Applied Soft Computing Journal 89 (2020) 106075

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

A wildfire warning system applied to the state of Acre in the Brazilian Amazon

I.D.B. Silva [b,*], M.E. Valle [a], L.C. Barros [a], J.F.C.A. Meyer [a]

[a] Department of Applied Mathematics, University of Campinas-UNICAMP, Campinas-SP, Brazil
[b] Center of Exact and Technological Sciences, Federal University of Acre-UFAC, Rio Branco-AC, Brazil

ARTICLE INFO

Article history:
Received 6 July 2018
Received in revised form 30 December 2019
Accepted 5 January 2020
Available online 15 January 2020

Keywords:
Wildfire indexes
Fuzzy system
Machine learning
Forest fires

ABSTRACT

In this paper, we present a dynamic wildfire warning map that combines both spatial and weather information. In particular, our wildfire early warning model is obtained by aggregating two indexes called wildfire risk and wildfire danger. The wildfire risk index, which is based on georeferenced features such as altitude and forest type, measures the fuel necessary for a wildfire to start at a certain location on a map. The wildfire danger uses weather conditions to yield temporal information concerning the possibility of a wildfire to spread. Machine learning techniques and fuzzy logic operations are used to determine the wildfire risk and danger indexes from available data. Although both wildfire risk and wildfire danger indexes can be used separately, using concepts from fuzzy logic, they can be combined to yield a wildfire warning system that takes into account both weather and static information. We illustrate the wildfire early warning model by considering weather and geographical data for the state of Acre.

1. Introduction

Forest fires that occur in the Amazon as a whole are the result of severe droughts that have occurred in the region and cut–burn practices for cleaning land that often become uncontrollable [1]. The environmental impact of such burning affects the quality of the soil and air, the forest is increasingly prone to new fires, biodiversity is destroyed, and many other negative effects occur [2].

The statistics on the occurrence of forest fires in Brazil and many other countries show that most fires start from fire sources arising, directly or indirectly, from human activities. Indeed, the vast majority of forest fires occur because of human causes [3]. Furthermore, the spread and the intensity of a forest fire are usually associated with the following environmental factors: vegetation cover, relief, weather conditions, and fuel type [3–5]. Each of them has a distinct contribution to the degree of risk and danger of local fires. The use of accurate weather data is important for planning, prevention, and fighting forest fires [3,6]. One of the preventive measures is the use of reliable risk and danger indices.

In Brazil, the first index have been proposed in the early 1970s by Soares using meteorological data and the occurrence of fires in the Monte Alegre farm at the central region of the state of Paraná, Brazil [3]. A forest fire index currently used in Brazil is based on the early works of Soares and it is referred to as the formula of Monte Alegre (FMA) [7–10]. In few words, the FMA yields a non-negative real number based on the relative humidity and the number of days that do not rain significantly. The FMA is then used to obtain crisp sets that describe null, low, medium, high, and very high wildfire risk. In this index, human factors are not explicitly included. At this point, we would like to recall that the FMA have been conceived as a forest fire index for a specific region. Thus, its applicability is somewhat limited to humid subtropical weather which is far different from the hot and very humid equatorial weather at the Amazon Rain Forest. This remark instigated us to pursuit a fire forest index which could be used as a wildfire warning at the Amazon. The following presents a brief review of fire forest indices from the literature in a worldwide context.

1.1. Fire forest indices: a literature review

Several countries have proposed, either through government agencies or through independent researchers, forest fire information systems that are based on wildfire risk or danger indices. For example, since 1960s the *Canadian forest service* has been using the *Canadian forest fire hazard assessment system* (CFFDRS), which is based on two concepts: the *fire weather index* (FWI) and the *fire behavior prediction* (FBP). The first seeks to anticipate

* Corresponding author.
E-mail addresses: isaacdayanbs@gmail.com (I.D.B. Silva), valle@ime.unicamp.br (M.E. Valle), laeciocb@ime.unicamp.br (L.C. Barros), joni@ime.unicamp.br (J.F.C.A. Meyer).

https://doi.org/10.1016/j.asoc.2020.106075
1568-4946/

the potential for daily fire ignition. The latter estimates the fire evolution pattern originated from a single source [11].

Another system that couples several forest fire risk indices is the *national fire hazard assessment system* (NFDRS) developed in the United States. The NFDRS takes into account the effect of fuel moisture content, weather, and slope on the spreading rate of fire by a semi-physical propagation model [12–14]. As in FMA, CFFDRS and NFDRS also do not include human factors explicitly.

The fire risk system proposed by Fiorucci (2008) [4] was based on the CFFDRS and NFDRS systems and involves three models: Fuel load, fuel humidity, and potential fire spread. This model stands out from the others because it uses differential calculus to estimate the fuel moisture, but the extensive amount of parameters to be estimated makes it difficult to be applied in real situations.

A common approach to ascertaining the wildfire risk and hazard is the use of formulas based on linear combination of the variables available on georeferenced maps manipulated by geographic information system (GIS) softwares [6,15]. Furthermore, statistical techniques are usually applied to obtain hazard or risk indices associated to the probability of fire occurrence [16–18]. For instance, statistical data were used to obtain the fire hazard index developed in Australia by Mc Arthur [19]. More recently, Bolourchi and Uysal proposed a fuzzy rule-based system that yields the probability of fire using inputs from a wireless sensor network [20]. Garcia–Jimenez et al. improved the fuzzy system of Bolourchi and Uysal by using overlapping functions in the inference mechanism to deal with uncertainty and vagueness in the formulation of the fuzzy rule base [21]. Despite their wide applicability, probability distributions, differential equations, or mathematical relations used to obtain the wildfire risk or hazard indices are often assumed known *a priori* in the aforementioned approaches.

Alternative approaches to predict forest fire risk are based on machine learning systems that learn from available data. Alonso-Betanzos et al. and Goldarag used multilayer neural networks to classify the presence or absence of heat sources in regions of Spain and Iran, respectively [16,22]. Briefly, they used as input weather variables that were extrapolated/interpolated in space from positions of meteorological stations. The output of the network is a real number in the unit interval [0, 1] that is interpreted as a forest fire risk. In a similar vein, Adab et al. Justino et al. and Matin et al. proposed wildfire risk or danger models whose parameters are estimated empirically using a large number of georeferenced input variables [23–25].

At this point, we would like to point out that San-Miguel-Ayanz proposed a classification of wildfire indices in long-term and short-term according to the time variation of the variables [26]. Short-term indices use variables that are modified almost continuously such as weather information. Long-term indices depend on variables that are modified a little or not at all in the short and medium term. For instance, the type of vegetation and topography of the studied area are variables of long-term indices. Short and long-term indices can be combined to yield a dynamic index. In a similar fashion, in this paper we present an early warning system that combines two wildfire indices which can be classified according to the long-term and short-term criteria of San-Miguel-Ayanz. Furthermore, the wildfire indexes are obtained using soft computing and machine learning techniques. As a consequence, the new approach can be derived from data but it can also deal with the uncertainty inherent to them.

1.2. Contributions and organization of the paper

In this paper, spatial variables, such as altitude and forest type, are used to define a long-term index called *wildfire risk* and denoted by the symbol ρ. The wildfire risk, which can be used to determine a map of a spatial risk, is obtained using a classifier whose output measures the potential of a fire to ignite at a certain location. In this paper, we propose to use the fuzzy k-nearest neighbors (fuzzy *k*NN) classifier which yields a value in the unit interval where 0 represents no risk of fire while 1 is the largest risk of fire. The temporal information produces the short-term index called *wildfire danger* and denoted by π. The wildfire danger uses weather variables such as relative humidity and rainfall to yield a time series. Finally, the wildfire risk and danger indices can be combined in a new index denominated *wildfire warning* and denoted by the symbol $\mathbb{W}$. The wildfire warning is used for early warning and decision-making.

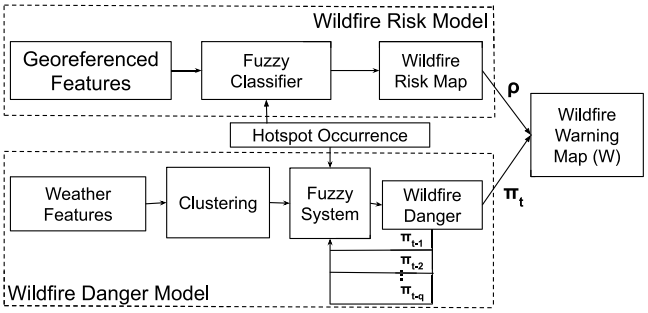


Fig. 1. Block diagram of the model of wildfire warning system which combines the spatial wildfire risk ρ and the temporal wildfire danger π_t.

Since they are based on machine learning techniques, both wildfire risk and wildfire danger – and thus the wildfire warning – can be estimated from a representative set of data. Moreover, the relationship between the input variables and the risk or danger indices is not defined *a priori*, but rather learned from data. In our opinion, one advantage of our approach is that the input and output variables can be easily obtained — they do not require physical, chemical experiments nor empirical formulas. On the downside, it requires a large amount of available data. Another positive aspect of our approach is that the amount of input variables and the parameters to be estimated or fine-tuned is small compared to many other machine learning models. Furthermore, the use of fuzzy system helps the model to deal with uncertainty inherent to the complex weather variables. In fact, despite using few input variables, the results obtained in the experiments are quite promising. The block diagram depicted in Fig. 1 summarizes the wildfire warning model used in our experiments.

This paper is organized as follows. The next section describes the methodology. Namely, the wildfire risk, which is determined using a fuzzy classifier, is described in Section 2.1. We discuss the wildfire danger, which is given by a fuzzy system and machine learning techniques, in Section 2.2. We finish the methodological section with the wildfire warning index, which is obtained by combining the wildfire risk and danger indexes. Section 3 applies the methodology in a case study for the State of Acre-Brazil. The paper finishes with some concluding remarks in Section 4.

2. Methodology — wildfire risk, danger and warning models

In this section, we present the wildfire risk, danger, and warning models, respectively.

2.1. Wildfire risk model

The wildfire risk index is related to the propitious (or necessary) conditions for a wildfire to start. In mathematical terms, the wildfire risk is characterized by a mapping $\rho : \Omega \to [0, 1]$ such

that $\rho(\mathbf{x}) \in [0, 1]$ represents the risk of a wildfire to ignite at a certain location with features $\mathbf{x} = (x_1, \ldots, x_N) \in \Omega$, where $\Omega \subseteq \mathbb{R}^m$ is a set of geographically referenced data such as altitude, forest typology, and distance to the closest watercourse or road. Like San-Miguel-Ayanz [26], we assume that the feature set does not change significantly in a certain period of time. Furthermore, the feature set must be strongly related to the occurrence of wildfires but it should also be easily measurable. The following references describe some features which are strongly related to the occurrence of forest fires [5,27,28].

Let us assume that the wildfire risk map can be determined from a representative data set

$$\mathcal{T}_L = \{(\mathbf{x}_\ell, \omega_\ell) : \ell = 1, \ldots, L\}, \tag{1}$$

referred to as the *training set*. Here, $\mathbf{x}_\ell = (x_{1\ell}, x_{2\ell}, \ldots, x_{N\ell})$, for $\ell = 1, \ldots, L$, is the feature vector containing georeferenced input information from the model and $\omega_\ell \in \{0, 1\}$ indicates the occurrence (or not) of hotspot[1] at the location with feature $\mathbf{x}_\ell$ over a certain period of time. Precisely, the values $\omega_\ell = 0$ and $\omega_\ell = 1$ represent respectively the absence or the presence of a hotspot at $\mathbf{x}_\ell$ on the period of time.

We shall interpret the wildfire risk mapping ρ as a fuzzy set on Ω. In other words, the value $\rho(\mathbf{x})$ corresponds to the pertinence of $\mathbf{x}_\ell \in \Omega$ to the fuzzy set of all hotspot points. From a practical point of view, the wildfire risk $\rho : \Omega \to [0, 1]$ corresponds to a fuzzy classifier determined from the training set $\mathcal{T}_L$. In principle, any binary fuzzy classifier, including evolving fuzzy classifiers which can process streaming data only and in real time [30] – is a candidate model. For simplicity, however, in this paper we only use of the fuzzy k-nearest neighbor (fuzzy *k*NN) algorithm.

Using the fuzzy *k*NN algorithm, the wildfire risk $\rho(\mathbf{x})$ of a feature vector $\mathbf{x} \in \Omega$ corresponds to a weight average of the membership of its k nearest neighbors [31]. Formally, given an unclassified input vector $\mathbf{x} \in \Omega$, let

$$\left\{(\mathbf{x}_{(1)}, \omega_{(1)}), (\mathbf{x}_{(2)}, \omega_{(2)}), \ldots, (\mathbf{x}_{(k)}, \omega_{(k)})\right\} \subseteq \mathcal{T}_L,$$

denote the subset of the training set composed by the k nearest neighbors of $\mathbf{x}$. In other words, the following inequality holds true for all $i \leqslant k$ and $j > k$:

$$\|\mathbf{x} - \mathbf{x}_{(i)}\| \leqslant \|\mathbf{x} - x_{(j)}\|,$$

where $\|\cdot\|$ denotes a certain norm. In our approach, we considered the norm derived from the Mahalanobis metric that takes into account the correlation between the training data. In fact, the Mahalanobis distance between $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\|\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y})^T C^{-1} (\mathbf{x} - \mathbf{y}), \tag{2}$$

where C denotes the covariance matrix of all the vectors $\mathbf{x}_\ell$, $\ell = 1, \ldots, N$, of the training set [32,33].

Using the set of the k nearest neighbors, the fuzzy *k*NN assigns the membership degree to the unlabeled input $\mathbf{x}$ by means of the equation

$$\rho(\mathbf{x}) = \frac{\sum_{j=1}^{k} \omega_{(j)} \left(1/\left|\left|\mathbf{x} - \mathbf{x}_{(j)}\right|\right|\right)^2}{\sum_{i=1}^{k} \left(1/\left|\left|\mathbf{x} - \mathbf{x}_{(i)}\right|\right|\right)^2}, \tag{3}$$

In words, the membership degree of $\mathbf{x}$ in the fuzzy set of hotspot points is determined by an average of the pertinence of the k nearest neighbors weighted by the inverse of the square of the Mahalanobis distance. Summarizing, Algorithm 1 and the diagram shown in Fig. 2 illustrates the proposed methodology.

[1] Formally, a hotspot is a point on the planet – determined by a satellite – that are emanating high thermal radiation. In this work, we use data from the reference satellite NOAA-15 [29].

Algorithm 1: Wildfire Risk Algorithm

Input: Training set $\mathcal{T}_L = \{(\mathbf{x}_\ell, \omega_\ell) : \ell = 1, \ldots, L\}$, the input $\mathbf{x} \in \Omega$, and the nearest neighbors parameter k.

Output: The wildfire risk $\rho(\mathbf{x})$.

1. Compute the covariance matrix C of $\{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$;
2. Determine the Mahalanobis distance d_ℓ between the input $\mathbf{x}$ and all feature vectors $\mathbf{x}_\ell$, $\ell = 1, \ldots, L$;
3. Sort the distances in an ascending order such that $d_{(j)} \leqslant d_{(j+1)}$;
4. Define

$$\rho(\mathbf{x}) = \frac{\sum_{j=1}^{k} \omega_{(j)}/d_{(j)}^2}{\sum_{j=1}^{k} 1/d_{(j)}^2}.$$

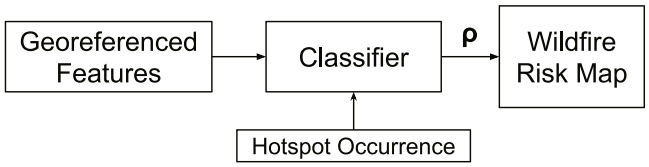


Fig. 2. Block diagram of the wildfire risk model obtained from the classifier.

We suggest to determine k using a cross-validation technique such as the (*10-fold cross validation*) [33]. In this case, the available data is divided into 10 disjoint subsets. Nine subsets are used for training while the remaining set is used for validation. The process is repeated until all 10 subsets have been used for validation. Then, we select the value of k that results in the smallest average error. The final wildfire risk map is determined using the entire data set $\mathcal{T}_L$.

Finally, we would like to point out that the geographic coordinates (i, j) of a certain location are not used to derive the wildfire risk $\rho(x)$, but only its features. Notwithstanding, the geographic coordinates can be used to plot a *wildfire risk map*, that is, we obtain the wildfire risk map by evaluating ρ at the features $\mathbf{x} \equiv \mathbf{x}(i, j)$ of the point at coordinates (i, j).

2.2. Wildfire danger model

According to Flannigan [34], fires are strongly linked to weather and climatic conditions. Weather is a critical factor in determining fire behavior and the size of the burned area. In this section, we present the wildfire danger modeling that depends only on climatic information.

Suppose we have a representative training set $\mathcal{T}_T = \{(\mathbf{y}_t, F_t) : t \in 1, \ldots, T\}$ of historical data from a certain location, where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Mt})$ denotes an M-dimensional weather feature vector at time $t \in \{1, \ldots, T\}$. Accordingly to San-Miguel-Ayanz, the weather features, such as relative air humidity and rainfall, may change significantly from one instant of time to the following. Also, let $F_t \equiv F(t)$ be an indicator of the occurrence of hotspot at the instant of time $t \in \{1, \ldots, T\}$, that is, $F_t = 1$ if there is a hotspot in time t and $F_t = 0$ otherwise.

Based on the set of historical data $\mathcal{T}_T$, we determine the wildfire danger index as a function of the weather features $\pi_t \equiv \pi(\mathbf{y}_t) \in [0, 1]$. Like the wildfire risk, the wildfire danger $\pi_t \in [0, 1]$ measures the potential of a fire to start at the instant t, but it depends only on interpolated weather features. In the following, we describe a methodology for determining the wildfire danger π_t using a fuzzy partition of the weather features set as well as the corresponding danger factors $\{\phi_t : t = 1 \ldots, T\}$. Due to the temporal information, however, the wildfire danger is more complex than the wildfire risk presented in the previous section.

First of all, the weather features $\{\mathbf{y}_t, t = 1, \ldots, T\}$ are grouped using a fuzzy clustering technique, an unsupervised machine learning approach. Despite the many fuzzy clustering techniques such as fuzzy c-means and evolving fuzzy clustering, in this paper we only considered the *subtractive clustering* which is a fast and robust clustering technique that does not require any *a priori* knowledge of the number of clusters [35]. Any way, the fuzzy clustering must yield the center (or location) and spread of the fuzzy clusters. In the subtractive clustering, the location and spread of a cluster are determined using a single parameter $r_{max} \in [0, 1]$, which specifies the range of influence of the clusters [35]. We would like to point out that r_{max} can be obtained by an exhaustive search if the user does not have any information about this parameter. In other words, the parameter r_{max} can be determined from a finite set of candidate values by applying the cross-validation technique on the training set.

Let C denote the number of clusters and $\mathbf{c}_i = (c_{1i}, \ldots, c_{Mi}) \in \mathbb{R}^M$, for $i = 1, \ldots, C$, be the centers of the clusters yielded by the fuzzy clustering applied to the historical data. In addition, let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_M) \in \mathbb{R}^M$ denote the M-dimensional vector with the spread of each component of the weather feature set. Note that the temporal evolution of the weather information is not taken into account to determine the centers of the clusters and the spread of each weather features. We anticipate that this remark is particularly useful when there are long gaps of missing data for training the wildfire danger model.

Using the cluster center and spread, we define a fuzzy set whose Gaussian membership function μ_i is given by the following equation for any $\mathbf{y}_t = (y_{1t}, \ldots, y_{Mt})$ and $i = 1, \ldots, C$:

$$\mu_i(\mathbf{y}_t) = \exp\left[-\frac{1}{2}\sum_{j=1}^{M}\frac{(c_{ij} - y_{jt})^2}{\sigma_j^2}\right]. \tag{4}$$

The fuzzy sets characterized by the membership functions $\mu_1, \ldots, \mu_C$ define a fuzzy partition of the weather feature set. Furthermore, by considering the indicators of the occurrence of hotspot, we can assign to each fuzzy set a wildfire danger index f_i given by the equation

$$f_i = \frac{\sum_{t\in B_i} F_t \mu_i(\mathbf{y}_t)}{\sum_{t\in B_i} \mu_i(\mathbf{y}_t)}, \quad \forall i = 1, \ldots, C \tag{5}$$

where $B_i \subseteq \{1, \ldots, T\}$ is the set of instants t for which $\mu_i(\mathbf{y}_t)$ attains its maximum value. Formally, we have

$$B_i = \left\{t \in \{1, \ldots, T\} : \mu_i(\mathbf{y}_t) = \max_{k=1:C}\{\mu_k(\mathbf{y}_t)\}\right\}. \tag{6}$$

Note that f_i is given by a weighted average of the indicators of the occurrence of hotspot F_t for some instant t. In the following, we standardized the parameters $f_1, \ldots, f_m$ so that the maximum value is 1. Moreover, since a hotspot usually occurs in conditions of low humidity and rainfall, we zeroed the values f_i below a threshold $\tau > 0$. Broadly speaking, the threshold τ limits the fuzzy sets that contribute to the occurrence of the hotspot. Concluding, the *wildfire danger factor* of the ith fuzzy set, denoted by ϕ_i, is defined by the following equation for $i = 1, \ldots, C$:

$$\phi_i = \begin{cases} \dfrac{f_i}{\max_{j=1:C}\{f_j\}}, & \text{if } f_i \geqslant \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

We would like to point out that (7) improved significantly the performance of the model because it deals with unbalanced occurrence of hotspot and, thus, it incorporates further information about the danger of wildfires. The threshold value on (7) may depend on the region under study and it can be determined in several ways, such as: information from a specialist, from the literature, or by exhaustive search.

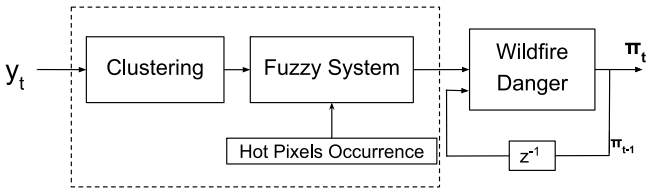


Fig. 3. Block diagram of the wildfire danger model.

The danger factors $\phi_1, \ldots, \phi_C$ and the fuzzy clusters characterized by the membership functions $\mu_1, \ldots, \mu_C$ are used to estimate the current instantaneous wildfire danger as follows. Given the current weather feature vector $\mathbf{y}_t$, with $t > T$, define the instantaneous wildfire danger $\tilde{\pi}(\mathbf{y}_t)$ by means of the equation

$$\tilde{\pi}(\mathbf{y}_t) = \frac{\sum_{i=1}^{C}\phi_i\mu_i(\mathbf{y}_t)}{\sum_{i=1}^{C}\mu_i(\mathbf{y}_t)}. \tag{8}$$

Note that the instantaneous wildfire danger $\tilde{\pi}(\mathbf{y}_t)$ does not take into account past information but only the weather features $\mathbf{y}_t$. On the one hand, the temporal evolution of the weather information is not totally necessary to determine the instantaneous wildfire danger function $\tilde{\pi}$. Hence, missing data does not significantly effect the training phase of the instantaneous wildfire danger model. On the other hand, the value $\tilde{\pi}(\mathbf{y}_t)$ may change significantly from one instant of time to the following. Most importantly, since the instantaneous wildfire danger function $\tilde{\pi}$ is static, it is unable to characterize a time window corresponding, for instance, to the driest period of the year. In view of this fact and in agreement with some danger indexes from literature [26, 36,37], instead of using the instantaneous danger index $\tilde{\pi}(\mathbf{y}_t)$, we considered an exponentially weighted moving average (EWMA) of the past danger risks [38]. Formally, we define the current wildfire danger $\pi_t \equiv \pi(\mathbf{y}_t)$ by means of the equation

$$\pi_t = \beta\tilde{\pi}(\mathbf{y}_t) + (1 - \beta)\pi_{t-1}. \tag{9}$$

where $\tilde{\pi}(\mathbf{y}_t)$ is the instantaneous wildfire danger, π_{t-1} is the past wildfire danger, and $\beta \in (0, 1]$ is the weighting factor known as "smoothing constant". In words, the current wildfire danger π_t depends on the current weather features $\mathbf{y}_t$ as well as on past wildfire dangers, which are incorporated implicitly by the term π_{t-1} in (9). In contrast to the memoryless instantaneous wildfire danger $\tilde{\pi}(\mathbf{y}_t)$, π_t is less susceptible to changes. Finally, the parameter β can be determined, for example, using an exhaustive search on the training set with a cross-validation technique.

A block diagram for the computation of the wildfire danger index is shown in Fig. 3. The training and evaluation algorithms are given respectively by Algorithms 2 and 3. Finally, we define the *weather warning* α_t by means of the equation

$$\alpha_t = \begin{cases} 1, & \text{if } \pi_t > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

As the name suggest, the weather warning α_t is a binary variable which can be used for decision-making purposes by the responsible authorities.

Finally, we would like to point out that weather information are usually collected from several meteorological stations located in the area of study. In this case, the weather feature vectors $\mathbf{y}_t$ as well as the indicators of the occurrence of hotspot of a certain location can be obtained by interpolating the data collected from the nearest meteorological stations. Furthermore, the methodology described above can be applied on the interpolated data to yield the wildfire danger index at a certain location of interest. We shall illustrate this remark in the Experiments and Discussion Section.

Algorithm 2: Wildfire Danger Training Algorithm

Input: Training set $\mathcal{T}_T = \{(\mathbf{y}_t, F_t) : t = 1, \ldots, T\}$ and threshold τ.
Output: Cluster centers $\mathbf{c}_1, \ldots, \mathbf{c}_C$, spread vector $\sigma = (\sigma_1, \ldots, \sigma_M)$, and danger factors $\phi_1, \ldots, \phi_C$.

1. Apply a fuzzy clustering algorithm, such as the subtractive clustering method, to obtain the centers $\mathbf{c}_1, \ldots, \mathbf{c}_C$ and spread vector σ from $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$;
2. Determine the sets $B_1, \ldots, B_C$ using (6), where the membership functions are evaluated using (4);
3. Compute $f_1, \ldots, f_C$ using (5);
4. Determine the wildfire danger factors $\phi_1, \ldots, \phi_C$ by means of (7);

Algorithm 3: Wildfire Danger Evaluation Algorithm

Input: The current weather feature vector $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Mt})$, past wildfire danger π_{t-1}, parameter $\beta \in (0, 1)$, cluster centers $\mathbf{c}_1, \ldots, \mathbf{c}_C$, spread vector $\sigma = (\sigma_1, \ldots, \sigma_M)$, and danger factors $\phi_1, \ldots, \phi_C$.
Output: The current wildfire danger.

1. Determine instantaneous wildfire danger $\tilde{\pi}(\mathbf{y}_t)$ using (8), where the membership functions are evaluated using (4);
2. Define the current wildfire danger index using (9);

2.3. Temporal evolution of wildfire warning

In the previous subsection, we presented the wildfire risk and danger models. On the one hand, the wildfire risk depends only on georeferenced information which either do not change or have low variation over time. On the other hand, the wildfire danger depends on weather and, thus, it varies significantly on time. It turns out that a useful early warning wildfire system must take into account both dynamic weather and static spatial information. The *wildfire warning*, denoted by $\mathbb{W}$, combines both georeferenced and weather information. In the following, we describe how the wildfire risk ρ and wildfire danger π_t are appropriately combined to yield the wildfire warning $\mathbb{W}$ at location with features $\mathbf{x} \in \Omega$ and at time $t \in \{1, \ldots, T\}$.

First of all, recall that the wildfire risk $\rho(\mathbf{x})$ represents the degree of membership of $\mathbf{x} \in \Omega$ in the fuzzy set of all hotspots. In a similar fashion, the wildfire danger $\pi(\mathbf{y}_t)$ can be also interpreted as a membership degree of time $\mathbf{y}_t$ into the fuzzy set of hotspots. Therefore, we can combine these two memberships values using an "AND" connective from fuzzy logic such as a triangular norm or t-norm for short [39]. In his paper, we use the Łukasiewicz triangular norm denoted by t_L. The Łukasiewicz t-norm was chosen, in parts, because a low wildfire risk yields null wildfire warning during the temporal simulation. For more details on t-norms the reader can refer to Appendix A.1.

In mathematical terms, given the spatial and the current weather feature vectors $\mathbf{x}$ and $\mathbf{y}_t$, the wildfire warning $\mathbb{W}(\mathbf{x}, \mathbf{y}_t)$ combines the wildfire risk $\rho(\mathbf{x})$ and the wildfire danger $\pi(\mathbf{y}_t)$ as follows using the Łukasiewicz t-norm t_L:

$$\mathbb{W}(\mathbf{x}, \mathbf{y}_t) = t_L(\rho(\mathbf{x}), \pi(\mathbf{y}_t)) \equiv \max\{0, \rho(\mathbf{x}) + \pi(\mathbf{y}_t) - 1\}. \tag{11}$$

From the increasing property of t-norms, we conclude that if $\rho(\mathbf{x}) \geqslant \rho(\mathbf{x}')$ and $\pi(\mathbf{y}_t) \geqslant \pi(\mathbf{y}'_t)$ then $\mathbb{W}(\mathbf{x}, \mathbf{y}_t) \geqslant \mathbb{W}(\mathbf{x}', \mathbf{y}'_t)$. In other words, the greater the wildfire risk or the wildfire danger, the higher the wildfire warning value. Furthermore, the wildfire warning attains its highest value in regions such that the wildfire risk ρ is maximum and $\pi(\mathbf{y}_t)$ is also maximum, that is, the weather features favors a wildfire. This remark is particularly important for the wildfire modeling because fires occur when both spatial and weather features are favorable. Indeed, according to the specialized literature [5], the most affected areas are those that have specific spatial characteristics and a history of occurrences of fires.

Fig. 4. Localization of State of Acre in South America and Brazil [41].

3. Experiments and discussion

Let us now validate the wildfire risk, danger, and warning models by considering data from the State of Acre, Brazil, an important state localized in the Brazilian Amazon. Precisely, with 164, 123, 712 km^2 and 733,559 inhabitants, the State of Acre is situated in the northwestern region of Brazil. It has 22 municipalities and the two largest ones are Rio Branco and Cruzeiro do Sul. Rio Branco, the capital of Acre, has 336,038 inhabitants and it is located in the eastern part of the State. The second largest city, Cruzeiro do Sul, has 78,507 inhabitants and it resides in the extreme western area of the State. Fig. 4 situates the State of Acre in the South America, including Brazil, and the neighboring countries Peru and Bolivia [40].

The occurrence of fires in several states of the Amazon Region, including Acre, has been the object of attention of Brazilian media in recent years [42]. We chose to study the State of Acre because it is localized in a part of the Amazon region that has strategic importance for Brazil. In addition, there is an increase in the amount of hotspots in this state. In recent years, the number of hotspots is reaching values close to those registered in the year 2005, in which an extreme drought and forest fires occurred [29,43]. Fig. 5 shows the amount of hotspots detected by the reference satellite NOAA-15 in the State of Acre and Brazil from 1999 to 2017, respectively. Note, from Fig. 5, that the number of hotspots between the years 2011 to 2017 increased by approximately 31% throughout Brazil. In the State of Acre, the number of hotspots increased more than 229%. A reliable wildfire warning system is essential for planning actions to prevent and control forest fires for the years to come.

3.1. Wildfire risk applied for the state of Acre

Let us illustrate the wildfire risk index described in Section 2.1 by considering the State of Acre, Brazil, as a case of study. Precisely, using georeferenced data provided by the *Ecological Economic Zoning of the State of Acre* and the *Technology Foundation of the State of Acre* (FUNTAC, in Portuguese), we shall construct a wildfire risk map of the State of Acre [1].

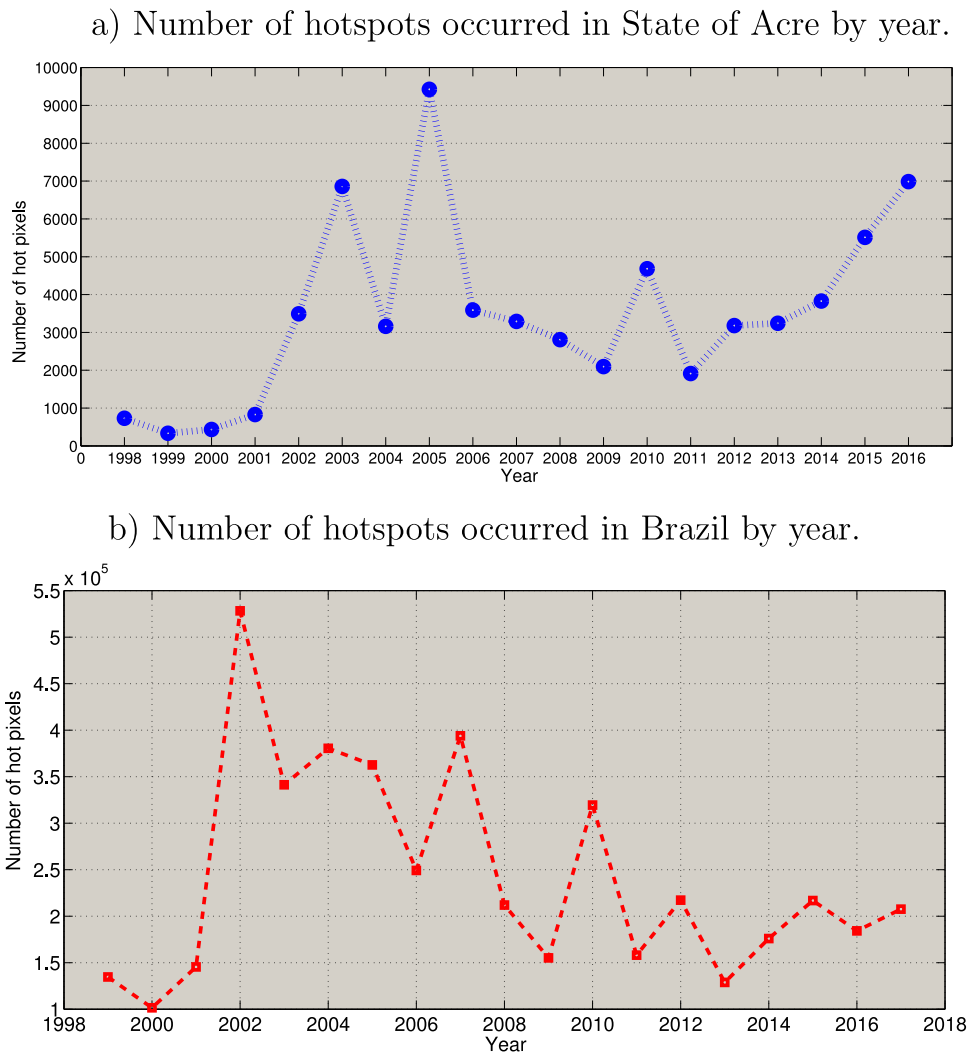


Fig. 5. Evolution of number of hotspots occurred in State of Acre and Brazil by year [29].

3.1.1. Data acquisition and pre-processing

First of all, we crossed the georeference maps with a rectangular mesh with 307,302 cells covering the State of Acre as well as with maps of hotspots generated by the National Institute of Spatial Research (INPE, in Portuguese) [29]. As a consequence, we obtained a database with information about the hotspots in the period from 2010 to 2018. We used the following four variables as spacial features: altitude, forest typology, distance to the closest watercourse, and the distance to the closest road. We would like to point out that, although we have not used the feature "population" explicitly, it is implicitly referenced in the variable forest typology, since there is a characteristic of this variable called anthropic area. The population density is also strongly related to the distance to the closest road.

The occurrence of hotspot from the eastern and western regions of the State of Acre are significantly different. Thus, we shall divide our data into two different scenarios: The west and east of the state of Acre, denoted respectively using the sub-indices "w" and "e".

For each scenario, the feature vectors have been normalized to have zero mean and variance one. Furthermore, because the number of samples from class ω_0 (absence of hotspot) is much larger than the number of data from class ω_1 (presence of hotspot), we have not used all the data but a subset with an equal number of samples from both classes. In order to avoid distortions, the subset have been chosen randomly.

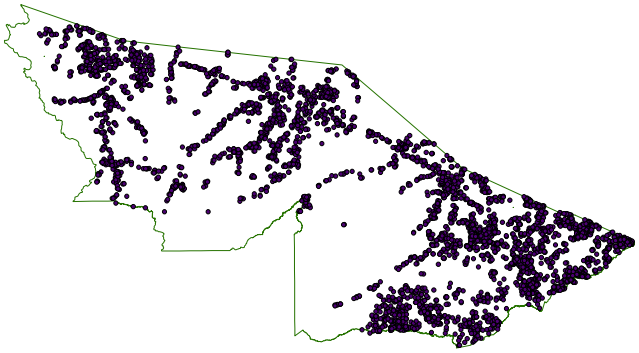
For each scenario, the balanced data has been divided into a training set $\mathcal{T}_L$ and a test set $\mathcal{A}_L$. Recall that the training set is used to synthesize the wildfire risk map ρ, while the test set $\mathcal{A}_L$ is used to evaluate the proposed methodology. In our experiments, the data from 2011 to 2018, which corresponds to approximately 85%, have been used for training. The data from 2010, which corresponds to the remaining 15%, formed the test set $\mathcal{A}_L$.

3.1.2. Results and discussion

As suggested in Section 2.1, the number k of the nearest neighbors has been determined using the cross-validation technique with ten folders on the training set. Specifically, for each $k \in \{1, 3, 5, \ldots, 47, 49\}$, we computed the average *Hamming loss* among the 10 validation sets. For the western region of the state of Acre, the lowest average Hamming loss was 16.03%, attained by considering $k_w = 29$. For the east, the number $k_e = 21$ of nearest neighbors provided the lowest average Hamming loss rate of 11.42%.

The risk maps ρ_w and ρ_e for the western and eastern regions of the State of Acre, obtained using $k_w = 21$ and $k_e = 29$, yielded a Hamming loss on the whole test data ($\mathcal{A}_w \cup \mathcal{A}_e$) of 16.41%. Moreover, the accuracy rate of the fuzzy kNN was 83.59%. These values suggest that it is possible to predict the occurrence (or not) of a hotspot for a certain period of time by considering the spatial information of any point in the domain.

a) Hotspot occurrence Map for State of Acre in 2010 [29].



b) Wildfire risk map obtained for State of Acre.

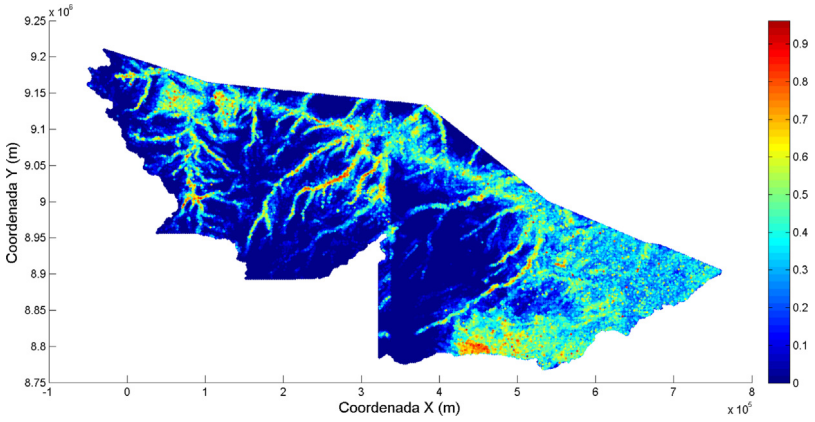


Fig. 6. Hotspot occurrence map and wildfire risk map for the State of Acre in 2010.

Fig. 6 shows the map of the State of Acre with all the hotspots that occurred in the year 2010. This figure also shows the wildfire risk ρ obtained from the data from 2011 to 2018. Note, from Fig. 6, that the wildfire risk map accompanies the occurrence of hotspot, which can be found predominantly in the urban areas of the state (east side of the state and region around Cruzeiro do Sul) as well as in a few rural areas in the middle of the state.

In order to quantitatively evaluate the wildfire risk map, we computed a measure of similarity between the predicted wildfire risk and the presence or absence of hotspot in a testing (unused for training) data. We considered three different measures of similarity, namely, complement of the relative distance of Hamming (S_h) – that is an example of a strong measure of similarity – S_1 and S_2 that are part of a class of measures of similarities studied by [44]. Further information on the measure of similarity and the formulas for S_h, S_1, S_2 can be found in Appendix A.1. The similarities $S_h = 0.8897$, $S_1 = 0.9043$, and $S_2 = 0.9022$, which characterizes a fairly high similarity between real hotspot and predicted wildfire risk, have been obtained. The high similarities between the wildfire risk map and true hotspots inform us that the cells associated to places where the phenomenon of hotspots occurred have a high risk of fire.

The fact that the *k*NN classifier has learned to identify hotspots using data from 2011 to 2018 and performed quite well in 2010 supports the hypothesis that hot spots are always happening in regions with similar characteristics in the State of Acre. Another interesting feature of the risk map generated by this approach is that it yielded intermediate risk values close to large rivers and it produced large values in locations with human presence in the interior of the state. These two remarks agree with the specific literature [3,5,9] that claims the human presence responsible for much of the occurrence of wildfire.

Table 1
Classifiers performance on validation data.

Classifier	Mean error
Decision Tree	0.2141
Naive Bayes	0.2291
Discriminant Analysis	0.2258
k-Nearest Neighbors	0.1449

We would like to conclude this subsection by pointing out that we evaluated the performance of the risk map obtained using classifiers different from the fuzzy *k* nearest neighbors, namely decision tree, naive Bayes, and discriminant analysis. Table 1 summarizes the mean error produced by these three classifiers and the fuzzy *k*-NN on a validation set obtained by randomly picking 15% of the training set without replacement. This table justifies our choice of the fuzzy *k*-NN classifier.

3.2. Wildfire danger for the State of Acre

Let us now apply the wildfire danger model presented in Section 2.2 to estimate the possibility of a wildfire to spread using weather information collected from the six weather stations located at the State of Acre.

3.2.1. Data acquisition and pre-processing

First of all, we considered the relative air humidity and rainfall as weather information. Furthermore, the rainfall data suffers

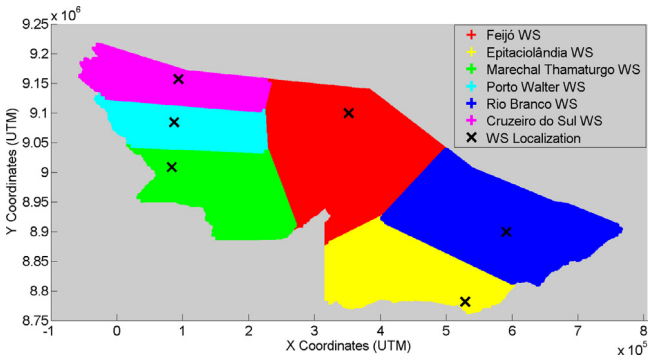


Fig. 7. Localization of weather stations and their region of influence in State of Acre.

from rapid variations which may degrade significantly the performance of a machine learning models. Therefore, we smoothed the rainfall data using a moving average. We would like to point out that we have not considered the temperature as a weather feature because we have not obtained statistically significant improvement on the preliminary experiments using the average daily temperature. We believe that this is a characteristic of the equatorial weather at the Amazon Rain Forest which hardly presents a significant thermal variation. This important issue, however, requires further investigation and it must be addressed if the model is applied for other regions.

In our experiments, we used the relative air humidity and rainfall collected from 2010 to 2018 at the weather stations (WS) localized at Rio Branco, Cruzeiro do Sul, Feijó, Epitaciolândia, Marechal Thaumaturgo, and Porto Walter, and provided by National Institute of Meteorology [45]. Fig. 7 shows the location of the weather stations (marked with "×") as well as their region of influence in the state of Acre. The region of influence of a certain weather station corresponds to its Voronoi cell. For example, the region of influence of the Rio Branco WS corresponds to all points closer to Rio Branco WS than any other weather station. In our experiments, we synthesized a wildfire danger for each region of influence using the weather data collected from the corresponding weather station. Also, using data provided by the National Institute of Spatial Research [29], we counted the number of hotspots in the region of influence of a certain weather station from 2010 to 2018.

Unfortunately, the data collected by the weather stations from 2010 to 2018 present serious problems of missing data. In fact, apart from the lack of eventual measurements, there are long gaps without measurements of relative air humidity and rainfall from certain weather stations. Fig. 8 summarizes the number of data missing at all meteorological stations from 2010 to 2018. Note that, although the largest amount of data have been collected in 2013, there is missing data only from two weather station in 2010, namely Porto Walter WS and Marechal Thaumaturgo WS.

Such as in the previous subsection, we divided the data into a training set and a testing set. Hopefully, the training of the wildfire danger model do not depend on the temporal evolution of the weather features. Therefore, we can simply discard the long periods of missing data during the training phase. The temporal evolution of the wildfire danger index is only established in the testing phase by means of an exponentially weighted moving average in (9). In view of this remark, we opted to use data from 2011 to 2018 for training while the data from the year of 2010 has been used for testing. Moreover, to predict the wildfire danger at a region determined by the Voronoi diagram of the meteorological stations, the weather data (humidity and precipitation) have been spatially interpolated using the inverse distance weighting (IDW) technique [46].

Table 2
Information about the fuzzy sets obtained using the *subtractive clustering* for all six weather stations from 2011 to 2018.

WS city	Number of cluster centers	dispersions (σ_U, σ_P)	r_{max}
Feijó	4	(0.5194,0.3147)	0.1
Epitaciolândia	2	(0.2683, 0.1786)	0.4
Marechal Thaumaturgo	2	(0.3246, 0.2243)	0.2
Porto Walter	4	(0.5114, 0.2591)	0.1
Rio Branco	4	(0.4768, 0.3610)	0.1
Cruzeiro do Sul	4	(0.4527, 0.3728)	0.1

3.2.2. Results and discussion

Applying the subtractive clustering method on the training data from each weather station, we obtained the information summarized in Tables 2 and 3. Precisely, the values of the radius $r_{\max}$ used in the subtractive clustering as well as the spread of the relative air humidity and rainfall precipitation can be found in Table 2. At this point, we would like to point out that the radius $r_{\max}$ have been determined from an exhaustive search applied on a 10-fold cross validation. The coordinates U_i and P_i of the centers of the clusters are shown in Table 3. This table also shows the value f_i and the danger factor ϕ_i corresponding to each fuzzy cluster. Note that, in agreement with specific literature [3,5], the fuzzy sets whose centers have low values of air humidity and rainfall values are associated with larger wildfire danger factors.

Due to the low computational cost and the absence of complementary information from experts or from the literature, the threshold has been determined by an exhaustive search on $\tau \in \{0.1, 0.2, \ldots, 0.9, 1\}$. Precisely, by using 10-fold cross-validation on the training set, we obtained the threshold $L = 0.4$ that maximizes the average performance for all weather stations. Similarly, the number β in (9) have been also determined by an exhaustive search $\beta \in \{0.1, 0.2, \ldots, 0.9\}$. The optimal value $\beta = 0.9$ adopted for all weather stations was obtained by applying 10-fold cross-validation on the training set.

Fig. 9 shows the wildfire danger index and the percentage of hotspots in the regions covered by the six weather stations in the year 2010. For comparison purposes, this figure also shows a rescaled version of the index provided by the Monte Alegre formula. We would like to recall that, in the State of Acre, the dry season spans from June to November. Accordingly, the occurrence of hotspots increases significantly in June. Note from Fig. 9 that the wildfire danger index increases accordingly in the same period. During the rainy period, which corresponds to almost the first half of the year, the wildfire danger is predominantly zero which means there is a low chance of a wildfire occurrence. Also, from Fig. 9, the wildfire danger model visually outperformed the Monte Alegre formula.

In order to quantitatively evaluate the wildfire danger estimate, we compared the weather warning α_t given by (10) with the occurrence F_t of hotspot in the same time instant t. At this point, recall that π_t measures the possibility while α_t acts as an alert for a wildfire start at time t. Table 4 presents the values of AUC – area under ROC curve – for both wildfire danger model and Monte Alegre formula using the test set from the six weather stations at the State of Acre. The interested reader is invited to consult Appendix A.2 for more details of the hit rate and the ROC curve. Concluding, the wildfire danger model provided a more reliable wildfire index than the Monte Alegre formula for the regions covered by the six weather stations.

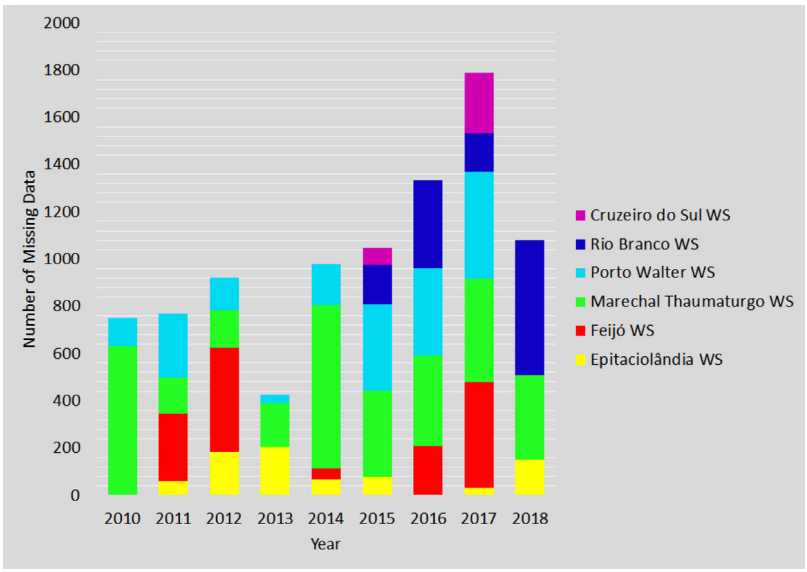
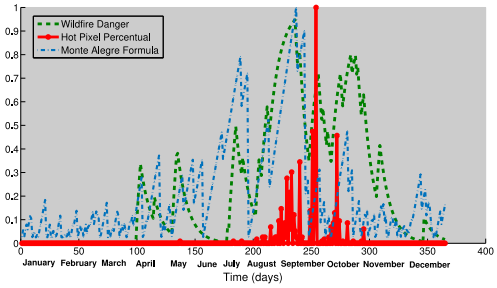
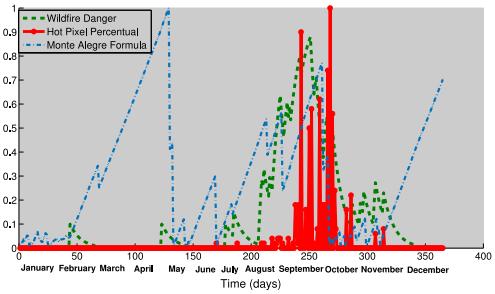


Fig. 8. Number of missing data in all weather stations per year.

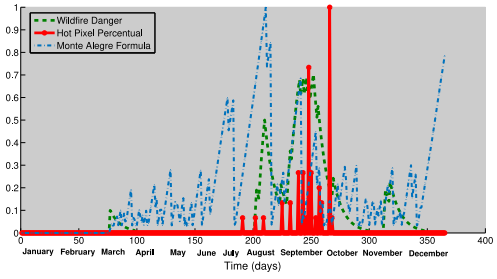
a) Feijó WS Data.



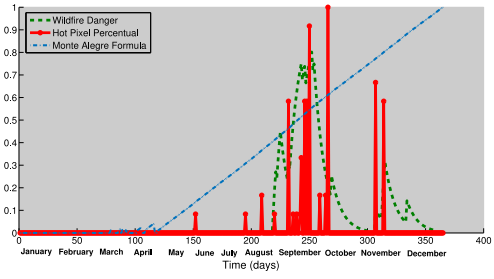
b) Epitaciolândia WS Data.



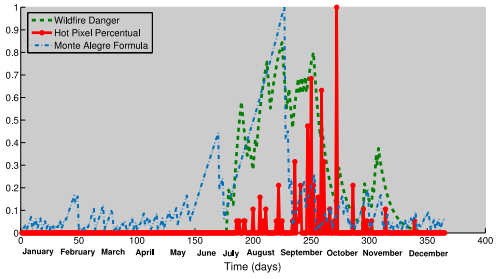
c) Marechal Thaumaturgo WS Data.



d) Porto Walter WS Data.



e) Rio Branco WS Data.



f) Cruzeiro do Sul WS Data.

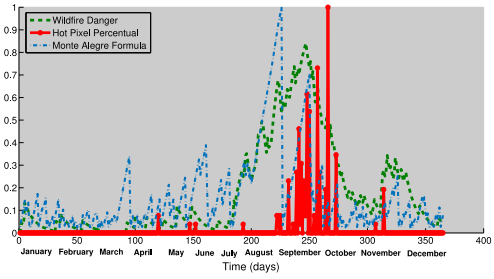


Fig. 9. Wildfire danger and Monte Alegre Formula for six weather stations in 2010.

Table 3
Information about the fuzzy sets obtained using the *subtractive clustering* for six weather stations data from 2011 to 2018.

WS city	Cluster	Humidity (%)	Rainfall (mm)	f_i	ϕ_i
Feijó	1	86.42	2.18	0.003	0
	2	81.21	0.52	0.39	0
	3	90.33	3.20	0.001	0
	4	76.5	0	1	1
Epitaciolândia	1	82	2.38	0.004	0
	2	66	0	1	1
Marechal Thaumaturgo	1	83.70	1.62	0.05	0
	2	72.88	0.18	1	1
Porto Walter	1	84.54	1.84	0.14	0
	2	79.21	0.56	0.37	0
	3	89.25	4.10	0	0
	4	74.21	0	1	1
Rio Branco	1	85.75	1.31	0.26	0
	2	79.75	0.62	0.36	0
	3	91.17	2.42	0.004	0
	4	74	0.18	1	1
Cruzeiro do Sul	1	85.25	0.72	0.09	0
	2	78.92	0.24	0.45	0.45
	3	91	1.23	0.22	0
	4	72.79	0	1	1

Table 4
AUC values for both models considering weather station data in the six cities.

Weather station city	AUC (Danger model)	AUC (Monte Alegre Formula)
Feijó	0.91	0.76
Epitaciolândia	0.92	0.62
Marechal Thaumaturgo	0.89	0.73
Porto Walter	0.83	0.67
Rio Branco	0.87	0.69
Cruzeiro do Sul	0.87	0.82

Table 5
Similarity measures between the warning index and the occurrence of hotspot evaluated in 2010.

Measure	S_1	S_2	S_3	S_4	S_h
08 August	0.97	0.97	0.97	0.96	0.96
23 August	0.90	0.90	0.90	0.88	0.88
28 August	0.89	0.89	0.89	0.88	0.88
06 September	0.98	0.98	0.98	0.97	0.97
08 September	0.98	0.98	0.99	0.97	0.97
22 September	0.98	0.98	0.98	0.97	0.97

3.3. Wildfire warning evolution: Study for the State of Acre, Brazil

Let us now construct an evolving wildfire warning map for the state of Acre. Briefly, the wildfire warning map is obtained by combining the wildfire risk map presented in Section 3.1 and the wildfire danger index determined in Section 3.2. Recall that the wildfire risk map and the wildfire danger index have been determined using data from 2011 to 2018. Furthermore, both wildfire risk map and wildfire danger index have been tested on data from 2010. Also, the wildfire danger index have been determined using the weather information of six weather stations at the State of Acre. By partitioning the domain into the six Voronoi regions shown in Fig. 7, the wildfire danger indexes have been extrapolated for the entire State of Acre.

3.4. Results and discussion

Using the Łukasiewicz t-norm, we obtained a sequence of wildfire warning maps which reveals the evolution of wildfire warning during the year of 2010. In general, the wildfire warning is null (zero) for most of the year. The wildfire warning begins to increase only in the driest period which, for the Amazon region, spans from the months of June to November [5]. This behavior agrees with the actual phenomena of occurrence of hotspots for the studied region. Figs. 10 and 11 show six frames (days) of the wildfire warning and hotspots map sequence covering the dry season. Quantitatively, Table 5 provides the numerical values yielded by five similarity measures (see Appendix A.1 for details). The similarity measures compare the true occurrence of hotspots and wildfire warning maps in six days of 2010. More generally, Table 6 provides the average and minimum values of the six similarity measures over the whole year of 2010. Note that the five similarity measures yielded very similar values. Finally, Fig. 12 shows the Hamming similarity measure between the true occurrence of hotspots and the wildfire warning map by the days of the 2010. These results reflect a high similarity between the wildfire warning model and the phenomenon studied.

Concluding, we believe that the wildfire warning model can be useful for the authorities responsible for fighting forest fires. In particular, the wildfire warning attains higher values in places more prone to fire and only in the period of the year in which there are occurrences of hotspot.

4. Concluding remarks

In this paper we presented a methodology for constructing a dynamic wildfire warning map of a certain region of interest. The dynamic wildfire warning map is obtained by combining two indices, the wildfire risk and the wildfire danger.

The wildfire risk index, which depends on georeferenced information such as altitude, forest typology, distance to watercourse or road, measures the possibility of a fire to ignite at a certain point. The wildfire risk is derived from historical data using a fuzzy *k*NN classifier.

The wildfire danger takes into account the weather and weather conditions to measure the possibility of the occurrence of hotspot at a certain time instant t. Like the wildfire risk index, the wildfire danger is determined from historical data using machine learning techniques. Precisely, the wildfire danger is modeled by a fuzzy system that uses information provided by a fuzzy clustering technique such as the *subtractive clustering* method.

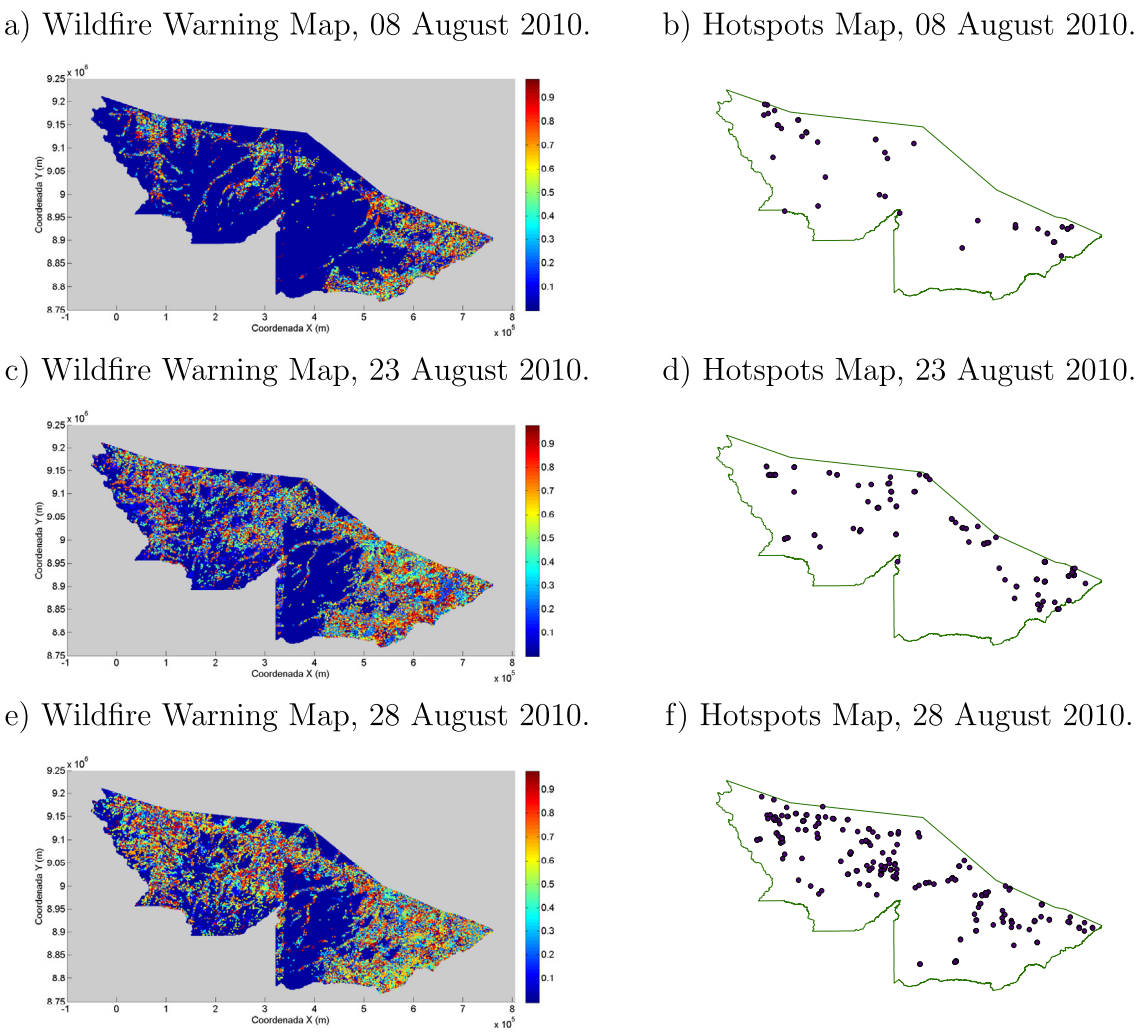


Fig. 10. Wildfire warning map and hotspot occurrence map for three days in August of 2010.

Table 6
Average and minimum values of the similarity measures between the warning index and the occurrence of hotspot during the year of 2010.

	S_1	S_2	S_3	S_4	S_h
Average	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02
Minimum	0.8874	0.8874	0.8874	0.8868	0.8873

The wildfire warning couples the wildfire risk (spatial information) and wildfire danger (temporal information) to yield a dynamic map of a certain region. Broadly speaking, the wildfire warning is defined as the degree of truth of the propositions "the spatial AND the weather information allows for a fire to ignite at a location **x** at time t". Formally, it is given by the evaluation of a triangular norm at the wildfire risk and the wildfire danger indices. We believe that the dynamic wildfire warning map provides an early warning of forest fire that can be used by local authorities and policymakers for decision making purposes.

We validated our approach by considering the state of Acre, Brazil, which is located in the Brazilian Amazon and has been the object of attention in recent years. Precisely, we used historical data from 2011 to 2018 to synthesize the wildfire risk map of the state of Acre, Brazil. The similarity between the obtained wildfire risk map and the real map indicating the occurrence of hotspot in the year 2010 was at least 83%. Historical data from 2011 to 2018 of six weather stations of Acre has been used to determine the fuzzy system for the wildfire danger index, which outperformed the formula of Monte Alegre in our experiments. Finally, the wildfire danger combined with the wildfire risk model, yielded a wildfire warning system that agrees with the occurrence of hotspot in the year 2010 with similarity measures greater than 0.89. Some frames of the dynamic wildfire warning map for the state of Acre have been given in Section 3.3 for illustrative purposes.

Concluding, one important advantage of the proposed approach is the use of machine learning techniques. Since the dynamic wildfire warning map is determined using historical data, it can be gradually improved by incorporating new data. Moreover, because it does not require any *a priori* knowledge, the wildfire warning model presented in this paper can be easily adapted for other regions. On the downside, wildfire risk and danger models presented in this paper are data dependent. In particular, due to the uncertainty inherent to the relation between the number of hotspot and the real occurrence of wildfire, the main obstacle of our approach is the selection of the variables that describe the occurrence of the phenomenon.

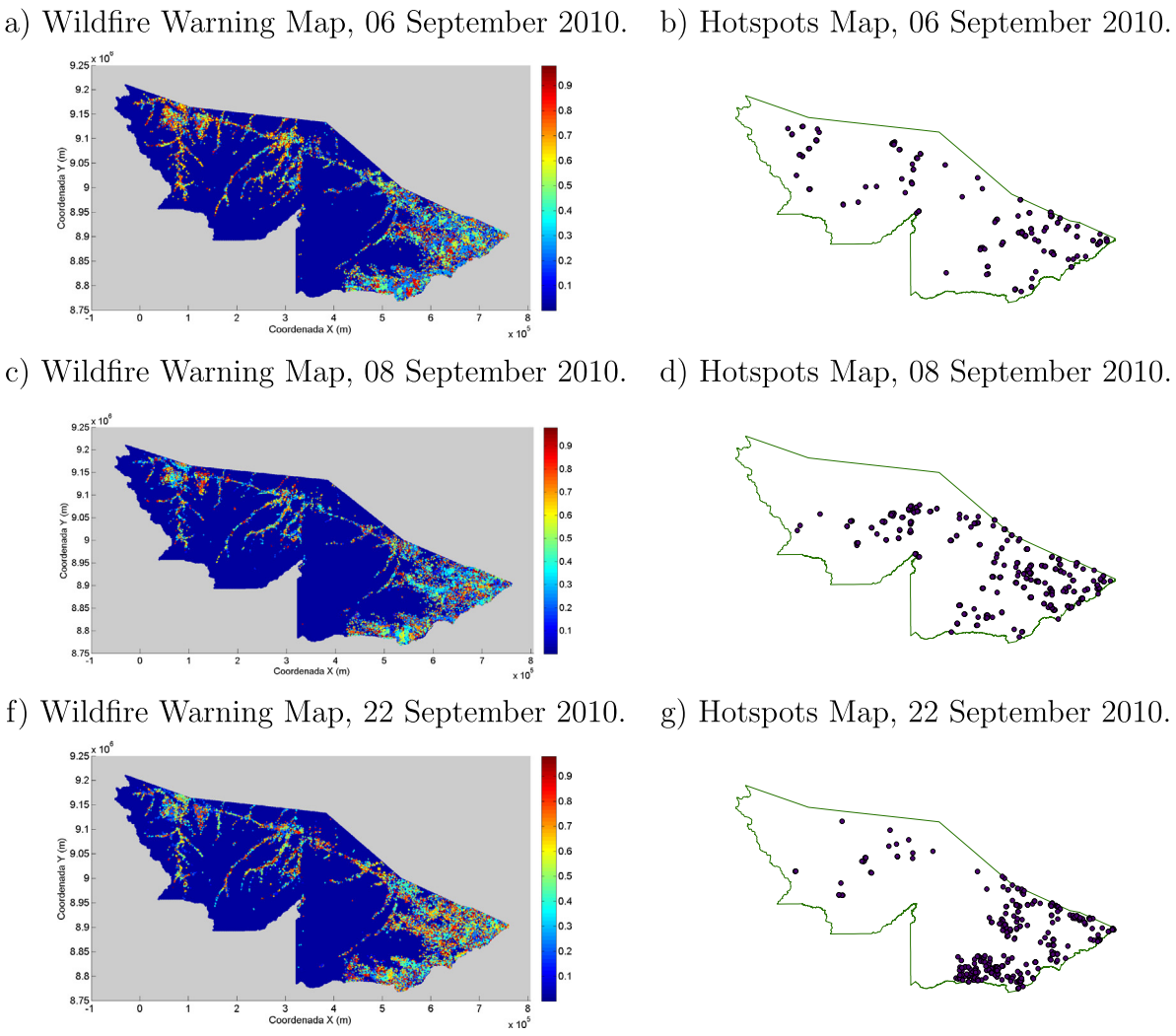


Fig. 11. Wildfire warning map and hotspots occurrence map for three days in September of 2010.

In the future, we plan to incorporate in the proposed models further variables such as air temperature and data obtained from satellite images like the normalized difference vegetation index, road and population density, and elevation. Another improvement can be achieved with the use of burn scars record. A more systematic selection of variables from input variables should also be performed. We would like to point out that we also need to evaluate the performance of other fuzzy grouping and classification methods in both wildfire risk and danger models. We should also consider using weather forecasting for future wildfire danger and warning estimation.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106075.

CRediT authorship contribution statement

I.D.B. Silva: Investigation, Data curation, Methodology, Software, Writing - original draft. **M.E. Valle:** Conceptualization, Methodology, Writing - review & editing, Software, Supervision. **L.C. Barros:** Conceptualization, Supervision, Methodology. **J.F.C.A. Meyer:** Conceptualization, Supervision, Methodology.

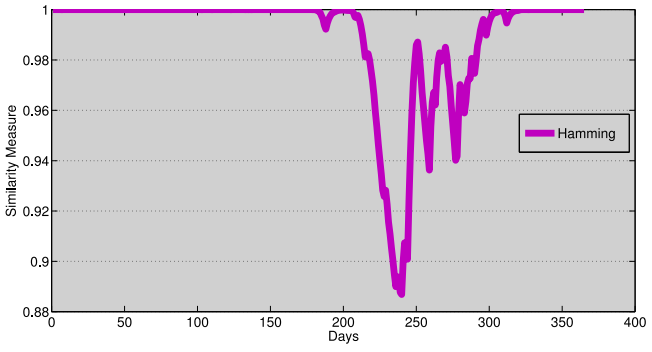


Fig. 12. Hamming similarity measure between the true occurrence of hotspot and the wildfire warning map for each day of 2010.

Acknowledgments

We thank the Federal University of Acre, Brazil for funding the grammar review of this paper and CAPES, Brazil for the doctoral scholarship granted to the first author. This work was supported by CNPq, Brazil under grants 310118/2017-4 and 305862/2013-8.

Appendix. Mathematical definitions and concepts

A.1. Fuzzy subset and similarity measure

A fuzzy set A on a universe of discourse U is characterized by its membership function $A : U \to [0, 1]$, where $A(u)$ denotes the degree of membership of the element $u \in U$ into the fuzzy set A. We denote by $\mathcal{F}(U)$ the family of all fuzzy sets of U. We say that a fuzzy set $A \in \mathcal{F}(U)$ is a subset of another fuzzy set $B \in \mathcal{F}(U)$, and write $A \subseteq B$, if $A(u) \leqslant B(u)$, $\forall u \in U$. The standard complement of a fuzzy set $A \in \mathcal{F}(A)$, denoted by $\bar{A}$, is the fuzzy set defined by $\bar{A}(u) = 1 - A(u)$, $\forall u \in U$ [47–50].

A fuzzy set E of the Cartesian product $U_1 \times U_2 \ldots \times U_n$ is called a fuzzy relation on $U_1 \times U_2 \times \cdots \times U_n$. The value $E(u_1, u_2, \ldots, u_n)$ represents the degree to which the elements $u_1 \in U_1$, $u_2 \in U_2$, $\ldots$, $u_n \in U_n$ are related. We speak of a binary fuzz relation if the Cartesian product is formed of only two sets $U_1 \times U_2$, that is, a binary relation is a fuzzy set $E : U_1 \times U_2 \to [0, 1]$. A binary relation $E : U \times U \to [0, 1]$ is symmetric if $E(x, y) = E(y, x)$ for all $(x, y) \in U \times U$. Also, we say that $E : U \times U \to [0, 1]$ is locally reflexive if $E(x, x) \geqslant E(x, y)$ for all $x, y \in U$.

A fuzzy similarity measure, also known as fuzzy equivalence or equality index, measures the degree to which fuzzy sets are equal. Formally, a similarity measure is a symmetric and locally reflective binary fuzzy relation [51]. Examples of fuzzy similarity measure are given below for fuzzy sets $A, B \in \mathcal{F}(U)$ on a finite universe of discourse $U = \{u_1, \ldots, u_n\}$, where $a_i = A(u_i)$ and $b_i = B(u_i)$ for all $i = 1, \ldots, n$:

$$S_h(A, B) = 1 - \frac{1}{n}\sum_{i=1} |a_i - b_i|, \tag{A.1}$$

$$S_1(A, B) = \frac{1}{n}\left[n + \sum_{i=1}^{n}(a_i \cdot b_i) - \max\left\{\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i\right\}\right] \tag{A.2}$$

$$S_2(A, B) = \frac{\sum_{i=1}^{n}[a_i \cdot b_i + (1 - a_i) \cdot (1 - b_i)]}{n + \sum_{i=1}^{n}(a_i \cdot b_i) - \min\left\{\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i\right\}} \tag{A.3}$$

$$S_3(A, B) = \frac{1}{n}\left\{n + \sum_{i=1}^{n} \max\{a_i + b_i - 1, 0\} - max\left[\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i\right]\right\} \tag{A.4}$$

$$S_4(A, B) = \frac{1}{n}\left\{\sum_{i=1}^{n}[a_i \cdot b_i + (1 - a_i) \cdot (1 - b_i)]\right\} \tag{A.5}$$

Finally, the connective "AND" is described in fuzzy logic by a triangular norm or t-norm. An operator $t : [0, 1] \times [0, 1]$ is a t-norm if it satisfies the following conditions [48]:

1. $t(1, x) = x$,
2. $t(x, y) = t(y, x)$,
3. $t(x, t(y, z)) = t(t(x, y), z)$,
4. if $x \leqslant u$ and $y \leqslant v$ then $t(x, y) \leqslant t(u, v)$.

The Łukasiewicz t-norm, defined by

$$t_L(x, y) = \max\{0, x + y - 1\}, \quad \forall x, y \in [0, 1], \tag{A.6}$$

is an example of t-norm.

A.2. The ROC curve

The *receiver operating characteristic* (ROC) analysis originated from the development of statistical decision theory between 1950 and 1960 in psychology and in the evaluation of radar signal detection. This analysis is a standard methodology to evaluate the performance of a classification system and also used to evaluate medical diagnoses [52–54]. The development of methodologies for the statistical analysis of the ROC data was driven by various applications of the ROC analysis to medicine [55,56]. Furthermore, the ROC curves are commonly used in machine learning to present the results of binary decision problems [57].

Table A.7
Confusion matrix for a binary classifier.

	prediction ($y = 0$)	prediction ($y = 1$)
real($y = 0$)	TN	FP
real($y = 1$)	FN	TP

By making an analogy to the diagnosis of a disease, suppose that a cut-off value is set for a certain diagnostic test in a study that determines the classification of the individuals as sick or healthy [55,58]. High cut-off values lead to a very sensitive and very specific test. In contrast, low cut-off values invert this behavior [55,58].

An effective way of evaluating the performance of a given classifier is by discriminating the hits (or errors) committed for each class, this can be done from a confusion matrix or contingency table illustrated in A.7. The elements of the main diagonal represent the correct decisions: number of true positives (TP) and true negatives (TN); while the elements outside this diagonal represent the errors made: number of false positives (FP) and false negatives (FN) [59].

From Table A.7, we can calculate some important metrics that evaluate the performance on the positive and negative classes:

$$FFP = \frac{FP}{TN + FP} \quad \text{(Fraction of false positives)}$$

$$FTP = \frac{TP}{TP + FN} \quad \text{(Fraction of true positives)}$$

$$FFN = \frac{FN}{TP + FN} \quad \text{(Fraction of false negatives)}$$

$$FTN = \frac{TN}{TN + FP}. \quad \text{(Fraction of true negatives)}$$

Generally speaking, the cut-off value is evaluated by plotting the fraction of false positives (FFP) by the fraction of true positives (FTP). Recall that the fraction of false positives FFP is the quotient of the number of sick individuals that have been misclassified as healthy by the number of cases. Similarly, the fraction of true positives refers to the ratio of the number of healthy individuals that have been correctly classified as healthy by the number of cases. The ROC curve is the graphical representation of the pairs of points (FFP, FTP) obtained by varying the cut-off value [58].

ROC curve can be used to find a threshold for a classifier which maximizes sensitivity at the highest acceptable false positive rate (or cost). However, different problems have distinct optimal classifier thresholds. It is possible to use ROC curves to assess the performance of the classifier over its entire operating range [52].

The most widely-used measure is the area under the curve (AUC). The AUC is 0.5 for a classifier with random guessing and the curve follows the diagonal and assumes the value 1.0 for the perfect classifier. Bradley (1997) investigated the use of the *area under the ROC curve* (AUC) as a performance measure for machine learning algorithms and concluded that it provides a single effective measure to evaluate the performance of a binary classifier [52,60].

A.2.1. Monte Alegre's model

Let us now address the Monte Alegre's model which have been used for comparison purpose in Section 3. The Monte Alegre's model yields an index, denoted by FMA, which depends on the past relative air humidity and the precipitation of the day. Precisely, the index FMA is determined recursively using Table A.8,

Table A.8
Monte Alegre's Model.

Precipitation in the nth day (mm)	FMA index
$\leqslant 2.4$	$\text{FMA}(n) = 100/H_n + \text{FMA}(n-1)$
2.5 to 4.9	$\text{FMA}(n) = 100/H_n + 0.7\text{FMA}(n-1)$
5.0 a 9.9	$\text{FMA}(n) = 100/H_n + 0.4\text{FMA}(n-1)$
10.0 a 12.9	$\text{FMA}(n) = 100/H_n + 0.2\text{FMA}(n-1)$
$\geqslant 12.9$	$\text{FMA}(n) = 0$

where n denotes the number of days of the observation period and H_i is the relative air humidity measured at 1 pm of the ith day, for $i = 1, \ldots, n$.

References

[1] G.E.A. Acre, Zoneamento Ecológico-Econômico do Estado do Acre, Fase II (Escala 1:250.000): Documento Síntese, SEMA, Ed. Rio Branco, 2010.

[2] G.E.A. Acre, Plano integrado de prevenção, controle e combate às queimadas e aos incêndios florestais do Estado do Acre, third ed., SEMA, Ed. Rio Branco, 2013.

[3] R.V. Soares, A.C. Batista, Incêndios Florestais controle, efeitos e uso do fogo, Edited by the authors, Curitiba-PR, Brasil, 2007.

[4] P. Fiorucci, F. Gaetani, R. Minciardi, Development and application of a system for dynamic wildfire risk assessment in Italy, Environ. Model. Softw. 23 (6) (2008) 690–702.

[5] D.C. Nepstad, A.G. Moreira, A.A. Alencar, Floresta em Chamas, Edited by the authors, Brasília-DF, Brasil, 1999.

[6] J.R.S. Nunes, R.V. Soares, A.C. Batista, Manual de prevenção e combate a incêndios florestais 2ª edição, Edited by the authors, Curitiba-PR, Brasil, 2008.

[7] J.R.S. Nunes, R.V. Soares, A.C. Batista, Ajuste da fórmula de Monte alegre alterada (fma+) para o estado do paraná, Floresta 37 (2007) 1–14.

[8] J.R.S. Nunes, R.V. Soares, A.C. Batista, FMA+ - um novo índice de perigo de incêndios florestais para o estado do paraná, Floresta 36 (1) (2006) 75–91.

[9] J.R.S. Nunes, R.V. Soares, A.C. Batista, Incêndios florestais no Brasil o Estado da Arte, Editado pelos autores, Curitiba, 2009.

[10] C.A. Alvares, I.R. Cegatta, L.A.A. Vieira, R.F. Pavani, E.M. de Mattos, P.C. Sentelhas, J.L. Stape, R.V. Soares, Forest fire danger: application of Monte alegre formula and assessment of the historic for piracicaba, SP, Sci. Florestalis 42 (2014) 521–532.

[11] F.C.F.D. Group, Development and Structure of the Canadian Forest Fire Behavior Prediction System, Tech. Rep. STX-3, Canadian Forestry Service, Forestry CA, Ottawa, Ontario, 1992.

[12] R.C. Rothermel, A Mathematical Model for Predicting Fire Spread in Wildland Fuels, Tech. Rep. 114, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 1972, p. 40.

[13] J.E. Deeming, R.E. Burgan, J.D. Cohen, Aids to Determining Fuel Models for Estimating Fire Behavior, Tech. Rep. 39, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 1977, p. 63.

[14] H. Anderson, Aids to Determining Fuel Models for Estimating Fire Behavior, Tech. Rep. 122, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 1982, p. 22.

[15] F. Sivrikaya, et al., Evaluation of forest fire risk with GIS, Pol. J. Environ. Stud. 23 (1) (2014) 187–194.

[16] Y.J. Goldarag, A. Mohammadzadeh, A.S. Ardakani, Fire risk assessment using neural network and logistic regression, J. Indian Soc. Remote Sens. 44 (6) (2016) 885–894.

[17] H.K. Preisler, D.R. Brilinger, R.E. Burgan, J.W. Benoit, Probability based models for estimation of wildfire risk, Int. J. Wildland Fire 13 (2) (2004) 133–142.

[18] M. Bisquert, J.M. Sánchez, V. Caselles, Modeling fire danger in Galicia and Asturias (Spain) from MODIS images, Remote Sens. 6 (1) (2014) 540–554.

[19] A.G. McArthur, A. Forestry, T. Bureau, Weather and Grassland Fire Behaviour, Canberra : Forestry and Timber Bureau, 1966, prepared for the Country Fire Authority and Victorian Rural Fire Brigades Association Group Officers Study Period, Geelong, 13th–15th August, 1965.

[20] P. Bolourchi, S. Uysal, Forest fire detection in wireless sensor network using fuzzy logic, in: 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks, 2013, pp. 83–87.

[21] S. Garcia-Jimenez, A. Jurio, M. Pagola, L.D. Miguel, E. Barrenechea, H. Bustince, Forest fire detection: A fuzzy system approach based on overlap indices, Appl. Soft Comput. 52 (2017) 834–842.

[22] A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, J. Canda, E. Jimenez, J.L. Legido, S.M. niz, C. Paz-Andrade, M.I. Paz-Andrade, A neural network approach for forestal fire risk estimation, in: Proceedings of the 15th European Conference on Artificial Intelligence, ECAI'02, IOS Press, Amsterdam, The Netherlands, 2002, pp. 643–647.

[23] H. Adab, K.D. Kanniah, K. Solaimani, Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques, Nat. Hazards 65 (3) (2013) 1723–1743.

[24] F. Justino, F. Stordal, A. Clement, E. Coppola, A. Setzer, D. Brumatti, Modelling weather and climate related fire risk in Africa, Am. J. Clim. Change 2 (4) (2013) 209–224.

[25] M. Matin, V. Chitale, M. S. R. Murthy, K. Uddin, B. Bajracharya, S. Pradhan, Understanding forest fire patterns and risk in Nepal using remote sensing, geographic information system and historical fire data, Int. J. Wildland Fire 26 (2017) 276.

[26] J. San-Miguel-Ayanz, J.D. Carlson, M. Alexander, K. Tolhurst, G. Morgan, R. Sneeuwjagt, M. Dudley, Current methods to assess fire danger potential, in: Wildland Fire Danger Estimation and Mapping, World Scientific, 2011, pp. 21–61.

[27] A.C. Batista, D.S. Oliveira, R.V. Soares, Zoneamento de Risco de Incêndios Florestais para o Estado do Paraná, FUPEF, Curitiba-PR, Brasil, 2002.

[28] A.C. Batista, A.F. Tetto, F. Deppe, L. Grodzki, J.T. Grassi, Análise dos impactos das mudanças climáticas sobre o risco de incêndios florestais no Estado do Paraná, Sci. Forestalis 42 (104) (2014) 491–501.

[29] D.P.I. INPE, Divisão de processamento de imagens, 2015, URL Disponível em: http://www.dpi.inpe.br/. (20 March 2015).

[30] P.P. Angelov, X. Zhou, Evolving fuzzy-rule-based classifiers from data streams, IEEE Trans. Fuzzy Syst. 16 (6) (2008) 1462–1475.

[31] J.M. Keller, M.R. Gray, J.A. Givens Jr., A fuzzy K-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. 15 (4) (1985) 580–585.

[32] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., John Wiley and Sons, New York, USA, 2001.

[33] S. Haykin, Neural Networks and Learning Machines, third ed., Prentice-Hall, Upper Saddle River, NJ, 2009.

[34] M. Flannigan, B. Wotton, Chapter 10- climate, weather, and area burned, in: E.A. Johnson, K. Miyanishi (Eds.), Forest Fires, Academic Press, San Diego, 2001, pp. 351–373.

[35] S.L. Chiu, in: D. Dubois, H. Prade, R. Yager (Eds.), Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification, John Wiley & Sons, New York, NY, USA, 1997.

[36] D. Paton, P.T. Buergelt, S. McCaffrey, F. Tedim, Wildfire Hazards, Risks, and Disasters, Elsevier, 2014.

[37] M.R. Barbosa, J.C.S. Seoane, M.G. Buratto, L.S. de O. Dias, J.P.C. Raivel, F.L. Martins, Forest fire alert system: a Geo Web GIS prioritization model considering land susceptibility and hotspots – a case study in the Carajás national forest, Brazilian amazon, Int. J. Geogr. Inf. Sci. 24 (6) (2010) 873–901.

[38] S.W. Roberts, Control chart tests based on geometric moving averages, Technometrics 1 (3) (1959) 239–250.

[39] W. Pedrycz, F. Gomide, Fuzzy Systems Engineering, John Wiley & Sons, Inc., New Jersey, 2007.

[40] BRASIL, Instituto Brasileiro de Geografia e Estatística, 2013, URL http://cod.ibge.gov.br/MOG. (26 May 2013).

[41] Google Inc, Google earth pro 7.1.8.3036 (32–bit), 2017, URL www.google.com.br/intl/pt-BR/earth/. (24 March 2017).

[42] E. Bertoni, Queimadas recordes avançam sem controle, caderno Cotidiano, Folha de São paulo, 2015, URL http://folha.com/no1699987. (15 January 2015).

[43] N.V. Pantoja, I.F. Brown, Estimativas de áreas afetadas pelo fogo no leste do Acre associadas à seca de 2005, in: Anais Do XIII Simpósio Brasileiro de Sensoriamento Remoto, SBSR, Natal, Brasil, 2009, pp. 6029–6036.

[44] H.N.B. De Baets, H. De Meyer, A class of rational cardinality-based similarity measures, J. Comput. Appl. Math. 132 (2001) 51–69.

[45] M.A.P.A. Brasil, Instituto nacional de meteorologia, 2011, disponível em: http://www.inmet.gov.br. Acesso em: (20 June 2015).

[46] X.A. C., K.C. W., S.B. R., Daily gridded meteorological variables in Brazil (1980–2013), Int. J. Climatol. 36 (6) (2016) 2644–2659.

[47] M.E. Valle, A.C. Souza, On the recall capability of recurrent exponential fuzzy associative memories based on similarity measures, Mathware Soft Comput. Mag. 22 (1) (2015) 33–39.

[48] L. Barros, R.C. Bassanezi, W.A. Lodwick, A First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics, Vol. 347, first ed., Springer-Verlag Berlin Heidelberg, 2017, p. 299.

[49] L. Barros, R. Bassanezi, P. Tonelli, Fuzzy modelling in population dynamics, Ecol. Modelling 128 (1) (2000) 27–33.

[50] M.S. Peixoto, L.C. de Barros, R.C. Bassanezi, Predator–prey fuzzy models, Ecol. Modelling 214 (1) (2008) 39–44.

[51] B. De Baets, H. De Meyer, Transitivity-preserving fuzzification schemes for cardinality-based similarity measures, European J. Oper. Res. 160 (3) (2005) 726–740, decision Analysis and Artificial Intelligence.

[52] S. Wang, et al., Optimizing area under the ROC curve using semi-supervised learning, Pattern Recognit. 48 (1) (2015) 276–287.
[53] J.A. Swets, ROC analysis applied to the evaluation of medical imaging techniques, Invest Radiol. 14 (2) (1979) 109–121.
[54] K. Hajian-Tilaki, Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation, Caspian J. Intern. Med. 4 (2) (2013) 627–635.
[55] M.J.P. Castanho, L.C. Barros, A. Yamakami, L.L. Vendite, Fuzzy receiver operating characteristic curve: An option to evaluate diagnostic tests, IEEE Trans. Inf. Technol. Biomed. 11 (2007) 244–250.
[56] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36.
[57] J. Davis, M. Goadrich, The relationship between precision-recall ans ROC curves, in: Proceedings of the International Conference on Machine Learning, Pittsburgh, PA, 2006, pp. 233–240.
[58] A.C.S. Braga, Curvas ROC: aspectos funcionais e aplicações (Tese de doutorado), Universidade do Minho, 2000.
[59] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874, ROC Analysis in Pattern Recognition.
[60] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159.