Implementation of a Random Forest Classifier to Examine Wildfire Predictive Modelling in Greece Using Diachronically Collected Fire Occurrence and Fire Mapping Data

Conference Paper · June 2021

DOI: 10.1007/978-3-030-67835-7_27

Implementation of a Random Forest Classifier to Examine Wildfire Predictive Modelling in Greece Using Diachronically Collected Fire Occurrence and Fire Mapping Data

Alexis Apostolakis(✉), Stella Girtsou, Charalampos Kontoes, Ioannis Papoutsis, and Michalis Tsoutsos

National Observatory of Athens, Athens, Greece
{alex.apostolakis,sgirtsou,kontoes,ipapoutsis,mtsoutsos}@noa.gr

Abstract. Forest fires cause severe damages in ecosystems, human lives and infrastructure globally. This situation tends to get worse in the next decades due to climate change and the expected increase in the length and severity of the fire season. Thus, the ability to develop a method that reliably models the risk of fire occurrence is an important step towards preventing, confronting and limiting the disaster. Different approaches building upon Machine Learning (ML) methods for predicting wildfires and deriving a better understanding of fires' regimes have been devised. This study demonstrates the development of a Random Forest (RF) classifier to predict "fire"/"non fire" classes in Greece. For this a prototype and representative for the Mediterranean ecosystem database of validated fires and fire related features has been created. The database is populated with data (e.g. Earth Observation derived biophysical parameters and daily collected climatic and weather data) for a period of nine years (2010–2018). Spatially it refers to grid cells of 500 m wide where Active Fires (AF) and Burned Areas/Burn Scars (BSM) were reported during that period. By using feature ranking techniques as Chi-squared and Spearman correlations the study showcases the most significant wildfire triggering variables. It also highlights the extent by which the database and selected features scheme can be used to successfully train a RF classifier for deriving "fire"/"non-fire" predictions over the country of Greece in the prospect of generating a dynamic fire risk system for daily assessments.

Keywords: Fire/non-fire classification · Machine Learning (ML) · Random Forest (RF) · Burned Scar Mapping (BSM) · EFFIS · FIRMS

1 Introduction

Forest fires affect severely the natural and rural ecosystems, human lives, critical infrastructures and assets in general. Last year's bush fires in New South Wales of Australia burned about 1.65 million hectares and claimed the lives of 6 people [1]. In the Amazonian forested ecosystems, the Brazilian Space Agency has reported during August 2019 a significant increase in fire occurrences of the order of 83% compared to the same

J. Lokoč et al. (Eds.): MMM 2021, LNCS 12573, pp. 318–329, 2021.
https://doi.org/10.1007/978-3-030-67835-7_27

period of the previous year. Furthermore, the extreme drought and heat wave events during the summer seasons of 2017 and 2018 have resulted in extraordinary fire events that affected severely the Mediterranean ecosystems [2]. The problem becomes considerably important if we account for the climate change scenarios which suggest substantial warming and increase of heat waves, drought and dry spell events across the entire Mediterranean [3] in the future years. In this regard, the access to validated information on the spatiotemporal patterns of wildfire behavior, from past fire occurrences and fire triggering factors, are of great importance for the implementation of the proper disaster risk mitigation policies.

In this context, a number of fire danger rating and prevention systems have been developed in Europe ingesting weather, topography, fuel, and fire ignition data [4, 5]. As stated in [6, 7] the models which have been used for predicting the fire ignition susceptibility in an area are classified into three groups; theoretical (or physics-based), semi-empirical, and empirical models. Theoretical and semi-empirical models are entirely or partly based on equations that describe the physics of the related to the fire ignition physical phenomena like fluid mechanics, combustion and heat transfer. On the other hand the empirical models are purely based on the statistical correlations between data extracted from historical fire records and their related environmental, biophysical, morphological, fuel, and climatic/weather data. These historical data can be regarded as vectors of variables which are known or believed to influence the ignition of a fire, and provide the essential input knowledge for the model to identify which are the areas and the time periods at high risk. Empirical models, gaining knowledge from long lasting histories of fire events and their triggering parameters have used either statistical correlations or ML methods as in the present study [7, 8].

ML algorithms learn directly from the data and develop their own internal model without being necessary to provide any expert knowledge or simulate precisely the physical parameters that feed the model running. Moreover, ML models detect and automatically formulate the complex mathematical relations that exist between the diverse input parameters and this is an important advantage over the physical-based models where the mathematics of those relations should be known in advance [8].

Regarding the parameters that affect the presence of wildfires, several studies referenced in [9] and [10], show that vegetation proxies as NDVI, topography (altitude, slope, aspect), soil moisture, fuel, and meteorological data are considerably influencing factors for fire risk and fire ignition.

Aiming to advance further the relevant research, this work provides the foundations of a data driven model that covers the Greece's national needs for predicting "fire"/"non fire" prone areas at the enhanced spatial resolution of 500 m. A prototype feature database was generated and organized in the form of a datacube structure, using a nine-year lasting record (2010–2018) of multi-source and multi-sensor data such as EO based essential environmental and ecosystem related indexes, meteo records, fuel classes, morphological features with diverse spatial and temporal resolutions. Furthermore, a Random Forest (RF) algorithm [8], has been parameterized and successfully trained to produce a satisfactory prediction of "fire"/"non fire" class mapping on a daily basis over Greece.

2 Study Area - Training Data Set

2.1 Study Area

The area of interest covers the Greece's territory (131.957 km^2) located in the southeast of the Mediterranean climatic zone, with mild and rainy winters, warm and dry summers and extended periods of sunshine throughout most of the year [11].

A major part of the country, up to 58.8% of the total surface, represents low altitude areas (0–500m) which are prone to fire ignition [10]. The topography and the dominant north winds in combination with the vegetation types in the central and southern parts of Greece are between the prime drivers for fire ignition during the summer period [10]. Last but not least the vegetation cover that makes Greece particularly prone to fire hazard and fire risk such as coniferous and mixed forests, sclerophyllous vegetation, natural grasslands, transitional woodlands, semi natural and pasture areas correspond to approximately 72% of the total surface of the country (source: Copernicus CORINE Land Cover 2018 of Greece) [12]. It is worth noting that Greece consists the typical case of the Mediterranean ecosystem in regard to fire risk and expected damages similar to the ones of France, Italy, Portugal, Spain, and Turkey, with the most severe events being associated with strong winds at lower altitudinal zones ($<$1000 m) during hot dry summer periods [12], and extended heat waves that tend to be longer and drier due to the climate change [13].

The official annual statistics [14, 15] confirm that the average number of fires and the corresponding burned areas in Greece have increased by a factor of 4 during the last decades. This makes obvious that wildfires continue to constitute the major threat for the environment and the society in Greece, often with significant cost in human lives, as in the extreme case of the fire in Mati on July 23, 2018, which claimed the lives of 102 people. This highlights the need for setting and put in operation reliable fire management systems providing knowledge on the contemporary fire regimes and supporting the preparedness and emergency response procedures.

2.2 Data Resources

Forest Fire Inventory. A reliable forest fire inventory is vital for the prediction of wildfires regimes in an area, given that new fire occurrences in the same location are used to happen under similar weather and environmental conditions. For this, an exhaustive forest fire inventory of fire occurrences and burn scar maps was compiled by exploiting diachronic data generated by the FireHub system of BEYOND [16], as well as the NASA FIRMS and the European Forest Fire Information System (EFFIS/JRC) data.

To be noted that the Centre of Earth Observation Research and Satellite Remote Sensing BEYOND of NOA runs the operational system FireHub that offers (a) the so called "Diachronic Burn Scar Mapping service" providing polygons of Burned Areas at high spatial resolution (10 m–30 m) for more than 35 years which cover the entire country (the Diachronic NOA-BSM product), and (b) the "Early Detection and Real Time Fire Monitoring Service" that is systematically archiving for the last 10 years the daily observations of Active Fire (AF) at the spatial resolution of 500 $\times$ 500 m (the AF product) [17]. The database was created by applying the proper masks to exclude areas which are unlikely for fire ignition (e.g. grid cells laid on urban, agriculture and water areas) [18] (Fig. 1).

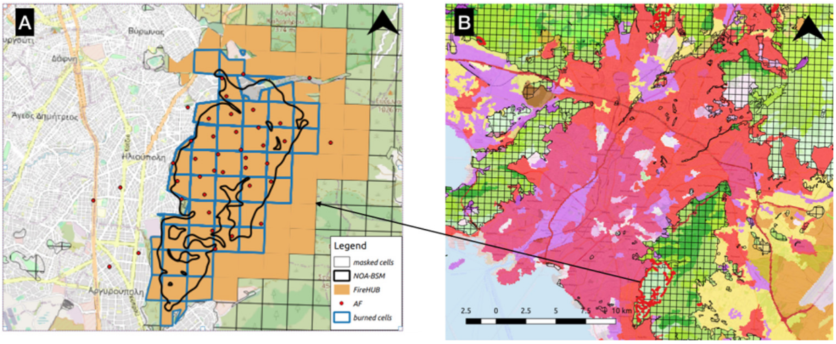

Fig. 1. (A) Blue cells validated as true positives (burned) from NOA-BSM (black polygon), AF (red points) and FireHub spotted active fires (orange cells). (B) Map shows how urban areas where excluded from the current study. (Color figure online)

The process for labeling the grid cells as “fire”/“non-fire” in order to create the training and validation dataset for the prediction algorithm (see Sect. 3), was initially based on FireHub evidences. At this stage, every grid cell intersecting with a BSM-NOA polygon and the corresponding AF detections had been labeled as fire cell. In a following step any remaining AF evidences were checked spatially and temporally against the EFFIS and FIRMS datasets. This resulted in a new set of “fire” cells which were also added in the database and used for the analysis. In conclusion, this process has returned a set of 12.978 “fire” cells which had been reported by independent observations of the three operating systems as FireHub, EFFIS, and FIRMS corresponding to the period 2010–2018. Each fire cell was assigned a unique fire event ID and the corresponding date of fire occurrence.

Additionally, an equivalent dataset of 12.585 “non-fire” cells spanning the same period (2010–2018) was created. The latter dataset was generated through a simple random selection spatially expanding over the entire Greece. It is worth-noting that the “non-fire” cells are recognized as areas where there is no presence of a recent fire outbreak. In practice and as explained in [19] these “non-fire” samples are representing pseudo-absence data because it is undefined if the conditions were in favor of a fire outbreak or not at their locations.

Meteorological Data. The meteorological data were derived from ERA5-Land, a reanalysis dataset providing hourly temporal resolution and native spatial resolution at 9 km. We obtained temperature, wind and precipitation datasets for a total of eight years (2010–2018) and computed the following parameters: maximum temperature, minimum temperature, mean temperature, dominant wind direction, maximum wind speed of the dominant direction, maximum wind speed, wind direction of the maximum wind speed, accumulated precipitation of the past 7 days.

EO Vegetation Data. Several studies consider NDVI as being of paramount importance in wildfires modelling [17]. In this study, the vegetation condition is represented by using the NDVI as proxy (Normalized Differential Vegetation Index) parameter, the latter being produced on 8-day intervals from NASA's MODIS dataset at 500 m spatial resolution. A number of 504 MODIS images were downloaded, stored and processed in order to calculate the NDVI data for the study period.

Topographic and Land Use Data. Topographic and Digital Elevation Model derived parameters are important factors for fire susceptibility prediction and mapping [10]. For the purposes of the study the Copernicus EU-DEM v1.1 at 25 m spatial resolution was used. This DEM was processed for deriving the morphological related features such as slope, aspect and curvature. Previous studies have shown that these features influence significantly the probability for fire occurrence [10]. Moreover, the land cover category assigned to each fire cell was retrieved from the CORINE 2012, for the cells representing areas that were burned before 2014, and the CORINE 2018 if the area was burned after 2014. Moreover, grid cells containing multiple land cover classes, were assigned the class with the highest fire proneness according to [10].

Table 1. Products and features extracted for the training of the model

Product	Source	Spatial resolution	Temporal resolution	Features
Wind (u-comp, v-comp)	ERA5-Land	9 km	Hourly	Dominant dir., Max speed of the dom. Dir., Max wind speed, Wind dir of the max speed
Temperature	ERA5- Land	9 km	Hourly	Max temperature, Min temperature, Mean temperature
Precipitation	ERA5- Land	9 km	Hourly	7-day accumulated precipitation
NDVI	NASA -MODIS	500 m	8-day	NDVI
DEM	Copernicus	25 m	Static	DEM, slope, aspect, curvature
Corine	Copernicus	100 m	6-year	CLC
"fire"/"non-fire" cells	NOA, EFFIS, NASA	500 m	Daily	

3 Methodological Approach and Implementation

3.1 Data Archiving and Modelling

The previous section makes obvious that the feature database has incorporated dynamic (weather/environmental) and less dynamic or static (DEM, land cover) data in multiple spatiotemporal resolutions. The multisource and multi type character of the data, together with its high spatiotemporal diversity and big volume, rendered the file storage and the manipulation of the data layers a challenging process. For this, an innovative earth observation datacube architecture was employed [20] for storing and managing the data and also for extracting and exploiting the features to be analyzed. This datacube is a one of its kind asset; it contains analysis ready spatio-temporal data in a structured and easily accessible architecture, providing a single access point to key environmental parameters that serve as wildfire drivers.

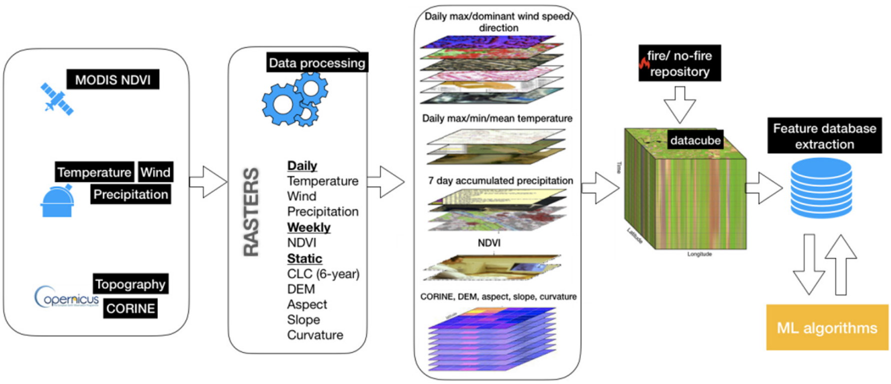

Fig. 2. System's architecture and processing blocks

Figure 2 shows the main blocks of the system's architecture and the corresponding processing steps. It encompasses access to the data, data downloading and manipulation, data ingestion and processing, information layer creation, feature space creation, ML algorithm training, implementation and validation. The Data processing block refers to a multitude of actions for the conversion of raw and/or archived FIRMS, EFFIS, ERA5, FireHub and Copernicus data so as to create the analysis ready data. Actually, this processing block has been designed to invoke a suite of properly developed Python scripts running over the datacube offering flexibility, scalability, reproducibility, transferability and capability of the system to respond to any scale, from local to national.

3.2 Feature Ranking

Guided by the need to validate the quality and adequacy of the feature database for addressing the "fire"/"no fire" (fire presence/absence) classification problem, it was necessary to apply feature ranking and examine the order of correlation between the independent variables, namely the wildfire related features of Table 1, with the dependent variable which was set by two distinct classes: the "fire" class and the "non-fire" class. For this, filter, wrapper, and embedded techniques for feature selection (and ranking) have been applied [21].

Table 1 shows the selected factors influencing fire prediction, the corresponding data sources and the features extracted and used as input to the "fire"/"non-fire" classification problem.

Firstly, for evaluating the categorical features (Corine Land Cover, dominant wind direction and wind direction of the maximum wind speed) the chi-squared filter method based on the Chi-squared statistics test was used [22]. For the remaining of the features (numerical), a correlation filter method based on Spearman's rank correlation coefficient has been used for evaluating the dependencies of the various feature combinations. It is worth noting that the Spearman's correlation has been selected as it can detect linear and non-linear monotonic relations [23] but also the correlations between the ordinal data. Actually the resulted number is ranging in the set $[-1, 1]$ indicating stronger correlation when the number is closer to 1 or -1, and weaker when the number is near to zero.

Furthermore, and in order to allow for a comparative ranking for all the features (numerical and categorical), three more feature ranking methods of the wrapper and embedded type were employed using Random Forest. The selected algorithms were the Sequential Feature Selection [24] using as score metric the Area Under the learning Curve (AUC) [25], the feature ranking through the measurement of the node impurity, and finally the permutation importance [22].

3.3 Random Forest (RF) Algorithm Implementation

As a final step and in order to demonstrate that the quality of the generated feature database is suited for letting a ML algorithm to distinguish between "fire" and "non-fire" classes, a RF model was chosen for implementation [26] and evaluation of its classification output. RF is ideal for this purpose because it is one of the most used and tested algorithms for predicting fire occurrence, susceptibility and risk [8]. Moreover RF is the most common representative of the family of decision trees ensembles which are commonly used in all sorts of classification problems and it can handle directly categorical data like wind direction and Corine Land Cover.

Many tests of the RF classifier were performed and a wide range of results was produced. The experiments with the highest scores were those where the shuffling method was used to create the training and test dataset. Actually, using shuffling provided the training set with indicative "fire" cells from almost every burned area. Considering that burned scars are small regions with very similar prevailing conditions, shuffling makes it easier for the algorithm to recognize fire cells and reduces the model's ability to adapt properly to new, previously unseen data.

In a further analysis step, and in order to increase the difficulty by simulating entirely unknown conditions for the test dataset, the shuffling process was left out. The target was to avoid neighboring "fire"/"non fire" cells, from the same day, and the same fire event, to be included both in the training and test datasets. Thus, following the k-fold cross validation methodology [26, 27] the feature space was split into 10 "folds" under the rule that the cell samples of a specific day cannot be distributed in more than one folds. This approach effectively allowed us to assess the model's capacity to generalize better, in the context of an operational forecasting application.

Furthermore, for tuning the RF model's internal parameters (decision trees number, trees depth, number of features in each tree) a random search hyper-parameterization process was performed [28, 29]. The metric that was selected for maximization was the recall [30] for the fire class, because the target was to achieve better performance in fire class prediction rather than in the "non-fire" class. The results that are presented in the next section indicate a significant learning potential of the RF algorithm using as input the specific feature database.

4 Results

4.1 The Spearman's Correlation

Figure 3 shows in a heat map the returned Spearman's rank correlation coefficient considering all possible pairs of numeric features.

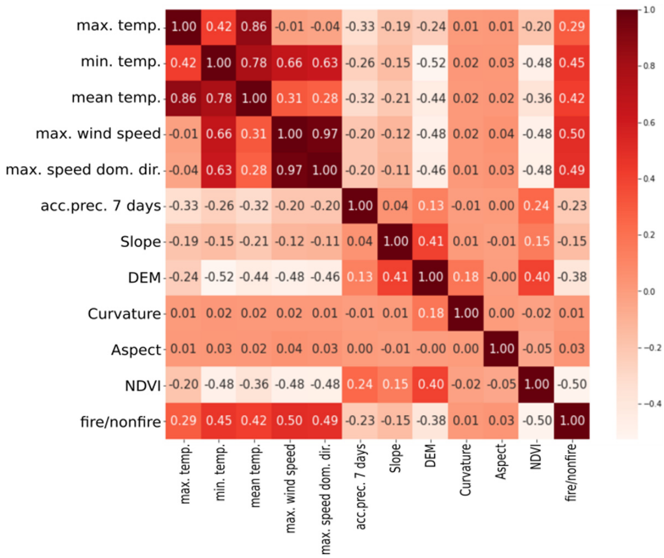

Fig. 3. Spearman's correlation results

Notable but not surprising the observations on this heat map indicate that: (1) the wind speed, the temperature, the NDVI and the DEM elevation are ranked as more meaningful and influencing features for "fire"/"non fire" class prediction, (2) The DEM elevation and the accumulated rainfall have inverse correlations with the dependent variable "fire"/"non-fire", (3) The two wind speed and the three temperature features are highly correlated one another.

4.2 The Chi-Squared Tests

The p-values of the chi-squared tests for the categorical inputs (Corine Land Cover, dominant wind direction and wind direction of the maximum wind speed) were extremely small ($<10^{-100}$) so the standard null hypothesis that these variables are independent was rejected beyond doubt. Consequently, it was showed that all the categorical variables are dependent with a high statistical significance with the "fire"/"non-fire" variable.

4.3 Comparative Feature Ranking

The ranking results from Sequential Feature Selection (SFS), RF impurity and the permutation importance are shown in Table 2; one can deduce that NDVI, wind speed and temperature variables are ranked between the first three places.

Table 2. Feature Ranking returned by Sequential Feature Selection, RF impurity and Permutation importance

Ranking	SFS AUC	RF impurity	Permutation importance
1	NDVI	max. speed of the dom. dir	NDVI
2	maximum wind speed	NDVI	max. temp
3	max. temp	max. temp	maximum wind speed
4	CLC	maximum wind speed	mean temp
5	acc. prec. of past 7 days	min. temp	max. speed of the dom. dir
6	DEM	mean temp	DEM
7	Aspect	acc. prec. of past 7 days	acc. prec. of past 7 days
8	Curvature	DEM	wind dir. of the max. speed
9	Slope	wind dir. of the max. speed	dominant wind direction
10	min. temp	dominant wind direction	CLC
11	max. speed of the dom. dir	CLC	Aspect
12	wind dir. of the max. speed	Aspect	Curvature
13	mean temp	Slope	min. temp
14	dominant wind direction	Curvature	Slope

4.4 The RF Classifier Results

The metrics of precision, recall, and f1-score returned from the application of RF are shown in Table 3.

The first two rows indicate classification results based on training and test data that were randomly produced at the rates of 90% and 10% respectively of the feature dataset while using the shuffling process. In contrast, the two last rows indicate the mean values

Table 3. RF metrics results

Training/Test set	Class	Precision	Recall	F1-Score
Training/Test with shuffling	Fire	0.94	0.92	0.93
	Non fire	0.92	0.94	0.93
k-fold cross validation (folds include entire days)	Fire	0.75	0.77	0.76
	Non fire	0.77	0.75	0.76

of the k-fold cross validation metrics without shuffling. The significance from invoking shuffling in the classification results is obvious.

In the first case (with shuffling) the results are quite satisfactory. The model is capable to distinguish quite well the "fire" from the "non-fire" class. In the case of the k-fold cross validation without shuffling, where training and test cells belong to different fire events, the mean recall value for the class fire remains still quite high as it was targeted. The difference and lower performance results in the k-fold cross validation case without invoking shuffling was expected as explained in 3.3.

5 Conclusion

This study showcased the creation of a prototype feature database that has proven useful for understanding the fire regime and predicting the "fire"/"non fire" classes in Greece. Moreover the study showed that this national scale feature database has been a reliable input for ML modelling by training successfully a RF algorithm, in order to distinguish between the two categories. As shown the parameters which are highly significant in the prediction have been *NDVI*, *maximum wind speed* and *maximum temperature* followed by the *DEM derived parameters* and *accumulated precipitation of the past 7 days.* The promising results which have been returned from the application of the RF classifier guide our steps towards building an enhanced fire risk model through invoking, testing, merging and validating results from different data driven ML approaches. Moreover, in order to address the daily requirements for emergency response there is need to translate the algorithm's "fire"/"non-fire" predictions to a number of fire risk classes. In this regard, it is necessary to conduct additional tests that examine the performances of various ML algorithms e.g. Artificial Neural Networks (ANN), Convolutional NN (CNN) or other ensemble algorithms like boosting trees. Furthermore, additional EO derived data (i.e. LST, soil moisture, proximity classes) could be considered for integration in the model in order to enhance the classification performance and introduce new aspects of the expected fire occurrences in Greece's ecosystem. This latter is an ongoing research study at the premises of the BEYOND Center of Excellence and subject of validation by the Greek Fire Brigade authority using fire data from the fire season of 2020.

Finally, since an ML Model can easily be transferred between feature databases with same features, a fire risk ML model developed for Greece could operate on a feature database with data for another location. Especially if this area has similar ecosystem properties like for example Mediterranean countries, the performance of the model could be similar. The challenge in this case would be the creation of harmonized data features with the ones used in the original development. Even though the level of performance of the ML model on another dataset cannot be "a priori" guaranteed, it is definitely worth to test the potential of transferring a developed for fire prediction ML model for Greece to another country or larger area.

Acknowledgements. This paper has been supported by using data and resources from the following Projects funded from EC and the Greek Government - Ministry of Development & Investments : (1) FRAMEWORK SERVICE CONTRACT FOR COPERNICUS EMERGENCY MANAGEMENT SERVICE RISK AND RECOVERY MAPPING- The European Forest Fire Information System (EFFIS) JRC/IPR/2014/G.2/0012/OC; (2) FRAMEWORK SERVICE CONTRACT FOR COPERNICUS EMERGENCY MANAGEMENT SERVICE RISK AND RECOVERY MAPPING - Program Call for tender JRC/IPR/2014/G.2/0012/OC; and (3) CLIMPACT: Flagship Initiative for Climate Change and its Impact by the Hellenic Network of Agencies for Climate Impact Mitigation and Adaptation.

References

1. The Copernicus Emergency Management Service Monitors Impact of Fires in Australia | Copernicus Emergency Management Service. https://emergency.copernicus.eu/mapping/ems/copernicus-emergency-management-service-monitors-impact-fires-autralia. Accessed 31 July 2020
2. European Commission: JRC Tecnical Report Forest Fires in Europe, Middle East and North Africa 2018 (2018)
3. Castellari, S., Kurnik, B.: Climate change, impacts and vulnerability in Europe 2016, no. 1. (2017)
4. Fares, S., et al.: Characterizing potential wildland fire fuel in live vegetation in the Mediterranean region. **74**, 1 (2017). https://doi.org/10.1007/s13595-016-0599-5
5. Lambert, J., Drenou, C., Denux, J.-P., Balent, G., Cheret, V.: Monitoring forest decline through remote sensing time series analysis. GISci. Remote Sens. **50**(4), 437–457 (2013). https://doi.org/10.1080/15481603.2013.820070
6. Pastor, E., Zárate, L., Planas, E., Arnaldos, J.: Mathematical models and calculation systems for the study of wildland fire behaviour. Progress Energy Combust. Sci. **29**(2), 139–153 (2003). https://doi.org/10.1016/S0360-1285(03)00017-0.
7. Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X., Xu, C.: Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. Sci. Total Environ. **630**, 1044–1056 (2018). https://doi.org/10.1016/j.scitotenv.2018.02.278
8. Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D.: A review of machine learning applications in wildfire science and management (2020)
9. Pourtaghi, Z.S., Pourghasemi, H.R., Aretano, R., Semeraro, T.: Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. Ecol. Indic. **64**, 72–84 (2016). https://doi.org/10.1016/j.ecolind.2015.12.030
10. Kontoes, C., Keramitsoglou, I., Papoutsis, I., Sifakis, N., Xofis, P.: National scale operational mapping of burnt areas as a tool for the better understanding of contemporary wildfire patterns and regimes. Sensors **13**(8), 11146–11166 (2013). https://doi.org/10.3390/s130811146
11. ΕΜΥ: Εθνική Μετεωρολογική Υπηρεσία. https://www.emy.gr/emy/el/. Accessed 31 July 2020

12. EEA: State of the environment report (SOER) No 1/2010 : The European environment: State and outlook 2010. Synthesis (2010)
13. Rivera, A., Bravo, C., Buob, G.: Climate Change and Land Ice (2017)
14. Kailidis, D., Karanikola, P.: Forest Fires 1900–2000. Giahoudi Press, Thessaloniki (2004)
15. Forest Fires in Europe 2006 | EU Science Hub. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/forest-fires-europe-2006. Accessed 13 Aug 2020
16. Beyond Centre of Excellence for EO based monitoring of Natural Disasters. https://www.beyond-eocenter.eu/. Accessed 31 July 2020
17. Kontoes, C., Papoutsis, I., Themistocles, H., Ieronymidi, E., Keramitsoglou, I.: Remote Sensing Techniques for Forest Fire Disaster Management: The FireHub Operational Platform, Book Chapter No. 6, Integrating Scale in Remote Sensing and GIS (2017)
18. SEVIRI Monitor - NOA GIS. https://195.251.203.238/seviri/. Accessed 31 July 2020
19. Massada, A.B., Syphard, A.D., Stewart, S.I., Radeloff, V.C.: Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA (2013). https://doi.org/10.1071/WF11178
20. Killough, B.: Overview of the open data cube initiative. In: International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2018-July, pp. 8629–8632 (2018). https://doi.org/10.1109/IGARSS.2018.8517694
21. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024
22. Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M.: Benchmark for filter methods for feature selection in high-dimensional classification data. Comput. Stat. Data Anal. **143**(September) (2020). https://doi.org/10.1016/j.csda.2019.106839
23. Dodge, Y.: The Concise Encyclopedia of Statistics, p. 502. Springer, Heidelberg (2010)
24. Feelders, A., Verkooijen, W.: On the Statistical Comparison of Inductive Learning Methods (1996)
25. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. **30**(7), 1145–1159 (1997). https://doi.org/10.1016/S0031-3203(96)00142-2
26. Ho, T.K.: Random decision forests. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, vol. 1, pp. 278–282 (1995). https://doi.org/10.1109/ICDAR.1995.598994
27. Stone, M.: Cross-Validatory Choice and Assessment of Statistical Predictions (1974)
28. Tonini, M., D'andrea, M., Biondi, G., Esposti, S.D., Trucchia, A., Fiorucci, P.: A machine learning-based approach for wildfire susceptibility mapping. The case study of the Liguria region in Italy. Geoscience **10**(3), 18 (2020). https://doi.org/10.3390/geosciences10030105
29. Bergstra, J., Ca, J.B., Ca, Y.B.: Random Search for Hyper-Parameter Optimization Yoshua Bengio (2012)
30. Kent, A., Berry, M.M., Luehrs, F.U., Perry, J.W.: Machine literature searching VIII. Operational criteria for designing information retrieval systems. Am. Doc. **6**(2), 93–101 (1955). https://doi.org/10.1002/asi.5090060209