Wildfire dynamics from ECOSTRESS data and machine learning: The case of South-Eastern Australia's black summer

Ivone K Masara[1], Shakthi Bharathi Murugesan[1], Soe Win Myint[1], Yuanhui Zhu[2], and Joshua B Fisher[3]

[1]School of Geographical Sciences and Urban Planning, Arizona State University
[2]Arizona State University
[3]Chapman University

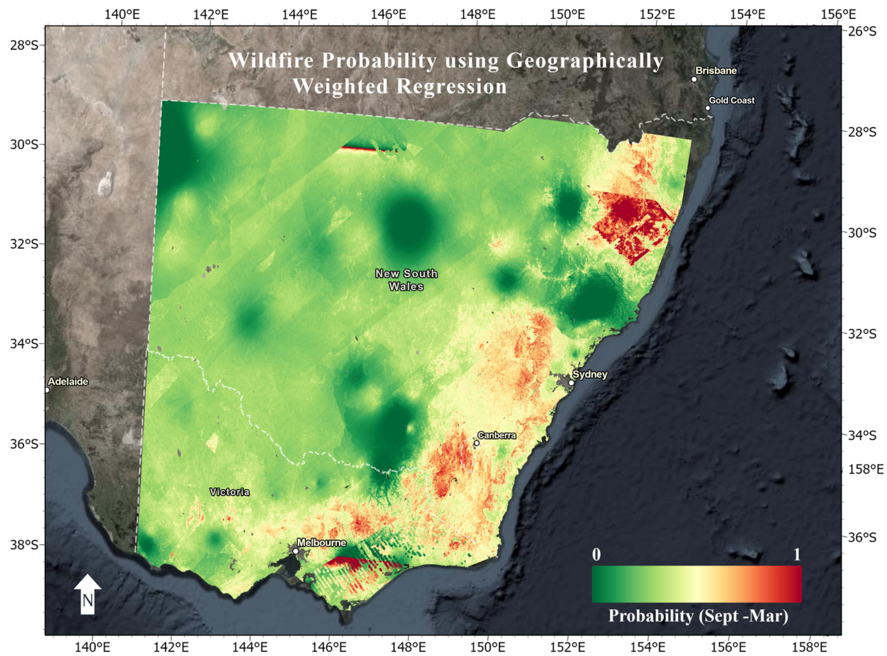
November 26, 2022

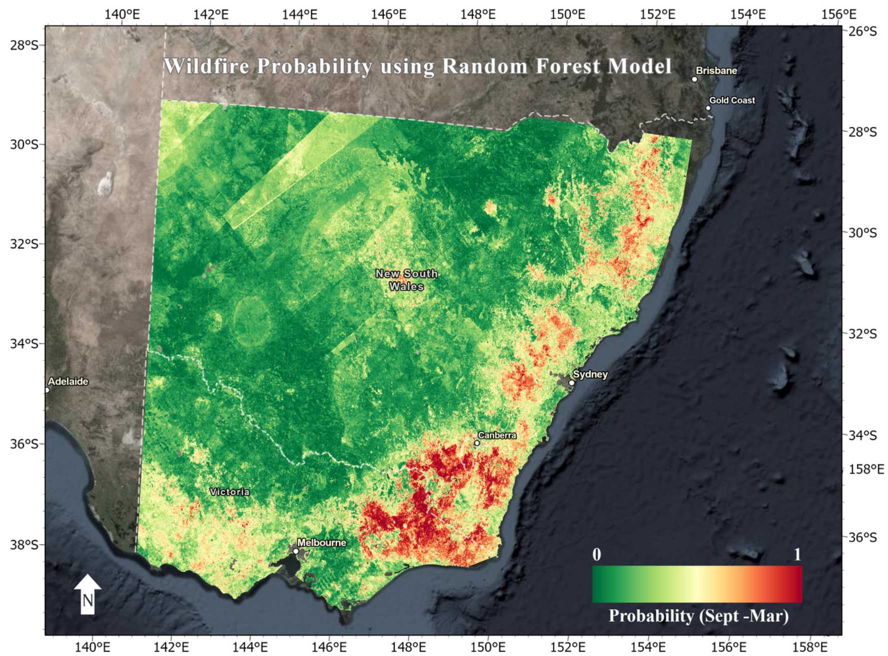
Abstract

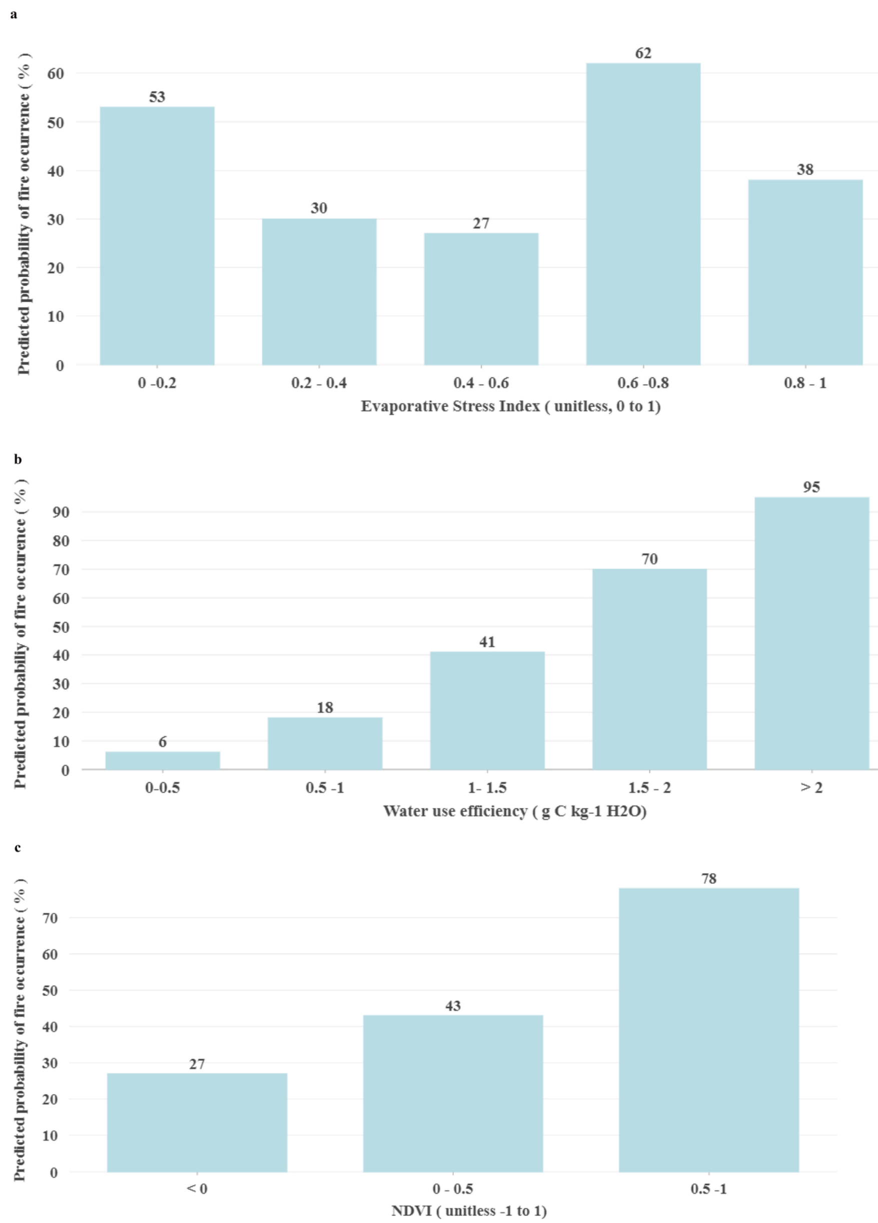
In 2019–20 Australia was devastated by the worst wildfires observed in decades. NASA's ECOsystem Space-borne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission, launched in 2018, captured many dynamics of the fires at high resolution, including ecosystem stress prior to the fires. We aimed to determine the predictive capacity of ECOSTRESS observations for fire occurrence and intensity in Southeast Australia. We found that ECOSTRESS data (evaporative stress index and water use efficiency) were highly predictive of fire dynamics (25-65% occurrence prediction accuracy for ESI; and, 40-95% occurrence prediction for WUE > 1 gCkg-1H2O alone, depending on their levels) with the ESI coefficient averaging approximately three times stronger than general topographic variables or meteorological variables. Our results, based on a logistic regression model, had an overall predictive accuracy of 83%, suggesting high potential of using ECOSTRESS data to project and examine fires in Australia and other similar regions of the world.

Hosted file

`essoar.10510596.1.docx` available at https://authorea.com/users/555759/articles/606187-wildfire-dynamics-from-ecostress-data-and-machine-learning-the-case-of-south-eastern-australia-s-black-summer


Wildfire Probability using Geographically Weighted Regression
Brisbane
Gold Coast
New South Wales
Adelaide
Sydney
Canberra
Victoria
Melbourne
N
0
1
Probability (Sept -Mar)


Wildfire Probability using Random Forest Model
Brisbane
Gold Coast
New South Wales
Adelaide
Sydney
Canberra
Victoria
Melbourne
N
0
1
Probability (Sept -Mar)


a
Predicted probability of fire occurence (%)
53
30
27
62
38
0 -0.2
0.2 - 0.4
0.4 - 0.6
0.6 -0.8
0.8 - 1
Evaporative Stress Index (unitless, 0 to 1)
b
Predicted probability of fire occurence (%)
6
18
41
70
95
0-0.5
0.5 -1
1- 1.5
1.5 - 2
> 2
Water use efficiency (g C kg-1 H2O)
c
Predicted probability of fire occurence (%)
27
43
78
< 0
0 - 0.5
0.5 -1
NDVI (unitless -1 to 1)

Wildfire dynamics from ECOSTRESS data and machine learning: The case of South-Eastern Australia's black summer

Shakthi B. Murugesan[1], Ivone K. Masara[1], Soe W. Myint[1], Yuanhai Zhu[1], Joshua B. Fisher[2]

[1]School of Geographical Sciences and Urban Planning, Arizona State University

[2] Schmid College of Science and Technology, Chapman University, 1 University Drive, Orange, CA, 92866, USA

Corresponding Author: Ivone Masara (**imasara@asu.edu**)

Key Points

- We tested the predictability of fire occurrence for the 2019-20 Australian wildfires.
- Logistic regression models captured the wildfire dynamics with a high accuracy.
- Evaporative stress index and water use efficiency were the most significant predictors.

Abstract

In 2019–20 Australia was devastated by the worst wildfires observed in decades. NASA's ECOsystem Space-borne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission, launched in 2018, captured many dynamics of the fires at high resolution, including ecosystem stress prior to the fires. We aimed to determine the predictive capacity of ECOSTRESS observations for fire occurrence and intensity in Southeast Australia. We found that ECOSTRESS data (evaporative stress index and water use efficiency) were highly predictive of fire dynamics (25-65% occurrence prediction accuracy for ESI; and, 40-95% occurrence prediction for WUE > 1 $gCkg^{-1}H_2O$ alone, depending on their levels) with the ESI coefficient averaging approximately three times stronger than general topographic variables or meteorological variables. Our results, based on a logistic regression model, had an overall predictive accuracy of 83%, suggesting high potential of using ECOSTRESS data to project and examine fires in Australia and other similar regions of the world.

Plain Language Summary

Wildfires have been on the increase all over the world in recent times. In particular, Australia's catastrophic wildfire event that occurred in 2019-2020 was one of the worst seen in this modern era. NASA's ECOSTRESS, launched in 2018, captures the variations in temperature of plants. It addresses how water availability affects the vegetation, drought estimation, and agricultural vulnerability. Using this high resolution ECOSTRESS data, we have found that the drought condition of south-eastern Australia is one of the main reasons behind the wildfire spreading during the 2019-20 season. This suggests that the ECOSTRESS data could potentially be used to examine and predict fire occurrence across different regions of the world.

Key Words

Wildfire; susceptibility; machine learning; ECOSTRESS; evaporative stress; water use efficiency; Australia

1. Introduction

Wildfires have been increasing in severity and intensity worldwide (Doerr et al., 2016). This has led to aggravating social, ecological and economic consequences in many communities (Alexander et al., 2020). Among the common sources of wildfires is human activity and natural causes such as lightning strikes (Lewis et al., 2015). Recent studies show that climate change is also making our environment increasingly vulnerable to devastating wildfires (Lim et al., 2019; Halofsky et al., 2020). These studies have indicated that climate change-induced environmental changes such as an increase in temperature, land cover change and precipitation variability are highly likely to alter the frequency and intensity of wildfires. Furthermore, an increase in the amount of greenhouse gases and other aerosols from wildfire emissions and changes in the surface reflectance produced by fires contribute to ongoing climate change which is expected to increase substantially in the future (Brown et al., 2018). This makes the development of reliable susceptibility models of wildfire danger necessary for assurance of public safety, natural resource management, and planning of risk management. The models could help identify areas with higher fire risk, that even with limited resources, authorities could choose to focus on monitoring specific areas (Whitburn et al., 2016).

Variables to be considered in the prediction of wildfires vary by region because of different influencing factors. These variables can be categorized into topography, vegetation, climate, and human activities (Ganteaume et al., 2013; Nami et al., 2018; Parisien et al., 2012). Topographic effects on wildfire (e.g., slope, aspect, and elevation) are primarily indirect (Jaafari et al., 2017, Parisien et al., 2012) by influencing the type of vegetation, local climate, and human accessibility (Jaafari et al., 2017; Nami et al., 2018). Climate variables (rainfall and temperature) exert direct and indirect influences on wildfire events (Jaafari et al., 2017; Nami et al., 2018; Parisien et al., 2012). Vegetation (land cover), on the other hand, affects wildfire and fire spread through fuel characteristics such as vegetation type, water availability and evapotranspiration, affecting the moisture in the plants and fuel load (Nami et al., 2018). As suggested by Lim et al., (2019), the inclusion of climate change is critical in wildfire prediction variables.

The robustness and sensitivity of models rely heavily on available data. Historically, wildfires have been collected mainly by post-fire field surveys, which are time-consuming and often lack ignition points (Lim et al., 2019). However, with the introduction of remote sensing methods and satellite monitoring systems, spatially comprehensive datasets are available on demand. This has helped researchers in quantifying climate, topographical and human factors towards the contribution of wildfires around the world. Fire data products, such as from the Moderate Resolution Imaging Spectroradiometer (MODIS) are currently freely available online, enabling access to timely information worldwide (Wulder et al., 2012).

In June 2018, NASA launched ECOSTRESS primarily to address how water availability affects key climate biomes around the world, drought estimation, and agricultural vulnerability (Fisher et al 2020). The instrument measures thermal infrared radiance at 5 spectral bands in the 8-12.5 m range, with approximately 70 by 70 meters spatial resolution on the ground. Data from ECOSTRESS show how ecosystems change with climate and create a crucial link between the water cycle and plant health, both natural and human inflicted. The products derived from ECOSTRESS data utilized in this study include four core products: L4_ESI_PT-JPL; L4_WUE; L2_LSTE, and L3_ET_PT-JPL. ESI is evaporative stress index, WUE is water use efficiency, LST is land surface temperature, and ET is evapotranspiration. L3 products are derived from L2 data, and L4 products are derived from L3 data.

In the bid to predict wildfire events, fire weather indices (FWI) were amongst the first probability mapping trials developed by scientists. It was commonly used to define an area's seasonal and long-time forest fire hazard. FWI are produced from environmental factors such as weather data (dry bulb temperature, humidity and wind speed, etc.) to calculate fire danger rating and fuel moisture content (McArthur 1966; Fosberg 1978; Srock et al., 2018). This led to the development of spatial prediction of wildfire susceptibility using Geographic Information Systems (GIS) and Remote Sensing (RS), implemented in different approaches, such as fuzzy logic and the analytical network process (ANP) (Tonini et al., 2020). However, the use of conventional parametric statistical modeling techniques, such as fuzzy logic, by the weighting of inputs may be problematic because of subjective ranking (Satir et al., 2016).

An alternative approach is to learn the complex nonlinear relationships associated with fire directly from observational and numerical modeling of data. This can be done using machine learning (ML) algorithms (Bui et al., 2018). ML algorithms use statistics to find patterns in massive amounts of data. Recently, ML algorithms such as Neural Network, Support Vector Machine, Random Forest (RF) and Logistic Regression (LR) classifiers have achieved reasonably reliable results in various natural hazard susceptibility mapping studies (Satir et al., 2016).

This study aimed to: (1) establish a geospatial database, including MODIS MCD64A1 fire product, digital elevation model (DEM), slope, aspect, ECOSTRESS data (i.e., evapotranspiration (ET), evaporative stress index (ESI), land surface temperature (LST), water use efficiency (WUE)), NDVI generated from Sentinel-2 data, and rainfall data; (2) quantify wildfire probability using geographically weighted regression (GWR), logistic regression (LR), and random forest (RF) algorithms; (3) determine the importance of explanatory variables; (4) assess model performance; and, (5) assess vulnerability of major cities.

2. Study Area

Our study focuses on the south-eastern region of Australia (Figure 1a). The climate over the region is temperate with December and January being the hottest months (Geoscience Australia, 2020). In recent years the south-eastern part has been experiencing lots of sporadic wildfires. However, the 2019-2020 bushfire season was unprecedented in intensity and devastation, hence it is widely known as **'Black Summer'**. Throughout the summer, multiple fires scorched large tracts of land in Victoria and New South Wales of Australia, resulting in 34 fatalities and huge losses of land and wildlife (Bushfires in Victoria - Research Guides, 2020). Fires were ignited in September 2019 and were contained by early March 2020. The state of New South Wales had the highest number of homes lost (2,439) followed by Victoria (396). The Black Summer was the worst bushfire season on the state of Victoria's record. New South Wales also experienced the longest continuous burning in the history of Australia's bushfire. It consumed more than 4 million hectares.

The most predominant land cover type in Southeast Australia is hummock grasslands (23%) and eucalypt woodlands (Geoscience Australia, 2020). In general, Australia is known to be the lowest elevation continent in the world, with an elevation averaging 330 meters. The highest points on the other continents are all more than twice the height of Australia's highest peak, Mount Kosciuszko, which is 2,228 meters above sea level.

3. Datasets and Modeling

3.1 Data Acquisition & Processing

Fire occurrences between the period of September 2019 and March 2020 were collected from the MODIS MCD64A1 product as a polygon shapefile and mapped. Rainfall data were obtained in the form of a CSV file for all seven months from the Bureau of Meteorology, Australia and all CSV files were then combined into a single file using a python script. A point feature class was created and converted into a raster using the Inverse Distance Weighting (IDW) interpolation tool. The National 9 second (~250 m) DEM of Australia was downloaded from Geoscience Australia and their derivatives (slope and aspect) were created by running the Slope and Aspect tools in ArcGIS Pro 2.5 respectively. Sentinel-2 L2A (16 bit) data was downloaded from the Sentinel Hub EO browser at a resolution of 10 m and NDVI was mapped using band 4 and 8. All variable raster images were clipped to extract the study area. ECOSTRESS data products, including evapotranspiration (ET), evaporative stress index (ESI), Land Surface Temperature (LST), and Water Use Efficiency (WUE), acquired from NASA LPDAAC AppEARS, were used to model wildfire dynamics. A mosaic dataset in a raster format was created for each variable over the seven-month period (September 2019 - March 2020).

3.2 Building a Dataset

Our approach was designed to set up a three-week time lag for data collection prior to a wildfire event in the 4^{th} week and predict wildfire probability in the following week (5^{th} week). We computed mean values of the select data in three weeks to minimize or eliminate gaps. The Australian bushfire started to spread in the first week of September 2019 and faded early April 2020. The fires ceased at the end of October 2019 in south eastern Australia and reignited in late November 2019. To understand the impact of change in climate condition of the country after the first fire and to effectively assess the fire influencing factor, we built models for the first week of September (the week wildfire started), last week of November, and the first week of December (the weeks when the second fire started). In addition, we also built a general model for the entire wildfire period. The datasets were built accordingly for each model.

3.3 Modeling Fire Dynamics Using Machine Learning Approach

Machine Learning is based on algorithms capable of learning from and making predictions on data, through the modeling of the hidden relationships between a set of input and output variables, representing the predisposing factors (explanatory variables) and the occurrences of the phenomenon (dependent variable) (Tonini et al., 2020). Here, the main approach to modeling fire dynamics was Logistic Regression (LR). Additionally, we evaluated Random Forests Classifier (RF) and Geographically Weighted Regression (GWR) algorithms to create models that fit relationships between wildfire events and the explanatory variables. Refer to supplementary information for more details about the RF and GWR methods and results. The fit relationships from these models were then used in the susceptibility mapping and assessment of variable influence.

Logistic Regression is a machine learning method that defines a set of independent input variables to estimate the occurrence probability value of predictor variables that are dependent upon the independent variable. (Hosmer and Lemeshow, 2000). The input dataset provided to the logistic regression function acts as explanatory variables to predict the probability of wildfire locations within the study area. The equation used is:

P= 1/1+exp (-b0+ b1V1+b2v2+…..bnVn)

where P is probability value, b is the coefficient of each explanatory variable, and V is explanatory variable.

3.4 Performance Assessment Procedure

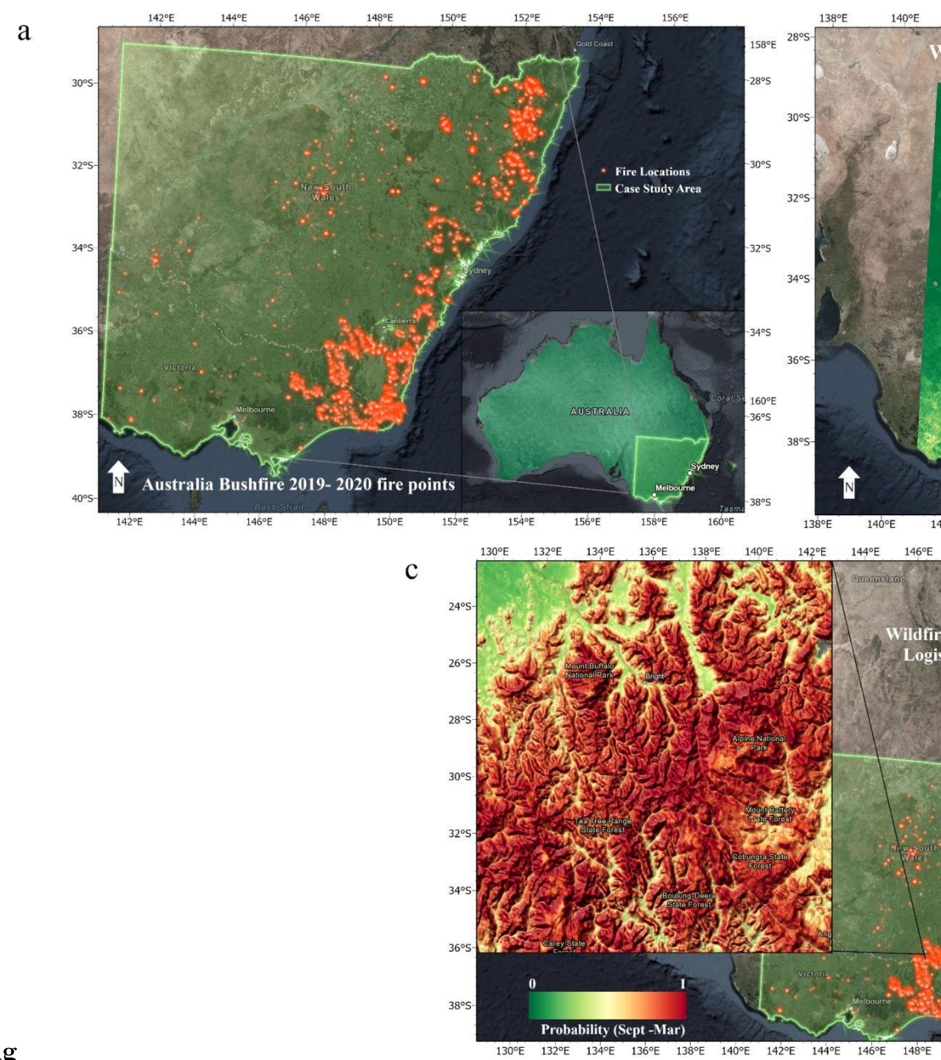
Overall accuracy, sensitivity and specificity are calculated by creating a confusion matrix to assess the performance of the model results. A confusion matrix is a performance measurement for machine learning classification problems where output can be two or more classes. It is a table with four different combinations of predicted and actual values: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). True Positive is defined as it is vulnerable to fire, True Negative as it is not vulnerable to fire, False Positive as vulnerable to fire but it is not vulnerable to fire, and False Negative as not vulnerable to fire but it is vulnerable to fire. Accuracy is calculated using the formula:

(TP + TN)/ Total * 100

Further, predictive performance is determined by calculating the sensitivity, specificity, positive predictive ability, and negative predictive ability of the model.

4. Results

The main results of this study are presented as: (1) susceptibility mapping, which includes probability of fire occurrences values produced by logistic regression; (2) identifying the cause of the fire spread by understanding the importance of explanatory variables; (3) assessing the model performance to evaluate stability and consistency; and, (4) assessing the cities that might be at risk of wildfire spread.



4.1 Susceptibility Mapping

Figure 1a-c. Wildfire Susceptibility mapping for southeastern Australia: a) Ground fire points 2019-2020 from MODIS; b) Wildfire probability maps based on logistic regression from a suite of explanatory variables; c) Zoom-in example at 70 m x 70 m resolution. The maps have a probability scale of 0-1. Areas at probability value 0 are not vulnerable to fire whereas areas at probability value 1 are most vulnerable.

Figure 1a shows the ground fire points of Australian bushfire 2019-2020 extracted from MODIS. For susceptibility mapping, the probability for each pixel to burn within 3 weeks under the consideration of a set of explanatory variables was given as an output from three models (RF and GWR susceptibility maps are presented in Supplementary information, Figure S1 and S2). These values were used to map the wildfire susceptibility in the fire season. Susceptibility maps of 70 m resolution for all three models were produced. The logistic regression susceptibility map is shown below in Figure 1b. We find that the probability of wildfire occurrence is towards the coastal area as experienced in the 2020 wildfire event, showing the more susceptible areas. Figure 1c shows the zoomed in map of the Alpine national park region where the fire was very high during the season.

4.2 Importance of Explanatory Variables

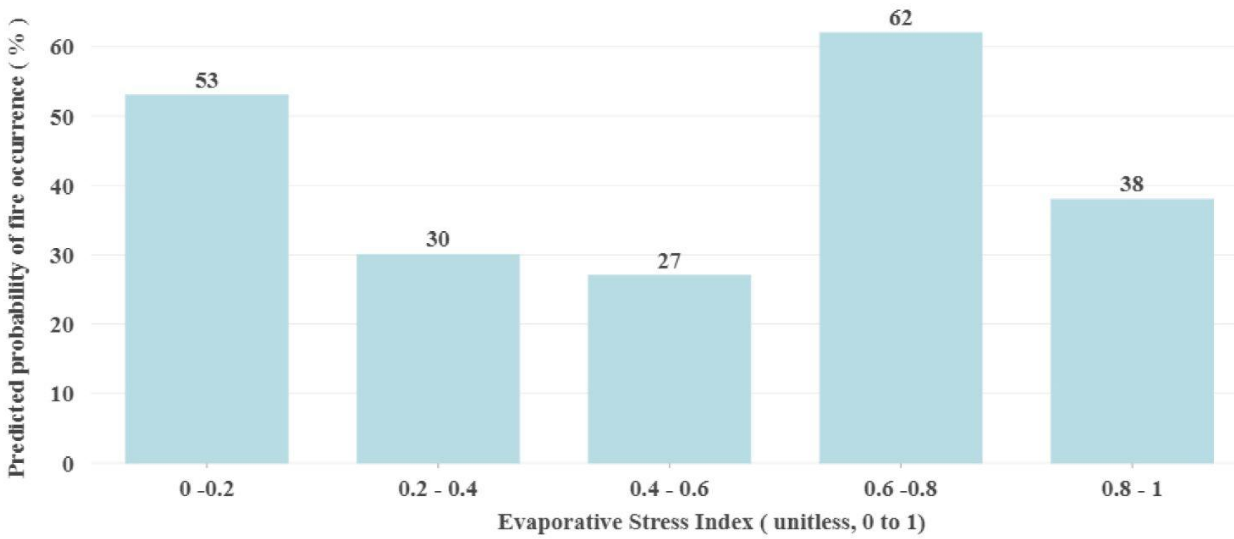
The importance of explanatory variables was ranked using correlation coefficients produced from the models. General models with data from the period of September to March 2020 were created along with monthly models (September, November and December) for deeper understanding. The logistic regression models show evaporative stress index, NDVI, rainfall and water use efficiency as the top-ranking variables in the models.

Explanatory Variables	Logistic Regresssion Models											
	General			September			November			December		
	Coefficient	Std.Dev	Probability	Coefficient	Std.Dev	Probability	Coefficient	Std.Dev	Probability	Coefficient	Std.Dev	Probability
Evaporative Stress Index	3.11	0.14	0.00*	1.63	1.10	0.14	1.70	1.84	0.35	3.79	3.46	0.27
Water Use Efficiency	0.66	0.02	0.00*	0.23	0.06	0.00*	0.76	0.16	0.00*	0.41	0.31	0.19
Evapotranspiration	0.00	0.00	0.02*	-0.01	0.00	0.01*	0.00	0.00	0.55	0.00	0.00	0.69
Land Surface Temperature	0.00	0.00	0.77	0.00	0.00	0.06	0.00	0.00	0.5	0.00	0.00	0.13
NDVI	0.29	0.13	0.02*	9.18	1.63	0.00*	8.27	1.26	0.00*	10.76	2.25	0.00*
Rainfall	-0.09	0.01	0.00*	-0.91	0.11	0.00*	-0.21	0.12	0.11	-0.68	0.26	0.01*
Slope	0.14	0.01	0.00*	0.11	0.02	0.00*	0.12	0.02	0.00*	0.17	0.05	0.00*
Elevation	0.00	0.00	0.00*	0.00	0.00	0.00*	0.00	0.00	0.42	0.00	0.00	0.01*
Aspect	0.00	0.00	0.00*	0.00	0.00	0.00*	0.00	0.00	0.01*	0.00	0.00	0.61

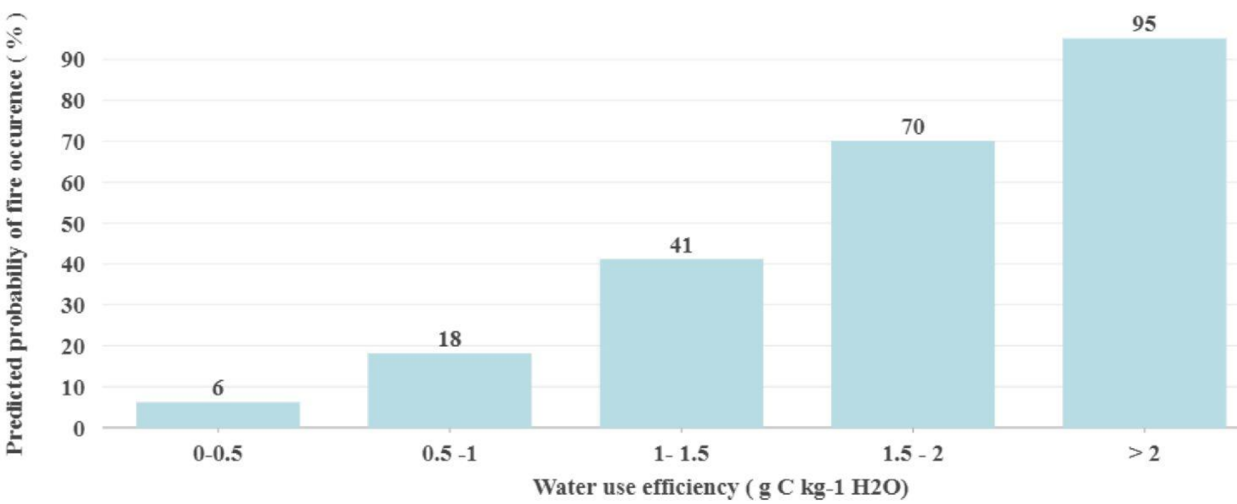
Table 1. Correlation coefficients, standard deviation and probability values of the explanatory variables from the logistic regression model (General, September, November and December); * represents statistically significant.

It is informative to note that the ECOSTRESS data, particularly evaporative stress index (ESI) and water use efficiency (WUE), are found to be top-ranking, having more weight than the rest of the variables (Table 1). In the general logistic regression model, the ESI coefficient averages approximately three times more than the rest of the variables. This makes it evident that south-eastern Australia was likely in drought condition during the wildfire season, i.e., vegetation is stressed due to lack of water which is indicated by the ESI variable. Water use efficiency (the ratio of carbon uptake to water use) with an average of 1.88 g C kg^{-1} H_2O over the study area during fire season was significantly correlated to wildfire with a correlation coefficient of 0.66. The low value of evapotranspiration caused WUE to increase during this period.

a


Predicted probability of fire occurrence (%)
0
10
20
30
40
50
60
53
30
27
62
38
0 -0.2
0.2 - 0.4
0.4 - 0.6
0.6 -0.8
0.8 - 1
Evaporative Stress Index (unitless, 0 to 1)

b


Predicted probability of fire occurence (%)
0
10
20
30
40
50
60
70
80
90
6
18
41
70
95
0-0.5
0.5 -1
1- 1.5
1.5 - 2
> 2
Water use efficiency (g C kg-1 H2O)

c

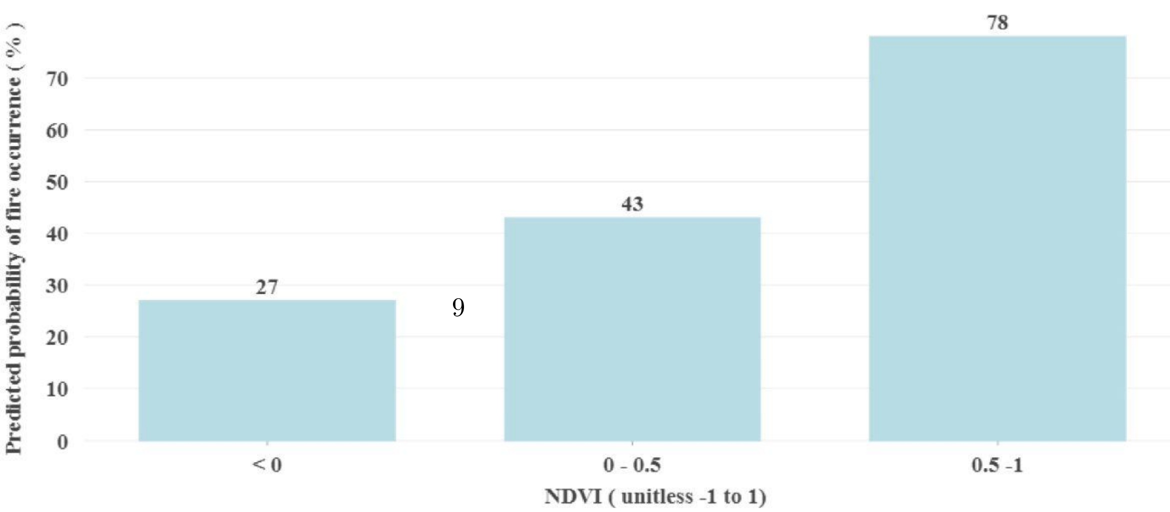

Predicted probability of fire occurrence (%)
0
10
20
30
40
50
60
70
27
43
78
< 0
0 - 0.5
0.5 -1
NDVI (unitless -1 to 1)

Figure 2a-c. The influence of main predictor variables for fire occurrence: a) evaporative stress index; b) water use efficiency; and, c) NDVI.

NDVI is a measure of greenness, which, in the model, acts as a 'switch' that indicates if the region is vegetated or not. Even though NDVI has been identified as one of the top four variables, we do not consider it as an insightful predictor—rather we consider it a conditional dictator (i.e., one cannot have fire if there is no vegetation). From NDVI, 77% of fires occurred in highly vegetated areas (NDVI > 0.5). 95% of the vegetation burned during wildfire has water use efficiency values greater than 2 g C kg^{-1} H_2O (Figure 2).

It is also noted that rainfall might not have played a major role in the fire occurrence. Although precipitation was a dominant factor in all three models, it was not statistically significant. For November and December, despite high rainfall in a given region, the predicted wildfire probability was also high. As for the month of September, regions that had high rainfall did not have wildfires. But since it was the starting week of the wildfire, rainfall would not have played any role in the wildfire spread. So, rainfall deficiency might not be the sole cause of the wildfire spread.

4.3 Performance Assessment

In the wildfire susceptibility modeling accuracy assessment, pixels were classified as fire and non-fire using a threshold probability value of 0.5 (any pixel with a value above 0.5 is considered a fire pixel and vice versa). In order to evaluate the performance of logistic regression models, the accuracy assessment was performed by creating a confusion matrix. The effectiveness of the LR model was specified by evaluating the sensitivity and specificity of the model.

To assess the performance of the models, true positive, true negative, false positive and false negative metrics are used to compute the confusion matrix, which is subsequently used to determine the other evaluation metrics, i.e., sensitivity, specificity, accuracy, positive predictive ability, and negative predictive ability. The overall general model obtained an accuracy of 83%, sensitivity was 81%, specificity was 84%, positive predictive ability was 83%, and negative predictive ability was 83%. The monthly models – September, November, and December –produced an accuracy of 83%, 85%, and 88% respectively.

The performance results demonstrate the effectiveness of the LR models. Even though two of the time-specific models generate a slightly higher accuracy, the general model can be used effectively to predict wildfire probability for any other areas elsewhere if time is not an issue or important factor. Since the general model uses far more data over the entire wildfire period, it can be expected to be more consistent, robust, or reliable.

4.4 Assessing the Susceptibility of Cities

The south-eastern cities that fall on wildfire-affected areas were also mapped based on the spread of fire pixels. A 5 km buffer for each city was created and the mean predicted fire probability value of each pixel covering the buffer was derived. Figure 3 shows the cities susceptible to wildfire. The Gipps land region (Alpine and East Gippsland) in Victoria State, mid north coast and Coffs Harbor regions (Blue mountains, Oberon and Lithgow) in New South Wales are predicted as highly susceptible areas followed by greater Sydney and Melbourne areas. This information can be used to help policy makers, fire managers, forest rangers, and city planners to assess, manage, prepare, and mitigate wildfires in connection to fire susceptibility of nearby cities.

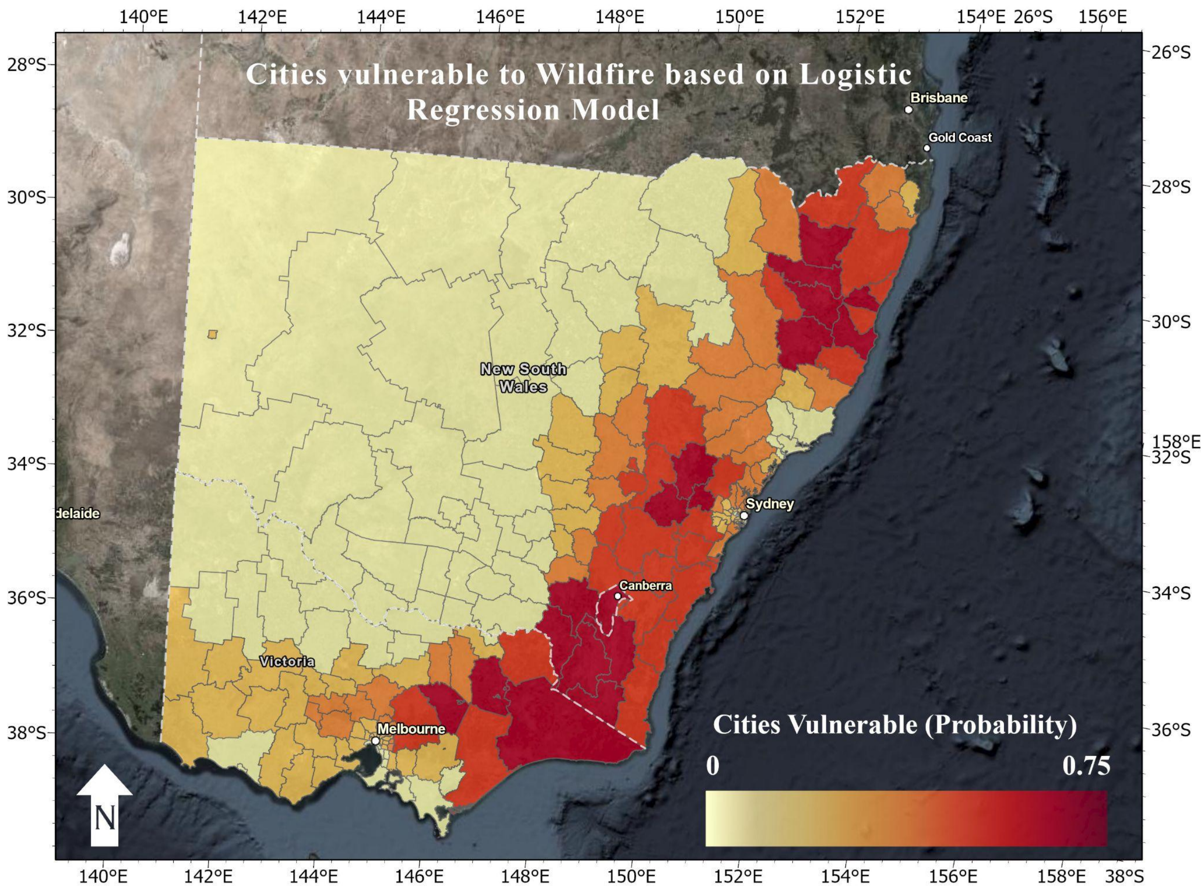


Figure 3. Map showing the ranking of wildfire susceptibility in different districts of the study area. Maroon represents regions more vulnerable to wildfire while yellow represents regions less vulnerable to wildfire.

5. Discussion

In this study, fire points were used as a dependent variable whereas various biophysical factors like slope, elevation, aspect, rainfall, NDVI, ECOSTRESS data including evapotranspiration, evaporative stress index, water use efficiency and land surface temperature (Fisher et al., 2020) were selected to determine the wildfire dynamics over southeastern Australia from September 2019 - March

2020 (Ganteaume et al., 2013; Nami et al., 2018; Parisien et al., 2012). NASA's ECOSTRESS was used to primarily address how water availability affects the key climate biomes around the world; this research quantifies the impact that drought has on wildfire.

Climate change is a major factor in the increase of wildfire events (Lim et al. 2019). There are few studies which show that the change in temperature and land cover are likely to be the cause of wildfire spread (Lim et al., 2019; Halofsky et al., 2020). However, our data showed that the land surface temperature of south-eastern Australia during the wildfire season did not have much increase compared to the pre-fire season. It is also shown by our results that the land surface temperature was not statistically significant. Additionally, from our results, the evaporative stress index, which is a drought indicator, is seen as the most contributing factor, followed by water use efficiency, NDVI, and rainfall.

We employed three different methods to predict wildfire occurrences: logistic regression, geographically weighted regression and random forest models. 70 m resolution wildfire susceptibility maps are generated using all three models and they are found to be useful in visualizing the intensity of fire during the wildfire season. With a high resolution, it helps the firefighters, other authorities or even the general public to locate the affected areas more accurately.

The model evaluation process revealed that the Random Forest model had an accuracy of 91%, Geographically Weighted Regression model accuracy 85% while Logistic Regression models had 83% accuracy. However, we adhered to the logistic regression model in this paper. Although we find that the performance of the three models implemented in this research is operational, the Logistic regression model is considered the most effective as it shows consistent results across all months. The Random forest model involves a lot of over-fitting (Liaw and Wiener, 2002) and the results are relatively inconsistent. Though the Geographically weighted regression model has shown a good performance and higher accuracy, it is questionable for predicting fire at a global level or examining the strength of predictor variables because GWR generates a local regression model at each point (pixel) based on locally associated similar or more homogeneous pixels around each pixel, their R2 values are generally higher than one regression model for the entire image (Fotheringham et al., 2002; Oliveira et al., 2012). Since they all are local models, their predictions are generally more accurate at each point leading to higher overall accuracy. Unlike logistic regression models, it does not represent the relation between the dependent and explanatory variables globally.

We also developed monthly models to understand how change in climate conditions affect wildfire spread. It is found that, for the month of December, the evaporative stress index coefficient is nearly three times more than the September month. That shows plants had become even more stressed during the second fire (month of November and December). This makes ECOSTRESS data a valuable factor in the analysis of wildfire prediction now and in the future.

6. Conclusion

Our research developed an effective model to predict wildfires in other time periods or other similar areas in the world. We conclude that the logistic regression model using ECOSTRESS data can be employed effectively to predict wildfire probability for the 2019-2020 Australian fires. This can be achieved without prior knowledge of geospatial processing and machine learning algorithms. The analytical procedures, data, regression models, and susceptibility mapping approach for cities can help policy makers, fire managers, and city planners to assess, manage, prepare, and mitigate wildfires in the future.

7. Acknowledgements and Data Availability

Data in this project is extracted from different sources. Climate data from Australian Government http://www.bom.gov.au/climate/data/, ECOSTRESS data from Land Processes Distributed Active Archive Center (LP DAAC) AppEEARS - https://lpdaacsvc.cr.usgs.gov/appeears/, MODIS Thermal Anomalies/Fire- https://lpdaac.usgs.gov/products/mod14a2v006/, NDVI from https://scihub.copernicus.eu/dhus/#/home and DEM from https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/66006.

ArcGIS Pro Software by Esri was used to curate the maps and perform Analysis.

This research is funded by NASA Award # 80NSSC20K0168. JBF was supported in part by NASA ECOSTRESS.

References

1. Abdel-Rahman, Elfatih & Mutanga, Onisimo & Adam, Elhadi & Ismail, Riyad. (2014), Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machine classifiers. ISPRS Journal of Photogrammetry and Remote Sensing. 88. 48–59. 10.1016/j.isprsjprs.2013.11.013.

2. Alan F. Srock , Joseph J. Charney, Brian E. Potter and Scott L. Goodrick. (2018), The Hot-Dry-Windy Index: A New Fire Weather Index.MDPI Atmosphere Journal.

3. Alexander I. Filkov, Tuan Ngo, Stuart Matthews, Simeon Telfer &Trent D. Penman.(2020), Impact of Australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends. Journal of Safety Science and Resilience,Volume 1, Issue 1,Pages 44-56,ISSN 2666-4496.

4. Brown, C. D., Liu, J., Yan, G., & Johnstone, J. F. (2015), Disentangling legacy effects from environmental filters of postfire assembly of boreal tree assemblages. Ecology, 96, 3023–3032. https://doi. org/10.1890/14-2302.1

5. BushFires in Australia Research Guides, (2020) https://guides.slv.vic.gov.au/bushfires

6. Caceres, Claudia. (2011), Using GIS in Hotspots Analysis and for Forest Fire Risk Zones Mapping in the Yeuare Region, Southeastern Honduras.

Available at http://www.gis.smumn.edu/GradProjects/CaceresC.pdf.

7. Doerr, S. H., & Santín, C. (2016), Global trends in wildfire and its impacts: perceptions versus realities in a changing world. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, *371*(1696), 20150345. https://doi.org/10.1098/rstb.2015.0345

8. Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.-H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. Water Resources Research 56(4): 1-20.

9. Fosberg, M. (1978), Weather in Wildland Fire Management: The Fire Weather Index. Conference on Sierra Nevada Meteorology, Lake Tahoe, CA, 1–4.

10. Fotheringham, A.S., Brunsdon, C., & Charlton, M.(2002), Geographically weighted regression: The analysis of spatially varying relationships. New York, John Wiley & Sons: 269 p.

11. Ganteaume, Anne., Camia, Andrea & Jappiot, Marielle & San-Miguel-Ayanz, J. & Long-Fournel, Marlène & Lampin, Corinne. (2013), A Review of the Main Driving Factors of Forest Fire Ignition Over Europe. Environmental management. 51.

12. G. Hulley, S. Hook, J. Fisher, & C. Lee, "ECOSTRESS, A NASA Earth-Ventures Instrument for studying links between the water cycle and plant health over the diurnal cycle," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 5494-5496, DOI: 10.1109/IGARSS.2017.8128248.

13. Geoscience Australia,(2020) Website: https://info.australia.gov.au/about-australia/our-country/our-natu

14. Halofsky, J.E., Peterson, D.L. & Harvey, B.J. (2020), Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. *fire ecol* 16. https://doi.org/10.1186/s42408-019-0062-8

15. Hamid Reza Pourghasemi, Candan Gokceoglu, Spatial Modeling in GIS

and R for Earth and Environmental Sciences, Elsevier, 2019, Pages 607-619, ISBN 9780128152263, https://doi.org/10.1016/B978-0-12-815226-3.00028-4.

16. Hosmer, D.W. and Lemeshow, S. (2000), Applied logistic regression. 2nd Edition, John Wiley & Sons, Inc., New York. doi:10.1002/0471722146

17. Jaafari, Abolfazl & Mafi-Gholami, Davood & Zenner, Eric. (2017), A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Ecological Informatics. 39. 10.1016/j.ecoinf.2017.03.003.

18. Jaafari, Abolfazl & Pourghasemi, Hamid Reza. (2019), Factors influencing regional scale wildfire probability in Iran: An application of random forest and support vector machines. 10.1016/B978-0-12-815226-3.00002-8.

19. Kennedy Were, Dieu Tien Bui, Øystein B. Dick, Bal Ram Singh (2015), A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecological Indicators,Volume 52, 2015, Pages 394-403,ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2014.12.028.

20. Lee, C. M., J. B. Fisher, & S. J. Hook (2020), Mapping vegetation health around the world, *Eos, 101,* https://doi.org/10.1029/2020EO146736. Published on 08 July 2020.

21. Lewis, S. L., Edwards, D. P., Galbraith, D. (2015) Increasing human dominance of tropical forests. Science 349: 827–832.

22. Liaw, A. and Wiener, M. (2002), Classification and Regression by Random Forest. R News, 2, 18-22.

23. Lim, Chul-Hee & Kim, You & Won, Myungsoo & Kim, Sea Jin & Lee, Woo-Kyun. (2019), Can satellite-based data substitute for surveyed data to predict the spatial probability of forest fire? A geostatistical approach to forest fire in the Republic of Korea. Geomatics, Natural Hazards and Risk.

24. Marj Tonini, Mirko D'Andrea , Guido Biondi , Silvia Degli Esposti , Andrea Trucchia and Paolo Fiorucci (2020),A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy.

25. MCARTHUR, A.G. 1966. Weather and grassland fire behavior. Dep. of Nat. Dev., For. and Timber Bur. Leafl. No. 100. Canberra, Australia. 23 p.

26. Michael A. Wulder, Jeffrey G. Masek, Warren B. Cohen, Thomas R. Loveland, Curtis E. Woodcock. (2012), Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment Volume 122,Pages 2-10,ISSN 0034-4257,https://doi.org/10.1016/j.rse.2012.01.010.

27. Nami, M.H., Jaafari, A., Fallah, M. *et al.* Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* 15, 373–384 (2018). https://doi.org/10.1007/s13762-017-1371-6

28. Ngoc-Thach, Nguyen & Dang Ngo, Toan & Pham, Xuan-Canh & Nguyen, Hong-Thi & Bui, Hang & Hoang, Nhat-Duc & Tien Bui, Dieu. (2018), Spatial Pattern Assessment of Tropical Forest Fire Danger at Thuan Chau area (Vietnam) using GIS-Based Advanced Machine Learning Algorithms: A comparative study. Ecological Informatics. 46. 10.1016/j.ecoinf.2018.05.009.

29. Parisien, Marc-André & Snetsinger, Susan & Greenberg, Jonathan & Nelson, Cara & Schoennagel, Tania & Dobrowski, Solomon & Moritz, Max. (2012), Spatial variability in wildfire probability across the western United States. International Journal of Wildland Fire. 21. 313-327. 10.1071/WF11044.

30. Sandra Oliveira, José M.C. Pereira, Jesús San-Miguel-Ayanz &Luciano Lourenço(2014), Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression, Applied Geography,Volume 51,2014, Pages 143-157, ISSN 0143-6228, https://doi.org/10.1016/j.apgeog.2014.04.002.

31. Satir, Onur & Berberoglu, Suha & Donmez, Cenk. (2016). Mapping regional forest fire probability using artificial neural network models in a Mediterranean forest ecosystem. Geomatics, Natural Hazards and Risk. 7. 1645-1658. 10.1080/19475705.2015.1084541.

32. Simon Whitburn, Martin van Damme, Lieven Clarisse, Solène Turquety, Cathy Clerbaux, et al.. Doubling of annual ammonia emissions from the peat fires in Indonesia during the 2015 El Niño. Geophysical Research Letters, American Geophysical Union, 2016, 43 (20), pp.11,007-11,014. ff10.1002/2016GL070620ff. ffinsu-01375904

33. Tonini, M.; D'Andrea, M.; Biondi, G.; Degli Esposti, S.; Trucchia, A.; Fiorucci, P. A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. *Geosciences* **2020**, *10*, 105. https://doi.org/10.3390/geosciences10030105

34. Tien Bui, Dieu., Hung Van Lec., Nhat-Duc Hoang.(2018). GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method.,Ecological Informatics.

35. (2020), ArcGIS Pro (Version 2.5). Esri Inc. https**://www.esri.com**/en-us/arcgis/products/arcgis-pro/overview