

7-2021

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

Abdullah Al Saim

*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>

 Part of the Geographic Information Sciences Commons, Nature and Society Relations Commons, Physical and Environmental Geography Commons, and the Remote Sensing Commons

---

### Citation

Saim, A. (2021). Machine Learning & Big Data Analyses for Wildfire & Air Pollution Incorporating GIS & Google Earth Engine. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/4226>

---

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu), [uarepos@uark.edu](mailto:uarepos@uark.edu).

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

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)

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<sub>2</sub> 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.

©2021 by Abdullah Al Saim  
All Rights Reserved

## **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

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.

## TABLE OF CONTENTS

<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
<b>WILDFIRES IN ARKANSAS .....</b>	<b>3</b>
<b>IMPACT OF CALIFORNIA WILDFIRES ON AIR QUALITY .....</b>	<b>6</b>
<b>REFERENCES.....</b>	<b>8</b>
<b>CHAPTER 2: MACHINE LEARNING FOR MODELING WILDFIRE</b>	
<b>SUSCEPTIBILITY IN ARKANSAS .....</b>	<b>10</b>
<b>ABSTRACT.....</b>	<b>10</b>
<b>INTRODUCTION.....</b>	<b>11</b>
<b>STUDY AREA.....</b>	<b>15</b>
<b>DATA ACQUISITION AND PROCESSING .....</b>	<b>17</b>
<b>Multiple Linear Regression.....</b>	<b>20</b>
<b>Random Forest Classification.....</b>	<b>23</b>
<b>RESULTS AND DISCUSSIONS .....</b>	<b>24</b>
<b>OLS and GWR Outputs .....</b>	<b>24</b>
<b>RF Outputs .....</b>	<b>29</b>
<b>CONCLUSIONS .....</b>	<b>34</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>36</b>
<b>REFERENCES.....</b>	<b>36</b>

<b>CHAPTER 3: BIG DATA ANALYSES FOR DETERMINING THE SPATIO- TEMPORAL TRENDS OF AIR POLLUTION DUE TO WILDFIRES IN CALIFORNIA USING GOOGLE EARTH ENGINE .....</b>	<b>42</b>
<b>ABSTRACT.....</b>	<b>42</b>
<b>INTRODUCTION.....</b>	<b>43</b>
<b>STUDY AREA.....</b>	<b>47</b>
<b>DATASETS AND METHODS .....</b>	<b>49</b>
<b>AERONET .....</b>	<b>49</b>
<b>MODIS MAIAC .....</b>	<b>49</b>
<b>MODIS DT and DB .....</b>	<b>51</b>
<b>MODIS Active Fire Product .....</b>	<b>52</b>
<b>Interpolation of AOD from AERONET .....</b>	<b>52</b>
<b>Sentinel-5 Precursor .....</b>	<b>53</b>
<b>Big Data Analyses .....</b>	<b>54</b>
<b>RESULTS AND DISCUSSIONS .....</b>	<b>55</b>
<b>Seasonal Trends of AOD .....</b>	<b>57</b>
<b>MODIS Thermal Anomaly.....</b>	<b>60</b>
<b>Aerosol Index and NO<sub>2</sub> Concentration from Sentinel-5P .....</b>	<b>68</b>
<b>CONCLUSIONS .....</b>	<b>69</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>70</b>

<b>REFERENCES.....</b>	<b>70</b>
<b>OVERALL CONCLUSIONS .....</b>	<b>75</b>

## LIST OF FIGURES

Figure 1.1: Projected increase of “very large fire week” in the United States .....	2
Figure 2.1: Numbers of documented wildfires and burned areas due to wildfires in Arkansas from 1981 to 2018 .....	13
Figure 2.2: Physiographic provinces of Arkansas .....	16
Figure 2.3: Selected explanatory variables .....	25
Figure 2.4: Scatter plots of the fire density versus the selected 15 variables .....	26
Figure 2.5: Identified variables for predicting fire susceptibility in Arkansas .....	30
Figure 2.6: Documented fire incidences plotted on top of the modeled fire likelihood .....	35
Figure 3.1: Geomorphic provinces of California and location of the 21 AERONET stations used in this study .....	48
Figure 3.2: Scatter plots of MODIS and AERONET at 470 nm and 550 nm wavelengths for all seasons and the fire season .....	56
Figure 3.3: Monthly trends of AOD derived from MODIS and AERONET at 470 nm and 550 nm wavelengths .....	59
Figure 3.4: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2012 and 2013 retrieved at 470 nm and 550 nm wavelengths .....	61
Figure 3.5: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2014 and 2015 retrieved at 470 nm and 550 nm wavelengths .....	63
Figure 3.6: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2016 and 2017 retrieved at 470 nm and 550 nm wavelengths .....	65

Figure 3.7: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2018 and 2020 retrieved at 470 nm and 550 nm wavelengths .....	67
Figure 3.8: MODIS thermal anomaly and Sentinel-5P Aerosol Absorbing Index (AAI) for 2020 wildfires .....	68
Figure 3.9: MODIS thermal anomaly and Sentinel-5P NO <sub>2</sub> concentration for 2020 wildfires ....	69

## **LIST OF TABLES**

Table 2.1: Ordinary Least Square Results .....	27
Table 2.2: Performace results of OLS and GWR .....	28
Table 2.3: Variables used in the analysis and their importance and percentage.....	32

## **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.

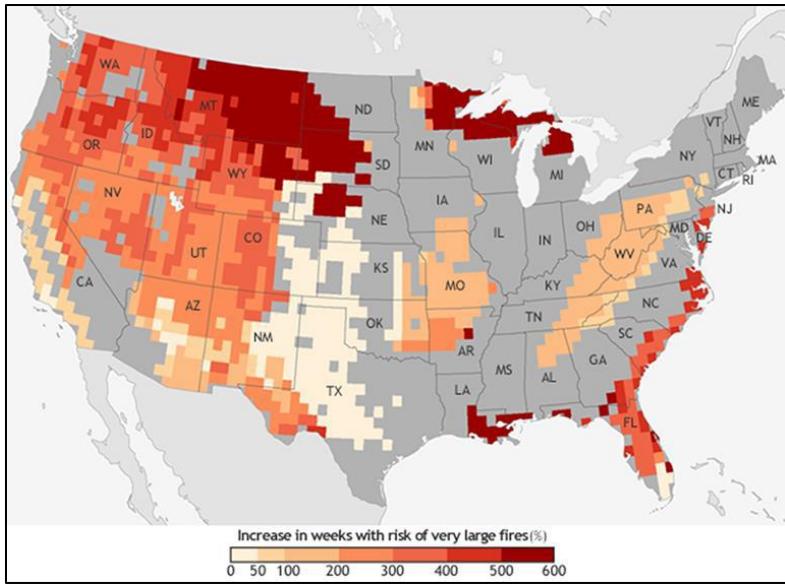


Figure 1.1: Projected increase of “very large fire week” in the United States (NOAA Climate.gov map based on data from Barbero et al., 2015).

Another effect of fires is the air pollution. Wildfires emit a large volume of different types of gases like CO<sub>2</sub>, CO, CH<sub>4</sub>, NO<sub>2</sub>, 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

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

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

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

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<sub>2</sub>, SO<sub>2</sub>, CO<sub>2</sub>, 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).

To accurately measure the impact of wildfires on air quality, it is 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

methane ( $\text{CH}_4$ ), carbon dioxide ( $\text{CO}_2$ ), sulfur dioxide ( $\text{SO}_2$ ), and Nitrogen dioxide ( $\text{NO}_2$ ) concentrations. The combination of these satellites can provide a holistic idea about the impact of wildfires on air quality over a broad region. Moreover, cloud computing like Google Earth Engine (GEE) has provided the opportunity to analyze petabytes of data without downloading them on a local machine, allowing researchers to perform various investigations with Big Data acquired by numerous satellites.

This part of the research uses the GEE platform to analyze the daily MODIS thermal anomaly data and the AOD measurements from 2010 to 2020 in order to investigate the temporal and spatial changes in air quality over California during large fire events. AOD measurements at 470 and 550 nm wavelengths were cross-validated with the available ground-based Aerosol Robotic Network (AERONET) measurements at 470 and 550 nm wavelengths. Sentinel-5P data have been employed to determine the changes in other major air pollutants during 2020 wildfires. This research highlights the effect of fire on air quality during wildfires, as well as assists in visualizing the spatial distribution of air pollution in the source region and its surroundings.

## REFERENCES

- Ag Facts. (n.d.). Retrieved June 12, 2021, from <https://www.arfb.com/pages/education/ag-facts/>
- Barbero, R., Abatzoglou, J. T., Larkin, N. K., Kolden, C. A., & Stocks, B. (2015). Climate change presents increased potential for very large fires in the contiguous United States. *International Journal of Wildland Fire*, 24(7), 892–899.
- Climate and Weather—Encyclopedia of Arkansas*. (2019, November 18). <https://encyclopediaofarkansas.net/entries/climate-and-weather-4579/>
- Fernandes, A., Riffler, M., Ferreira, J., Wunderle, S., Borrego, C., & Tchepel, O. (2019). Spatial analysis of aerosol optical depth obtained by air quality modelling and SEVIRI satellite observations over Portugal. *Atmospheric Pollution Research*, 10(1), 234–243.

*Forestry*. (n.d.). Retrieved June 12, 2021, from <https://www.arfb.com/pages/arkansas-agriculture/commodity-corner/forestry/>

Gupta, P., Doraiswamy, P., Levy, R., Pikelnaya, O., Maibach, J., Feenstra, B., Polidori, A., Kiros, F., & Mills, K. (2018). Impact of California fires on local and regional air quality: The role of a low-cost sensor network and satellite observations. *GeoHealth*, 2(6), 172–181.

Harrison, M. A. 402 N. W. S. S. 136, & Us, A. 72601 P.-2502 C. (2018, January 14). *Arkansas Forestry Commission Shares Wildfire Outlook for 2018—Buffalo National River (U.S. National Park Service)*. <https://www.nps.gov/buff/learn/news/arkansas-forestry-commission-shares-wildfire-outlook-for-2018.htm>

Kassianov, E., Cromwell, E., Monroe, J., Riihimaki, L. D., Flynn, C., Barnard, J., Michalsky, J. J., Hodges, G., Shi, Y., & Comstock, J. M. (2021). Harmonized and high-quality datasets of aerosol optical depth at a US continental site, 1997–2018. *Scientific Data*, 8(1), 1–10.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). *World map of the Köppen-Geiger climate classification updated*.

## **CHAPTER 2**

### **MACHINE LEARNING FOR MODELING WILDFIRE SUSCEPTIBILITY IN ARKANSAS**

#### **ABSTRACT**

Wildfire is a common natural hazard that is influenced by both natural and anthropogenic factors. Contemporary changes in the landscape and the climatic conditions have resulted in a dramatic increase in the fire frequency and intensity in the United States. Arkansas is a natural state with 56% of its area covered by forest. Each year between 1981 and 2018, about 1,000 wildfires occurred and burned more than 10,000 acres in Arkansas. Machine Learning (ML) techniques have been effectively used in identifying fire-susceptible areas. This study employs Multiple Linear Regression (MLR) and Random Forest (RF) methods to address the common natural and anthropogenic factors that influence wildfires and ultimately to model fire-susceptibility in Arkansas. To investigate the relationship between the explanatory variables and the fire density, MLR has been applied to the 15 variables known to contribute to Oklahoma's fire density, and then the identified significant variables were incorporated into RF to train the model for predicting wildfire-prone areas in Arkansas. Oklahoma's severe wildfires have occurred under similar climatic conditions to Arkansas, and RF has a higher predictive ability compared to MLR, thus, they were used for training the model. Among the 15 explored variables, potential evapotranspiration, soil moisture, Palmer Drought Severity Index, and dry season precipitation are found to be the most significant factors contributing to the fire density. The obtained R-squared values from RF are significant, with 0.99 for the training regression and 0.92 for the validation. Results show that the Ouachita National Forest and the Ozark Forest, in

west-central and west Arkansas, respectively, have the highest susceptibility to wildfires. The southern part of Arkansas has low-to-moderate fire susceptibility, while the eastern part of the state has the lowest fire susceptibility. These outcomes will ultimately support Arkansas' fire preparedness plan to reduce its economic loss and save lives.

**KEYWORDS:** Fire susceptibility mapping, Machine Learning, Multiple Linear Regression, Random Forest, Arkansas.

## INTRODUCTION

Forests are one of the influential natural resources for sustainable development and achieving environmental and ecological balance. Since the dawn of civilization, humans have been obtaining tangible and intangible benefits from the forest resource. However, this resource is depleting gradually due to natural and anthropogenic reasons (Valdez et al., 2017). Forest fire is one of the prominent causes of forest depletion worldwide, and due to recent climate change and human intervention, the number of wildfires, severity, and the extent of forest fires have increased drastically. Wildfires are uncontrollable and unwanted burning of combustible vegetation in an area that causes large environmental and ecological destruction such as the release of large amounts of CO<sub>2</sub> in the atmosphere, unexpected changes in the landscape, extensive soil erosion, and alteration of soil properties (Akther & Hassan, 2011; Adab et al., 2013). Furthermore, properties and infrastructures may also be destroyed by wildfires, adding to the economic loss of a country.

The wildfire dynamic is complex, and recent analyses of fire occurrences have shown that the likelihood of wildfire in the western US is becoming more extreme than before. Over the years, changes in the landscape and climatic conditions have increased the average fire frequency

and intensity. Even areas that were originally less prone to wildfires are now experiencing constant and severe fire events due to human influence, topography, and change in weather conditions (Parisien et al., 2012). The new fire-prone areas in the US have exceeded almost tens of millions of acres compared to previous records. The budget for fire response and suppression, along with the estimated property and structure loss to fire, is significantly increasing (Steelman, 2016). Fire-susceptibility modelling has become an important component in fire-control planning, management, and policymaking. Many models try to estimate the fire occurrence and ignition source by analyzing long term data from human-induced factors or by examining various topological and climatic factors. Fire Science has become an essential field of research, and these investigations have helped to identify the possible factors that may influence wildfires (Rodrigues et al., 2018). A comprehensive perception of the driving factors of wildfires can be achieved by combining both the natural and the human-induced factors.

Sustainable forest resource management requires a complete understanding of the wildfire's drivers, such as the ignition points, climatic conditions, and human influence. Factors contributing to fires in one area may not exist or have minimal impact on fires in another area. Due to the complexity of fire events, assessments of the long-term fire-prone regime is needed to get a holistic idea about the drivers of fires, helping the evaluation regarding the susceptibility over areas (Oliveira et al., 2012). Remote sensing-based analysis has recently become popular in investigating fire occurrences due to the vast availability of data and high computing capabilities. The advancement in satellite imagery has made it possible to monitor vegetation change and observe various climatic factors over time, which is crucial for fire forecasting and modelling (Chowdhury & Hassan, 2013).

As a natural state, Arkansas is blessed with enriched biodiversity and natural resources.

About 56% of the state is covered by forest, approximately 11.8 billion trees (Nowak & Greenfield, 2018). Every year, nearly more than 1,000 wildfires have occurred between 1981 and 2018, and in some years, this number has exceeded 4,000 events (Figure 2.1-top). The severity of these wildfires varies from year to year, but on a yearly average more than 10,000 acres have burned, and in some years, this number has surpassed 90,000 acres (Figure 2.1-bottom). The topography, the climate, and the vegetation cover in Arkansas have made it susceptible to wildfires. Although the wildfire frequency in the state is rising, its fire-susceptibility research is still limited.

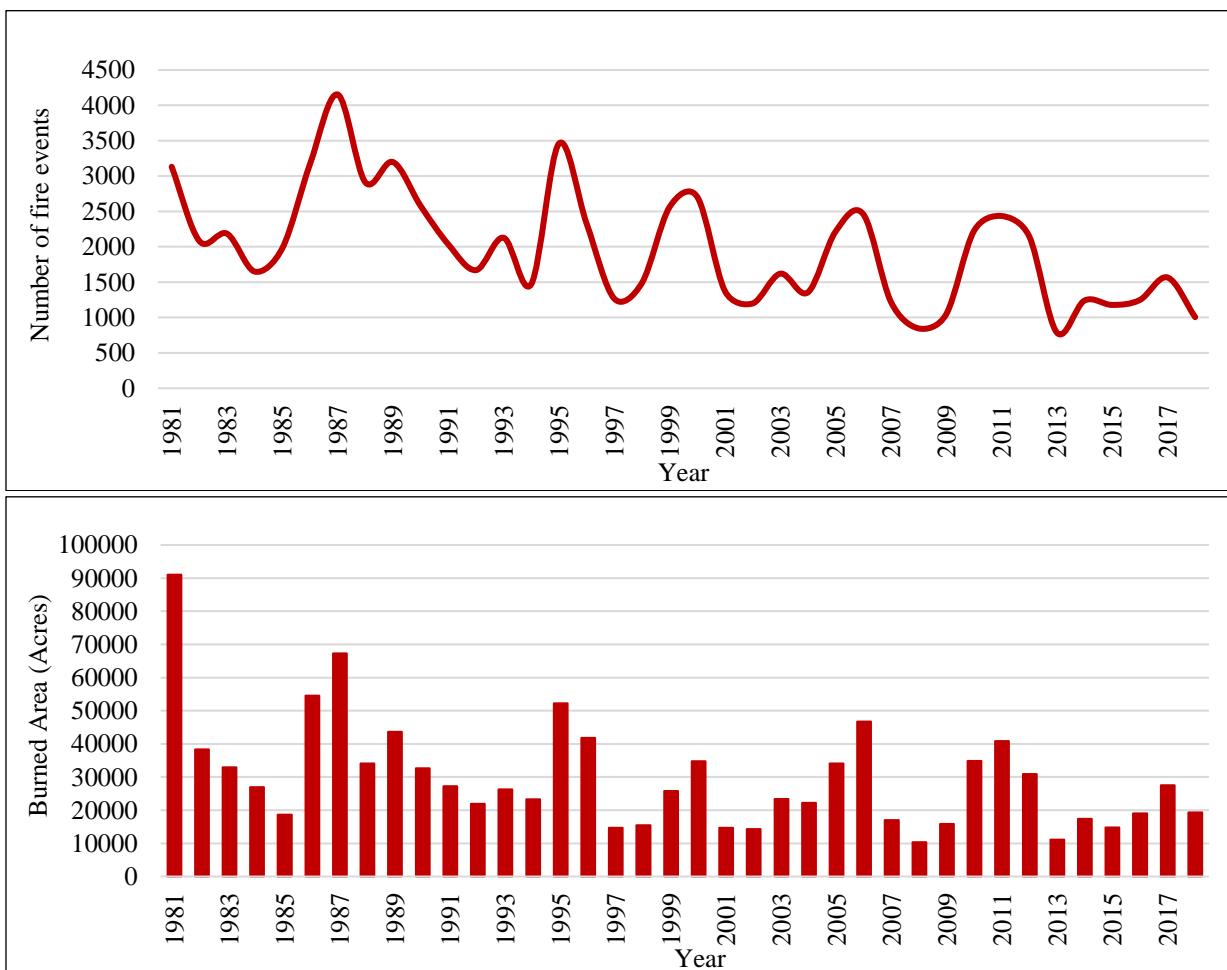


Figure 2.1: Numbers of documented wildfires (top) and burned areas due to wildfires (bottom) in Arkansas from 1981 to 2018 (Wildfire Statistics, n.d.).

Various techniques have been developed to investigate wildfires, and new methods are being invented to fill up the previous work limitations. One of the popular approaches to analyze wildfires is the use of statistical methods to map the fire ignition probability over variant spatial and temporal scales. Examples of statistical methods include kernel density interpolation, which calculates the probability of density of functions of random variables; weights-of-evidence modelling, a quantitative method to provide evidence to the predictive power of an independent variable to a dependent variable; and Bayesian belief network analysis, representing the conditional dependence among variables (Chuvieco et al., 2004; Dickson et al., 2006; Dlamini, 2010). Recently, ML algorithms have become very popular due to their significant development. Common ML methods includes Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) (Chang et al., 2020). Nowadays, ML techniques are being broadly applied in natural hazard studies, such as wildfires, landslides, drought and flood monitoring, and earthquake occurrence (Wang et al., 2015; Rouet-Leduc et al., 2017, 2017; Taalab et al., 2018; Park & Kim, 2019; Jaafari & Pourghasemi, 2019). Geospatial analysis, coupled with statistical models and ML, can provide more accurate measurements as they can capture more spatial variabilities and they are proficient at handling spatial autocorrelations.

In this study, several variables represented by GIS and remotely sensed data related to wildfires were tested and applied to produce a fire-susceptibility map for Arkansas. These variables represent various aspects of topology, climate, and human activities, and they are selected based on literature reviews (Syphard et al., 2008; Aldersley et al., 2011; Parisien et al., 2012; Yang et al., 2015; Rodrigues et al., 2018). All GIS variables for the state were standardized to a 10x10-km grid. This allowed a systemic approach to analyze the statistical

relationship between the determined variables and fire events. Two different regression methods were then applied: Linear Regression Model (LRM) and ML-based RF. Initially, all selected variables were run through Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) over Oklahoma. Due to its extensive wildfire experience by both natural and anthropogenic causes making it an ideal study for factors of fire behaviour, Oklahoma has been selected for analyzing the relationship between the dependent and the explanatory variables, using OLS regression and GWR (Carlson et al., 2002; Reid et al., 2010; Weir et al., 2012; Gorte & Economics, 2013; Balch et al., 2018). This process identified the statistically significant variables that were then applied to RF to predict wildfire-susceptibility in Arkansas because RF has proved to be a better approach than MLR for predicting wildfires (Arpacı et al., 2014; Rodrigues & de la Riva, 2014). The specific goals of this study are: (1) identify the relevant driving factors for wildfire by analyzing the fire density in Oklahoma and (2) use those explanatory variables to investigate fire susceptibility in Arkansas. This study can help the forest department and the state government to implement state-wide precaution measures and reduce the damage and impact of wildfires across the state.

## **STUDY AREA**

Arkansas is situated in the south-central part of the United States. It is mainly covered by forests and lakes and has major mountain valleys (Figure 2.2). The enriched forest covers more than half of the state. Oak covers about 42%, and shortleaf pine occupies 29% of the forest type. About 58% of the timberland is owned by private landowners such as farmers and ranchers (Hodgdon & Tyrrell, 2003). Nearly 13% of the state's forested acreage is considered for national forest status, and Forest Resource companies lease or own about 23% of the state's timberland (Clutter et al., 2005). The state's Forest Resource provided job opportunities to about 64,789

people in 2011, amounting to \$3.6 billion of labor income. Additionally, 75% of the energy needed for the forest industries is generated from wood waste (Pelkki, 2005). This highlights the value and importance of the forest resource in Arkansas.

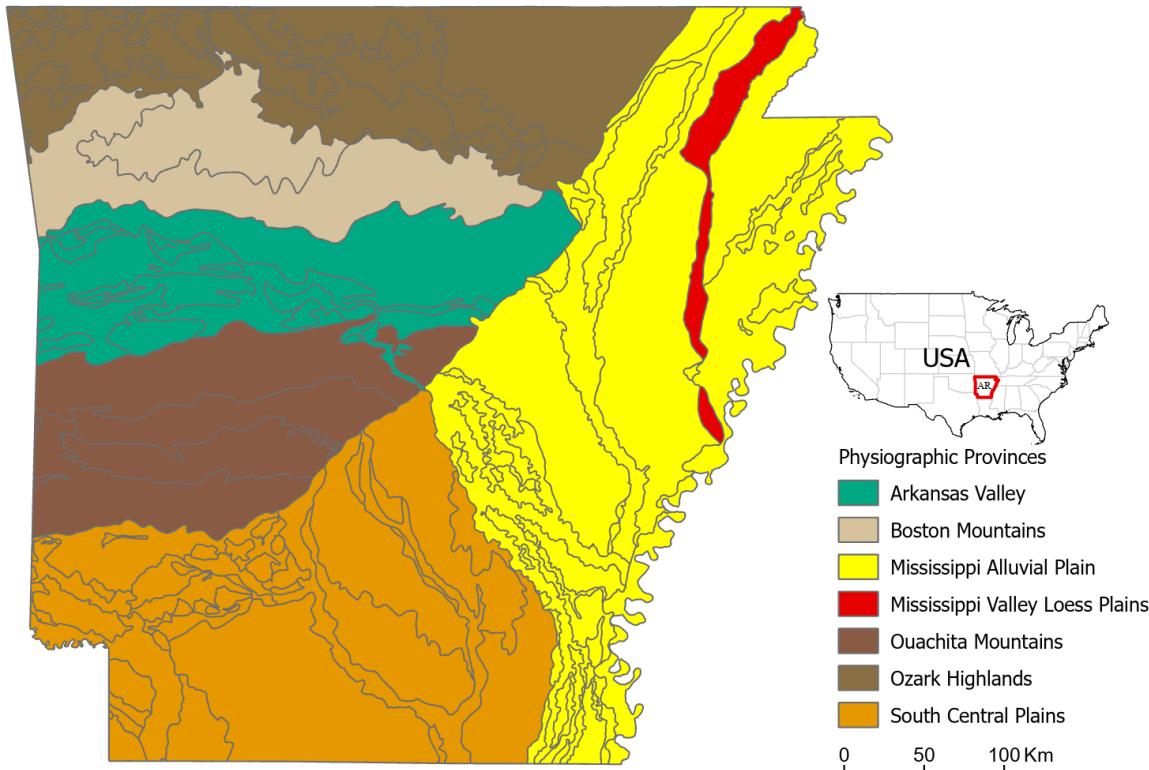


Figure 2.2: Physiographic provinces of Arkansas.

In terms of population, the three main population centers are Northwest Arkansas, the Little Rock metropolitan area, and Northeastern Arkansas. Northwest Arkansas, consisting of Fayetteville, Springdale, Bentonville, and Rogers, has more than 525,000 people living in its region, making it the second most populated area in Arkansas (He et al., 2016). The Little Rock metropolitan area is located at the center of the state, and it includes Jacksonville, Little Rock, Benton, Cabot, Conway, and Maumelle, and approximately 739,000 people live in this area.

Northeastern Arkansas, including Jonesboro, is the third main population hub of the state with a population over 100,000 (Rowden & Aly, 2018).

## **DATA ACQUISITION AND PROCESSING**

The study approach consisted of three major steps: (1) data acquisition and processing for training and prediction, (2) exploring the relationship between fire and selected variables, and (3) geospatial modelling to produce a fire susceptibility map for Arkansas. Four different types of remotely sensed data were acquired for this study including vegetation, climate, topography data, and human intervention factors. The fire incident data have been obtained from the Monitoring Trends in Burn Severity (MTBS) website. MTBS has been collecting fire incident data from 1984 to present, and it is monitored by U.S. Geological Survey-Center for Earth Resources Observation and Science (EROS) and the U.S. Department of Agriculture-Forest Service Geospatial Technology and Applications Center (GTAC). Using this dataset, fire density has been calculated for Oklahoma using a kernel density tool. The spatial distribution of fire incidence is highly clustered, and for this reason, density has been estimated by the K<sup>th</sup> nearest neighbour.

Vegetation distribution and composition are affected by topographic factors, which themselves are indirectly related to the flammability of the forest as well as influences for the climate of the region. Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) with a spatial resolution of 1 arc-second data was used to derive the elevations for the training and the predicting processes. From the elevation data, the slope of the study area was extracted. Later, it was concluded that human-induced fire tends to occur in the gentler sloped area.

Vegetation data were acquired from the National Land Cover Database (NLCD), a database that provides national data on landcover at a 30-m resolution. In this study, the 2016 data were used to infer the vegetation condition of the study area. Considering the context of this study, the landcover map was used as a representative of burnable vegetation. NLCD data divided the landcover into 16 classes, but for the sake of this analysis, the data were reclassified into 7 classes only to satisfy the training and predicting purposes. Furthermore, to precisely measure the impact of various vegetation types, data from the Landfire Program were used to divide the forest class into herb, shrub, tree, and wildland urban interface.

Climatic conditions are also very important for modelling fire susceptibility. Conditions like precipitation, temperature, lighting, and soil moisture have significant impacts on fire ignition and propagation, so they were considered in this analysis. Daymet Climate data from 1990 to 2015 were used to calculate the average dry season precipitation and annual average temperature (e.g., Thornton et al., 2018). The average annual soil moisture from 1990 to 2015 were estimated from the TerraClimate dataset, with a resolution of 4 km. A thorough examination of the model performance revealed that annual measures were more suitable than seasonal measures in some cases; however, the opposite was true in some others. Precipitation, directly and indirectly, influences fire density, and in this study, the influence of precipitation was more noticeable in the dry season (June - October). For temperature and soil moisture, the model showed higher sensitivity for annual average than seasonal variations.

In addition, climate water deficit (DEF), potential evapotranspiration (PET), and Palmer Drought Severity Index (PDSI) were used in this study to better understand wildfire drivers. DEF is the difference between PET and actual evapotranspiration, as defined by Stephenson (1998), and helps to estimate drought stress on soils and plants, as well as influences the vegetation

cover. PET is the amount of water evaporated by transpiration and evaporation from an area, given a uniform vegetation cover and being well supplied with water. PDSI is designed to calculate agricultural drought by adding precipitation to the top two layers of the soil and using a temperature driven evapotranspiration algorithm to remove moisture. All these metrics are commonly used in fire research (Littell et.al., 2009; Miller et al., 2012), and thus they were taken into account in this study. Geospatial data of 4-km resolution for these variables were collected for the 1990-2015 period from the TerraClimate dataset for calculating the annual average.

Human intervention is another important determinant in all wildfire studies. For example, accessibility to protected areas can be represented by road density because roads serve as pathways to the vegetated range and people can intentionally or unintentionally start a fire. Recent investigations have also shown that electricity lines can act as a source of fire ignitions (e.g., Miller et al., 2017). To determine the effects of road and electricity line networks in the study area, their data were extracted from the Topologically Integrated Geographic Encoding and Referencing system (TIGER) dataset, and then Euclidean distances to the electricity lines and kernel density of the road lines were calculated.

Gridding the study area was found to be an effective way to incorporate all the variables for the model in one attribute. This not only facilitated the desired analysis but also allowed extracting the required information from different GIS layers to a standard form. Therefore, a fishnet gridding was applied to produce a 10x10-km grid for training and predicting purposes, and then spatial joining was applied to join the required data from different layers to one attribute table. In regression analysis, the data need to be normally distributed, so various transformations, such as log transformation and square root transformation, were applied to standardize the variables.

## Multiple Linear Regression

The quality and robustness of the predicted model depend mainly on the training data, making variable selection for training the model vital. To identify significant variables, explanatory regression was applied. Through trial and error, the explanatory regression provides a means to narrow down the factors that are the best representative for explaining the dependent variables. Variables to run OLS were therefore selected based on the results of the explanatory regression.

OLS has been proven to be an effective method to investigate the relationships among variables in wildfire research. OLS determines the dependent variables ( $Y$ ) by utilizing the independent variable ( $X$ ) and produces an unbiased minimum sum of error square in  $Y$  (Fernandes & Leblanc, 2005). Different assumptions like normality, homogeneity, and independence of residual are also taken into account. These assumptions are essential for accurate modelling because violation of these can lead to inefficient and biased results (Montgomery et al., 2013). The general equation of OLS is given below:

$$Y = \beta_0 + \sum_{j=1}^p X_j \beta_j + \varepsilon \quad \dots \text{eq. (1)}$$

Where,  $Y$  is the dependent variable,  $\beta_0$  is the intercept of the model,  $\beta_j$  is the slope coefficient for the  $j^{th}$  independent variable  $X_j$  ( $j=1$  to  $p$ ), and  $\varepsilon$  is the random error term with expectation 0 and variance  $\sigma^2$ . Hence, the above equation can be simplified as:

$$Y = X\beta + \varepsilon \quad \dots \text{eq. (2)}$$

OLS provides various statistical outputs, aiding in model performance evaluation. Coefficients represent the type and strength of the relationship between the individual explanatory and dependent variables; it can be either positively or negatively correlated.

Analysis of variance (ANOVA) measures whether the model is statistically significant or not by calculating the p-value. The T-test (*p-value*) developed by Gössset is employed to identify the statistically significant parameters (e.g., Gössset, 1908).

$$T = \frac{Z}{s} = \frac{(\bar{X} - \mu)}{\frac{\hat{\sigma}}{\sqrt{n}}} \quad \dots \dots \dots \text{eq. (3)}$$

Where,  $\bar{x}$  is the sample mean from  $X_1, X_2, \dots, X_n$  out of a size  $n$ ,  $\mu$  is the population mean,  $\sigma$  is the standard deviation of the population,  $S$  is the standard error of the mean, and  $Z$  is the standardized statistic.

Here, the sample mean follows the tendency of normal distribution. The consistency in the relationship between the explanatory and dependent variables over geographic and data spaces is measured by Koenker (BP) Statistic. If BP results are statistically significant, this indicates robust probabilities. Multicollinearity is an important factor to consider when developing a model. It is defined as the inter-correlation among model variables, and it reveals redundancy in model variables that has to be removed. Higher multicollinearity leads to higher variance and covariance, eventually leading to unreliable statistical results. Variance Inflation Factor (VIF) checks for multicollinearity, and it is calculated for each variable by  $(1 - R^2)$  for the regression of that variable. Higher VIF expresses higher multicollinearity, so in this research, a threshold less than 7.5 for VIF was used. R-Squared and Akaike's Information Criterion (AICc) are used to infer overall model performance.

Through the smoothing of local regression, GWR was developed, and was later improved with the advancement of statistical measures (Brunsdon et al., 1999; Páez et al., 2002; Fotheringham et al., 2003). GWR incorporated spatial heterogeneity, making it different from

OLS, as OLS does not consider local variations and geographical data that may vary over space, making it difficult to fit everything into one global equation and may not even represent the true situation (Hanham & Spiker, 2005). Spatial data are non-stationary, their structure affects the correlation between variables, and the association between variables may vary in space; these are the three principals of GWR (Hanham & Spiker, 2005). GWR applies multiple calibrated local regression at each sample point to capture the spatial variation (Zhang et al., 2004). The general framework of GWR is:

$$Y_i = \beta_0(u_i v_i) + \beta_1(u_i v_i)x_1 + \beta_1(u_i v_i)x_1 + \beta_2(u_i v_i)x_2 + \dots + \beta_p(u_i v_i)x_p + \varepsilon_i; \\ i=1, 2, 3, \dots, n \quad \text{eq. (4)}$$

Where,  $(u_i, v_i)$  is point of coordinates at point  $i$ ,  $Y_i$  is the value of random variable,  $\varepsilon_i$  is the random error term, and  $X_i$  is the value of a fixed variable which is known and does not contain errors.

Three different smoothing functions are available in GWR: (1) predefined bandwidth, (2) corrected Akaike Information Criterion (AICc), and (3) cross validation. GWR provides important diagnostic statistics to understand local information by providing a standardized residual, the difference between the observed and the calculated values by the model, and this helps to evaluate the performance of the model. GWR also provides local variations of the adjusted R-squared, which demonstrates how well the model is in regard to explaining the dependent variable, and the leverage value, which measures the influence of an explanatory variable on the model calibration.

## **Random Forest Classification**

The RF classifier is a popular ML technique utilized in numerous research areas. By using Leo Breiman's RF algorithm in the model building and prediction process, this algorithm generates a “forest”, a group of decision trees, randomly, and these decision trees select a subset of randomly selected variables for classification (Breiman, 2001). RF is becoming popular in geospatial studies because of its flexibility and high predictive and computational capabilities in classification, regression, and unsupervised learning. Furthermore, it is a non-parametric nature-based method with the capability of determining the importance of input variables, and it provides an algorithm for estimating the missing values. Studies have shown that this method has a potential capability in fire risk prediction (e.g., McKenzie et al., 2000; Amatulli et al., 2006; Lozano et al., 2008; Oliveira et al., 2012; Rodrigues & de la Riva, 2014; Song et al., 2017). The Forest-based classification uses the training dataset to train the model based on the given explanatory values. This model can then be applied to predict unknowns based on a prediction dataset with the same explanatory variables used in the training process.

Training the model was accomplished for prediction by establishing a relationship between the explanatory variables and the variable to predict via building a “forest”. RF provides some measures to evaluate the performance of the model. In the building process, some data were excluded for validation and then the trained model predicted that excluded data and compared them with the observed value to estimate its accuracy. RF gives the option to predict a variable as a categorical or a continuous variable, as it has the capability to predict features to a fully different location, but it must have all the associated explanatory variables of the prediction area that were used in the training process. Following the training, numerous different diagnostic methods can be followed to assess the model performance, such as forest characteristics, Out Of

Bag (OOB) errors, and a summary of variable importance. When 100% of the dataset are used in training the model, then the OOB error can help to understand the model performance, as it is calculated by the unseen subset of data that are not used in building the “forest”. The variable importance is estimated via the Gini coefficient by calculating the number of times a variable cause split and the impact of that split divided by the number of “trees”. This process can exclude a portion of insignificant data while training the model.

## **RESULTS AND DISCUSSIONS**

### **OLS and GWR Outputs**

The dependent variable was defined as fire density for both training and prediction. For training the model, the fire-density of Oklahoma was selected. The average fire density is higher in both the northeast and in the southern parts, and especially in the southeastern portion of Oklahoma, where the density of wildfire is relatively higher than in the western side. After many trials and errors, the following predictors were selected for the analysis: slope, soil moisture, dry season precipitation (June-October), average maximum temperature (Tmax), PET, DEF, PDSI, portion of cultivated area, portion of tree area, portion of herb area, portion of shrub area, portion of barren area, portion of wildland urban interface, road lines density, and distance to electricity lines (Figure 2.3 a-j).

After running the OLS, the model showed the expected relation with fire density. The topographic parameter of slope was positively correlated with fire density, suggesting that fire tends to spread at a higher slope than at a lower slope. Vegetation was also positively correlated with fire density. All subclasses of vegetation, including trees, shrubs, and herbs were positively correlated with fire density, with trees having the strongest influence. Wildland urban interface (WUI) and barren area also showed a positive correlation with fire density, as well. Cultivated

land showed a negative correlation with fire density (Table 2.1), most likely because cultivated areas are frequently monitored and are in private ownership, making the chance of ignition extremely low. Relations between the fire density and the selected 15 variables are shown in figure 2.4.

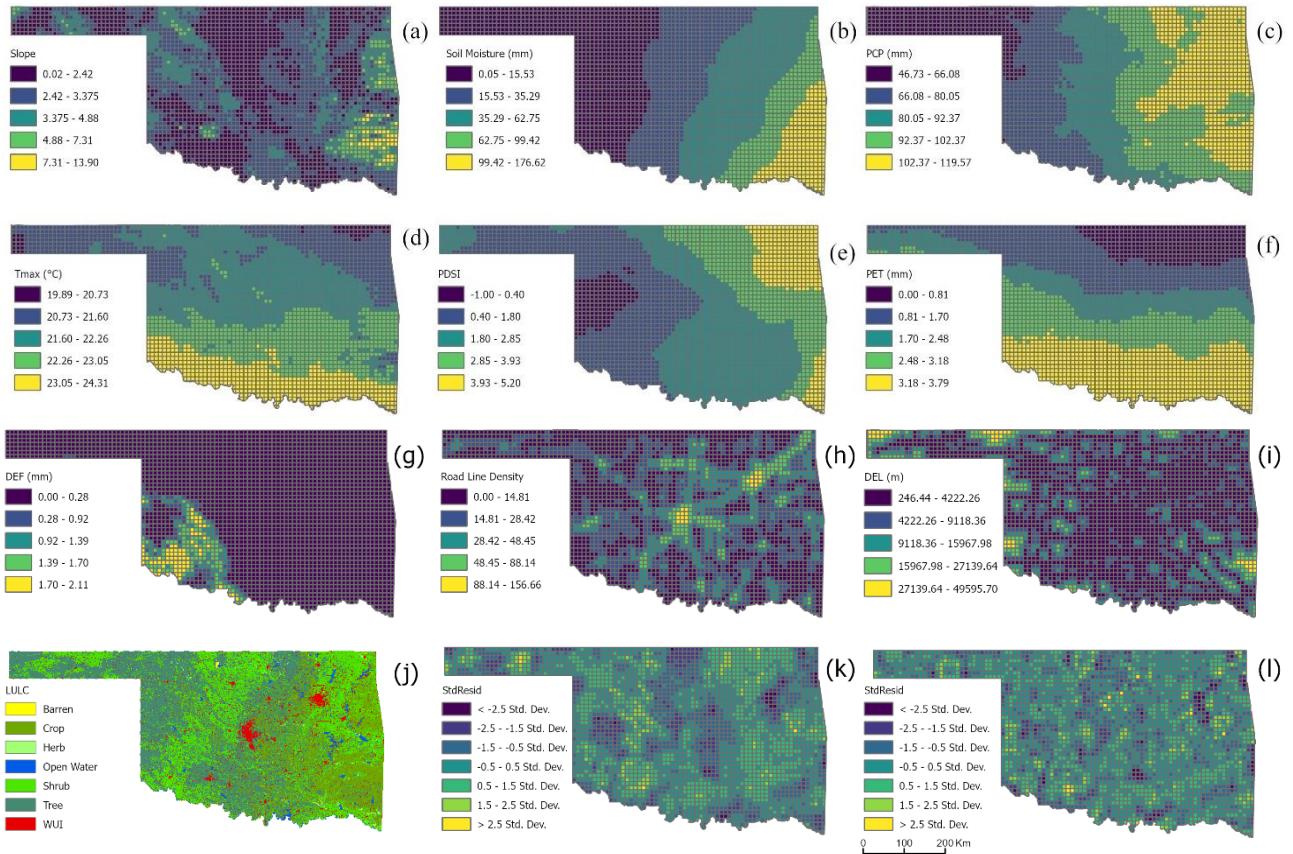


Figure 2.3: Selected explanatory variables (grid cell=10x10 km): (a) slope, (b) soil moisture, (c) dry season precipitation (PCP), (d) average maximum temperature (Tmax), (e) Palmer Drought Severity Index (PDSI), (f) potential evapotranspiration (PET), (g) climate water deficit (DEF), (h) road line density, (i) Euclidean distance to electricity lines (DEL), (j) land use/cover, (k) standard deviation (std) of residual from OLS, and (l) standard deviation (std) of residual from GWR.

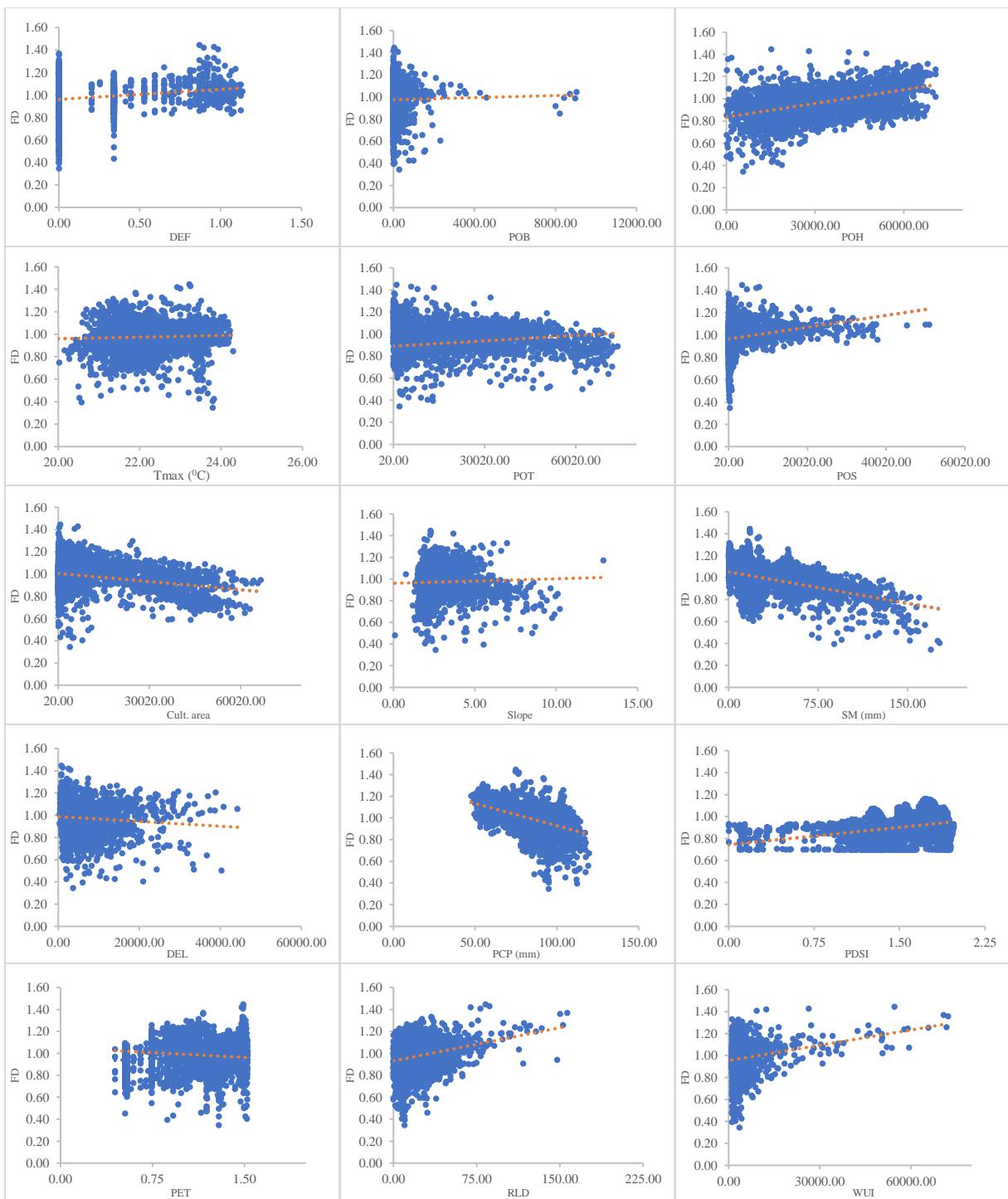


Figure 2.4: Scatter plots of the fire density versus the selected 15 variables. Soil moisture (SM), Euclidean distance to electricity lines (DEL), dry season precipitation (PCP), cultivated area (Cult. area), average maximum temperature (Tmax), potential evapotranspiration (PET), climate water deficit (DEF), road line density (RLD), portion of herb (POH), portion of tree (POT), PDSI (Palmer Drought Severity Index), portion of shrub (POS), wildland urban interface (WUI), and portion of barren (POB).

Table 2.1: Ordinary Least Square Results

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	-10.14432	1.357473	-7.472948	0.0000*	1.226492	-8.271007	0.0000*	.....
Slope	0.082004	0.028799	2.847436	0.0044*	0.028678	2.859514	0.0043*	2.8997
SM	-0.016795	0.001446	-11.61456	0.0000*	0.001380	-12.17155	0.0000*	5.5536
DEL	-0.000018	0.000004	-4.012207	0.0001*	0.000004	-4.424301	0.0000*	1.3836
PCP	0.044053	0.002921	15.084160	0.0000*	0.003131	14.070245	0.0000*	4.5952
Cult. area	-0.000002	0.000000	-5.893633	0.0000*	0.000000	-5.629685	0.0000*	2.3833
Tmax	0.347790	0.061850	5.623101	0.0000*	0.055513	6.264970	0.0000*	5.1519
PET	-0.368855	0.047081	-7.834447	0.0000*	0.043862	-8.409510	0.0000*	5.7289
DEF	0.672038	0.063530	10.578241	0.0000*	0.064860	10.361411	0.0000*	1.9731
PDSI	0.381437	0.105376	3.619785	0.0003*	0.108099	3.528603	0.0004*	2.9396
RLD	0.018232	0.001874	-9.730890	0.0000*	0.002036	-8.956737	0.0000*	2.1991
POH	0.000042	0.000002	20.829400	0.0000*	0.000002	19.821224	0.0000*	2.3023
POT	0.000054	0.000003	16.681918	0.0000*	0.000003	16.333287	0.0000*	7.4380
POS	0.000020	0.000005	3.657684	0.0003*	0.000005	4.250684	0.0000*	2.0090
WUI	0.000040	0.000006	7.108861	0.0000*	0.000006	6.419036	0.0000*	2.3549
POB	0.000058	0.000026	2.251220	0.0244*	0.000011	5.354982	0.0000*	1.0427

Soil moisture (SM), Euclidean distance to electricity lines (DEL), dry season precipitation (PCP), cultivated area (Cult. area), average maximum temperature (Tmax), potential evapotranspiration (PET), climate water deficit (DEF), road line density (RLD), portion of herb (POH), portion of tree (POT), PDSI (Palmer Drought Severity Index), portion of shrub (POS), wildland urban interface (WUI), portion of barren (POB), significant p-values <0.01 are denoted by \*, StdError is the standard deviation error, t-statistic is the ratio between estimated and hypothesized values relative to StdError, probability and robust probability (Pr) are significant when p-values <0.01, and VIF (variance inflation factor) with values <7.5 are considered.

Dry season precipitation and average maximum temperature tended to have a direct relationship with fire density. As the precipitation in the dry season decreased, the fire density increased. Additionally, fire density also increased as the maximum temperature increased. Other positively related factors to fire density included PDSI and DEF. Average soil moisture, however, showed a negative relation with fire density, proving that soil moisture inversely affected fire incidences. Furthermore, PET showed a negative correlation with fire density, along with human-induced factors like distance to electric lines, but the road lines density was positively associated with fire density (Table 2.1).

The P-value for all these explanatory variables is less than 0.1, indicating that these variables are statistically significant. In addition, multicollinearity is tested by VIF, retrieving values less than 7.5. Choosing the threshold for VIF has much divergence in the literature. A VIF value between 5 and 10 is usually recommended as values out of this range would indicate the existence of collinearity and thus redundancy (e.g., Petter et al., 2007; Cenfetelli & Bassellier, 2009; Hair, 2009; Kline, 2015). Among the 15 selected variables, only four have VIF values over 5; however, these are the most significantly important factors in modelling fire susceptibility, so a threshold of 7.5 has been applied in this study. Robust probability value is also less than 0.1 (Table 2.1), adding more confidence to the regression results. The performance of the model is evaluated by the adjusted R-squared value, calculated to be 0.51, meaning the model can explain nearly half of the observed variation by these explanatory variables (Table 2.2).

Table 2.2: Performance results of OLS and GWR

OLS Results		GWR Results	
Adjusted R-Squared	0.505702	Adjusted R-Squared	0.8703
Joint Wald Statistic	2888.262294	Multiple R-Squared	0.8941
Koenker (BP) Statistic	120.552969	Sigma-Squared	0.3408
Jarque-Bera Statistic	72.179020	Sigma-Squared MLE	0.2783
AICc	8572.895584	AICc	5171.1933

AICc is Akaike's Information Criterion.

OLS was able to explain about 50% of the variance of the dependent variable with the explanatory variables; however, GWR provided better results as the adjusted R-squared value increased from 0.51 to 0.87 (Table 2.2). This demonstrates that the variables are spatially related, and GWR is superior to OLS in this case for fire density modelling. In this study, the golden search approach was applied to determine the suitable neighbour size. This approach tries to find a suitable number of neighbours based on the lowest AICc value. The findings determined the suitable number of neighbours as 31 with the AICs value of 5171, and GWR provided better

results again compared to OLS, providing confidence to use these variables in a non-parametric method to predict fire density at a fully different location (Table 2.2).

## RF Outputs

Fire distribution showed an irregular pattern in the training dataset, and the results of OLS and GWR indicated that fire density is influenced by both natural and human-related factors. The adjusted R-squared value inferred a nonlinear relationship (Table 2.2), suggesting that a non-parametric method is more suitable for this study (e.g., Oliveira et al., 2012). This explains why the results from RF have shown better performance than linear regression. Initially, the model was trained based on the dataset of Oklahoma, and all 15 variables were used in forest-based regression to train the model (Figures 2.2 a-j). For training, a total number of 300 “trees” were selected to test the model, and 20% of the data were excluded for validation purposes. The adjusted R-squared value for the training was 0.98 with a standard error of 0.002, and a p-value of 0.00, indicating that the result is statistically significant. For validation, the adjusted R-squared value was 0.92 with a standard error of 0.01, and a p-value of less than 0.00. The validation data shares more than 95% of the training data value range, indicating that the model is well-validated. Overall, the results show that the model is well-fitted for prediction. Therefore, the same 15 explanatory variables for the prediction area were fed to the model to predict wildfire densities across Arkansas (Figure 2.5).

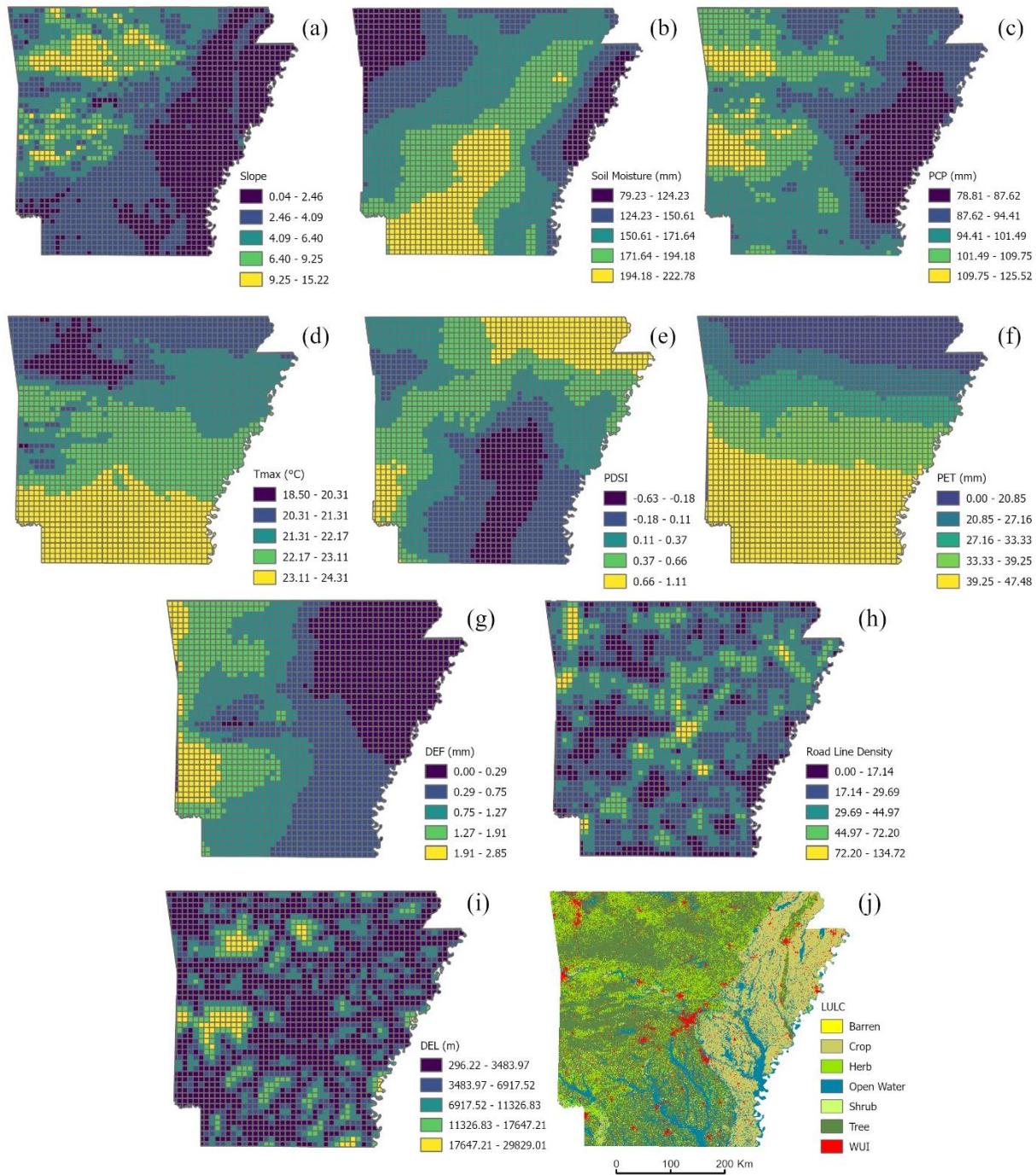


Figure 2.5: Identified variables for predicting fire susceptibility in Arkansas (grid cell=10x10 km): (a) slope, (b) average soil moisture, (c) dry season precipitation (PCP), (d) average maximum temperature (Tmax), (e) Palmer Drought Severity Index (PDSI), (f) potential evapotranspiration (PET), (g) climate water deficit (DEF), (h) road line density, (i) Euclidean distance to electricity lines (DEL), and (j) land use/cover.

To predict features at a different location, the model needed all the associated explanatory variables that were used in the training process. Thus, a 10x10-km grid was created for Arkansas and spatial joining was applied to join the necessary data from different GIS layers to one attribute table.

A total of 300 “trees” were used in the prediction process, Model Out of Bag (OOB) Errors provided Mean Squared Error (MSE) and the percentage of variation to assess the accuracy of the model. In the prediction, no percentage of the data were excluded for validation, and thus the OOB results were utilized to evaluate the model performance. The MSE value for 50% of the “trees” was 924.42, and the percentage of variation explained was 90.14; for 100% of the “trees”, the MSE was 913.81, and the percentage of variation explained was 90.26. The difference in values for both cases was minimal, indicating the number of “trees” had minimal effect on the model outcome.

Of the 15 variables, PET showed the highest importance in the RF (Table 2.3); however, in MLR, it showed a negative strong relationship. The influence of PET on wildfires was explained in details in previous work (e.g., Kane et al., 2015). Wildland fuels such as live vegetation, organic soil, and dead fuels are affected by water deficits; therefore, increasing the water deficit may make fuels drier and more susceptible to ignition and burning (Dimitrakopoulos et al., 2010). It has been found that summer evapotranspiration has a significant correlation with wildfires in southwest and southern California (Abatzoglou & Kolden, 2013). Soil moisture directly influences the dryness of fuels, and affects the dead fuels that are generally found in the ground, so it acts like a proxy to drought (Chuvieco et al., 2004). It has been found that under low wet conditions at the surface, burnable fuels are more susceptible to burning (Bartsch et al., 2009). PDSI shows significant importance in explaining

wildfires, as previous wildfires in the western states have shown a correlation with PDSI (Collins et al., 2006). PDSI with precipitation and temperature have been used previously to explain fire events in the 12 ecoregions of the western US (Littell et al., 2009). Human factors, as well, have a major influence on fire density. Many reports have identified faulty electricity lines acted as a source of ignition.

Table 2.3: Variables used in the analysis and their importance and percentage

Variable	Importance	%
PET	4497387.15	22
SM	3508886.60	17
PDSI	2466707.62	12
PCP	2082161.71	10
POS	1830527.78	9
Cult. area	1689034.18	8
Slope	1023721.58	5
POH	841701.74	4
Tmax	754822.33	4
POT	690317.75	3
RLD	417814.53	2
WUI	255724.12	1
DEL	216439.31	1
POB	188394.61	1
DEF	53002.19	0

Soil moisture (SM), Euclidean distance to electricity lines (DEL), dry season precipitation (PCP), cultivated area (Cult. area), average maximum temperature (Tmax), potential evapotranspiration (PET), climate water deficit (DEF), road line density (RLD), portion of herb (POH), portion of tree (POT), portion of shrub (POS), wildland urban interface (WUI), portion of barren (POB), and PDSI (Palmer Drought Severity Index).

Climatic factors have a significant impact on fire distribution (Drever et al., 2008).

Average maximum temperature and dry season precipitation have demonstrated noticeable importance in the model. In this study, average dry season precipitation showed more importance than average maximum temperature. Temperatures above 20°C promoted wildfires and showed a positive correlation in the linear regression.

The only topographic factor included in this fire modelling is the slope; initially, elevation and aspect factors were also included, but they were eventually excluded due to multicollinearity. The slope of the area showed a notable effect on the RF results, as well as vegetation covers like portions of tree, shrub, and herb, which acted as proxies for fuel. In the linear regression, these variables demonstrated a significant correlation with fire density. Cultivated area, barren, and WUI also showed some importance in the RF results.

The final outcome from the RF modelling is the fire susceptibility map for Arkansas (Figure 2.6), showing 5 categories of fire likelihood across the entire state. Results indicate that the Ouachita National Forest and the Ozark Forest are highly susceptible to wildfires, while the eastern side of the state is the least sensitive to wildfire, and the southern part of the state showing only a moderate risk of fire. The Ouachita National Forest is close to Little Rock, a densely populated area in central Arkansas, making this area at a high risk of wildfire. Northwest Arkansas is at medium risk of wildfire, as it is close to the Ozark Forest, a high-risk zone for wildfires. It is noteworthy that most of the historic fire incidences are coincident with the high-risk modelled category, which adds more confidence to the established model. The predicted wildfire susceptibility showed a good correlation within the high and high-medium susceptible areas, while the model did not accurately predict some fire incidences in the southern part of the state. In fact, most of these incidences are just prescribed burns done to maintain the forest and agricultural production, but they are still documented as wildfires in the archived data. The southern part is known as the West Gulf Coastal Plain and has a gentle slope with rich soil properties. It includes extensive areas of loblolly-shortleaf pine forests and agricultural fields. Forest harvesting related to hardwood products is predominated in this area.

## CONCLUSIONS

Wildfires are one of the major hazards in the United States, costing the national economy billions of dollars every year. Recent human influences and contemporary climate changes heightened the severity and frequency of wildfires across the States. Consequently, it is important to regularly evaluate the relationship between the fire events and the factors that may trigger wildfires. In this study, MLR and RF were used to assess the connection between forest fires and 15 identified variables, including slope, soil moisture, dry season precipitation (June–October), average maximum temperature (Tmax), PET, DEF, PDSI, portion of cultivated area, portion of tree area, portion of herb area, portion of shrub area, portion of barren area, portion of wildland urban interface, road lines density, and distance to electricity lines. Fire density and the average of different explanatory variables were used to investigate the relationship through OLS, GWR and RF, with RF outperforming OLS and GWR. This study demonstrated the prediction capability of RF.

Initially, the key explanatory variables were selected for explaining fire density using MLR, providing a robust statistical means to measure the association and finding the most significant variables for predictions. The model was then trained with the identified variables, and a portion of data was excluded for validation to evaluate the model performance. In training and validation, the model provided significant statistical outputs, highlighting high precision in the prediction process by the RF model.

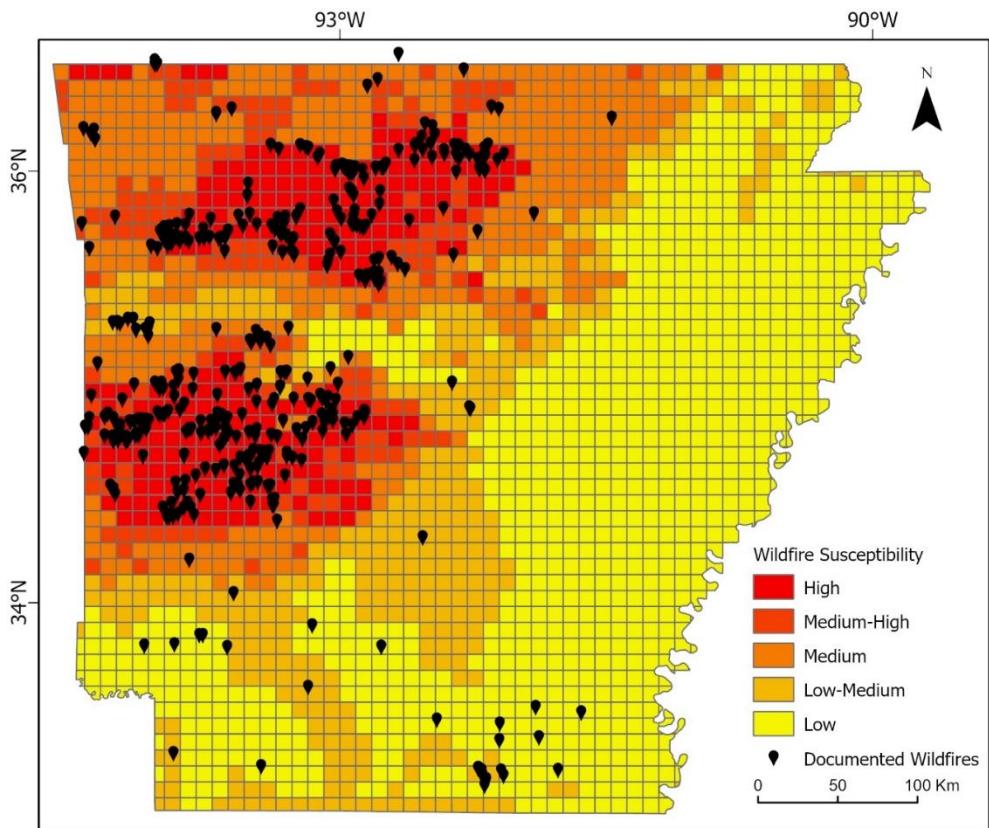


Figure 2.6: Documented fire incidences plotted on top of the modeled fire likelihood (grid cell=10x10 km).

The produced fire susceptibility categories were cross-checked with the documented fire incidences, and this showed corresponding results (Figure 2.6). The model incorporated both physical and human factors to predict fire-susceptible areas. The modelling results revealed a strong relationship between the fire density and the explanatory variables used in the analysis. The relationship between fire density and contributing factors may vary by location, but the model still can be used to identify the main sources of fire. Fire prevention and planning can benefit from these results to better prepare for future fire events and minimize the loss due to fires by taking necessary precaution measures in a given area according to the local fire drivers.

## ACKNOWLEDGEMENTS

This study was partially funded by USGS and ArkansasView grant #GR908148UAF awarded to Dr. Mohamed Aly. Thanks to USGS, NASA, the Climatology Lab, and the United States Census Bureau for providing necessary data for conducting this research.

## REFERENCES

- Abatzoglou, J. T., & Kolden, C. A. (2013). Relationships between climate and macroscale area burned in the western United States. *International Journal of Wildland Fire*, 22(7), 1003–1020.
- Adab, H., Kanniah, K. D., & Solaimani, K. (2013). Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards*, 65(3), 1723–1743.
- Akther, M. S., & Hassan, Q. K. (2011). Remote sensing-based assessment of fire danger conditions over boreal forest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(4), 992–999.
- Aldersley, A., Murray, S. J., & Cornell, S. E. (2011). Global and regional analysis of climate and human drivers of wildfire. *Science of the Total Environment*, 409(18), 3472–3481. <https://doi.org/10.1016/j.scitotenv.2011.05.032>
- Amatulli, G., Rodrigues, M. J., Trombetti, M., & Lovreglio, R. (2006). Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences*, 111(G4). <https://doi.org/10.1029/2005JG000133>
- Arpacı, A., Malowerschnig, B., Sass, O., & Vacik, H. (2014). Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography*, 53, 258–270. <https://doi.org/10.1016/j.apgeog.2014.05.015>
- Balch, J. K., Schoennagel, T., Williams, A. P., Abatzoglou, J. T., Cattau, M. E., Mietkiewicz, N. P., & St Denis, L. A. (2018). Switching on the Big Burn of 2017. *Fire*, 1(1), 17.
- Bartsch, A., Balzter, H., & George, C. (2009). The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. *Environmental Research Letters*, 4(4), 045021.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Brunsdon, C., Aitkin, M., Fotheringham, S., & Charlton, M. (1999). A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems. *Geographical and Environmental Modelling*, 3(1), 47–62.

- Carlson, J., Burgan, R. E., Engle, D. M., & Greenfield, J. R. (2002). The Oklahoma Fire Danger Model: An operational tool for mesoscale fire danger rating in Oklahoma. *International Journal of Wildland Fire*, 11(4), 183–191.
- Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, 689–707.
- Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W., & Guo, Z. (2020). Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models. *Remote Sensing*, 12(3), 502.
- Chowdhury, E. H., & Hassan, Q. K. (2013). Use of remote sensing-derived variables in developing a forest fire danger forecasting system. *Natural Hazards*, 67(2), 321–334. Agricultural & Environmental Science Collection; ProQuest Central.  
<https://doi.org/10.1007/s11069-013-0564-7>
- Chuvieco, E., Aguado, I., & Dimitrakopoulos, A. P. (2004). Conversion of fuel moisture content values to ignition potential for integrated fire danger assessment. *Canadian Journal of Forest Research*, 34(11), 2284–2293.
- Clutter, M., Mendell, B., Newman, D., Wear, D., & Greis, J. (2005). *Strategic factors driving timberland ownership changes in the US South*.
- Collins, B. M., Omi, P. N., & Chapman, P. L. (2006). Regional relationships between climate and wildfire-burned area in the Interior West, USA. *Canadian Journal of Forest Research*, 36(3), 699–709.
- Dickson, B. G., Prather, J. W., Xu, Y., Hampton, H. M., Aumack, E. N., & Sisk, T. D. (2006). Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecology*, 21(5), 747–761.
- Dimitrakopoulos, A., Mitsopoulos, I., & Gatoulas, K. (2010). Assessing ignition probability and moisture of extinction in a Mediterranean grass fuel. *International Journal of Wildland Fire*, 19(1), 29–34.
- Dlamini, W. M. (2010). A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software*, 25(2), 199–208.
- Drever, C. R., Drever, M. C., Messier, C., Bergeron, Y., & Flannigan, M. (2008). Fire and the relative roles of weather, climate and landscape characteristics in the Great Lakes-St. Lawrence forest of Canada. *Journal of Vegetation Science*, 19(1), 57–66.
- Fernandes, R., & G. Leblanc, S. (2005). Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sensing of Environment*, 95(3), 303–316.  
<https://doi.org/10.1016/j.rse.2005.01.005>

- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.
- Gorte, R., & Economics, H. (2013). *The rising cost of wildfire protection*. Headwaters Economics Bozeman, MT.
- Gösset, W. S. (1908). The probable error of a mean. *Biometrika*, 6, 1–25.
- Hair, J. F. (2009). Multivariate data analysis.
- Hanham, R., & Spiker, J. S. (2005). Urban sprawl detection using satellite imagery and geographically weighted regression. In *Geo-spatial technologies in urban environments* (pp. 137–151). Springer.
- He, W., Goodkind, D., & Kowal, P. (2016). US Census Bureau, international population reports. *An Aging World: 2015*, 16–1.
- Hodgdon, B., & Tyrrell, M. (2003). Literature review: An annotated bibliography on family forest owners. *GISF Research Paper*, 2.
- Jaafari, A., & Pourghasemi, H. R. (2019). Factors influencing regional-scale wildfire probability in Iran: An application of random forest and support vector machine. In *Spatial modeling in GIS and R for Earth and environmental sciences* (pp. 607–619). Elsevier.
- Kane, V. R., Cansler, C. A., Povak, N. A., Kane, J. T., McGaughey, R. J., Lutz, J. A., Churchill, D. J., & North, M. P. (2015). Mixed severity fire effects within the Rim fire: Relative importance of local climate, fire weather, topography, and forest structure. *Forest Ecology and Management*, 358, 62–79.
- Kline, R. B. (2015). Principles and practice of structural equation modeling. Guilford publications.
- Littell, J. S., McKenzie, D., Peterson, D. L., & Westerling, A. L. (2009). Climate and wildfire area burned in western US ecoregions, 1916–2003. *Ecological Applications*, 19(4), 1003–1021.
- Lozano, F. J., Suárez-Seoane, S., Kelly, M., & Luis, E. (2008). A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment*, 112(3), 708–719.
- McKenzie, D., Peterson, D. L., & Agee, J. K. (2000). Fire frequency in the interior Columbia River Basin: Building regional models from fire history data. *Ecological Applications*, 10(5), 1497–1516.
- Miller, C., Plucinski, M., Sullivan, A., Stephenson, A., Huston, C., Charman, K., Prakash, M., & Dunstall, S. (2017). Electrically caused wildfires in Victoria, Australia are over-

represented when fire danger is elevated. *Landscape and Urban Planning*, 167, 267–274. <https://doi.org/10.1016/j.landurbplan.2017.06.016>

Miller, J. D., Skinner, C., Safford, H., Knapp, E. E., & Ramirez, C. (2012). Trends and causes of severity, size, and number of fires in northwestern California, USA. *Ecological Applications*, 22(1), 184–203.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2013). *Introduction to Linear Regression Analysis* (5th ed.). John Wiley & Sons, Incorporated.

Nowak, D. J., & Greenfield, E. J. (2018). US Urban Forest Statistics, Values, and Projections. *Journal of Forestry*, 116(2), 164–177. Agricultural & Environmental Science Collection; eLibrary; ProQuest Central; STEM Database. <https://doi.org/10.1093/jofore/fvx004>

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. (2012a). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012b). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>

Páez, A., Uchida, T., & Miyamoto, K. (2002). A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A*, 34(4), 733–754.

Parisien, M. A., Snetsinger, S., Greenberg, J. A., Nelson, C. R., Schoennagel, T., Dobrowski, S. Z., & Moritz, M. A. (2012). Spatial variability in wildfire probability across the western United States. *International Journal of Wildland Fire*, 21(4), 313–327. <https://doi.org/10.1071/wf11044>

Park, H., & Kim, K. (2019). Prediction of severe drought area based on random forest: Using satellite image and topography data. *Water*, 11(4), 705.

Pelkki, M. H. (2005). *An economic assessment of Arkansas' forest industries: Challenges and opportunities for the 21st century*. Arkansas Agricultural Experiment Station.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 623–656.

Reid, A. M., Fuhlendorf, S. D., & Weir, J. R. (2010). Weather variables affecting Oklahoma wildfires. *Rangeland Ecology & Management*, 63(5), 599–603.

Rodrigues, M., & de la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, 57, 192–201. <https://doi.org/10.1016/j.envsoft.2014.03.003>

- Rodrigues, M., Jiménez-Ruano, A., Peña-Angulo, D., & de la Riva, J. (2018). A comprehensive spatial-temporal analysis of driving factors of human-caused wildfires in Spain using Geographically Weighted Logistic Regression. *Journal of Environmental Management*, 225, 177–192. <https://doi.org/10.1016/j.jenvman.2018.07.098>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18), 9276–9282.
- Rowden, K. W., & Aly, M. H. (2018). GIS-based regression modeling of the extreme weather patterns in Arkansas, USA. *Geoenvironmental Disasters*, 5(1), 1–15.
- Song, C., Kwan, M.-P., Song, W., & Zhu, J. (2017). A comparison between spatial econometric models and random forest for modeling fire occurrence. *Sustainability*, 9(5), 819.
- Steelman, T. (2016). U.S. wildfire governance as social-ecological problem. *Ecology and Society*, 21(4). JSTOR. <http://www.jstor.org/stable/26270036>
- Stephenson, N. (1998). Actual evapotranspiration and deficit: Biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography*, 25(5), 855–870.
- Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., & Clayton, M. K. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, 17(5), 602–613. <https://doi.org/10.1071/wf07087>
- Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, 2(2), 159–178.
- Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelm, N., Wei, Y., Devarakonda, R., & Cook, R. B. (2018). Daymet: Annual Climate Summaries on a 1-km Grid for North America, Version 2. *ORNL DAAC, Oak Ridge, TN*.
- Valdez, M. C., Chang, K.-T., Chen, C.-F., Chiang, S.-H., & Santos, J. L. (2017). Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. *Geomatics, Natural Hazards and Risk*, 8(2), 876–892. <https://doi.org/10.1080/19475705.2016.1278404>
- Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130–1141.
- Weir, J. R., Reid, A. M., & Fuhlendorf, S. D. (2012). *Wildfires in Oklahoma*. *Wildfire Statistics*. (n.d.). Arkansas Department of Agriculture. Retrieved July 21, 2021, from <https://www.agriculture.arkansas.gov/forestry/wildfire-statistics/>
- Yang, J., Weisberg, P. J., Dilts, T. E., Loudermilk, E. L., Scheller, R. M., Stanton, A., & Skinner, C. (2015). Predicting wildfire occurrence distribution with spatial point process models

and its uncertainty assessment: A case study in the Lake Tahoe Basin, USA.  
*International Journal of Wildland Fire*, 24(3), 380–390. <https://doi.org/10.1071/wf14001>

Zhang, L., Bi, H., Cheng, P., & Davis, C. J. (2004). Modeling spatial variation in tree diameter–height relationships. *Forest Ecology and Management*, 189(1–3), 317–329.

## **CHAPTER 3**

### **BIG DATA ANALYSES FOR DETERMINING THE SPATIO-TEMPORAL TRENDS OF AIR POLLUTION DUE TO WILDFIRES IN CALIFORNIA USING GOOGLE EARTH ENGINE**

#### **ABSTRACT**

California has a long history of large-scale severe fire events emitting smoke and plumes into the atmosphere, which have resulted in regional air pollution. With the advancement in Earth observation satellites and cloud computing capabilities in remote sensing, monitoring air pollution became possible through a long-term analysis of satellite images. This research utilizes the Google Earth Engine (GEE) platform to navigate its geospatial datasets of MODIS MCD19A2 Version 6 level 2 and Sentinel-5P products to investigate the impact of (2010-2020) wildfires on air quality in California. MODIS MCD19A2 Version 6 level 2 uses an advanced MAIAC algorithm to produce 1-km resolution images, which retrieves Aerosol Optical Depth (AOD) at 470 nm and 550 nm wavelengths. Historical fire events are cross-validated using MODIS 1-km MYD14A1 V6 dataset, and AOD values are validated using ground-based sun photometers AERONET AOD measurements. The Mean Absolute Error (MAE), the Relative Mean Bias (RMB), and the Root Mean Square Error (RMSE) are used to check uncertainty of the analysis, and the linear regression is used to verify the correlation between satellite and ground measurements. The average monthly MODIS MAIAC AOD at 470 nm and 550 nm tends to overestimate AOD by 19% and 4%, respectively, compared to the AERONET AOD values. 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. During the fire season (May-

October), more accurate correlations with a correlation coefficient above 0.8 and the adjusted R-squared value above 0.65 are found for both wavelengths. Sentinel-5P data shows that the 2020 wildfires significantly raise the NO<sub>2</sub> concentration in its surrounding areas. This study identifies the historical spatio-temporal trends of air pollution due to (2010-2020) wildfires in California, which can help in making informed decisions in prevention and mitigation programs.

**KEYWORDS:** Aerosol Optical Depth, Google Earth Engine, MODIS MAIAC, Sentinel-5P, AERONET, Air Pollution, California.

## INTRODUCTION

Wildfires are a common natural ecological disturbance shaping the landscape of the United States; they can be defined as the uncontrollable and unwanted burning of combustible vegetation, ultimately creating a wide range of ecological and environmental destruction. Wildfires not only destroy forest resources but also have secondary impacts such as the release of carbon dioxide and particulate matter into the air, extensive soil erosion and property alterations, and large-scale landscape alteration (Akther & Hassan, 2011). Studies have suggested that wildfire patterns are changing across the United States as time progresses (Liu et al., 2010; Jolly et al., 2015). In addition, the United States fire season being 78 days longer during 1987-2002 compared to 1970-1986; Between 1984 and 2010, the occurrence of large fires (>5000 ha) has also increased in the southwestern and southeastern United States (Preisler & Westerling, 2007; Barbero et al., 2014; Dennison et al., 2014). In the western United States, the number of large wildfires of burned areas more than 1,000 acres is increasing over years. Unfortunately, recent climate change and landscape development have made California prone to larger and more damaging wildfires. The state has recorded an average of 4,183 wildfires per

year between 2010 and 2015, resulting in about 108,256 acres of burned areas (Fire, 2018). The burning of biomass leads to a huge amount of smoke emission into the atmosphere, and this smoke can travel a long distance and cause large-scale air pollution.

Atmospheric aerosols emitted from wildfires play an important role in the Earth's climate and hydrological cycle. Aerosols consisting of a mixture of solid and liquid particles are suspended in the atmosphere (Qi et al., 2013). They can alter the radiative balance of the Earth-atmosphere by scattering and absorbing solar radiation, as well as change the microphysical properties of the clouds like ice nuclei and cloud condensation nuclei (Jianping et al., 2006; Andrews et al., 2011). The lifetime of aerosols varies from few days to weeks, obscuring our understanding of its role (Lu & Guo, 2012). Wildfires, however, greatly alter the aerosol properties of the atmosphere within a very short period. This can cause large uncertainty to the radiative forcing of the atmosphere system. Therefore, long term spatio-temporal observation of the aerosol optical characteristics is needed to better understand the impact of wildfires on the aerosol distribution. Aerosol Optical Depth (AOD), which quantifies the vertical column of atmospheric aerosol through their potential radiative properties, is an important component for measuring the aerosol optical properties (Bright & Gueymard, 2019). It can be measured by ground-based monitoring stations and through space-based satellite observations. Ground-based stations have highly accurate temporal coverage, but cover a limited reach; while satellite observations have a larger spatial coverage, but low spectral resolution (Bilal et al., 2015). Recent advancements in Earth Observing Satellites have enabled assessing aerosol loading and properties at finer spatial and temporal resolutions. Some of the popular satellites that are used in atmospheric aerosol monitoring are Multi-angle Imaging Spectroradiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS), Visible Infrared Imaging Radiometer Suite

(VIIRS), Ozone Monitoring Instrument (OMI), Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), and Polarization and Directionality of the Earth's Reflectance (POLDER) (Kaufman et al., 2002; Remer et al., 2005; Torres et al., 2007; Winker et al., 2010; Kahn & Gaitley, 2015).

MODIS provides AOD measurements of the globe and has been used successfully for long term monitoring. MODIS has on-boarded the Terra and Aqua satellites and acquires data spanning 36 spectral bands (McPhetres & Aggarwal, 2018). It retrieves the AOD values using dark target (DT) ocean algorithms, DT land algorithms, and deep blue (DB) algorithms. Ocean AOD values are retrieved using DT ocean algorithms at seven wavelengths (470, 550, 660, 870, 1200 and 2100 nm), and land AOD values are obtained using DT land algorithms across three wavelengths (470, 550, and 660 nm). MODIS Collection 6 AOD acquires data at 10-km resolution for Aqua, validated at 550 nm, and 3-km resolution for Terra. MODIS DT and DB (10 km) AOD products have a coarse spatial resolution and MODIS DT (3 km) AOD products have low retrieval accuracy (Hsu et al., 2013; Levy et al., 2013; Sayer et al., 2013; Sayer et al., 2015; McPhetres & Aggarwal, 2018). To overcome these limitations, MAIAC algorithm has been developed.

An advanced and improved MAIAC algorithm uses a time-series analysis and a combination of image-based and pixel-based processes to obtain aerosol values (Lyapustin et al., 2011a,b). MAIAC, combined with MODIS MCD19A2 Version 6 level 2 products, provide daily images with AOD values at 1-km pixel resolution. These newly improved products were successfully employed to obtain AOD data over bright urban areas in New England, the United States, South America, metropolitan areas across Mexico City, and regions in Israel (Just et al., 2015; Kloog et al., 2015; Superczynski et al., 2017; Martins et al., 2017). MAIAC was applied in

monitoring the biomass burning aerosol, as it has improved cloud detection capabilities, enabling it to measure AOD over partially cloudy days, as compared to the conventional 10-km resolution MODIS AOD products that cannot retrieve data on such days (Mhawish et al., 2019). This gives ideal means to monitor long-term changes and determine the impact of fires on air quality. MODIS also provides thermal anomaly data, which are extensively used in fire identification and monitoring (Huang et al., 2012; McCarley et al., 2020). Using the MODIS 4 and 11-micrometer radiances, MYD14A1 V6 dataset provides high resolution (1 km) fire mask composite (Kodandapani & Theme, n.d.). The fire radiative power (FRP) is linearly related to biomass burning, identifying active fire events (Wooster et al., 2005; Freeborn et al., 2011). Ground-based monitoring provides regular measurements of AOD at higher temporal and spectral resolutions. The Aerosol Robotic Network (AERONET) obtains AOD measurements from sun photometers, which provide global coverage (Bilal et al., 2015). Often, AOD values from AERONET and MODIS are compared for analyzing the atmospheric property of a region. The integration of MODIS and AERONET data provides an excellent capability for monitoring the impacts of wildfires on air quality. Long-term analysis of satellite images requires large storing capabilities and high computing capabilities; therefore, Google Earth Engine (GEE), with its archive of petabytes of free Earth Observation (EO) data, makes an ideal platform for analyzing long-term spatial and temporal changes.

The recently launched Sentinel-5 Precursor (5P) provides additional ways to monitor atmospheric parameters like air quality, ozone & Ultraviolet (UV) radiation, and climate monitoring at a finer resolution. This is capable of measuring ozone ( $O_3$ ), methane ( $CH_4$ ), formaldehyde ( $CH_2O$ ), aerosol, carbon monoxide (CO),  $NO_2$ , and  $SO_2$  in the atmosphere, which can provide additional information about the impacts of wildfire on atmospheric compositions.

Wildfires are anticipated to abruptly change the atmospheric composition within a short period, and historical analysis of AOD over California using high spectral resolution images is still limited. This study utilized the GEE platform to access available MODIS and Sentinel-5P datasets from 2010 to 2020 and evaluated the performance of the MODIS 1-km resolution MYD14A1 V6 dataset in detecting previous wildfires. MODIS MAIAC-derived AOD measurements were validated using ground-based AERONET AOD records. This study can improve understanding of the short-term and long-term effects of wildfires on air quality, fulfilling its objectives to investigate historical trends of AOD between 2010 and 2020 and to assess the impact of wildfires on air quality during the fire season (May-October) in California.

## **STUDY AREA**

California has a broad range of ecosystems, topology, climate, vegetation, and fuel distribution, all contributing to its fire regime. The state can be divided into eleven distinct geomorphic provinces (Figure 3.1); from north to south, they are: Klamath Mountains, Cascade Range, Modoc Plateau, Basin and Range, Coast Ranges, Central Valley, Sierra Nevada, Transverse Ranges, Mojave Desert, Peninsular Ranges, and Colorado Desert (Norris & Webb, 1990). The climate of California varies on a north-south and east-west gradient, with the northwest tending to be cooler and wetter than the southeast. The northwest portion of the state contains mainly forest, while shrublands are dominating in the southeastern portion. The climatic condition coupled with landscape formation determine the size and severity of fires. The diversity in climatic conditions, including the relative humidity and wind speeds and directions combined with the variability in the vegetation cover and topography, has made California susceptible to wildfires (Jin et al., 2015; Williams et al., 2019; Williams & Abatzoglou, 2016). The main driver of wildfires in the southern portion is Santa Ana winds, which can trigger

wildfires year-round but tend to be highly active during December (Dennison et al., 2008). Devastating and long-lasting fires are seen in fall, in comparison to spring fires, which can be easily contained. Fires not only destroy the vegetation of an area but also change its landscape, and increased fire events can lead to the generation of non-native species and the extinction of native species (Rother & De Sales, 2021).

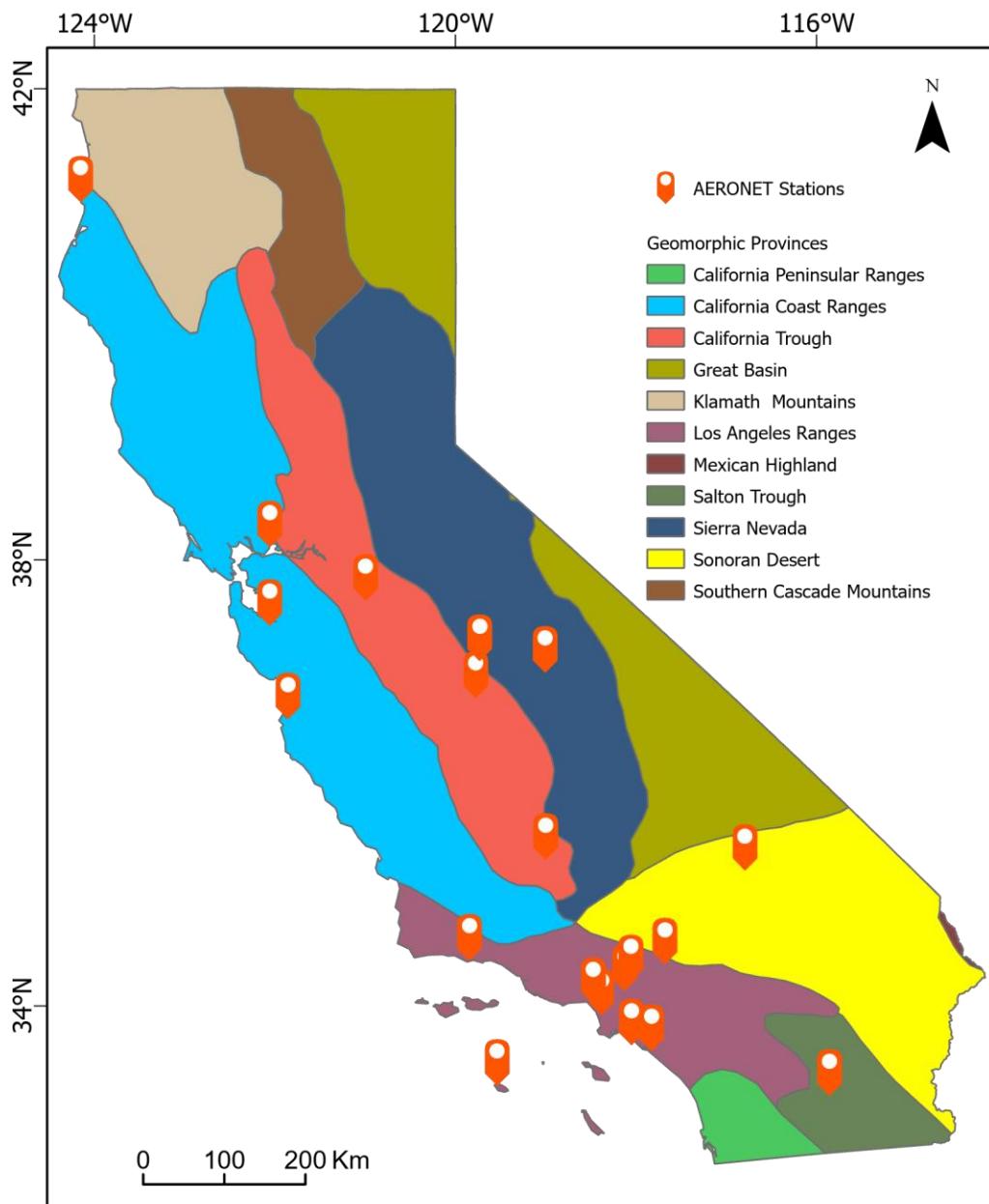


Figure 3.1: Geomorphic provinces of California and location of the 21 AERONET stations used in this study.

## **DATASETS AND METHODS**

### **AERONET**

The AERONET is a ground-based measurement of global aerosol properties using sun photometers. These properties are measured using solar direct and diffuse properties (Holben et al., 1998; Petrenko et al., 2012). The 21 stations that have data for the 2010-2020 period were used in this analysis for validating satellite measurements of AOD. AOD is measured using the Beer-Lambert-Bouguer equation based on direct measurements of the solar radiance (Holben et al., 1998; Sherman et al., 2016). There are approximately 700 well-calibrated sun photometers around the globe that measure AOD at 0.34 to 1.64-micrometer wavelengths every 15 minutes at three different levels: (a) Level 1.0 (unscreened), (b) Level 1.5 (cloud screened), and (c) Level 2.0 (cloud screened and quality assured). Level 2.0 data are characterized by very low uncertainty (0.01-0.02), and they are widely used to validate satellite-derived AOD measurements (McPhetres & Aggarwal, 2018), and this is the level used in this study.

### **MODIS MAIAC**

MODIS, a part of the A-Train constellation, has a 2330 km-wide swath and provides near-daily global coverage by every single instrument (Terra/Aqua). Terra MODIS crosses the equator at 10:30 (local time), and Aqua MODIS crosses the equator at 13:30 (local time), and for both satellites, the retrieval algorithms have been updated over time. Three operation algorithms: DT, DT, and MAIAC are currently being used to measure AOD (Mhawish et al., 2019).

MAIAC provides AOD data at a 1km-horizontal resolution by processing deriving column water vapor, cloud masking, and retrieval of aerosol parameters over land (Lyapustin et al., 2018). It takes advantage of time-series analysis of pixel groups in fixed  $25 \times 25$  km<sup>2</sup> blocks, helping it separate surface properties that are relatively static in a short duration compared to

aerosol and clouds, which change rapidly over time (Lyapustin et al., 2011a, 2011b, 2018). MAIAC took the MODIS Top-Of-Atmosphere (TOA) L1B reflectance and grid it to a 1-km resolution. A sliding window approach was used to gain 5 days (over poles) and 16 days (over equator) days of MODIS radiance observations. Relatively low AOD conditions (e.g., AOD value less than 0.5 globally) are required for the retrieval of the surface Bidirectional Reflectance Distribution Function (BRDF). A stable surface condition for 5-16 days is optimal for BRDF retrieval. MAIAC can identify most of the aerosol emission sources, including high-intensity plumes, at 1-km resolution, as it combines cloud masking with the detection of absorbing aerosols (smoke or dust) (Mhawish et al., 2019). The Minimum reflectance method is used to characterize the surface reflectance spectral ratios (SRC) 0.47/2.13 and 0.47/0.55, the center of the MAIAC aerosol retrieval process (Lyapustin et al., 2018). The SRC is obtained by using four or more days of low AOD and inverting available measurements in  $25 \times 25$  km<sup>2</sup> blocks (Lyapustin et al., 2011). After obtaining the SRC, AOD is measured from the last MODIS measurement (Lyapustin et al., 2012). The MAIAC algorithm uses different band combinations (0.47, 0.55, 0.65, and 2.13  $\mu\text{m}$ ) based on the surface brightness to detect the aerosol type. Spatio-statistical filtering detects residual cloud and improves the overall quality of the AOD values at 1-km resolution. Uncertainty of the surface spectral BRDF is used to characterize AOD retrieval error, although it is not currently reported (Hu et al., 2014).

MAIAC uses geographically prescribed aerosol models, based on the aerosol climatology obtained from AERONET. The current MAIAC-retrieved aerosol properties do not account for seasonal variations, and aerosol models are static, making it one of the limitations of the MAIAC C6 aerosol product (Mhawish et al., 2019). In the current study, AODs with the highest quality at 0.47 and 0.55  $\mu\text{m}$  were used.

## MODIS DT and DB

The MODIS DT algorithm uses a statistical relationship between the visible bands at 0.47 and 0.65  $\mu\text{m}$  and the shortwave infrared band at 2.12  $\mu\text{m}$  to retrieve aerosol over dark vegetated land surfaces (Levy et al., 2013). Dark pixels with a TOA reflectance between 0.01 and 0.25 in the 2.12  $\mu\text{m}$  are selected in the DT algorithm for the AOD measurement process. After eliminating the cloud, water, snow/ice, and other bright pixels, the selected pixels are organized in a 20x20 (pixel) array. Generally, the algorithm drops the brightest 50% and darkest 20% pixels in the 0.65  $\mu\text{m}$  channel. The average of the remaining pixels is used to determine the AOD, and the pixel number determines the retrieval quality. Overland, the DT algorithm uses three options: low-, moderate- and high-absorbing fine aerosol models plus one coarse aerosol model. The C6 DT product provides AOD measurements at two spatial resolutions 10-km and 3-km by using a similar algorithm; the main difference is created based on the pixel selection process for retrieving AOD at the 10 km and 3 km (Mhawish et al., 2019).

In contrast to the DT algorithm, the DB algorithm measures AOD over bright surfaces at 10-km resolution. Based on the surface type, it uses the 0.412 and/or 0.47/0.65  $\mu\text{m}$  wavelengths (Hsu et al., 2013; Tao et al., 2017). DB uses the static seasonal database of spectral surface reflectance, view geometry, and previous MODIS measurements to retrieve AOD over bright surfaces; and for vegetated surfaces, the DB algorithm uses statistical spectral ratios similar to the DT algorithm. Based on the geographical location and season, the DB algorithm selects the aerosol model; after removing residual clouds and filtering data for bad quality, the DB algorithm measures AOD at 1-km and aggregates it to 10-km (Bilal et al., 2015).

## **MODIS Active Fire Product**

MODIS Thermal Anomalies/Fire products use the radiance from MODIS 4 and 11  $\mu\text{m}$  riances to calculate the thermal anomaly. Fire detection strategy relies on adequate fire strength, relative background surface temperature, and sunlight reflection. The Level 2 MODIS active fire product has 1 km spatial resolution and is detected by Terra (MOD14) and Aqua (MYD14) satellites (Yin et al., 2019). Up to four observations can be made per day, depending on the cloud cover. Terra (MOD14) passes the equator at 10:30, and 22:30 (local time) (UTCC8); while Aqua (MYD14) passes the equator at 01:30 and 13:30 (local time). In collection 6, MODIS active fire product radiance sensed at mid-infrared wavelength (3.9  $\mu\text{m}$ ) channel, and the observed radiance is used to calculate Fire Radiative Power (FRP). MOD14/MYD14 products provide FRP estimation and its corresponding confidence level. Using the nearest neighbor resampling technique, the MOD14 and MYD14 are re-projected to a 1-km MODIS sinusoidal equal area to calculate the fire radiative energy (McCarley et al., 2020). In this study, The MYD14A1 V6 dataset from 2010 to 2020 was used to identify fire events.

## **Interpolation of AOD from AERONET**

MODIS retrieves AOD values at 470 and 550 nm wavelengths. The AERONET sun photometers measure AOD at six wavelengths (380, 440, 550, 675, 870, and 1020 nm). As there is no common wavelength between MODIS and AERONET, the AOD values have to be interpolated to a common wavelength to compare the results. In this study, AERONET AOD values were interpolated to 470 and 550 nm to compare with MODIS AOD values. The Angstrom exponent is typically used to determine the optical thickness of aerosol (Holben et al., 1998; Sherman et al., 2016). A larger particle size corresponds to lower values of the Angstrom

exponent, and high values of the Angstrom exponent resemble smaller particle size. The relationship between the Angstrom exponent and the wavelength can be expressed as follows:

$$\tau(\lambda) = \beta\lambda^{-\alpha} \quad \dots \dots \dots \text{eq. (1)}$$

$$\alpha = -\frac{\ln(\tau_1/\tau_2)}{\ln(\lambda_1/\lambda_2)} \quad \dots \dots \dots \text{eq. (2)}$$

Where,  $\alpha$  is the Angstrom exponent,  $\beta$  is the turbidity coefficient, equal to the AOD at 1  $\mu\text{m}$ , and  $\tau(\lambda)$  is the AOD at a given wavelength.

### **Sentinel-5 Precursor**

Sentinel-5 Precursor (Sentinel-5P) satellite onboarded the TROPOspheric Monitoring Instrument (TROPOMI), a passive hyperspectral nadir-viewing imager, launched on 13 October 2017. Sentinel-5P crosses the equator at 13:30 (local time) and has a repeat cycle of 17 days. It is a near-polar orbiting sun-synchronous satellite with a swath width of approximately 2,600 km and an along-track resolution of 7 km, which helps to capture the globe daily. In this study, NRTI/L3\_AER\_AI products were used. Based on Rayleigh scattering in the Ultraviolet (UV) spectral range for a pair of wavelengths, the NRTI/L3\_AER\_AI provides near real-time high-resolution imagery of the Absorbing Aerosol Index (AAI), which is then compared with a pair of measurements at the 354 and 388 nm wavelengths. When there are UV-absorbing aerosols like dust and smoke in the atmosphere, the AAI value is positive. This makes it ideal for monitoring the development of episodic aerosol plumes from wildfires. Sentinel-5P also provides near real-time high-resolution imagery of NO<sub>2</sub> concentrations (NRTI/L3\_NO2), as the substance enters the atmosphere mainly through fossil fuel combustion and biomass burning. An algorithm, used in OMI for the DOMINO-2 product, EU QA4ECV NO<sub>2</sub>, and later adapted for TROPOMI, is applied to process TROPOMI NO<sub>2</sub> data (Verhoelst et al., 2021). A 3-dimensional global TM5-MP

chemistry transport model is applied in the retrieval process of NO<sub>2</sub> measurements. This study used the NRTI/L3 products for monitoring the AAI and NO<sub>2</sub> concentrations during wildfires, both with a 1-km resolution.

## **Big Data Analyses**

The latest advancements in Earth Observing (EO) satellites have made it possible to acquire data on a global scale with higher spatial and variant temporal resolutions. The increasing number of sensors has made it difficult to process this huge amount of satellite data on modern computing and analysis infrastructures, as it requires a lot of storage and processing power. GEE is a computing platform that uses a dedicated High-Performance Computing (HPC) infrastructure for analyzing and visualizing petabytes of geospatial datasets acquired by EO satellites (Ravanelli et al., 2018). Through GEE, anyone can have access to more than 30 years of publicly free-data archives, including scientific datasets and historical images from many different satellites. GEE has the potential to overcome the limitations related to downloading, storing, and processing large-scale remotely-sensed data, referred to as Geo Big Data (Gorelick et al., 2017). This study employed GEE to calculate the monthly average AOD and the thermal anomaly of 11 years from daily MODIS MCD19A2.006 products and MODIS MYD14A1 V6 products, respectively, as well as to collect and process Sentinel-5P AAI and NO<sub>2</sub> data.

For the validation of the spatially and temporally MODIS AOD data, MODIS AOD and AERONET AOD were plotted against each other as shown in Figure 3.2. Linear regression was applied to calculate the slope and intercept. Uncertainty of the analysis was determined by the RMB, the RMSE and the MAE. According to eq. (3), if the RMB value is greater than 1, then the retrieved values are overestimated; and if the value is less than 1, then the opposite is true.

$$RMB = (\overline{AOD}_{(MODIS)} / \overline{AOD}_{(AERONET)}) \dots \text{eq. (3)}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (AOD_{MODIS\ i} - AOD_{AERONET\ i})^2} \quad \dots \dots \dots \text{eq. (4)}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |AOD_{MODIS\ i} - AOD_{AERONET\ i}| \quad \dots \text{eq. (5)}$$

## RESULTS AND DISCUSSIONS

Monthly average AOD data of 11 years from MODIS MAICA and AERONET were plotted against each other to determine the correlation (Figure 3.2). The average monthly MODIS MAIAC at 470 nm tended to overestimate the AOD than the monthly average ground based AERONET AOD value at 470 nm by approximately 19%. Overall, the data shows a good correlation with a 0.80 correlation coefficient, and a 0.65 adjusted R-squared, a 0.03 MAE, and a 0.04 RMSE. Better performance is observed in the fire season from MODIS MAICA measurements. The adjusted R-squared increased from 0.65 to 0.70 as well as the correlation coefficient that rose to 0.84, but the mean overestimation remained the same 19% (Figure 3.2 top). Also, the RMSE and MAE increased to 0.05 and 0.04, respectively. The same analysis was conducted at 550 nm to determine which wavelength measurements were more suitable. At 550 nm, the monthly average MODIS MAIAC measurements and AERONET measurements revealed a good correlation. Compared to the measurements at 470 nm, the correlation coefficient decreased slightly from 0.80 to 0.78, and the adjusted R-squared decreased from 0.65 to 0.60. MODIS also showed overestimation of AOD measurements at 550 nm when compared to AERONET AOD at 550 nm, but the overestimation was only by 4%. The RMSE and MAE at 550 nm, showed lower values than at 470 nm. The RMSE and MAE values were 0.03 and 0.02, respectively.

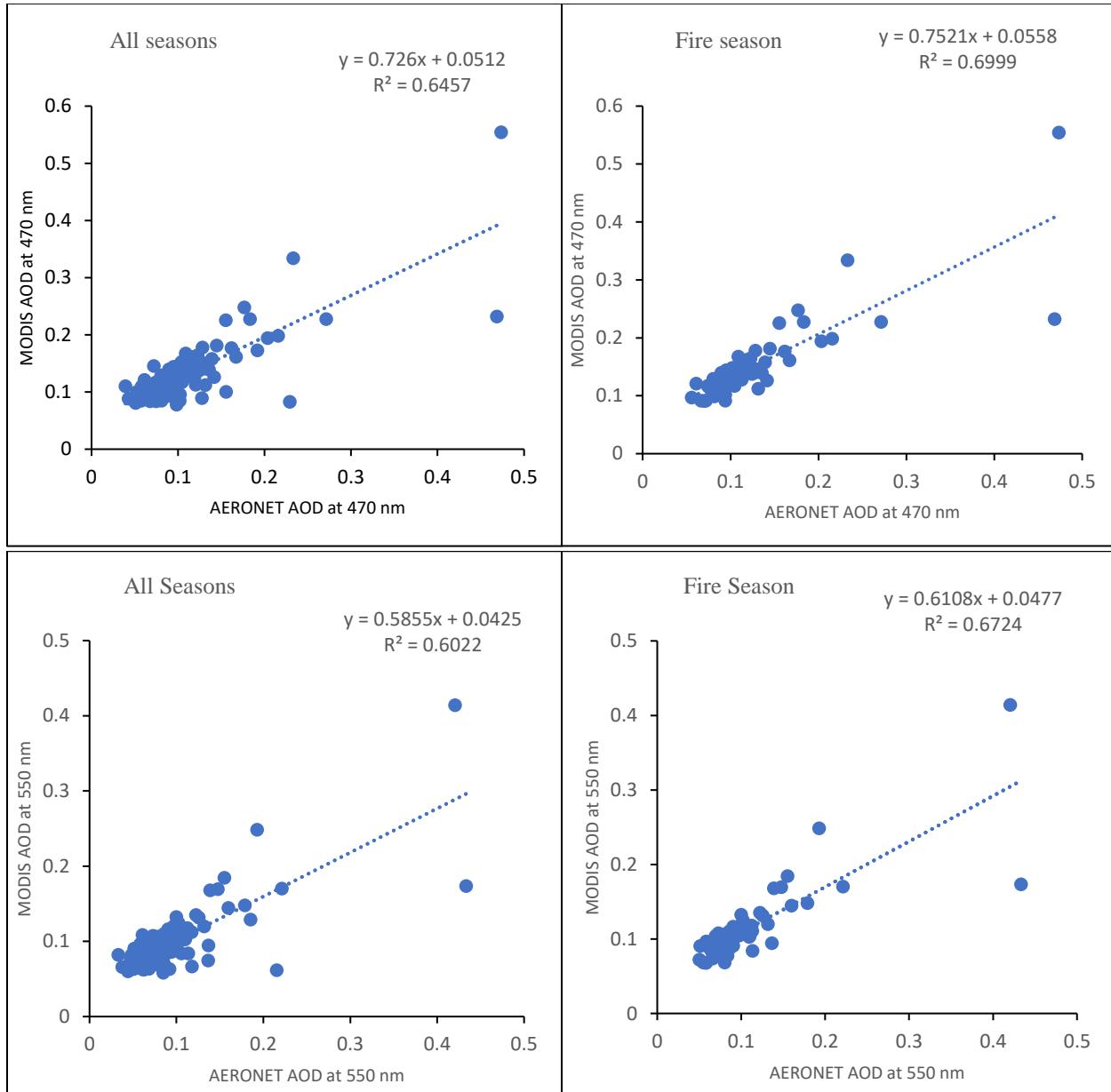


Figure 3.2: Scatter plots of MODIS and AERONET at 470 nm (top) and 550 nm (bottom) wavelengths for all seasons and the fire season.

Like the results at 470 nm, a better correlation was found between the fire season measurements at 550 nm. The correlation coefficient increased from 0.78 to 0.82 and the adjusted R-squared rose from 0.60 to 0.67 (Figure 3.2 bottom). Compared to the measurements at 470 nm, the RMSE and MAE were less for measurements at 550 nm wavelength, reading 0.04

and 0.02, respectively. The scatter plots (Figure 3.2) show that the AOD measurements over 0.4 have the greatest deviation. This deviation is observed in both all season and fire season measurements, and these values actually represent the 2020 fire events (August - September). The 2020 wildfire season in California was a record-setting event. According to the California Department of Forestry and Fire protection, it was the largest wildfire season since pre-1800. The AERONET stations are not evenly distributed throughout California, as a result a deviation between MODIS and AERONET stations is observed in these two months' measurements. The AERONET stations were able to pick up the regional initial increase of AOD at the beginning of fire events, while the broad fire impact was captured at a later time by MODIS.

These results demonstrate the capabilities of MODIS MAIAC in estimating AOD for large areas over a long time period effectively. During the fire season, MODIS AOD measurements were more similar to the ground based AERONET measurements of AOD at both wavelengths. As the AOD values varied over space and time, subtle AOD fluctuation was difficult to capture for the satellite, but during wildfire events, a large amount of smoke and plumes entered the atmosphere, an event easily identified by the satellite.

### **Seasonal Trends of AOD**

Monthly average AOD values derived from MODIS and AERONET at 470 nm showed an increasing trend. The AOD values ranged from 0.04 to 0.47 for AERONET measurements and MODIS measurements ranged from 0.08 to 0.55 at 470 nm. In some fire events, the average AOD value reached more than 0.2 for both MODIS and AERONET. For instance, the recorded AOD value was above 0.45 during the 2020 fire events. The trend lines show an increasing pattern for both MODIS and AERONET AOD, and it is expected that the current trend will exceed the AOD value of 0.15 soon. During the fire season, AOD values were much higher than

during usual times, ranging from 0.5 to 0.47 and from 0.1 to 0.55 for MODIS and AERONET at 470 nm, respectively. The AOD values in the fire season showed a rising pattern with a projected average value of 0.2 (Figure 3.3 top). Like the measurements at 470 nm, the value of AOD at 550 nm also showed a growing trend. For AERONET and MODIS-derived AOD values at 550 nm, the data ranged from 0.03 to 0.43 and from 0.05 to 0.41, respectively (Figure 3.3 bottom). The expected average value was approaching 0.15 for both MODIS and AERONET-based measurements.

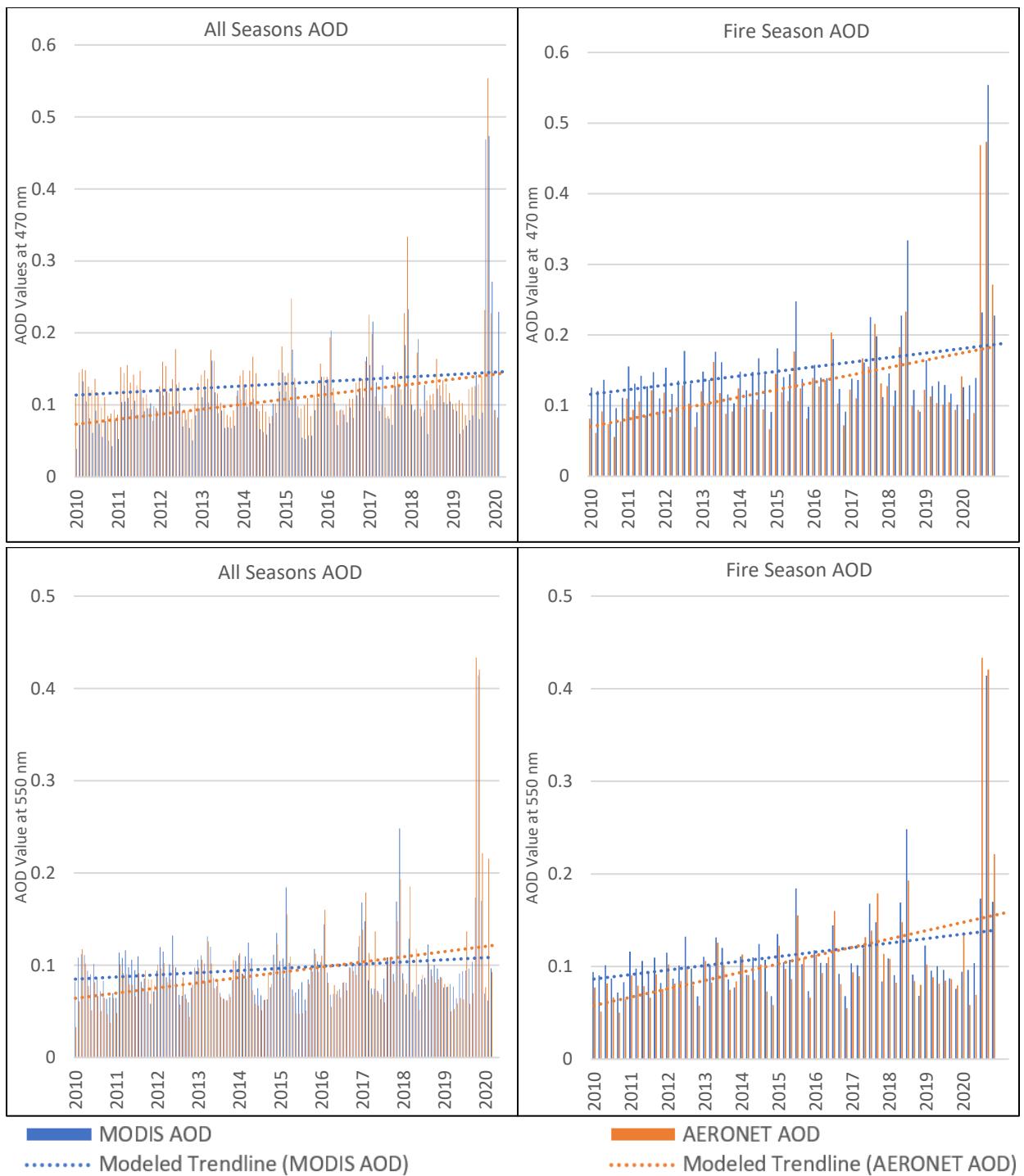


Figure 3.3: Monthly trends of AOD derived from MODIS and AERONET at 470 nm (top) and 550 nm (bottom) wavelengths.

As seen from both instruments, the trend line showed that the AOD will pass the 0.15 mark soon. Measurements at 470 nm and 550 nm from both instruments showed that the

wildfires will increase the AOD value in the atmosphere. The recent increase in fire events is causing more smoke and plumes to enter the atmosphere, and thus increasing the aerosol particles. This also suggests that major fire events can significantly increase up to 3 times the average AOD values.

### **MODIS Thermal Anomaly**

MODIS thermal anomaly had been used to discover previous major fire events. These thermal anomalies were plotted on top of the MODIS AODs to determine the regional impact of wildfires on air quality. In all major fire events, we can see a visible change in the AOD concentration in the atmosphere. MODIS Thermal Anomalies/Fire Version 6 products (MYD14A1.006: Aqua) were able to capture fire events, and the MCD19A2 V6 data product with 1-km resolution showed the visual change in the AOD at a regional level.

In August 2012, northern California experienced heavy multiple wildfire events: the Bagley fire had burned 11,083 acres, the Chips fire had burned 63,100 acres, the Fort Complex fire burned 6,683 acres, the North Pass fire burned 17,820 acres, and the Ponderosa fire had burned 28,089 acres. Thermal anomalies from MODIS were detected in several areas in northern California. This burning caused a higher AOD concentration in the atmosphere, as was visible from the MODIS data. These fire events caused the raise of AOD value in northern California up to 1.46 at 470 nm and 1.83 at 550 nm wavelengths. After the fire events, the AOD values were still higher in northern California compared to other regions. In August 2013, the Rim fire in Tuolumne caused 257,314 acres of burned areas, increasing the AOD level. The Corral Complex fire was also detected by the MODIS thermal anomaly and caused 12,503 acres of burned areas, raising the AOD values in northern California to more than 2.3 at 470 and 550 nm wavelengths (Figure 3.4)



Figure 3.4: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2012 and 2013 retrieved at 470 nm (a, d, c, g, h, and i) and 550 nm (d, e, f, j, k, and l) wavelengths.

The Happy Complex fire in August 2014 in Klamath National Forest in northern California caused 134,056 acres of burned areas with a damage cost of \$86.7 million. The AOD values at 470 and 550 nm wavelengths were over 1.9 for two months in northwestern California. There were multiple fire events in northern California in August 2015. Among them, the Gasquet complex, Nickowitz, Fork Complex, Mad River Complex were devastating. These fires caused more than 147,000 acres of burned areas and caused the AOD at 470 and 550 nm to raise to more than 2.8. The Rough fire in eastern California also caused serious damage, including burned areas of 151,623 acres, and like the fires in northern California, this fire increased the AOD values to over 2.0 at 470 nm and 3.0 at 550 nm in eastern California (Figure 3.5).

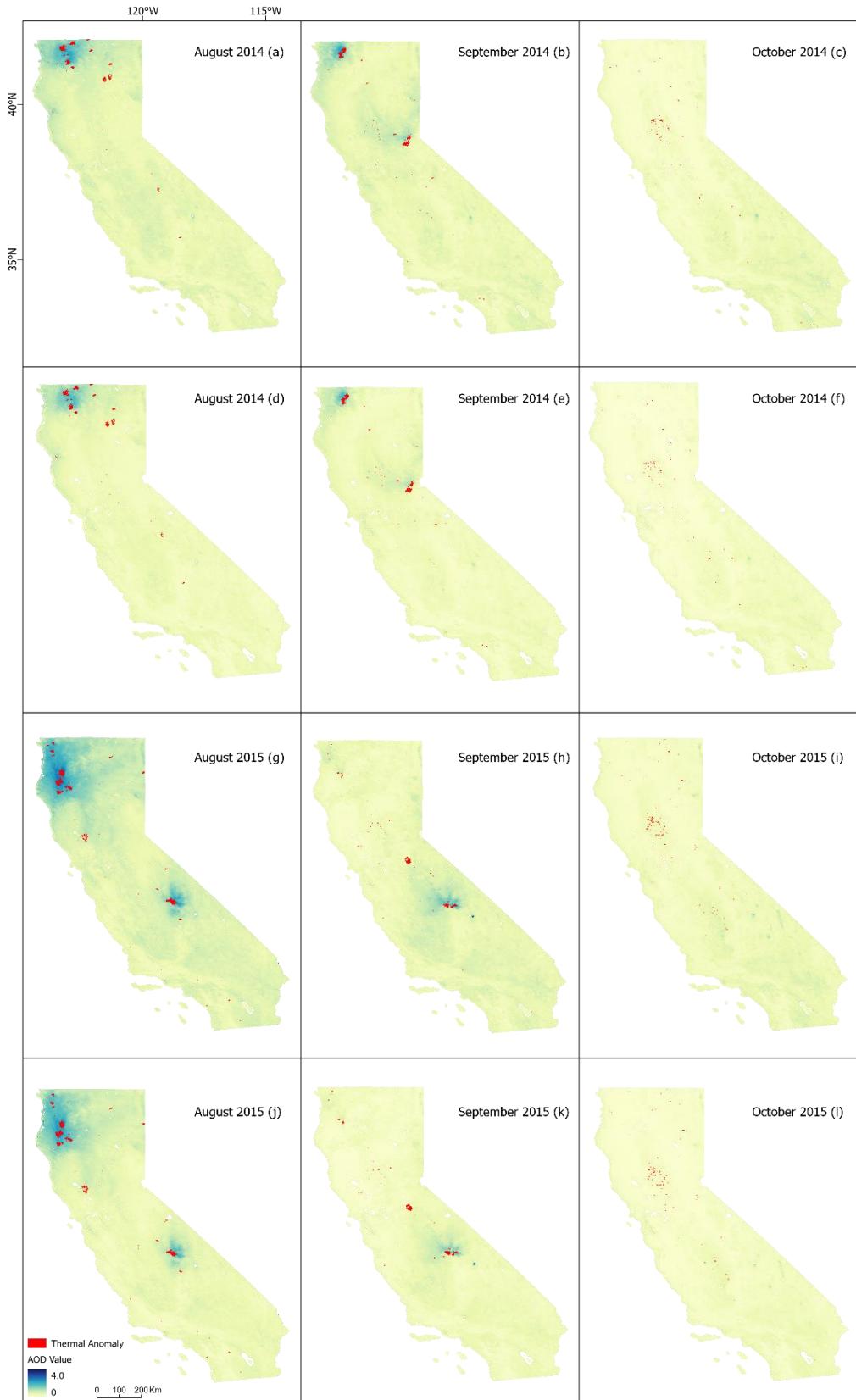


Figure 3.5: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2014 and 2015 retrieved at 470 nm (a, d, c, g, h, and i) and 550 nm (d, e, f, j, k, and l) wavelengths.

Fire events continued in 2016. The Rey Fire, Chimney Fire, and Blue Cut Fire occurred near the western coast of Santa Barbara and San Bernardino in August. In total, these fires caused more than 115,000 acres of burned areas. During these fire events, the AOD values at 470 and 550 nm wavelengths in the atmosphere were more than 2.0. The Miller Complex fire occurred, 17 miles east of Cave Junction, Oregon, in August 2017, burning 39,715 acres of land. Another notable fire was the Helena Fork Fire that occurred near Junction City, located in Trinity County, and burned 21,846 acres of land. The Eclipse Complex fire near the Klamath Nation Forest burned more than 78,698 acres of land. These fires increased the AOD values in northern California significantly. The Nuns, Adobe, Norrbom, Pressley, Patrick, and Oakmont fires near Napa and Sonoma County burned 56,556 acres of land and raised the AOD values to over 3.0 near the western coast of California (Figure 3.6).



Figure 3.6: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2016 and 2017 retrieved at 470 nm (a, d, c, g, h, and i) and 550 nm (d, e, f, j, k, and l) wavelengths.

The Mendocino Complex Fire of 2018 in northern California lasted more than three months and burned 459,123 acres of land, at an estimated loss of more than \$257 million. The Stone Fire happened 10 miles southwest from Canby, Big Canyon, and burned 39,387 acres of land. For two months, northern California experienced a higher AOD value, more than 2.3 (Figure 3.7 a, b, c, d, e, and f). Compared to other years, California experienced a significant reduction of fire events in 2019. The major fire events included: Kincade Fire in Sonoma County (burned 77,758 acres), Tucker Fire in the Modoc National Forest (burned 14,217 acres), and Walker Fire in the Plumas National Forest (burned 54,612 acres). Then, California encountered a disastrous fire season in 2020. The Creek Fire in September burned more than 379,895 acres with an estimated loss of \$193 million. The burning emitted a huge amount of smoke, which crossed the Atlantic Ocean and impacted the weather of Europe. The AOD values at 470 and 550 nm wavelengths were over 2.9 (Figure 3.7 g, h, I, j, k, and l).

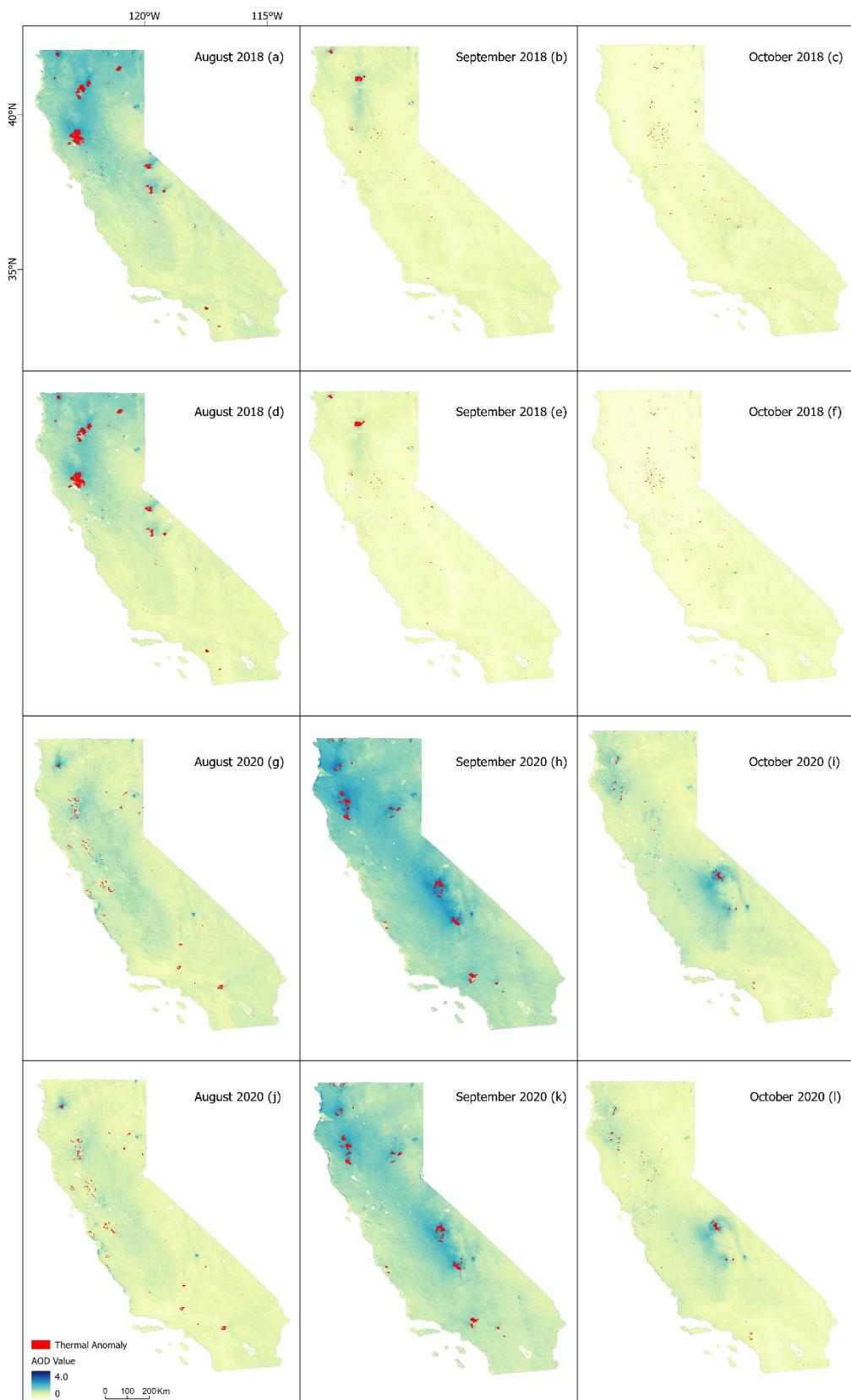


Figure 3.7: MODIS thermal anomaly and MODIS MAIAC AOD concentration for 2018 and 2020 retrieved at 470 nm (a, d, c, g, h, and i) and 550 nm (d, e, f, j, k, and l) wavelengths.

## Aerosol Index and NO<sub>2</sub> Concentration from Sentinel-5P

Sentinel-5P data along with AOD measurements were used in this study for investigating the NO<sub>2</sub> change. Sentinel-5P can measure atmospheric concentration of methane (CH<sub>4</sub>), carbon dioxide (CO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), and Nitrogen dioxide (NO<sub>2</sub>), allowing this study to investigate the impact of 2020 wildfires on the atmospheric Aerosol and NO<sub>2</sub> change (Figures 3.8 and 3.9). The analysis showed a significant rise in the NO<sub>2</sub> concentration as well as in the aerosol index near the 2020 wildfires. Positive values of AAI near the fire events indicated significant amount of smoke and dust in the atmosphere. Due to air circulation, the higher concentrated NO<sub>2</sub> tended to spread out across nearby areas.

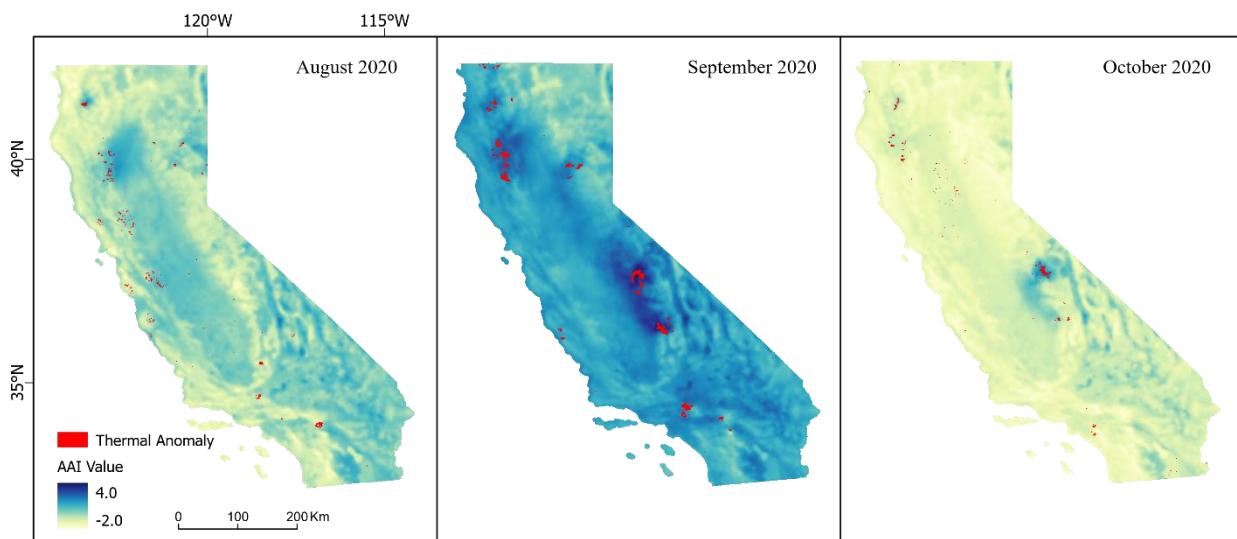


Figure 3.8: MODIS thermal anomaly and Sentinel-5P Aerosol Absorbing Index (AAI) for 2020 wildfires.

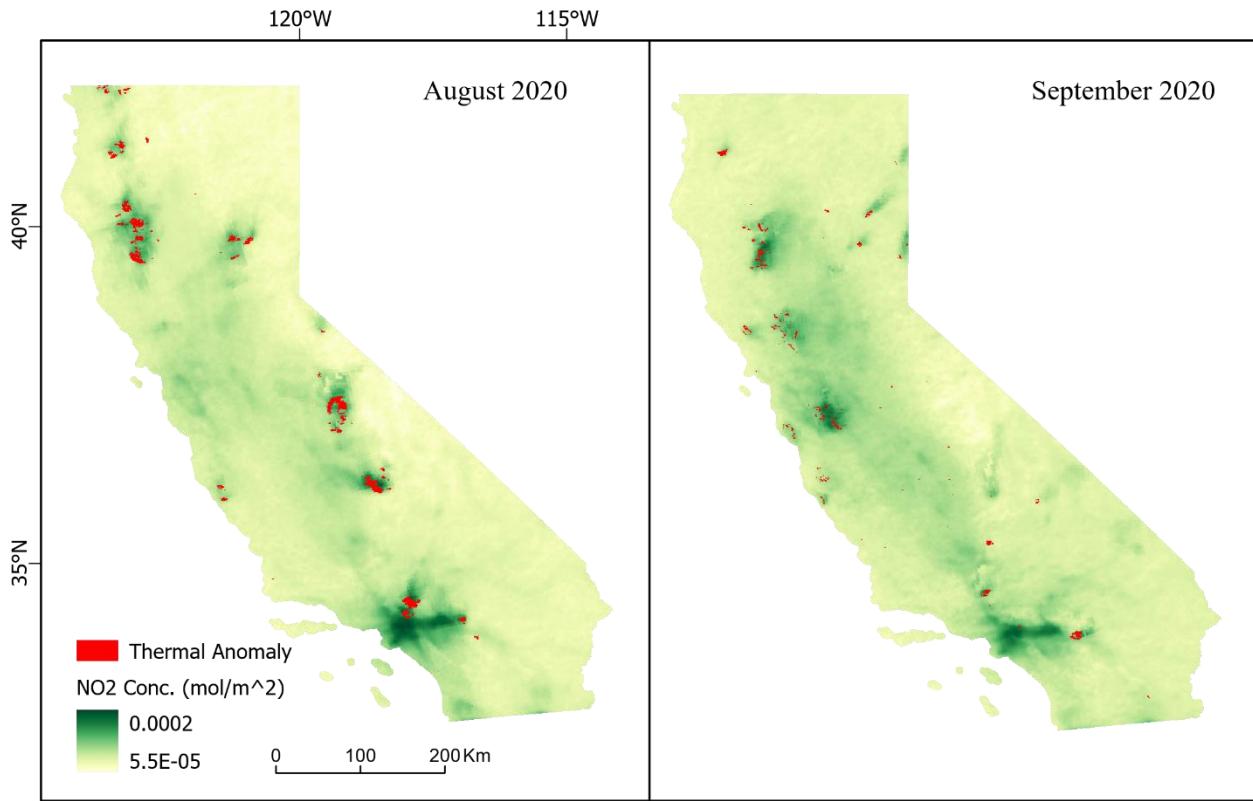


Figure 3.9: MODIS thermal anomaly and Sentinel-5P NO<sub>2</sub> concentration (mol/m<sup>2</sup>) for 2020 wildfires.

## CONCLUSIONS

Wildfires have multidimensional impacts on the ecosystem and human development.

Among many detrimental impacts of wildfires, air pollution is the one emphasized in this study.

Due to the complex dynamics of air movement, air pollution caused by wildfires is studied very limitedly; furthermore, to investigate air pollution, a long-term and large-scale investigation is needed. With the advancements in satellite imagery and cloud computing, spatio-temporal trends of air pollution due to wildfires can be analyzed effectively. This study used the MODIS MAIAC 1-km resolution of daily satellite images from 2010-2020 wildfires in California to retrieve the AOD at 470 and 550 nm wavelengths to determine the monthly trends of AOD and to examine the influence of wildfires on the neighboring areas. Measurements from MODIS were validated using ground based AERONET measurements. MODIS thermal anomaly data were overlaid

with MODIS AOD data to visualize the atmospheric effects of wildfires. Overall, the satellite-retrieved AOD showed a good correlation with AERONET data acquired over California for the study period. The correlation coefficients are 0.80 and 0.78 for MODIS-derived AOD at 470 and 550 nm wavelengths, respectively. During the fire seasons (May-October), MODIS AOD measurements showed a better performance with increased correlation coefficient to 0.84 and 0.82 at 470 and 550 nm wavelengths, respectively. Data acquired by the recently launched Sentinel-5P were then utilized to determine the effects of the 2020 wildfires on the atmospheric NO<sub>2</sub>. It was ultimately determined that wildfires significantly increased the NO<sub>2</sub> concentration over a broad area in California. This research demonstrates the capability of Big Data analyses for determining air pollution trends due to wildfires, and the results can improve understanding of the wildfires impacts on air quality and can assist in the mitigation programs.

## ACKNOWLEDGEMENTS

This study was partially funded by USGS and ArkansasView grant #GR908148UAF awarded to Dr. Mohamed Aly. Thanks are also due to USGS, NASA, the European Space Agency, Google Earth Engine, and the United States Census Bureau for providing necessary data for conducting this research.

## REFERENCES

- Akther, M. S., & Hassan, Q. K. (2011). Remote sensing-based assessment of fire danger conditions over boreal forest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(4), 992–999.
- Andrews, E., Ogren, J., Bonasoni, P., Marinoni, A., Cuevas, E., Rodríguez, S., Sun, J., Jaffe, D., Fischer, E., & Baltensperger, U. (2011). Climatology of aerosol radiative properties in the free troposphere. *Atmospheric Research*, 102(4), 365–393.
- Barbero, R., Abatzoglou, J., Steel, E., & Larkin, N. K. (2014). Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environmental Research Letters*, 9(12), 124009.

- Bilal, M., Nichol, J. E., & Nazeer, M. (2015). Validation of Aqua-MODIS C051 and C006 operational aerosol products using AERONET measurements over Pakistan. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 2074–2080.
- Bright, J. M., & Gueymard, C. A. (2019). Climate-specific and global validation of MODIS Aqua and Terra aerosol optical depth at 452 AERONET stations. *Solar Energy*, 183, 594–605.
- Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophysical Research Letters*, 41(8), 2928–2933.
- Dennison, P. E., Moritz, M. A., & Taylor, R. S. (2008). Evaluating predictive models of critical live fuel moisture in the Santa Monica Mountains, California. *International Journal of Wildland Fire*, 17(1), 18–27.
- Fire, C. (2018). *Top 20 deadliest California wildfires*.
- Freeborn, P. H., Wooster, M. J., & Roberts, G. (2011). Addressing the spatiotemporal sampling design of MODIS to provide estimates of the fire radiative energy emitted from Africa. *Remote Sensing of Environment*, 115(2), 475–489.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.
- Holben, B. N., Eck, T. F., Slutsker, I. al, Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y., & Nakajima, T. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, 66(1), 1–16.
- Hsu, N., Jeong, M., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., Huang, J., & Tsay, S. (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, 118(16), 9296–9315.
- Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., & Puttaswamy, S. J. (2014). Estimating ground-level PM<sub>2.5</sub> concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment*, 140, 220–232.
- Huang, X., Li, M., Li, J., & Song, Y. (2012). A high-resolution emission inventory of crop burning in fields in China based on MODIS Thermal Anomalies/Fire products. *Atmospheric Environment*, 50, 9–15.
- Jianping, H., Yujie, W., Tianhe, W., & Yuhong, Y. (2006). Dusty cloud radiative forcing derived from satellite data for middle latitude regions of East Asia. *Progress in Natural Science*, 16(10), 1084–1089.

- Jin, Y., Goulden, M. L., Faivre, N., Veraverbeke, S., Sun, F., Hall, A., Hand, M. S., Hook, S., & Randerson, J. T. (2015). Identification of two distinct fire regimes in Southern California: Implications for economic impact and future change. *Environmental Research Letters*, 10(9), 094005.
- Jolly, W. M., Cochrane, M. A., Freeborn, P. H., Holden, Z. A., Brown, T. J., Williamson, G. J., & Bowman, D. M. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications*, 6(1), 1–11.
- Kahn, R. A., & Gaitley, B. J. (2015). An analysis of global aerosol type as retrieved by MISR. *Journal of Geophysical Research: Atmospheres*, 120(9), 4248–4281.
- Kaufman, Y. J., Tanré, D., & Boucher, O. (2002). A satellite view of aerosols in the climate system. *Nature*, 419(6903), 215–223.
- Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A. C., Schwartz, J., & Broday, D. M. (2015). Estimating daily PM<sub>2.5</sub> and PM<sub>10</sub> across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmospheric Environment*, 122, 409–416.
- Kodandapani, N., & Theme, W. (n.d.). *Amplified Fire Occurrences in Response to Drought and Vegetation Stress in the Western Ghats of India*.
- Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., & Hsu, N. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6(11), 2989–3034.
- Liu, Y., Stanturf, J., & Goodrick, S. (2010). Trends in global wildfire potential in a changing climate. *Forest Ecology and Management*, 259(4), 685–697.
- Lu, G., & Guo, X. (2012). Distribution and origin of aerosol and its transform relationship with CCN derived from the spring multi-aircraft measurements of Beijing Cloud Experiment (BCE). *Chinese Science Bulletin*, 57(19), 2460–2469.
- Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012). Discrimination of biomass burning smoke and clouds in MAIAC algorithm. *Atmospheric Chemistry and Physics*, 12(20), 9679–9686.
- Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., & Korkin, S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *Journal of Geophysical Research: Atmospheres*, 116(D3).
- Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11(10), 5741–5765.
- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., & Reid, J. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres*, 116(D3).

- McCarley, T. R., Hudak, A. T., Sparks, A. M., Vaillant, N. M., Meddens, A. J., Trader, L., Mauro, F., Kreitler, J., & Boschetti, L. (2020). Estimating wildfire fuel consumption with multitemporal airborne laser scanning data and demonstrating linkage with MODIS-derived fire radiative energy. *Remote Sensing of Environment*, 251, 112114.
- McPhetres, A., & Aggarwal, S. (2018). An evaluation of MODIS-retrieved aerosol optical depth over AERONET sites in Alaska. *Remote Sensing*, 10(9), 1384.
- Mhawish, A., Banerjee, T., Sorek-Hamer, M., Lyapustin, A., Broday, D. M., & Chatfield, R. (2019). Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia. *Remote Sensing of Environment*, 224, 12–28.
- Norris, R. M., & Webb, R. W. (1990). *Geology of California*.
- Petrenko, M., Ichoku, C., & Leptoukh, G. (2012). Multi-sensor aerosol products sampling system (MAPSS). *Atmospheric Measurement Techniques*, 5(5), 913–926.
- Qi, Y., Ge, J., & Huang, J. (2013). Spatial and temporal distribution of MODIS and MISR aerosol optical depth over northern China and comparison with AERONET. *Chinese Science Bulletin*, 58(20), 2497–2506.
- Ravanelli, R., Nascetti, A., Cirigliano, R. V., Di Rico, C., Leuzzi, G., Monti, P., & Crespi, M. (2018). Monitoring the impact of land cover change on surface urban heat island through Google Earth Engine: Proposal of a global methodology, first applications and problems. *Remote Sensing*, 10(9), 1488.
- Remer, L. A., Kaufman, Y., Tanré, D., Mattoo, S., Chu, D., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R., & Kleidman, R. (2005). The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, 62(4), 947–973.
- Rother, D., & De Sales, F. (2021). Impact of Wildfire on the Surface Energy Balance in Six California Case Studies. *Boundary-Layer Meteorology*, 178(1), 143–166.
- Sayer, A., Hsu, N., Bettenhausen, C., Jeong, M., & Meister, G. (2015). Effect of MODIS Terra radiometric calibration improvements on Collection 6 Deep Blue aerosol products: Validation and Terra/Aqua consistency. *Journal of Geophysical Research: Atmospheres*, 120(23), 12–157.
- Sayer, A. M., Hsu, N., Bettenhausen, C., & Jeong, M. (2013). Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. *Journal of Geophysical Research: Atmospheres*, 118(14), 7864–7872.
- Sherman, J. P., Gupta, P., Levy, R. C., & Sherman, P. J. (2016). An evaluation of MODIS-retrieved aerosol optical depth over a mountainous AERONET site in the southeastern US. *Aerosol and Air Quality Research*, 16(12), 3243–3255.

- Tao, M., Chen, L., Wang, Z., Wang, J., Che, H., Xu, X., Wang, W., Tao, J., Zhu, H., & Hou, C. (2017). Evaluation of MODIS Deep Blue aerosol algorithm in desert region of East Asia: Ground validation and intercomparison. *Journal of Geophysical Research: Atmospheres*, 122(19), 10–357.
- Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., & Levelt, P. (2007). Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview. *Journal of Geophysical Research: Atmospheres*, 112(D24).
- Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., Fjæraa, A. M., Granville, J., Niemeijer, S., & Cede, A. (2021). Ground-based validation of the Copernicus Sentinel-5p TROPOMI NO<sub>2</sub> measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmospheric Measurement Techniques*, 14(1), 481–510.
- Williams, A. P., & Abatzoglou, J. T. (2016). Recent advances and remaining uncertainties in resolving past and future climate effects on global fire activity. *Current Climate Change Reports*, 2(1), 1–14.
- Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., & Lettenmaier, D. P. (2019). Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future*, 7(8), 892–910.
- Winker, D., Pelon, J., Coakley Jr, J., Ackerman, S., Charlson, R., Colarco, P., Flamant, P., Fu, Q., Hoff, R., & Kittaka, C. (2010). The CALIPSO mission: A global 3D view of aerosols and clouds. *Bulletin of the American Meteorological Society*, 91(9), 1211–1230.
- Wooster, M. J., Roberts, G., Perry, G., & Kaufman, Y. (2005). Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research: Atmospheres*, 110(D24).
- Yin, L., Du, P., Zhang, M., Liu, M., Xu, T., & Song, Y. (2019). Estimation of emissions from biomass burning in China (2003–2017) based on MODIS fire radiative energy data. *Biogeosciences*, 16(7), 1629–1640.

## OVERALL CONCLUSIONS

Wildfires have become a growing concern across the world. Due to recent climate and landscape changes, the intensity and the frequency of fire events have increased drastically. Areas that were less susceptible in the past are now experiencing infrequent and extensive wildfires, and studies are showing that the trends of rising wildfire numbers are likely continue in the future. The effective and sustainable management of land and forest resources depends on preventive measures from natural disasters, and for this, there is no alternative but to identify hazardous areas. Even though Arkansas is not experiencing a tremendous number of wildfires, there are possibilities of the increasing threat in the future. Furthermore, Arkansas is highly dependent on its forest and agricultural resources, so it is vital to protect these resources from future threats. The first portion of the research was focused on identifying the wildfire-susceptible areas in Arkansas. There are different methods to analyze wildfires, such as GIS, a powerful tool to investigate complex phenomena like wildfire with a rich repository of spatial and temporal data. Different statistical methods were integrated with GIS, providing a means to combine natural and anthropogenic variables. Furthermore, ML has become extremely popular in the scientific community due to its superior computation capabilities. Recently, ML has been coupled with GIS, adding a new dimension in the field of GIS.

In the second chapter of this thesis, GIS-based multiple regression, and ML-based RF were utilized successfully to identify wildfire susceptible areas in Arkansas. First, OLS was applied to Oklahoma's datasets to identify the significant variables that were then used in investigating wildfires in Arkansas. After running OLS, GWR was employed on the same datasets to capture spatial variability. The adjusted R-squared values for OLS and GWR were 0.51 and 0.87, respectively. Results indicated that GWR performed better than OLS, and GWR

was better in incorporating the spatial variability. Results also suggested that a nonparametric method like RF would lead to better outcomes than multiple regression. Therefore, RF was utilized to identify the wildfire-susceptible areas in Arkansas. Statistical outputs of RF, such as the adjusted R-squared, were above 0.9 for training and validation, and the model was able to explain more than 90% of the variation. The output of the model suggested that the Ouachita National Forest and the Ozark Forest are highly susceptible to wildfires and the eastern side of the state is the least sensitive to wildfire. However, Northwest Arkansas and Southern Arkansas have moderate and low moderate susceptibility to wildfires. This study will contribute to the Arkansas Forestry Commission under the Arkansas Department of Agricultural and will help in the proper land and fire management policy.

Wildfires not only burn forests and destroy properties but also have a negative impact on the air quality. Smokes from wildfires can travel a great distance and affect other areas, even inducing different respiratory diseases. Monitoring air quality with satellite data is challenging as air is very dynamic; therefore, daily measurements of air quality are needed, but for traditional GIS uses, this can be a very tedious process as it requires downloading and processing terabytes of satellite data. To overcome this barrier, Google has introduced a cloud-based platform, GEE, to perform long-term satellite data analysis without downloading them to a local machine. This platform also uses the power of a supercomputer to perform complex analytical tasks, which can be very difficult to execute on a local machine.

The third chapter of this thesis explored the capability of GEE to monitor the air quality during 2010-2020 wildfires in California. California is well known for its large-scale wildfire events every year due to its climatic condition and landscape formation. In this study, MODIS products were extensively utilized to identify thermal anomalies of wildfires and AOD

measurements in order to determine the impact of those fires on air quality. Recently, a superior and improved MAIAC algorithm has been developed for MODIS AOD products to overcome the barrier of coarse resolution. Previously, the resolution for AOD products was 10-km, a very low retrieval accuracy, but with the advanced MAIAC products, the spatial resolution has been increased to 1-km. MODIS AOD retrieved values at 470 and 550 nm wavelengths using dark target (DT) ocean, DT land, and deep blue (DB) algorithms. Furthermore, the MODIS MYD14A1 V6 dataset provides fine-resolution fire composite data, typically used to detect fire incidences. AOD measurements from the satellite may have a greater spatial resolution, but could be misleading due to reflectance, a property influenced by cloud covers, which can affect the AOD measurements. For this, ground-based sun photometers, AERONET, were used to measure the AOD values. The newly launched Sentinel-5P measures other atmospheric parameters that can also explain the impact of wildfire on the atmosphere.

In this study, Sentinel-5P and MODIS AOD measurements and fire mask data were employed to determine the impact of wildfires on air quality over California from 2010 to 2020. The AOD measurements were cross-validated using the AERONET measurements at 470 and 550 nm wavelengths for the fire season (May-October) and for a whole year. AERONET and MODIS MAIAC measurements showed a good correlation. For the monthly average, the adjusted R-squared values were 0.65 and 0.60 at 470 nm and 550 nm, respectively. For the fire season, the adjusted R-squared values were approximately 0.70 and 0.67 at 470 nm and 550 nm, respectively. It was found that during major fire events, the AOD level increased three times compared to average. Furthermore, large-scale and intensive fire events were raising the average AOD values in the atmosphere. Sentinel-5P data showed that during a forest fire, NO<sub>2</sub> contraction increased significantly near the fire event. This analysis can help the California

Department of Forestry and Fire Protection (CALFIRE) take preventive measures and necessary actions to control air pollution from wildfires.