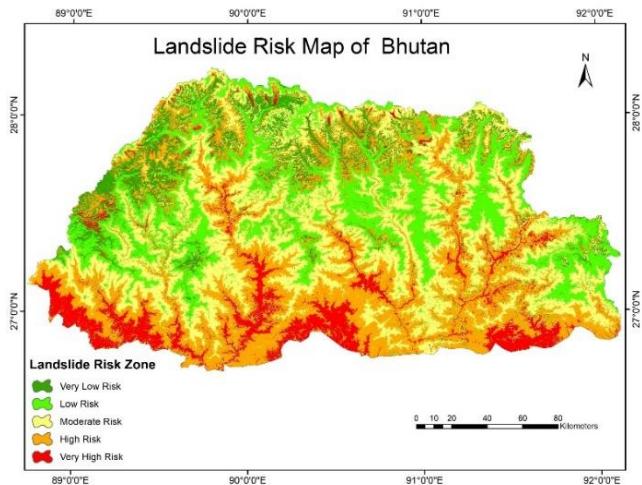


Fig. 4. Landslide Risk Map of Bhutan.

B. Validation result of the landslide risk map

Validation was performed using 30% of the total landslide data, employing two methods namely, accuracy and AUC. The accuracy is calculated using Equation 8 and the accuracy of the landslide risk map is 0.8852. Likewise, the prediction rate is deduced from the AUC value, computed using Equation 7. This AUC is also considered the prediction rate and it reflects the likelihood of future landslide occurrences. The AUC for the landslide risk map is 0.9457 as shown in Fig. 5. Following [20], the accuracy is classified as "very good (0.8-0.9)" while the AUC is categorized as "excellent (0.9-1.0)" for the prediction of landslides.

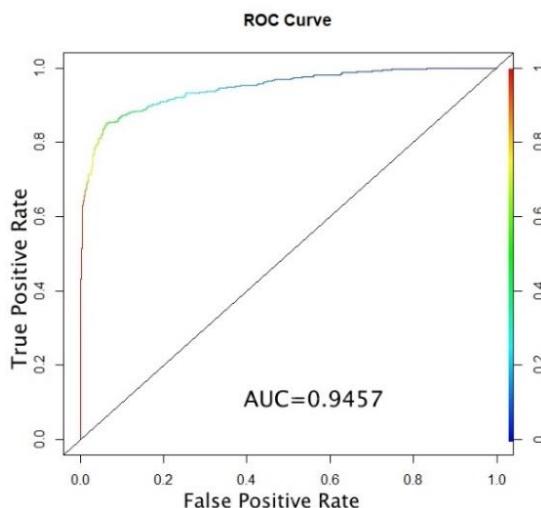


Fig. 5. ROC Curve for Prediction rate curve

C. Bhutan Landslide Information Mobile App

The "Bhutan Landslides Information" (BLI) mobile app has been designed with a user-friendly interface to ensure ease of use for all users. The interface as shown in Fig. 6 is intuitive and straightforward, providing clear navigation and accessible features. Key elements of the interface include:

- Interactive Maps:** The app features interactive maps that allow users to view the exact location of a landslide and check high-risk zones. This feature ensures accurate spatial data collection and improves the

reliability of the information stored in the database. Thus, when the user clicks on Risk Map on the main interface then the user can view the risk map in detail.

- User Input Forms:** The app includes well-structured forms as shown in Fig. 6 for users to input landslide information. These forms include fields for uploading an image, selecting the location via GPS, selecting the date and time, entering remarks, choosing a district from dropdown lists, and entering a phone number (optional) and email ID (optional). The forms are designed to be simple and easy to fill out, minimizing the time required to report a landslide.
- Search and Filter Options:** Users can view landslide information based on the district. The search and filter options are easy to use, allowing users to quickly find relevant information about the landslide and view detailed information once clicked on the screen.
- Data Export:** The app includes a feature to download the collected data. This functionality is crucial for further analysis and reporting, making the data easily accessible and usable for different purposes.

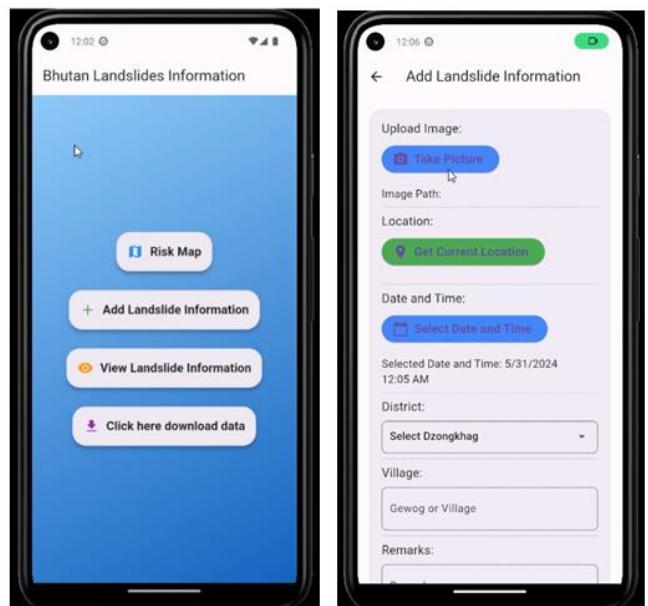


Fig. 6. Main interface

The development of the mobile app and its integration with Firebase has successfully addressed the need for a comprehensive landslide risk map and database in Bhutan. The app's user-friendly interface, real-time data synchronization, and robust database design have ensured the fulfillment of the project's objectives. This initiative represents a significant step forward in enhancing landslide risk management and improving the safety and resilience of communities in Bhutan.

To improve the BLI mobile app's usability, functionality, and applicability, user testing was implemented and carried out with selected stakeholders. This process was implemented to ensure the app meets the needs of users engaged in landslide risk management. A feedback survey, administered via Google Forms, captured perspectives on the app's user-friendliness, accuracy, and potential impact on decision-making during the monsoon season. Survey results indicated strong support for the app's digital landslide risk map and data

upload feature, with an average user-friendliness rating of 4.7/5 and a usefulness rating of 4.9/5. Additionally, stakeholders suggested valuable feature improvements, such as an offline data upload mode, more detailed risk maps, and real-time notifications.

V. CONCLUSION

The study aimed to assess the severity of landslides in various places in Bhutan, generating a landslide risk map to provide informative insights. A landslide risk map covering the entire Bhutan using twelve independent factors and landslide inventory data. The landslide inventory was prepared from high-resolution satellite images, relevant offices, and on-site visits. The landslide data were partitioned into 70% for training independent factors and 30% for validation purposes.

The landslide risk map was prepared employing a logistic regression model. The generated landslide risk map classified landslides into five zones using natural break classification. The percentages of different zone areas are 9.44% for very high-risk zones, 29.85% for high-risk zones, 30.91% for moderate-risk zones, 21.05% for low-risk zones, and 9.76% for very low-risk zones. Validation utilized accuracy and AUC methods, yielding an accuracy of 0.8851 and an AUC of 0.9457, with 1 representing a perfectly accurate map. The AUC value also informs the prediction rate for the generated map. Since the value of accuracy and AUC are closer to 1, the generated map is considered to be perfect. Similarly, the ground truthing also shows correct information for most places in the high-risk zone.

The generated map will help decision makers and planners for infrastructure development such as planning of new road construction. The generated map is expected to be uploaded in a mobile app whereby users can locate their geolocation. The geolocation can help users to identify their landslide risk zone.

This study addressed the critical need for comprehensive and accessible landslide risk information in Bhutan by developing a digital landslide risk map and database through the BLI mobile app. Users can download the BLI mobile app to contribute landslide information, which will record the landslide location information. Researchers and decision-makers can use this data for their work. Additionally, users can view the landslide map, which will display different risk levels corresponding to their location, allowing them to identify the landslide risk zone in their area.

ACKNOWLEDGMENTS

The researchers extend our sincere thanks to the Coalition for Disaster Resilient Infrastructure (CDRI) for providing financial support for the project (Application ID: 2212081599). The sincere gratitude also goes to Jigme Namgyel Engineering College allowing us to use laboratory facilities and for granting leave for site investigations and meetings with various agencies. Further, the researchers also thank the Department of Geology and Mine, Bhutan for sharing landslide data. Additionally, the researchers are grateful to our mentor, Dr. Tshewang Lhendup (Ph.D.), President of Jigme Namgyel Engineering College under the Royal University of Bhutan for his mentorship.

REFERENCES

- [1] A. N. Khan, A. E. Collins, and F. Qazi, "Causes and extent of environmental impacts of landslide hazard in the Himalayan region: a case study of Murree, Pakistan," *Natural Hazards*, vol. 57, pp. 413-434, 2011, doi: 10.1007/s11069-010-9621-7.
- [2] C. Huggel, J. J. Clague, and O. Korup, "Is climate change responsible for changing landslide activity in high mountains?," *Earth Surface Processes and Landforms*, vol. 37, no. 1, pp. 77-91, 2012, doi: 10.1002/esp.2223.
- [3] M. Dilley, *Natural disaster hotspots: a global risk analysis*, vol. 5, World Bank Publications, 2005.
- [4] A. Dikshit, R. Sarkar, B. Pradhan, S. Acharya, and A. M. Alamri, "Spatial landslide risk assessment at Phuentsholing, Bhutan," *Geosciences*, vol. 10, no. 4, Article 131, 2020, doi: 10.3390/geosciences10040131.
- [5] EM-DAT, *The International Disaster Database*, 2020.
- [6] T. Thongley and C. Vansarochana, "Landslide identification and zonation using the index of entropy technique at Ossey Watershed Area in Bhutan," *Applied Environmental Research*, vol. 43, no. 1, pp. 102-115, 2021, doi: 10.35762/AER.2021.43.1.8.
- [7] N. Chettri, K. Tempa, L. Gurung, and C. Dorji, "Association of climate change to landslide vulnerability and occurrences in Bhutan," in *Impact of Climate Change, Land Use and Land Cover, and Socio-economic Dynamics on Landslides*, pp. 3-37, 2022, doi: 10.1007/978-981-16-7314-6_1.
- [8] A. Dikshit, R. Sarkar, B. Pradhan, S. Segoni, and A. M. Alamri, "Rainfall-induced landslide studies in the Indian Himalayan region: a critical review," *Applied Sciences*, vol. 10, no. 7, Article 2466, 2020, doi: 10.3390/app10072466.
- [9] D. Gyelmo, "Are Bhutan's monsoon landslides becoming more deadly?," *Third Pole*, 2023. [Online]. Available: <https://www.thethirdpole.net/en/climate/bhutan-landslides-climate-change/>.
- [10] Y. Tshering and C. Anutariya, "Enabling semantic interoperability in Bhutan's e-government: An ontology-based framework," in *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Bangkok, Thailand, 2022, pp. 1-6, doi: 10.1109/JCSSE54890.2022.9836286.
- [11] C. H. Mason and W. D. Perreault Jr., "Collinearity, power, and interpretation of multiple regression analysis," *Journal of Marketing Research*, vol. 28, no. 3, pp. 268-280, 1991, doi: 10.1177/002224379102800302.
- [12] J. I. Daoud, "Multicollinearity and regression analysis," in *Journal of Physics: Conference Series*, vol. 949, no. 1, p. 012009, IOP Publishing, 2017.
- [13] D. T. Bui et al., "Flood spatial modeling in northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model," *Remote Sensing*, vol. 11, no. 13, Article 1589, 2019, doi: 10.3390/rs11131589.
- [14] Lee, S. A. R. O. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, 26(7), 1477-1491.
- [15] Zhu, L., & Huang, J. F. (2006). A GIS-based logistic regression method for landslide susceptibility mapping in regional scale. *Journal of Zhejiang University-Science A*, 7(12), 2007-2017.
- [16] Lin, L., Lin, Q., & Wang, Y. (2017). Landslide susceptibility mapping on a global scale using the method of logistic regression. *Natural Hazards and Earth System Sciences*, 17(8), 1411-1424.
- [17] D. T. Bui, T. A. Tuan, H. Klempe, B. Pradhan, and I. Revhaug, "Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree," *Landslides*, vol. 13, no. 2, pp. 361-378, 2016, doi: 10.1007/s10346-015-0557-6.
- [18] X. Sun et al., "Landslide susceptibility mapping using logistic regression analysis along the Jinsha river and its tributaries close to Derong and Deqin County, southwestern China," *ISPRS International Journal of Geo-Information*, vol. 7, no. 11, Article 438, 2018, doi: 10.3390/ijgi7110438.
- [19] M. Sciarra, L. Coco, and T. Urbano, "Assessment and validation of GIS-based landslide susceptibility maps: a case study from Feltrino stream basin (Central Italy)," *Bulletin of Engineering Geology and the Environment*, vol. 76, pp. 437-456, 2017, doi: 10.1007/s10064-016-0954-7.
- [20] D. M. Zhao et al., "Comparative performance assessment of landslide susceptibility models with presence-only, presence-absence, and

- pseudo-absence data," *Journal of Mountain Science*, vol. 17, no. 12, pp. 2961-2981, 2020.
- [21] W.-T. Sung, I. G. T. Isa, and S.-J. Hsiao, "An IoT-Based Aquaculture Monitoring System Using Firebase," *Computers, Materials & Continua*, vol. 76, no. 2, pp. 2185-2218, 2023.
- [22] Y. Tshering, S. P. Kamishetty, and H. Sarwarzadah, "Event management for social service website using Ruby on Rails: BTO event view application developed implementing collaborative technique," *World Journal of Research and Review*, vol. 13, no. 6, pp. 1-9, Dec. 2021, doi: 10.31871/WJRR.13.6.5.
- [23] Y. Tschering, S. Dorji, and T. Wangmo, "User-friendly Android-based contact application: Case of accessible contact in educational organization," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 9, pp. 1803-1808, 2022.
- [24] Thongley, T., & Vansarochana, C. (2021). Landslide susceptibility assessment using frequency ratio model at Ossey watershed area in Bhutan. *Engineering and applied science research*, 48(1), 56-64.

Deep Residual Neural Networks with Self-Attention for Landslide Susceptibility Mapping in Uttaradit Province, Thailand

1st Kritchayan Intarat

*Research Unit in Geospatial Applications (Capybara Geo Lab)
Faculty of Liberal Arts,
Thammasat University
Pathumthani, Thailand, 12121
intaratt@tu.ac.th*

<https://orcid.org/0000-0002-0728-753X>

2nd Nithima Nuengjamnong

*Research Unit in Geospatial Applications (Capybara Geo Lab)
Faculty of Liberal Arts
Thammasat University
Pathumthani, Thailand, 12121
nithima.n@arts.tu.ac.th*

<https://orcid.org/0000-0002-5665-0737>

3rd Jojinda Sae-Jung

*Research Unit in Geospatial Applications (Capybara Geo Lab)
Faculty of Liberal Arts
Thammasat University
Pathumthani, Thailand, 12121
jinda.s@arts.tu.ac.th*

<https://orcid.org/0009-0006-0425-7550>

4th Woraman Jangsaeng

*Center of Excellence in Digital Earth Applications and Emerging Technology (CoE: DEET), Sirindhorn International Institute of Technology
Thammasat University
Pathumthani, Thailand, 12120
m6622040878@g.siit.tu.ac.th*

<https://orcid.org/0009-0004-6351-8787>

5th Patimakorn Yoomee

*Research Unit in Geospatial Applications (Capybara Geo Lab)
Faculty of Liberal Arts,
Thammasat University
Pathumthani, Thailand, 12121
st124146@ait.ac.th*

<https://orcid.org/0009-0006-2940-8351>

6th Teerapong Panboonyuen

*Research Unit in Geospatial Applications (Capybara Geo Lab)
Faculty of Liberal Arts,
Thammasat University
Pathumthani, Thailand, 12121
teerapong.pa@chula.ac.th*

<https://orcid.org/0000-0001-8464-4476>

Abstract—Landslides pose serious threats, causing infrastructural damage and economic losses. Accurate prediction of landslide susceptibility is crucial for risk management. Traditional methods often fail to handle the complex, nonlinear relationships among geological, hydrological, and meteorological factors. This paper proposes an enhanced ResNet152 model incorporating multihead self-attention mechanisms to improve landslide susceptibility mapping in Uttaradit Province, Thailand, using geospatial data. The integration of self-attention allows the model to capture intricate dependencies within the input data, focusing on critical features that traditional convolutional layers might miss. This enhancement leads to a performance boost, with the ResNet152+Att model achieving a precision of 91.44%, recall of 91.74%, F1-score of 91.59%, and ROC-AUC of 0.9678, outperforming the baseline ResNet152. Using self-attention in deep residual networks significantly improves model accuracy and interpretability, providing a more robust framework for landslide susceptibility assessment. The implementation code is available at <https://kaopanboonyuen.github.io/geoai-landslides>.

Index Terms—Landslide, Deep learning, ResNet, Attention mechanism, Uttaradit

I. INTRODUCTION

Landslides are natural hazards that can cause significant damage to infrastructure, loss of lives, and economic setbacks. Various factors, such as geological, hydrological, and meteorological factors, influence their frequency and intensity [1]. Practical assessment and prediction of landslide susceptibility are crucial for risk mitigation and management. Traditional

methods for landslide susceptibility assessment, including statistical and heuristic approaches, have shown limitations in handling the complex, nonlinear relationships among the contributing factors [2].

Recently, deep learning (DL) methods have been making waves across various fields [3], [4], including natural hazard assessment. Deep learning, a subset of machine learning, excels in automatically learning representations from data, making it a decisive tool for spatial and temporal prediction tasks [5]. Convolutional Neural Networks (CNNs), a well-known DL architecture, have effectively handled spatial data and exhibited promise in various geospatial applications, especially in landslide susceptibility prediction [1], [6]. Residual Network, or “ResNet,” introduced by [7], is a DL architecture that significantly mitigates the vanishing gradient problem, enabling the training for very deep networks [8]. The key innovation in ResNet is using identity shortcut connections that bypass one or more layers, allowing the network to learn residual functions relative to the layer inputs. This approach has been highly effective, as demonstrated by ResNet models like ResNet50, ResNet101, and ResNet152, which have achieved state-of-the-art results in numerous benchmarks, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [7], [8].

In recent studies, ResNet architectures have shown remarkable performance improvements in various fields, including

image recognition [7], medical imaging [9], and image segmentation [10]. Introducing identity mappings in ResNet101 and ResNet152 further enhanced their capability to learn deeper representations [8]. Advanced training techniques have also been proposed to optimize these networks [11], and the impact of residual connections has been widely studied [12]. In the case of landslide susceptibility, previous studies have revealed that such models outperform traditional methods, leveraging their superior feature extraction capabilities [13], [14]. Integrating ResNet into landslide susceptibility assessment frameworks can thus significantly enhance prediction accuracy and reliability [15].

This study aims to improve the ability to predict landslide susceptibility by enhancing the ResNet model in Uttaradit Province, located in the lower-northern region of Thailand. First, we intend to capture the intricate patterns and interactions among multiple landslide conditioning factors using the set of ResNet models. The ResNet 50, ResNet101, and ResNet152 models will be trained and validated using a comprehensive dataset, including topographical, geological, and hydrological variables and historical landslide occurrences, and statistically evaluated to deliver the best-predicted model of landslide. Next, the attention module will be added to the best ResNet model called “residual attention network.” Finally, the predicted results will be compared using statistical testing. The results of this study may contribute to the existing knowledge on landslide susceptibility assessment and provide a robust framework for hazard mitigation planning.

II. MATERIAL AND METHODS

A. Study area: Uttaradit, Thailand

Uttaradit Province is located in the lower northern region of Thailand (Fig. 1). It has an area of approximately 7,838 km². The topography can be divided into three types: 1) Lowland Areas along the Nan River, which are low-lying areas on both sides, making up about 20% of the total area. 2) Flat areas between valleys along the Khlong Tron, Pad River, and various streams, interspersed with mountainous terrain, accounting for another 20% of the total area. 3) Mountainous and high-altitude regions located in the north and east of the province, comprising approximately 60% of the total area. Given these physical characteristics, Uttaradit Province has a high potential for landslides. The average annual rainfall is approximately 1,300 – 1,400 mm, and the yearly temperature is around 27°C.

B. Dataset used in this study

- Landslide occurrence information was imported from the Department of Mineral Resources, which includes 4,192 recorded landslide occurrence points over the past 30 years from 1989 to 2019 in Uttaradit Province. Non-landslide occurrence points were generated by calculating a 1-km buffer from each landslide point [16]. Such a technique can ensure a balanced dataset to match the number of landslide occurrences. 70% of both data sets will be employed for model training and the remaining 30% for model validation.

- Intrinsic factors incorporated with the landslide susceptibility are those influenced by natural phenomena and anthropogenic disturbance. Natural factors input into the computations were the digital elevation model (DEM), slope, aspect, curvature, geology type, stream power index (SPI), topographic wetness index (TWI), topographic ruggedness index (TRI), and soil, normalized difference vegetation index (NDVI), normalized difference water index (NDWI). Additionally, the distance from streams and the distance from faults are considered. The anthropogenic disturbance factors include the distance from transportation.

C. Multicollinearity

Since independent variables may cause multicollinearity issues in linear models, an occurrence where two or more independent variables are highly correlated [17] can lead to inaccuracies in predicting the probability of landslide occurrences within the model. This issue can be resolved by analyzing the variance inflation factor (VIF) (1) and tolerance (TOL). These methods eliminate multicollinearity in linear models by testing the relationships between the independent variables.

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (1)$$

VIF_i stands for the variance inflation factor for the ith variable, and R_i² is the coefficient of determination (R₂) for the ith variable. Generally, an acceptable VIF value is 10, and a TOL value is 0.10 [18]. Independent variables with a VIF more significant than ten may indicate multicollinearity issues in the linear model [19]. The selected factors are the DEM, slope, aspect, curvature, SPI, distance from streams, faults, transportation, and geology. These will be used as variables to train the models to assess landslide susceptibility in the study area.

D. Residual Network and Residual Attention Network models

Residual network (ResNet) architecture revolutionized the field of DL by handling the vanishing gradient matter and empowering the deep networks’ training [7]. ResNet’s core concept lies in using residual blocks, which include shortcut connections that bypass one or more layers. This approach addresses the vanishing gradient problem by allowing gradients to flow directly through these shortcuts during back-propagation. These connections pass the input directly to subsequent layers without modifications (identity mappings) or apply a linear transformation if the dimensions differ [8]. The architecture of ResNet supports significant depth, with models such as ResNet50, ResNet101, and ResNet152, which consist of layers according to their numbers, respectively.

Apart from ResNet itself, the residual attention network (Fig. 2) combines the strengths of residual learning and attention mechanisms. It consists of multiple attention modules stacked together, allowing the model to focus on relevant features while mitigating the vanishing gradient problem common

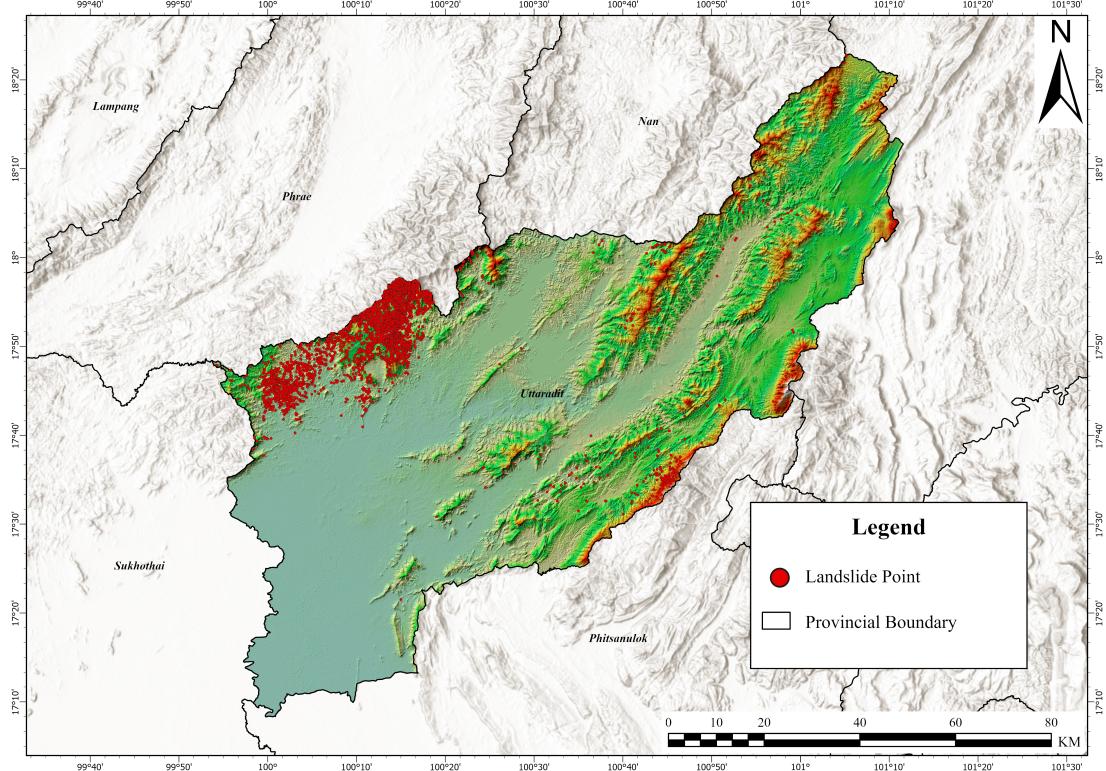


Fig. 1. Study location (Uttaradit Province, Thailand).

in deep networks [20]. The attention mechanism dynamically weights the importance of different spatial regions, which is crucial for tasks requiring detailed spatial analysis e.g. landslide susceptibility prediction [21]. Identical hyperparameters were set for every DL model as in Table I.

TABLE I
MODEL'S STATISTICAL PERFORMANCE

Hyperparameters	Value
Bach_size	2,048
Epoch	2,500
Learning Rate	0.001
Loss Function	Cross Entropy
Optimizer	Adam

E. Self-Attention Mechanism

The self-attention mechanism has been widely adopted in various DL models due to its capability to capture long-range dependencies and enhance feature representation. In our study, we integrate self-attention into deep residual neural networks to improve the performance of landslide susceptibility mapping using geospatial data.

The self-attention mechanism operates by computing the attention scores between all pairs of elements in an input sequence. This allows the model to weigh the importance of each element relative to others, thus capturing contextual information effectively. The attention scores are calculated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \quad (2)$$

where Q (queries), K (keys), and V (values) are matrices derived from the input features, and d_k is the dimensionality of the keys. The scaled dot-product attention calculates the similarity between the queries and keys, scales the result by the square root of the critical dimensionality to prevent tremendous values, and applies the softmax function to obtain the attention weights. These weights are then used to produce a weighted sum of the values [22].

To integrate self-attention into the deep residual network, we modify the residual blocks by incorporating self-attention layers. Specifically, given an input feature map X , the modified residual block can be expressed as:

$$\hat{X} = \text{Attention}(XW_Q, XW_K, XW_V), \quad (3)$$

$$Y = \text{ReLU}(\hat{X} + X), \quad (4)$$

where W_Q , W_K , and W_V are learnable weight matrices for the queries, keys, and values, respectively. The output of the self-attention mechanism \hat{X} is added to the original input X through a residual connection, followed by a ReLU activation function.

Incorporating self-attention into our deep residual network allows the model to focus dynamically on the most relevant geospatial features for landslide susceptibility mapping. This

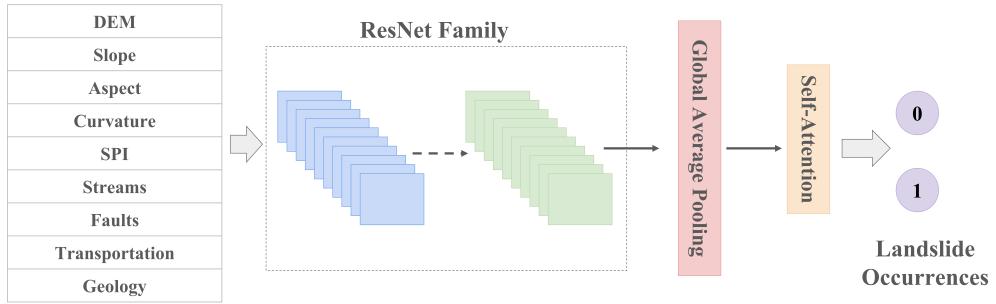


Fig. 2. The architecture of residual attention network.

results in enhanced feature representations and improved predictive performance.

III. RESULTS AND DISCUSSIONS

During the prediction stage, the four susceptibility DL models were categorized into five classes: very high, high, moderate, low, and very low. This classification followed the natural breaks approach [23], [24]. The performance of the different models was evaluated using metrics e.g. precision, recall, F1 score, area under the receiver operating characteristic curve (ROC), and training time. The results are outlined in Table II and Fig. 3.

TABLE II
MODEL'S STATISTICAL PERFORMANCE

Model	Precision	Recall	F1-score	ROC	Time
ResNet50	89.93%	89.15%	89.54%	0.9622	108 min 47 s
ResNet101	89.09%	90.77%	89.92%	0.9612	164 min 15 s
ResNet152	90.92%	91.45%	91.18%	0.9645	213 min 28 s
ResNet152+Att	91.44%	91.74%	91.59%	0.9678	361 min 33 s

The ResNet50 model achieved a precision of 89.93%, recall of 89.15%, F1-score of 89.54%, and ROC of 0.9622, with a training time of 108 min and 47 s. These results indicate robust performance in predicting landslide susceptibility with relatively fast training time compared to the other models. The ResNet101 model showed slightly improved recall and F1-score compared to ResNet50, with values of 90.77% and 89.92%, respectively, and a precision of 89.09%. The ROC for ResNet101 was 0.9612, and the training time increased to 164 min and 15.08 s, reflecting the increased complexity of the model.

The ResNet152 model exhibited improved performance over ResNet101, achieving a precision of 90.92%, recall of 91.45%, F1-score of 91.18%, and a ROC of 0.9645, with a training time of 213 min and 27.99 s, indicating a further increase in training time due to the more profound architecture [7], [25]. Adding attention layers to the ResNet152 model (ResNet152+Att) resulted in the highest performance across all evaluated metrics. The ResNet152+Att model achieved a precision of 91.44%, recall of 91.74%, F1-score of 91.59%, and ROC of 0.9678. However, this model required the longest training time, taking

361 min and 33.08 s. These results indicate that incorporating attention mechanisms enhances the model's ability to focus on critical features, thus improving its predictive accuracy and robustness in landslide susceptibility mapping. The landslide susceptibility maps from DL models are displayed in Fig. 4

The significant increase in training time for deeper models and those incorporating attention mechanisms highlights the trade-off between model complexity and computational efficiency. Although the performance of the ResNet152+Att model has improved from ResNet152, with the ROC increasing from 0.9645 to 0.9678 (an increase of approximately 0.34%) and the F1 score increasing from 91.18% to 91.59% (an increase of approximately 0.45%), it requires the most training time. Specifically, the ResNet152+Att model's training time of 361 min and 33.08 s is about 1.7 times longer than the ResNet152 model's training time of 213 min and 27.99 s.

To mitigate these computational demands, applying pruning techniques to ResNet152+Att could effectively reduce the number of parameters and inference time while preserving accuracy. Filter pruning removes less important filters in convolutional layers and can maintain the model's architecture and predictive capability. On the other hand, unstructured pruning can eliminate individual weights and further reduce model size. Pruning allows for a more efficient model optimized for faster deployment without significantly sacrificing accuracy. Despite this, performance improvement is not significantly more significant than other models. Therefore, models like ResNet50, ResNet101, and ResNet152 are more suitable, balancing prediction accuracy and computational resource usage. Pruning techniques make these models even more feasible for real-world, time-sensitive applications by minimizing computational demands. This balance is critical in real-world scenarios where timely and efficient processing is essential, ensuring that even with slightly lower accuracy, the models can still deliver reliable results within a reasonable timeframe.

The predictions from four models indicate that areas highly susceptible to landslides are primarily located in the western part of the province, particularly in the Eastern Phi Pan Nam Mountains region. These mountains are characterized by limestone, and historically, numerous landslides have occurred. The frequency of landslide occurrences significantly influences the

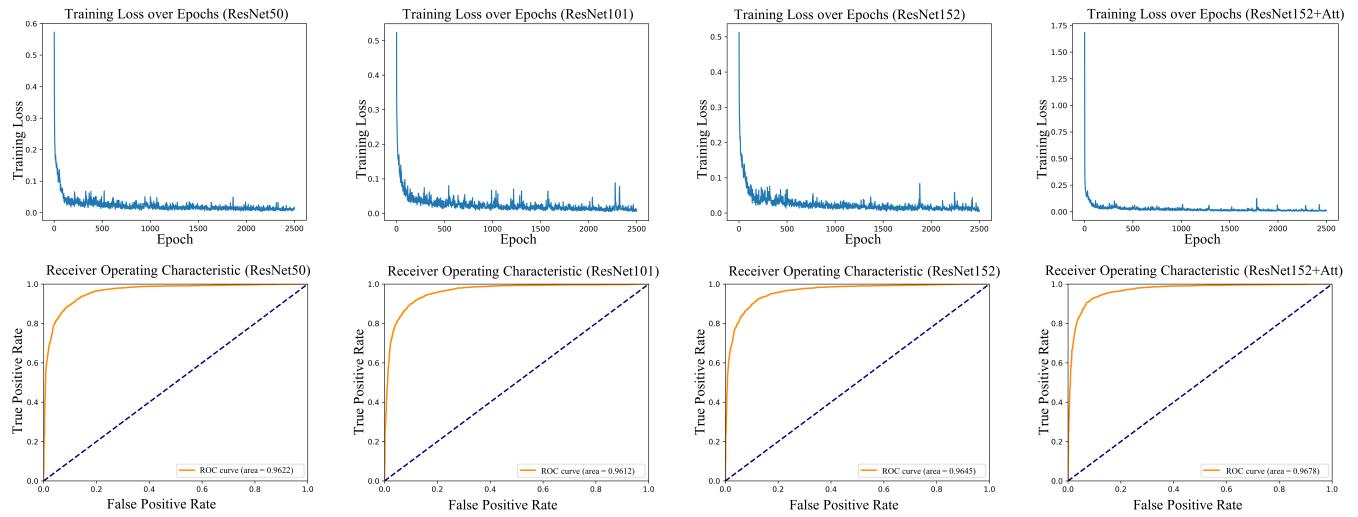


Fig. 3. Loss and ROC curves from the DL models.

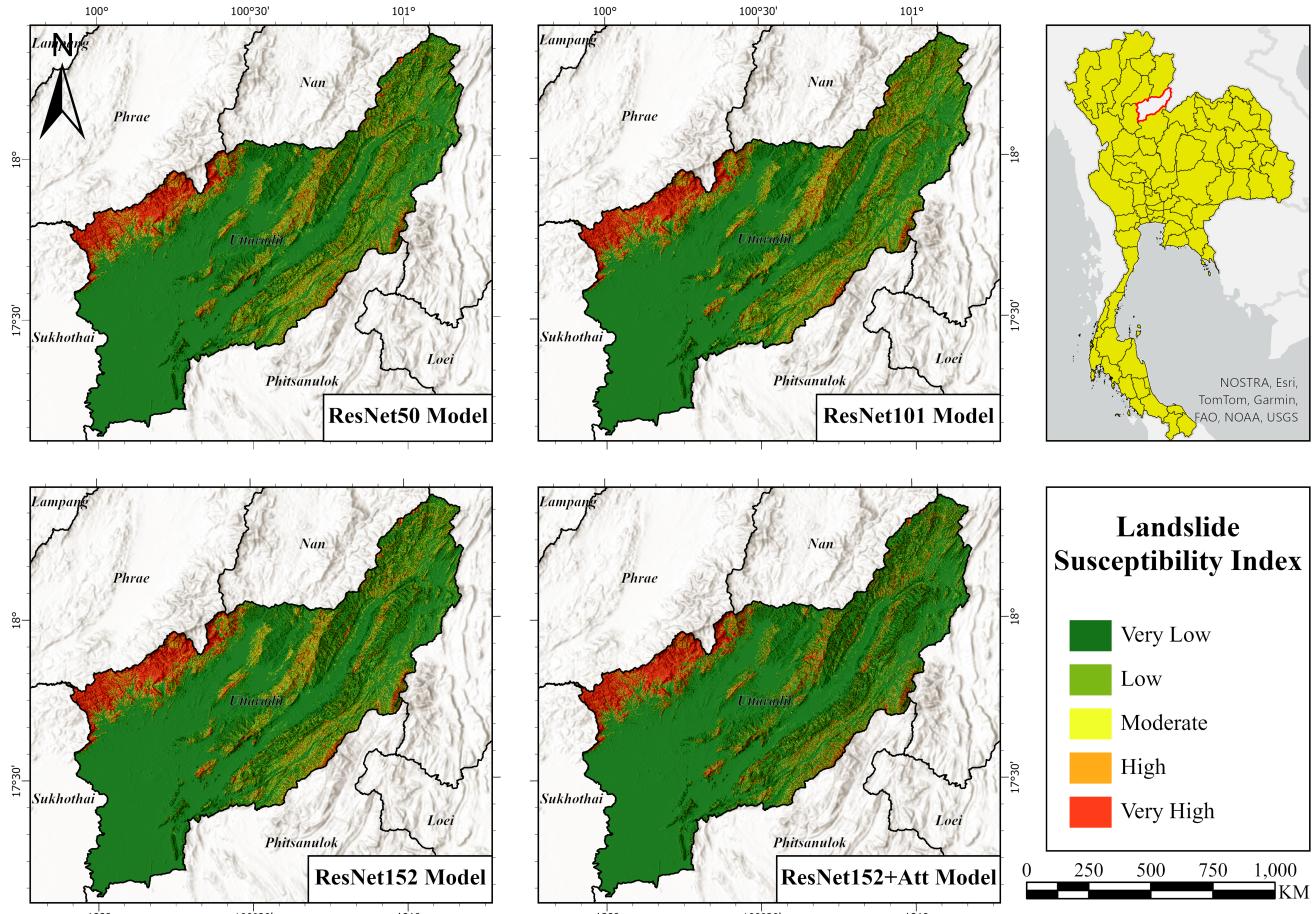


Fig. 4. Landslide susceptibility maps obtained from DL models.

model's predictions [1]. Additionally, the factors contributing to landslide risk in this area are mainly intrinsic [24], which were selected for the comprehensive data analysis. The only extrinsic factor considered is the distance from roads. Thus, landslide sensitivity is primarily based on physical factors of the area.

According to the region's physical characteristics, it is evident that the western part of the province consists of complex mountainous terrain undergoing physical weathering of rocks. As previously mentioned, it is evident that weathering and erosion processes drive landslide occurrences. The weathering process changes the rock's exterior size and shape without altering its internal composition. Subsequently, erosion and weathering cause the rock material to break down. Debris and various materials are then transported by gravitational forces, accumulating in lower-lying areas and creating landslide-prone zones [19].

IV. CONCLUSION

In this paper, the results reveal that the more profound the ResNet model, the more improved the prediction performances of landslide susceptibility. Notably, the ResNet152+Att outperformed the baseline ResNet152 model, achieving a higher F1-score and ROC. However, this model required significantly more training time. While the ResNet152+Att model offers the best accuracy, ResNet50 and ResNet101 provide a more practical balance of performance and training efficiency. These findings highlight the importance of considering accuracy and efficiency when selecting models for real-world predictions. Future work should integrate other models, e.g., Grad-CAM (Gradient-weighted Class Activation Mapping) or SHAP (SHapley Additive exPlanations) and multi-temporal analysis to improve predictions over time.

ACKNOWLEDGMENT

This research was supported by the Faculty of Liberal Arts, Thammasat University, Research Unit in Geospatial Applications (Capybara Geo Lab), and partially funded by the Faculty of Liberal Arts' Research Fund, Thammasat university, Fiscal year 2566, Contract No. 9/2566.

REFERENCES

- [1] M. Azarafza, M. Azarafza, H. Akgün, P. M. Atkinson, and R. Derakhshani, "Deep learning-based landslide susceptibility mapping," *Scientific reports*, vol. 11, no. 1, p. 24112, 2021.
- [2] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, and F. Guzzetti, "A review of statistically-based landslide susceptibility models," *Earth-science reviews*, vol. 180, pp. 60–91, 2018.
- [3] T. Panboonyuen, N. Nithisopa, P. Pienroj, L. Jirachuphun, C. Watthanasisirikrit, and N. Pornwiriyakul, "Mars: Mask attention refinement with sequential quadtree nodes for car damage instance segmentation," in *International Conference on Image Analysis and Processing*. Springer, 2023, pp. 28–38.
- [4] T. Panboonyuen, "Semantic segmentation on remotely sensed images using deep convolutional encoder-decoder neural network," 2019.
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [6] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE geoscience and remote sensing magazine*, vol. 5, no. 4, pp. 8–36, 2017.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [8] ———, "Identity mappings in deep residual networks," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 2016, pp. 630–645.
- [9] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.
- [13] I. Yilmaz, "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from kat landslides (tokat—turkey)," *Computers & Geosciences*, vol. 35, no. 6, pp. 1125–1138, 2009.
- [14] M. Abedini, B. Ghasemian, A. Shirzadi, H. Shahabi, K. Chapi, B. T. Pham, B. Bin Ahmad, and D. Tien Bui, "A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment," *Geocarto International*, vol. 34, no. 13, pp. 1427–1457, 2019.
- [15] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, and F. Guzzetti, "A review of statistically-based landslide susceptibility models," *Earth-science reviews*, vol. 180, pp. 60–91, 2018.
- [16] Y. Huan, L. Song, U. Khan, and B. Zhang, "Stacking ensemble of machine learning methods for landslide susceptibility mapping in zhangjiajie city, hunan province, china," *Environmental Earth Sciences*, vol. 82, no. 1, p. 35, 2023.
- [17] H. Yu, W. Pei, J. Zhang, and G. Chen, "Landslide susceptibility mapping and driving mechanisms in a vulnerable region based on multiple machine learning models," *Remote Sensing*, vol. 15, no. 7, p. 1886, 2023.
- [18] D. T. Bui, O. Lofman, I. Revhaug, and O. Dick, "Landslide susceptibility analysis in the hoa binh province of vietnam using statistical index and logistic regression," *Natural hazards*, vol. 59, pp. 1413–1444, 2011.
- [19] H. Yu, W. Pei, J. Zhang, and G. Chen, "Landslide susceptibility mapping and driving mechanisms in a vulnerable region based on multiple machine learning models," *Remote Sensing*, vol. 15, no. 7, p. 1886, 2023.
- [20] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.
- [21] Y. Ma, S. Xu, T. Jiang, Z. Wang, Y. Wang, M. Liu, X. Li, and X. Ma, "Learning a deep attention dilated residual convolutional neural network for landslide susceptibility mapping in hanzhong city, shaanxi province, china," *Remote Sensing*, vol. 15, no. 13, p. 3296, 2023.
- [22] S. Khallaghi, "Advancing the use of deep learning techniques for earth observation and environmental monitoring," Ph.D. dissertation, Clark University, 2024.
- [23] T. Zeng, B. Jin, T. Glade, Y. Xie, Y. Li, Y. Zhu, and K. Yin, "Assessing the imperative of conditioning factor grading in machine learning-based landslide susceptibility modeling: a critical inquiry," *Catena*, vol. 236, p. 107732, 2024.
- [24] M. A. Hussain, Z. Chen, R. Wang, S. U. Shah, M. Shoaib, N. Ali, D. Xu, and C. Ma, "Landslide susceptibility mapping using machine learning algorithm," *Civ. Eng. J.*, vol. 8, pp. 209–224, 2022.
- [25] J. Park, J.-k. Kim, S. Jung, Y. Gil, J.-I. Choi, and H. S. Son, "Ecg-signal multi-classification model based on squeeze-and-excitation residual neural networks," *Applied Sciences*, vol. 10, no. 18, p. 6495, 2020.

The Development of Satellite-based Active Fire Database and Hotspot Clustering to Enhance Forest Fire Management Practices in Thailand

1st Woranut Chansury
Geo-Informatics and Space Technology Development Agency (Public Organization)
 Bangkok, Thailand
 woranut@gistda.or.th

5th Chatchaya Buaniam
Department of National Parks, Wildlife and Plant Conservation
 Bangkok, Thailand
 kunying_b@hotmail.com

2nd Nuntikorn Kitratporn
Geo-Informatics and Space Technology Development Agency (Public Organization)
 Bangkok, Thailand
 nuntikorn.kit@gistda.or.th

6th Itthirit Pintong
Department of National Parks, Wildlife and Plant Conservation
 Bangkok, Thailand
 ightplaza555@gmail.com

3rd Anusorn Rungsimanich
Geo-Informatics and Space Technology Development Agency (Public Organization)
 Bangkok, Thailand
 anusorn50@gistda.or.th

7th Nattanon Chaisak
Department of National Parks, Wildlife and Plant Conservation
 Bangkok, Thailand
 chwarzkiserz@outlook.com

4th Teansiri Moolchan
Geo-Informatics and Space Technology Development Agency (Public Organization)
 Bangkok, Thailand
 teansiri@gistda.or.th

Abstract— This research aims to develop centralized database of satellite-based active fire hotspot and develop hotspot clustering algorithm to support firefighting operation. The first result is the active fire hotspot database which is automatically retrieving data from 4 satellites under Moderate Resolution Imaging Spectroradiometer: MODIS (Aqua and Terra) and Visible Infrared Imaging Radiometer Suite: VIIRS (Suomi NPP and NOAA20). The second result is the hotspot clustering model based-on DBSCAN technique which showed 99% agreement with the location of actual fire event. The percentage of overlapping area was 76.72% in forest. The clustering of hotspots will assist in identifying extensive fire zones for forest rangers, facilitating their entry to these regions to extinguish the fire and avert its propagation. This strategy will alleviate the effects on forest regions, fauna, and ecosystems. Furthermore, at a national level, it can facilitate policy-making by assisting agencies in prioritizing and distributing resources for efficient fire management.

Keywords— Active fire hotspot, Fire clustering, DBSCAN, Forest fire control

I. INTRODUCTION

Forest fires significantly accelerate the decline of wooded areas in Thailand, with evidence showing that human activities are the main trigger for wildfires. The intensity of these flames has profoundly disrupted the natural balance. Wildfires in Thailand primarily occur from December to April, with a peak frequency in March. The northern region is the most affected, leading to persistent smoke and increased particulate matter concentrations that detrimentally affect respiratory health.

The increasing number of wildfires and affected areas have been reported over the years, indicating a worsening situation. The data on fire suppression from 2017 to 2020 reveals a consistent surge in firefighting operations on wildfires. In 2017, 4,650 activities covered an area of 120.67 km². Although the number of operations in 2018 decreased to 3,786 with an area of 89.23 km², the number rose again in 2019, to 7,323, covering 242.84 km². The number of

firefighting operations further reached 7,550, occupied over 279.70 km² in 2020 [1] (Fire Control Department, 2020).

The Department of National Parks, Wildlife, and Plant Conservation utilizes satellite data to monitor fires, specifically the active fire hotspots. Historically, hotspot detection relied on Moderate Resolution Imaging Spectroradiometer (MODIS) product, which has a spatial resolution of 1,000 meters. Currently, hotspot data with an improved spatial resolution of 375 meters from Visible Infrared Imaging Radiometer Suite (VIIRS) is employed to enhance monitoring capabilities. VIIRS surpasses MODIS in detecting smaller and less severe fires, while still provides sub-daily observation. However, the improvement in spatial resolution resulted in a drastic increase of hotspots which overwhelm firefighting operation. Although a single hotspot pinpoints a fire locations, multiple hotspots spatially close together may indicate large fire events. Such knowledge is essential for resource allocation and planning during operation. Despite the existing fire alert system, such as NASA FIRMS, the hotspots has yet to be further analyzed, hindering the delivery of actionable information to personnel in the field during major fires in remote forest areas.

To enhance fire control efficacy, a centralized fire data system integrating satellite information and additional spatial analysis should be implemented. This study aims to develop and assess hotspot clustering methods, as well as deploy the result for operational uses. This facilitates strategizing, assessing situations, prioritizing activities, and performing firefighting operations effectively.

II. MATERIALS AND METHODS

A. Satellite Data on Active Fire Hotspots

The VIIRS, as technically described by [2], is used to gather hotspots data for active fire clustering. The VIIRS sensors are onboard Suomi-NPP, NOAA-20, and NOAA-21, with a plan to be installed with new satellite into a series of frequent and long-term observations. The active fire hotspots data is provided via the Fire Information for Resource

Management System (FIRMS) which is a data collection system that provides up-to-date and comprehensive global information on hotspots. FIRMS [3] is a software application developed by the University of Maryland and funded by NASA and the Food and Agriculture Organization of the United Nations. The hotspots datasets were obtained via FIRMS for this study.

B. Centralize Database for Active Fire Hotspots

Precise fire data is crucial for further research and the development of forest fire innovations. After establishing data pipeline from FIRMS system, a local centralized data system was developed. The original dataset was assessed for high quality flag before its observation timestamp is converted to local time zone. The coordinate system is modified to Universal Transverse Mercator (UTM) utilizing the World Geodetic System 1984 (WGS84) spheroid and datum, specifically in Zone 47. The hotspots were analyzed with land use to identified potential land use related to each hotspot. This process provide detailed information on significant places and forestry agencies' jurisdictional limits in different regions. Lastly, data dictionary was created that provides an elaborate description of the dataset's structure and organization. To ensure that both the data owner and users have a comprehensive understanding of the data and can use it effectively, specify the fields' titles, data formats, and data width capacity.

C. Hotspot cCustering with Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Various data clustering techniques are available and consider as one type of unsupervised machine learning [4],[5]. The density-based method, specifically DBSCAN, is widely applied for hotspots clustering, such as in Canada [6] and Indonesia [7].

DBSCAN method described by [8] requires optimization of two key hyperparameters, namely the epsilon (ϵ) value, which is the search radius, and the minimum number of points needed to make a cluster (minPts). The DBSCAN algorithm (Fig. 1) functions by randomly selecting an unexamined area and assessing the quantity of data points inside the specified search radius (ϵ). If the number of data points within the designated radius meets or exceeds the minimal threshold (minPts), we designate the location under consideration as a core point and mark it as verified. The same approach is then employed to evaluate all locations within the designated radius until the process is complete. Border points are designated as locations within the ϵ radius of a core point that possess fewer data points than the minPts criterion. Points that are neither core nor border as noise are considered as noise. Finally, a distinct cluster ID is assigned to each cluster.

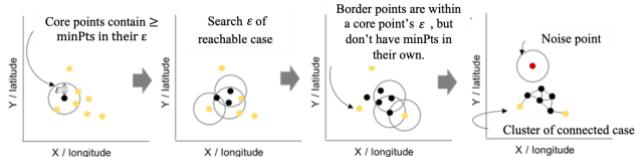


Fig. I. Clustering based on density: DBSCAN

The optimal search radius can vary depending on the objective, [9] previously evaluated suitable radius, ranges from 750 to 1,500 meters. In this study, the diagonal distance

between the centroid of the VIIRS satellite data grid was calculated. The uniform geographic grid dimension of 375 meters per grid was determined. According to the data presented in Fig. 2, the optimal radial distance between hotspot coordinates must exceed 530 meters but not beyond 750 meters ($530.3 < \epsilon \leq 750.0$). These values were evaluated for optimal setting. The minimum number of points (minPts) is set to be greater than or equal to the data's dimensionality. This study utilizes coordinates X (longitude) and Y (latitude) for calculations, establishing a minimum of 3 places for analysis. Additionally, temporal duration between detected hotspots should be considered also since the VIIRS datasets repeated multiple times a day. The period of 12-hour and 24-hour were examined to determine whether how many repeated observation should be integrated.

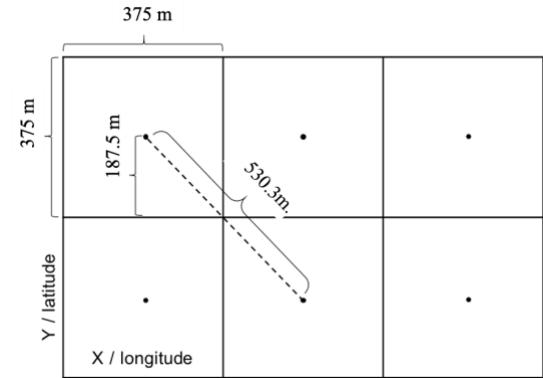


Fig. II. An active cluster of fires is situated in the middle of a 375×375 meter grid, as defined by the dimensions of the VIIRS satellite, which is used to determine the search radius (ϵ).

D. Hotspot Cluster Delineation

To delineate the area of hotspot clusters from DBSCAN, convex hull algorithm is used. A study by [10] applied convex hull to hotspot cluster points and suggested it to be an efficient approach. The convex hull converts data points within the same clusters from a point representation to a polygon representation. This involves identifying the minimal boundary that encompasses all points without any extensions. After computing the convex hull, we incorporate a buffer of 187.5 meters to match the dimensions of the data grid. Fig. 3 illustrates the convex hull and buffer processes.

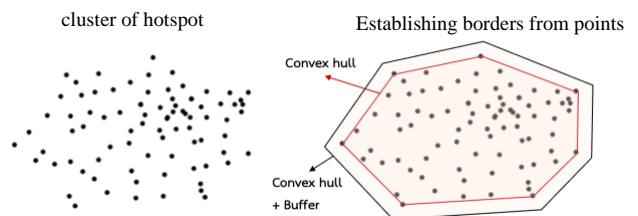


Fig. III. Convex Hull Algorithm

E. Validation of Hotspot Clusters

The model validation entails comparing the percentage of overlapped area of clustered active fire hotspots with the burned regions derived from Sentinel-2 satellite imagery. Sentinel-2 is a high-resolution satellite, with a spatial resolution of 10 to 60 meters. Equation 1 determines the overlap percentage.

$$\text{Overlap Area (\%)} = \frac{\text{Brunt Area}_{\text{sentinel-2}}}{\text{Burn Area}_{\text{cluster}}} \times 100 \quad (1)$$

The burned area was then identified using the Normalized Burn Ratio (NBR) index from before and after fire event and calculated the change. Equation 2 derives the NBR index from the short-wavelength infrared bands, while Equation 3 calculates burn index difference, or Different Normalized Burn Ratio (dNBR).

$$\text{NBR} = \frac{\rho_{\text{SWIR1}} - \rho_{\text{SWIR2}}}{\rho_{\text{SWIR1}} + \rho_{\text{SWIR2}}} \quad (2)$$

Where ρ_{SWIR1} : B08 - Sentinel-2 spectral bands
 ρ_{SWIR2} : B12 - Sentinel-2 spectral bands

$$d\text{NBR} = \text{NBR}_{\text{pre-fire}} - \text{NBR}_{\text{post-fire}} \quad (3)$$

A total of 54 validate samples of DBSCAN-based hotspot clustering of VIIRS dataset were used. The pre-fire images were from within a month before and post-fire images captured within 14 days following hotspot detection.

Additionally, active fire hotspots covered a period of five years, from 2017 to 2021, comprising a total of 799,809 hotspots were obtained. The developed DBSCAN methodology was applied, and a statistical analysis was conducted. Descriptive statistical analysis involves summarizing the characteristics of hotspot clustering. This entails calculating the mean, minimum, maximum, variability, and median, in addition to analyzing the distribution of hotspot clusters across the 5 years.

III. RESULTS

A. Database of Satellite-based Active Fire Hotspot

A database system integrates data about active fires from satellites was developed. Although this study uses only VIIRS-based observation, the centralized system was developed to support all available active fire hotspot dataset. This dataset consists of data gathered by four satellites: Aqua and Terra from the MODIS system, and Suomi-NPP and NOAA-20 from the VIIRS system, utilizing automated data entry, as illustrated in Table I.

TABLE I. SUMMARIZES THE ACQUISITION AND STORAGE OF HOTSPOT DATA FROM THE AQUA, TERRA, SUOMI-NPP AND NOAA-20 SATELLITES.

Sensor/Satellite	Launch Year	Image Acquisition Frequency	Data Source
MODIS/AQUA	2002	daily	NASA
MODIS/TERRA	1999	daily	NASA
VIIRS/Suomi-NPP	2011	daily	NASA
VIIRS/NOAA-20	2017	daily	NASA

B. The DBSCAN-based Hotspot Clustering

The DBSCAN algorithm was employed to create an active fire clustering model using active fire data from the VIIRS system. Several hyperparameter settings were tested based on 5-year hotspot dataset from VIIRS hotspots observed by Suomi-NPP satellite and the optimal values were determined to be a search radius of 550 meters, a minimum of 3 hotspots per cluster, a temporal occurrence of 12-hour

period. Fig. 4 illustrates an example of the clustering results, with the aggregated heat spots indicated in red and a convex hull boundary of 187.5 meters indicated in orange.

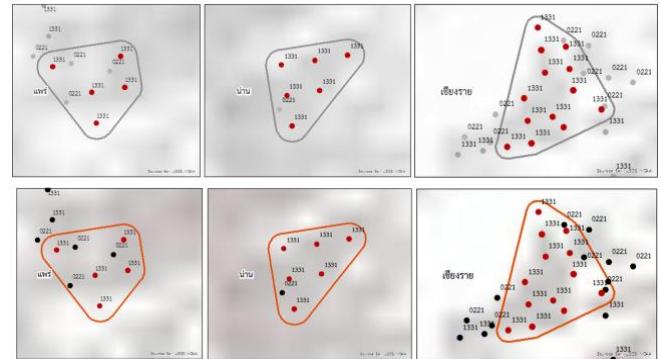


Fig. IV. On March 26, 2023, the DBSCAN algorithm was employed to cluster the hotspots derived from the VIIRS satellite system.

C. Validation of the Active Fire Hotspot Clustering

The validation process included assessing the degree of overlap between active fire clusters and burn scar regions retrieved from Sentinel-2 satellite. The results showed that 99% of all samples were found with areas that had burn scars. Table 2 displays the results of the area overlap percentage, providing a detailed breakdown of the average, middle value, and variability. The forest regions exhibited an average overlap percentage of 76.72. The median value suggests that most hotspot clusters had an overlapped area larger than that indicated by the mean value.

TABLE II. THE PERCENTAGE OVERLAP OF REGIONS BETWEEN ACTIVE FIRE CLUSTERS AND BURN SCARS OBSERVED IN SENTINEL-2 SATELLITE IMAGERY.

	Average	Median	Standard Deviation
Forest Area	76.72	83.50	25.35

Fig. 5 illustrates the examples validation between hotspot clusters, delineated by red boundary, and the burned scars from Sentinel-2, shown in white. The results show that hotspot clusters often cover larger areas than the actual burn scars. This is because the spatial resolution of the hotspot used for clustering isn't very good at finding small changes in space, like ravines. Although fires may occur simultaneously in proximity, agricultural plots are dispersed over slopes and do not constitute broad, connected areas. This behavior causes clusters to form, even among separate fires, resulting in a reduced overlap ratio compared to clusters in forested areas. Therefore, this clustering method is not suitable for agricultural fire events and should be strictly applied for wildfire occurred in the forested areas.

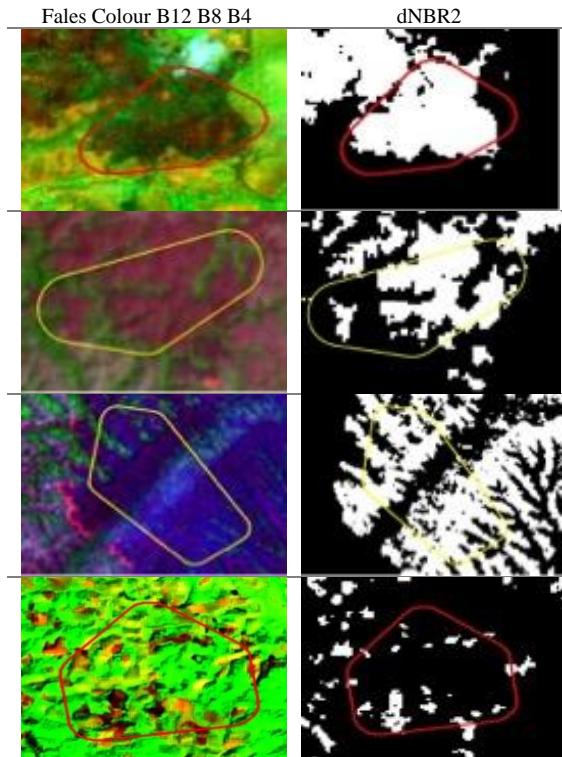


Fig. V. Example of burn scar calibration in forested regions from active fire clusters in northern provinces

D. Statistical analysis of DBSCAN active fire clusters

The research team utilized hotspot data gathered from the VIIRS onboard the Suomi-NPP satellite system over a five-year period, from 2017 to 2021. The aggregate total of hotspot accrued throughout this period amounted to 799,809.

Table III demonstrates that with a search radius (ϵ) of 550 meters, the model produced a total of 49,479 clusters comprising of 204,879 active fire hotspots. The allocation of hotspots among clusters shows the mean of roughly 4 points per cluster, whereas both the minimum and median are 3 points per cluster. The maximum number of hotspots within a single cluster is 109 points.

TABLE III. THE STATISTICAL VALUES FOR THE NUMBER OF HOTSPOTS PER CLUSTER USING A SEARCH RADIUS OF 550 METERS FOR 5 YEARS FROM 2017 TO 2021.

Value Statistic	Result
Number of Cluster	49,479
Active Fire Hotspot (Mean)	4.141
Standard Deviation	2.308
Minimum	3.000
25 Percentile	3.000
50 Percentile	3.000
75 Percentile	4.000
Maximum	109

The statistical study revealed that the quantity of hotspots generated by clustering via the DBSCAN method exhibits a positively skewed distribution, as illustrated in Fig. 5. Nearly 50% of the clusters have only 3-4 points each. Using different search radius ($530.3 < \epsilon \leq 750.0$) in experiments showed that the statistical distribution stayed mostly the same, even though the total number of clusters changed. Nonetheless, employing a wider search radius of 750 meters yielded an increased number of clusters and affected their sizes, ultimately leading to an overestimation. Consequently, the research team selected a search radius of 550 meters.

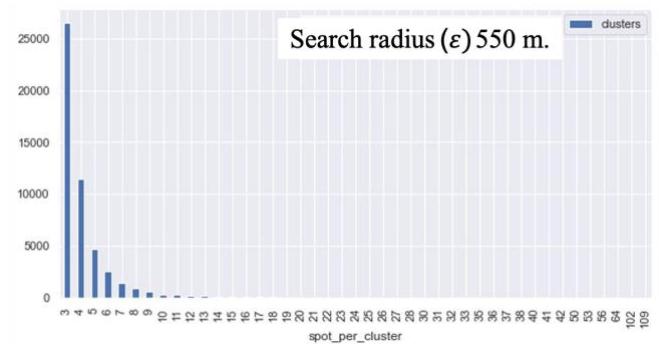


Fig. VI. A histogram of the frequency distribution of hotspot within clusters, using a search radius of 550 meters. The x-axis represents the number of cluster, while the y-axis indicates the number of hotspots within clusters.

The spatial distribution of hotspot clusters for the preceding five years (2017–2021) was analyzed, as shown in Fig. 6. It was found that Samut Songkhram and Phuket were the provinces without any hotspot clusters. Provinces with a minimal number of clusters encompassed those in the southern region, along with Chanthaburi and Trat in the eastern region. In contrast, Mae Hong Son province had a disproportionately high density of hotspot clusters relative to its area, signifying a serious condition. Provinces exhibiting a significant concentration of hotspot clusters comprised Chiang Mai, Chiang Rai, Phayao, Nan, Lamphun, Lampang, Phrae, Tak, Sukhothai, Uttaradit, Kamphaeng Phet, Phichit, Nakhon Sawan, Phetchabun, Chaiyaphum, Uthai Thani, Kanchanaburi, and Nakhon Nayok.

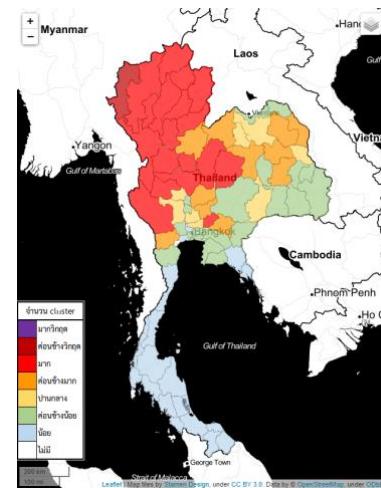


Fig. VII. Distribution of number of DBSCAN-based hotspot cluster within the provinces of Thailand

IV. CONCLUSIONS

The construction of a geographical and temporal database utilizing data from four satellites—Aqua and Terra in the MODIS system, and Suomi-NPP and NOAA-20 in the VIIRS system—necessitates the direct acquisition of signals transmitted by these satellites. This method ensures rapid data collection through automated data flow and implements essential adjustments to enhance the data for practical application in real-world contexts. A fire clustering model using the DBSCAN algorithm was also developed and validated. The model efficiently clusters hotspots corresponding to actual burn zones, utilizing a search radius of 550 meters and a minimum of three hotspots occurring within 12-hour period. This model categorized active fires into

clusters with 99% of samples coincide with detected burned area. Analyzing the area intersect between the calculated clusters and the actual burned areas revealed a substantial overlap rate of 76.72% in forested areas. We attribute the mismatched of overlap results to the limitations of satellite technology in detecting spatial alterations at a granular level. The DBSCAN-based hotspots cluster developed in this study can support not only the resource planning and allocation during fire operation, but also prioritization of prone areas to large fire events.

V. FUNDING:

This research was funded by the National Research Council of Thailand for the year 2022 under the project “Developing Geospatial Innovation for Thailand Forest Fire Clustering and Spread Prediction from Hotspot Analysis”

ACKNOWLEDGMENT

We acknowledge the Department of National Parks, Wildlife and Plant Conservation and Geo-Informatics and Space Technology Development Agency for access to geospatial data platforms.

REFERENCES

- [1] Department of fire control. (2020). *Annual Report on Forest Fire Management in Thailand (2017-2020)*. Bangkok: Department of National Parks, Wildlife, and Plant Conservation.
- [2] Lee, T., Liang, S., & Sweeney, T. (2014). **The VIIRS fire product**. *Journal of Applied Remote Sensing*, 8(1), 083626.
- [3] Davies, S., Elvidge, C., & Tuttle, B. (2009). The Fire Information for Resource Management System (FIRMS): A comprehensive tool for global fire monitoring. *International Journal of Remote Sensing*, 30(18), 4757-4771.
- [4] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C.-T. (2017). **A review of clustering techniques and developments**. *Neurocomputing*, 267, 664-681.
- [5] Xu, D., & Tian, Y. (2015). **A comprehensive survey of clustering algorithms**. *Annals of Data Science*, 2(2), 165-193.
- [6] Nikonorov, T., Chen, J. M., & Adelabu, S. (2015). **Application of DBSCAN clustering algorithm for fire detection in Canada**. *Journal of Environmental Informatics*, 26(1), 11-21.
- [7] Sitanggang, I. S., Hasmadi, M., & Saad, M. A. (2018). **Spatial-temporal DBSCAN algorithm for forest fire hotspots clustering in Indonesia**. *IEEE Access*, 6, 16622-16631.
- [8] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). **A density-based algorithm for discovering clusters in large spatial databases with noise**. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226-231). AAAI Press.
- [9] Briones-Herrera, C. I., Martínez-Torres, J., Martínez-López, M., & Jáuregui-Haza, U. (2020). **Determination of the optimal radius levels for identifying fire boundaries using MODIS and VIIRS heat hotspots**. *Applied Sciences*, 10(15), 5153.
- [10] Arte, S., Ling, Y., & Korytkowski, M. (2019). Efficient fire cluster mapping using DBSCAN and convex hull algorithms with MODIS data. *Remote Sensing*, 11(3), 282.

Performance evaluation of Mini-UAV Photogrammetry for Rooftop Area Measurement

Theerasak Ooppakarn

*Department of Geographic Information Science, School of Information and Communication Technology, University of Phayao
Phayao, Thailand
theerasak@uru.ac.th*

Sawarin Lerk-u-suke*

*Department of Geographic Information Science, School of Information and Communication Technology, University of Phayao
Phayao, Thailand
sawarin.le@up.ac.th*

Abstract— This study evaluates the efficacy of mini unmanned aerial vehicle (UAV) photogrammetry for measuring rooftop areas by integrating UAV-derived data with high-precision terrestrial surveying data. The research conducted at Uttaradit Rajabhat University in Uttaradit province, Thailand, included the extraction of 38 selected rooftop footprint samples from a 2D orthomosaic image and a 3D Digital Surface Model (DSM). The findings indicate that both measurement approaches exhibit high precision in comparison to ground surveying, with the 2D approach achieving 98.63% accuracy and the 3D approach achieving 95.00% accuracy. Both methods demonstrated a strong positive correlation with ground-truth data (correlation coefficient = 0.99). The 2D method exhibited lower error metrics, with a Mean Absolute Error (MAE) of 2.14 m² and a Mean Relative Error (MRE) of 1.36%, compared to the 3D method's MAE of 6.42 m² and MRE of 4.99%. The use of Real-Time Kinematic (RTK) positioning significantly improved measurement accuracy. This study confirms that UAV photogrammetry provides reliable, efficient, and precise data comparable to terrestrial surveying for measuring rooftop areas. These findings present significant insights for applications in urban planning, construction, and planning and installation of renewable energy systems, including solar panel positioning and rooftop photovoltaic panel installation.

Keywords: UAV photogrammetry, rooftop area detection, DSM, RTK

I. INTRODUCTION (HEADING I)

A. UAV photogrammetry:

Unmanned Aerial Vehicles (UAVs) perform as a cost-effective and efficient instrument for surveying and mapping applications [1]. UAVs can be provided with high-resolution cameras and centimeter-level ground sampling distance (GSD), as well as multi-sensor attachments, to acquire aerial imagery of the terrestrial surface. Photogrammetry and computer vision techniques can subsequently process this imagery to produce digital elevation models (DEMs) and orthomosaic images. These sophisticated methods produced three-dimensional representations of the terrain and constructed surrounding [2], [3], [4]. Recent research has shown the potential and increasing popularity of employing UAV photogrammetry to extract comprehensive information regarding building facades and rooftop features [5]. This method let researchers use the special features of unmanned aerial vehicles (UAVs), which can take high-resolution pictures from above. They then used advanced photogrammetry methods to look at this data and make accurate 3D models and measurements of the building's parts [6]. Moreover, researchers have effectively utilized UAVs for

various geospatial and environmental monitoring applications, including flood mapping and solar plant monitoring, highlighting their versatility and efficacy in meeting diverse research and operational demands [7], [8].

B. Problem statement:

Accurate and precise observation of the rooftop details is crucial for multiple applications, including solar panel installation, energy conservation assessments, and building maintenance and management. Conventional approaches to survey the rooftop information, such as tape measurement and field surveys, tend to be expensive, labor-intensive, and frequently lack the necessary depth for these applications [9]. These approaches may be insufficient for extensive or rapidly evolving building areas [10], [11], necessitating a more efficient and scalable solution to facilitate decision-making and optimization processes. This approach improves the positioning of solar panels, thereby increasing the efficiency and cost-effectiveness of renewable energy systems [12], [13], [14]. Nonetheless, current methods for assessing rooftop areas, such as manual surveys or LIDAR scans, can be highly labor-intensive and often fail to deliver the necessary level of detail for these applications [15]. An improved and scalable approach is required to provide precise and up-to-date information to facilitate these essential tasks.

C. Objective of the study:

This study aims to evaluate the potential of using mini-UAV photogrammetry to measure rooftop areas and to assess the performance of this approach compared to traditional methods, such as ground surveys with a total station.

II. LITERATURE REVIEW

A. Mini-UAV technology in surveying and mapping:

Several studies have explored the use of mini-UAV for surveying and mapping applications. Conducted a comparative study of modern UAV platforms for topographic mapping, demonstrating the effectiveness of UAV photogrammetry for generating high-resolution orthomosaic image and digital surface models [1]. Similarly, explored the use of UAV for flood-risk management, highlighting their cost-effectiveness and flexibility in providing rapid deployment and revisit capabilities for mapping small project areas [7].

The use of small UAVs for mapping and surveying has been investigated in a various study. demonstrated the efficiency of UAV photogrammetry in producing high-resolution orthomosaic images and digital surface models through a comparative analysis of contemporary UAV

platforms for topographic mapping [1]. Examined the use of UAVs for flood risk management in a similar manner, emphasizing their affordability and adaptability in offering rapid implementation and examine capabilities for mapping small project areas [7]. Utilizing techniques like normal estimation from the generated point cloud, researchers have been investigating of applying UAV-based photogrammetry to segment the surfaces of buildings and rooftop area. For applications including solar panel installation, energy efficiency evaluations, and building planning, these techniques have made it possible to precisely define and quantify individual roof segments [16]. UAV-based photogrammetry can facilitate more economical and effective decision-making in these areas by offering precise measurements of rooftop area, which will improve building energy performance and rooftop space use [17].

B. Accuracy, reliability and limitations of UAV photogrammetry:

Previous studies have demonstrated the efficacy of UAV-based photogrammetry in producing high-resolution orthomosaic images and digital surface models, emphasizing its potential for various mapping and monitoring applications [18], [19], [20], [21]. Nonetheless, the accuracy and reliability of this method for quantifying roof section areas have not been thoroughly investigated.

Conventional techniques like ground surveys or LiDAR surveys can be labor-intensive, expensive, and usually insufficient in detail, rendering them unsuitable for expansive or rapidly evolving building environments [12], [13], [22], [23]. In these cases, a more efficient and scalable approach is required to deliver precise and current information.

The reliability and accuracy of UAV photogrammetry for surveying and mapping applications have been extensively studied. Numerous studies have evaluated the precision of UAV-generated point clouds and 3D models against GNSS measurements and terrestrial laser scanning, emphasizing the difficulties in reliably identifying vertical features, particularly in plain areas. These studies emphasize the necessity of accurate preparation, sensor calibration, and data processing methodologies to guarantee the reliability of UAV photogrammetry for specific purposes.

III. METHODOLOGY

A. Equipment and tools

The study utilized a DJI Mavic 3M RTK to capture aerial imagery. Pix4D Mapper software was used to process the captured imagery and generate 3D point clouds, digital surface models, and orthomosaic image. Topcon Hiper V Plus was used to process the RTK data, ground control point (GCP) and the check point (CP) measurement.

The DJI Mavic 3M RTK camera was calibrated with the parameters detailed in Table 1. The targets used for the GCP and CP in this study will have a size of 60 cm x 60 cm. These larger targets were selected to ensure they can be easily identified and measured in the UAV image, providing reliable GCP and CP for the photogrammetric process.

TABLE I. CALIBRATION PARAMETER OF THE CAMERA

Parameters	Value
Focal Length	12 mm
Pixel Size x	0.00327 mm
Pixel Size y	0.00327 mm
f_x	3719.337845
f_y	3736.556777
c_x	2641.923361
c_y	1971.747665
k₁	-0.076011
k₂	-0.120803
k₃	0.232768
p₁	0.000179
p₂	0.000482
Date	06 July 2024

B. Study Area and Data Collection

The study area is located at the Uttaradit Rajabhat University campus in Uttaradit Province, Thailand, which covers an area of approximately 0.32 sq. km². The site has mostly flat terrain with an average elevation of around 70 m above the mean sea level (MSL). This location was chosen due to the presence of diverse building types and roof structures, making it suitable for evaluating the potential of mini-UAV and photogrammetry techniques in measuring rooftop area.

The UAV will fly at an altitude of 90 m above ground level (AGL), capturing images with an 80% forward overlap and a 70% side overlap to ensure high-quality data acquisition. We will establish the UAV-captured imagery, or GCP, across the study area using a Topcon Hiper V Plus GNSS receiver, which will provide accurate geospatial reference for the photogrammetric data processing. The UAV operates at a speed of 10 m/s with a flight time of 33 minutes. During the data collection, we programmed the UAV's camera to capture images automatically at a 2 second intervals, resulting in an average ground sampling distance of 2.46 cm per pixel at the 90 m flight altitude.

TABLE II. PARAMETERS OF THE FLIGHT PLAN

Flight's Parameters	Data
Area Covered	0.316 sq.km
Flight height	90 m
GSD	2.5 cm
Forward overlap	80%
Side overlap	70%
Number of strips	16
Number of images	942

Real time kinematic (RTK) GNSS technology integrated with the UAV will enable precise georeferencing of the acquired imagery with the National continuously operating reference stations (CORS). GCPs and CPs surveyed with high-precision GNSS equipment will be used RTK GNSS

surveying at various locations across the study area to further enhance the georeferencing accuracy.

The rooftop data was collected through *in situ* measurements using a high-precision Topcon GM-50 total station, which has an accuracy of approximately 0.3 cm. These highly accurate ground-based measurements of the roof's detail was used as the reference data for assessing the accuracy of the rooftop areas derived from the UAV photogrammetric process.

There was only slight cloud cover and a gentle breeze, which did not pose any significant obstacles for the RTK surveying process. The minimal cloud cover and light winds allowed for uninterrupted satellite signals, ensuring accurate and reliable RTK measurements throughout the survey period.

C. Data Processing and Analysis

1) Image processing:

The aerial imagery captured by the mini-UAV was processed by using Pix4D Mapper software to generate a 3D point cloud, digital surface model (DSM), and the orthomosaic image. The processing was used 9 GCPs and 17 CPs to generated the results and validate their positional accuracy. This approach can assess the impact of the number of GCPs on the overall accuracy of the photogrammetric products. Figure 1 shown the photogrammetric workflow of the study.

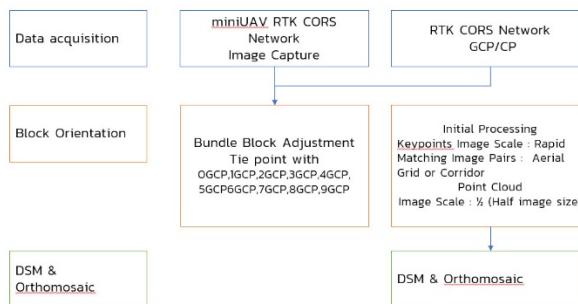


Fig. 1. Photogrammetric workflow for the study

The photogrammetric processing started with the critical step of image orientation and tie point generation. This process involves identifying and matching common features across the overlapping aerial images, establishing the spatial relationships between them. This will be followed by bundle adjustment, which optimizes the interior and exterior camera parameters to refine the 3D reconstruction and minimize distortions.

The initial processing step using a full keypoints image scale configuration to ensure that all the important visual features in the images are detected and matched. This was combined with matching the image pairs with an aerial grid, further enhancing the spatial alignment of the overlapping imageries.

A dense point cloud was generated at half the image scale, leveraging three-photo matching to achieve an optimal point density. This approach produced a highly detailed and accurate 3D representation of the scene, capturing the intricate roof structures and geometries in fine detail. Once the point cloud is generated, DSM will be created, representing the topographic surface including the building and other objects.

Then, orthomosaic image generated will have a spatial resolution of 2.5 cm with a high-quality aerial representation

of the study area. These high-quality photogrammetric outputs, including DSM and the 2.5 cm GSD of orthomosaic image, will then provide the fundamental data for the analysis and quantification of the rooftop area.

The accuracy of the UAV-based photogrammetric method will be evaluated by comparing the derived orthophoto and DSM to the ground truth measurements obtained using the RTK GNSS survey. The root mean square error (RMSE) was calculated to quantify the discrepancies between the photogrammetric and reference measurements at the CPs. The accuracy is shown as the RMSE of X, Y, and Z coordinates and orthomosaic image for different numbers of GCP used in the photogrammetric workflow. The results indicate that the photogrammetric processing achieves an accuracy better than 30 cm, which provides valuable insights into the viability of using the mini-UAV with the installed RTK module.

2) Roof segmentation and area calculation

The rooftop structures and geometries captured in the high-resolution DSM generated from the UAV imagery will be used to delineate and measure the individual rooftop segment. 38 selected rooftop footprints were collected and manually extracted using GIS software (See Fig 2). The rooftop footprint will then be compared to the ground surveys that were collected by using the total station on both 2D and 3D measurements. This comparison will be conducted to assess the accuracy and reliability of the UAV photogrammetric approach and quantify the rooftop areas that were calculated using the DSM and GIS software.

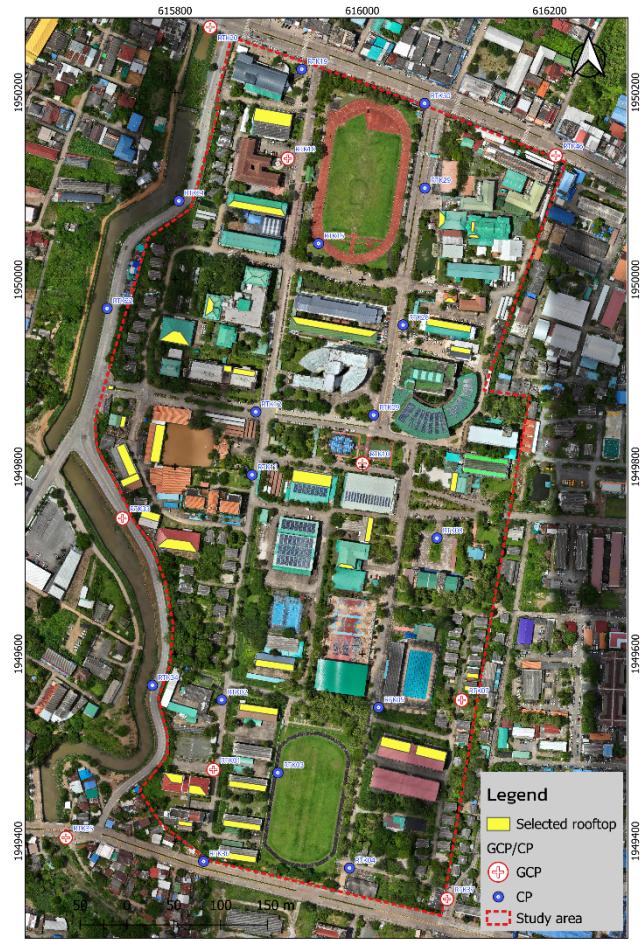


Fig. 2. Rooftops for performance evaluation

IV. RESULTS

Positional accuracy of the orthomosaic image and DSM was evaluated using RMSE as an error indicator. Results showed horizontal accuracy ranging from 0.179 to 0.208 m and vertical accuracy between 0.206 and 0.257 m. The number of Ground Control Points (GCPs) affected accuracy, with 4 GCPs achieving consistent accuracy, maintaining errors below 0.208 m without significant variation. (See Fig 3.).

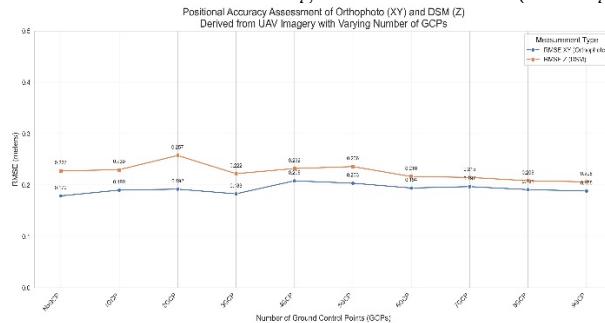


Fig. 3. Position Accuracy of Orthophoto and DSM

The building rooftop area calculated by counting pixels that fall within each manually delineated rooftop segment polygon. The method offers a computationally efficient alternative to Triangulated Irregular Network (TIN) based approaches, while maintaining high accuracy, particularly for areas more than 250 cells. The algorithm operates on a cell-by-cell basis, utilizing a 3x3 cell neighborhood for each calculation.

The accuracy has been validated against TIN-based calculations, showing high correlation (Spearman's rank correlation $r^2 > 0.999$) for polygons with more than 250 cells. This makes it particularly suitable for landscape-scale analyses where computational efficiency is crucial. By implementing this method, researchers can efficiently calculate surface areas across large landscapes, enabling more accurate assessments of topographic complexity and its influence on ecological processes.

The results, when compared to ground measurements, demonstrated higher accuracy and reliability. Table 3 presents the area calculated through different methods: one using a total station and the other computed from UAV photogrammetric-based processing (mini-UAV).

TABLE III. COMPARISON OF ROOFTOP AREA MEASUREMENTS OBTAINED FROM GROUND SURVEYS AND UAV PHOTOGRAMMETRIC-BASED ROOF AREA EXTRACTION (m^2)

Roof ID	Total station		mini-UAV	
	T2DArea	T3DArea	OrthoArea2D	DSMArea3D
roof01	565.98	571.44	560.84	575.37
roof02	409.89	422.72	412.23	427.67
roof03	5.23	6.29	5.22	6.69
roof04	9.53	11.36	9.35	10.90
roof05	78.18	83.52	83.75	97.84
roof06	36.58	46.08	36.75	51.15
roof07	30.64	38.18	33.31	52.36
roof08	617.21	678.92	621.33	689.93
roof09	340.87	353.00	343.83	363.98
roof10	119.36	147.04	119.31	147.07

roof11	107.41	118.03	106.71	118.14
roof12	256.92	266.57	256.93	267.51
roof13	54.05	54.05	53.76	57.21
roof14	133.55	143.32	134.54	146.23
roof15	23.39	23.50	22.85	25.38
roof16	56.76	62.04	56.93	65.90
roof17	397.27	412.20	396.53	412.40
roof18	312.66	336.56	314.17	343.06
roof19	149.73	153.35	150.44	164.84
roof20	160.80	164.32	163.20	167.66
roof21	355.05	362.90	356.12	384.86
roof22	523.26	538.12	529.28	553.19
roof23	122.04	123.10	124.87	127.77
roof24	146.96	150.37	149.95	154.36
roof25	298.57	314.52	299.48	326.37
roof26	289.73	299.47	294.94	311.26
roof27	266.14	270.82	264.08	269.41
roof28	265.93	271.20	261.59	267.36
roof29	281.94	286.88	270.75	288.60
roof30	68.93	72.94	70.45	86.25
roof31	147.72	171.83	149.71	172.98
roof32	150.17	173.58	146.94	169.08
roof33	341.68	395.06	339.29	420.55
roof34	338.43	391.02	341.98	403.70
roof35	63.68	64.63	63.89	66.72
roof36	63.46	64.36	63.45	66.09
roof37	47.62	55.10	47.62	55.29
roof38	60.76	66.48	61.57	73.28

A. Evaluation of mini-UAV Photogrammetry for Roof Area Measurement

This study evaluated the accuracy of roof area measurements obtained through UAV photogrammetry compared to ground-based survey methods. We calculated roof area on both 2D orthomosaic image (OrthoArea2D) and the Digital Surface Model (DSMArea3D).

B. Accuracy (%)

Both UAV approaches offer extremely accurate measurements, according to the accuracy percentages. With assessments of 98.63% and 95.00% accuracy, respectively, the 2D orthomosaic image approach (OrthoArea2D) and the 3D DSM approach (DSMArea3D) demonstrate slightly better accuracy. This implies that the 2D method might be adequate and precise enough for rooftop area estimations.

C. Correlation

The measurements from UAV photogrammetric products show surprisingly strong positive correlations with ground-based measurements, with values approaching 1. This indicates that the measurements of UAV products consistently represent variations in actual roof areas, regardless of the roof area. The slightly increased correlation with the 2D technique further supports its reliability for this application.

D. Mean Absolute Error (MAE)

The MAE provides an understanding of the magnitude of errors in measurement. The 2D methodology exhibits a lower Mean Absolute Error (MAE), signifying that its inaccuracies are less than those of the total station's rooftop area measurement. This indicates that the 2D method has greater consistency in accuracy across various rooftop areas than the 3D method.

E. Mean Relative Error (MRE) (%)

The MRE offers insights into the relative magnitude of measurement errors. The 2D comparison method shows a reduced MRE, indicating that the measurement inaccuracies are lower compared to the ground survey measurements. This indicates that the 2D comparison preserves the precision more reliably across various rooftop areas than the 3D method.

F. Root Mean Square Error (RMSE)

The RMSE values, which prioritize greater errors, reflect a comparable trend to the MAE. The 2D comparison indicate s a lower RMSE, signifying that it is less susceptible to significant measurement errors than the 3D comparison method.

Furthermore, the rooftop area obtained from mini UAV photogrammetric processing will be compared with ground-based total station measurements, and the variations will be examined to evaluate the utility of UAV photogrammetry in precisely quantifying the rooftop area.

TABLE IV. SUMMARIZES THE KEY METRICS USED TO EVALUATE THE ACCURACY OF THE ROOF SEGMENT AREA MEASUREMENTS

Metric	Roof segment area	
	OrthoArea2D vs. T2DArea	DSMArea3D vs. T3DArea
Accuracy (%)	98.6314	95.0069
Correlation	0.9998	0.9993
MAE	2.1487	6.4242
MRE (%)	1.3686	4.9931
RMSE	3.1346	8.9574

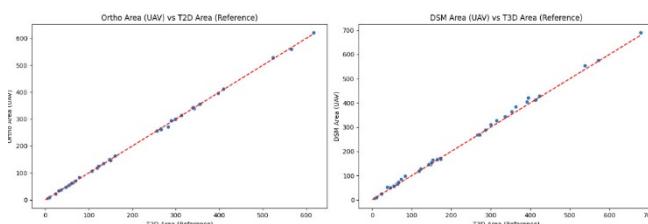


Fig. 4. Scatter plot comparing ground survey measurements with UAV photogrammetric-based roof area extraction

Table 3 and Figure 4 show a high level of agreement between the UAV-photogrammetric-based products and ground survey-based measurements. This indicates that the UAV photogrammetric approach can reliably quantify rooftop areas. These findings highlight the potential of mini-UAV platforms equipped with RTK positioning and photogrammetric techniques to provide cost-effective and accurate measurements of roof geometries such as area and shape.

V. DISCUSSION

In this paper, the results indicate that UAV photogrammetry, particularly when integrated with RTK positioning, provides a dependable and accurate alternative to conventional ground-based techniques for measuring rooftop areas. Both the 2D orthomosaic and 3D DSM methodologies achieved high accuracy levels, with the 2D method, calculated on orthomosaic image, indicating a significant advantage in precision and consistency. The high precision validates that UAV-based measurements are appropriate for applications necessitating precise rooftop data. The significant relationship with ground survey data further indicates the efficacy of UAV photogrammetry in accurately capturing real-world measurements.

Selection between the 2D and 3D approaches needs to be determined by the specific requirements of the application. The 2D orthomosaic method is especially effective for obtaining a precise measurement of rooftop area, demonstrating reduced mean absolute error (MAE) and mean relative error (MRE). On the other hand, the 3D DSM method, while marginally less precise, offers additional volumetric data, making it appropriate for conditions in which the vertical feature and elevation variations of rooftops are important. The capability of mini-UAV photogrammetry to generate both 2D and 3D data offers a comprehensive resource for various spatial analysis.

Accurate rooftop measurements are crucial for efficient solar panel placement and rooftop photovoltaic (PV) installations. The findings offer actionable insights for optimizing energy efficiency and resource allocation in renewable energy projects.

Future research could explore the use of machine learning algorithms to automate the segmentation and measurement of rooftop areas based on UAV data. This could reduce manual effort and improve consistency, making UAV photogrammetry even more accessible for large-scale applications.

REFERENCES

- [1] M. H. Chaudhry, A. Ahmad, and Q. Gulzar, "A comparative study of modern UAV platform for topographic mapping," IOP Conference Series Earth and Environmental Science, vol. 540, no. 1, p. 012019, Jul. 2020, doi: 10.1088/1755-1315/540/1/012019.
- [2] R. C. Erenoglu, O. Erenoglu, and N. Arslan, "Accuracy Assessment of Low Cost UAV Based City Modelling for Urban Planning," Tehnicki Vjesnik - Technical Gazette, vol. 25, no. 6, Dec. 2018, doi: 10.17559/tv-20170904202055.
- [3] S. I. Jiménez-Jiménez, W. Ojeda-Bustamante, M. Marcial-Pablo, and J. Enciso, "Digital Terrain Models Generated with Low-Cost UAV Photogrammetry: Methodology and Accuracy," ISPRS International Journal of Geo-Information, vol. 10, no. 5, p. 285, Apr. 2021, doi: 10.3390/ijgi10050285.
- [4] "UAS RTK / PPK photogrammetry as a tool for mapping the urbanized landscape, creating thematic maps, situation plans and DEM," *Acta Montanistica Slovaca*, no. 26, pp. 649–660, Feb. 2022, doi: 10.46544/ams.v26i4.05.
- [5] S. K. P. Kushwaha, K. R. Dayal, A. Singh, and K. Jain, "BUILDING FACADE AND ROOFTOP SEGMENTATION BY NORMAL ESTIMATION FROM UAV DERIVED RGB POINT CLOUD," *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W17, pp. 173–177, Nov. 2019, doi: 10.5194/iss-archives-xlii-2-w17-173-2019.

- [6] B. Ruzgienė and C. Aksamituskas, "THE USE OF UAV SYSTEMS FOR MAPPING OF BUILT-UP AREA," *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, pp. 349–353, Aug. 2013, doi: 10.5194/isprsarchives-xl-1-w2-349-2013.
- [7] L. Hashemi-Beni, J. Jones, G. Thompson, C. Johnson, and A. Gebrehiwot, "Challenges and Opportunities for UAV-Based Digital Elevation Model Generation for Flood-Risk Management: A Case of Princeville, North Carolina," *Sensors*, vol. 18, no. 11, p. 3843, Nov. 2018, doi: 10.3390/s18113843.
- [8] S. Gallardo-Saavedra, L. Hernández-Callejo, and O. Duque-Perez, "Technological review of the instrumentation used in aerial thermographic inspection of photovoltaic plants," *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 566–579, Jun. 2018, doi: 10.1016/j.rser.2018.05.027.
- [9] J. Melius, R. Margolis and S. Ong, "Estimating Rooftop Suitability for PV: A Review of Methods, Patents, and Validation Techniques", www.nrel.gov/publications.
- [10] N. Haala, M. Kölle, M. Cramer, D. Laupheimer, and F. Zimmermann, "Hybrid georeferencing of images and LiDAR data for UAV-based point cloud collection at millimetre accuracy," *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 4, p. 100014, Apr. 2022, doi: 10.1016/j.photo.2022.100014.
- [11] N. Kolibarov, D. Wästberg, V. Naserentin, D. Petrova-Antanova, S. Ilieva, and A. Logg, "Roof Segmentation Towards Digital Twin Generation in LoD2+ Using Deep Learning," *IFAC-PapersOnLine*, vol. 55, no. 11, pp. 173–178, Jan. 2022, doi: 10.1016/j.ifacol.2022.08.068.
- [12] X. Huang, K. Hayashi, T. Matsumoto, L. Tao, Y. Huang, and Y. Tomino, "Estimation of Rooftop Solar Power Potential by Comparing Solar Radiation Data and Remote Sensing Data—A Case Study in Aichi, Japan," *Remote Sensing*, vol. 14, no. 7, p. 1742, Apr. 2022, doi: 10.3390/rs14071742.
- [13] E. Fakhraian, M. A. Forment, F. V. Dalmau, A. Nameni, and M. J. C. Guerrero, "Determination of the urban rooftop photovoltaic potential: A state of the art," *Energy Reports*, vol. 7, pp. 176–185, Sep. 2021, doi: 10.1016/j.egyr.2021.06.031.
- [14] Mr. R. Ahamed, "ASSESSMENT OF SOLAR POTENTIAL ROOFTOPS USING GIS AND REMOTE SENSING TECHNIQUES," *Sab*, Feb. 2019, [Online]. Available: https://www.academia.edu/37436948/ASSESSMENT_OF_SOLAR_POTENTIAL_ROOFTOPS_USING_GIS_AND_REMOTE_SENSING_TECHNIQUES
- [15] Q. Yang *et al.*, "A new technical pathway for extracting high accuracy surface deformation information in coal mining areas using UAV LiDAR data: An example from the Yushen mining area in western China," *Measurement*, vol. 218, p. 113220, Jun. 2023, doi: 10.1016/j.measurement.2023.113220.
- [16] Z. Chen, B. Yang, R. Zhu, and Z. Dong, "City-scale solar PV potential estimation on 3D buildings using multi-source RS data: A case study in Wuhan, China," *Applied Energy*, vol. 359, p. 122720, Feb. 2024, doi: 10.1016/j.apenergy.2024.122720.
- [17] A. Anand and C. Deb, "The potential of remote sensing and GIS in urban building energy modelling," *Energy and Built Environment*, vol. 5, no. 6, pp. 957–969, Jul. 2023, doi: 10.1016/j.enbenv.2023.07.008.
- [18] F. Agüera-Vega, F. Carvajal-Ramírez, and P. Martínez-Carricundo, "Assessment of photogrammetric mapping accuracy based on variation ground control points number using unmanned aerial vehicle," *Measurement*, vol. 98, pp. 221–227, Dec. 2016, doi: 10.1016/j.measurement.2016.12.002.
- [19] O. E. Mora, A. Suleiman, J. Chen, D. Pluta, M. H. Okubo, and R. Josenhans, "Comparing sUAS Photogrammetrically-Derived Point Clouds with GNSS Measurements and Terrestrial Laser Scanning for Topographic Mapping," *Drones*, vol. 3, no. 3, p. 64, Aug. 2019, doi: 10.3390/drones3030064.
- [20] K. Nakano, Y. Tanaka, H. Suzuki, K. Hayakawa, and M. Kurodai, "ON PRACTICAL ACCURACY ASPECTS OF UNMANNED AERIAL VEHICLES EQUIPPED WITH SURVEY GRADE LASER SCANNERS," *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2020, pp. 343–348, Aug. 2020, doi: 10.5194/isprs-archives-xliii-b1-2020-343-2020.
- [21] P. Kavaliauskas, D. Židanavičius, and A. Jurelionis, "Geometric Accuracy of 3D Reality Mesh Utilization for BIM-Based Earthwork Quantity Estimation Workflows," *ISPRS International Journal of Geo-Information*, vol. 10, no. 6, p. 399, Jun. 2021, doi: 10.3390/ijgi10060399.
- [22] K. R. Dayal and I. M. Chauhan, "UAV PHOTOGRAMMETRY FOR FEATURE EXTRACTION AND MAPPING OF CORRUGATED INDUSTRIAL ROOFTOPS," *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4, pp. 129–133, Sep. 2018, doi: 10.5194/isprs-archives-xlii-4-129-2018.
- [23] D. A. Jacques, J. Gooding, J. J. Giesekam, A. S. Tomlin, and R. Crook, "Methodology for the assessment of PV capacity over a city region using low-resolution LiDAR data and application to the City of Leeds (UK)," *Applied Energy*, vol. 124, pp. 28–34, Jul. 2014, doi: 10.1016/j.apenergy.2014.02.076.

Optimizing Building Footprint Extraction from UAV Imagery using Pretrained Deep Learning Models: A Case Study in University of Phayao, Thailand

Kannika Maomool

*Department of Geographic Information Science, School of Information and Communication Technology, University of Phayao
Phayao, Thailand
kannika.maomool@gmail.com*

Samniang Suttara

*Western University
Kanchanaburi, Thailand
lookniang@gmail.com*

Sawarin Lerk-u-suks*

*Department of Geographic Information Science, School of Information and Communication Technology, University of Phayao
Phayao, Thailand
sawarin.le@up.ac.th*

Abstract— This study investigates an optimization of deep learning models, the ArcGIS pretrained model (Building Footprint Extraction—USA), for detecting building footprints from UAV orthorectified images, emphasizing the influence of parameters including threshold, padding, and tile size. Experimental results show increasing the threshold value significantly reduces the false positives, with higher thresholds (0.7–0.9) achieving correctness levels of up to 90%, while the lower threshold values result in correctness around 40–60%. Padding values of 600–800 improve detection accuracy by minimizing false positives through the acquisition of additional contextual information surrounding building structures. Additionally, larger tile sizes improve detection efficiency, while smaller tiles provide finer resolution but can increase false positives. The findings reveal that balancing these parameters is critical for optimizing building footprint extraction, contributing valuable insights for geospatial analysis in urban planning and remote sensing applications.

Keywords— deep learning, pretrained model, building footprint, segmentation, extraction

I. INTRODUCTION

Building footprints are crucial spatial data utilized in various applications, such as urban planning, infrastructure monitoring, disaster management, and other geographic information systems. Accurate representation of features is essential for assessing building density, land use/land cover patterns, and the change detection. In recent years, numerous techniques have been developed to extract the building footprints, ranging from traditional ground surveys to remote sensing methods. Imagery derived from aerial platforms, satellites, and the UAV (Unmanned Aerial Vehicle) photogrammetry has obtained significant attention due to its ability to provide high-resolution, up-to-date data. Aerial imagery, with its wide coverage and moderate resolution, has a long time used for building footprint extraction, while satellite imagery offers extensive temporal coverage at various resolutions. However, UAV photogrammetry has emerged as a particularly effective tool, providing high-detail, low-cost, and flexible data collection, especially in urban environments [1], [2]. These methods allow for the generation of 3D models, orthophotos, and digital surface models, which can be integrated into Geographic Information Systems (GIS) for further analysis and decision-making. Previous studies highlighted the complexities in accurately extracting building

footprints [3], [4], [5], [6], particularly in dense urban areas where the diversity of structures, shadows, and occlusions presents significant challenges. Extracting footprints from remote sensing data is not a simple task [7]; it requires advanced knowledge of image processing techniques, photogrammetry, and understanding of limitations of the data. Automated methods often struggle with variability in building shapes, materials, and surrounding environments [8], while manual or semi-automated approaches require considerable expertise and time. The extraction of precise and reliable building footprints demands not only technical proficiency but also experience in interpreting the outputs from various sources [9], [10] like LIDAR, UAV imagery, aerial photographs, and satellite data. The use of deep learning and machine learning techniques have grown in popularity for extracting building footprints from earth observation imagery. Convolutional Neural Networks (CNNs) and Active Contour Model (ACM) are highly proficient in image segmentation tasks [8], [9], [10], [11]. These models can autonomously identify and outline building perimeters from high-resolution satellite and UAV data [12], [13].

However, implementing these methods requires a significant understanding of mathematical concepts such as linear algebra, optimization, and probability, as well as digital image processing techniques. Pre-processing the data including noise handling, shadows, or varying resolutions requires expertise, and without this knowledge and experiences, the accuracy of the models can be compromised. Furthermore, the training these models often requires large, annotated datasets and computational resources, to perform accurate results which may be difficult to obtain [14].

Esri's ArcGIS platform is a comprehensive suite of tools that provides powerful capabilities for object detection and feature extraction, particularly valuable for those new to the field of deep learning, remote sensing, and geoinformation technology [15]. Within ArcGIS, ArcGIS Pro and ArcGIS Image Analyst offer user-friendly environments where complex spatial analysis tasks, including object detection, can be executed with minimal coding experience [16]. The Deep Learning Tools in ArcGIS Pro allow users to apply pre-trained deep learning models to extract features such as buildings, roads, or vegetation from high-resolution imagery. These tools are designed with an intuitive graphic user interface (GUI), making it easy for beginners to start using advanced

techniques like convolutional neural networks (CNNs) without diving into detailed programming.

Proprietary software solutions provide an approachable entry point into building footprint extraction for novices and non-experts in the field. These tools provide friendly GUI and pre-configured workflows and pre-trained data model that abstract the complexities of deep learning models and data processing pipelines. By leveraging such software, researchers can focus on the application and interpretation of results rather than the intricacies of algorithm implementation.

In This paper, we propose an optimum parameter for building footprint detection; rooftop boundary extraction in UAV orthorectified images. The results show statistical analysis of yields

II. METHODS

A. Study area

The University of Phayao (UP) was previously governed by Naresuan University (NU) before attaining autonomous status, which acknowledges the significance of equitable opportunity distribution and the promotion of educational equality. The institution, in collaboration with the provincial governor and delegates from Phayao province, including those from both public and private sectors, has addressed issues related to income and education for the population of Phayao Province [17]. The University of Phayao Act B.E. 2010 was enacted and published in the Government Gazette on July 16, 2010. separated from Naresuan university It is a fully autonomous, non-governmental university. The University of Phayao located in Mae Ka sub-district, Muang Phayao district, Phayao province. The location consisted of 8.25 sq. km. of land.

Fig. 1 presents the study area, outlined by the red-line border, represents one of the building groups at the University of Phayao. This building is of particular interest due to several distinctive attributes, including its expansive rooftop, clear and unobstructed surface, and simple geometric structure with regular facets. These characteristics make the building readily identifiable by human interpretation and well-suited for solar rooftop photovoltaic system installations. Additionally, its structural simplicity, coupled with the building's location within a group of similar structures in the university, enhances its potential for efficient solar energy generation.



Fig. 1. The study area.

B. UAV orthorectified image

The UAV orthorectified image was captured using a DJI Phantom 4 Pro Version 2.0, flown at an altitude of 90 meters

above ground, in full compliance with government regulations for aerial operations. The flight mission was carefully planned to achieve an 80 percent overlap and a 70 percent sidelap, ensuring sufficient data redundancy for high-precision photogrammetric processing. The captured images were processed using Agisoft Metashape Professional Version 2.0.1 (Educational license, Single), where the Structure-from-Motion (SfM) technique was applied to align the images based on common tie points. This was followed by the generation of a dense point cloud and the creation of a digital surface model (DSM). Orthorectification was performed to correct any geometric distortions resulting from camera orientation, terrain undulation, and perspective effects. The final orthorectified image is both georeferenced and uniformly scaled, making it an accurate and reliable dataset for spatial analysis and related work. The tested UAV orthorectified image (Fig. 2), with a resolution of 27,825 x 12,636 pixels and a ground sampling distance (GSD) of 2.5 cm, was used without compression to preserve the full quality and detail necessary for accurate and reliable analysis.



Fig. 2. The tested UAV orthorectified image.

C. Pretrained Deep Learning model

This study utilized ArcGIS pretrained model for building footprint extraction (Building Footprint Extraction - USA) which was released in February 2021. The model is constructed by the Mask R-CNN algorithm, recognized for its high efficacy in segmentation tasks within computer vision. Trained on 8-bit very high-resolution aerial and satellite imagery with a ground sampling distance ranging from 0.1 to 0.4 meters, the model demonstrated an average precision score of 0.79 [18].

Leveraging this model facilitated the efficient identification and extraction of building footprints, streamlining the process with a pretrained model fitted to a high and very high-resolution imagery. ArcGIS pretrained model were used to automate the detection of buildings within the imagery. By leveraging deep learning techniques, the models efficiently identified the spatial characteristics of buildings, facilitating the rapid and accurate extraction of building footprint [19]. This approach significantly reduced the manual effort typically required for digitizing geographic features and ensured consistent and precise results across the study area. The configured parameters for building footprint detection comprised padding values of 200, 400, 600, and 800. The threshold varies from 0.1 to 0.9, increasing in increments of 0.1, and the tile sizes utilized were 512, 1024, and 2048.

III. RESULTS AND DISCUSSION

The experimental results reveal the correction of building footprint extraction using ArcGIS pretrained model depend

on the different selected padding, threshold, and tile size values (See Table I, II, II, and IV). Lower threshold values tend to result in a higher number of detected objects, but often with lower corrected footprint. As the threshold increases, fewer objects are detected, but the corrected footprint improves. This suggests that a balance must be considered between sensitivity (detecting more objects) and precision (correctly identifying rooftops). The use of padding also plays a crucial role in the model's performance. Smaller padding values (e.g., 200) tend to produce a wider range of detected objects but can also lead to more false positives. On the other hand, larger padding values (e.g., 800) generally result in more precise detections.

Increasing the threshold value generally improves the footprint's correctness. At lower thresholds, the model tends to detect more objects, but many of them are false positives, leading to lower correctness. As the threshold increases, the model becomes more selective, resulting in fewer detections but higher correctness. This is because higher thresholds filter out noise, ensuring that only objects with higher confidence scores are detected as true building footprints. For example, at a threshold of 0.9, correctness consistently reaches 90% or higher across different padding values, compared to lower thresholds, where correctness hovers around 40-60%. Higher thresholds improve correctness by reducing false positives, while the effect of tile size depends on balancing computational efficiency and detection resolution.

The results also show how padding, threshold, and the correctness impact the detection of building footprints using ArcGIS pretrained model (See in Fig. 3). This is a clear trade-off: as the threshold increases, the model focuses more on detecting true footprints and reducing false positives, but it may miss a few actual footprints. These results highlight the importance of balancing number of footprints and correctness to achieve optimal detection performance, depending on the specific goals of the task.

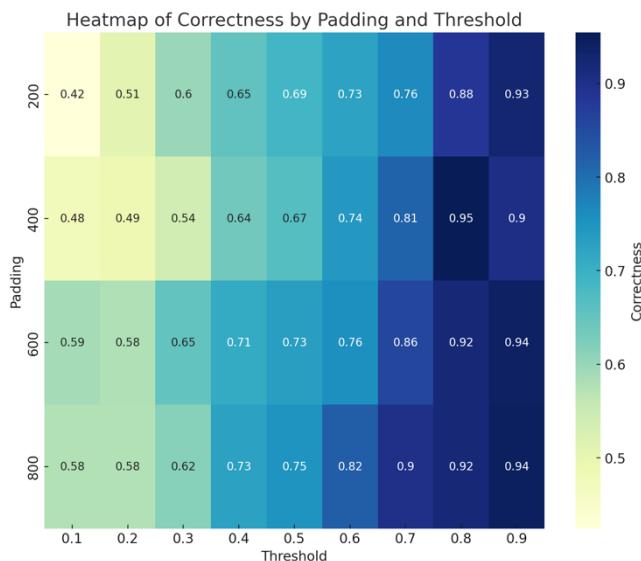


Fig. 3. Heatmap of Correctness by Padding and Threshol value

TABLE I. RESULTS FROM SETTING PADDING IS 200

Padding	Threshold	Tiled size	Detected footprints	Corrected footprints
200	0.1	512	138	33
200	0.2	512	69	34
200	0.3	512	53	32
200	0.4	512	48	32
200	0.5	512	45	31
200	0.6	512	38	27
200	0.7	512	31	23
200	0.8	512	25	22
200	0.9	512	18	17
200	0.1	1024	70	32
200	0.2	1024	70	32
200	0.3	1024	56	32
200	0.4	1024	47	29
200	0.5	1024	43	29
200	0.6	1024	38	27
200	0.7	1024	32	24
200	0.8	1024	24	22
200	0.9	1024	17	16
200	0.1	2048	59	34
200	0.2	2048	59	34
200	0.3	2048	53	33
200	0.4	2048	47	32
200	0.5	2048	44	31
200	0.6	2048	37	28
200	0.7	2048	35	28
200	0.8	2048	28	24
200	0.9	2048	20	1

TABLE II. RESULTS FROM SETTING PADDING IS 400

Padding	Threshold	Tiled size	Detected footprints	Corrected footprints
400	0.1	1024	80	33
400	0.2	1024	76	33
400	0.3	1024	61	31
400	0.4	1024	51	32
400	0.5	1024	46	30
400	0.6	1024	35	26
400	0.7	1024	30	24
400	0.8	1024	23	22
400	0.9	1024	16	14
400	0.1	2048	61	33
400	0.2	2048	61	33
400	0.3	2048	56	32
400	0.4	2048	48	31
400	0.5	2048	44	30
400	0.6	2048	35	26
400	0.7	2048	29	24
400	0.8	2048	21	20
400	0.9	2048	15	14

TABLE III. RESULTS FROM SETTING PADDING IS 600

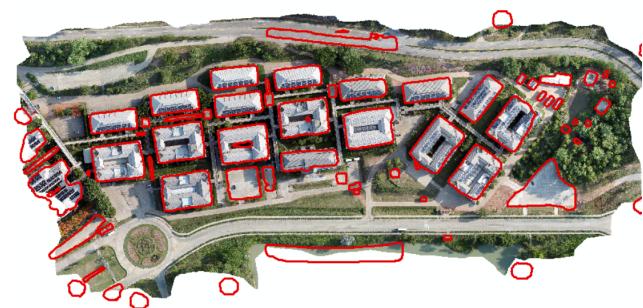
<i>Padding</i>	<i>Threshold</i>	<i>Tiled size</i>	<i>Detected footprints</i>	<i>Corrected footprints</i>
600	0.1	2048	56	33
600	0.2	2048	57	33
600	0.3	2048	49	32
600	0.4	2048	45	32
600	0.5	2048	41	30
600	0.6	2048	37	28
600	0.7	2048	28	24
600	0.8	2048	25	23
600	0.9	2048	17	16

TABLE IV. RESULTS FROM SETTING PADDING IS 800

<i>Padding</i>	<i>Threshold</i>	<i>Tiled size</i>	<i>Detected footprints</i>	<i>Corrected footprints</i>
800	0.1	2048	59	34
800	0.2	2048	59	34
800	0.3	2048	52	32
800	0.4	2048	44	32
800	0.5	2048	40	30
800	0.6	2048	34	28
800	0.7	2048	30	27
800	0.8	2048	25	23
800	0.9	2048	18	17

Figure 4 demonstrates the false positives on the building footprint detection model, particularly under different threshold and padding configurations. At lower threshold value (0.2), the model detects a higher number of false positives, resulting in lower correctness. Because the model is less selective, detecting many objects that are not actual building footprints. As the threshold increases (e.g., 0.7 or 0.9), the number of false positives decreases significantly (See in Fig. 5), leading to improved correctness across different padding values. Additionally, the padding parameter influences the results; for example, padding values of 400 to 800 generally perform better in terms of reducing false positives, as seen by the higher correctness levels. This indicates that configuring a higher threshold and appropriate padding can minimize false positives and improve the overall accuracy of the model.

According to the results, optimal parameters for building footprint extraction is achieved by using larger padding values (600–800) combined with higher tile sizes (1024–2048 pixels) and a higher threshold setting (e.g., 0.7–0.9). This configuration enhances detection accuracy by encompassing a broader contextual area around building structures, reducing false positives and enabling the model to maintain high precision in footprint delineation.



Padding = 200, Threshold = 0.2, Tile size = 512

Fig. 4. Small padding and low threshold occurred false positive object detection (red border)



Fig. 5. High padding and high threshold reduced false positive object detection (red border)

IV. CONCLUSION

This study illustrates the effectiveness of using ArcGIS pretrained model for building footprint extraction from UAV imagery, representing a robust and efficient approach that significantly reduces manual effort in geospatial analysis. The results emphasize the importance of optimizing parameters such as threshold, padding, and tile size to achieve a balance between detection sensitivity and precision, thereby improving the accuracy of building footprint identification.

A key finding is the trade-off between sensitivity and precision. Lower threshold values tend to detect a higher number of objects but often result in higher false positives, thereby reducing overall correctness. Conversely, increasing the threshold significantly reduces false positives, enhancing detection accuracy. Additionally, padding shows a critical role, with both extremely large and small padding values affecting the model's performance by affecting detection accuracy and the ability to extract building footprint. Tile size is another important factor that influences detection products. Smaller tile sizes provide finer resolution, which is useful for detecting smaller or more complicated features, but they also require more processing and can increase false positives if the threshold is not appropriately set. Larger tile sizes, while less detailed, improve detection efficiency and reduce false positives by processing larger areas.

Researchers who want to use pretrained models should carefully consider model adaptation to local contexts, especially when the data comes from different geographic regions or architectural types. Pretrained models, like those in ArcGIS, are designed for datasets with specific characteristics, such as The U.S. urban buildings, which may not apply to another location. Additional local-specific data can improve model accuracy by learning the unique structural features and environmental variables in the location. Building footprint extraction sensitivity and precision should be balanced by

exploring appropriate parameter settings, including higher threshold values, larger padding, and adequate tile sizes. The pretrained models are powerful and versatile geospatial analytic tools because they improve detection accuracy and are suitable for getting started in feature extraction using machine and deep learning techniques.

The study contributes valuable insights into building footprint detection using deep learning in UAV imagery, providing a practical, adjustable tool for urban planning, land use analysis, and infrastructure management. Future research can build upon these findings by exploring customized fine-tuning techniques to expand the model's applicability globally, therefore maximizing its utility across various geographic and structural contexts.

ACKNOWLEDGMENT

We gratefully acknowledge the School of Information and Communication Technology (ICT), University of Phayao, for providing access to the ArcGIS software license used in this research.

REFERENCES

- [1] "A Review of Building Extraction From Remote Sensing Imagery: Geometrical Structures and Semantic Attributes," *IEEE Journals & Magazine | IEEE Xplore*, 2024. <https://ieeexplore.ieee.org/document/10458126/>
- [2] M. Awrangjeb, X. Hu, B. Yang, and J. Tian, "Editorial for Special Issue: 'Remote Sensing based Building Extraction,'" *Remote Sensing*, vol. 12, no. 3, p. 549, Feb. 2020, doi: 10.3390/rs12030549.
- [3] Y. Huang and Y. Jin, "Aerial Imagery-Based Building Footprint Detection with an Integrated Deep Learning Framework: Applications for Fine Scale Wildland-Urban Interface Mapping," *Remote Sensing*, vol. 14, no. 15, p. 3622, Jul. 2022, doi: 10.3390/rs14153622.
- [4] W. Liu *et al.*, "Building Footprint Extraction From Unmanned Aerial Vehicle Images Via PRU-Net: Application to Change Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2236–2248, Jan. 2021, doi: 10.1109/jstARS.2021.3052495.
- [5] C. Ayala, R. Sesma, C. Aranda, and M. Galar, "A Deep Learning Approach to an Enhanced Building Footprint and Road Detection in High-Resolution Satellite Imagery," *Remote Sensing*, vol. 13, no. 16, p. 3135, Aug. 2021, doi: 10.3390/rs13163135.
- [6] J. Yin, F. Wu, Y. Qiu, A. Li, C. Liu, and X. Gong, "A Multiscale and Multitask Deep Learning Framework for Automatic Building Extraction," *Remote Sensing*, vol. 14, no. 19, p. 4744, Sep. 2022, doi: 10.3390/rs14194744.
- [7] G. Pasquali, G. C. Iannelli, and F. Dell'Acqua, "Building Footprint Extraction from Multispectral, Spaceborne Earth Observation Datasets Using a Structurally Optimized U-Net Convolutional Neural Network," *Remote Sensing*, vol. 11, no. 23, p. 2803, Nov. 2019, doi: 10.3390/rs11232803.
- [8] W. Nurkarim and A. W. Wijayanto, "Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework," *Earth Science Informatics*, vol. 16, no. 1, pp. 515–532, Nov. 2022, doi: 10.1007/s12145-022-00895-4.
- [9] Y. Sun, X. Zhang, X. Zhao, and Q. Xin, "Extracting Building Boundaries from High Resolution Optical Images and LiDAR Data by Integrating the Convolutional Neural Network and the Active Contour Model," *Remote Sensing*, vol. 10, no. 9, p. 1459, Sep. 2018, doi: 10.3390/rs10091459.
- [10] K. Zhang, J. Yan, and S. -c. Chen, "Automatic Construction of Building Footprints From Airborne LIDAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2523–2533, Aug. 2006, doi: 10.1109/tgrs.2006.874137.
- [11] K. Bittner, S. Cui, and P. Reinartz, "BUILDING EXTRACTION FROM REMOTE SENSING DATA USING FULLY CONVOLUTIONAL NETWORKS," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, pp. 481–486, May 2017, doi: 10.5194/isprs-archives-xlii-1-w1-481-2017.
- [12] M. Awrangjeb, C. Zhang, and C. S. Fraser, "Automatic extraction of building roofs using LIDAR data and multispectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, pp. 1–18, Jun. 2013, doi: 10.1016/j.isprsjprs.2013.05.006.
- [13] "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Journals & Magazine | IEEE Xplore*, Dec. 01, 2017. <https://ieeexplore.ieee.org/document/8113128>
- [14] A. Khan, A. Laghari, and S. Awan, "Machine Learning in Computer Vision: A Review," *ICST Transactions on Scalable Information Systems*, p. 169418, Jul. 2018, doi: 10.4108/eai.21-4-2021.169418.
- [15] "Desktop GIS Software | Mapping Analytics | ArcGIS Pro." <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview?>
- [16] "ArcGIS | Geospatial Platform - GIS Software for Business & Government," *Esri*. <https://www.esri.com/en-us/arcgis/geospatial-platform/overview>
- [17] "University Of Phayao History : University Of Phayao." https://www.up.ac.th/en/Intro_history.aspx
- [18] "Introduction to the model—ArcGIS pretrained models | Documentation." <https://doc.arcgis.com/en/pretrained-models/latest/imagery/introduction-to-building-footprint-extraction-usa.htm>
- [19] "Pretrained Deep Learning Models | Image Feature Extraction & More." <https://www.esri.com/en-us/arcgis/deep-learning-models>

An Implementation of the NOAA Satellite Data Reception Ground Station for Weather and Climate Enthusiasts

Wichapol Utumporn
*Electrical and Computer Engineering
Dept., Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
wichapol.utu@dome.tu.ac.th*

Dahmmaet Bunnjaweht
*Electrical and Computer Engineering
Dept., Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
dahmmaet@engr.tu.ac.th*

Suwabhat Somboonsub
*Electrical and Computer Engineering
Dept., Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
suwabhat.som@dome.tu.ac.th*

Warinthon Kiadtikornthaweeyot Evans
*Industrial Engineering Dept.,
Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
kwarin@engr.tu.ac.th*

Watchara Amasiri
*Electrical and Computer Engineering
Dept., Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
awatchar@engr.tu.ac.th*

Pirapat Waritkraikul
*Electrical and Computer Engineering
Dept., Faculty of Engineering,
Thammasat School of Engineering,
Thammasat University,
Pathumthani 12120, Thailand
kupp.soy@gmail.com*

Abstract—During the dry season in Thailand, the risk of forest fires increases significantly, posing a serious threat to both the environment and public health. There is a pressing need for weather and climate enthusiasts to access comprehensive data on forest fires, climate patterns, and environmental conditions. This work aimed at implementing a satellite ground station to exemplify enthusiasts' alternative data access other than published data on governmental websites. The goal was to achieve good results with low-cost software-defined radio (SDR) hardware, specifically the RTL-SDR, commercial-off-the-shelf components, and open-source software. Raspberry-NOAA-V2 program suite offers a powerful solution for individuals with minimal or no coding experience, enabling hobbyists to easily set up an automatic satellite imagery reception. With the free reverse proxy functionality from ngrok, the station owner could establish and manage access to the image archive without needing any network knowledge, while their peer enthusiasts could access it from anywhere. Such an approach was considered to offer hobbyists with inquisitive interest to apply technology to build an operational ground station to obtain real-time meteorological data. While the ground station employed inexpensive components, the weather image reconstruction results were relatively satisfactory. The implemented station was a good hands-on activity for weather and climate enthusiasts. By combining the strengths of technology, real-time data, and community awareness, a municipality can create more resilient ecosystems with local engagement that are better prepared to face the challenges of climate variability.

Keywords— NOAA, satellite images, ground station, software defined radio, RTL-SDR

I. INTRODUCTION

In the era of rapid technological innovation and urbanization, many cities have started to integrate technology into urban management to enhance livability and sustainability. These cities leverage a variety of technologies to optimize resource utilization, improve public services, and promote community engagement. Many cities face a major issue in effectively managing forest fires, which are becoming

more common and intense because of climate change and human activity. Addressing this issue necessitates not only improved monitoring and response technologies but also a collaborative approach involving several data sources and community participation. Hot spots, areas with a high fire risk or ongoing forest fires, require precise and fast information to provide efficient management and response.

Knowledge-sharing websites, such as forums, community-driven FAQs, and collaborative wikis, have made information more accessible to everyone. They provide a forum for users, regardless of technical background, to seek advice, exchange their expertise, and contribute to the collective development of knowledge. This helps transform the collaborative setting into a data-sourcing community that helps individuals learn and contribute to the larger digital world. This is where the collaboration of weather and climate enthusiasts and crowdsourcing data can have a transformative impact [1]-[2]. Enthusiasts who monitor local weather conditions and climate trends can provide valuable real-time observations that, when combined with data from professional meteorological sources, result in a more complete understanding of fire risks and environmental conditions. Public awareness is another critical component in the management of forest fires or environmental hazards. Engaging the community through educational initiatives and interactive platforms allows individuals to understand their role in fire prevention and response.

The Polar Orbiting Environmental Satellites (POES) and the METEOR series are essential for weather observation and environmental monitoring. The National Oceanic and Atmospheric Administration (NOAA) operates the POES system, which consists of satellites in polar orbits that circle the Earth from pole to pole. These satellites collect extensive and continuous data on atmospheric conditions, sea surface temperatures, and land surface features. Their polar orbits allow them to capture global data with high temporal resolution, making them invaluable for weather forecasting, climate monitoring, and disaster management [3]. Russian

METEOR series focuses on space-based environmental monitoring. These satellites operate in Low Earth Orbits (LEO) and are equipped with various sensors to collect data on atmospheric conditions, oceanography, and terrestrial environments. METEOR satellites contribute to weather forecasting, climate studies, and environmental observation, complementing the data provided by other satellite systems [4]. Both POES and METEOR contribute significantly to the understanding of the Earth's weather and climate systems.

Flightradar24 is an example of a technology-enabled enthusiast community that uses crowdsourced data. It has revolutionized air traffic tracking by combining data from multiple sources into a single platform. This system relies heavily on Automatic Dependent Surveillance-Broadcast (ADS-B) technology [5]. Satellite Networked Open Ground Station (SatNOGS) is a revolutionary example in the low-cost LEO satellite ground station community. SatNOGS is an open-source hardware and software platform aimed at creating a satellite ground station network. The SatNOGS's concept is to make better use of the deployed ground stations around the world, hence resolving both underutilization and coverage issues [6]. However, Flightradar24 has higher popularity compared to SatNOGS; this can be attributed to several factors such as real-time information, user-friendly interface, community engagement, commercial applications, and marketing and media presence. In contrast, while valuable and impactful, SatNOGS appeals to a more specialized audience interested in satellite communications and space technology. Its particular focus, paired with the more technical nature of its data and setup, may contribute to its lower public profile. Nonetheless, SatNOGS demonstrates how crowdsourcing can improve technological skills and add to a larger pool of data, resulting in more comprehensive and accessible information for both users and researchers.

This paper presents an implementation of an inexpensive NOAA satellite ground station, which could be executed with support from open-source projects, commercial-off-the-shelf parts, and knowledge-sharing online communities. Undoubtedly, the weather data obtained does not replace the precision of government weather stations. Engaging weather and climate enthusiasts with available technology offers educational benefits enriching the collective understanding of meteorological patterns and phenomena. This hands-on experience can deepen understanding and awareness of weather and climate issues. Public awareness is another critical component in the management of forest fires and environmental hazards.

The paper is organized as follows. Section II begins by providing background information. Section III describes how

the ground station was implemented. Section IV presents the implementation results and discussion. Finally, Section V presents the conclusions.

II. BACKGROUND

Software-defined radio offers a flexible, upgradeable solution with a longer lifespan for both military and civilian wireless communications. It can also deliver adaptable and potentially cost-effective terminals for end users [7]. DVB-T dongles that use the Realtek RTL2832U chip can serve as inexpensive software-defined radios, as this chip enables the transfer of raw I/Q samples to the host computer [8]. Over the years, RTL-SDR has gained immense popularity and has made radio spectrum access widely available to everyone. The RTL-SDR can be used in various forms of receivers, ranging from terrestrial radio to satellite signals.

The amateur satellite ground station can be built with commercial-off-the-shelf hardware and open-source software and can receive data from weather satellites in the amateur radio frequency band at 137 MHz. The Automatic Picture Transmission (APT) system offers a reduced resolution, geometrically-corrected analog data stream from the Advanced Very High Resolution Radiometer (AVHRR) sensor. The analog APT signal is transmitted continuously and can be received in real-time by relatively simple, affordable ground station equipment [9]-[10]. The NOAA APT transmitted image resolution is about 4 km, while the High Resolution Picture Transmission (HRPT) image resolution is about 1.1 km.

This work focused only on the meteorological data received from NOAA-15, NOAA-18, and NOAA-19. The basic parameters of these NOAA satellites in APT mode are tabulated in Table 1 [11]- [14].

TABLE I. BASIC PARAMETERS OF NOAA SATELLITES IN APT MODE

Satellite	Parameters			
	Frequency (MHz)	Modulation	Baseband (kHz)	Polarization
NOAA 15	137.6200	FM	1.7	RHCP
NOAA 18	137.9125	FM	1.7	RHCP
NOAA 19	137.1000	FM	1.7	RHCP

There are a variety of ground station configurations for receiving NOAA weather images. Table 2 tabulates some of them to outline different setup possibilities. The diagram of a typical ground station for this purpose is illustrated in Fig. 1.

TABLE II. COMPARISON WITH SOME PREVIOUS WORK ON NOAA WEATHER GROUND STATION

Comparison Aspects	[15]	[16]	[17]	[18]	[19]	This Work
Antenna	X-Quad Yagi	Quadrifilar Helicoidal	Double Cross Dipole	Quadrifilar Helicoidal	V-Dipole	Eggbeater
Computing Unit	PC	PC	PC	Raspberry Pi	PC	Embedded PC
Receiver	Icom PCR	RTL-SDR	NI USRP	RTL-SDR	RTL-SDR	RTL-SDR
Software	SPECTRAN, WXtoIMG, SatPC32, Icom proprietary	SDR Sharp, Audacity, WXtoIMG, Orbitron	Labview, WXtoIMG	rtl_fm, SoX, noaa_apt, Predict	Cubic SDR, WXtoIMG, G-Predict	SDR Sharp, VB-Cable, WXtoIMG, Raspberry-NOAA-V2
Web Accessible	No	No	Yes	No	No	Yes

In addition to the literature on ground stations referenced in Table 2, work focused on decoding mechanism, for example, those implemented using GNU Radio [20] and MATLAB [21].

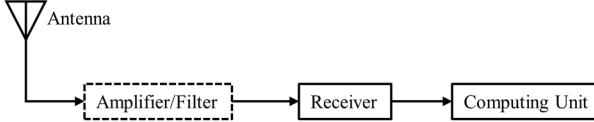


Fig. 1. Diagram of typical weather ground station.

III. IMPLEMENTATION

The implementation setup of the outdoor unit for receiving the signal from the NOAA's weather satellites is shown in Fig. 2. The eggbeater antenna with reflectors was chosen to work with an RTL-SDR V3 dongle. The antenna was installed on the rooftop of a seven-story building topped with a high-density polyethylene waterproof roof membrane. For a non-permanent and non-invasive antenna mounting structure, a metal tripod with additional weight for extra stability was employed. The ground station computing unit was a Mini Tree fanless industrial computer with Intel® Core™ i5-4278U processor running the Linux Mint Debian Edition 5.



Fig. 2. A picture of the outdoor unit of the implemented ground station.

The core processing software was Raspberry-NOAA-V2. This software suite is an open source code and a series of scripts that enable 64-bit Debian systems, as well as 32-bit Raspberry Pi devices as an automated NOAA and METEOR weather satellite station equipped with an SDR such as an RTL-SDR. The software uses WXtoIMG and meteor_decoder to decode the satellites' data, a program named predict to predict satellite passes, and several automatically generated cron scripts to schedule recording and processing [22].

To add network accessibility to a web portal, ngrok functionality was utilized as depicted in Fig. 3. ngrok is a globally distributed reverse proxy that provides security, protection, and performance enhancement for applications and network services, regardless of the deployment location [23]. ngrok demonstrates environment independence by facilitating traffic delivery to services across various platforms, whether hosted on cloud services, a Raspberry Pi, or a conventional computer, without requiring any modifications to the original environment's networking configuration.

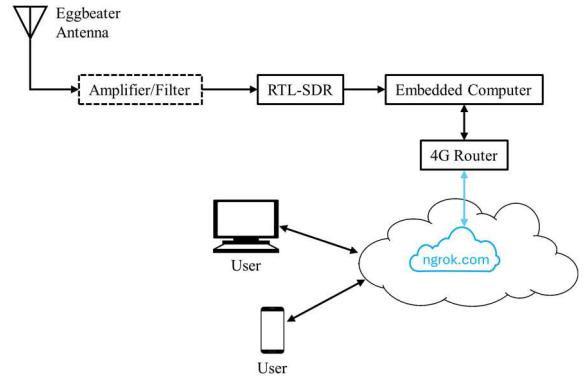


Fig. 3. Conceptual diagram of the implemented ground station.

To improve signal quality, a low noise amplifier (LNA) and/or filter are often used. However, integrating an LNA involves additional considerations. The LNA needs a power supply, which must be carefully chosen and implemented to avoid adding extra noise from the switched mode supply. Specifications and compatibility issues between amplifier characteristics and the front end of the chosen SDR were also critical. If the proper LNA setup was achieved, then the reduction of the gain on the SDR software had to be reconfigured. To avoid additional components, more cabling, and fine-tuning the overall operation, amateurs working on ground stations usually start with a simple setup without supplemental signal enhancers and blockers.

IV. RESULTS AND DISCUSSION

Once the antenna, receiver, and software configuration were completed, the station demonstrated its efficiency through the use of relatively lightweight components, which facilitated rapid re-deployment, if needed. The station could be quickly mobilized for various operational needs. The Raspberry-NOAA-V2 suite provided automated satellite tracking, decoding, processing, and archiving. The output JPG image was annexed with its icon, station's and satellite's information, and enhancement processing as illustrated in Fig. 4.

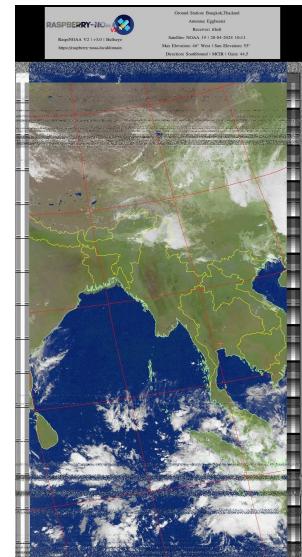


Fig. 4. Decoded image obtained from the NOAA-19 on 20 April 2024 at 10:11 AM.

Storm advisories are essential in tropical countries because they provide critical information about severe weather, allowing people to take necessary measures. On 31 May 2024, the Thai Meteorological Department (TMD) issued a weather advisory for the tropical storm "MALIKSI". Fig.5 shows a weather radar image on 1 June 2024 at 7 PM on the TMD website for public information. The image collected from the NOAA-15 satellite by the implemented ground station on the same day is depicted in Fig. 6.

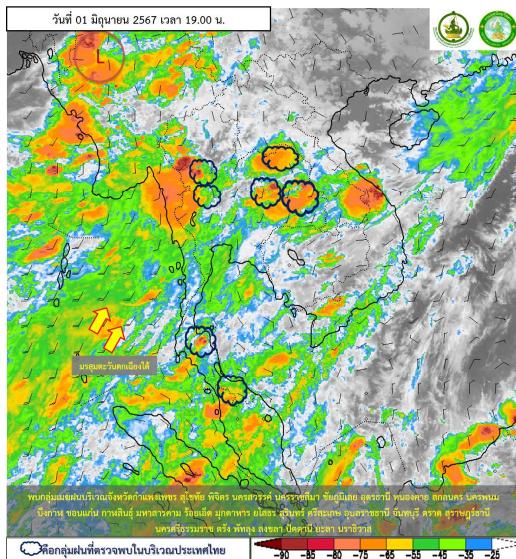


Fig. 5. Weather radar image provided by the TMD on 1 June 2024 at 7 PM.

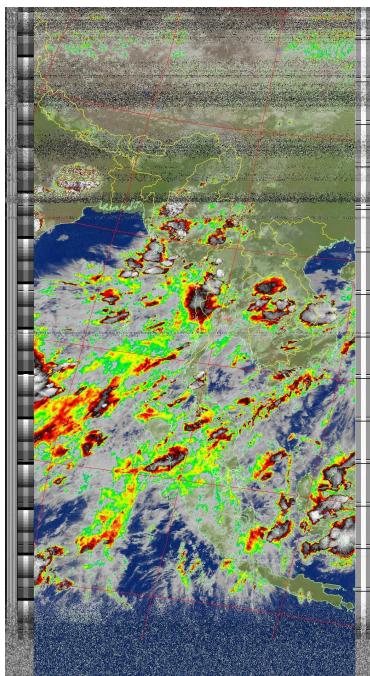


Fig. 6. Decoded image from the NOAA-15 on 1 June 2024 at 7.19 PM.

The burning season is an annual climatic occurrence during the driest period of the year. Forest fires are widespread in Northern Thailand, and often occur throughout the dry season. The Forest Fire Control Division (FFCD), Department of National Parks, Wildlife and Plant Conservation aims to limit the damage caused by forest fires by using all available prevention and suppression strategies. In addition to the

departmental prepared data, the FFCD website portal provides forest fire followers with links to related agencies such as Geo-Informatics and Space Technology Development Agency (GISTDA) in Thailand [24], ASEAN Specialised Meteorological Centre (ASMC) in Singapore [25] and NASA's Fire Information for Resource Management System [26] in the USA.

Fig. 7 shows the hotspot map provided by the GISTDA whose data was derived from the Suomi National Polar-orbiting Partnership (Suomi NPP) with the Visible Infrared Imaging Radiometer Suite (VIIRS). Fig. 8 illustrates the hotspot map from the ASMC using the data collected from NOAA-20.

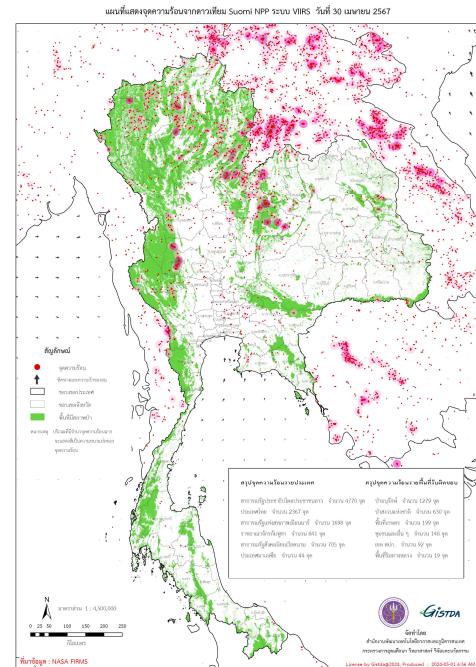


Fig. 7. A map of the hotspot on 30 April 2024 by the GISTDA.

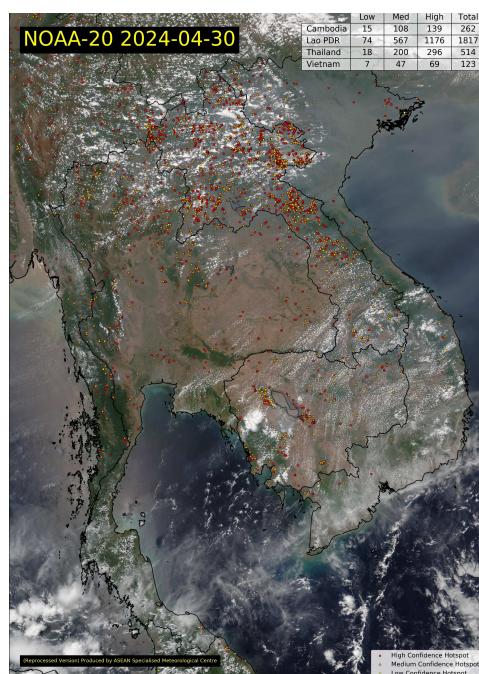


Fig. 8. A map of the hotspot on 30 April 2024 by the ASMC.

For comparison, the thermal images acquired from the implemented ground station are shown in Fig. 9 and Fig. 10.

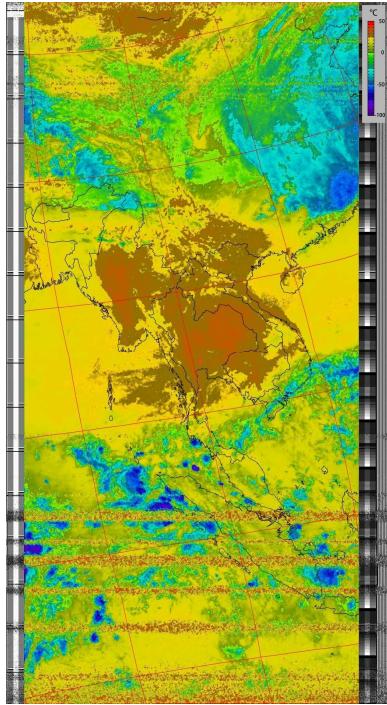


Fig. 9. Decoded thermal image obtained from the NOAA-19 on 30 April 2024 at 9.50 AM.

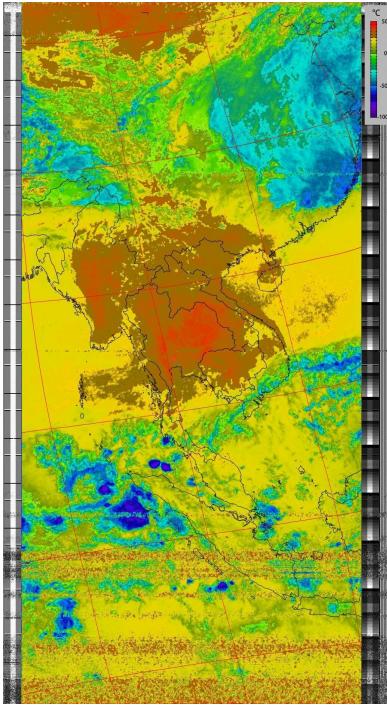


Fig. 10. Decoded thermal image obtained from the NOAA-18 on 30 April 2024 at 11.09 AM.

More advanced enthusiasts may upgrade the station to receive the data in the HRPT mode for higher image resolution [27]-[28]. The basic parameters of the HRPT are shown in Table 3 [11]-[14]. The better resolution comes with the cost of acquiring a better receiver and a higher-performance computing unit. Nonetheless, the HRPT resolution of 1 km could not compete with the VIIRS system on board Suomi

NPP and NOAA-20 with the resolution of 375 m. as depicted earlier in Fig. 7 and Fig. 8.

TABLE III. BASIC PARAMETERS OF NOAA SATELLITES IN HRPT MODE

Satellite	Parameters			
	Frequency (MHz)	Modulation	Data Rate (kbps)	Polarization
NOAA 15	1702.5	BPSK	665	LHCP
NOAA 18	1707.0	BPSK	665	RHCP
NOAA 19	1698.0	BPSK	665	RHCP

Regardless of operating in the APT or the HRPT modes, the NOAA satellites continuously generate a substantial volume of data. Therefore, it is essential to either periodically clear the data drive or attach an external storage. Alternatively, network storage configuration or crowdsourcing platform ability must be deployed.

As discussed earlier, the satellite image from the ground station built by a hobbyist could not match the quality provided by the professional agencies. Many enthusiasts may choose to assemble their ground stations to capture and evaluate weather satellite data directly. This hands-on method enables them to obtain real-time data that is not necessarily accessible via public channels. Setting up personal weather stations and satellite receivers allows these enthusiasts to track specific weather events, monitor climates, and contribute to more localized weather observations.

V. CONCLUSION

This paper presents a method for establishing a ground station to receive APT images from operational NOAA satellites. The ground station allowed for a straightforward implementation with the help of knowledge-sharing websites, making it a simpler and affordable solution. This meteorological station was capable of receiving and decoding good-quality imagery. In addition to educational purposes, the implemented ground station could serve as an alternative to real-time information access for weather and climate enthusiasts. They could also share this weather information with interested community. NOAA-19 is the last of the NOAA series of weather satellites equipped with an APT transmitter. A newer generation of meteorological satellites transmits images with higher resolutions and higher data rates, increasing the complexity and cost of the enthusiasts' ground station. It is simply a matter of time before community-driven, knowledge-sharing websites and open-source tool developers offer alternatives to minimize those obstacles. While government-provided data is vital to the broader public, allowing devoted enthusiasts to access and use their data collection methods constitutes a significant step forward in personal weather monitoring. This dual technique broadens the scope of meteorological information and offers a wider range of viewpoints on climate and atmospheric conditions. Therefore, raising attention and understanding of the effects of climate change on public well-being will facilitate both behavioral change and societal support for environmental awareness.

ACKNOWLEDGMENT

The authors would like to thank various vendors and developers for offering free access to their tools, such as

Raspberry-NOAA-V2, ngrok, and GitHub. The authors also would like to thank the Faculty of Engineering for the facility support and conference expenses.

REFERENCES

- [1] National Meteorological Service, "Crowdsourcing for Weather and Climate Monitoring." <https://blog.metoffice.gov.uk/2022/07/07/crowdsourcing-for-weather-climate-monitoring/> (accessed 2 September 2024).
- [2] M. Giazz et al., "Meteonetwork: An Open Crowdsourced Weather Data System," *Atmosphere*, vol. 13, no. 6, p. 928, 2022. [Online]. Available: <https://www.mdpi.com/2073-4433/13/6/928>.
- [3] National Oceanic and Atmospheric Administration, "POES Space Environment Monitor." <https://www.ngdc.noaa.gov/stp/satellite/poes/index.html> (accessed 29 July 2024).
- [4] World Meteorological Organization, "Satellite: Meteor-M N2-3." https://space.oscar.wmo.int/satellites/view/meteor_m_n2_3 (accessed 29 July 2024).
- [5] Flightradar24. "How Flight Tracking Works." <https://www.flightradar24.com/how-it-works> (accessed 11 August 2024).
- [6] Libre Space Foundation. "Crowd-sourced Satellite Operations." <https://network.satnogs.org/> (accessed 1 July 2024).
- [7] T. Ulversoy, "Software Defined Radio: Challenges and Opportunities," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 531-550, 2010.
- [8] Osmocom. "RTL-SDR." <https://osmocom.org/projects/rtl-sdr/wiki> (accessed 15 July 2024).
- [9] National Oceanic and Atmospheric Administration, NOAA KLM User's Guide, 2014. [Online]. Available: https://www.star.nesdis.noaa.gov/mirs/documents/0.0_NOAA_KLM_Users_Guide.pdf.
- [10] National Oceanic and Atmospheric Administration, User's Guide for Building and Operating Environmental Satellite Receiving Stations, 2009. [Online]. Available: https://www.noaasis.noaa.gov/NOAASIS/pubs/Users_Guide-Building_Receive_Stations_March_2009.pdf.
- [11] National Oceanic and Atmospheric Administration, "POES Performance Status." <https://www.ospo.noaa.gov/operations/poes/status.html> (accessed 29 July 2024).
- [12] World Meteorological Organization, "Satellite: NOAA-15." https://space.oscar.wmo.int/satellites/view/noaa_15 (accessed 29 July 2024).
- [13] World Meteorological Organization, "Satellite: NOAA-18." https://space.oscar.wmo.int/satellites/view/noaa_18 (accessed 29 July 2024).
- [14] World Meteorological Organization, "Satellite: NOAA-19." https://space.oscar.wmo.int/satellites/view/noaa_19 (accessed 29 July 2024).
- [15] V. Dascal, P. Dolea, O. Cristea, and T. Palade, "Advanced VHF Ground Station for NOAA Weather Satellite APT Image Reception," *Acta Technica Napocensis*, vol. 53, no. 3, pp. 1-7, 2012.
- [16] S. Mahmood, M. T. Mushtaq, and G. Jaffer, "Cost Efficient Design Approach for Receiving the NOAA Weather Satellites Data," in 2016 IEEE Aerospace Conference, 2016: IEEE, pp. 1-6, doi: 10.1109/AERO.2016.7500854.
- [17] C. Bosquez, A. Ramos, and L. Noboa, "System for Receiving NOAA Meteorological Satellite Images Using Software Defined Radio," in 2016 IEEE ANDESCON, 2016: IEEE, pp. 1-4, doi: 10.1109/ANDESCON.2016.7836233.
- [18] D. P. Silvaa and J. Eduardo, "piNOAA: An Independent Daily Earth Observation Service Using a Raspberry-Pi Data Processing Platform," in Proc. 73rd Int. Astronautical Congr., 2022.
- [19] A. Waseem, M. A. U. Kamil, A. Bhardwaj, A. Razim, M. A. Qadeer, and T. M. Ghazal, "Weather Satellite Tracking with RTL-SDR: NOAA 15 Image Reception and Signal Decoding," in 2024 2nd International Conference on Cyber Resilience (ICCR), 2024: IEEE, pp. 1-7, doi: 10.1109/ICCR61006.2024.10533107.
- [20] L. C. Nicolae Crisan, "NOAA Signal Decoding and Image Processing Using GNU-Radio," *Acta Technica Napocensis*, vol. 54, no. 2, pp. 36-39, 2013.
- [21] C. Patil, T. Chavan, and M. Chaudhari, "Hardware and Software Implementation of Weather Satellite Imaging Earth Station," in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016: IEEE, pp. 664-670, doi: 10.1109/ICACCI.2016.7732122.
- [22] J. Karimi. "Raspberry-NOAA-V2." <https://github.com/jekhokie/raspberry-noaa-v2> (accessed 5 October 2023).
- [23] ngrok. "What is ngrok?" <https://ngrok.com/docs/what-is-ngrok/> (accessed 29 July 2024).
- [24] Geo-Informatics and Space Technology Development Agency, "Decision Support System for Disaster Management." <https://disaster.gistda.or.th> (accessed 1 May 2024).
- [25] ASEAN Specialised Meteorological Centre, "Satellite Images: Polar-Orbiting Satellite." <http://asmc.asean.org/satellite-polar/> (accessed 1 May 2024).
- [26] National Aeronautics and Space Administration, "Fire Information for Resource Management System." https://firms.modaps.eosdis.nasa.gov/map/#d:24hrs,24hrs;l:fires_all,aqua_crc,topo;@110.1,14.6,4.7z (accessed 1 May 2024).
- [27] C.-T. Tsai and J.-C. Juang, "A Software-Defined Radio Based Adaptive HRPT/APT Signal Receiver," *Journal of Aeronautics, Astronautics and Aviation*, vol. 46, no. 4, pp. 219-223, 2014.
- [28] Supriyono, Kurdianto, A. A. P. B. Dwi, H. Gunawan, and F. Armin, "Software Defined Radio Development for NOAA Satellite Signal Reception HRPT/AHRPT Mode," in 2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2023: IEEE, pp. 382-389.

Monitoring Land Use Changes and Comparing Watershed Quality Classes: A Case Study of Mueang Phetchabun District, Phetchabun Province

Kittikun Nupat

*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
Kittikun.nu@gmail.com*

Niti Iamchuen

*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
niti.ia@up.ac.th*

Thidapath Anucharn

*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
thidapath.an@up.ac.th*

Phongsakorn Hongpradit
*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
67025224@up.ac.th*

Thitisorn Sriprom
*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
67025213@up.ac.th*

Nakarin Chaikae*
*School of Information and Communication Technology
University of Phayao
Phayao, Thailand
Nakarin.ch@up.ac.th*

Abstract— This study aims to investigate land use changes in Mueang Phetchabun District, Phetchabun Province, Thailand, between 2014 and 2024, using Remote Sensing (RS) and Geographic Information System (GIS) technologies. The analysis focuses on land use changes and their comparison with watershed quality classifications. The results reveal that, over the past decade, forest areas expanded by 109.75 km² (10.67%), built-up areas grew by 56.67 km² (52.37%), and water bodies increased by 4.32 km² (45.36%). Conversely, agricultural land saw a significant reduction of 170.74 km² (15.44%). A further comparison of land use changes with watershed quality classifications indicates that by 2024, agricultural land overlapping with Class 1A watershed quality areas declined by 26.77 km² (4.94%), with Class 1B by 39.58 km² (24.64%), and with Class 2 by 55.68 km² (13.71%). Simultaneously, forest areas overlapping with Class 1A increased by 25.18 km² (4.65%), with Class 1B by 36.93 km² (22.99%), and with Class 2 by 51.33 km² (12.64%). Built-up areas overlapping with Class 1A rose by 1.34 km² (0.25%), with Class 1B by 2.57 km² (1.6%), and with Class 2 by 4 km² (0.99%). Water bodies overlapping with Class 1A grew by 0.26 km² (0.05%), with Class 1B by 0.08 km² (0.05%), and with Class 2 by 0.36 km² (0.09%). These findings highlight trends in forest restoration within key conservation zones, alongside community and water resource expansion, underscoring the need for sustainable resource planning. Comparing land use changes with watershed quality classifications provides essential information for relevant agencies to develop balanced and sustainable land management strategies.

Keywords— *Land Use Change, Remote Sensing, Geographic Information System, Phetchabun*

I. INTRODUCTION

Thailand, as a member of the United Nations, has committed to the Sustainable Development Goals (SDGs), notably Goal 13, which emphasizes urgent action to address climate change and its impacts [1]. In recent years, Thailand has experienced significant climate-related challenges that have led to widespread natural disasters, causing extensive damage to homes, communities, and agricultural lands. Rising temperatures during the summer season, coupled with dry vegetation, have increased the likelihood of large-scale forest fires, contributing to escalating pollution levels each year. Communities near these fires face severe health impacts from

hazardous air quality. In Phetchabun Province, where mountainous terrain surrounds the area, residents experience extreme weather conditions in both summer and winter. Between 2021 and 2023, air quality measurements in Phetchabun consistently showed elevated PM2.5 levels, with a significant rise in pollution-related health risks in 2023 [2].

Mueang Phetchabun District, located within this province, represents a critical area for studying land use changes. The district's land use patterns are influenced by the Phetchabun Provincial 5-Year Development Plan (2023–2027), which promotes tourism, agriculture, and sustainable natural resource management [3]. Land use in this area is closely connected with natural resources, particularly forests, which provide essential habitats, support biodiversity, and act as significant carbon sinks by absorbing CO₂. This sequestration role is crucial in mitigating greenhouse gas emissions and aligns with the objectives of SDG Goal 13 [1].

Given the substantial impact that land use changes can have on the environment, economy, and society, it is essential to develop effective tools to monitor, analyze, and manage these shifts. This study investigates land use changes in Mueang Phetchabun District over the period from 2014 to 2024, using Remote Sensing (RS) and Geographic Information Systems (GIS) to analyze trends in land use and to compare these changes with watershed quality classifications. The insights generated aim to guide sustainable natural resource and environmental management strategies for this district.

II. METHODOLOGY

A. Study Area

The study area selected for this research is Mueang Phetchabun District in Phetchabun Province, Thailand. This district features a diverse topography, ranging from lowland areas suitable for agriculture to mountainous regions with steep slopes. The ecological landscape includes forests, rivers, and natural water resources, making it a fertile region ideal for agriculture and forest conservation. In recent years, Mueang Phetchabun District has faced environmental challenges, such as soil and water degradation, increased deforestation for agricultural expansion, and land use practices that are not

well-suited to its geographic features. Comparing land use with watershed quality classifications is an essential process that can guide the restoration of altered areas and inform planning and development aligned with watershed conservation and sustainable resource use [4].

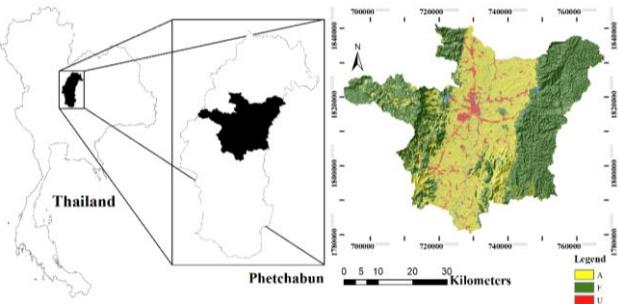


Fig. 1. Example of a figure caption. (figure caption)

B. Equipment and Methods for Land Use Classification

Data preparation for this analysis included obtaining satellite imagery from the United States Geological Survey's (USGS) Earth Explorer platform, accessible at [5]. <https://www.earthexplorer.usgs.gov/>. This platform enables efficient access to remote sensing data and offers multiple search filters for satellite or aerial imagery. For this study, the researcher selected February images from 2014 and 2024, sourced from the LANDSAT-8 and LANDSAT-9 satellites, respectively, to facilitate a supervised classification analysis. February was chosen for its typically clear skies and minimal interference from wildfire smoke, providing optimal conditions for accurate data capture.

The study employed Erdas Imagine software, a robust tool for image management and analysis, recognized for its efficient, user-friendly processing capabilities [6]. Supervised classification was used, allowing flexible and precise land-use categorization [7]. The Maximum Likelihood Classification method was selected due to its statistical foundation, which assesses the probability of pixel values across spectral bands to identify the most probable land-use category [8].

Classification was conducted across four primary land-use categories: agricultural areas (A), forest areas (F), urban and built-up areas (U), and water bodies (W). False color composite (FCC) techniques were used to enhance differentiation, with unique RGB band combinations set at 5:4:3 to highlight forests and water bodies, 6:5:4 for agricultural areas, and 7:6:4 for urban and built-up areas. Google Maps satellite imagery was also utilized for field validation purposes. To confirm classification accuracy, field surveys were cross-verified using Google Maps satellite imagery.

False color composite (Table II) techniques were applied with the following R:G:B band combinations: 5:4:3 for distinguishing forests and water bodies. 6:5:4 for identifying agricultural areas. 7:6:4 for detecting urban and built-up areas. Additionally, Google Maps satellite imagery was used for field survey verification.

TABLE I. TABLE TYPE STYLES

False color Composite Bands	
Band Number	Description
Band 1	Coastal/Aerosol
Band 2	Visible Blue
Band 3	Visible Green
Band 4	Visible Red
Band 5	Near - Infrared
Band 6	Short - Infrared
Band 7	Short - Infrared

(https://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/composites.html)

C. Equipment and Methods for Land Use Classification

The accuracy assessment for classifying land use into four categories was conducted using reference data from the Land Development Department (LDD), Ministry of Agriculture and Cooperatives, a reliable government agency in Thailand [9]. This dataset, representing the current land use, has been recently updated, making it a dependable source for validating the supervised land use classification results.

The study utilized LANDSAT-8 satellite imagery from 2014 and LANDSAT-9 imagery from 2024. After classification, these images were validated against the LDD's land use data by implementing a stratified random sampling method with 256 sample points, as defined by Equation 1. The goal was to achieve an overall map accuracy of 80% with a margin of error of 5%. This approach relied on a simple random sampling [10]. Sample size was calculated using a binomial probability model (Equation 1), to assess the classification accuracy in terms of overall accuracy and Kappa coefficient [11].

$$n = \frac{Z^2(p)(q)}{E^2} \quad (1)$$

where:

n = Sample size

Z = Z-score corresponding to the desired confidence level (1.96 for 95% confidence and round-up to 2)

p = Estimated proportion of the attribute of interest

q = the perfect of 100 - p

E = Margin of error (5% in this case)

$$K = \frac{n \sum_{i=1}^k nii - \sum_{i=1}^k n + i + \dots}{n^2 - \sum_{i=1}^k n + i + \dots} \quad (2)$$

D. Analysis of Land Use Changes

The analysis of land use changes was conducted by comparing the classified land use data derived from the accuracy-assessed LANDSAT-8 satellite imagery of 2014 and LANDSAT-9 satellite imagery of 2024. This comparison aimed to identify changes in land use over the 10-year period across four main categories: water bodies, urban and built-up areas, agricultural land, and forest areas. By examining the differences in these categories between 2014 and 2024, the study provides insights into trends in land use change within the region [12], [13].

E. Comparison of land use and watershed quality classes

This study compares watershed classification and land use patterns using data from Thailand's National Environmental Board, part of the Ministry of Natural Resources and Environment. Thailand's watershed classification system divides watersheds into five quality classes, ranking their importance in terms of water resource conservation and water management needs [14].

The analysis focused on Class 1 watershed areas, split into two key subcategories. Class 1A: These areas are the most critical for conservation, highly sensitive to environmental change, and primarily designated as protected natural forest. Land use in Class 1A zones is either heavily restricted or prohibited to prevent water source degradation and soil erosion. Class 1B: Also, high-priority, these areas allow slightly more land use flexibility than Class 1A. Controlled land use is permitted to preserve the watershed's environmental integrity, with carefully managed agricultural or developmental activities allowed within sustainable limits. Class 2 watersheds, which support secondary water source conservation. Typically found on high mountain slopes or moderate gradients (30-50%), these areas are vulnerable to soil erosion, with lower fertility levels. Both Class 1 and Class 2 watershed areas are essential for maintaining ecosystems, preventing natural disasters, and conserving soil and water resources. The environmental roles of Classes 1A, 1B, and 2 are significant, and inadequate conservation or management in these watersheds could lead to severe soil erosion, depletion of water sources, increased flood risk in low-lying regions, and biodiversity loss. By comparing land use changes with watershed classifications in Classes 1A, 1B, and 2, this study offers valuable insights for sustainable resource planning and management to mitigate long-term environmental impacts [15].

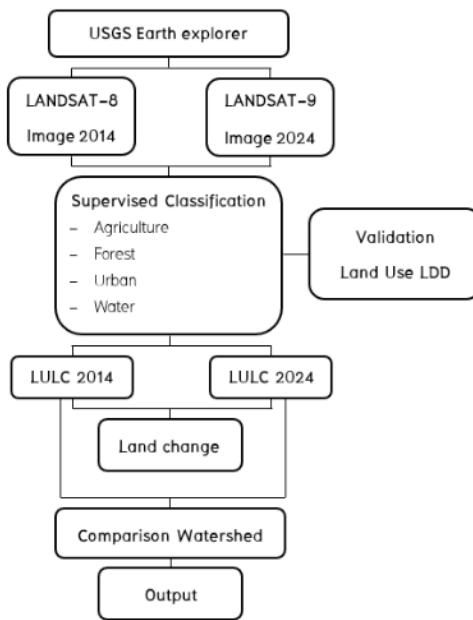


Fig. 2. Process

III. RESULT

A. Characterization and Classification Methods of Land Use

Defining features is a crucial step that enables the model to accurately classify different land types. In this study, four primary land use categories were established: agricultural areas, forest areas, urban and built-up areas, and water bodies. These categories were selected to align with the classifications provided by the Land Development Department of Thailand's Ministry of Agriculture and Cooperatives. The official data classifies land into five categories: agricultural areas, forest areas, urban and built-up areas, water bodies, and barren land. To enhance clarity and precision in classification, barren land was combined with agricultural areas for the purpose of this study.

For the classification process, supervised classification was employed, allowing for flexible categorization and improved accuracy. Specifically, the Maximum Likelihood Classification (MLC) method was chosen due to its statistical foundation and widespread use. MLC analyzes the probability of each pixel value across different spectral bands to determine the most likely land use category for each pixel.

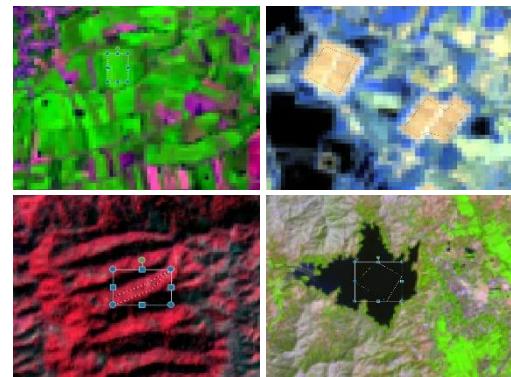


Fig. 3. Training Area

B. Land Use Classification

The results of the land use classification conducted using LANDSAT-8 satellite imagery from 2014 indicate that the majority of the area is as agricultural land, encompassing 1,276.66 km², which accounts for 56.7% of the total area. This is followed by forest areas, which cover 918.35 km² or 40.78% of the total area. Next in rank are built-up areas, which occupy 51.55 km², representing 2.29% of the total area. Finally, water bodies constitute 5.20 km², accounting for just 0.23% of the total area.

TABLE II. LAND USE CLASSIFICATION 2014

Classification 2014		
Class	Area (Sq. km)	Average (%)
A	1,276.66	56.70
F	918.35	40.78
U	51.55	2.29
W	5.20	0.23

The results of the land use classification conducted using LANDSAT-9 satellite imagery from 2024 reveal that the majority of the area is classified as agricultural land, covering 1,105.92 km², which accounts for 49.11% of the total area.

This is followed by forest areas, which encompass 1,028.10 km², representing 45.66% of the total area. Next in rank are built-up areas, covering 108.23 km², or 4.81% of the total area. Lastly, water bodies occupy 9.52 km², accounting for 0.42% of the total area.

TABLE III. LAND USE CLASSIFICATION 2024

Classification 2024		
Class	Area (Sq. km)	Average (%)
A	1,105.92	49.11
F	1,028.10	45.66
U	108.23	4.81
W	9.52	0.42

C. The results of the accuracy assessment of land use classification.

The accuracy assessment of the land use classification derived from LANDSAT-8 satellite imagery from 2014 revealed an overall accuracy of 0.78. Additionally, the Kappa statistic for the results was calculated to be 0.61, indicating a substantial agreement between the classified data and the reference data.

TABLE IV. ACCURACY ASSESSMENT 2014

LULC 2014	LULC LDD.					Accuracy	Kappa
	A	F	U	W	Total		
A	104	5	4	0	113	0.92	0
F	28	98	0	0	126	0.78	0
U	12	0	1	0	13	0.08	0
W	5	0	3	2	10	0.2	0
Total	149	103	8	2	262	0	0
Accuracy	0.7	0.95	0.12	1	0	0.78	0
Kappa	0	0	0	0	0	0	0.61

The accuracy assessment of the land use classification derived from LANDSAT-9 satellite imagery from 2024 revealed an overall accuracy of 0.8. Additionally, the Kappa statistic for the results was calculated to be 0.65, indicating a substantial agreement between the classified data and the reference data.

TABLE V. ACCURACY ASSESSMENT 2024

LULC 2024	LULC LDD.					Accuracy	Kappa
	A	F	U	W	Total		
A	96	9	8	0	113	0.85	0
F	16	108	2	0	126	0.86	0
U	9	0	4	0	13	0.31	0
W	7	0	1	2	10	0.2	0
Total	128	117	15	2	262	0	0
Accuracy	0.75	0.92	0.27	1	0	0.80	0
Kappa	0	0	0	0	0	0	0.61

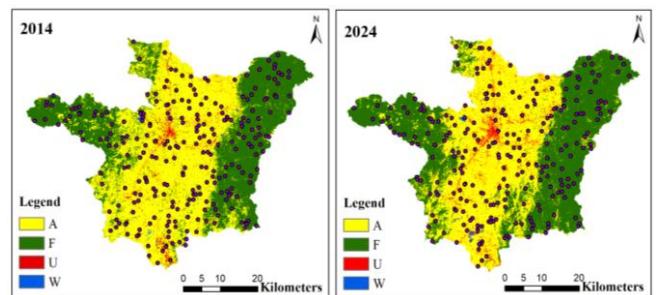


Fig. 4. Supervised Classification and Accuracy Random Point 2024 - 2024

D. Analysis of Land Use Changes (2014–2024).

The analysis of land use changes between 2014 and 2024 reveals notable trends. Forest areas increased by 109.75 km², rising from 918.35 km² in 2014, which reflects an increase of 10.67%. Similarly, built-up areas expanded by 56.67 km², growing from 51.55 km² in 2014, equivalent to a 52.37% increase. Water bodies also experienced growth, with an increase of 4.32 km² from 5.20 km² in 2014, representing a 45.36% rise. Conversely, agricultural areas saw a decline of 170.74 km², declining from 1,276.66 km² in 2014, which translates to a reduction of 15.44%.

TABLE VI. LAND USE CHANGE 2014 – 2024

Land Use Change 2014 - 2024 (Sq km)				
Classification	LULC 2014	LULC 2024	Different	Average (%)
A	1,282.94	1,111.43	171.51	15.43
F	915.30	1,029.15	113.85	11.06
U	48.29	101.82	53.53	52.57
W	5.12	9.22	4.10	44.44

TABLE VII. CHANGE DETECTION

LULC 2024				
LULC 2014	A	F	U	W
A	1028.61	197.55	52.34	4.14
F	80.22	831.10	3.58	0.12
U	2.23	0.18	45.80	0.07
W	0.16	0.00	0.08	4.89

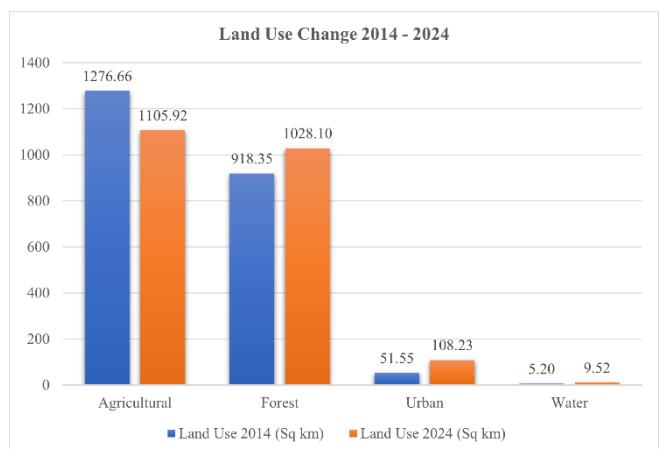


Fig. 5. Comparative graph of land use changes from 2014 to 2024

E. Comparison of Land Use and Watershed Quality Classes.

The analysis comparing watershed quality classes with land use classifications derived from LANDSAT-8 satellite imagery in 2014 indicated significant overlaps among various land use types. Agricultural land intersected with Watershed Quality Class 1A over an area of 75.05 square kilometers, accounting for 13.85% of that class. Additionally, it overlapped with Watershed Quality Class 1B, covering 103.32 square kilometers (64.32%), and with Watershed Quality Class 2, totaling 184.42 square kilometers (45.41%). For forested areas, the overlap with Watershed Quality Class 1A was substantial, at 465.48 square kilometers, which represents 85.87% of that class. The overlap with Watershed Quality Class 1B was 54.79 square kilometers (34.11%), while the area intersecting with Watershed Quality Class 2 was 218.72 square kilometers (53.85%). Community and built-up areas showed an overlap of 1.2 square kilometers with Watershed Quality Class 1A (0.22%), 2.42 square kilometers with Class 1B (1.51%), and 2.91 square kilometers with Class 2 (0.72%). Lastly, water bodies overlapped with Watershed Quality Class 1A by 0.31 square kilometers (0.06%), with Class 1B by 0.11 square kilometers (0.07%), and with Class 2 by 0.09 square kilometers (0.02%). These findings highlight the intricate relationship between land use and watershed quality, emphasizing the importance of integrated land management strategies.

TABLE VIII. WATERSHED COMPARISON CLASSIFICATION 2014

Water shed	LULC 2014							
	A		F		U		W	
	<i>km²</i>	(%)	<i>km²</i>	(%)	<i>km²</i>	(%)	<i>km²</i>	(%)
1A	75.05	13.85	465.48	85.87	1.20	0.22	0.31	0.06
1B	103.32	64.32	54.79	34.11	2.42	1.51	0.11	0.07
2	184.42	45.41	218.72	53.85	2.91	0.72	0.09	0.02

The analysis comparing watershed quality classes with land use classifications derived from LANDSAT-9 satellite imagery in 2024 revealed significant overlaps among various land use types. Agricultural land intersected with Watershed Quality Class 1A over an area of 48.27 square kilometers, accounting for 8.91% of that class. It also overlapped with Watershed Quality Class 1B, covering 63.75 square kilometers (39.68%), and with Watershed Quality Class 2, totaling 128.74 square kilometers (31.70%). For forested areas, the overlap with Watershed Quality Class 1A was substantial, at 490.66 square kilometers, representing 90.52% of that class. The overlap with Watershed Quality Class 1B was 91.72 square kilometers (57.10%), while the area intersecting with Watershed Quality Class 2 was 270.05 square kilometers (66.49%). Community and built-up areas showed an overlap of 2.54 square kilometers with Watershed Quality Class 1A (0.47%), 4.99 square kilometers with Class 1B (3.11%), and 6.91 square kilometers with Class 2 (1.7%). Lastly, water bodies overlapped with Watershed Quality Class 1A by 0.57 square kilometers (0.1%), with Class 1B by 0.19 square kilometers (0.12%), and with Class 2 by 0.45 square kilometers (0.11%). These results highlight the relationship between land use and watershed quality, indicating areas that may require more focused land management practices.

TABLE IX. WATERSHED COMPARISON CLASSIFICATION 2024

Water shed	LULC 2024							
	A		F		U		W	
	<i>km²</i>	(%)	<i>km²</i>	(%)	<i>km²</i>	(%)	<i>km²</i>	(%)
1A	75.05	13.85	465.48	85.87	1.20	0.22	0.31	0.06
1B	103.32	64.32	54.79	34.11	2.42	1.51	0.11	0.07
2	184.42	45.41	218.72	53.85	2.91	0.72	0.09	0.02

The comparison of land use with watershed quality classes between 2014 and 2024 reveals notable changes in overlap areas. In 2024, agricultural land that overlapped with Watershed Quality Class 1A decreased by 26.77 square kilometers, accounting for 4.94%. The overlap with Class 1B diminished by 39.58 square kilometers, representing a 24.64% reduction, while the overlap with Class 2 fell by 55.68 square kilometers, or 13.71%. Conversely, forested areas that intersected with Class 1A increased by 25.18 square kilometers from 2014, reflecting a growth of 4.65%. Overlaps with Class 1B rose by 36.93 square kilometers, marking a 22.99% increase, and overlaps with Class 2 increased by 51.33 square kilometers, representing a 12.64% growth. Additionally, community and built-up areas saw an increase in overlap with Class 1A of 1.34 square kilometers, or 0.25%, while overlaps with Class 1B rose by 2.57 square kilometers (1.6%), and overlaps with Class 2 increased by 4 square kilometers (0.99%). Water bodies also experienced changes, with overlaps in Class 1A increasing by 0.26 square kilometers (0.05%), overlaps with Class 1B growing by 0.08 square kilometers (0.05%), and overlaps with Class 2 rising by 0.36 square kilometers (0.09%).

IV. CONCLUSION

The changes in land use over time indicate a notable decrease in agricultural land, from 56.98% in 2014 to 50.38% in 2024. This shift reflects significant alterations in land utilization. Conversely, forested areas have increased from 40.65% to 45.76% over the same ten-year period. This growth suggests some level of success in the management and implementation of various projects and activities by governmental agencies, in line with the development plans for Phetchabun Province (2023–2027) aimed at sustainable natural resource and environmental management. However, challenges remain, as forest resources continue to face threats, such as wildfires caused by foraging and hunting activities, as well as field burning, particularly during the annual dry season. Another significant reason for forest degradation is the encroachment into forest areas for monoculture cultivation. The Phetchabun Province is home to a diverse array of ethnic groups, particularly the Hmong. Research on the impacts of commercial agricultural production systems on the Hmong community indicates that this ethnic group primarily engages in agriculture, focusing on the cultivation of single crops such as cabbage, corn, beans, ginger, and fruit trees for trade [16]. Additionally, the Hmong reside in high mountainous regions that overlap with forested areas, leading to encroachment and deforestation for agricultural purposes or for selling land to investors. The growth of the industrial and tourism sectors has resulted in increased private sector investment in the area, consequently causing an increase in community and built-up areas from 2.14% to 3.39%. This rise signifies urban expansion and infrastructure development in the region. The

increase in population and local economic development has led to greater demand for land.

This study compares land use with watershed quality classes 1A, 1B, and 2, highlighting a significant decrease in agricultural land overlapping these classes. Specifically, agricultural areas have diminished, while forested regions and water bodies have increased. This trend suggests a positive movement toward the restoration and conservation of natural resources and the environment in critical areas, consistent with conservation policies and sustainable development goals. Over the past decade, stricter conservation measures have been implemented to mitigate the ecological impacts of land use in areas vulnerable to degradation, as outlined in the Phetchabun Province Development Plan (2023–2027) focusing on sustainable management of natural resources and the environment. The growth of community and built-up areas within watershed quality classes is primarily driven by population expansion, infrastructure development, and economic growth, particularly in the agricultural and tourism sectors. These changes emphasize the necessity for relevant authorities to develop management strategies for areas with usage constraints. It is crucial to consider conservation and restoration of vital natural resources and watershed quality classes. This includes urban planning that minimizes negative impacts on ecosystems, promoting suitable community development, ensuring adequate water resource management for agriculture, and encouraging conservation-oriented tourism. Ultimately, this study highlights the trends in land use change, shifts in land utilization within watershed quality classes, and the urgent need for effective governmental policies aimed at sustainable management of natural resources and environmental conservation in the future.

REFERENCES

- [1] United Nations, "Sustainable Development Goals," [Online]. Available: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Accessed: Oct. 28, 2024.
- [2] Pollution Control Department, Thailand, "Watershed Classification and Environmental Risks," 2023, [Online]. Available: <https://www.pcd.go.th>. Accessed: Oct. 28, 2024.
- [3] Phetchabun Provincial Government, "5-Year Provincial Development Plan (2023-2027)," Phetchabun Province, Thailand, 2023.
- [4] Pollution Control Department, Thailand, "Watershed Classification and Environmental Risks," 2023, [Online]. Available: <https://www.pcd.go.th>. Accessed: Oct. 28, 2024.
- [5] United States Geological Survey, "Earth Explorer," [Online]. Available: <https://www.earthexplorer.usgs.gov/>. Accessed: Nov. 3, 2024.
- [6] Erdas Imagine, "Erdas Imagine Software," Hexagon Geospatial, [Online]. Available: <https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine>. Accessed: Nov. 3, 2024.
- [7] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 6th ed. Hoboken, NJ, USA: Wiley, 2015.
- [8] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2016.
- [9] Land Development Department, Ministry of Agriculture and Cooperatives, "Present Land Use Data," [Online]. Available: <https://www.ldd.go.th/>. Accessed: Nov. 3, 2024.
- [10] Southern Research Station, USDA Forest Service, *Forest Science in the South*, 2008 (Science Update SRS-016). Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2008.
- [11] K. Fitzpatrick-Lins, "Comparison of sampling procedures and data analysis for land use and land cover maps," *Photogrammetric Engineering & Remote Sensing*, vol. 47, no. 3, pp. 343-351, 1981.
- [12] Land Development Department, Ministry of Agriculture and Cooperatives, "Present Land Use Data," [Online]. Available: <https://www.ldd.go.th/>. Accessed: Nov. 3, 2024.
- [13] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2009.
- [14] National Environmental Board, Ministry of Natural Resources and Environment, "Watershed Quality Classification in Thailand," [Online]. Available: <https://www.onep.go.th/>. Accessed: Nov. 3, 2024.
- [15] M. E. Baker and M. J. Wiley, "Multi-scale control of flooding and riparian forest conservation in a watershed," *Journal of Environmental Management*, vol. 90, no. 2, pp. 531-544, 2009.
- [16] K. Wongwatthanaphong, "Impacts of Commercial Agricultural Production System on Hmong Ethnic Groups," *Journal of Southern Technology*, vol. 9, no. 1, pp. 69-84, Jan.-Jun. 2559.

Perspective on mortality associated with particulate matter in upper northern Thailand over the 2014–2023 period

Patsanun Lawongyer
*School of Energy and Environment
 University of Phayao
 Phayao, Thailand
 ppatsanun.pim@gmail.com*

Sittichai Pimonsree*
*School of Energy and Environment
 University of Phayao
 Phayao, Thailand
 sittichai007@hotmail.com*

Patipat Vongruang
*School of Public Health
 University of Phayao
 Phayao, Thailand
 patipat7@hotmail.com*

Abstract— The upper northern region of Thailand has long been facing issues with particulate matter (PM) problems. During the study period from 2014 to 2023, PM10 and PM2.5 concentrations exceeded both the Thailand and the WHO guidelines, significantly impacting the local people's health, particularly through diseases associated with PM. This study aims to analyze diseases related to mortality from PM. The findings revealed that the highest mortality rates related to PM-associated diseases were observed in the circulatory system (I00-I99), including heart failure, unspecified (I50.9), essential (primary) hypertension (I10), and heart failure (I50), accounting for 17%, 9%, and 8% of all causes of death, respectively. Additionally, chronic kidney disease, unspecified (N18.9), and senile degeneration of the brain, not elsewhere classified (G31.1) also show high mortality rates in this region, accounting for 8% and 6% of all causes of death, respectively. Correlation analysis revealed a strong relationship between PM10 concentration and heart failure, unspecified (I50.9), with a correlation coefficient of 0.857 ($p = 0.007$). Therefore, the relevant agencies should closely monitor these disease groups and find solutions promptly.

Keywords— particulate matter, diseases, mortality, upper northern Thailand

I. INTRODUCTION

Particulate matter (PM) is a significant pollutant that adversely affects human health. Due to its extremely small size, PM can be deeply inhaled into the respiratory tract and lungs, and enter the bloodstream, circulating throughout the body. This leads to a wide range of harmful health effects, particularly death from diseases related to PM [1-3]. Several epidemiological studies have demonstrated that PM is associated with increased mortality from various causes, including all-cause mortality excluding accidents, cardiovascular diseases, respiratory diseases, diabetes, and neurological disorders [2-6]. The adverse health effects of air pollution may be attributable to both short-term and long-term exposure.

The Upper Northern Thailand has long suffered from severe fine particulate matter pollution, primarily due to biomass burning, which includes agricultural burning and forest fires, contributing more than 70% to overall emission sources [7]. This widespread dispersal of PM across the area makes it one of the region's most significant factors affecting illness and mortality. The study on the relationship between PM2.5 and mortality in Chiang Mai found an 8.9% increase in the risk of death from chronic obstructive pulmonary disease (COPD) and an 8.6% increase in the risk of death from coronary artery disease [8].

This research project is supported by National Research Council of Thailand (NRCT): (Contact No. N41A661134)

It is essential to study diseases associated with PM in Upper Northern Thailand, a region known for its distinct PM pollution and high mortality rates from PM-related diseases. The findings provide critical information to relevant agencies for planning and implementing preventive measures to mitigate the health impacts of air pollution on local communities. Additionally, residents in each area need to be informed about the risks of PM-related diseases so they can take preventive actions. Many studies around the world found many impact of PM on many health outcomes. However, there are few research on relationship between air pollution and only some specific diseases in Northern Thailand. Therefore, this study aims to overview investigate significant mortality-related diseases that may be linked to PM exposure in Northern Thailand over the 2014–2023 period. A preliminary correlation analysis was conducted on mortality rates of selected diseases associated with PM, selected through a literature review.

II. METHODS

A. Data sources

We collected hourly data from 14 monitors of the Pollution Control Department (PCD) in the 8 provinces of upper northern Thailand, including Chiang Mai (35T, 36T), Lamphun (68T), Lampang (37T, 38T, 39T, 40T), Phrae (69T), Nan (67T, 75T), Phayao (70T), Chiang Rai (57T, 73T), and Mae Hong Son (58T), over the period from 2014 to 2023, including the pollutants PM10 and PM2.5 (Fig.2). Quality control of PM data was conducted by following steps: 1) completeness of daily data, comprising at least 18 hours (75% of the day) [5], [9], [10] 2) completeness of monthly data, comprising more than 25 days and 3) completeness of yearly data, comprising all 12 months were included in the analysis. The data from stations exhibiting a correlation coefficient greater than 0.5 were averaged across all stations.

We obtained the cause of death according to the International Classification of Diseases, 10th revision (ICD-10) as reported by the Thailand Ministry of Public Health in Health Region 1, which includes Chiang Mai, Lamphun, Lampang, Phrae, Nan, Phayao, Chiang Rai, and Mae Hong Son provinces, for the years 2014 to 2023.

B. Statistical analysis

Data analyses were performed using descriptive statistics for each variable. The Crude death rate was used to analyze the mortality rate per 100,000 population in each area and disease. Correlation analysis was used to examine the association between PM and selected diseases every year from 2014 to 2023, to identify diseases that may have a

potential correlation with PM. Diseases associated with PM were identified through a literature review of global studies that reported significant impacts of PM on mortality-related diseases. Results were considered significant at $p < 0.05$.

III. RESULTS

A. Particulate matter in Upper Northern Thailand

Data analysis found a serious PM problem in Upper Northern Thailand. PM10 annual mean concentrations from 14 monitors in upper northern Thailand have exceeded the WHO guideline every year (**Fig.1**). At station 36T, PM10 and PM2.5 levels exceed the WHO guidelines, and PM10 levels also surpassed Thailand's standard in 2016. Additionally, the variation of both PM10 and PM2.5 are similar, with correlation coefficient (r) of 0.63. Most PM are fine particles with an average ratio of PM2.5 to PM10 of 0.56 and a standard deviation of 0.22.

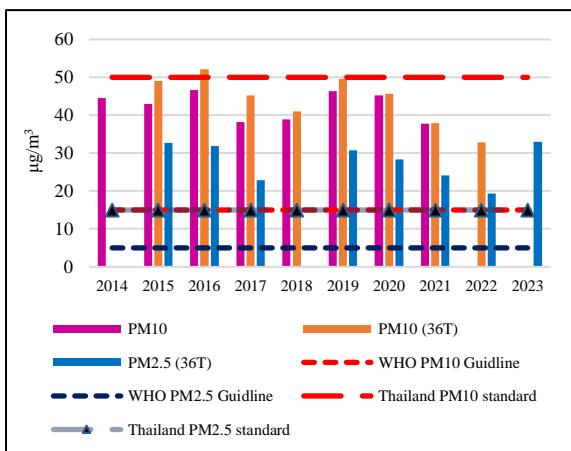


Fig. 1 Annually mean particulate matter concentration in upper northern Thailand from 2014 to 2023.

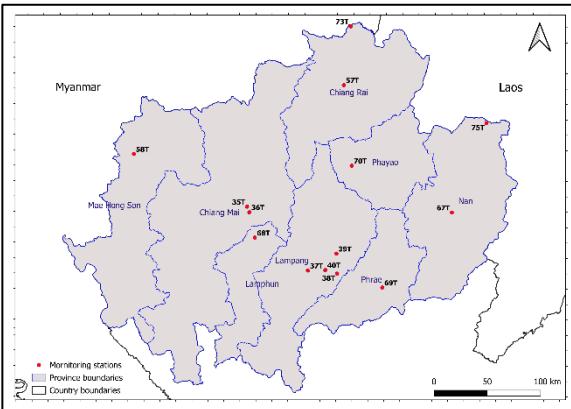


Fig. 2 The locations of air quality stations in upper northern Thailand

B. Cause of death in Upper Northern Thailand

The cause of death in the top 10 diseases in upper northern Thailand was reported in **Table 1**. The highest average mortality rate per 100,000 population for Senility NOS (R54) was 79.52. Following this, the mortality rates per 100,000 population for Heart failure, unspecified (I50.9), Sepsis, unspecified (A41.9), Essential (primary) hypertension (I10), and Heart failure (I50) were 38.88, 27.86, 21.33, and 18.68, respectively. It is noteworthy that 7 diseases associated with PM are among the top 10 causes of death in upper northern Thailand.

The diseases associated with PM that have the highest mortality rates in the top five are unspecified heart failure (I50.9), unspecified hypertension (I10), and heart failure (I50), accounting for 17%, 9%, and 8% of all deaths, respectively. These are followed by senile degeneration of the brain, not elsewhere classified (G31.1), and chronic kidney disease, unspecified (N18.9), which account for 8% and 6% of all deaths, respectively. The mortality rate for the top 10 diseases associated with PM accounts for 71% of all causes of death, which is a significant proportion. Numerous studies confirm that these groups of diseases are associated with PM [1,4,11-13]. (**Fig.3**)

TABLE I. TOP 10 CAUSES OF DEATH IN UPPER NORTHERN THAILAND FROM 2014 TO 2023.

The top 10 cause of death	Average mortality rate per 100,000 population
Senility NOS (R54)	79.52
Heart failure, unspecified (I50.9)	38.88
Sepsis, unspecified (A41.9)	27.86
Essential (primary) hypertension (I10)	21.33
Heart failure (I50)	18.68
Senile degeneration of brain, not elsewhere classified (G31.1)	18.24
Chronic kidney disease, unspecified (N18.9)	13.51
Respiratory failure, unspecified (J96.9)	13.25
Other ill-defined and unspecified causes of mortality (R99)	11.78
Liver cell carcinoma (C22.0)	11.08

Note: Bold letter is death by diseases associated with PM

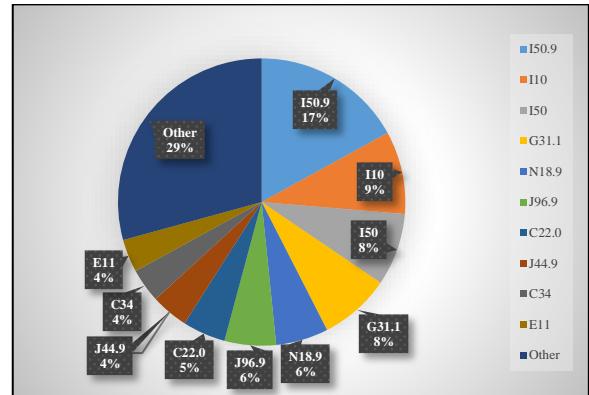


Fig. 3 Percentage of deaths among the top 10 diseases associated with particulate matter in upper northern Thailand (Health region 1) from 2014 to 2023.

The results in **Fig.4 and 5** indicate that comparing the average mortality rate per 100,000 population for the top 10 diseases linked to PM during the periods 2014-2019 and 2020-2023 shows that the four diseases with the highest average mortality rates in both periods are heart failure, unspecified (I50.9), essential (primary) hypertension (I10), heart failure (I50), and senile degeneration of the brain, not elsewhere classified (G31.1). Additionally, essential (primary) hypertension (I10), heart failure (I50), and senile degeneration of the brain, not elsewhere classified (G31.1) show an increasing trend in the period 2020-2023. Moreover, some diseases with increased mortality rates in the 2020-2023 period include type 2 diabetes mellitus (E11). Therefore, these diseases should be closely monitored now and in the future.

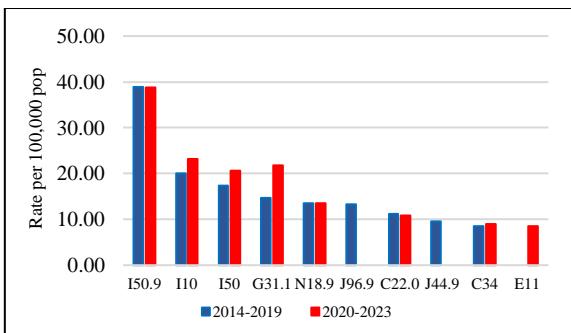


Fig. 4 Compare the average mortality rate per 100,000 population among the top 10 diseases associated with particulate matter in upper northern Thailand from the periods 2014-2019 and 2020-2023.

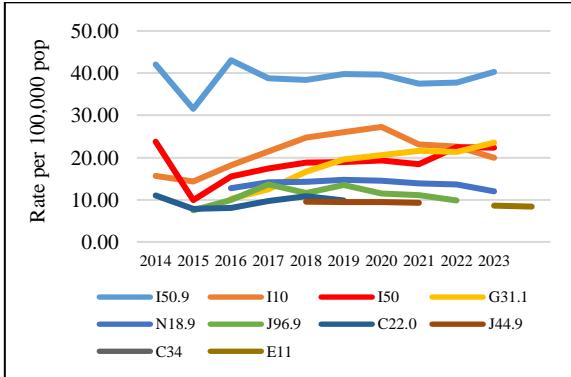


Fig. 5 Trend the average mortality rate per 100,000 population among the top 10 diseases associated with particulate matter in upper northern Thailand from 2014 to 2023.

C. Association between particulate matter and cause of death

The trend of the average mortality rate per 100,000 population among the top 5 diseases associated with PM10 is consistent and particularly evident in heart failure, unspecified (I50.9) (**Fig.6**). A correlation analysis revealed a strong relationship between PM10 concentration and heart failure, unspecified (I50.9) of 0.857.

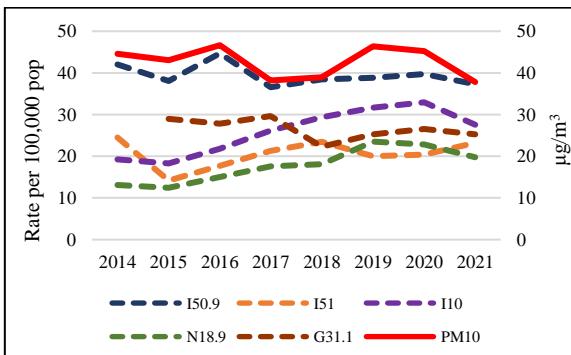


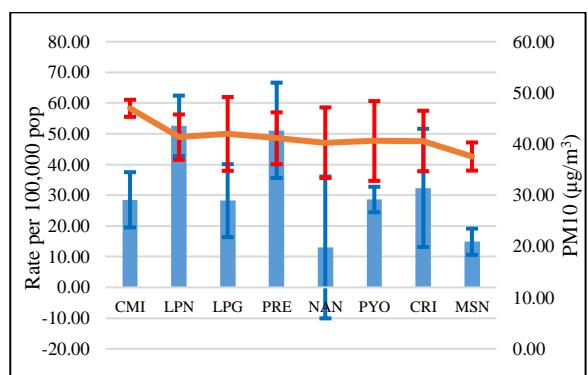
Fig. 6 Trend the average mortality rate per 100,000 population among the top 5 diseases associated with PM10 in upper northern Thailand. (Health region 1) from 2014 to 2021.

The average mortality rate per 100,000 population among the top 5 diseases associated with PM10 in 8 provinces, as shown in **Fig. 7**, indicates that each province exhibits varying trends in the average mortality rate for each disease, which in some areas correspond to the levels of PM10. The findings show that heart failure, unspecified (I50.9), has a high average mortality rate per 100,000 population, with 52.606 in Lampang and 51.136 in Phrae, respectively, which is more

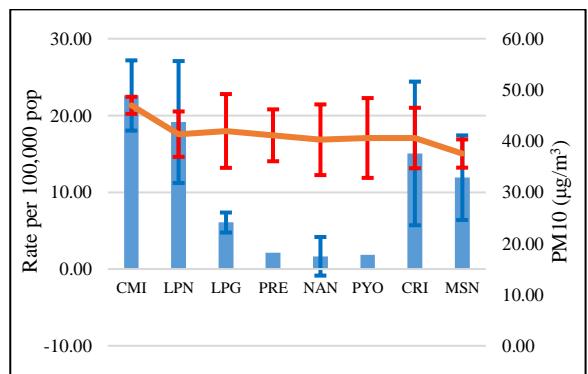
than four times higher than the mortality rate in Nan; heart failure (I50) has a high mortality rate, with 22.603 per 100,000 population in Chiang Mai and 19.145 in Lamphun, which is more than 13 times higher than the mortality rate in Nan. Essential (primary) hypertension (I10) has a high mortality rate, with 42.489 per 100,000 population in Phrae and 33.559 in Phayao, which is more than five times higher than the mortality rate in Mae Hong Son; senile degeneration of the brain, not elsewhere classified (G31.1), has a high mortality rate, with 36.715 per 100,000 population in Lampang, which is more than 55 times higher than the mortality rate in Mae Hong Son; and has a high mortality rate, with 19.664, 15.167, and 14.840 per 100,000 population in Phrae, Lampang, and Phayao, respectively, which is more than five times higher than the mortality rate in Nan. It is evident that each area has different mortality rates for each disease group. The mortality rates of each disease vary across provinces and may be influenced by various local factors. Therefore, it is crucial that each area closely monitors different diseases, depending on which has high mortality rates.

A correlation analysis revealed a strong relationship between PM10 concentration and diseases in 8 provinces, PM10 concentration was associated with Heart failure, unspecified (I50.9) and Essential (primary) hypertension (I10) of 1.000 in Lamphun; Heart failure (I50) of 1.000 in Nan and chronic kidney disease, unspecified (N18.9) of 1.000 in Lamphun. The study results indicate differences in mortality rates across provinces and an unclear trend in the relationship between mortality rates and PM levels, suggesting that other factors may influence mortality and warrant further investigation.

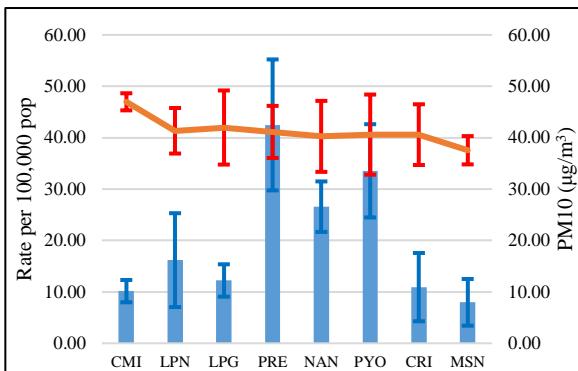
Heart failure, unspecified (I50.9)



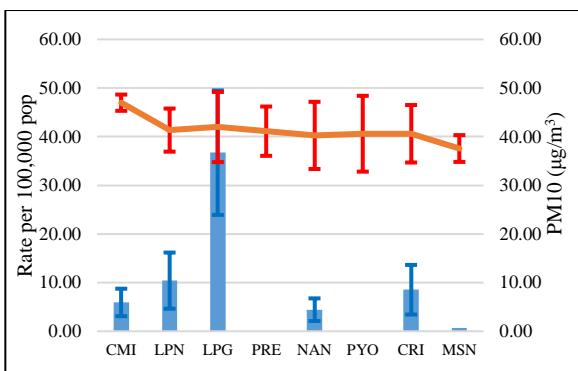
Heart failure (I50)



Essential (primary) hypertension (I10)



Senile degeneration of brain, not elsewhere classified (G31.1)



Chronic kidney disease, unspecified (N18.9)

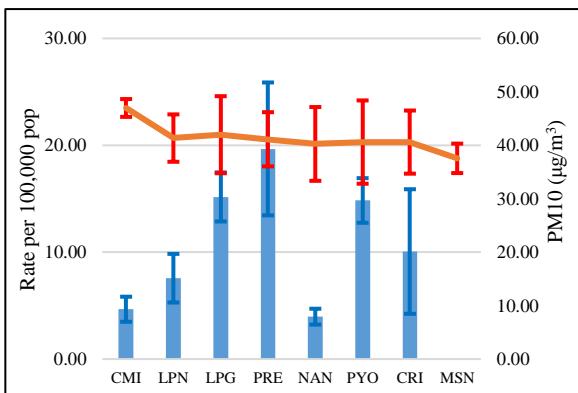


Fig. 7 Average mortality rate per 100,000 population among the top 5 diseases associated with average PM10 in 8 provinces of upper northern Thailand from 2014 to 2021.

Note: CMI-Chiang Mai, LPN-Lamphun, LPG- Lampang, PRE-Phrae, NAN-Nan, PYO-Phayao, CRI-Chiang Rai, MSN- Mae Hong Son.; The bar graph represents the death rate per 100,000 population, while the line graph shows the concentration level of PM.; Standard Deviation and Error Bar.

IV. DISCUSSION

According to the findings of this study, the levels of PM2.5 in Upper Northern Thailand from 2014 to 2023 generally exceeded both Thailand and WHO standards, while PM10 levels remained below the Thailand's standard. Furthermore, the region experiences high mortality rates from diseases linked to particulate matter exposure. Several factors may contribute to this, including demographic changes, such as an aging population, which increases the

likelihood of age-related diseases. Age itself may also be a factor in susceptibility to certain illnesses. Additionally, behavioral factors, including diet, alcohol consumption, smoking, lack of exercise, and stress, significantly contribute to the development of NCDs such as hypertension, diabetes, cancer, cardiovascular diseases, COPD, kidney disease, and dementia. Fine particulate matter is a major environmental factor affecting various body systems. It has been linked to skin diseases [14-16], eye diseases and conjunctivitis [17], muscle remodeling and dysfunction [18], lower bone mineral density (BMD) and weaker bones [19], impact on gut microbiota and intestinal function [20], and reduced live birth rates as well as impaired sperm function [21,22].

Additionally, the study shows that particulate matter plays a role in developing these conditions. Fine particles can enter the lungs, causing inflammation that can spread through the bloodstream to affect the entire body [12]. This process can trigger coagulation mechanisms, cause endothelial dysfunction, and promote cardiovascular changes, such as increased blood pressure and heart rate variability. The study discovered that PM exposure is linked to respiratory disorders such as asthma, lung cancer, COPD, and pneumonia. This is due to the interaction between PM_{2.5} compounds and the respiratory epithelium, which induces an inflammatory response. Long-term exposure to this pollutant can lead to various patterns of respiratory disease development [23]. In addition to impairing lung function and increasing the risk of lung cancer, PM also impacts the human nervous system. There is evidence that PM_{2.5} exposure increases the risk of Parkinson's disease, Alzheimer's disease, memory loss, and dementia. Prolonged exposure to PM_{2.5} can lead to grey matter atrophy, disrupting the tight junctions between neighboring endothelial cells, increasing neuroinflammation, and damaging astrocytes, ultimately causing neurodegeneration.

Additionally, there are reports of kidney damage linked to PM_{2.5} exposure. Studies have shown that PM_{2.5} activates inflammatory pathways, such as the NLRP3 inflammasome, which worsens acute kidney injury [13]. Long-term studies in India have also found that higher PM_{2.5} levels are associated with a decline in the estimated glomerular filtration rate (eGFR), a key indicator of worsening kidney function over time [24].

This study suggests that in addition to the diseases caused by PM that are monitored by the Ministry of Public Health, such as cardiovascular diseases, respiratory diseases, eye diseases, skin diseases, and cancer, there are other causes of death related to PM exposure, including senile degeneration of the brain (G31.1) and chronic kidney disease (N18.9). These diseases show high mortality rates in the Upper Northern region of Thailand. The mortality rates for PM-related diseases also vary between provinces. This study also aligns with the GBD 2021 report, which found that the diseases with the highest death rates are ischemic heart diseases (I20-I25), COVID-19 (U07), and stroke (I60-I69), with age-standardized death rates per 100,000 of 108.7, 94.0, and 87.4, respectively. Ischemic heart diseases and stroke fall within the category of cardiovascular diseases (I00-I99), which should be monitored both in Northern Thailand and globally. Additionally, the GBD report highlights that ambient PM is among the leading risk factors worldwide and at each level of the socio-demographic index [25]. Therefore,

relevant authorities should increase their efforts to monitor these additional diseases. One strength of this study is that it provides valuable preliminary data on diseases linked to PM exposure. Additionally, as numerous factors influence mortality, further in-depth studies are necessary to clarify the relationship between PM and health impacts. Future studies should control confounding factors such as meteorological variables (temperature and relative humidity), day of the week, and long-term trends and should employ advanced statistical methods to establish a clearer relationship between PM exposure and disease-specific mortality.

V. CONCLUSION

Fine particulate matter has been a significant issue in upper northern Thailand over the past decade and is a risk factor affecting the local population, particularly concerning mortality associated with PM. Over the 2014–2023 period, the highest mortality rates have been observed in diseases of the circulatory system (I00-I99), including heart failure, unspecified (I50.9), essential (primary) hypertension (I10), and heart failure (I50), accounting for 17%, 9%, and 8% of all causes of death, respectively. Additionally, chronic kidney disease, unspecified (N18.9), and senile degeneration of the brain, not elsewhere classified (G31.1) also show high mortality rates in this region, accounting for 8% and 6% of all causes of death, respectively. Notably, heart failure, unspecified (I50.9) is significantly correlated with PM10, with a correlation coefficient of 0.857. Therefore, these conditions should be closely monitored by relevant agencies, and further research is needed to estimate the health effects of air pollutants and their relationship with causes of death associated with PM, to prevent PM-related diseases in the area.

ACKNOWLEDGMENT

Thank Pollution Control Department for providing meteorological and air pollutant data. Additionally, we extend our gratitude to Ministry of Public Health for providing health data.

REFERENCES

- [1] X. Xue, J. Chen, B. Sun, B. Zhou and X. Li, "Temporal trends in respiratory mortality and short-term effects of air pollutants in Shenyang, China," *Environ Sci Pollut Res*, vol.25, pp. 11468–11479, January 2018.
- [2] J. Cai, C. Peng, S. Yu, Y. Pei, N. Liu, Y. Wu, Y. Fu and J. Cheng, "Association between PM2.5 exposure and all-cause, non-Accidental, accidental, different respiratory diseases, sex and age mortality in Shenzhen, China," *Int. J. Environ. Res. Public Health*, vol.16, pp.1-26, January 2019.
- [3] T. Liu, H. Meng, M. Yu, Y. Xiao, B. Huang, L. Lin, H. Zhang, R. Hu, Z. Hou, Y. Xu, L. Yuan, M. Qin, Q. Zhao, X. Xu, W. Gong, J. Hu, J. Xiao, S. Chen, W. Zeng, X. Li, G. He, Z. Rong, C. Huang, Y. Du and W. Ma, "Urban-rural disparity of the short-term association of PM2.5 with mortality and its attributable burden," *The Innovation*, vol.2, pp. 1-8, September 2021.
- [4] A. Zanobetti, F. Dominici, Y. Wang and J. D. Schwartz, "A national case-crossover analysis of the short-term effect of PM2.5 on hospitalizations and mortality in subjects with diabetes and neurological disorders," *Environmental Health*, vol.13, pp.1-11, May 2014.
- [5] L. M. T. Luong, D. Phung, P. D. Sly, L. Morawska and P. K. Thai, "The association between particulate air pollution and respiratory admissions among young children in Hanoi, Vietnam," *Science of The Total Environment*, vol. 578, pp. 249-255, August 2016.
- [6] S. E. Yoo, J. S. Park, S. H. Lee, C. H. Park, C. W. Lee, S. B. Lee, S. D. Yu, S. Y. Kim and H. Kim, "Comparison of short-term associations between PM2.5 components and mortality across six major cities in South Korea," *Int. J. Environ. Res. Public Health*, vol. 16, pp.1-13, August 2019.
- [7] P. Vongruang and S. Pimonsree, "Biomass burning sources and their contributions to PM10 concentrations over countries in mainland Southeast Asia during a smog episode," *Atmospheric Environment*, vol. 228, pp. 117414, May 2020.
- [8] C. Pothirat, W. Chaiwong, C. Liwsrisakun, C. Bumroongkit, A. Deesomchok, T. Theerakittikul, A. Limskon, P. Tajarermuang and N. Phetsuk, "Acute effects of air pollutants on daily mortality and hospitalizations due to cardiovascular and respiratory diseases," *J Thorac Dis*, vol. 11, pp. 3070-3083, July 2019.
- [9] N. T. T. Nhungh, C. Schindler, T. M. Dien, N. Probst-Hensch, L. Perez and N. Künzli, "Acute effects of ambient air pollution on lower respiratory infections in Hanoi children: An eight-year time series study," *Environment International*, vol. 110, pp. 139-148, October 2017.
- [10] T. Yang, T. He, J. Huang and G. Li, "Impact of birth season on the years of life lost from respiratory diseases in the elderly related to ambient PM2.5 exposure in Ningbo, China," *Environmental Health and Preventive Medicine*, vol. 26, pp.1-9, June 2021.
- [11] B. Bowe, Y. Xie, Y. Yan and Z. Al-Aly, "Burden of cause-specific mortality associated with PM2.5 air pollution in the United States," *JAMA Network Open*, vol. 2, pp. e1915834, November 2019.
- [12] C. Krittawong, Y. K. Qadeerb, R. B. Hayesc, Z. Wangd, G. D. Thurston, S. Virani and C. J. Lavieh, "PM2.5 and cardiovascular diseases: State-of-the-Art review," *International Journal of Cardiology Cardiovascular Risk and Prevention*, vol. 19, pp. 200217, September 2023.
- [13] H. Pei, X. Dai, Z. He, Z. Tang, Y. Zhu and R. Du, "PM2.5 exposure promotes the progression of acute kidney injury by activating NLRP3-mediated macrophage inflammatory response," *Ecotoxicology and Environmental Safety*, vol. 278, pp. 116454, June 2024.
- [14] S. Deyvapriya, P. Halai, O. Kiss, B. Mambwe, R. Watson, and A. Langton, "P30 Exposure to particulate matter induces structural and pigmentary changes to human skin," *British Journal of Dermatology*, vol. 190, pp. e93-e93, May 2024.
- [15] H. B. Kim, M. G. Choi, B. Y. Chung, J. Y. Um, J. C. Kim and C. W. Park, "Particulate matter 2.5 induces the skin barrier dysfunction and cutaneous inflammation via AHR - and T helper 17 cell - related genes in human skin tissue as identified via transcriptome analysis," *Experimental Dermatology*, vol. 32, pp. 547-554, December 2022.
- [16] H. W. Na, H. S. Kim, H. J. Choi, N. Cha, Y. R. Seo, and Y. D. Hong, "Transcriptome Analysis of Particulate Matter 2.5-Induced Abnormal Effects on Human Sebocytes," *International Journal of Molecular Sciences*, vol. 23, pp. 11534, September 2022.
- [17] P. Upaphong, C. Thonusin, O. Wanichthanoolan, N. Chattipakorn and S. C. Chattipakorn, "Consequences of exposure to particulate matter on the ocular surface: Mechanistic insights from cellular mechanisms to epidemiological findings," *Environmental Pollution*, vol. 345, pp. 123488, March 2024.
- [18] B. Sivakumar and G. A. Kurian, "PM2.5 toxicity in blood impairs cardiac redox balance and promotes mitochondrial dysfunction in rat heart that further aggravates ischemia reperfusion injury by modulating PI3K/AKT/mTOR/NF- κ B signaling axis," *Journal of Biochemical and Molecular Toxicology*, vol. 38, pp. e23718, May 2024.
- [19] Q. Ge, S. Yang, Y. Qian, J. Chen, W. Yuan, S. Li, P. Wang, R. Li, L. Zhang, G. Chen, H. Kan, S. Rajagopalan, Q. Sun, H. F. Zheng, H. Jin, and C. Liu, "Ambient PM 2.5 Exposure and Bone Homeostasis: Analysis of UK Biobank Data and Experimental Studies in Mice and in Vitro," *Environmental Health Perspectives*, vol.131, pp. 107002, October 2023.
- [20] Y. S. Son, N. Son, W. D. Yu, A. Baek, Y. J. Park, M. S. Lee, S. J. Lee, D. S. Kim and M. Y. Son, "Particulate matter 10 exposure affects intestinal functionality in both inflamed 2D intestinal epithelial cell and 3D intestinal organoid models," *Frontiers in Immunology*, vol.14, pp. 1168064, June 2023.
- [21] S. Leathersich, C. Roche, M. Walls, E. Nathan, and R. Hart, "O-075 Particulate matter (PM2. 5 and PM10) exposure prior to oocyte collection is associated with decreased live birth rates in subsequent frozen embryo transfers," *Human Reproduction*, vol. 39, July 2024.
- [22] X. Wei, Z. Zhang, Y. Gu, R. Zhang, J. Huang, F. Li, Y. He, S. Lu, Y. Wu, W. Zeng, X. Liu, C. Liu, J. Liu, L. Ao, F. Shi, Q. Chen, Y. Lin, J.

- Du, G. Jin, Y. Xia, H. Ma, Y. Zheng, R. Huo, J. Cao, H. Shen and Z. Hu, "Inter- and trans-generational impacts of real-world PM2.5 exposure on male-specific primary hypogonadism," *Cell Discovery*, vol. 10, pp. 1-16, April 2024.
- [23] A. Garcia, E. E. Santa-Helena, A. D. Falco, J. D. P. Ribeiro, A. Gioda, and C. R. Gioda, "Toxicological effects of fine particulate matter (PM2.5): Health risks and associated Systemic Injuries—Systematic Review," *Water Air Soil Pollut*, vol. 234, pp. 1-23, May 2023.
- [24] S. Mandal, R. Jagannathan, S. Anand, M. Kabayam, V. Narayan, and V. Mohan, "Long-term exposure to PM2.5 associated with decline in kidney function in a large urban cohort in two Indian cities," In *ISEE Conference Abstracts*, vol. 2023, September 2023.
- [25] Institute for Health Metrics and Evaluation (IHME), *Global Burden of Disease 2021: Findings from the GBD 2021 Study*. Seattle, WA: IHME, 2024.

Author Index

Ali, Nor Sukor	185
Amasiri, Watchara	248
Anucharn, Thidapath	102, 254
Anurakjaturong, Chitsanu	174
Aobpaet, Anuphao	27
Aungthitipan, Poramed	16
Auynirundronkool, Kridsakron	86
Bakker, Mark	27
Bantayan, Nathaniel	144
Batubara, Mario	108
Bidin, Mohamed Noor Azman	185
Boonanun, Nitis	174
Boonchaisri, Siriwat	196
Boonlerd, Viphada	139
Boonriam, Warin	6
Boonsuk, Boonchuang	6
Boonthalarath, Suriyawate	114
Buaniam, Chatchaya	232
Budiantoro, Poki Agung	108
Bujal, Noor Ropidah	185
Bunnjaweht, Dahmmaet	248
Chaichakan, Chatthip	22
Chaikaew, Nakarin	102, 114, 196, 254
Chaisak, Nattanon	232
Chaiwongsen, Panupong	6
Chaiwongsai, Jirabhorn	1, 74, 90
Chalamkate, Tuvachit	49
Chanchai, Chaturawit	22
Chansury, Woranut	139, 232
Chantee, Surachai	90
Chanthakhhot, Aitsara	16
Chayawat, Chompunut	1
Chitapornpan, Sukhuma	22, 37
Chompuchan, Chuphan	201
Choosumrong, Sittichai	191
Chuenkamol, Saruda	49
Deeprasertkul, Prattana	74, 90
Dehghani, Alireza	31
Do, Thi Viet Huong	210
Evans, Warinthorn Kiadtikornthaweeeyot	248

Fathrio, Ibnu	108
Hashim, Mokhtar	185
Hoang, Anh-Huy	152
Homklin, Supreeda	22
Hongpradit, Phongsakorn	102, 254
Hongthong, Sukanya	16
Hooper, Andy	27
Horanont, Teerayut	179
Hussadin, Areewan	49
Iamchuen, Niti	80, 102, 254
Imman, Saksit	16, 65, 69, 159
Intana, Sorrachat	37
Intarat, Kritchayan	49, 226
Jaemklang, Bantita	96
Jangsaeng, Woraman	226
Jantaput, Chattakan	196
Jantawan, Watcharin	196
Janthakot, Athicha	16
Jeefoo, Phaisarn	43, 90, 114, 162, 206
Jindasee, Pongsak	74
Jusoh, Shaizan	185
Kantawong, Krittika	1, 74
Kanya, Kankaew	96
Khobkhun, Boonyasith	1
Khosungnoen, Sumitta	168
Khrueawan, Pranorm	196
Khrutsuwan, Wilawan	174
Kitratporn, Nuntikorn	162
Kitratporn, Nuntikron	232
Kleawkla, Achara	196
Kreetachat, Torpong	16, 65, 69, 159
Kristiana, Yopie	108
Lathif, Musthofa	108
Lawongyer, Patsanun	260
Lazecky, Milan	27
Lerk-U-Suke, Sawarin	11, 74, 162, 237, 243
Li, Yilin	31
Loungon, Wattana	69
Malithong, Pensri	11
Malithong, Sutti	11
Manajitprasert, Supaporn	86

GIS IDEAS 2024

Author Index

Maomool, Kannika	243
Maskhunthod, Pattaranun	16
Md Dali, Syazana	185
Meeprom, Phurith	86
Moolchan, Teansiri	232
Moukomla, Sitthisak	86
Naeije, Marc	27
Nak-Eiam, Sorasak	6
Nguyen Thi Thu, Nga	125
Nguyen, Phuong Nghi	210
Nguyen, Thi Thanh Ninh	210
Nguyen, Tien-Thanh	152
Niemnil, Sommart	27
Noikeaing, Sirilux	139
Nuangjamnong, Nithima	49, 226
Nuangjamnong, Panu	162
Nuangjumnong, Panu	139
Nuntakwang, Arthit	6
Nupat, Kittikun	254
Ooppakarn, Theerasak	237
Ountuang, Sopita	174
Paengwangthong, Wipop	206
Panboonyuen, Teerapong	226
Peekhunthod, Thanakrizt	43
Petchprayoon, Pakorn	139
Phaensanthia, Buachomphu	96
Phan Quo, Yen	125
Phinyoyang, Athiwat	96, 168
Photong, Chaiwat	22
Phunak, Kodchakon	168
Phuphaniat, Marut	206
Pimonsree, Sittichai	260
Pinkaeo, Montree	90
Pinmongkhonkul, Sitthisak	6
Pintong, Itthirit	232
Piyathamrongchai, Kampanart	191
Pooriwitayawattana, Kitthiya	174
Pragot, Wanawan	22
Prasomsup, Wilawan	96, 168
Prasongsri, Kednipa	179
Prathumchai, Kulapramote	174
Promaooh, Chayapol	49
Rachata, Napa	74
Rangsaritvorakarn, Nattaphon	80

Rattanaburi, Pheeraphong	139
Rattanasupa, Sansita	139
Rojanavasu, Pornthep	1, 90
Rungsipanich, Anusorn	232
Sae-Jung, Jojinda	226
Senasutthiphany, Pemmanee	11
Simons, Wim J.F.	27
Singharath, Mesha	43
Sirisakorn, Kasidet	201
Sittichok, Ketvara	201
Skawsang, Sukij	162
Slangen, Aimée	27
Somboonsub, Suwabhat	248
Somprasert, Somanat	22, 37
Soonthornrangsang, Jenny	27
Sriprom, Thitisorn	102, 254
Subarkah, Heru	108
Suk-Ueng, Krittawit	11
Sukawattanavijit, Chanika	1
Sulaeman, Asep	108
Suriyachai, Nopparat	65, 69, 159
Suttara, Samniang	243
Suwannahong, Kowit	16, 69
Suwanpat, Pimsiri	22
Tanarat, Satawat	37
Tansomros, Pornmongkol	16
Test Submission, Krittika	150
Thanarat, Preut	69
Thianthai, Porutai	162
Thirakultomorn, Tinn	96, 168
Thongley, Thongley	220
Thongta, Kornganok	22
Threenet, Ekawit	196
Titayavan, Manas	6
Tong Thi, Hanh	125
Tookhokkruad, Supattra	174
Tran, Thuong	144
Trieu, Hoa	144
Tshering, Younten	220
Udon, Natima	191
Utumporn, Wichapol	248
Van, Ngoc Truc Phuong	55
Vongruang, Patipat	260
Vossepoel, Femke	27

GIS IDEAS 2024

Author Index

Vu, Danh-Tuyen	152
Waritkraikul, Pirapat	248
Weerasai, Khatiya	159
Wiriyaaalongkorn, Winai	196
Wongcharee, Surachai	16, 69
Woottipriyatorn, Pannatorn	102
Xu, Junke	31
Yoomee, Patimakorn	226
Yutthapan, Apantri	139

