Machine Learning & Big Data Analyses for Wildfire & Air Pollution
Incorporating GIS & Google Earth Engine

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Geography

by

Abdullah Al Saim
Jahangirnagar University
Master of Science in Environmental Sciences, 2019
Jahangirnagar University
Bachelor of Science in Environmental Sciences, 2017

July 2021
University of Arkansas

PREVIEW

This thesis is approved for recommendation to the Graduate Council.

Mohamed H. Aly, Ph.D.
Thesis Director

W. Fred Limp, Ph.D.
Committee Member

Jackson D. Cothren, Ph.D.
Committee Member

ABSTRACT

The climatic condition, the vegetation type, and the landscape of the United States have made it susceptible to wildfires. This research is divided into two parts based on the analysis of two different aspects of wildfires of two distinct regions. The first part of the study investigates the wildfire susceptibility in Arkansas. Arkansas is a natural state, and it is heavily dependent on its forest and agricultural resources. During the last 30 years, more than 1,000 wildfires occurred in Arkansas and caused more than 10,000 acres of burned areas. Therefore, identifying wildfire-susceptible areas is crucial for ensuring sustainable forest and agricultural resources. Geographic Information System (GIS)-based Machine Learning (ML) can effectively identify fire-prone areas. In this research portion, Multiple Linear Regression (MLR) and Random Forest (RF) methods are applied to 15 layers of GIS data representing natural and anthropogenic factors that influence wildfires. These 15 variables are selected based on the relationship between fire density and explanatory variables. After identifying all variables, geospatial data are prepared and incorporated in RF for training and predicting wildfire-susceptible areas in Arkansas. The obtained R-squared values from RF are 0.99 for the training regression and 0.92 for the validation. Research outcomes suggest that potential evapotranspiration, soil moisture, Palmer Drought Severity Index, and dry season precipitation are the most contributing factors to wildfires in Arkansas among the 15 considered variables. Outputs also indicate that the Ouachita National Forest and the Ozark Forest have the highest susceptibility to wildfires, the southern part of Arkansas has low-to-moderate fire-susceptibility, and the eastern part of the state has the lowest fire susceptibility. The second part of this research investigates the impact of wildfires on air quality over California, which has been chosen for this analysis because of its extensive history of large and severe wildfires. This portion employs the Google Earth Engine (GEE)

PREVIEW

platform to navigate its geospatial datasets of Moderate Resolution Imaging Spectroradiometer (MODIS) MYD14A1 V6, MCD19A2 Version 6 level 2, and Sentinel-5 Precursor (Sentinel-5P) to validate fire incidents and determine the effect of wildfires on the atmosphere from 2010 to 2020. MODIS MCD19A2 uses an advanced Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm to produce 1-km resolution images and retrieves Aerosol Optical Depth (AOD) at 470 nm and 550 nm wavelengths. These retrieved AOD values from MODIS are validated using the ground-based sun photometers Aerosol Robotic Network (AERONET), and the uncertainty is checked using the Mean Absolute Error (MAE), the Relative Mean Bias (RMB), and the Root Mean Square Error (RMSE). Linear regression shows good correlations between AERONET and MODIS. The correlation coefficient and the adjusted R-squared value vary from 0.78 to 0.80 and from 0.60 to 0.65, respectively, for AOD values at 550 nm and 470 nm. Results from Sentinel-5P indicate that the 2020 fire events in California raised the NO_2 concentration in its atmosphere. This research can improve understanding of the long-term effects of wildfires on air quality and the predictive methodologies that can be used for preemptive measures.

PREVIEW

©2021 by Abdullah Al Saim
All Rights Reserved
PREVIEW

ACKNOWLEDGEMENTS

First and foremost, all praises and thanks be to Almighty, who has given me good health and showered his blessings throughout my degree to complete my research work. His benevolence has made me prosperous in all my academic pursuits.

I would like to express my sincere gratitude to my advisor and thesis director, Dr. Mohamed Aly, for continuously supporting me in my academic journey. I could not thank him enough for accepting me as one of his students and believing in my aptitudes. His patience, encouragement, motivation, and insights have supported me throughout my research and writing of this thesis. At the beginning of the COVID-19 pandemic, when the normal flow of life got disrupted, and everyone was feeling down, he continuously motivated me, guided me to stay on the right track, and supported me to focus on my goals. I would also like to thank him for his compassion, empathy, and great sense of humor when I needed it the most. I have learned a deal of GIS, remote sensing, and etiquette of writing and presenting, which has boosted my confidence and inspired me to strive for excellence in scientific research. Above all, he has taught me to be a good human being with genuine intentions. I will surely carry out these lessons and virtues in my future life.

I would like to extend my thanks to both my committee members, Dr. Fred Limp and Dr. Jackson Cothren, for being a part of this journey. I valued their encouragement, time, insightful comments, and questions. Their well-designed courses have laid the foundation of the methodology of this research.

I am extremely grateful to my family members for their endless love, prayers, and caring. Without their constant support and uncountable sacrifices, I would not be standing here where I am right now. They have put all their faith in my capabilities, which have been my main strength

PREVIEW

along this course. They have always prioritized my interests and prosperity above their emotions. This journey was not smooth, and I had to sacrifice my vacation, family festivals, and personal life to get my degree.

Last but not least, I would like to credit USGS, NASA, the Climatology Lab, and the United States Census Bureau for providing data and supporting this study. This research was partially funded by USGS and ArkansasView grant #GR908148UAF awarded to Dr. Mohamed Aly. I am also grateful to John Wilson for letting me take my office computer home during the pandemic. I am grateful to the Center for Advanced Spatial Technologies (CAST) and the InSAR Lab at the University of Arkansas for the computers and software, which were integral components of this research. I would also like to thank all the InSAR group members for extending their helping hands and being good friends.

PREVIEW

TABLE OF CONTENTS

PREVIEW

PREVIEW

PREVIEW

LIST OF FIGURES

PREVIEW

PREVIEW

LIST OF TABLES

PREVIEW

CHAPTER 1

INTRODUCTION

The United States is known for its major hotspots of wildfires because of its vegetation structure, climatic condition, and urban development. Every year, millions of dollars of property damage and destruction are caused by wildfires in the United States. With increasing population, development, and climate change, it is crucial to identify wildfire-susceptible areas for states that can be affected in the future. Arkansas is known as the "Natural State" and is blessed with many natural resources. The economy of the state depends mainly on its forest and agricultural resources. Even though Arkansas is not recognized as a substantial wildfire-prone area, it has the potential to become susceptible with the potential to cause heavy damage to Arkansas' forest and agricultural resources. There are limited studies on wildfires in Arkansas, presenting a bigger concern. As the climate is changing, a significant increase in the number of wildfires and burned areas is expected. It has been projected that about 200% to 300% increase in the number of "very large wildfires weeks" in Arkansas are expected by mid-century (2041-2070) compared to the recent past (1971-2000) (Figure 1.1). The state is not equipped with the essential resources to control large-scale fire events. Furthermore, private forest owners take minimal necessary control measures to prevent fire events. Proper land management is vital to prevent large-scale fire events, and for this, identification of wildfire susceptible areas is crucial.

PREVIEW

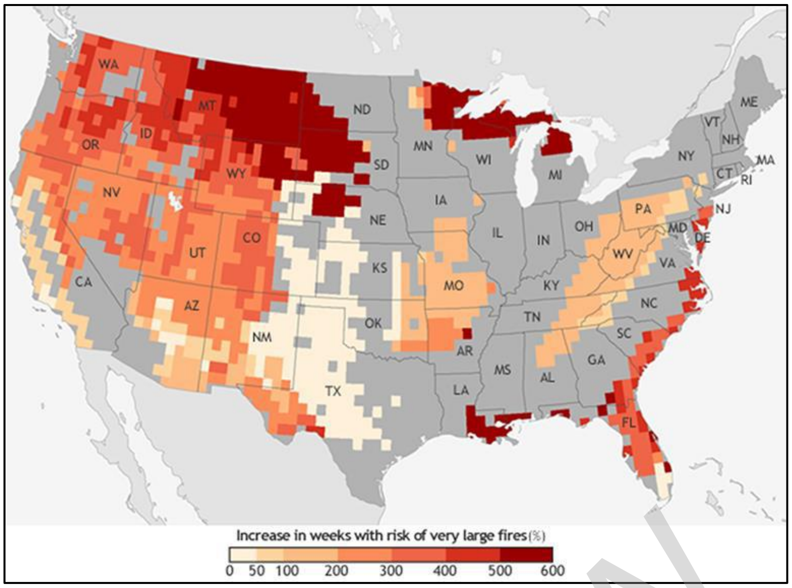

Figure 1.1: Projected increase of “very large fire week” in the United States (NOAA Climate.gov map based on data from Barbero et al., 2015).

Another effect of fires is the air pollution. Wildfires emit a large volume of different types of gases like CO_2, CO, CH_4, NO_2, volatile organic compounds, and fine particulate particles of different diameters between and less than 2.5 and 10 micrometers. These pollutants negatively alter air quality and affect human health. Additionally, these materials are uplifted into the higher atmosphere through heat and transported to other regions affecting weather and the climate system. For instance, smoke from the 2020 California wildfires crossed the Atlantic Ocean and traveled to Europe. As the numbers and intensity of the wildfires are increasing, air pollution caused by them is also becoming a major concern.

Geospatial analysis of natural hazards is a very popular and effective way to analyze the multidimensional effects of hazards like wildfires. Higher computational power, availability of data, and the introduction of new advanced methods have made this analysis more appealing and practical. Nowadays, numerous satellites are equipped with different types of powerful and high-level sensors, which can capture more accurate measurements at higher resolutions. At the same time, Geographic Information System (GIS) software has improved significantly. The

PREVIEW

introduction of cloud computing and Machine Learning (ML) algorithms into the geospatial software and application has added a new dimension to the field of GIS. This research explores these new areas of GIS, such as ML and cloud computing, to analyze the recently growing concern of wildfires in Arkansas and California.

The wildfire-susceptibility research on Arkansas will surely contribute to state organizations like the Arkansas Department of Emergency Management (ADEM), the Arkansas Department of Agriculture-Arkansas Forestry Commission, the Governor's Office, and the academic intuitions interested in this kind of research. The air quality research on California will provide additional insights to the California Department of Forestry and Fire Protection (CALFIRE), the California Air Resources Board (CARB), and the California Department of Public Health (CDPH). Both studies will add value to federal agencies like the Federal Emergency Management Agency (FEMA), the United States Geological Survey (USGS), the Natural Resources Commission Service (NRCS), and the United States Department of Agriculture (USDA).

WILDFIRES IN ARKANSAS

Arkansas has a subtropical climate, which is influenced by its topography and its proximity to the western plains and the Gulf of Mexico to the south (Kottek et al., 2006). The main characteristics of Arkansas' weather are hot humid summers and drier winters. Arkansas has both mountain and flat plains. The south and the eastern parts of the state are mainly flat, and the northwestern part is mountainous. The natural division of Arkansas is the Ozark Plateau, which lies in the northwestern and north-central part of Arkansas and has rugged hills and deep valleys. This part of the state is also the coolest and the driest, with a mean January and July temperatures of 35°F and 79°F, respectively. The Ouachita Mountains range from east to west

PREVIEW

and receive the most rainfall, with an average of 58 inches per year. The Arkansas River Valley is lower than the Ozark Plateau and the Ouachita Mountains. Its average annual precipitation is 51 inches, and the mean January and July temperatures are 40°F and 82°F, respectively. The Mississippi Alluvial Plain extends in the eastern part of the state, with an average annual precipitation of 46 inches. The temperature in this plain ranges from 36°F to 82°F. The West Gulf Coastal Plain lies in the southeastern and southern parts of Arkansas and has the highest average temperature of the state, ranging from 44°F to 82°F (Climate and Weather - Encyclopedia of Arkansas, 2019).

The climatic conditions of the state favor the growth of forest lands, blessing the state with natural forest resources. About 19 million acres of forest area covers Arkansas, which covers 56% of the state. In the 1950s and 1960s, Arkansas lost 20% of its forest land, but from 1978, the forest land has grown more than 1 million acres. The forest area of the state is mostly in the southwest, Ozarks, and Ouachita regions. These three regions contain almost 88% of the total forest lands. The majority portion of the forest land is owned by private ownership, with the trees consisting of mostly hardwood timber types with a majority of Pine and Oak trees. Arkansas contains the largest national forest, located in the south within the Ozark-St. Francis and Ouachita National Forests with an area of 2.5 million acres. This forest resource is affected by different types of disturbances, events that affect or kill at least 25% of the trees in an area. Weather and fire are the major disturbances in Arkansas, and studies have shown 60% of the fire disturbance has occurred in the pine and pine/oak-dominated forest areas (Forestry, n.d.).

Besides forestry, Arkansas' economy is heavily dependent on agricultural productions. Agricultural products contribute around $16 billion annually to the state's economy. In fact, Arkansas ranked number one in rice production in the United States; furthermore, the topography

PREVIEW

and the climate of this region also favor various agricultural productions like soybeans, cotton, poultry, and feed grains. Almost 95% of the total land resources are utilized by agriculture and forestry. In Arkansas, 14.5 million acres are farmland, 6.2 million acres are croplands, and 8.3 million are used as livestock and hays. These facts point out the significance of proper land management and policies. Natural hazards like wildfires can heavily affect states like Arkansas, which are heavily dependent on forestry and agriculture (Ag Facts, n.d.).

In 2012, 2,148 wildfires were recorded and burned 34,423 acres of land. Approximately 1,566 wildfires were recorded in 2017, burning more than 27,549 acres of land. It was found that the 2017 wildfires were larger when comparing the size of wildfires of that year with the 10-year average. The 10-year average size of wildfires is 15.5 acres, while the average of 2017 wildfires in Arkansas was 17.6 acres. According to the National Interagency Coordination Center 2018 wildfire outlook, Northwest Arkansas was predicted to have higher wildfire activities than normal based on projected dry air and drought conditions. The main reasons for Arkansas wildfires are arson and unintentional outdoor burning of grass, leaves, and trash. Natural events like lightning strikes also cause wildfires, but anthropogenic reasons like campfires, cigarettes, and other fire-causing equipment are the primary reason for fires in Arkansas (Harrison & Us, 2018).

In this study, a GIS-based Multiple Regression and ML technique has been utilized to identify the wildfire susceptible areas in Arkansas. Numerous layers of GIS data were used to analyze previous wildfire events. ML techniques, such as Random Forest (RF), give higher flexibility and predictive power over common statistical measurements like Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR). Successful prediction from ML depends on training of the model. The better the model has trained, the more accurately it can

PREVIEW

predict a feature. In this study, Oklahoma has been chosen as a training study site due to its history of large and widespread wildfires in the past, making it ideal for analyzing the relationship between wildfires and contributing factors. Additionally, Oklahoma has a similar climatic condition to Arkansas. After successfully finding the proper variables to explain wildfires, RF was applied to identify wildfire-susceptible areas in Arkansas.

IMPACT OF CALIFORNIA WILDFIRES ON AIR QUALITY

Due to long, dry summers and high temperatures, wildfires have become a common natural event in California. Every year, California is experiencing wildfires due to natural and anthropogenic reasons. Approximately 2,434 to 3,672 wildfires had been recorded between 2007 and 2016, which burned 25,438 to 434,667 acres of land. Moreover, due to population increase, more people are living near the wildfire danger areas; this can claim lives, destroy houses and properties, and influence the regional climate. Wildfires not only destroy infrastructure but also change the air quality. Biomass burning of wildfires causes the release of different materials like CO, NO_2, SO_2, CO_2, and particulate matter (PM) like PM2.5. These increases in air pollutants can cause cardiorespiratory diseases and other human health effects. In fact, wildfires can increase air pollutants in an area abruptly within a short time. In Indonesia, it has been found that large and long-term fire affects areas with 1,000 micrograms per cubic meter of PM10 for several days and 150 micrograms per cubic meter for long periods. In Europe, it has been reported that biomass smoke is responsible for at least 40,000 premature deaths per year. Between 2008 and 2012, hospital admission due to respiratory issues were approximately 5,200 to 8,500 annually. The estimated economic cost of short-term premature deaths and hospital admission is $11 to $20 billion dollars per year; and $76 to $130 billion dollars per year for long-term, exhibiting the economic damage of air pollutants wildfires (Gupta et al., 2018).

PREVIEW

To accurately measure the impact of wildfires on air quality, it essential to have persistent and stable measuring systems. Ground-based air quality measuring instruments have low spatial resolution and high temporal coverage, making it difficult to measure the effect of air pollution due to wildfires, as wildfire smoke can affect a broad region for a short period, and ground-based instruments can only capture a fraction of the effects. Advancements in the satellite observation systems have allowed us to characterize the vertical column of the atmosphere with higher spatial and temporal resolutions. Satellite images can capture the widespread effect of air pollution during a wildfire event. Air quality modeling with satellite technology has added new opportunities in the different transport and pollution source research. The interest in air quality monitoring and modeling has been growing in the research community due to the availability of accurate satellite measurements, and these, combined with ground measurements, can provide an unbiased understanding of the effect of air pollution during a wildfire event (Fernandes et al., 2019).

Aerosol Optical Depth (AOD) characterizes the atmospheric aerosol and has important influences on the climatic processes; it has a complex interaction with cloud particles and can change the radiative forcing of the atmosphere. The spatial and temporal variations in AOD in the atmosphere have become a growing concern in the scientific community for accurate climate modeling and climate change research (Kassianov et al., 2021). Moderate Resolution Imaging Spectroradiometer (MODIS) provides global measurements of AOD at a high spatial resolution, making it ideal for analyzing AOD trends over a large area for a long period. MODIS retrieves the AOD measurements at 470 and 550 nm wavelengths. Along with these measurements, MODIS also provides thermal anomaly data, used to detect active fire and monitor fire events. Furthermore, Sentinel-5P has the capabilities to measure other atmospheric components like

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.

PREVIEW