Predictive modelling of sea debris around Maltese coastal waters

Mark Dingli

Supervisor: Dr Kristian Guillaumier

May 2024

L-Università ta' Malta
Faculty of Information &
Communication Technology

Abstract

The accumulation of sea surface debris around the coastal waters of Malta, presents numerous ecological and environmental challenges that negatively affect both marine ecosystems and human activities. This is exacerbated by the absence of an effective system that can predict their movement, making it more challenging to address and mitigate this issue effectively.

The primary objective of this project was to develop a system that can predict dispersion patterns of sea surface debris around Malta's coast. To achieve this, we developed a comprehensive machine learning and physics-based pipeline. This pipeline uses historical sea surface current data to predict future conditions, while also having the ability to visualise the movement of debris.

Central to this system is the integration of LSTM and GRU models, trained to predict the next 24 hours of sea currents within a specific area. These predictions were subsequently utilised by the Lagrangian model to visualise the movement of surface debris, offering insights into future dispersion patterns.

A comparative evaluation was conducted for both models, examining the accuracy of their predictions and the quality of the simulations generated by the Lagrangian model, based on these predictions. The results indicated that the LSTM model outperformed the GRU model. This was evidenced by the LSTM's enhanced precision in forecasting the movements of sea surface currents, thereby providing a more reliable basis for the subsequent simulation of debris dispersal patterns.

Overall, this project offers a novel approach to addressing the challenge of sea surface debris around Malta. By harnessing the power of machine learning in tandem with a physics based Lagrangian model, we have established a framework that not only predicts sea surface currents with notable accuracy, but also visualises the movement of surface marine debris, allowing us to make more informed decisions about our environment and our effect on it.

Acknowledgements

I would like to extend my deepest gratitude to several individuals whose support and guidance were invaluable in the completion of this project.

Foremost, I express my heartfelt appreciation to my supervisor, Dr Kristian Guillaumier, for his unwavering guidance, encouragement, and insightful critiques throughout the whole project. His expertise and dedication were instrumental in navigating the complexities of this project and in pushing the boundaries of my academic capabilities.

My sincere thanks also go to Dr Adam Gauci, who not only provided the essential data for this project but also offered his expertise and support. His contributions have been pivotal in enriching the quality of this work.

Additionally, I owe a profound debt of gratitude to my family, especially my parents and sister. Their unwavering support and love have been the cornerstone of my resilience throughout this journey. Their belief in my potential and constant encouragement has been instrumental in empowering me to pursue my goals.

Lastly, I must express profound appreciation to my girlfriend, Ilenia, and my close friends. Their understanding, patience, and encouragement have provided me with the motivation needed to persevere through the challenges of this project.

To all mentioned, and to those who contributed, I am eternally thankful. Your roles in this academic endeavour have left an indelible mark on both the project and my personal growth.

Contents

List of Figures

List of Tables

No table of figures entries found.

List of Tables

No table of figures entries found.

List of Abbreviations

FYP   Final year project (Style: Abbreviations)

LSTM   Long short-term memory

GRU   Gated recurrent unit

NetCDF   Network Common Data Form

Note that the List of Abbreviations should be sorted on the acronym list.

1 Introduction

This project is an integration of machine learning techniques with a physics based Lagrangian model [1] to address the environmental issues of sea debris. At the core of this project is a pipeline that harnesses historical data to forecast future conditions, specifically predicting the next 24 hours of sea surface currents. These predictions serve as inputs for a Lagrangian model [1], enabling it to simulate the movement of surface marine debris. Finally, a comparative evaluation of both LSTM and GRU models is conducted, focusing on their predictive accuracy and the quality of the visualizations. This project introduces an approach of merging machine learning with a physics-based model, offering valuable insights to marine conservation efforts and improving decision-making for managing marine debris around the Maltese Islands.

1.1   Problem Definition

Sea surface debris around the coastal waters of Malta presents a significant environmental challenge. Predominantly composed of plastics, which constitute 82% of all man-made floating items encountered in the Mediterranean sea [2], this debris endangers marine life, disrupts ecological balances, and undermines the ecological integrity of coastal areas [3]. This problem is further aggravated by the lack of an effective system that can predict and forecast the movement of this surface debris, since as of writing, there exists no system that adequately addresses this challenge specifically for the coastal areas around Malta. This further underscores the need for a system that can accurately predict and visualise the dispersion patterns of sea surface debris.

1.2   Motivation

The geological characteristics of the Mediterranean sea makes it difficult for surface debris to escape the area naturally, leading to the accumulation of sea surface debris [4]. The current absence of a predictive system tailored to the coastal regions of Malta impedes effective interventions to mitigate environmental harm. This gap opens an opportunity for the implementation of a system that through the application of Machine Learning and physics-based modelling, aims to address an urgent

ecological issue, which is widely recognised as a global crisis [5]. By fulfilling this need, the project aims to provide accurate predictions that can guide effective cleanup operations and inform strategies for long-term marine conservation around the surrounding waters of Malta.

1.3   Aims and Objectives

The aim of this project is to create a system enhanced with Machine Learning for simulating and predicting the movement of marine debris in the coastal waters of Malta, thereby supporting marine conservation efforts. To achieve this aim, the following objectives have been identified:

01. Data integration: To preprocess and integrate the sea surface currents datasets ensuring compatibility and consistency for input into both models.

02. Lagrangian model development: To utilize develop a Lagrangian physics-based model for simulating the movement of surface marine debris, employing historical data to ensure accurate simulations.

03. AI models development: To develop and fine-tune both LSTM and GRU models for the prediction of future sea surface currents. These models will serve as a crucial component of the forecasting system, leveraging their respective strengths in time series data processing to ensure robust and accurate predictions.

04. Integrating the AI models with the Lagrangian model: To integrate the model's predictions into the Lagrangian model. This integration aims to create future simulations and visualisations of marine debris movement, enhancing the project's predictive capabilities for marine conservation.

05. Comparison of AI models: To conduct a comparative evaluation of both LSTM and GRU models, focusing on their predictive accuracy and the quality of the final visualizations.

1.4   Proposed Solution

This project aims to develop an integrated pipeline for predicting and simulating the movement of marine debris around Malta's coastal waters. The process begins with the preprocessing of the sea surface currents datasets that will be used as input for

the subsequent modelling stages. A Lagrangian model will be developed to visualise the debris movement. This approach is designed to clarify both the expected input from the AI models and the expected nature of the ensuing visualizations. The core of the solution involves developing and fine-tuning two types of machine learning models: LSTM and GRU. These models will undergo extensive testing to determine the optimal architecture and hyperparameters, aiming to accurately predict sea surface currents for a future 24-hour period.

Upon establishing the predictive models, the pipeline integrates these predictions into the Lagrangian model, transforming the predicted data into dynamic visualisations of future debris movement. The project culminates in a comparative analysis of the LSTM and GRU models, evaluating their effectiveness through various metrics, including their predictive accuracy and the quality of the generated visualisations. By analysing the results and visualisations, this project aims to provide actionable insights for effective cleanup operations and strategies for long-term marine conservation around the coastal waters of Malta.

1.5   Summary of Results

To force Word to automatically update the cross-referencing, select the entire document by pressing CTRL-A on your keyboard, followed by F9.

1.6   Document Structure

The remainder of this document is organised into the following chapters:

Background: Here, the foundational elements of the project are discussed. This chapter includes a thorough overview of the utilized datasets, an explanation of the Lagrangian model's principles and capabilities, and an insight into the Machine Learning models.

Literature Review: In the literature review, we will delve into existing research and findings relevant to marine debris, the use of Lagrangian models, and the application of different AI models in environmental forecasting, establishing the scientific grounding for the project's methodologies.

Methodology: This section details the processes undertaken in implementing the FYP. It includes the steps involved in data integration, the development and integration of the Lagrangian and AI models, and the comparative evaluation of the AI models.

Evaluation: A comprehensive outline of the strategies employed to test and evaluate the effectiveness and reliability of the implementation is presented in this section. This will be followed by the presentation and discussion of the results.

Conclusion: This FYP is concluded by summarizing conducted work, revisiting the aims and objectives, acknowledging any encountered limitations, highlighting obtained results, and finally suggesting any proposals for future work.

2 Background and Literature Review

In this chapter, we provide a comprehensive background for the project, while also presenting an overview of pertinent academic papers and literature to underpin the project's scientific foundations.

2.1 Background

This section is divided into five distinct segments where we collectively provide detailed information regarding this project. First, we begin with an exploration of the impacts of marine debris on ecosystems and also delve into the specific datasets used in this project. Then, we discuss the intricacies of the physics based Lagrangian Model, provide an explanation of time series analysis, and finally round off with an exposition on deep learning models, specifically LSTMs and GRUs.

2.1.1 *The impact of marine debris on ecosystems*

The environmental and ecological impact of marine debris, particularly in coastal and marine ecosystems, has been extensively researched, as evidenced by [6] and [7]. Several studies in this area reveal significant negative effects, ranging from harm to marine wildlife due to ingestion and entanglement [8], to the disruption of natural habitats [9]. The impact on coastal ecosystems extends beyond the environment, affecting economic sectors reliant on marine health, such as tourism and fishing as discussed in [9]. Further research delves into the long-term ecological consequences, highlighting the urgent need for effective management and mitigation strategies as seen in [10]. These studies collectively emphasize the critical nature of addressing marine debris for ecosystem sustainability and conservation.

2.1.2 *The Dataset*

The dataset forms the backbone of any project, with its selection and preprocessing being crucial for creating subsequent models. In this project we utilize a single type of dataset which is provided by the Department of Geosciences at the University of Malta.

This dataset consists of sea surface currents velocity data, recorded in hourly increments across four years, spanning from January 2020 to December 2023. These

data points are derived from a model generated by high-frequency (HF) radar systems [11], located on the northern parts of the Maltese islands and southern Sicily. The locations of these radar systems, depicted in Figure 2.1 and identified from [12], provide a temporal snapshot of the sea surface currents movements.
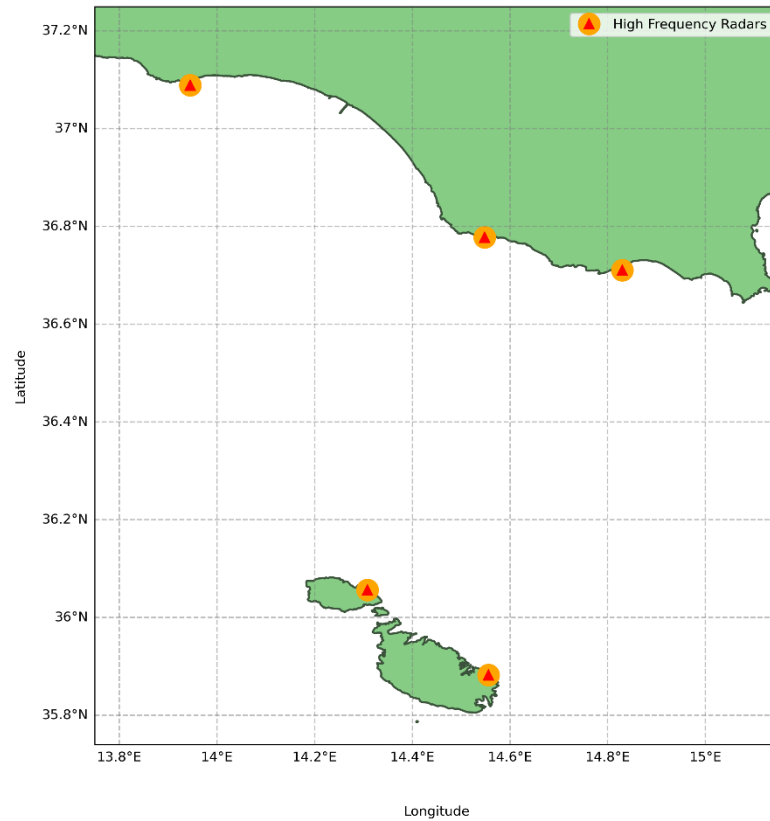

Figure 2.1 High Frequency Radars Locations

The data is composed of several variables including longitude, latitude, and time, coupled with eastern and northern sea current velocities; denoted as 'u' and 'v'. The variable 'u' signifies the east-west current component, indicating the velocity at which surface currents travel horizontally, either eastwards (positive 'u') or westwards (negative 'u'). Similarly, 'v' represents the north-south current component, denoting vertical movement towards the north (positive 'v') or south (negative 'v'). The data's geographical scope is defined within the boundaries of 14.15 to 14.81 degrees longitude and 35.79 to 36.30 degrees latitude. This coverage translates into a grid of 52 latitude points by 43 longitude points, for a total of 180 data points, as detailed in Figure 2.2. The dataset is in NetCDF format [13], a commonly used standard for

climate and meteorological data, ensuring compatibility with the Lagrangian Model employed in the simulations.
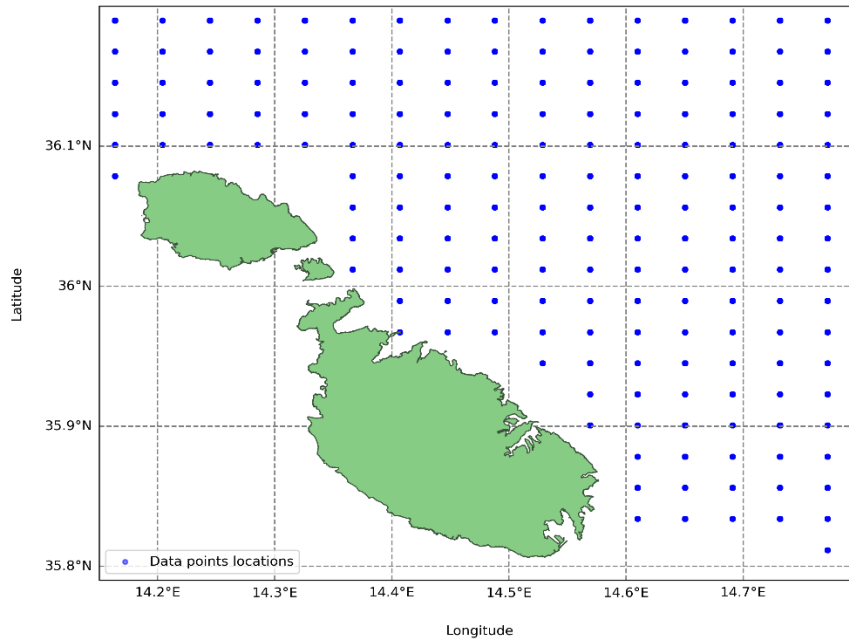

Figure 2.2 Radar Data Points Locations for Dataset

This dataset is integral to the project, providing comprehensive environmental parameters essential for the subsequent development of both the physics-based simulations and the Machine Learning models.

2.1.3 Physics-Based Lagrangian Model

The practice of tracking ocean surface movements in a Lagrangian framework dates back to the earliest days of oceanography, with early methods involving observing the drift of ships or the paths of specially designed floats to document current movement, as outlined by [14]. The physics based Lagrangian model [1] plays a pivotal role in environmental simulations. By offering a dynamic method to trace individual particle trajectories within fluid mediums, the model ensures precise tracking of the particle's temporal movement. Its broad applicability spans from localized studies, to complex, global-scale environmental systems, underlining its adaptability. This is evident in its varied applications, which include tracking oil spills diffusion [15], mapping floating plastic debris [16], simulating jellyfish migrations [17], analysing smoke dispersion [18], and many other.

The physics based Lagrangian model [1] operates by representing particles within a fluid medium, tracking their position and properties as they move with the fluid's flow. The model calculates the trajectory of each particle by integrating the velocity field of the fluid, which may vary in time and space. This approach enables the simulation of dispersal patterns of particles, such as marine debris by accounting for both advection and diffusion processes. Advection represents the movement of particles by the flow of the fluid [19]. Diffusion, on the other hand, models the dispersion of particles through random motion [19]. This is done by employing techniques such as random walks or Gaussian distributions. This inclusion of randomness enhances the realism of the simulation.

To facilitate these simulations, several Python toolkits like OceanParcels [20], PyGnome [21], and Flexpart [22] have been developed, each specifically designed for simulating the movement of particles using the Lagrangian model framework. These toolkits enable the customization and execution of particle tracking simulations, leveraging data on ocean currents, wind fields, and other environmental phenomena.

OceanParcels [20] is distinguished by several features that make it suitable for our project. One of its notable capabilities are custom kernels. These are user-defined functions that allow for tailored simulation scenarios at each time step. Through custom kernels, users can implement complex behaviours and interactions of particles within the fluid, such as particle reflection or response to environmental variables like temperature and wind. Another significant feature is particle initialization. This feature enables the creation of particles at specific locations, times, and with distinct properties, allowing for more detailed and accurate simulations.

All these attributes render OceanParcels [20] as an optimal choice for this project. By integrating these features, this toolkit facilitates the development of comprehensive simulations. This is crucial for understanding and predicting the movement of marine debris, thereby enhancing our strategies for marine conservation and debris management.

2.1.4 Time Series Modelling

Time series modelling is a technique used to predict future data points by analysing the trends, cycles, and patterns in a series of data points collected over an interval of time [23]. Its main focus is on analysing historical data to uncover the underlying

structure of the dataset, which can then be used to forecast future trends. This method is particularly powerful for its ability to incorporate the sequence and time dependence within the dataset. By examining how values are interconnected over time, time series models can forecast future values based on the inherent temporal dynamics present in the historical data [24]. This form of predictive modelling assumes that past patterns are indicative of future behaviours, making it an indispensable tool in a variety of fields ranging from weather forecasting [25] to stock market predictions [26].

While time series modelling is a powerful tool for forecasting future data, it also encounters several limitations. Time series data often exhibit seasonality and trends, which can complicate the forecasting process [27]. Outliers, missing sequences of data, and anomalies can also significantly impact the accuracy of forecasting models, requiring careful identification, and handling. The capacity of these models to integrate external influential factors and variables is also somewhat limited, often necessitating the integration of additional features for enhanced predictive accuracy [28]. Additionally, time series models require significantly more data for training, which can be a challenge in situations where data is limited [28]. These challenges highlight the importance of adopting a methodical approach to time series modelling, emphasizing the need to carefully consider the specific context and characteristics of the data being analysed when utilizing time series models for effective forecasting.

In the context of this project, we harness time series modelling to predict sea surface current velocities. Accurate predictions require a detailed analyses of the data sequences to discern patterns that could forecast future predictions. The historical hourly data of surface currents form a time series, which is inherently continuous but sampled at discrete intervals. To address this, deep learning models, a subset of artificial neural networks, are employed due to their proficiency in handling vast amounts of sequential data and their capacity to learn complex temporal patterns [29]. Through training on past sea surface current data, these models are equipped to predict future values.

2.1.5 Deep Learning Models

3 Methodology

4 Evaluation

4.1 Writing the Evaluation Chapter

The evaluation component of an FYP is critical.

- *One has to make sure and explain why all tests used to evaluate the system are relevant, using evidence from the literature about similar systems, and justifying any deviations from standard approaches.*

- *Demonstration that system works as intended (or not, as the case may be).*

- *Include comprehensible summaries of the results of all critical tests that have been made.*

- *The student must also critically evaluate the system in the light of these tests results, describing its strengths and weaknesses.*

- *Ideas for improving it can be carried over into the Future Work section.*

- *Comparison of practical with theoretical results and their interpretation.*

- *Comparison with published work when available.*

5 Conclusion

5.1 Writing the Conclusions Chapter

The Conclusions section should be a summary of the project and a restatement of its main results, i.e. what has been learnt and what it has achieved. An effective set of conclusions should not introduce new material. Instead, it should draw out, summarise, combine, and reiterate the main points that have been made in the body of the report and present opinions based on them.

5.2 Writing the Future Work Chapter

Whether by the end of the project all the original aims and objectives have been completed or not, there is always scope for future work. Also, the ideas will have evolved during the project beyond the original target. The Future Work section is for expressing these ideas.

References

References

[1] C. Kehl *et al*, "Efficiently simulating Lagrangian particles in large-scale ocean flows — Data structures and their impact on geophysical applications," *Comput. Geosci.*, vol. 175, pp. 105322, 2023. Available: https://www.sciencedirect.com/science/article/pii/S0098300423000262. DOI: 10.1016/j.cageo.2023.105322.

[2] G. Suaria and S. Aliani, "Floating debris in the Mediterranean Sea," *Mar. Pollut. Bull.*, vol. 86, *(1)*, pp. 494-504, 2014. Available: https://www.sciencedirect.com/science/article/pii/S0025326X14004056. DOI: 10.1016/j.marpolbul.2014.06.025.

[3] M. Compa *et al*, "Risk assessment of plastic pollution on marine diversity in the Mediterranean Sea," *Sci. Total Environ.*, vol. 678, pp. 188-196, 2019. Available: https://www.sciencedirect.com/science/article/pii/S0048969719318984. DOI: 10.1016/j.scitotenv.2019.04.355.

[4] J. Mansui *et al*, "Predicting marine litter accumulation patterns in the Mediterranean basin: Spatio-temporal variability and comparison with empirical data," *Prog. Oceanogr.*, vol. 182, pp. 102268, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0079661120300069. DOI: 10.1016/j.pocean.2020.102268.

[5] P. G. Ryan, "A brief history of marine litter research," in *Marine Anthropogenic Litter*, M. Bergmann, L. Gutow and M. Klages, Eds. 2015, Available: https://doi.org/10.1007/978-3-319-16510-3_1. DOI: 10.1007/978-3-319-16510-3_1.

[6] P. R. Pawar, S. S. Shirgaonkar and R. B. Patil, "Plastic marine debris: Sources, distribution and impacts on coastal and ocean biodiversity," 2016.

[7] S. Katsanevakis, "Chapter 2 - marine debris, a growing problem: Sources, distribution, composition, and impacts," in Anonymous New York: Nova Science Publishers, 2008, pp. 53-100.

[8] D. W. Laist, "Impacts of marine debris: Entanglement of marine life in marine debris including a comprehensive list of species with entanglement and ingestion records," in *Marine Debris: Sources, Impacts, and Solutions*, J. M. Coe and D. B. Rogers, Eds. 1997, Available: https://doi.org/10.1007/978-1-4613-8486-1_10. DOI: 10.1007/978-1-4613-8486-1_10.

[9] C. M. Rochman *et al*, "The ecological impacts of marine debris: unraveling the demonstrated evidence from what is perceived," *Ecology*, vol. 97, *(2)*, pp. 302-312, 2016. . DOI: 10.1890/14-2070.1.

[10] P. Agamuthu *et al*, "Marine debris: A review of impacts and global initiatives," *Waste Manag. Res.*, vol. 37, *(10)*, pp. 987-1002, 2019. . DOI: 10.1177/0734242X19845041.

[11] J. Harlan *et al*, "The Integrated Ocean Observing System High-Frequency Radar Network: Status and Local, Regional, and National Applications," *Marine Technology Society Journal*, vol. 44, *(6)*, pp. 122-132, 2010. Available: https://www.ingentaconnect.com/content/mts/mtsj/2010/00000044/00000006/art00017. DOI: 10.4031/MTSJ.44.6.6.

[12] Anonymous.*"Portus 3.0."* portus.research.um.edu.mt. https://portus.research.um.edu.mt/?p=13.833 (accessed Apr 06, 2024).

[13] Anonymous.*"UNIData | NETCDF."* unidata.ucar.edu. https://www.unidata.ucar.edu/software/netcdf/ (accessed Apr 06, 2024).

[14] E. van Sebille *et al*, "Lagrangian ocean analysis: Fundamentals and practices," *Ocean Modelling*, vol. 121, pp. 49-75, 2018. Available: https://www.sciencedirect.com/science/article/pii/S1463500317301853. DOI: 10.1016/j.ocemod.2017.11.008.

[15] S. A. Lonin, "Lagrangian model for oil spill diffusion at sea," *Spill Science and Technology Bulletin*, vol. 5, *(5-6)*, pp. 331-336, 1999. Available: http://inis.iaea.org/search/search.aspx?orig_q=RN:32026786.

[16] L. C. -. Lebreton, S. D. Greer and J. C. Borrero, "Numerical modelling of floating debris in the world's oceans," *Marine Pollution Bulletin*, vol. 64, *(3)*, pp. 653-661, 2012. Available: https://dx.doi.org/10.1016/j.marpolbul.2011.10.027. DOI: 10.1016/j.marpolbul.2011.10.027.

[17] M. N. Dawson, A. S. Gupta and M. H. England, "Coupled biophysical global ocean model and molecular genetic analyses identify multiple introductions of cryptogenic species," *Proceedings of the National Academy of Sciences*, vol. 102, *(34)*, pp. 11968-11973, 2005. Available: https://doi.org/10.1073/pnas.0503811102. DOI: 10.1073/pnas.0503811102.

[18] D. Hertwig *et al*, "Development and demonstration of a Lagrangian dispersion modeling system for real-time prediction of smoke haze pollution from biomass burning in Southeast Asia," *Journal of Geophysical Research. Atmospheres*, vol. 120, *(24)*, pp. 12605-12630, 2015. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JD023422. DOI: 10.1002/2015JD023422.

[19] R. G. Williams and M. J. Follows, *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms.* 2011Available: https://www.cambridge.org/core/product/31EF28FEF48A172FF746B3E654F9455A. DOI: 10.1017/CBO9780511977817.

[20] Anonymous."*OceanParcels.*" oceanparcels.org. https://oceanparcels.org (accessed Apr 06, 2024).

[21] Anonymous."*PyGNOME.*" gnome.orr.noaa.gov. https://gnome.orr.noaa.gov/doc/pygnome/index.html (accessed Apr 06, 2024).

[22] I. Pisso *et al*, "The Lagrangian particle dispersion model FLEXPART version 10.4," 2019. Available: http://hdl.handle.net/11250/2634384. DOI: 10.5194/gmd-12-4955-2019.

[23] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," vol. abs/1302.6613, 2013. Available: https://api.semanticscholar.org/CorpusID:17070340.

[24] T. Raicharoen, C. Lursinsap and P. Sanguanbhokai, "Application of critical support vector machine to time series prediction," in . DOI: 10.1109/ISCAS.2003.1206419.

[25] S. Raksha *et al*, "Weather forecasting framework for time series data using intelligent learning models," in . DOI: 10.1109/ICEECCOT52851.2021.9707971.

[26] A. Chatterjee, H. Bhowmick and J. Sen, "Stock price prediction using time series, econometric, machine learning, and deep learning models," in . DOI: 10.1109/MysuruCon52639.2021.9641610.

[27] P. Wang *et al*, "Interval time series forecasting: A systematic literature review," vol. 43, *(2)*, pp. 249-285, 2024. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3024. DOI: 10.1002/for.3024.

[28] S. Jadon, J. Milczek and A. Patankar, "Challenges and approaches to time-series forecasting in data center telemetry: A survey," Cornell University Library, arXiv.org, Ithaca, Feb 11,. 2021.

[29] A. Alsharef *et al*, "Time Series Data Modeling Using Advanced Machine Learning and AutoML," vol. 14, *(22)*, 2022. . DOI: 10.3390/su142215292.