BURNED AREA PREDICTION IN SOUTHERN ASIA USING MACHINE LEARNING WITH LAND AND ATMOSPHERIC PARAMETERS

Amir Mustofa Irawan[1], Mercè Vall-llossera[1,2], Carlos López-Martínez[1,2], Adriano Camps[1,2,3], David Chaparro[4], Gerard Portal[2], Miriam Pablos[1]

[1]CommSensLab – UPC, Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain
[2]Institut d'Estudis Espacials de Catalunya IEEC, 08034 Barcelona, Spain
[3]ASPIRE Visiting International Professor, UAE University CoE, 15551 Al Ain, UAE
[4]German Aerospace Center, Microwaves and Radar Institute, Münchener Strasse 20, 82234 Wessling, Germany

ABSTRACT

In the work a random forest model has been implemented as an interpretable machine learning tool in the effort to estimate the burned areas caused by fire outbreaks in India, Pakistan, and Myanmar in April and May 2022. The proposed model combines environmental and atmospheric (including upper tropospheric) factors suggested to drive patterns of burned areas, and determines the weight of each factor on the propagation of fires. Results demonstrate that the model mimics the actual burned area by considering a combination of vegetation, atmosphere, and human-related variables and improves accuracy by approximately 7% after adding jet stream features. This approach could lead to implement a semi-operational forecast system that may be tested in multiple demonstration sites.

Index Terms— Random forest, burned area, passive microwave sensing, jet stream, soil moisture, vegetation optical depth.

1. INTRODUCTION

Wildfires affect the preservation of ecosystems and ultimately result in ecological harm and human casualties [1]. Early wildfire forecasting and detection can significantly reduce the destructive effects of fires by contributing to the establishment of early warning systems. However, wildfire propagation is a complex, non-linear process that includes the interactions of many variables and circumstances, and is challenging to forecast.

Therefore, understanding the environmental factors that drive patterns of burned areas and fire severity is critical. The impact of climate change on the water content of soils and vegetation may influence when and where large fires occur. In addition, persistent tropospheric ridging in jet streams has been found to be responsible for temperature extremes [2] that have led to wildfire activity. The combination of land, vegetation, and atmospheric (including upper tropospheric) driving effects must be understood to analyse the fire fuel conditions and resulting potential propagation of fires.

Machine learning (ML) and advanced simulation methods have been explored to assist in forecasting wildfires by using several remote sensing and meteorological variables [3]. This study focuses on the complementary integration of soil, vegetation, atmosphere, and human-related variables with a comparison of results to those from other ML algorithms.

2. METHODOLOGY

Table 1. Summary of variables used in this study, including Soil Moisture (SM), Vegetation Optical Depth (VOD), Vapor Pressure Deficit (VPD), Land Surface Temperature (LST), 300-hPa wind component (u_{300} and v_{300}), 500-hPa Geopotential Height Anomalies (ΔZ_{500}), Elevation, Land Use, Distance to Road, and Burned Area.

Source	Parameter	Resolution
BEC SMOS L3	SM	~25 km
	VOD	~25 km
ERA5 ECMWF	VPD	~25 km
	LST	~25 km
NCEP–DOE II	u_{300} and v_{300}	~210 km
	ΔZ_{500}	~210 km
SRTM	Elevation	90 m
GLC2000-JRC	Land use	1 km
Open Street Map	Distance to road	-
MODIS	Burned Area	1 km

Table 1 summarizes the data used in this study. In addition, anomalies are also computed for SM, VOD, VPD and LST with the following approach. First, a climatology is obtained by smoothing over an average year with a centered moving window of 61 days. Second, nine-day anomalies are defined

as the differences between the raw data and the climatology from the day of a fire outbreak backwards to the beginning of the window.

2.1. Spatiotemporal composites for upper-tropospheric features

The meridional circulation index (MCI) is a measure of the jet stream's North-South (meridional) meandering and it is calculated as [2]:

$$MCI = \frac{v * |v|}{u^2 + v^2}$$

where are u and v are the meridional and zonal components of the wind, respectively. An MCI of 0 indicates the wind is purely zonal, whereas an MCI of 1 (or -1) indicates the flow is from the south (or north).

The spatio-temporal composites are also considered of the 300-hPa Meridional Circulation Index (MCI_{300}) and 500-hPa geopotential height anomalies (ΔZ_{500}), which average the time-varying climatic fields $F(x,t)$ related to the location x and time t of N events $E(x_i, t_i)$, $i \leq N$. The relevant climatic field is restricted to a box of latitude and longitude of 100° and 60°, respectively, for each event, and the variable N of such cases is averaged to produce the final composite field. For this study, lead-lag composites have been built by combining fields from 20 days prior to each event to 20 days after (i.e., 41 daily composites) [2].

2.2. Model construction

The burned area sample collected in the study area spans India, Pakistan, and Myanmar and was selected to review the causes of the extraordinary fire outbreaks that struck these areas from April to May 2022 [4]. For this event, the sample consists of 13,777 wildfire point locations regionally distributed, as seen in Fig. 1.

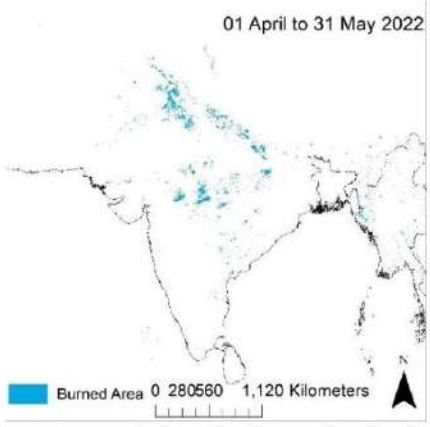

Fig. 1. Burned areas during the fire outbreaks in India and surrounding regions from April to May 2022.

Excessive noise and extraneous variability were removed by binning the data based on the MODIS fire pixel resolution (1 km) and assigning a value of average predictors to each pixel. As a result, the final sample contains 915 burned areas with average predictors based on the burned area bin resolution. The dataset is also randomly separated into training (80%), and testing (20%) subsets. The model prediction is designed to output a decimal logarithm of the burned area.

A random forest (RF) method is selected in this study to analyse and forecast the wildfires because this model can represent individual tree decisions [5] providing clear explanations of the importance of the variables used to derive the predictions. The normalisation of the data is not required because this approach is based on splitting data to generate predictions. Additional details on calculating the importance of the variables employing RF can be found in [6].

3. RESULTS

3.1. Development of burned area predictions based on the multisource Earth Observation Dataset

The RF-based variable importance that impacts wildfire propagation is shown in Fig. 2.

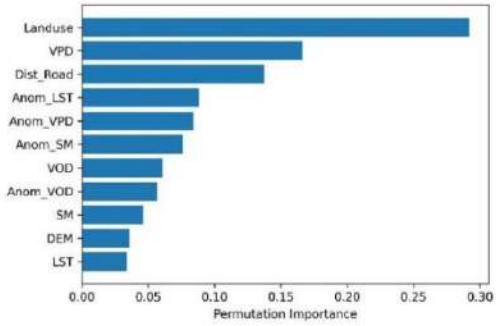

Fig. 2. Variable importance for forecasting burned areas, before incorporating jet stream features.

The results in Fig. 2 show that land use is the most important factor in determining the burned area value, according to the model, which is reasonable because most fires during this event occurred in croplands in India. As a validation test, the model's precision is significantly impacted when the land use category is interchanged or altered. VPD (i.e., atmospheric dryness) is identified as the second most important variable for predicting burned areas according to the model, which aligns with previous research [7]. The distance to roads is also a relevant metric (the third in importance), which is expected because fire ignitions are more frequent in areas surrounding roads and human settlements. The anomaly of SM is listed as the sixth variable in importance, which agrees with research from other regions that shows the important role of these variations in explaining fire outbreaks [8].

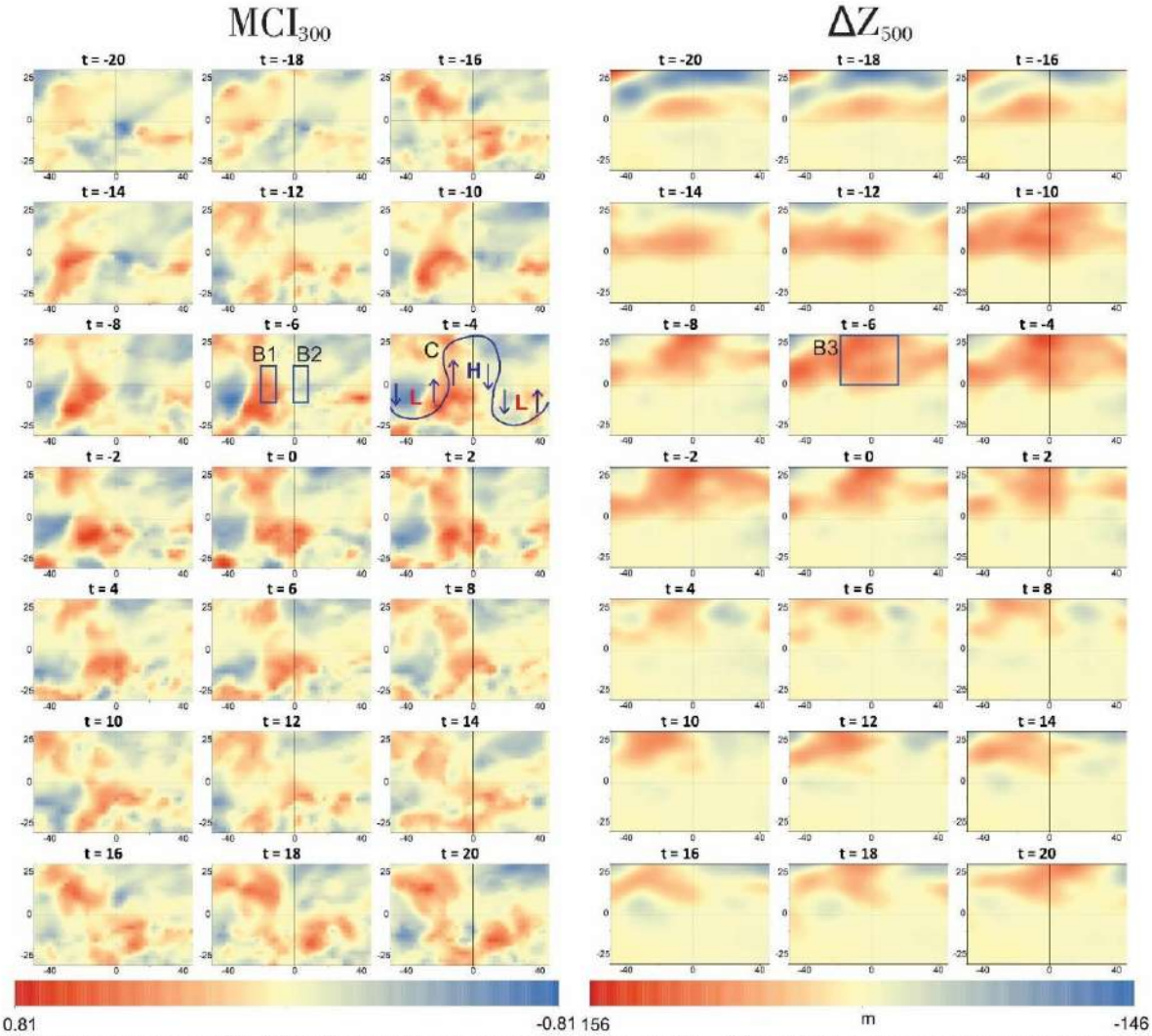

Fig. 3. (Left) Composite of the 300-hPa meridional circulation index (MCI_{300}) for all large fire spread events, with individual composites ranging from 20 days before each fire spread event (lag = -20) to 20 days after each fire spread event (lag = +20). MCI_{300} is composited relative to longitude and latitude for each large fire spread event. (Right) Similarly, the 500-hPa geopotential height anomalies (ΔZ_{500}) features a selected box that indicates the commencement of the ridge that begins in the centre of the event for MCI_{300} (B1 and B2) and ΔZ_{500} (B3), then adding these to the model as jet stream features. The illustration of the ridge (high area) appears on the left panels (C) with the timing of the ridge pattern such that the onset of the ridge starts to occur around at least t = -10 (strongly occurred at t = -6) and decays gradually after t = 4.

3.1. Adding jet stream features and ML model comparisons

A composite analysis is conducted for the ΔZ_{500} and MCI_{300} (see Fig. 3) to better understand the pattern between the jet stream and large wildfires ($\geq$ 5000 ha). Insight from a similar analysis and plotting mechanism can be found in [2]. The figures are centered on the locations of each large fire incident between 20 days before and after each event. To investigate the impact of the jet stream on the wildfire forecast, we add a box that indicates the beginning of when the ridge begins. The highest ΔZ_{500} positive value and the evident South (or North) MCI_{300} pattern strongly occur in the centre during the previous six days (t = -6) before the fire, which corresponds to the development of an upper-tropospheric ridge in the middle.

In addition, other ML algorithms, including supervised vector regressor (SVR), XGBoost, gradient boosting regressor (GBR), and a neural network (NN), were run and compared to the proposed RF model to determine their advantages. As presented in Fig. 4a, the RF model is evaluated by the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE). As it is seen by these metrics (Fig. 4a), the accuracy of the RF models improves by

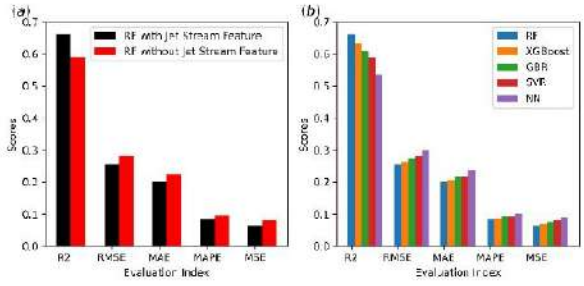

Fig. 4. (a) Accuracy of the RF models with and without the jet stream parameters. (b) Comparison of the evaluation metrics between multiple ML algorithms after adding the jet stream features.

approximately 7% (scores from 0.59 to 0.66) after adding the jet stream features (MCI_{300} and ΔZ_{500}). Comparing multiple ML models, as shown in Fig. 4b, the RF model performs the best for forecasting burned areas.

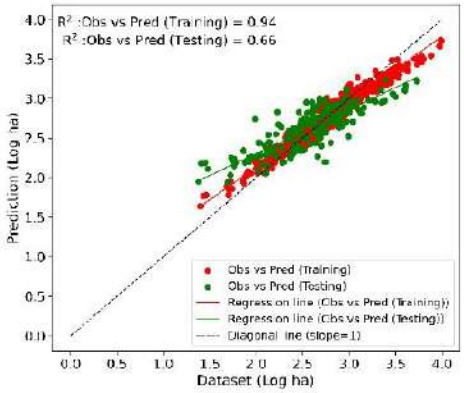

Fig. 5. Decimal logarithm of burned area estimation (originally in hectares) of the best RF model

Figure 5 shows the estimated burned areas based on the RF as the best model using the combination of land and atmospheric feature (including the jet stream features) during training and testing. As suggested by this figure, the model closely follows the actual burned area during most events except for several peak conditions. Future work will extend the period and optimise the ML techniques to better capture the pattern of the fires, specifically the large fires.

4. CONCLUSIONS

A random forest model leveraging a combination of land and atmospheric features (among others) is suggested to forecast the burned areas, as tested on a extraordinary fire season in southern Asia. These findings highlight the complexities of the process and the importance of accounting for key features, such as VPD, SM, upper-tropospheric, and human-related variables. When the upper tropospheric variable was included, the performance of the model improved by about 7%, suggesting some relevance of the jet stream variables. Future work will focus on improving the model performance in large fire scenarios.

5. AKNOWLEDGEMENTS

The project that gave rise to these results received the support of a fellowship from the "la Caixa" Foundation (ID 100010434), with fellowship code LCF/BQ/DI21/11860028. This work was also supported by Project PID2020-114623RB-C32 funded by MCIN/ AEI /10.13039/501100011033. In addition, D. Chaparro has received funding from the XXXIII Ramón Areces Postdoctoral grant.

6. REFERENCES

[1] T. A. Downing, M. Imo, J. Kimanzi, and A. N. Otinga, "Effects of wildland fire on the tropical alpine moorlands of Mount Kenya," *Catena (Amst)*, vol. 149, pp. 300–308, Feb. 2017, doi: 10.1016/J.CATENA.2016.10.003.

[2] P. Jain and M. Flannigan, "The Relationship between the Polar Jet Stream and Extreme Wildfire Events in North America," *J Clim*, vol. 34, no. 15, pp. 6247–6265, Aug. 2021, doi: 10.1175/JCLI-D-20-0863.1.

[3] F. Abid, "A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems," *Fire Technol*, vol. 57, no. 2, pp. 559–590, Mar. 2021, doi: 10.1007/S10694-020-01056-Z/METRICS.

[4] J. V. Gómez, "Com ens arribarà l'efecte dominó de l'onada de calor extrema a l'Índia i el Pakistan, per Jordi Vilardell Gómez," 2022. https://www.ccma.cat/324/com-ens-arribara-lefecte-domino-de-lonada-de-calor-extrema-a-lindia-i-el-pakistan/noticia/3162928/ (accessed Jan. 08, 2023).

[5] F. K. Dosilovic, M. Brcic, and N. Hlupic, "Explainable artificial intelligence: A survey,", *MIPRO 2018 - Proceedings*, pp. 210–215, Jun. 2018, doi: 10.23919/MIPRO.2018.8400040.

[6] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. Springer, 2012. Accessed: Dec. 07, 2022. [Online]. Available: https://link.springer.com/content/pdf/10.1007/978-1-4419-9326-7.pdf

[7] A. P. Williams *et al.*, "Observed Impacts of Anthropogenic Climate Change on Wildfire in California," *Earths Future*, vol. 7, no. 8, pp. 892–910, Aug. 2019, doi: 10.1029/2019EF001210.

[8] D. Chaparro, M. Piles, M. Vall-llossera, and A. Camps, "Surface moisture and temperature trends anticipate drought conditions linked to wildfire activity in the Iberian Peninsula," *Eur J Remote Sens*, vol. 49, pp. 955–971, Dec. 2016, doi: 10.5721/EUJRS20164950.