Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Estimating Forest Fire Losses Using Stochastic Approach: Case Study of the Kroumiria Mountains (Northwestern Tunisia)

Ahmed Toujani, Hammadi Achour & Sami Faïz

To cite this article: Ahmed Toujani, Hammadi Achour & Sami Faïz (2018) Estimating Forest Fire Losses Using Stochastic Approach: Case Study of the Kroumiria Mountains (Northwestern Tunisia), Applied Artificial Intelligence, 32:9-10, 882-906, DOI: 10.1080/08839514.2018.1514808

To link to this article: https://doi.org/10.1080/08839514.2018.1514808

APPLIED ARTIFICIAL INTELLIGENCE
2018, VOL. 32, NOS. 9–10, 882–906
https://doi.org/10.1080/08839514.2018.1514808

Estimating Forest Fire Losses Using Stochastic Approach: Case Study of the Kroumiria Mountains (Northwestern Tunisia)

Ahmed Toujani[a,b], Hammadi Achour[a,c], and Sami Faïz[b]

[a]LRSP Laboratory, Silvo-Pastoral Institute of Tabarka, University of Jendouba, Jendouba, Tunisia; [b]LTSIRS Laboratory, National Engineering School, University of Tunis El Manar, Tunis, Tunisia; [c]Unity of Geomatics and Geosystems, University of Mannouba, Tunis, Tunisia

ABSTRACT

Kroumiria Mountains (northwestern Tunisia) have experienced major fires, making them the main loss reason of Tunisian forested areas. The ability of accurately forecasting or modeling forest fire areas may significantly aid optimizing fire-fighting strategies. However, there are still limitations in the empirical study of forest fire loss estimation because the poor availability and low quality of fire data. In this study, a stochastic approach based on Markov process was developed for the prediction of burned areas, using available meteorological data sets and GIS layers related to the forest under analysis. The Self-organizing map (SOM) was initially used to classify spatio-temporal factors influencing the fire behavior. Subsequently, the SOM clusters were incorporated into a Hidden Markov Model (HMM) framework to model their corresponding burned areas. Results achieved using a database of 829 forest fires records between 1985 and 2016, showed the appropriateness of the HMM approach for the prediction of burned areas compared with a state-of-the art machine learning methods. The transition probability matrix (TPM) and the emission probability matrix (EPM) were also analyzed to further understand the spatiotemporal patterns of fire losses.

Introduction

In the Mediterranean Basin, the forest fire is one of the most serious hazards (Turco et al. 2014). Each year, more than 50,000 fires have been recorded in the Mediterranean basin, burning an area of 800,000 ha, which correspond to 1.7% of the total woodland cover (Fao 2013). In Tunisia, forest resources are very limited (≈ 1 million of ha), and only account for 6% of the whole national's territory (Fao 2016). Historical data on forest fires in the country reveals that forest fires affected about one thousand ha per year between 1996 and 2010 and approximately 3167 ha per year from 2011 to 2014 (Fao 2016). In the Kroumiria region (Northwestern Tunisia), which represents more than

CONTACT Ahmed Toujani ahmed.toujani@iresa.agrinet.tn

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

70% of the country's woodland cover, the fires ravage each year hundreds of hectares and sometimes by up to 0.6% of the total country's woodland area (Campos et al. 2008), exceeding the national average of 0.3% (308 ha) (Fao 2013).

There are many approaches used to assess forest fire hazard areas. Among them, one commonly used approach is based on the identification of causal geospatial variables through qualitative rules and expert knowledge (Bisquert et al. 2012; Chou 1992; Dlamini 2011; Jung et al. 2013; Lozano et al. 2008; Oliveira et al. 2016; Parisien et al. 2012). The common weakness of these studies integrating the Multi-Criteria Decision Analysis (MCDA) and Geographic information system (GIS) for forest fire susceptibility mapping, is the difficulty in assigning different weights to various stakeholder interests. Indeed, it is not easy and strongly depends on expert opinion to decide which parameters or criteria have greater or lesser weights (Eugenio et al. 2016).

To overcome this problem and reduce the subjectivity apparent in these qualitative approaches, statistical methods such as spatial linear and logistic regression models have been widely applied for studying the relationships between fire occurrence data and explanatory factors (Nieto et al. 2012; Preisler et al. 2011; Tao et al. 2013). The fuzzy logic has been also used for fire risk estimation (Soto and Alvear 2012), as well as other advanced statistical models such as probabilistic graphs (Dlamini 2011) and geographically weighted regression (Martínez-Fernández, Chuvieco, and Koutsias 2013; Oliveira et al. 2016; Rodrigues, De La Riva, and Fotheringham 2014). Furthermore, several data-mining algorithms have been also proposed for fire behavior modeling, including sliding window (Ku-Mahamud and Khor 2009), genetic algorithm (Castelli, Vanneschi, and Popovic 2015), neural networks (Özbayoğlu and Bozer 2012; Bisquert et al. 2012; Safi and Bouroumi 2013), random forest and generalized additive model (Pourtaghi et al. 2016), maximum entropy (Arpaci et al. 2014; Parisien et al. 2012), regression trees (Amatulli et al. 2006; Lozano et al. 2008; Stojanova et al. 2011), and support vector machine (Xie and Shi 2014).

Hidden Markov Models (HMMs) have been widely applied as time series approaches in climate change and weather forecasting (Li et al. 2013). HMM is a stochastic model that has been proven useful in meteorological applications, including rainfall modeling and forecasting (Jones et al. 2013), wind speed estimation (Barber, Bockhorst, and Roebber 2010), and spatio-temporal weather patterns (Ailliot et al. 2012). Likewise, in the field of fire danger assessment, there are some efforts supporting the time series modeling, such as dynamic neural networks for short-term projection (Cheng and Wang 2008), and generalized additive model for human-caused fires (Woolford et al. 2009). However, most of previous works were focused on spatial fire danger assessment despite the temporal behavior of many important fire factors such as weather variables and dynamic human-induced changes. In addition the HMM techniques, that have been proven useful in variety of research areas (Testa et al. 2015; Thomas et al. 2013; Yu and Sun 2016), are still not widely employed for spatiotemporal assessment of the forest fire hazard.

Due to these shortcomings, an in-depth study is required to test the ability of the HMM for modeling forest fire hazard behavior as well as predicting the size of burned areas. This would provide a tremendous benefit for fire-fighters to allocate resources, especially in peak seasons where there may be multiple fires happening at the same time. To the best of our knowledge, this is the first study that mines multi-dimensional geographical data by explicitly considering their spatial an temporal dimensions. This allows us to discover more complex forest fire behavior associated to these dimensions. A key feature of using HMM for a spatio-temporal forecasting is that HMM states can be directly linked to real world situations. In this sense, hidden states can be used to effectively predict fire loss areas that are influenced by dynamic factors. That is to say, when the sequence of burned areas is hidden; it can be estimated based on observable fire factors that are varying in time and space. In order to accomplish this goal, an experimental study was carried out in the Kroumiria region (Northwestern Tunisia) using a set of forest fire occurrence data compiled from different sources between 1985 and 2016.

Materials and Methods

Study Area

The Kroumiria region is located in the extreme north-west of Tunisia between 7°52′E–37°29′N and 9°33′E–36°22′N, covering an area of approximately 2500 (Figure 1). It extends south of the Mediterranean Sea and north of Wadi Medjerda and east from the Algerian border to Jebel El-Abiod. For management purposes, the forested areas are subdivided into 78 parcels, ranging from 0.64 km^2 to 47 km^2. The parcels are management units, which represent homogeneous stands regarding ecological and topomorphological conditions and they are generally delimited by roads or natural limits such as valley bottoms, rocky ridge, etc. The region is mountainous with elevations ranging from 0 m on the northern coast to 1235 m (a.s.l) and characterized by a rather heterogeneous geology, dominated by clayey sandstone rocks. Terrain gradient, computed from Digital Elevation Model (DEM), ranges from 0° to 49° with an average of 17° and standard deviation of 10° (Kalboussi and Achour 2018). The climate is Mediterranean with annual mean temperature of 18.2°C (1975–2004). Maximum and minimum temperatures are 34.4°C and 5.6°C, respectively. Average annual rainfall is 912 mm, with 77% of total precipitation occurring in autumn and winter. While summers are dry and long, with a strong north/south gradient and only about 4% of total annual precipitation occurring between May and October (Chakroun et al. 2012). These climatic and physiographic conditions have led to the development of a rich flora and diverse vegetation. At present, the forest cover is dominated at low elevations

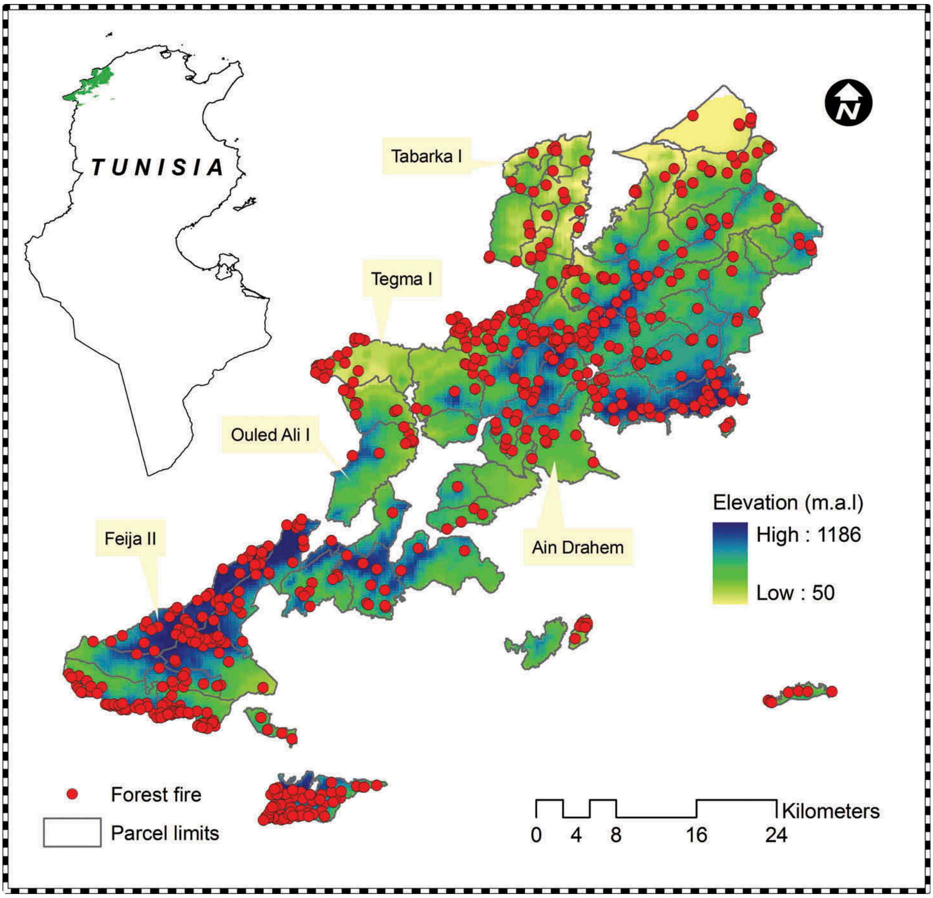

Figure 1. Forest fires occurred in Kroumiria parcels between 1985 and 2016.

(<100 m) by kermes oak (*Quercus coccifera*). At middle elevations (300–800 m a.s.l.), the forests mostly consist of pure stands of cork oak (*Quercus suber*), and at higher elevations (> 800 m), by zeen oak (*Quercus faginea*) formations. In the north-western border of the region, the natural vegetation is dominated by pure maritime pine (*Pinus pinaster*) forests (Kalboussi and Achour). Scattered within the region, many sectors were planted by stone pine (*Pinus pinea*), eucalyptus sp. and maritime pine. With the Mediterranean-climate characterized by dry and warm summers, these vegetation species are of high flammability, which is reflected in the study area by recurrent fire regimes (Pausas 2004).

Data Collection and Preprocessing

Forest Occurrence Data

In Tunisia, forest fires are classically documented by the Tunisian General Directorate of Forestry (DGF). In the Kroumiria region, where forest fires are very recurrent, the information has been systematically collected and recorded using a standard protocol since the beginning of the 80s. The typical report about each fire event included the burned area size, their location in the corresponding forest parcel, their date and duration, their major fuel types and the probable cause of their ignition. Fortunately, this primitive information exists for most forest fires since 1985. It should be noted, however, that some recorded fire points present missing and incomplete information. To overcome this problem, forest fire occurrences from the European Forest Fire Information System (EFFIS) and the Fire Information for Resource Management System (FIRMS) were also used to check the existing data and complete missing features and missing fires. Forest fire data from the EFFIS (2016) consist of a compilation of the extent of the burned areas (2010–2015) based on a semiautomatic classification of Moderate Resolution Imaging Spectroradiometer (MODIS) images, allowing the mapping burnt areas of about 40 ha or larger. However, a large number of fires smaller than the mentioned threshold have been also detected and mapped. Data delivered by the FIRMS (2015) consist of active fire points detected using MODIS sensor. The fire detection algorithm is fully automated and identifies pixels with one or more actively burning fires for the entire globe. Detection confidence is estimated in the detection procedure and ranges from 0 to 100% (Giglio et al. 2003). The confidence level is used to classify all fire pixels as low confidence [< 30%], nominal confidence [30–80%] or high confidence [> 80%]. Higher confidence levels can be applied to reduce the number of false alarms (errors of commission) at the expense of a lower detection rate (Giglio et al. 2010). In summary, a total of 829 forest fire occurrences recorded between 1985 and 2016 was used to set the time series (one time series representing burned areas for each parcel). As shown in Figure 2, many parcels were not affected by fires, and then they were discarded. Table 1 summarizes the main characteristics of some considered parcels in Kroumiria forests.

Environmental Predictors

We considered different sets and sources of environmental predictors of fire occurrences, which are summarized in Table 2. These variables were selected based on the investigation of two expert models, namely Dagorne's model (Dagorne et al. 1994) and Fire Weather Index model (FWI; Cortez and Moraiz 2007). The Dagorne's model, designed for the Mediterranean forest context, considers three main factors, including the combustibility index, human occupation index, and the topographic Index; while, the FWI model, designed specifically for Canadian's forest, focuses particularly on climatic variables (i.e., temperature, relative humidity, and wind).

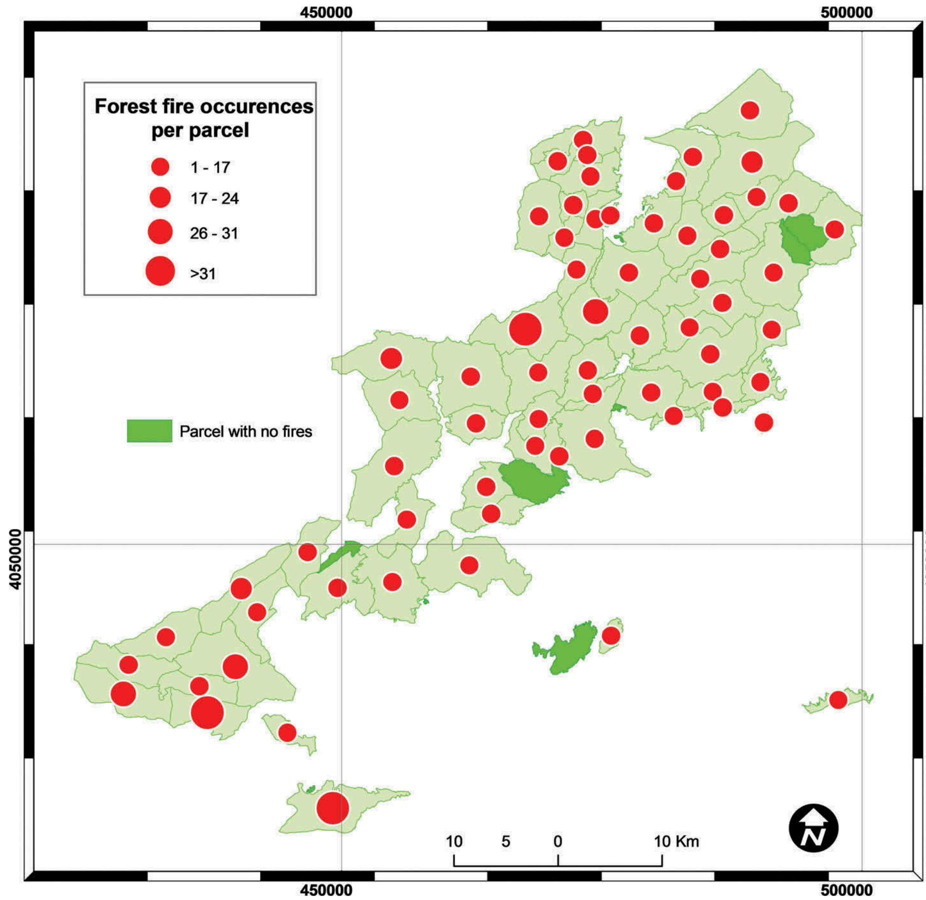

Figure 2. Forest fire occurrences per parcel.

Table 1. Forested area, number of fires, total burned area and largest burned in the period 1985–2016 for some considered parcels in Kroumiria forests.

Kroumiria parcel	Parcel ID	Forested area (ha)	First fire's year	Last fire's year	Number of forest fires	Total burned area (ha)	Largest burned area (ha)
Tabarka I	1	1760.675	1986	2015	5	455.715	450
Ain Drahem IX	50	4084.688	1987	2013	6	3.63	1.45
Ouled Ali I	53	1880.726	1994	2005	11	123.36	120
Feija II	47	2306.713	1993	2014	14	262.514	250
Tegma I	17	3314.273	1990	2015	19	313	300

Vegetation Layer. Vegetation data were obtained the national forest inventory map published by the DGF in 2010. This map classify the natural vegetation on its physiognomy, phenology, and floristic composition. More than 10 classes were identified for the Kroumiria region. We simplified those classes into four categories taking into account dryness of vegetation and dominant presence of deciduous species, grasses and weeds, which could act as fuel loads for fires. We

Table 2. Summary of the extracted parameters for fire factors.

Factors	Parameters	Description	Min	Median	Max	Mean	SD
Vegetation fuel	Flammability intensity	Flammability rating of tree cover for the area where the fire is located.	0	2	3	1.783	1.025
Topomorphological impact	Aspect (in degrees)	Solar exposure for fire location.	2	169	360	183.9	103.695
	Slope (in %)	Degree of slope for fire location.	0	8	24	8.351	4.743
	Elevation (in m)	Altitude of fire location.	15	531	1099	529.9	247.257
Anthropogenic impact	Road (in m)	Distance from fire location to the nearest road.	5.153	347.934	3739.171	572.619	616.001
	Settlement (in m)	Distance from fire location to the nearest settlement.	136.9	2200.7	29380.8	6099.1	7605.649
Meteorological impact	Relative Humidity (in %)	Daily mean of relative humidity.	26.71	34.04	34.42	68.86	6.55
	Wind speed (in m/s)	Daily mean of wind speed.	1.40	2.90	2.75	6.10	0.803
	Temperature (in °C)	Daily mean of temperature.	21.11	29.66	29.26	35.68	1.994
Other variables	Lookout-post (in m)	Distance from fire location to the nearby lookout post.	284.5	3602.3	18110.5	4224.0	2868.213
	Fire-wall (in m)	Distance from fire location to the nearby fire-wall.	4.955	1621.8	36854.5	7468.6	6304.116
	Tunisian-Algerian border (in m)	Distance from fire location to the adjacent Tunisan-Algerian border.	19.77	4793	20267.07	5892.12	4690.014

assigned a score value varying from 0 to 4 reflecting the level of the flammability (i.e., the ability of a species to ignite and sustain fire) for each vegetation class (Dimitrakopoulos 2001; Henaoui, Bouazza, and Amara 2013). A score of 0 was assigned to the class representing less flammable species such as Cistus salvifolius; while, a score of 3 was attributed to the class regrouping very flammable species, including *Pinus halepensis, Quercus suber, Quercus faginea*. Intermediate score-values of 1 and 2 were attributed to moderately flammable and flammable vegetation classes, respectively.

Topographical and Meteorological Variables. Elevation data are generated from the 1-arc second DEM (approximately 30 m) derived from the recently and freely available Shuttle Radar Topography Mission (SRTM) and retrieved from the

United States Geological Survey (USGS) website (http://www.earthexplorer. usgs.gov). To cover the entire study area, the tile "n36_e008_1arc_v3" was downloaded and then pit-filled using ArcGIS generic tools (ESRI, Inc., Redlands, CA, USA). From this DEM, we computed slope gradient and aspect grids using spatial analysis tool.

Meteorological data, including the mean temperature, relative humidity and wind speed were collected from the National Weather Service for fire event from 1985 to 2016. Missing values were completed with corresponding daily averaged data from the NASA's Surface meteorology and Solar Energy (SSE) database (NASA 2016). The SSE data set contains climatology parameters principally derived from an atmospheric model constrained to satellite observations. It is 22-year climatology (July 1983–June 2005) on a one-degree latitude by one-degree longitude grid.

Anthropogenic Variables. In most cases, forest fires have an anthropogenic origin, whether voluntary or involuntary (Renard et al. 2012). Fire ignitions recorded tend to be clustered around transportation networks and near urban areas (Catry et al. 2009; Martínez et al. 2009). We therefore considered four variables to describe the human footprint: (1) distance to road networks, (2) distance to settlements, (3) distance to lookout posts (i.e., watch tower), and (4) distance to fire-walls. Furthermore, as the Tunisian-Algerian border plays an important role in the spatial distribution of forest fires, we also computed, using Euclidean distance tool, the thematic layer "distance to the Tunisian-Algerian border."

Burned Area Modeling

An HMM consists of a doubly stochastic process representing a time-varying system (such as modeling of vegetation dynamics or phonological variations), in which the hidden stochastic process can be indirectly inferred by analyzing the sequence of observed symbols of another set of stochastic processes.

More formally, a discrete N-state HMM is defined by its components ($\lambda = (A, B, \Pi)$), where the state transition probability is $A = \{a_{ij}\}_{i,j=1,N}$, the distribution of observation probability is $B = \{b_i(o)\}_{i=1,N}$, and the distribution of initial state probability is $\Pi = \{\pi_i\}_{i=1,N}$. Given suitable estimates of the HMM parameters (modeling A and B matrices), solutions are well known for: (1) predicting observation and state sequences from the model by Monte Carlo sampling methods; (2) estimating the most likely behavioral "state" sequence given the model and an observation sequence (Viterbi algorithm); and (3) updating the model estimate given new observations (Baum-Welch algorithm). More details about the Viterbi or Baum Welch solutions are presented in the above references. The issues related to developing an HMM to estimate fire losses involved (1) determining hidden states applicable to assess the burned area sizes, (2) employing fire factors as observations to which they correspond, (3) and taking into account the spatial variability of fires to the different parcels, in addition to the temporal variability.

Figure 3 shows the distribution of the forest fires based on their amount of burned areas, with the majority of the fires presenting a small size. It indicates a positive skew, which is also the same case in Canada (Malarz, Kaczanowska, and Kułakowski 2002), Portugal (Cortez and Moraiz 2007) and Turkey (Ozbayoglu and Bozer 2011). Regarding the dataset, there are 807 fires less than 50 hectares in size, whose 74.10% of samples were less than 1 hectare. Thus, the logarithm function $y = \ln(x + 1)$ was applied to the area size to reduce skewness and improve symmetry (Figure 4). The log transformation is a common method used to improve regression results for right-skewed targets (Menard 2002). Thus, the transformed burned area, presented in Figure 4, will be the output target of this work.

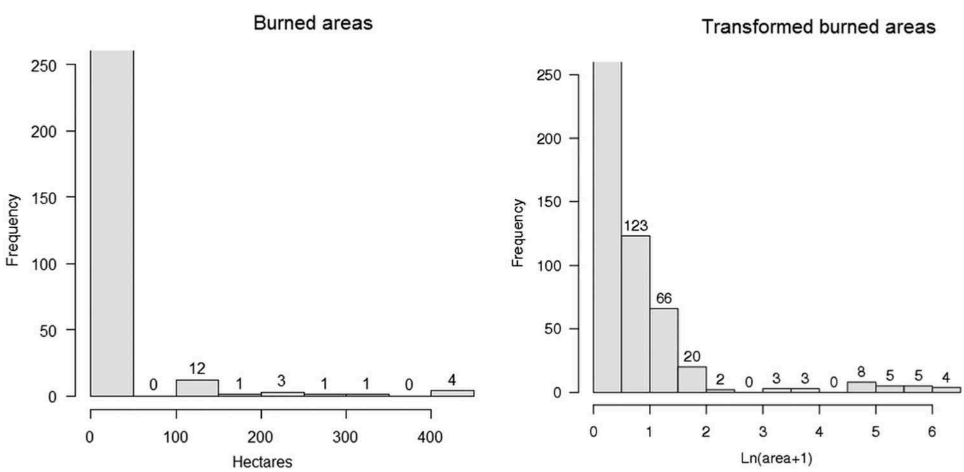

Figure 3. The histogram for the burned area (left) and respective logarithm transform (right).

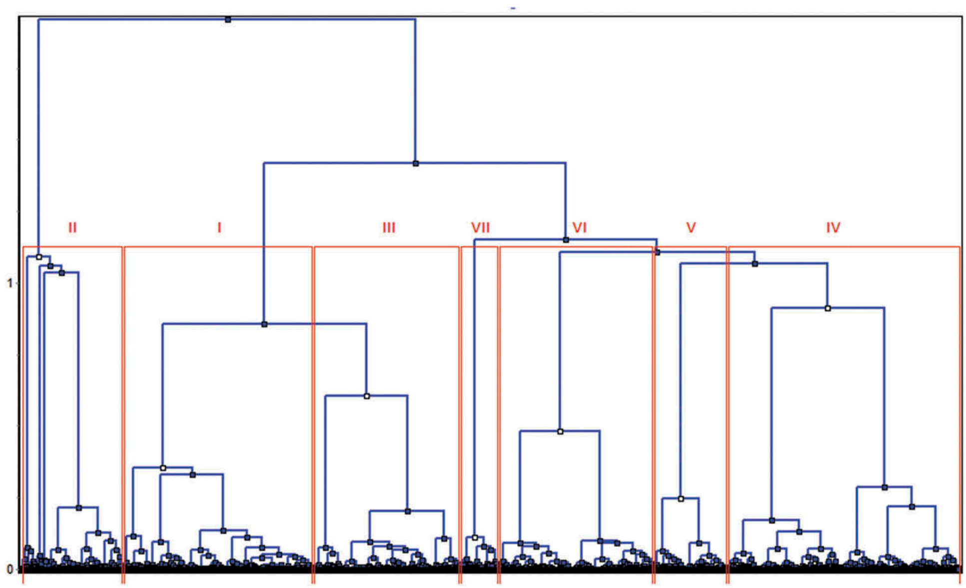

Figure 4. Seven clusters classified by the SOM.

By using a HMM, the burned area are understood as hidden states which are being searched for. However, in a HMM the hidden states are conventionally discrete variables. Instead, we perform discretization of the transformed value. We used four-state HMM according to the burned area classes, namely; limited, considerable, large, and massive. We first compute the z-score of transformed variable for each sample as $(x\text{-}\mu)/\sigma$, where μ is the mean and σ is the standard deviation. Then, we assign discretized values to samples according to their z-score using the following formula: "If |normalized burned area| < −1, then it is limited; else if −1 ≤|normalized burned area| <0, then it is considerable; else if 0 ≤|normalized burned area|<1, then it is large; else it is massive".

As we assumed that forest fire factors were suitable parameters to predict burned areas, they will serve as the observations in the HMM. In our research, 14 factor parameters were extracted for each selected fire spot: Month, year, nearby lookout post distance (m), adjacent fire-wall distance (m), Tunisian-Algerian border distance (m), vegetation flammability (0 to 3), temperature (Celsius), relative humidity (%) and wind speed (m/s), altitude (m), slope (%), aspect (0 (due north) to 360 (again due north)), closest road distance (m), and nearest settlement distance (m). In order to clearly bring out the annual variation of fire factors, the year parameter was treated as categorical variable. It was discretized into five ranges using equal-frequency discretization (1985–1995, 1996–2003, 2004–2007, 2008–2010, and 2011–2015 periods). After that, all fire factors data are stored as vector data.

The HMM with discrete observations is generally the most used since its robustness in training in comparison to continuous HMMs, where observations are formed by mixtures of Gaussians as opposed to discrete symbols (Rabiner 1989). Therefore, we used the Self Organized Map (SOM) to cluster the vector data of fire factors (Kohonen, Kaski, and Lappalainen 1997). The Self-organizing map is a nonparametric and nonlinear neural network that explores data using unsupervised learning. In the input layer, the neurons correspond to the variables describing the vector data. The output layer is most often organized as a grid of two-dimensional neurons.

Each neuron represents a group of similar data. The Euclidian distance (*dj* (*t*)) at neuron *j* on the SOM between weight at iteration time *t* and the input vector was calculated through learning processes:

$$d_j(t) = \sum_{i=0}^{P-1} \left[x_i - w_{ij}(t)\right] \tag{1}$$

where x_i is the value of parameter *i*, $w_{ij}(t)$ is the weight between parameter *i* and the neuron *j* on the SOM, and *P* is the number of the parameter.

The best matching neuron, which has the minimum distance, is chosen as the winner. For the best matching neuron and its neighboring neurons, the new weight vectors are updated as

$$w_{ij}(t+1) \ = w_{ij}(t) \ + a(t)\left[x(t) \ - w_{ij}(t)\right] \tag{2}$$

where *t* is the iteration time, and (*t*) is the learning rate. The learning rate accordingly decreases as the system converges. The final result of SOM learning is strongly influenced by the SOM dimension (M × N neurons) that is usually determined by considering the input vector data. There are several approaches to determine SOM dimension such as using the expert knowledge (Park and Chung 2006) or setting approximately the number of neurons to the number of input samples, if the data sets are small (Kohonen, Kaski, and Lappalainen 1997). In our research, according to Vesanto et al. (2000), the number of neurons is approximately equal to 5 × √(number of samples). Thus, by using the SOM, all vector data of fire factor parameters will be assigned with their corresponding cluster labels in order to be used as discrete HMM observations.

Before feeding the input vector data into the SOM network, certain preprocessing steps must be done (Hastie, Tibshirani, and Friedman 2009). As month and year variables were categorical, they were transformed into a 1-of-C encoding as recommended by Hsu, Chang, and Lin (2003). Using the boxplot rule, no outliers were identified. All attributes were normalized with respect to their standard deviation in order to represent them on the same scale, before being stored as data vectors. In the SOM processing, the network received the data vectors, and all weight vectors between the input parameter and the competitive nodes were computed (eq.2). The weights of the best-matching unit and other neurons in the SOM lattice are adjusted toward the input vector. Thus, the similarity among vectors of fire factor parameters is reflected on the SOM map. In order to decide on the correct number of clusters in the trained SOM, we utilized the Ward linkage clustering (Tasdemir and Milenov 2010). The Ward's dendrogram gives an aid to reveal the degree of association between the SOM units. The interpretation of the dendrogram and, consequently, of final clusters, is then reinforced.

Extracted clusters of fire factors are subsequently used as the observation symbols in an HMM classifier for burned area estimation. Each parcel is then presented by its sequence of burned areas as one HMM state sequence, and by its corresponding sequence of fire factor clusters as its matching observation sequence. Throughout this HMM modeling, the cluster of fire factors, in a given forest parcel, is the dynamic subject that is being observed and whose burned area size is to be determined if fire will occur.

Burned Area Estimation

Using the form of the forest fire model described above, HMM parameters are defined based on the learning set of observation sequences, where each of them (a sequence of clusters representing fire factors) for which the sequence burned areas is known (a state sequence). The HMM parameters, $\lambda = (A,B,\pi)$, were estimated using the Baum Welch algorithm (Rabiner 1989).

The Baum Welch is a modified version of Expectation-Maximization algorithm that is designed to optimize the parameters of an HMM so as to best model given training sequences. Thus it deals with maximizing the conditional probability of an observation sequence, $P(O|\lambda)$, occurring given an HMM (to be optimized). Once the model is trained on a series of observation sequences, the model can be utilized to compute the likelihood of a new sequence being generated by the model. This is where the strength of HMMs lie, since by using dynamic programming techniques this can be computed in just $O(N^2T)$ calculations (Rabiner 1989).

When the HMM have been established and their parameters estimated, the classification of a burned areas is done in the following way. The forest fire series, occurred in a given parcel, is represented at each date by a cluster implying its fire factors observed at that date. From the clusters representing the fire factors during a given succession of years, the classifier computes for each model, the probability that the corresponding burned area class emits the observed sequence of fire factor clusters. The burned area size is then assigned to the class whose model delivers the highest emission probability. A detailed description about how emission probabilities are computed using Viterbi algorithm can be found in Rabiner (1989).

As a practical use, given the coordinates (*x,y*) and the year of one fire, its factors can be carried out and then its corresponding cluster will be well-known. Thus, according to the series of past fires and its corresponding sequence of fire factors (sequence of SOM clusters), the most probable set of hidden states is reconstructed using Viterbi algorithm. As a result, the hidden state of the corresponding fire designates its estimated burned area size. In addition, fire losses can be analyzed against the other variable such as a time or a cluster of fire factors. For example, the analysis could show differences in fire behavior in certain year and with certain parameters of fire factors. The other use is the prediction of the next class of burned area according to the parcel in question.

Results and Discussion

SOM Discretization

Fire factors were patterned by SOM based on spatial parameters extracted from the fire spots stated above. In total 829 vectors of forest fire related factors were passed through the SOM after the preprocessing step (removing 12 outliers and normalizing the data). The selected size of the SOM was 10×10 nodes, giving an explained ratio of 51.83%. As showed in Figure 5, seven patterns of fire factors were identified according to the Ward linkage clustering (Tasdemir and Milenov 2010). The centroids of these clusters represent descriptive statistics for the parameters used

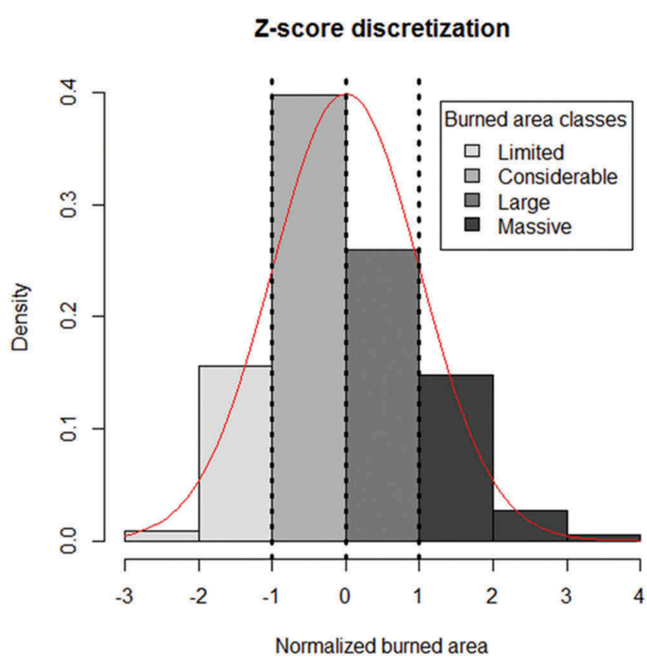

Figure 5. Four classes of burnt area after the discretization by Z-score method.

in the analysis. It is defined as vector consisted of the means of all variables for a given cluster. The mean values for all continuous variables for each cluster are summarized in Table 3. The profiles of the parameters can be correspondingly descripted according to the cluster centroids. For instance, cluster II characterized fires occurred in the Tunisia-Algeria border between 2010 and 2016. Fires in this cluster occurred only in June and after the postrevolution period. This spatiotemporal correlation can be explained, at partial, by the political instability and a widespread insecurity that Tunisia has experienced since 2010.

Cluster I was mainly typified by nearby road and populated areas, while cluster VI showed fires occurred far from the roads. Cluster III showed similar climatic features to those in cluster V, except for the relative humidity. Cluster VI showed fires occurred at high altitude and temperatures. In contrast, burned areas in cluster VII were correlated with distinct meteorological data from other clusters and were close to fire walls. The other interpretation that can be found from this primary classification is in consonance with the number of observations in each cluster. The clusters, having a small number of fires, notably cluster V and cluster VII, are in strong correlation with the period of 80's and 90's. According to their centroids, most fires are markedly away from roads and populated areas in comparison with clusters containing larger number of fires. Fires in clusters concerning the 2000's and 2010's decades (cluster I and cluster III) were associated with more adjacent roads and rural areas. This is particularly remarkable in these clusters since most fire incidents were recorded starting from 2002 and continuing until 2016. These primary indicators showed that the increasing trend of fire occurrences is related to the human causes. Similarly, in comparison with clusters presenting fires mainly occurred before 2002 (cluster V and cluster VII), the mean distance from firewalls is slightly reduced, as well, as the mean distance from the watch towers. This demonstrates that the forest surveillance and the firewall policy management were not effective in the last two decades. On the contrary, fire samples in cluster IV are not clearly distinguishable based on the years of occurrences. The flammability level was the highest compared to other clusters, which indicates the presence of highly flammable species such as Aleppo pine and maritime pine. Forest fires are then a recurrent phenomenon in the areas of cluster IV, due to the flammability of the vegetation during the Mediterranean dry season.

Transition and Emission Probabilities

HMMs are usually convenient to model temporal behavior as they infer optimal hidden states from observation sequences. Here, we aim to infer the estimated burned area from fire factors that correspond to each state. All in all, seven clusters representing fire factors were employed to infer four burned area classes. The HMM structure is then as follows: S = {*massive, large, considerable, limited*} and O = {*I, II, III, IV, V, VI, VII*}. In addition to define the HMM states and observations, we subsequently determined the

Table 3. Cluster centroids.

Factors and Periods	Attribute	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
Weather	RH	34.11	34.242	29.205	32.928	44.184	37.111	53.442
	Wind speed	2.88	2.506	2.615	2.803	2.673	2.263	3.608
	Temperature	29.647	27.601	30.067	29.585	26.054	30.182	24.899
Topomorphology	Slope	8.13	7.25	8.45	8.191	9.053	9.508	7.923
	Exposition	166.074	228.031	184.8	189.455	181.84	227.683	162.769
	Altitude	500.382	468.875	513.721	572.855	527.413	592.147	544.769
Vegetation Fuel	Flammability	1.03	0.811	1.418	2.763	0.326	0.533	0.091
Anthropogenic	Roads	224.718	2241.525	329.01	310.780	989.569	1038.112	587.965
	Settlements	937.001	22827.468	2249.452	1895.533	18666.203	4275.937	15185.671
Other fire factors	Fire wall	924.686	4361.953	1939.214	2635.983	9368.182	1306.770	15252.842
	Lookout posts	2647.335	11255.507	2254.045	4037.097	7010.294	1294.157	828.741
	Border	12844.349	1128.746	2578.557	3976.3	1820.669	7828.236	2533.891
Fire Years	1985–1993	0.03	0.074	0.053	0.144	0.452	0.239	0.578
	1994–2001	0.086	0.095	0.079	0.096	0.468	0.248	0.422
	2002–2005	0.308	0	0.325	0.17	0	0.068	0
	2006–2009	0.253	0	0.281	0.197	0	0.162	0
	2010–2012	0.157	0.347	0.105	0.207	0.016	0.188	0
	2013–2016	0.121	0.484	0.158	0.186	0.065	0.094	0
Fire Months	April	0	0	0	0	0.016	0	0
	May	0	0	0	0	0.048	0	0
	June	0.51	1	0	0.463	0.339	0.463	0
	July	0.237	0	0.474	0.399	0.306	0.399	0.453
	August	0.207	0	0.526	0.138	0.194	0.138	0.547
	September	0	0	0	0	0.065	0	0
	October	0	0	0	0	0.032	0	0
Forest Fires		189	95	114	188	62	117	64

initial state probabilities (π) based on the frequencies of first fire classes in each parcel. In 73 affected parcels, the first fires were classified into 29 massive fires, 22 large fires, 14 considerable fires, and 8 limited fires. The initial state distribution is then partitioned as $I = \{0.4, 0.3, 0.19, \text{and } 0.11\}$.

The learning set consists of 73 sequences with 829 fires in total, during the period from 1985 to 2016.

The sequence used to test the classifier was not used for training. Based on the Baum-Welch algorithm (Rabiner 1989), the estimated Transition Probabilities Matrix (TPM, Table 4) and Emission Probabilities Matrix (EPM, Table 5) were accordingly obtained. This requires sufficient amount of training sequences to provide fitting model parameters, which in our case could not be met by the available data set. To cope with this problem we applied the "leave-one-out" technique (Ko et al. 2009). All sequences excluding the one being classified were used to estimate the model parameters; this procedure was repeated for each tested sequence.

The TPM reports state transition probabilities, and the EPM reports the state dependent observation probability distributions. From the TPM, some dominant transitions between different states were clearly observed. In the parcel where a massive fire occurred, it is very possible that the next fire incident will be less intensive, giving high probabilities of being limited or considerable (0.360 and 0.422, respectively). This can be explained by the more effective prevention measures taken for the parcels where important losses and damage were recorded. The transitions from limited or considerable states to massive or large states confirm this interpretation, as their probabilities were lower with 0.151 for "considerable to massive," 0.104 for "considerable to large," 0.133 for "limited to massive," and 0.110 for "limited to large." In contrast, the probabilities estimated for staying in the same state of limited or considerable (0.325 and 0.42, respectively), and for the

Table 4. Transition probabilities matrix (TPM).

States (t-1) \ States(t)	Massive	Large	Considerable	Limited
Massive	0.134	0.152	0.360	0.422
Large	0.1	0.176	0.224	0.5
Considerable	0.151	0.104	0.325	0.42
Limited	0.133	0.110	0.316	0.441

Table 5. Emission probabilities matrix (EPM).

States \ Factors	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
Massive	0	0.329	0.177	0.045	0.271	0.034	0.144
Large	0.093	0.311	0.15	0.041	0.246	0	0.159
Considerable	0.332	0.025	0.09	0.212	0	0.234	0.107
Limited	0.328	0	0.041	0.220	0.126	0.139	0.146

transitions from limited to considerable states and reversely (0.316 and 0.441, respectively) were very high. This indicates that fire factors in most parts of the corresponding parcels were allied to the class of small fires. It can also be explained by the availability of continuous prevention and protection activities in these areas (Ben Jamâa and Abdelmoula 2004; Meddour-Sahar 2013).

We can also take advantage of the EPM for evidence suggesting that a class of burned area may have been associated with a cluster of fire factors. For example, we found that the observation of cluster II has an elevated likelihood of causing massive or large burned area with probabilities of 0.379 and 0.311, respectively. Then, the estimation of massive or large burned areas (two states) could be based solely on cluster analysis of fire factors, involving the cluster II (observation). This can be justified by the nearby Tunisia-Algeria border where fire fighters cannot gain access to burned areas. The other factors, that specify cluster II and cause important fire losses, could be the significant distance to roads and villages from these areas. On the contrary, cluster I and cluster VI are clearly associated with the classes of limited and considerable burned areas. The reason of this correlation is most likely because of the adjacency to roads and settlements (cluster I), and also because of the close lookout posts (cluster VI). The cluster V showed important probabilities of causing massive and large burned areas (0.271 and 0.246, respectively). This can be justified by the topomorphological conditions in the areas that belong to cluster IV. In addition to southwest aspect (the centroid for the aspect is 227.683 degree), fires of this cluster were associated to sites with higher altitude and steeper slopes (it has the highest centroids for the altitude and the slope degree with respectively 9.508 and 592.147). These areas are critical in terms of forest fire starts and spread, since more direct sunlight generally falls on the southwest slopes and higher relief (with resulting higher temperatures and sparser and lighter fuel loadings) (Dagorne et al. 1994).

Classification Performance

Since the recognition of burned area size is the main interest in this work, the confusion matrix is then used to store the correct and incorrect classifications made by the HMM model. In order to use the proposed HMM model as a classifier, the Viterbi algorithm was employed to decide the most likely "state" sequence given the model and its matching observation sequence. And then to recognize a burned area class for a given fire, two steps should be considered: in the first one, the clusters of fire factors, characterizing the fire in question and the previous fires occurred in the corresponding parcel, should be reorganized in an ordered sequence. Once we have constructed the sequence of clusters, the state sequence could be determined by the Viterbi process and the end-state will be the conforming burned area class.

Tables 6 and 7 show respectively the accuracies and the confusion matrix for burned area size classification using 10-fold cross-validation. These tables show that the method performed well for all burned area classes, that is, with 84% average class accuracy. These results fared well compared with the other models in the literature, even though there were not many studies performed for burned area estimation. In the study of Ozbayoglu and Bozer (2012), the highest accuracy level was above 65% using an MLP with only two inputs (humidity and wind speed) and three classes (small, medium, and big fire) in the output. In this regard it should be noted that the estimation results in these types of studies are very dependent on the quality of the data set. Cortez and Morais reported an accuracy rate of 60% in a similar study. In another study by Sakr, Elhajj, and Mitri (2011), the detection of fire days showed an accuracy of 92%. However, they did not perform burned area estimation.

Results from Table 6 indicate that the accuracy is relatively better for smaller fires with 85.81% and 89.84% for limited and considerable classes respectively comparing to 79.47% for large class and 79.47% for massive class. This constatation can be explained by analyzing the confusion matrix (Table 7) that shows an important number of massive and large fires classified as smaller fires (limited and considerable). The reason for missing classifications was that many fires able to be large or massive were successfully controlled. It should be noted that the estimation results of bigger fires are satisfactory compared to other models, even though there were not many studies performed for burned area estimation. The results from Cortez and

Table 6. The classification performance in terms of accuracy. The overall accuracy is calculated as the total number of correctly classified fires divided by the total number of test fires. The average accuracy is calculated as the sum of the accuracy figures in column Accuracy divided by the number of classes in the test set.

Class Accuracy	
Burned area size	Rate (%)
Limited	85.806
Considerable	89.841
Large	80.288
Massive	79.47
Overall Accuracy	84.801
Average Accuracy	83.852

Table 7. Confusion matrix for burned area classification.

Confusion matrix				
	Limited	Considerable	Large	Massive
Limited	133	11	7	4
Considerable	14	283	12	6
Large	18	11	167	12
Massive	15	7	9	120

Morais (2007) and Ozbayoglu and Bozer (2012) achieved lower accuracies of 46% and 57.24% for large fires, respectively. One explanation of this findings can be attributed to is the large number of very small fires compared to the number of big fires, since it can affect negatively the learning process and its results.

Comparison with Other Classification Techniques

Besides evaluating the used model's performance in terms of classification accuracy, it was also compared with other well-known classifiers, including back-propagation NN (Nikoskinen 2015), Random Forests (RF) (Hastie, Tibshirani, and Friedman 2009), second-degree Support Vector Machine (SVM-2) (Schölkopf and Smola 2002), and Isotonic Regression (IR) (Hoffmann 2009). To this end, we used the implementations provided by the Weka machine learning environment (Machine Learning Project 2016).

As we did for the HMM model, we performed some data pre-processing tasks and a preliminary analysis to tune the parameters for each considered techniques. The discrete variables were transformed into a 1-of-C encoding as done above. According to Hastie, Tibshirani, and Friedman (2009), all attributes were standardized to a zero mean and one standard deviation for the SVM and NN methods. The back-propagation NN utilized is a multilayer perceptron (MLP) with one hidden layer of eight nodes and logistic activation functions and one output node with a linear function (Hastie, Tibshirani, and Friedman 2009). Regarding the remaining methods, the SVM-2 and IR were used with their default parameters while the RF is developed with 500 trees (the default is 10).

Box-plots in Figure 6 show accuracies for the HMM approach and the classifiers tested. The HMM model obtained the best forecasting result compared to the other ML techniques. Thus, it becomes clear the superiority of stochastic approach for this study in relation to the standard supervised classification techniques. This indicates that the HMM was able to use the raw data and temporal information for better classification results. We speculate, therefore, that the hidden temporal fire factors, such as the dynamic aspect of climate change and the development of fire control policy, were modeled well by the HMM. Although, the percentage of correct classification for MLP is lower than those of all other considered ML techniques. A marked difference was observed among the remaining methods: SVM-2 and RF performed well on this problem, while IL performed less well. Thus, we assume that the complex relationships hidden in the data cannot be fittingly determined by a linear model, as well as all possible interactions between predictor variables (Tu 1996). This deduction can be supported by the outperformance of RF and SVM-2 that can produce nonlinear models. The poor performance of NN on this problem deserves further discussion. Makridakis

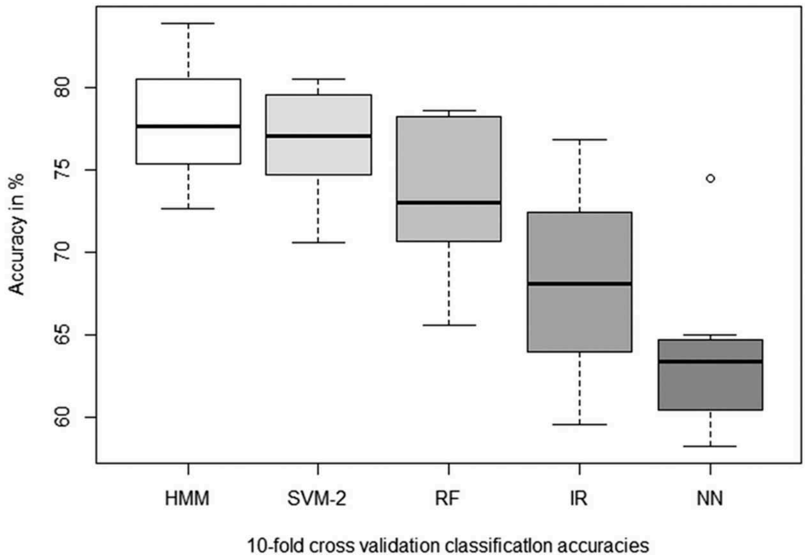

Figure 6. Boxplots of accuracy in the 10-fold cross-validation for the considered machine learning techniques (number of correct classifications/total sample size). On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the dots are the outliers.

and Hibon (2000) have indicated that handling of time series data in NN is a very complicated topic to achieve improved performance. It may be also due to the MLP architecture since that some complex, nonlinear functions cannot be efficiently represented by only basic hidden layers and thus low levels of nonlinear transformations (Nikoskinen 2015).

In addition to the superiority in accuracy of proposed HMM over compared ML algorithms, the HMM analysis is more informative by using transition and emission matrices. A primary exploration was done using Markov conditions (A matrices) to evaluate both the stochastic transitions among fire behaviors, and the importance of the various factors (B matrices). Such understanding provides insights into the model to better prevent fire and reduce damage when a fire occurs. However, despite their widespread adoption, machine learning models remain mostly black boxes (Ribeiro et al. 2016).

Conclusion

Using spatial and meteorological factors to reflect fire behavior, we have shown that hidden Markov models successfully predicted burned area sizes. Fire factors data were clustered using SOM before employing them as observations in the HMM. This clustering step permits to reduce variables of fire factors into clusters, allowing to use an HMM with discrete observations. It has also the advantage of exploring spatial and temporal patterns of fire factors. Model parameters allowed us to estimate expected behavioral states for fire losses (limited, considerable, large, and massive), as well as allowed to explore emission and transitions probabilities for evaluating the influence of effective factors on fire losses. The classification results of the study show that the proposed HMM outperformed the existing models and contributes to significant accuracy improvement, achieving an average of 83.85%. Its performance is also assessed with a state-of-the art machine learning methods evaluated using the same data sets, giving better or comparable results. In conclusion, data mining and stochastic processes in combination could be efficiently used to illustrate the behavioral processes and to monitor natural hazards.

Acknowledgements

The authors would like to thank the researcher Abdelaziz Hamdi and the forest engineers in Tunisian general directorate of forestry for helping in the identification of fire locations, and providing fire reports and forest inventories, respectively.

Funding

This work is part of a program of research supported by the laboratory of Silvo-pastoral resources (Silvopastoral Institute of Tabarka).

References

Ailliot, P., and V. Monbet. 2012. Markov-switching autoregressive models for wind time series. *Environmental Modelling and Software* 30:92–101. doi:10.1016/j.envsoft.2011.10.011.

Amatulli, G., M. J. Rodrigues, M. Trombetti, and R. Lovreglio. 2006. Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences* 111, no. G4.

Arpaci, A., B. Malowerschnig, O. Sass, and H. Vacik. 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography* 53:258–70. doi:10.1016/j.apgeog.2014.05.015.

Barber, C., J. Bockhorst, and P. Roebber. 2010. Auto-regressive HMM inference with incomplete data for short-horizon wind forecasting. In *Advances in Neural Information Processing Systems; Neural Information Processing Systems Foundation, Inc.: Barcelona, Spain, 2010;* 136–44.

Ben Jamâa, L., and K. Abdelmoula 2004, June. Les feux de forêts dans la suberaie tunisienne. In Colloque Vivexpo 2004:'Le chêne-liège face au feu.

Bisquert, M., E. Caselles, J. M. Sánchez, and V. Caselles. 2012. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *International Journal of Wildland Fire* 218:1025–29. doi:10.1071/WF11105.

Campos, P., H. Daly-Hassen, J. L. Oviedo, P. Ovando, and A. Chebil. 2008. Accounting for single and aggregated forest incomes: Application to public cork oak forests in Jerez Spain and Iteimia Tunisia. *Ecological Economics* 651:76–86. doi:10.1016/j.ecolecon.2007.06.001.

Castelli, M., L. Vanneschi, and A. Popovic. 2015. Predicting burned areas of forest fires: An artificial intelligence approach. *Fire Ecology* 111:106–18. doi:10.4996/fireecology.1101106.

Catry, F. X., F. Moreira, I. Duarte, and V. Acácio. 2009. Factors Affecting Post-Fire Crown Regeneration in Cork Oak (Quercus Suber L.) Trees. European Journal of Forest Research 128, no. 3: 231–240

Chakroun, H., F. Mouillot, M. Nouri, and Z. Nasr. 2012. Integrating MODIS images in a water budget model for dynamic functioning and drought simulation of a Mediterranean forest in Tunisia. *Hydrology and Earth System Sciences Discussions* 95:6251–84. doi:10.5194/hessd-9-6251-2012.

Cheng, T., and J. Wang. 2008. Integrated spatio-temporal data mining for forest fire prediction. *Transactions in Gis 12, No* 5 12:591–611. doi:10.1111/tgis.2008.12.issue-5.

Chou, Y. H. 1992. Management of wildfires with a geographical information system. *International Journal of Geographical Information Systems* 62:123–40. doi:10.1080/02693799208901900.

Cortez, P., and A. Morais. 2007. A Data Mining approach to predict Forest Fires using Meteorological data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 – Portuguese Conference on Artificial Intelligence, December Guimaraes, Portugal, pp. 512–523.

Dagorne, A., Y. Duché, J. Castex, and J. Ottavi. 1994. Protection des forêts contre l’incendie et système d’information géographique. *Forêts Méditerranéennes 15, No* 4:422–425.

Dimitrakopoulos, A. 2001. A statistical classification of mediterranean species based on their flammability components. *International Journal Of Wildland Fire 10, No* 2 10:113–118. doi:10.1071/WF01004.

Dlamini, W. M. 2011. Application of Bayesian networks for fire risk mapping using GIS and remote sensing data. *GeoJournal* 763:283–96. doi:10.1007/s10708-010-9362-x.

EFFIS, 2016. European forest fires information system. http://forest.jrc.ec.europa.eu/effis (Accessed March, 2016).

Eugenio, F. C., A. R. Dos Santos, N. C. Fiedler, G. A. Ribeiro, A. G. Da Silva, Á. B. Dos Santos, … V. R. Schettino. 2016. Applying GIS to develop a model for forest fire risk: A case study in Espírito Santo, Brazil. *Journal of Environmental Management* 173:65–71. doi:10.1016/j.jenvman.2016.02.021.

Fao, I. 2013. *The multiple dimensions of food security*, WFP, The State of Food Insecurity in the World 2013. Rome: FAO.

Fao, I. 2016. *The multiple dimensions of food security*, WFP, The State of ood Insecurity in the World 2016. Rome: FAO.

FIRMS, 2015. Fire information for resource management system FIRMS. URL: http://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/, (accessed September 2015).

Giglio, L., J. Descloitres, C. O. Justice, and Y. J. Kaufman. 2003. An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment* 87 (2–3):273–282. doi:10.1016/S0034-4257(03)00184-6.

Giglio, L., J. Randerson, G. Van der Werf, P. Kasibhatla, G. Collatz, D. Morton, and R. DeFries. 2010. Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences 7, No* 3 7:1171–1186. doi:10.5194/bg-7-1171-2010.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Unsupervised learning. In *The elements of statistical learning*. Springer, New York: 485–585.

Henaoui, S. E. A., M. Bouazza, and M. Amara. 2013. The fire risk of the plant groupings with Cistus in the area of Tlemcen (Western Algeria). *European Scientific Journal* ESJ, 9, no. 29.

Hoffmann, L. 2009. Multivariate isotonic regression and its algorithms. Thesis, Wichita State University, Wichita, Kansas, USA.

Hsu, C. W., C. C. Chang, and C. J. Lin 2003. A practical guide to support vector classification.

Jones, M. R., H. J. Fowler, C. G. Kilsby, and S. Blenkinsop. 2013. An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *International Journal of Climatology* 33:1178e1194. doi:10.1002/joc.3503.

Jung, J., C. Kim, S. Jaya Kumar, S. Kim, S. Han, D. H. Kim, and J. Heo. 2013. Forest fire risk mapping of Kolli Hills, India, considering subjectivity and inconsistency issues. *Natural Hazards* 653:2129–46. doi:10.1007/s11069-012-0465-1.

Kalboussi, M., and H. Achour. 2018. Modelling the spatial distribution of snake species in Northwestern Tunisia using maximum entropy (maxent) and Geographic Information System (GIS). *Journal Of Forestry Research 29, No* 1 29:233–245. doi:10.1007/s11676-017-0436-1.

Ko, A. H. R., P. R. Cavalin, R. Sabourin, and J. A. De Souza Britto. 2009. Leave-one-out-training and leave-one-out-testing hidden markov models for a handwritten numeral recognizer: The implications of a single classifier and multiple classifications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3112:2168–78. doi:10.1109/TPAMI.2008.254.

Kohonen, T., S. Kaski, and H. Lappalainen. 1997. Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural Computation* 96:1321–44. doi:10.1162/neco.1997.9.6.1321.

Ku-Mahamud, K. R., and J. Y. Khor. 2009. Pattern extraction and rule generation of forest fire using sliding window technique. *Computer and Information Science* 23:113.

Li, L., Z. Ma, L. Liu, and Y. Fan. 2013. Hadoop-based arima algorithm and its application in weather forecast. *International Journal of Database Theory and Application* 65:119–32. doi:10.14257/ijdta.2013.6.5.11.

Lozano, F. J., S. Suárez-Seoane, M. Kelly, and E. Luis. 2008. A multi-scale approach for modelling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment* 1123:708–19. doi:10.1016/j.rse.2007.06.006.

Machine Learning Project, 2016. Weka—Waikato environment for learning analysis. http://www.cs.waikato.ac.nz/ml/index.html (Accessed 11 Aout, 2016).

Makridakis, S., and M. Hibon. 2000. The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting* 164:451–76. doi:10.1016/S0169-2070(00)00057-1.

Malarz, K., S. Kaczanowska, and K. Kułakowski. 2002. Are forest fires predictable? *International Journal of Modern Physics C* 1308:1017–31. doi:10.1142/S0129183102003760.

Martínez, J., C. Vega-Garcia, and E. Chuvieco. 2009. Human-caused wildfire risk rating for prevention planning in spain. *Journal Of Environmental Management 90, No* 2 90:1241–1252. doi:10.1016/j.jenvman.2008.07.005.

Martínez-Fernández, J., E. Chuvieco, and N. Koutsias. 2013. Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Sciences* 132:311. doi:10.5194/nhess-13-311-2013.

Meddour-Sahar, O., R. Meddour, V. Leone, R. Lovreglio, and A. Derridj. 2013a. Analysis of forest fires causes and their motivations in North Algeria: The Delphi method. *iForest -Biogeosciences and Forestry* 26:247–54. doi:10.3832/ifor0098-006.

Menard, S. 2002. Applied logistic regression analysis No. 106. Sage. doi:10.1044/1059-0889(2002/er01)

NASA, 2016. Atmospheric science data center, surface meteorology and solar energy. http://eosweb.larc.nasa.gov/sse/, (accessed July 2016).

Nieto, H., I. Aguado, M. García, and E. Chuvieco. 2012. Lightning-caused fires in Central Spain: Development of a probability model of occurrence for two Spanish regions. *Agricultural and Forest Meteorology* 162:35–43. doi:10.1016/j.agrformet.2012.04.002.

Nikoskinen, T. 2015. *From neural networks to deep neural networks*. Aalto University, School of Science. http://sal.aalto.fi/publications/pdf-files/enik15_public.pdf.

Oliveira, T. M., A. M. Barros, A. A. Ager, and P. M. Fernandes. 2016. Assessing the effect of a fuel break network to reduce burnt area and wildfire risk transmission. *International Journal of Wildland Fire* 256:619–32. doi:10.1071/WF15146.

Özbayoğlu, A. M., and R. Bozer. 2012. Estimation of the burned area in forest fires using computational intelligence techniques. *Procedia Computer Science* 12:282–87. doi:10.1016/j.procs.2012.09.070.

Parisien, M. A., S. Snetsinger, J. A. Greenberg, C. R. Nelson, T. Schoennagel, S. Z. Dobrowski, and M. A. Moritz. 2012. Spatial variability in wildfire probability across the western United States. *International Journal of Wildland Fire* 214:313–27. doi:10.1071/WF11044.

Park, Y. S., and Y. J. Chung. 2006. Hazard rating of pine trees from a forest insect pest using artificial neural networks. *Forest Ecology and Management* 2221:222–33. doi:10.1016/j.foreco.2005.10.009.

Pausas, J. G. 2004. Changes in fire and climate in the eastern Iberian Peninsula Mediterranean basin. *Climatic Change* 633:337–50. doi:10.1023/B:CLIM.0000018508.94901.9c.

Pourtaghi, Z. S., H. R. Pourghasemi, R. Aretano, and T. Semeraro. 2016. Investigation of general indicators influencing on forest fire and its susceptibility modelling using different data mining techniques. *Ecological Indicators* 64:72–84. doi:10.1016/j.ecolind.2015.12.030.

Preisler, H. K., A. L. Westerling, K. M. Gebert, F. Munoz-Arriola, and T. P. Holmes. 2011. Spatially explicit forecasts of large wildland fire probability and suppression costs for California. *International Journal of Wildland Fire* 204:508–17. doi:10.1071/WF09087.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 772:257–86. doi:10.1109/5.18626.

Renard, Q., R. Pélissier, B. Ramesh, and N. Kodandapani. 2012. Environmental susceptibility model for predicting forest fire occurrence in the western ghats of India. *International Journal Of Wildland Fire 21, No* 4 21:368–379. doi:10.1071/WF10109.

Ribeiro, M.T., S. Singh, and C. Guestrin. 2016. "Why should i trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. ACM. doi:10.18653/v1/N16-3

Rodrigues, M., J. De La Riva, and S. Fotheringham. 2014. Modelling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography* 48:52–63. doi:10.1016/j.apgeog.2014.01.011.

Safi, Y., and A. Bouroumi. 2013. Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences* 76:271–86. doi:10.12988/ams.2013.13025.

Sakr, G. E., I. H. Elhajj, and G. Mitri. 2011. Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Engineering Applications of Artificial Intelligence* 245:888–94.

Schölkopf, B., and A. J. Smola. 2002. *Learning with Kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA,: MIT press.

Soto, M. C., and G. Alvear 2012. Forest fires in wildland-urban interface. Analysis and proposed actions. Case study: Central Chile-South America. Wildfire and Community: Facilitating preparedness and resilience, 106.

Stojanova, D., M. Ceci, A. Appice, and S. Džeroski 2011, September. Network regression with predictive clustering trees. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases pp.333–48. Springer Berlin Heidelberg. doi:10.1177/1753193411402002

Tao, J., Y. Zhang, X. Yuan, J. Wang, and X. Zhang. 2013. Analysis of forest fires in Northeast China from 2003 to 2011. *International Journal of Remote Sensing* 3422:8235–51. doi:10.1080/01431161.2013.837229.

Tasdemir, K., and P. Milenov 2010, April. An automated SOM clustering based on data topology. In ESANN.

Testa, A. C., J. K. Hane, S. R. Ellwood, and R. P. Oliver. 2015. CodingQuarry: Highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 161:170. doi:10.1186/s12864-015-1344-4.

Thomas, N. G., P. Pasquier, A. Eigenfeldt, and J. B. Maxwell 2013. A methodology for the comparison of melodic generation models using meta-melo. In ISMIR pp. 561–66. doi:10.3748/wjg.v19.i4.561

Tu, J. V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology* 49:1225–31. doi:10.1016/S0895-4356(96)00002-9.

Turco, M., M. C. Llasat, J. von Hardenberg, and A. Provenzale. 2014. Climate change impacts on wildfires in a Mediterranean environment. *Climatic Change* 1253-4:369–80. doi:10.1007/s10584-014-1183-3.

Vesanto, J., and E. Alhoniemi. 2000. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks* 113:586–600. doi:10.1109/72.846731.

Woolford, D. G., D. A. Stanford, R. J. Kulperger, D. Boychuk, and B. M. Wotton. 2009. Erlangian approximations for the transient analysis of a fluid queue model for forest fire perimeter. *Infor* 474:305.

Xie, D. W., and S. L. Shi. 2014. Prediction for burned area of forest fires based on SVM model. *Applied Mechanics and Materials* 513:4084–89. doi:10.4028/www.scientific.net/AMM.513-517.4084.

Yu, X., and S. Sun. 2016. HMM-DM: Identifying differentially methylated regions using a hidden Markov model. *Statistical Applications in Genetics and Molecular Biology* 151:69–81.