World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

Remotely sensed dry matter productivity and soil moisture content as potential predictors of arid rangeland wildfires: A case study of Kgalagadi District, Botswana

Issa Kaduyu *, Rejoice Tsheko, Justin H Chepete and Ednah Kgosiesele

Department of Agricultural and Biosystems Engineering, Botswana University of Agriculture and Natural Resources. Private Bag 0027, Gaborone.

World Journal of Advanced Engineering Technology and Sciences, 2022, 07(02), 143–156

Publication history: Received on 27 October 2022; revised on 06 December 2022; accepted on 08 December 2022

Article DOI: https://doi.org/10.30574/wjaets.2022.7.2.0143

Abstract

Fire is a critical tool for managing rangeland ecosystems; however, the increasing wildfire occurrence poses a considerable danger to rangeland ecosystem continuity. Predicting fire occurrence and mapping wildfire danger is critical in managing highly flammable rangelands. To identify potential remotely sensed variables for wildfire prediction, this study employed a Random Forest (RF) classifier using selected environmental variables to assess their possible use for wildfire prediction in Kgalagadi District, Botswana. The study used 107,883 active fire points from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor from 2015 to 2021. Datasets of remotely sensed Dry Matter Productivity (DMP), Soil Moisture (SM), Land Surface Temperature (LST), Live Fuel Moisture Content (LFMC), and Dead Fuel Moisture Content (DMFC) were analysed in ArcMap 10.7 Esri©. The RF model developed gave an Out of Bag (OOB) error of 9.91% and an overall accuracy of 90.15% for classifying fires and non-fire points using the test dataset. The results also showed a Kappa coefficient of 0.803, with 88.25% and 91.76% producer and user accuracies, respectively, for classifying fire points. The DMP was the most significant variable with Mean Decrease Accuracy (MDA)= 1,055.20 and Mean Decrease Gini (MDG)= 9.328.62), followed by SM (MDA= 828.39 and MDG= 15,745). The LFMC and DMFC were found to be weak in detecting fires. It is recommended that field studies be carried out in the study area to calibrate these to improve their contribution to accurate fire prediction, as most literature shows that they are significant in fire prediction.

Keywords: Wildfire prediction; Rangelands; Random Forest; Soil moisture; Dry Matter Productivity; Remote sensing

1. Introduction

There is a significant increase in wildfire incidences globally. Over 4 million square kilometres of land are burned annually, with 70% of the area burned in Africa and contributing 14 % to global greenhouse gas emissions [1–3]. Although wildfires have been reported to play a critical role in the continuous functioning of rangeland ecosystems, the devastating impact of natural wildfires and prescribed fires that often go out of control are far-reaching [4]. Moreover, the frequency, severity, and extent of wildfires have been predicted to increase in the next decade amidst a rapidly changing climate with prolonged droughts and lower precipitation [5, 6]. The expected increase in wildfires calls for accurate wildfire prediction methods and practical tools for wildfire management. Increasingly, control of wildfires is being carried out with acceptable risk management principles while considering analytical validity supported by the increasing wildfire studies [7]. Save for a handful of studies [8–11], wildfires in Africa are understudied, while the extent and frequency of wildfires in Southern African rangelands continue to grow unabated [11].

Developed countries across the world have, in the previous decades, conducted several studies and achieved considerable success in wildfire occurrence prediction based on different methods and various factors such as climate

* Corresponding author: Issa Kaduyu

variables and fuel characteristics applied in various empirical models [12–15]. Wildfire management in Africa remains entirely dependent on the central governments. In Botswana, the Department of Forest and Range Resources (DFRR) spends over US$1.5 million annually on the maintenance of about 10,000km of firebreaks [16], with more effort being given to fire suppression rather than prevention [17], though fire remains a vital tool in land use management in rural communities [18]. Despite efforts to manage wildfires in Botswana, over 1.8 million hectares have been burned annually in the last decade (2012-2021) [19], resulting in considerable devastation of wildlife and communities. Botswana is significantly affected by wildfires due to the high biomass accumulation during the summer and the favourable climate characterised as semi-arid [17]. In Botswana, wildfires are more frequent in the spring (August to October) compared to the winter due to the hot conditions [20]. Kgalagadi and Ghanzi Districts in Botswana are highly drought-prone areas due to their aridness and rainfall below 300mm annually [21]. A significant proportion (63%) of the Kgalagadi district is remote and is occupied by a Wildlife Management Area (WMA). Recently there has been an increase in dry biomass accumulation due to the country's reduced wildlife population in the WMAs [17]. Therefore, it is essential to develop wildfire prediction models for Southern African rangelands that harbour grassland savannahs with the highest fire potential in Africa [17]. The increasing drought incidents in the Kgalagadi District, coupled with the accumulation of dry grass biomass during the rains, make the area highly vulnerable to wildfires necessitating the need for handy tools for fire risk prediction.

Identifying potential predictor variables is critical in developing an accurate wildfire prediction model. The use of meteorological parameters and remotely sensed data in wildfire prediction were employed in several studies [22–26]. However, the application of these parameters and indices for wildfire prediction in Africa is limited by the sparse network of meteorological stations and the capacity to process imagery [27, 28]. To establish an efficient wildfire prediction model, it remains pertinent to determine a set of parameters and indices which can be used to predict the likelihood of wildfire occurrence in the Kgalagadi district. Machine learning methods combined with remotely sensed datasets are increasingly being used in predicting wildfires [24, 29–31]. Random Forest models have received significant attention in wildfire prediction studies and have been reported to produce promising results [24, 32]. Applying remotely sensed imagery in Botswana's wildfire management system is necessary since most of these fires occur in the remote wilderness where they burn unnoticed, exposing the spatial and temporal inadequacy of existing fire management methods such as firebreaks.

This study seeks to identify potential remotely sensed variables for wildfire prediction in Botswana. The hypothesis is that Dry Matter Productivity (DMP), Soil Moisture (SM), Land Surface Temperature (LST), Live Fuel Moisture Content (LFMC), and Dead Fuel Moisture Content (DFMC) could be used as potential predictors of rangeland wildfires.

2. Material and methods

The methodology comprises the study area description, data collection, and data analysis.

2.1. Study area

The study was conducted in Kgalagadi District (105,200 km^2), located southwest of Botswana, about 400km west of Gaborone city, lying between latitudes 20^{0}54' and 21^{0}20' and longitudes -23^{0}16' and -26^{0}46' (Figure 1). The district lies in the Kalahari/ Okavango basin, and it is characterised by a large portion of rangeland with savannah grasslands forming the primary land cover type in the area, covering 54.42% of the district's land area, and 60.3% of the wildfires in the area are experienced in grasslands. Kgalagadi is semi-arid with a relatively higher wildfire vulnerability since fuels are dry for most of the months during the year [20]. The area is predominantly a Wildlife Management Area (WMA), with most of the land area (63%) being covered by the Kgalagadi Transfrontier Park (KTP) and WMAs (Figure 1). Pastoral farming and ranching are also prominent in the district, with many ranches in the Eastern part of the district [33]. The low rainfall (<300mm annually) and low soil fertility inhibit successful arable farming; hence livestock production is the mainstay in the Kgalagadi district. Farmers rely on the grasslands to feed their livestock [33].

2.2. Data collection

This study developed a wildfire prediction model for the Kgalagadi District using LFMC, LST, SM, DMP, and DFMC as independent variables. Active fire points acquired between 2015 and 2021 were used to train and test a wildfire prediction model using a Random Forest (RF) classifier. The dependent variable was a combination of active fire points and non-fire points.

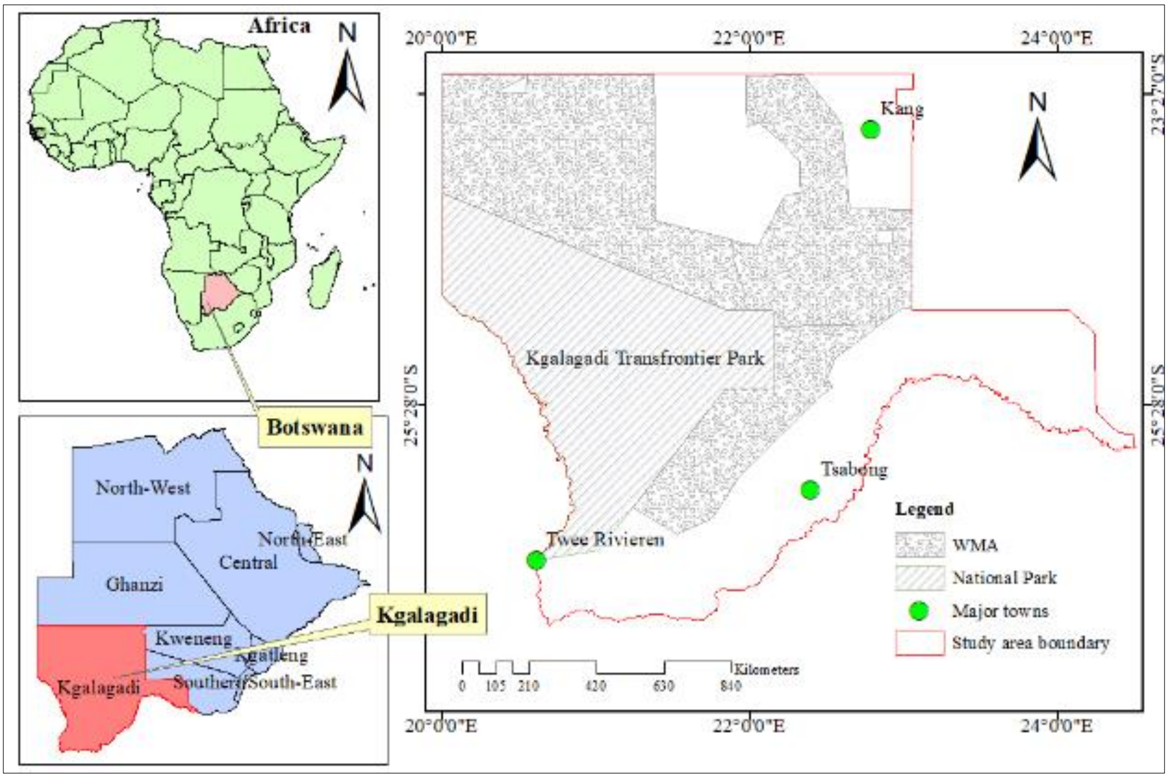

Figure 1 Study area location, Wildlife Management Area (WMA), National Park; the inset is a map indicating the location of Botswana in Africa

2.2.1. Dependent variable

Active fire points obtained using Visible Infrared Imaging Radiometer Suite (VIIRS) sensors at 375km resolution were obtained from the National Aeronautics and Space Administration (NASA) Fire Information and Resource Management System (FIRMS).

Table 1 Number of active fire points recorded in Kgalagadi between 2015 and 2021 (Source: FIRMS website)

Year	Number of fire points
2015	1210
2016	715
2017	17,154
2018	6,679
2019	382
2020	7,666
2021	65,845
Total	99651

The VIIRS data is processed by the University of Maryland in the United States of America (USA) using the standard quality Thermal Anomalies / Fire locations. The wildfire data was presented and supplied by FIRMS as point data containing the location (latitude and longitude coordinates), date, and time of capture. For this study, 99,651 random VIIRS fire points were used (Table 1).

A round 1000 m radius buffer zone was created around fire points for each year to avoid non-fire points from being created very close to the fire points, and non-fire points that fell within the buffer zone were excluded. The study applied the double random principle of time and space and randomly assigned dates and times of occurrence from the fire seasons to the randomly created fire points [25, 34].

All fire and non-fire points with missing values were excluded from the final dataset used for the study. The overall dataset contained fire points (n=80,860) and non-fire points (n=76,965). For analysis purposes, the points were assigned 1 and 0 for the fire and non-fire points, respectively. The final dependent variable for RF analysis contained 107,883 points. The study split the dataset into the training dataset (70% of 157, 825 = 107,883 observations) and the testing dataset (30% of 157, 825 = 53,942 observations), as commonly applied in most machine learning studies [35, 36].

2.2.2. Independent variables

The independent variables used in this study included DMP, SM, LFMC, and DFMC, obtained for the study period (2015-2021). Table 2 provides a detailed description of the independent variables used in this study.

Table 2 Description of independent variables

Independent variable	**Description**
Soil moisture (SM)	The Soil Moisture Active Passive (SMAP) surface soil moisture (0-0.05m depth) was used for fire prediction. Soil moisture data were acquired from NASA National Snow and Ice Data Centre Distributed Active Archive Centre (https://nsidc.org/data/data-programs/nsidc-daac). The SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture V003 was used for the study [37].
Dry Matter Productivity (DMP)	The DMP data was downloaded from the Copernicus Global Land Service site (https://land.copernicus.eu/global/) at a temporal resolution of 10 days and spatial resolution of 300m. The data aligned well with the land cover data obtained from the ESA World Cover project 2020 at a 10m resolution [38], with the lowest DMP values in bare areas, while the shrubland indicated the highest DMP for the study period (Figure 3).
Land surface temperature (LST)	The Moderate Resolution Imaging Spectroradiometer (MODIS) 1km resolution daily Land Surface Temperature/Emissivity (MOD11A1 v61) data downloaded from the NASA MODIS site (https://modis.gsfc.nasa.gov/data/dataprod/mod11.php). The MODIS Terra LST day data was used due to the 10:30 am overpass with clear sky compared to Aqua with 1:30 pm overpass time [39, 40]. The MOD11A1 has been applied in several other fire-related studies [39, 41] and other environmental studies[42, 43]. The MODIS-derived LST has been reported to correlate significantly with Landsat 8-derived LST with a Root Mean Square Error (RMSE) of 1.19K [44]. Similarly, an RMSE of 2.44K and bias of 1.43K were indicated for MODIS LST collection 6 data compared with in-situ station data in the Kalahari Desert [45].
Live Fuel Moisture Content (LFMC)	The empirical model Chuvieco et al. [47] proposed to compute live fuel moisture content for grasslands was applied since (60.3%) of the fires occur in the grasslands. $LFMC_g = -57.103 + 284.808 \times NDVI - 0.089 \times LST + 136.75 \times FD_g$ Equation 1 Where; NDVI is the Normalised Difference Vegetation Index obtained from the Copernicus Global Land service website (https://land.copernicus.eu/global/products/ndvi [46]. FDg is the function of the day was derived from Equation 2 [47], which accounts for the seasonal LFMC variations[24,25]; $FD_g = \left(\sin \left(1.5 \times \pi \times \frac{Dy + Dy^{\frac{1}{3}}}{365} \right) \right)^4 \times 1.3$ Equation 2 Where; Dy is the day of the year.
Dead Fuel Moisture Content (DFMC)	DFMC was estimated using the regression model proposed by Zormpas et al. [49]. Daily MODIS band 20 Brightness Temperature (BT) 1km data were obtained from the USGS Earth data site (https://appeears.earthdatacloud.nasa.gov/) [50]. The data was clipped to the study area, converted to degrees Celsius using ArcMap's raster calculator spatial analysis tool, and then

	daily DFMC values were then estimated using the equation (Equation 3). $DFMC = 19.832 - 0.4 \times BT$ Equation 3 Where BT is the brightness temperature

2.3. Data pre-processing

All datasets were re-sampled to 1000m spatial resolution to ensure consistent spatial resolution. The mean soil moisture was estimated for each 10-day decadal using the daily data. The soil moisture values for the study area from the SMAP ranged between 0.02cm^3/cm^3 to 0.5cm^3/cm^3 aligning with those indicated for SMAP-sentinel active-passive soil moisture retrievals [37] (Figure 2). The 10-day decadal DMP data was processed using ArcMap, and DMP values ranged from 0 kg/ha/day to 327 kg/ha/day (Figure 3). The daily LST datasets were converted to degrees Celsius using ArcMap's raster calculator spatial analysis tool (**Error! Reference source not found.**).

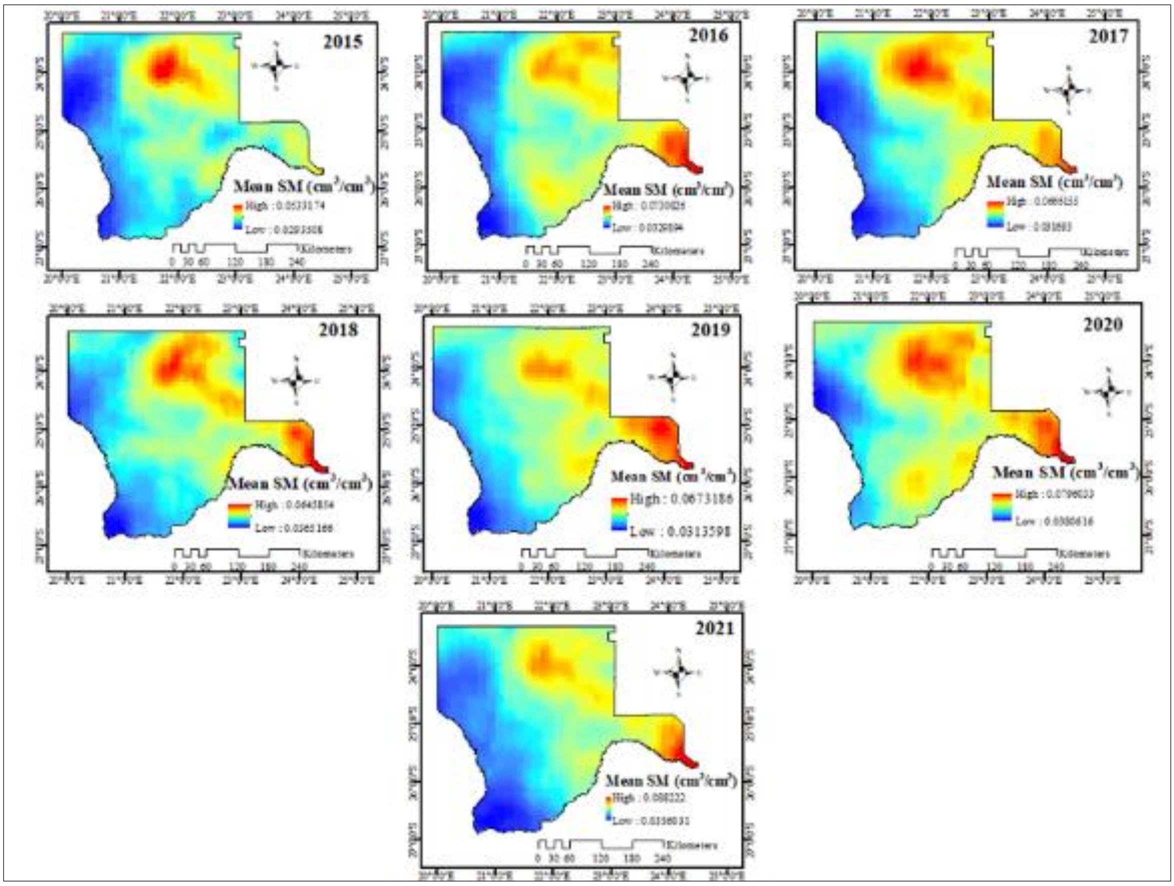

Figure 2 Thematic maps for the study area indicating the annual mean SM during the study period (2015-2021)

10-day decadal LFMC was estimated using ArcMap's raster calculator spatial analysis tool and obtained values between -50 to 350%, with most of the values falling between 0 and 200%. The negative LFMC values obtained were due to the very low NDVI for the study area, with NDVI values below 0.1 [47]. The daily DFMC estimates were determined using ArcMap, and areas covered with shrubs in the northeastern region had the highest DFMC (up to 7.6%) than the Southwestern region with bare lands, with negative DFMC values due to the high BT. Values of all the independent variables for the day/10-days decadal before the fires were extracted to the fire points and non-fire points in ArcMap and then exported to Microsoft Excel spreadsheets.

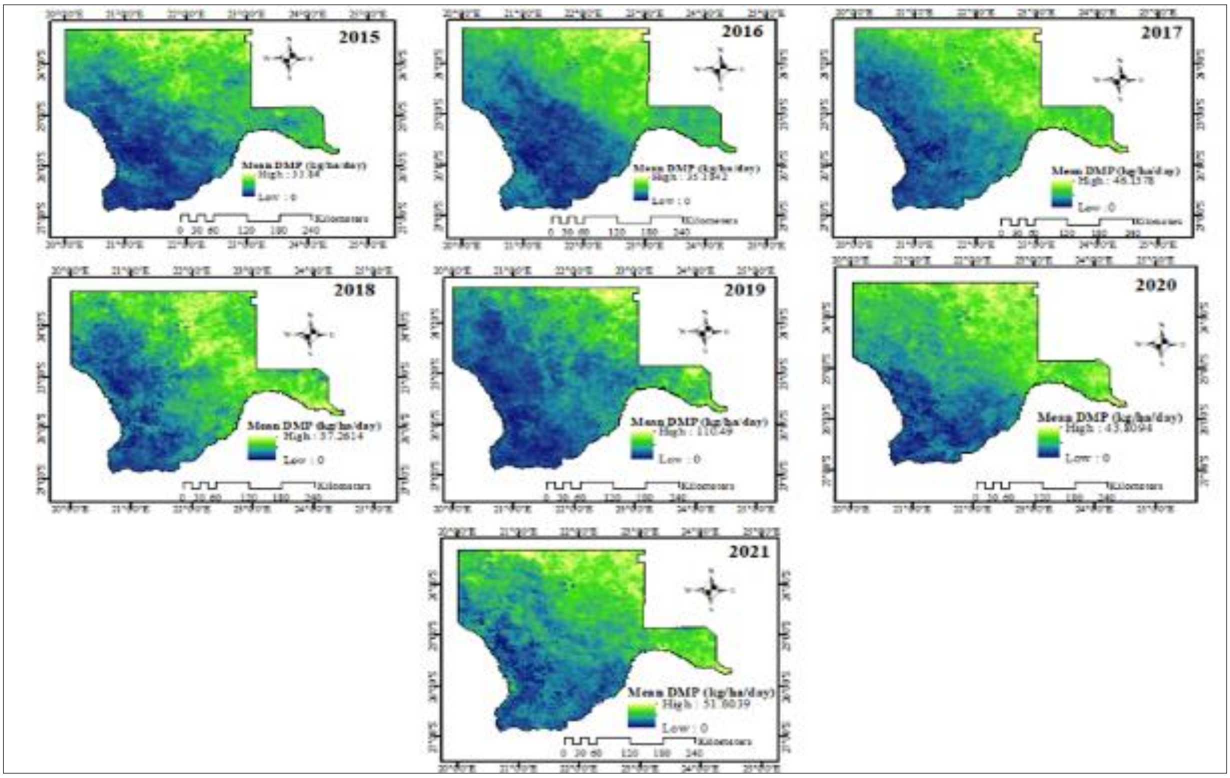

Figure 3 Dry matter productivity (kg/ha/day) variation for the study period 2015 to 2021

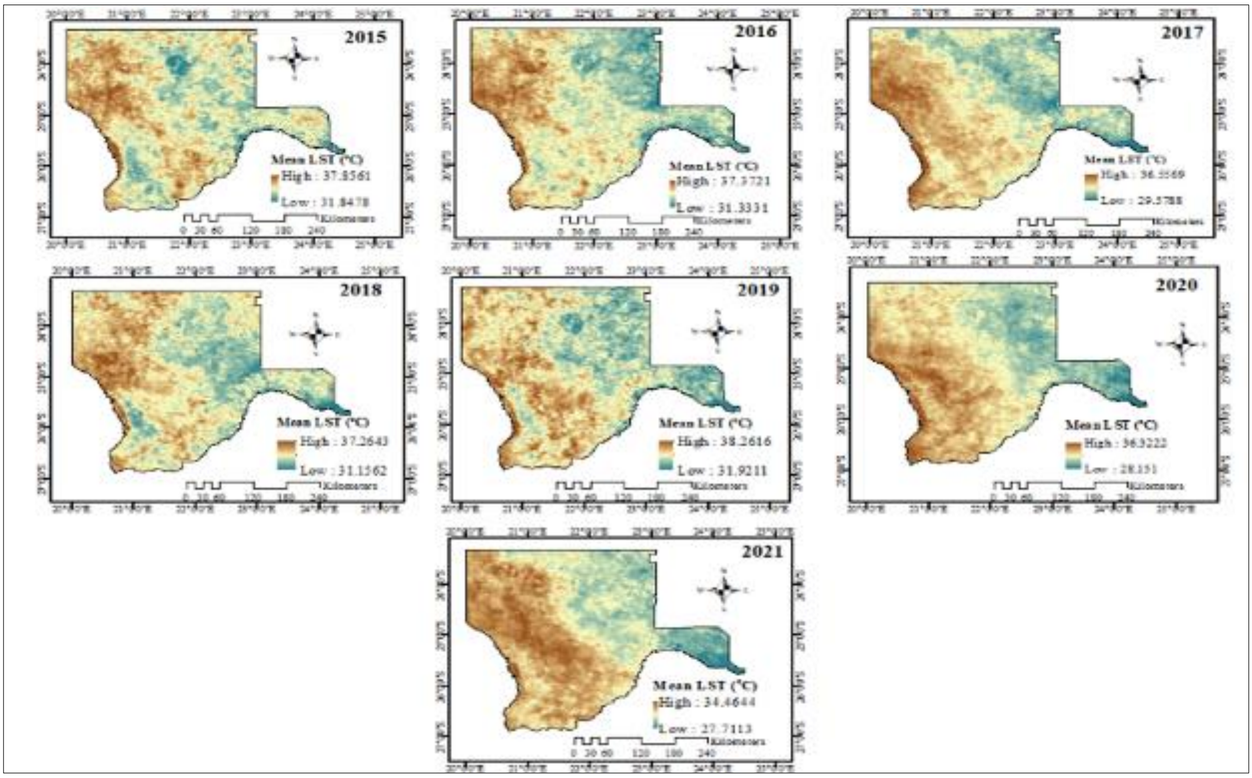

Figure 4 Annual mean Surface temperature during the study period (2015-2021)

2.4. Model training and testing

The Random Forest Classifier was used to train and classify the wildfire prediction model. The classifier was trained using the training dataset, and the testing dataset was used for the testing.

2.4.1. Random Forest model setup

The Random Forest (RF) classification algorithm in the R project for statistical computing software was used to predict wildfire occurrence in the study area using bootstrap samples drawn with replacement. The algorithm is designed to retain about one-third of the samples for validation, referred to as the Out-Of-Bag (OOB). At each node in the tree, the RF algorithm randomly samples some of the predictor variables referred to as 'mtry' to produce the best split for each predictor variable. The number of trees (mtree) and the number of variables at the nodes (mtry) are hyper-parameters. However, it is recommended that the mtry be the square root of the number of variables ($\sqrt{P}$ where P is the number of variables) [51].

RF also calculates the variable importance (VI) of the predictor variables by calculating the OOB error for each tree (t) and permuting each variable (X^j). In contrast, the other variables in the OOB data are left unchanged, and the OOB error (errOOB) is calculated in the permuted dataset [25, 52].

$$VI(X^j) = \frac{1}{ntree}\sum_t (err^{/}OOB_t^j - errOOB_t) \text{ ....Equation 4}$$

Where ntree is the number of trees in the forest and ∑ indicates the sum of all trees, for this study, RF classification was used; thus, the OOB error suggests the rate of misclassification by the forest. The aim is to minimise the OOB error by the RF algorithm. The RF can then be used to select contributing variables in the model.

2.4.2. RF model training

The *randomForest* package in the R project for statistical computing was used for implementing the Random Forest classification algorithm [53]. The caret package was used to streamline the training process. In this study, several trials were conducted with varying numbers of trees, and the optimal number of trees (ntree) was set at 900 with a mtry of 3 (Figure 4).

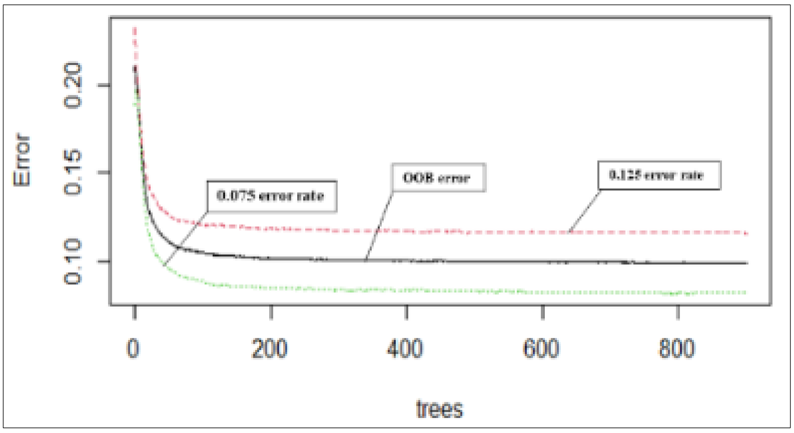

Figure 4 RF plot for the number of trees against the error rate

2.4.3. RF Model testing

The model was assessed using the test dataset, and the variable importance was obtained for each variable used. The model's overall accuracy, kappa coefficient, and user's and producer's accuracy were determined using the test results of the RF model. Mcnemar's Test P-Value was used to determine if the performance of the fire and non-fire point prediction from the model were equal.

2.4.4. Variable importance

A discriminant analysis was run within the RF to rank the importance of variables using the Mean Decrease Accuracy (MDA) and the Mean Decrease Gini (MDG). The RF model's final output was the predictor variables' relative importance. The variable importance was used to determine the groups of predictor variables that give better wildfire predictions. The variables with higher fire predicting power were then assessed for prediction accuracy by running RF models using the different variable combinations.

A summary of the Step by step workflow for identifying potential predictors is indicated in Figure 5 below.

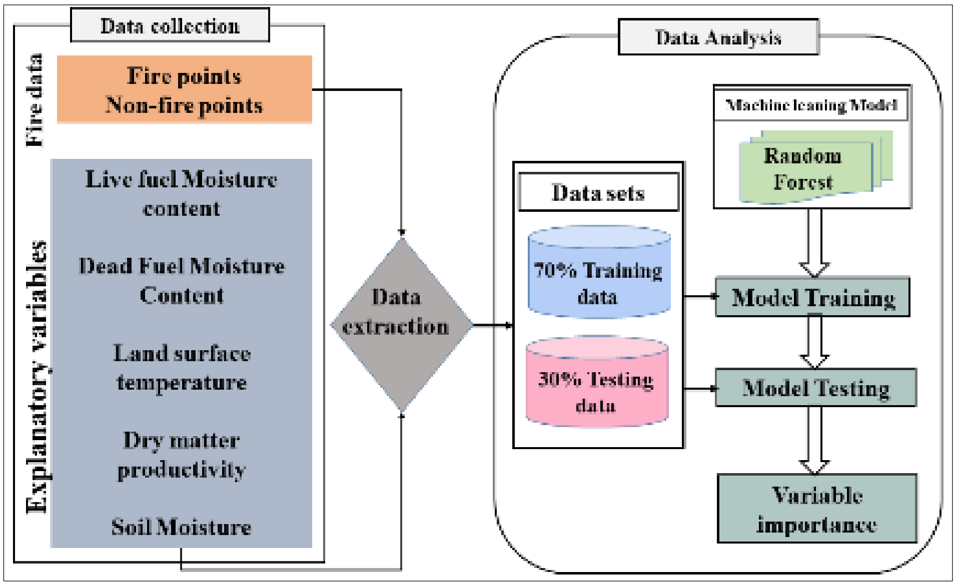

Figure 5 Workflow for wildfire prediction using remotely sensed variables

3. Results and discussions

3.1. Model training and testing

A confusion matrix and accuracy metrics in Table 3 below indicate the model predictions and actual outcome of the training and testing datasets used. Overall results showed an OOB error rate of 9.91%, which means a reasonably good model for predicting wildfires in the study area. The model was tested using a randomly selected testing dataset, and the different statistics from the testing are indicated in Table 3 and 4.

Table 3 Confusion matrix for RF classification model testing, the class errors, accuracy statistics, and overall error of the RF classification model

Prediction			
	Event	**Fires**	**Non-fires**
Reference	Fires	24 816	2 137
	Non-fires	3 177	23 812
Accuracy Metrics	Producer's Accuracy	0.8823	0.9207
	User's Accuracy	0.9176	0.8865
	Overall accuracy	0.9015	
	Kappa	0.803	
	Overall OOB error	9.91%	

Results indicate a substantial agreement between the fire and non-fire observers, with an overall kappa statistic of 0.803 (Table 3). The Kappa statistic obtained in this study is almost in perfect agreement. It is comparable to and higher than earlier fire prediction studies such as Santos et al. [54], who reported a substantial Kappa value of 0.65 for a RF fire prediction model of Minas Gerais, Brazil (2010). Le et al. [55] also found a 0.63 kappa value for their proposed deep neural computing model for predicting wildfires in tropical Vietnam. The results from this study indicate a promising and reliable RF model for predicting wildfires in the Kgalagadi District.

Although the RF model correctly classified 92.07% of the reference non-fire points, only 88.65% were identified as non-fire points by the classification model. In addition, the model achieved a user's accuracy of 91.76% for points classified as fires despite a lower producer's accuracy of 88.23% with 11.77% of fire points classified as non-fire points (Table 4). There was a significant difference (Mcnemar's Test P-value<0.05) between the prediction of fires and non-fire points by the model, indicating that it performs differently for the two classes (Table 4). The model exhibited a high probability of correctly predicting fires as real fires, indicated by the high User accuracy (Table 3). Tonini et al. [29] attribute the power of the RF models to discriminate burned areas in 75% of their study period in Greece to the good generalisation capabilities of the models.

Table 4 Statistics calculated from testing the RF model using the testing dataset

Statistic	Value
P-Value [Acc>NIR]	<0.001
Mcnemar's Test P-Value	<0.001
Detection rate	0.4414
Detection Prevalence	0.5003
Balanced Accuracy	0.9021

The study observed a Detection prevalence of 50.03% of the total predictions, which shows the number of positive events (correctly and incorrectly classified fires). The study found the detection rate to be 44.14% of the predictions, which indicates the fraction of points classified as real fires. The RF model showed a balanced fire prediction accuracy of 90.21%. The high balanced accuracy indicates the classifier's high sensitivity (a large proportion of correctly predicted fire points) and specificity (a large proportion of correctly predicted non-fire points). The performance of the RF model agrees with earlier studies that also showed high accuracies (>70%)of the RF model in wildfire studies [24, 25, 56, 57]. The model accurately predicted fire and non-fire points (Table 3), which indicates the RF model's high reliability and accuracy in predicting fire occurrences using the predictor variables. The accuracy is in the range reported by other authors for predicting fires by using different variables. For example, Karimi et al. [58] reported more than 80% accuracy when they used six vegetation indices derived from MODIS data to predict fire hazards in Golestan forests in Iran.

3.2. Variable importance in classification

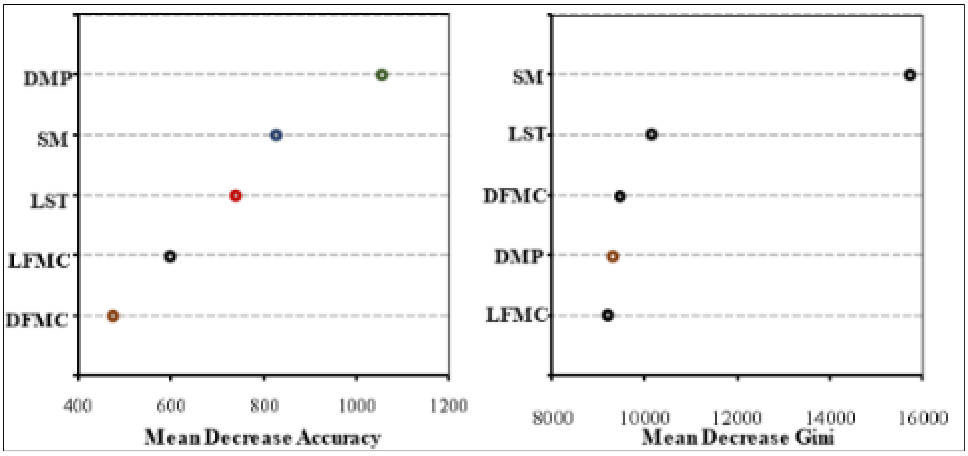

Figure 6 The MDA and MDG plots indicate the relative predictor variable importance calculated by the RF model. DMP-Dry Matter Productivity, SM-soil moisture, LFMC-Live Fuel Moisture Content, LST-Land Surface Temperature, and DFMC-Dead Fuel Moisture Content

The lack of dense meteorological stations or networks in Sub-Saharan African countries limits the meteorological factors' use in wildfire prediction studies, yet forest and vegetation composition maps use [59] are impossible in Botswana due to the lacking fuel maps. This study used validated remotely sensed global products such as DMP, SM, LFMC, DFMC, and LST to predict wildfires in Kgalagadi District. The RF analysis outcome was the relative importance of the wildfire prediction factors used in training the model. The variable importance increases with the magnitude of the values, as shown in

Figure **6** and Table 5.

DMP and surface SM were the most essential variables in predicting wildfires in the study area, with MDA and MDG greater than 800, respectively (

Figure **6**). LFMC and DFMC were the least important factors in wildfire prediction, with MDA and MDG of 600.28 and 9,208.96 and 478.43 and 9480.72, respectively. Noteworthy, despite the higher MDA (1055.20) observed for DMP, SM has a higher MDG (15745.69) than all predictor variables, followed by LST (10169.40). Overall, all variables were significantly important in predicting fire points than non-fire points, with DMP and SM having higher variable importance for classifying fires. Discriminant analysis results indicate that a combination of SM and DMP gives better predictions (80.22% accuracy) than LST, LFMC, and DFMC, with 73.45% prediction accuracy. The overall order predictor variable significance in predicting wildfires was DMP> SM> LST> LFMC> DFMC. These results differ from earlier studies showing LFMC and DFMC as critical variables in wildfire prediction in different environments [60, 61].

Table 5 Variable importance from the RF model for prediction of fire and non-fire points

	Non-fires	Fires	Mean Decrease Accuracy	Mean Decrease Gini
DMP	282.732	999.744	1 055.1978	9 328.616
LFMC	165.318	503.436	600.2813	9 208.956
LST	212.724	634.351	740.1304	10 169.398
SM	163.969	753.415	828.39	15 745.69
DFMC	282.584	484.285	478.4341	9 480.715

Fuel quantities available to burn any time are fundamental in successful wildfire prediction studies, yet quantifying fuels remains entirely labour-intensive. These results indicate that 1055.1978 additional points would be misclassified by the model with a reduction of 9328.616 in the purity of the decision tree nodes if DMP is removed (Table 5). In the arid Kgalagadi district, most of the fuel produced during the summer rains becomes dry immediately after the rainy season, increasing ignition potential during the August to November fire season, explaining the higher contribution of DMP to wildfire prediction by the RF model. Despite the limited use of the remotely sensed DMP product in wildfire prediction, the results indicate considerable potential for its use and application for mapping wildfire danger. Evidence from earlier long-term field studies and satellite-based studies also bespeak the substantial contribution of fuel availability to fires in the southern African dry grassland savannahs [62]. The cumulative DMP could also be used for identifying areas with significant fuel accumulation before the fire season for timely fire management activities to be carried out to prevent the effects of severe and mega-fires.

The availability of large quantities of dry matter produced during the rainy season and low SM seems to be a good recipe for wildfire ignition in the study area. The use of SM content in wildfire prediction has been suggested by several studies [63–65]. Results from this study agree with earlier studies that indicate the use of SM in wildfire prediction, with soil moisture having the highest (15,745.69) MDG of all parameters (Table 5). The high MDG indicates that SM has the highest contribution to the leaf nodes' purity at the decision tree's end. The substantial contribution of SM to prediction in the model is attributed to its effect on the fuel moisture contents, as shown in earlier studies [66, 67]. Rakhmatulina et al. [66] found that SM was the most critical environmental parameter in wildfire prediction in the Sierra Nevada. Every 1% increase in soil moisture resulted in a 0.6% increase in fuel moisture content [66]. The increasing availability of remotely sensed soil moisture data increased the possibility of using soil moisture as a wildfire danger prediction variable. However, there is a lack of remote sensors capable of capturing soil moisture data across large spatiotemporal domains [64]. Improving the availability of higher-resolution soil moisture data could help improve the prediction accuracy of wildfire danger. The LST was the third most important variable in predicting fires, with variable importance of 634.35 (Table 5). Adding LST to SM and DMP combination improved the prediction accuracy by 3.61%. The results from the RF model agree with those found by Bisquert et al. [68], indicating LST to be an essential factor in forest fire danger prediction using Artificial Neural networks (ANN) and Logistic Regression (LR). Adding the day of the year improved the performance of LST in fire prediction by separating high summer temperatures from winter [68]. Other studies have also applied the LST and LST anomalies in wildfire studies, arguing that higher LST and LST anomalies could indicate vegetation stress, which is a crucial indicator of fire danger and ignition [30, 69, 70]. The strong performance of LST in this study could be attributed to the fact that most fires occur after winter with increasing temperatures in the spring and summer seasons while, at the same time, the vegetation is generally dry. The high surface temperature during fire season could account for the increase in the purity of the decision tree nodes (MDG= 10,169.40) when LST is added.

Results from the RF model also indicated that live fuel moisture content was the fourth important variable, with a mean decrease accuracy of 600.28, showing a lower increase in misclassification if LFMC is removed. On the other hand, DFMC had a minor contribution to fire prediction by the RF model, with the lowest MDA of 478.43 (Table 5). However, adding DFMC to SM and DMP combination slightly improved prediction accuracy by 4.43%, while LFMC had a negative effect on the prediction accuracy. Fuel moisture content is the most used and studied fuel characteristic in wildfire danger rating systems and studies [60, 71]. This study applied the model proposed by Chuvieco et al. [47] to estimate LFMC using LST and NDVI. The low contribution of LFMC compared to DMP and SM could be attributed to the fact that the fraction of the day of the year in the proposed model is specific to Mediterranean areas, which somehow vary differently in the FMC across the year. Therefore, field studies are necessary to map fuel moisture contents to improve their performance in detecting wildfires in the study area.

4. Conclusion

Potential remotely sensed variables that could be used to predict forest fires were investigated using a Random Forest classifier based on VIIRS data from 2015 to 2021. The model exhibited excellent accuracy (OOB accuracy rate of 90.09%, Kappa of 80.3, and overall accuracy of 90.15%) in classifying fire and non-fire points. The RF model showed that DMP and SM are strong in predicting rangeland fires with MDA of 1,055.20 and 828.39 and MDG of 9.328.62 and 15,745, respectively. In contrast, LFMC and DFMC were weak in predicting rangeland fires, with MDA of 600.28 and 478.43 and MDG of 9,208.96 and 9480.72, respectively. The order of variable importance for predicting fire points was DMP> SM> LST> LFMC> DFMC with variable importance of 999.744, 753.415, 634.351, 503.436, and 484.285, respectively. The results of this study provide a possibility of using satellite-derived environmental variables to predict rangeland fire occurrence in the Kgalagadi District. It is recommended that field-based calibration and validation of fuel moisture content be carried out to improve their contribution to accurate fire prediction.

Compliance with ethical standards

Acknowledgments

The authors acknowledge the support of the Mobility of African Scholars for Transformative Engineering Training (MASTET) program.

Disclosure of conflict of interest

All authors declare no conflict of interest.

References

[1] Chuvieco E, Mouillot F, van der Werf G R, San Miguel J, Tanase M, Koutsias N, García M, Yebra M, Padilla M, Gitas I, Heil A, Hawbaker TJ and Giglio L. (2019). Historical background and current developments for mapping burned area from satellite Earth observation. Remote Sensing of Environment, 225, 45–64.

[2] Giglio L, Boschetti L Roy DP, Humber ML, Justice CO. (2018). The Collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of Environment, 217, 72–85.

[3] Ramo R, Roteta E, Bistinas I, van Wees D, Bastarrika A, Chuvieco E, van der Werf GR, (2021). African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data Proceedings of the National Academy of Sciences 118 9.

[4] Mapiye C, Mwale M, Chikumba N, Chimonyo M. (2008). Fire as a rangeland management tool in the savannahs of southern Africa: A review. Tropical and Subtropical Agroecosystem 8, 115–124.

[5] CCES. (2021). Wildfires and Climate Change Center of Climate and Energy Solutions from https://www.c2es.org/content/wildfires-and-climate-change/ accessed 19-12-2021.

[6] Halofsky JE, Peterson DL, Harvey BJ. (2020). Changing wildfire changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest USA. Fire Ecology 16, 1 4.

[7] Thompson MP and Calkin DE. (2011). Uncertainty and risk in wildland fire management: A review. Journal of Environmental Management, 92, 8, 1895–1909.

[8] Nhongo EJS, Fontana DC, Guasselli LA and Bremm C. (2019). Probabilistic modelling of wildfire occurrence based on logistic regression Niassa Reserve Mozambique. Geomatics, Natural Hazards and Risk, 10, 1, 1772–1792.

[9] Moyo T, Musakwa W, Nyathi NA, Mpofu E and Gumbo T. (2020). Modelling of natural fire occurrences: a case of South Africa. The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences (XLIII-B3-2020), Presented at the XXIV ISPRS Congress Commission III (Volume XLIII-B3-2020) - 2020 edition Copernicus GmbH 1477–1482.

[10] Cassidy, L Perkins J and Bradley J. (2022). Too much, too late: fires and reactive wildfire management in northern Botswana's forests and woodland savannas. African Journal of Range & Forage Science, 39, 1, 160–174.

[11] Andela N, van der Werf GR. (2014). Recent trends in African fires driven by cropland expansion and El Niño to La Niña transition Nature Climate Change, 4, 9, 791–795.

[12] Abolafia-Rosenzweig R, He C and Chen F. (2022). Winter and spring climate explains a large portion of interannual variability and trend in western US summer fire burned area. Environmental Research Letters, 17, 5 054030.

[13] Aragoneses E and Chuvieco E. (2021). Generation and Mapping of Fuel Types for Fire Risk Assessment Fire, 4, 3, 59.

[14] Cruz MG, Cheney NP, Gould JS, McCaw WL, Kilinc M, and Sullivan AL. (2021). An empirical-based model for predicting the forward spread rate of wildfires in eucalypt forests International Journal of Wildland Fire, 31, 1, 81–95.

[15] Newnham G, Opie K and Leonard J. (2017). A methodology for State-wide mapping of annual fuel load and bushfire hazard in Queensland (csiro: EP175130) CSIRO.

[16] Statistics Botswana. (2018). Botswana Environment Statistics Natural Disasters Digest 2017 Statistics Botswana Gaborone Botswana.

[17] Dube OP. (2013). Challenges of wildland fire management in Botswana: Towards a community inclusive fire management approach. Weather and Climate Extremes, 1, 26–41.

[18] Chuvieco E, Giglio L and Justice C. (2008). Global characterisation of fire activity: toward defining fire regimes from Earth observation data. Global Change Biology, 14, 7, 1488–1502.

[19] GWIS. (2022). GWIS - Statistics Portal Global Wildfire Information System from https://gwis.jrc.ec.europa.eu/apps/gwis.statistics/estimates accessed 16-3-2022

[20] Maabong KE and Mphale K. (2021). Wildfires in Botswana and Their Frequency of Occurrence. Atmospheric and Climate Sciences, 11, 04, 689–696.

[21] World Bank Group. (2022). Climate knowledge Portal. For Development Practitioners and Policymakers from https://climateknowledgeportal.worldbank.org/ accessed 4-10-2022

[22] Anderson K, Reuter G and Flannigan M D. (2007). Fire-growth modelling using meteorological data with random and systematic perturbations. International Journal of Wildland Fire, 16, 2, 174–182.

[23] Bergonse R, Oliveira S, Gonçalves A, Nunes S, DaCamara C, Zêzere JL. (2021). Predicting burnt areas during the summer season in Portugal by combining wildfire susceptibility and spring meteorological conditions. Geomatics Natural Hazards and Risk, 12, 1. 1039–1057.

[24] Latifah AL. Shabrina A. Wahyuni IN and Sadikin R. (2019). Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo 2019. International Conference on Computer Control Informatics and Its Applications (IC3INA). Presented at the 2019 International Conference on Computer Control Informatics and its Applications (IC3INA), 4–8.

[25] Su Z, Hu H, Wang G, Ma Y, Yang X, and Guo F. (2018). Using GIS and Random Forests to identify fire drivers in a forest city Yichun, China. Geomatics Natural Hazards and Risk, 9, 1, 1207–1229.

[26] Valdez MC, Chang KT, Chen CF, Chiang SH and Santos JL. (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.

[27] Dinku T. (2019). Chapter 7 - Challenges with availability and quality of climate data in Africa A. M. Melesse; W. Abtew; G. Senay (Eds.) Extreme Hydrology and Climate Variability. Elsevier, 71–80.

[28] Ifejika Speranza C, Akinyemi FO, Baratoux D, Benveniste J, Ceperley N, Driouech F and Helmschrot J. (2022). Enhancing the Uptake of Earth Observation Products and Services in Africa Through a Multi-Level Transdisciplinary Approach Surveys in Geophysics.

[29] Tonini M D, Andrea M, Biondi G, Degli Esposti S, Trucchia A and Fiorucci P. (2020). A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. Geosciences, 10, 3, 105.

[30] Chaparro D, Vall-llossera M, Piles M, Camps A, Rüdiger C and Riera-Tatché R. (2016). Predicting the Extent of Wildfires Using Remotely Sensed Soil Moisture and Temperature Trends. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9, 6, 2818–2829.

[31] Singh KR, Neethu KP, Madhurekaa K, Harita A and Mohan P. (2021). Parallel SVM model for forest fire prediction. Soft Computing Letters, 3, 100014.

[32] Chang C, Chang Y, Xiong Z, Ping X, Zhang H, Guo M and Hu Y. (2022). Model Comparisons for Predicting Grassland Fire Occurrence Probability in Inner Mongolia Autonomous Region China (preprint)Other Hazards (e.g. Glacial and Snow Hazards Karst Wildfires Hazards and Medical Geo-Hazards).

[33] Kgosikoma OE, and Batisani N. (2014). Livestock population dynamics and pastoral communities' adaptation to rainfall variability in communal lands of Kgalagadi South Botswana. Pastoralism, 4, 1, 19.

[34] Guo F, Su Z, Tigabu M, Yang X, Lin F, Liang H and Wang G. (2017). Spatial Modelling of Fire Drivers in Urban-Forest Ecosystems in China. Forests, 8, 6, 180.

[35] Pang Y, Li Y, Feng Z, Feng Z, Zhao Z, Chen S, Zhang H. (2022). Forest Fire Occurrence Prediction in China Based on Machine Learning Methods. Remote Sensing, 14, 21, 5546.

[36] Sharma L K, Gupta R, Fatima N. (2022). Assessing the predictive efficacy of six machine learning algorithms for the susceptibility of Indian forests to fire. International Journal of Wildland Fire, 31, 8, 735–758.

[37] Das N, Entekhabi RS, Dunbar RS, Kim S, Yueh S, Colliander A, O'Neill PE, Jackson T, Berg D, Bosch D, Walker J, Caldwell T and Cosh M. (2020). SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture Version 3. NASA National Snow and Ice Data Center Distributed Active Archive Center Boulder Colorado USA.

[38] Zanaga D, Van De Kerchove R, De Keersmaecker W, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li L, Tsendbazar NE, Ramoino F and Arino O. (2021). ESA WorldCover 10 m 2020 v100 Zenodo.

[39] Butt EW, Conibear L, Knote C and Spracklen DV. (2021). Large Air Quality and Public Health Impacts due to Amazonian Deforestation Fires in 2019. GeoHealth, 5, 7.

[40] Wan Z, Hook S and Hulley G. (2021). MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061 [Data set] NASA EOSDIS Land Processes DAAC

[41] Freeborn PH, Jolly WM, Cochrane MA and Roberts G. (2022). Large wildfire-driven increases in nighttime fire activity observed across CONUS from 2003–2020 Remote Sensing of Environment, 268, 112777.

[42] Xie C, Zhang X, Zhuang L, Zhu R and Guo J. (2022). Analysis of surface temperature variation of lakes in China using MODIS land surface temperature data. Scientific Reports, 12, 1, 2415.

[43] Roznik M, Brock Porth C, Porth L, Boyd M and Roznik K. (2019). Improving agricultural microinsurance by applying universal kriging and generalised additive models for interpolation of mean daily temperature. The Geneva Papers on Risk and Insurance - Issues and Practice, 44, 3, 446–480.

[44] Zhang Z and He G. (2013). Generation of Landsat surface temperature product for China 2000–2010. International Journal of Remote Sensing, 34, 20, 7369–7375.

[45] Duan SB, Li ZL, Li H, Göttsche FM, Wu H, Zhao W, Leng P, Zhang X and Coll C. (2019). Validation of Collection 6 MODIS land surface temperature product using in situ measurements Remote Sensing of Environment, 225, 16–29.

[46] Swinnen E Toté C. (2022). Algorithm Theoretical Basis Document Normalized Difference Vegetation Index (NDVI) Col-lection 300m Version 2 Copernicus Global Land Operations

[47] Chuvieco E, Cocero D, Riaño D, Martin P, Martínez-Vega J, de la Riva J and Pérez F. (2004). Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. Remote Sensing of Environment, 92, 3, 322–331.

[48] Castro FX, Tudela A and Sebastià MT. (2003). Modeling moisture content in shrubs to predict fire risk in Catalonia (Spain). Agricultural and Forest Meteorology, 116, 1, 49–59.

[49] Zormpas K, Vasilakos C, Athanasis N, Soulakellis N and Kalabokidis K. (2017). Dead fuel moisture content estimation using remote sensing. European Journal of Geography, 8, 5, 17–32

[50] Boschetti L, Vermote E and Wolfe R. (2015). MODTBGA MODIS/Terra Thermal Bands Daily L2G-Lite Global 1km SIN Grid V006 MODTBGA NASA EOSDIS Land Processes DAAC

[51] Probst P, Wright M and Boulesteix AL. (2019). Hyperparameters and Tuning Strategies for Random Forest. WIREs Data Mining and Knowledge Discovery, 9, 3.

[52] Grömping U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. The American Statistician, 63, 4, 308–319.

[53] R Core Team. (2022). R: A language and environment for statistical computing R Foundation for Statistical Computing Vienna Austria.

[54] Santos EE dos, Sena NC, Balestrin D, Fernandes Filho EI, Costa LM da and Zeferino L B. (2020). Prediction of Burned Areas Using the Random Forest Classifier in the Minas Gerais State. Floresta e Ambiente, 27.

[55] Le HV, Hoang DA, Tran CT, Nguyen PQ, Tran VHT, Hoang ND, Amiri M, Ngo TPT, Nhu HV, Hoang TV and Bui DT. (2021). A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. Ecological Informatics, 63, 101300.

[56] Çömert R, Matci DK and Avdan U. (2019). Object-based burned area mapping with random forest algorithm. International Journal of Engineering and Geosciences, 4, 2, 78–87.

[57] Leuenberger M, Kanevski M, Vega Orozco CD. (2013). Forest Fires in a Random Forest EGU2013-3238Presented at the EGU General Assembly Conference Abstracts.

[58] Karimi A, Abdollahi S, Ostad-Ali-Askari K, Eslamian S and Singh VP. (2021). predicting fire hazard areas using vegetation indexes case study: Forests of Golestan Province Iran. Journal of Geography and Cartography, 4, 1, 1–6.

[59] Shang C, Wulder MA, Coops NC, White JC, Hermosilla T. (2020). Spatially-Explicit Prediction of Wildfire Burn Probability Using Remotely-Sensed and Ancillary Data. Canadian Journal of Remote Sensing, 46, 3, 313–329.

[60] Chuvieco E, González I, Verdú F, Aguado I, Yebra M, Chuvieco E, González I, Verdú F, Aguado I. and Yebra M. (2009). Prediction of fire occurrence from live fuel moisture content measurements in a Mediterranean ecosystem. International Journal of Wildland Fire, 18, 4, 430–441.

[61] Matthews S. (2014). Dead fuel moisture research: 1991–2012. International Journal of Wildland Fire, 23, 1, 78–92.

[62] Lehsten V, Tansey K, Balzter H, Thonicke K, Spessa A, Weber U, Smith B and Arneth A. (2009). Estimating carbon emissions from African wildfires. Bio geosciences, 6, 3, 349–360

[63] Bartsch A, Balzter H and 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.

[64] Sharma S and Dhakal K. (2021). Boots on the Ground and Eyes in the Sky: A Perspective on Estimating Fire Danger from Soil Moisture Content. Fire, 4, 3, 45.

[65] Vinodkumar Dharssi I. (2019). Evaluation and calibration of a high-resolution soil moisture product for wildfire prediction and management. Agricultural and Forest Meteorology, 264 27–39.

[66] Rakhmatulina E, Stephens S, and Thompson S. (2021). Soil moisture influences on Sierra Nevada dead fuel moisture content and fire risks. Forest Ecology and Management, 496, 119379.

[67] Vinodkumar V, Dharssi I, Yebra M and Fox-Hughes P. (2021). Continental-scale prediction of live fuel moisture content using soil moisture information. Agricultural and Forest Meteorology, 307, 108503.

[68] Bisquert MM, Sánchez JM, Caselles V, Bisquert MM, Sánchez JM and Caselles V. (2011). Fire danger estimation from MODIS Enhanced Vegetation Index data: application to Galicia region (north-west Spain). International Journal of Wildland Fire, 20, 3 465–473.

[69] Yang H. (2021). Land Surface Temperature Anomalies and Fire Occurrence ArcGIS Story Maps from https://storymaps.arcgis.com/stories/06a6acff8a544fb187b2cb4ce262e614 accessed 7-4-2022

[70] Yang S. (2021). The Relationship between Land Surface Temperature Anomalies and Fire Occurrence in Cariboo Region in 2017 from http://dx.doi.org/10.14288/1.0396751 accessed 7-4-2022

[71] Jurdao S, Arevalillo JM, Chuvieco E and Yebra M. (2011). Development of a method to transform Life Fuel Moisture Content into ignition probability 3.