ABES

Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Hybrid Models Applied to Create a Classification Index of Fire Risk Levels in Brazil

Modelos Híbridos Aplicados à Construção de Índice de Classificação de Níveis de Risco de Fogo no Brasil

Pedro Antonio Galvão Junior[1], *Sandra Regina Monteiro Masalskiene Roveda*[1], *Henrique Ewbank de Miranda Vieira*[2]

ABSTRACT

Fire has always exerted a great attraction on humans. Fires generally provide social and environmental impacts at the places where they occur. Several Brazilian localities, especially in the driest months of the year, are more susceptible to this phenomenon. In this paper, an index able of classifying levels of fire risk in areas geographically located in Brazil. This paper presents an index capable of classifying fire risk levels elaborated from neuro-fuzzy systems. Data from the municipality of Sorocaba were used to test the proposed models. The results obtained by this index are promising, reaching values of mean absolute error below 3% when applied in the prediction of the risk of fire for the maximum period of up to 3 days. The proposed index can be used as a tool to support and assist various research agencies or institutes that need to identify the possibility of burning, corroborating the measures to reduce atmospheric emitters and meeting Goal 15 of Agenda 30 as defined by the UN in 2015, which aims to stimulate conservation actions and the recovery and sustainable use of ecosystems.

Keywords: fuzzy modeling; forecast model; machine learning; neuro-fuzzy model; artificial neural networks.

RESUMO

O fogo sempre exerceu grande atração sobre os seres humanos. As queimadas, de maneira geral, proporcionam impactos sociais e ambientais nos locais onde ocorrem. Diversas localidades brasileiras, especialmente nos meses mais secos do ano, estão mais suscetíveis a esse fenômeno. O estudo e o monitoramento do risco do fogo são uma poderosa ferramenta adotada no mapeamento e sensoriamento de áreas afetadas ao longo do território brasileiro e em outras partes do mundo. Este trabalho apresenta um índice para classificar os níveis de risco de fogo, elaborado com base nos sistemas *neuro-fuzzy*. Dados da cidade de Sorocaba foram utilizados para testar os modelos propostos. Os resultados obtidos mostram-se promissores, alcançando valores referentes à média de erros absolutos abaixo de 3%, aplicados na previsão do risco de queima pelo período máximo de até três dias. O índice proposto poderá ser utilizado como ferramenta de apoio e auxílio a diversos órgãos ou institutos de pesquisa que necessitam identificar a possibilidade de ocorrência de queimadas. Pode, assim, colaborar nas medidas para a redução de emissores atmosféricos, de modo a satisfazer o objetivo 15 da Agenda 30 definido pela Organização das Nações Unidas em 2015, o qual visa estimular ações de conservação, recuperação e uso sustentável de ecossistemas, especialmente.

Palavras-chave: modelagem *fuzzy*; modelo de previsão; *machine learning*; modelo *neuro-fuzzy*; redes neurais artificias.

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho" – Sorocaba (SP), Brazil.

[2]Centro Universitário Facens – Sorocaba (SP), Brazil.

Correspondence address: Pedro Antonio Galvão Junior – Rua Professor Tiberio Justo da Silva, 81 – Jardim Bandeirantes – CEP: 18133-000 – São Roque (SP), Brasil. E-mail: pedro.galvao@unesp.br

Conflicts of interest: the authors declare that there are no conflicts of interest.

Funding: none.

Received on: 12/01/2021. Accepted on: 06/14/2022.

https://doi.org/10.5327/Z2176-94781286

Introduction

In recent decades, fires have intensified throughout Brazil. This phenomenon has caused the emission of gases and pollutants into the atmosphere, contributing to changes in air quality and climate, influencing the quality of life of the population, and contributing to the occurrence of health problems in countless human beings. Fire has also influenced environmental patterns and processes worldwide by altering the distribution and structure of vegetation and creating new cycles of carbon emissions that directly affect the climate across the planet (Malmström, 2010).

The numerous impacts and alterations caused by fires are present in changes in soil characteristics that lead to the death of living beings, interfering with the environmental balance and causing incalculable environmental damage.

The analysis and monitoring of fire risk is a powerful tool adopted for mapping and sensing affected areas throughout the Brazilian territory and worldwide. Studies on the classification of indexes as fire risk indicators began in the mid-1930s in Canada and the United States. Through these studies, it was possible to divide indexes into cumulative and non-cumulative (Setzer et al., 2019).

Non-cumulative indexes are based only on the weather conditions prevailing on the day of data collection, while cumulative indexes take into account the climatic conditions of a succession of days for a maximum period of to 30 days (Soares and Batista, 2007).

In Brazil, since the mid-1960s it has been understood that adopting this type of resource is essential as a source of learning in order to create quick and effective actions against possible occurrences of fires caused by various natural or non-natural actions. The first classification of the level of fire risk was elaborated in Brazil in 1965, being motivated by the effect caused by the great fire of 1963 on Monte Alegre farm (Setzer et al., 2019).

Initially, a system whose main purpose was to serve as a form of protection against the occurrence of fires in localities of the state of Paraná was created by the company Klabin S.A, which was the most affected by the fire of 1963. Based on the information consolidated by Klabin, it was possible to establish one of the largest databases comprising data related to the occurrence of fires in Brazil.

Since the beginning of the 1980s, the National Institute for Space Research (INPE, acronym in Portuguese) has carried out daily activities related to the monitoring of fires throughout South America and especially Brazil. In 1998, the *Satellite Division and Environmental Systems* were created by INPE with the aim of developing, improving, and creating methods to estimate fire risk in a timely manner (Setzer et al., 2019).

In recent years, environmental issues have become increasingly evident worldwide due to the importance of the subject in association with the search for environmental preservation and ways to obtain an increasingly sustainable development.

After the consolidation of Agenda 21, the United Nations (UN) established in 2015 the Agenda 30, which refers to an action plan for the planet, for people, and for the future, consisting of 17 Sustainable Development Goals and 169 targets. Among its 17 goals, Goal 15 stands out, addressing how to protect, restore, and promote the sustainable use of terrestrial ecosystems, as well as how to manage the sustainability of forests to combat desertification, in order to stop and reverse land degradation and loss of biodiversity (supported by item 15.1). This item emphasizes the need to establish the conservation, recovery, and the sustainable use of ecosystems (especially forests), and stop land degradation by 2020 (ONU, 2015).

Establishing a way of learning and creating new knowledge through the classification of the risk levels of occurrence of fires over a period of days in the most diverse geographic areas of Brazil has become essential.

Thus, with the purpose of developing, improving, and creating methods to estimate fire risk in a timely manner, this paper presents a predictive model to establish a way of classifying fire risks in the Brazilian territory.

Literature review

The occurrence of fires, as well as their propagation, has a strong relationship with climatic factors, which influence the development of hotspots, especially in certain areas or regions of Brazil.

In this context, the use of meteorological data as a source of analysis has become an important tool for the elaboration of a prevention plan.

The adoption of preventive measures through the use of an index capable of reliably classifying the levels of fire risk is essential for planning prevention measures more efficiently, as well as adopting quick and effective actions in firefighting activities (Soares and Batista, 2007).

When thinking of how to estimate the possible occurrence of fires or the presence of fire risk, we are referring to the principles of the meteorological estimation called "fire risk" applied to vegetation burning. Burnings can be the result of the lack of rainfall or precipitation, which is caused by different types of climatic factors and decreases soil moisture (Soares and Batista, 2007).

To estimate or predict fire risk, we must initially consider the count of consecutive days without rain in a given location for the maximum amount of 120 consecutive days. This is a basic and simple way of attempting to establish a factor that allows stating that the possibility of occurrence of a fire is higher due to the lack of rainfall.

Other factors can be adopted in the creation of a method or index. Among them, the number of previous days that are analyzed in a given study to predict what may occur in the future is highlighted. According to the number of days, it is possible to identify values related to the number of days accumulated without rainfall, the volume of daily rainfall, and the rainfall accumulated for one or more days in relation to the presence of a previously observed fire risk (Setzer et al., 2019).

These factors are considered in the calculation and classification of fire risk. The values related to one or more factors must represent a period from 1 to 30 days.

Considering different studies that allowed us to design indicators related to the occurrence of fires, we highlight those shown in Table 1.

The Observed Fire Risk Index (RFO, acronym in Portuguese) was created by the Center for Weather Forecasting and Climate Studies/National Institute for Space Research (CPTEC/INPE, acronym in Portuguese) as a result of the analysis and recording of the occurrences of hundreds of thousands of fires in the main biomes present in the Brazilian territory, as well as in other South American countries. Its first version was applied to estimate the observed fire risk, which is dated from 2002 (Setzer et al., 2019).

Since then, this index is considered one of the main tools and sources of learning originating from the Forest Fire Monitoring Program by INPE. The index can be easily accessed and consulted through the website: https://queimadas.dgi.inpe.br/queimadas/portal.

The CPTEC/INPE uses different satellites (sensors) on a daily basis in order to produce essential data that can help to identify the so-called "thermal radiation", through which it is possible to recognize sources of high emission of radiation that are capable of causing a fire or that can be considered as a new heat source.

Methodology

Study area

The study was carried out in the municipality of Sorocaba, which stands out for being the fourth largest population in the countryside of the state of São Paulo and having the largest population in the southeastern region of the state, with approximately 695,000 inhabitants according to IBGE (2021). The municipality has an area of 449,872 km^2 and is located 87 km away from the state capital.

According to the Köppen climate classification, the climate is Cwa (subtropical), with mean temperatures of 37°C in summer and 22°C in winter, as well as annual rainfall of approximately 1,297 mm. The relief is classified as wavy, being characterized by the presence of slopes and high mountains, and an average altitude of 632 meters.

The highest altitude in the municipality (1,028 meters) is found at the headwaters of the Pirajibu River, Serra de São Francisco, close to the municipality of Alumínio, while the lowest altitude (539 meters) is found in the Sorocaba River valley.

The original natural vegetation is formed by the Atlantic Forest, with areas of dense ombrophilous forest consisting of Cerrado and secondary vegetation at various stages of succession. The municipality of Sorocaba was selected as the study area based on the variation in the level of fire risk presented in recent years, especially in the period defined for this paper. Figure 1 shows the map of the municipality of Sorocaba.

Historical rainfall data combined with spatial data referring to the record of the presence of fire risk for the municipality of Sorocaba were submitted to the processing of the Observed Fire Risk Index. Figure 2 shows the distribution of the 565 fires recorded in the municipality of Sorocaba.

Regarding rainfall data, the presence of the EDEN (Pirajibu) weather station, located in the neighborhood of Eden, met the needs of this study. Regarding the historical fire record, the municipality of Sorocaba presents an interesting variation of fire outbreaks per year among the 27 municipalities that comprise the metropolitan region of Sorocaba.

Model development

The Index of Level of Fire Risk in Brazil (ICRFB, acronym in Portuguese) was developed based on the analysis of the occurrence of observed fire risk and rainfall in the municipality of Sorocaba.

Through the combination of data from the National Institute for Space Research and data from the National Institute of Meteorology, the necessary conditions to build the ICRFB were obtained. Figure 3 highlights the sequence of processes and subprocesses used to build the ICRFB.

Table 1 – Studies on fire risk indicators

Study	Title	Authors	Type	Origin	Year
01	Development of a Forest Fire Risk Factor using Fuzzy Logic.	Isaac D. B. da Silva, Antonio C. F. Pontes Jr.	Paper	National	2011
02	Fire risk calculation method from the INPE program (2019).	Alberto W. Setzer, Raffi A. Sismanoglu, José G. M. dos Santos	Paper	National	2019
03	Assessment of the performance of three algorithms for land use classification using free geotechnologies.	Miqueias L. Duarte, Tatiana A. da Silva	Paper	National	2019
04	A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction.	Mohamed Ali Jallal, Aurora González-Vidal, Antonio F. Skarmeta, Samira Chabaa, Abdeouhab Zeroual	Paper	International	2020

Source: elaborated by the authors based on Silva and Pontes (2011); Duarte and Silva (2019); Setzer et al. (2019); Jallal et al. (2020).

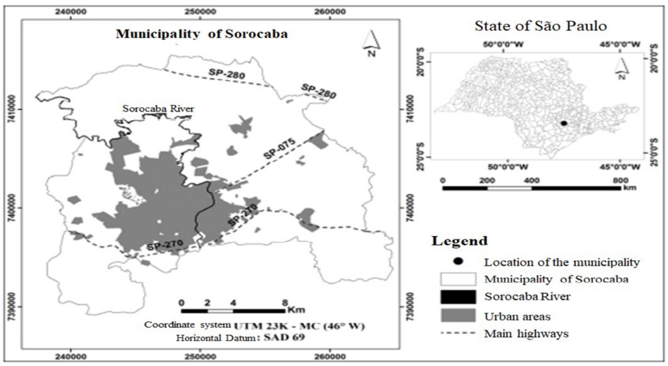

Figure 1 – Map of the municipality of Sorocaba.
Source: adapted from Lourenço et al. (2014).

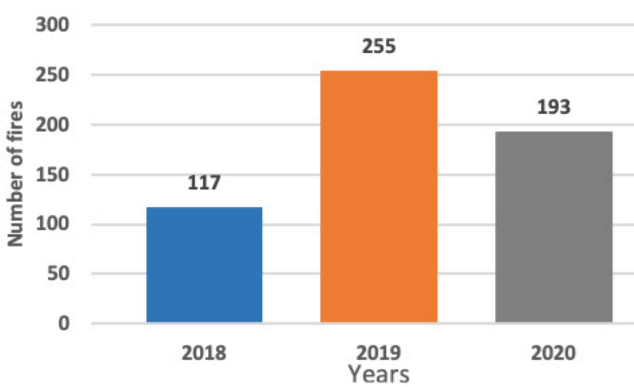

Figure 2 – Annual distribution of fires that occurred in the municipality of Sorocaba.

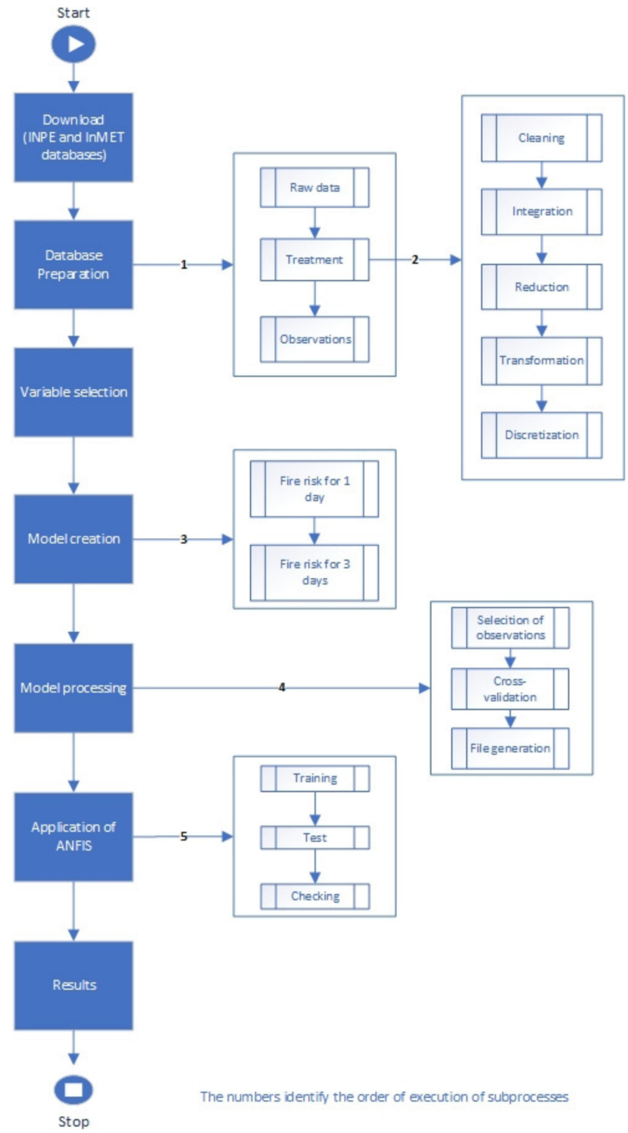

Figure 3 – Flowchart – Processes to build the ICRFB.

Variable selection

After understanding the elements that form the structure of the indexes presented above, the input variables selected for this study were defined, namely: day without rainfall and rainfall, which store data related to the daily occurrences of the presence or absence of rainfall.

The variable Observed Fire Risk was defined as the component responsible for presenting the result, being established based on the studies carried out by Lemos (2012).

However, if only these variables were used, the result would possibly be biased and not concrete due to the existence of values in the pre-established variables that only address the occurrence of fire and rainfall in the study area.

The absence of these phenomena was not identified. Thus, it became necessary to establish the other input variables that would be combined with the ones that were already defined, in order to increase the reliability of the process of analysis of the values in the databases. During the research, analysis, and simulations, new variables were recognized precisely to address issues related to the non-occurrence of fires and rainfall, namely: presence of rainfall and presence of fire.

The databases did not present any data related to the non-occurrence of these two natural phenomena in their observations. There were only observations regarding fire or rainfall that occurred in the municipality of Sorocaba for each day in 2018, 2019, and 2020.

The creation of these variables was conditioned to the need to offer a form of assistance for the implemented models that would ensure greater assertiveness in defining the result to be obtained, considering all the possibilities for each of the days under analysis, and knowing that there could be the two phenomena, only one phenomenon or none of them on the same date.

Given the needs highlighted above, the relevant variables were defined, as shown in Table 2.

Database

The data was obtained from the fire database of INPE, which was made available by its Fire Program. Only data related to the municipality of Sorocaba from the period of days between 01/01/2018 and 12/31/2020 were selected.

These data represent the occurrence of fire records identified throughout each day. Table 3 describes the representation of each variable in the fire database of INPE.

The database of daily rainfall kept by INMET was also used, which is associated with the Pirajbu weather station. Table 4 describes the variables selected from this database.

These data sets can be obtained directly by downloading the files corresponding to a period between 30 and 365 days, preferably in the format of values separated by commas.

Database preparation

As only the data of fires related to the municipality of Sorocaba stored in the database of INPE should be identified and there was the need to combine the rainfall data existing in the database of INMET, it was necessary to prepare the relevant databases for subsequent use.

For this stage, the Microsoft SQL Server 2019 Database Management System (DBMS) was used, following the steps presented in the methodology by Wong (2017). Thus, the data of fires in 2018, 2019, and 2020, in combination with the data referring to daily rainfall, were identified. Figure 4 provides a summary of the database preparation process. A brief summary of each step is described below.

- Raw data: data in the fire database of INPE that presented records of the fire outbreaks in every state and municipality in Brazil between 2018 and 2020 totaled more than 11 million records. Among them, 560 records related to the study area in association with the rainfall data found in the database of INMET were selected;

Table 3 – Description of the variables that make up the INPE Fireworks Database.

Name of variable	Description
Datehour	Displays the reference time of the passage of the satellite in Greenwich Mean Time (GMT);
Satellite	Identifies the name of the algorithm used and references the satellite that provided the image;
Country/State/ Municipality	Political-geographical entities in the Brazilian territory where the presence of fires were identified;
Biome	Name of the biome according to the Brazilian Institute of Geography and Statistics (IBGE, acronym in Portuguese);
DayWithoutRainfall	Number of days without rainfall until the detection of the heat source, which are measured within the range of values between 0 (zero) and 120 (one hundred and twenty);
Rainfall	Rainfall accumulated on the day until the detection of the heat source, which is measured within the range of values between 0 (zero) and 120 (one hundred and twenty);
ObservedFireRisk	Represents the value corresponding to the Observed Fire Risk for the day of detection of the heat source, which is identified within the range of values between 0.1 (zero point one) and 1 (one);
Latitude / Longitude	Displays the geographic coordinates of the center of the pixel that contains a possible fire source or presence of heat above 47º C.

Source: INPE (2019)

Table 4 – Description of the variables of the rainfall database.

Name of variable	Description
Days with Rainfall	Total of rainy days for the month referring to the precipitation date.
Monthly Rainfall	Total rainfall accumulated over the month according to the rainfall date.
Daily Rainfall	Value referring to the rainfall volume identified for the date of rainfall.

Source: INMET (2021).

Table 2 – Description of selected variables

Name of variable	Description	Value range	Unit	Type
DayWithoutRainfall	Represents the number of days without rainfall.	0-120	Days	Input
RainfallonthePreviousDay	Represents the amount of rainfall from the previous day.	0-120	Milimeters	Input
AccumulatedRainfallonthePreviousThreeDays	Represents the total amount of rainfall accumulated in the previous three-day period.	0-120	Milimeters	Input
PresenceofRainfall	Represents the identification of the presence or absence of rainfall.	0 or 1	True/False	Input
PresenceofFire	Represents the identification of the occurrence or not of a fire.	0 or 1	True/False	Input
ObservedFireRisk	Represents the value related to the Observed Fire Risk for 1 or 3 days.	0.1-1	Decimals	Output

- Treatment of data:
  - Cleaning: to input missing values, remove noise, and correct inconsistencies;
  - Integration: to unite data related to fires and rainfall, forming a single data source. In this task, the non-occurrence of fires and rainfall was also considered;
  - Reduction: to reduce the dataset to be analyzed, considering the occurrence of fire or rainfall according to the date, and not date and time;
  - Transformation: to standardize data by establishing a specific format for the study;
  - Discretization: to reduce the amount of continuous values existing in some variables.
- Observations: Result of the treatment of data, from which 1286 observations were selected to be applied to the models. In addition, the correlation between the record of fires, fire risk, and rainfall for the same period of days was identified.

Cross validation

Cross validation is a technique used to evaluate the generalization capacity of a model from a dataset, being widely used in problems where the modeling purpose is prediction. It aims to estimate the accuracy of the model in practice based on performance and assertiveness in relation to a new dataset (Ling et al., 2019).

One of the ways to divide these data is using the "holdout" method, which consists of randomly dividing the data into 70-30, that is, 70% of the data for the training step and 30% for the test or check steps. The disadvantage of using this technique is that the portion of data selected for training, testing or checking may be very similar, and thus present a good evaluation of the model.

However, when we put the model into production, the new data turns out to be very different from the data already known by the model, which generates bad results and practically makes this dataset unusable (Wong, 2017).

Another way of dividing the data is called "cross validation", which allows avoiding problems of randomness. Thus, variance can be avoided and the results are more robust.

The only disadvantage presented by this technique is directly related to possible impacts regarding processing time and performance (Ling et al., 2019). If the database presents a large number of observations to be analyzed, cross validation may become a longer process in relation to the processing time for the presentation of the results.

For this study, the *k-fold* cross validation technique was used, where a number of k (*folds* = 5) represents the amount of data clusters created, considering four combined groups that form a single group selected for the stage of data training, while the remaining group is used in the checking stage.

Out of the total of 1,286 observations identified in the stage of data preparation, 900 were distributed into five (5) groups consisting of 180 observations that were randomly selected, not allowing for the use of observations that were previously selected to comprise this new group. Figure 5 shows how k-fold cross validation was developed, as well as how observations are separated into the four (4) training groups, together with one (1) checking group.

ANFIS model

The neuro-fuzzy model is called a hybrid model for establishing a relationship between the fuzzy inference system and neural networks, allowing the information to be trained and creating the rule bases automatically.

This integration between the different techniques (neural networks, genetic algorithms, and fuzzy logic) allows creating a computational system that incorporates human knowledge, being able to learn and adapt it to the environment according to uncertainties and inaccuracies, which is known as adaptive neuro-fuzzy inference system (ANFIS).

The Takagi-Sugeno fuzzy inference system represents a dynamic system or a control that associates a set of linguistic rules in the antecedent ("if" part) with fuzzy propositions, while in the consequent ("then" part) functional expressions of the type and linguistic variables of the antecedent are presented instead of fuzzy sets, as used in the Mamdani model. The adaptive neuro-fuzzy inference system (ANFIS) uses the methodology pro-

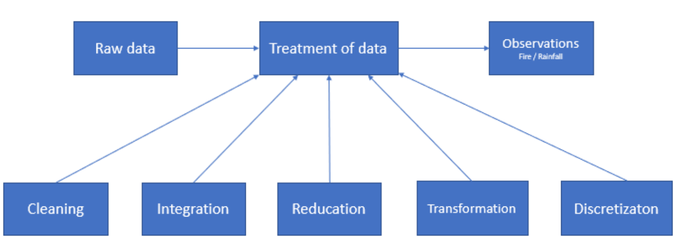

Figure 4 – Steps of the database preparation process.

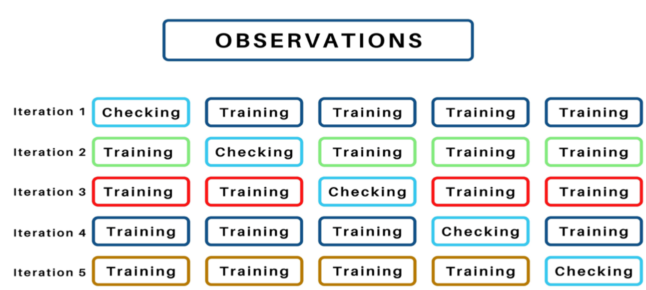

Figure 5 – *5-folds* cross validation.

posed by Takagi and Sugeno and can be divided into up to 5 layers, as shown in Figure 6.

Layers can be divided into linear and non-linear. The fuzzy logic is responsible for dividing these two blocks (non-linear and linear) (Cuevas et al., 2018).

Layer 1: corresponds to the input of values defined according to the set of input variables (input pairs) established for the model that will be submitted to the ANFIS. This layer is formed by the membership function, which specifies the degree of membership of the values to satisfy the linguistic terms (Eq. 1).

$$0\frac{1}{i} = u_{Ai}(x) \tag{1}$$

Layer 2: corresponds to the logical operator of the fuzzy set, which assigns degrees of membership to each of the variables, being described by "linguistic variables" (Eq. 2).

$$w^1 = o_i^1 \cdot O_i^2 \tag{2}$$

Layer 3: Calculates the normalized degree of activation of each node through the ratio between the applicability of the i-th rule (wi) and the sum of the applicability (Eq. 3).

$$O_i^3 = \underline{w}_i = \frac{wi}{\Sigma_J^H=1}\,\omega_J \tag{3}$$

Layer 4: Multiplies the output of the 3rd layer with a function fi, establishing a linear combination for each existing value within the set of input pairs. The values ($piqiri$) represent the adjustable parameters that determine the response of the function (Eq. 4).

$$O_i^4 = \underline{w}_i f_i = \underline{w}(\rho_i u_1 + q_i u_2 + r_i) \tag{4}$$

Layer 5: Calculates the sum of all nodes from the previous layer (Eq. 5).

$$O_i^5 = \Sigma_{\ i}\ \overline{w} f_i = \frac{\Sigma_i\ \overline{w} f}{\Sigma_i\ w} \tag{5}$$

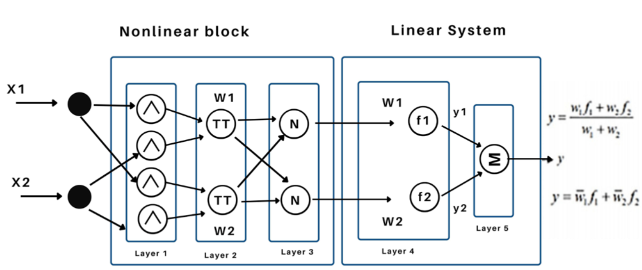

Figure 6 – Typical architecture of an adaptive neuro-fuzzy inference system (ANFIS).
Source: adapted from Cuevas et al. (2018).

The possibility of interaction between the existing layers in an ANFIS allows stating that the inference system has the same advantages compared to systems created by specialists (Mubarak et al., 2021).

The learning of the ANFIS system takes place through two sets of parameters directly involved in the training stage. One set is known as the antecedent variables, treated by constants that characterize the membership functions, while the other is known as the consequent variables, defined with linear functions applied to the output of the inference model.

The structure of an ANFIS system involves the use of a variable selection by determining the number of membership functions per variable and obtaining a set of fuzzy rules. When the amount of clusters for a given data set is not previously known, subtractive clustering is a fast and robust algorithm to obtain this number (Azad et al., 2019).

In addition, this technique allows the location of the cluster center, thus obtaining membership functions and rules from these cluster centers. With this information, it is possible to generate a Takagi-Sugeno fuzzy inference system that models the behavior of the data.

Models for fire risk prediction

After preparing the database and selecting the variables that are relevant to the purpose of this paper, several models were elaborated, trained, and tested in order to be able to identify the most assertive ones regarding mean absolute errors, the lowest level of complexity in its elaboration, and the processing time related to the amount of processed observations. Six different models were defined with the purpose of predicting the occurrence of the observed fire risk from one (1) up to three (3) days.

These models were named through a numerical sequence as Model 1, Model 2, Model 3, and so on, up to Model 6. Models 1, 2, and 3 present the prediction of fire risk for 1 day, while Models 4, 5, and 6 are designed to predict fire risk for 3 days.

In a practical way, the models present differences in their structures, that is, they do not have the same amount of input variables. This difference was established in order to identify changes in the behavior of each model, especially variations in the results.

The equality existing in each model refers to the output variable. All models have a single output variable called fire risk. Tables 5, 6, and 7 describe the structure of Models 1, 2 and 3, which were reused for the other models.

Results and Discussion

From the processing of the 1286 observations identified during the stage of data preparation, 386 observations (30% of the total) were randomly selected to comprise the data set conditioned to the training stage.

The other 900 observations were used in the cross validation process, which established the five (5) groups consisting of 180 observations also randomly selected without repetition. These groups were submitted to the training and data checking stages for each of the six (6) models.

The models were processed through the ANFIS hybrid model configured for the Fuzzy Inference System (FIS) applied to the subtractive clustering method, being processed over 100 epochs within the MATLAB computing environment through the toolbox called *AnfisEdit*.

The subtractive clustering method enabled better results in terms of data processing time due to its greater agility to read and process historical series that have more than four input items of data (Milan et al., 2021).

Table 8 shows the results obtained for the models developed to predict the presence of fire risk for one (1) day right after the training and data checking stages, according to the 5-fold cross validation technique.

Through the results shown in Table 8, Model 1, formed by five (5) input variables, presented the lowest mean absolute error and the highest standard deviation compared to the other two models.

This difference is caused by the amount of input data and their respective value ranges.

Table 9 shows the results referring to the models developed to predict the presence of fire risk for a maximum period of 3 days. The values were obtained at the end of the training and checking stages, according to the definitions of the 5-fold cross validation technique.

Based on the results shown in Tables 8 and 9, among the models created to predict the observed fire risk for one day, Model 3 was highlighted.

This model stands out for its simplicity, being formed by a small structure consisting of the input variables "day without rainfall" and "rainfall on the previous day", unlike Models 1 and 2, which consist of these and other input variables that did not initially exist in the database, such as the variable "accumulated rainfall on the previous three days", generated by calculating the sum of the rainfall volume of the three days prior to the date of analysis of the observed fire risk.

During the processing of each model submitted to the fuzzy inference system through the MATLAB computing environment, different types of membership functions were used by the inference system in relation to the set of values existing in the input and output variables.

Figure 7 highlights the list of membership functions defined for Model 3.

Figure 7 shows that the second input variable ("rainfall on the previous day") practically used only one membership function, unlike the first input variable ("day without rainfall"), which used three (3) Gaussian membership functions for its set of values.

Table 5 – Description of input and output variables selected in Model 1.

Name of variable	Value range	Unit	Type
DayWithoutRainfall	0-120	Days	Input
RainfallonthePreviousDay	0-120	Milimeters	Input
AccumulatedRainfallonthePreviousThreeDays	0-120	Milimeters	Input
PresenceofRainfall	0 and 1	True/false	Input
PresenceofFire	0 and 1	True/false	Input
ObservedFireRiskforOneDay	0.1-1	Risk level	Output

Table 6 – Description of input and output variables selected in Model 2.

Name of variable	Value range	Unit	Type
DayWithoutRainfall	0-120	Days	Input
RainfallonthePreviousDay	0-120	Milimeters	Input
AccumulatedRainfallonthePreviousThreeDays	0-120	Milimeters	Input
ObservedFireRiskforOneDay	0.1-1	Risk level	Output

Table 7 – Description of input and output variables selected in Model 3.

Name of Variable	Value range	Unit	Type
DayWithoutRainfall	0-120	Days	Input
RainfallonthePreviousDay	0-120	Milimeters	Input
ObservedFireRiskforOneDay	0.1-1	Risk level	Output

Table 8 – Summary of results – Models in the one-day period.

Models	Variables		Mean error	Standard deviation
	Input	Output		
Model 1	5	1	0.02	0.05
Model 2	3	1	0.03	0.01
Model 3	2	1	0.03	0.01

Table 9 – Summary of results – Models in the three-day period.

Models	Variables		Mean error	Standard deviation
	Input	Output		
Model 4	5	1	0.05	0.18
Model 5	3	1	0.05	0.01
Model 6	2	1	0.04	0.00

This behavior becomes evident when the values in each observation line selected by ANFIS present very close intervals, that is, the difference between these values is considered small. Thus, the same membership function ends up being used in the analysis and definition of results.

Regarding the output variable "observed fire risk for one day", it uses the Takagi-Sugeno linear output function, making it possible to establish a pattern of recognition and presentation for each line of output data recognized by the model.

Figure 8 shows the surface graph established by the ANFIS hybrid model according to the combination of some input variables, day without rain and precipitation existing in Model 3.

As Model 3 stood out among the set of models implemented to predict the presence of fire risk for one day, the models developed for three days also presented a possible candidate, Model 6.

Model 6 presented a mean absolute error of approximately 4%. It is worth mentioning that Model 6 is formed by the same set of input variables as Model 3.

Soon after the execution of the training and checking stages, all models were submitted to the evaluation stage of the fuzzy inference system, whose purpose is to ensure that the results returned by the ANFIS hybrid model are close to the so-called real values.

The evaluation process of the fuzzy inference system was carried out using the command *EvalFIS* of the MATLAB software. This functionality enabled the validation of the structure of the fuzzy inference system by taking into account the set of rules established for its learning.

At the end of the evaluation, the *EvalFIS* tool presents a comparison between the real data used by the system and the output data obtained. This comparison creates a list containing the values of the calculated outputs resulting from the combination of the so-called real data together with the calculated data. Thus, it becomes possible to concretely recognize the output results that the inference system is returning.

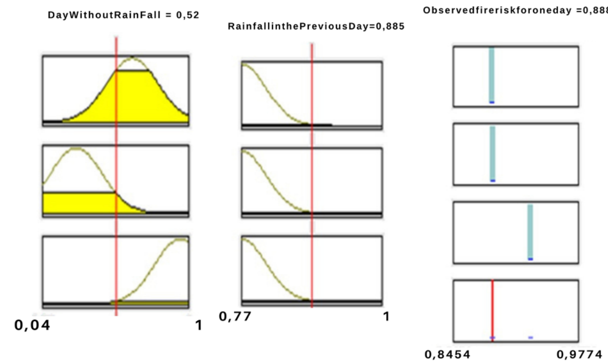

Figure 7 – Membership functions in Model 3.

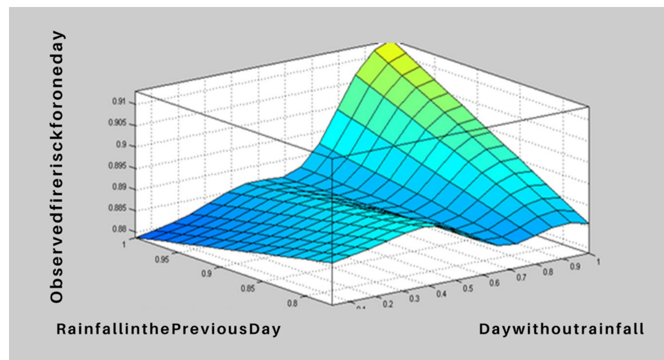

Figure 8 – Surface graph – "day without rainfall" and "rainfall".

Table 10 illustrates some values presented by the output variable "observed fire risk" for one day, which has real data, combined with the output generated after the evaluation of the fuzzy inference system belonging to Model 3.

Through the combination of the input variables together with the definition of the output variable "observed fire risk" for one or three days, it was possible to achieve the final objective of this study. Table 11 shows the levels of classification and their respective value intervals.

From the definition of the classification of levels of fire risk and intervals of values, the fuzzy matrix was applied in order to identify the number of observations belonging to each level.

The fuzzy matrix enabled verifying the degree of assertiveness of the results, besides recognizing the values related to the statistical classes "sensitivity" and "specificity".

Table 12 presents the fuzzy matrix conceived from the levels of classification.

The statistical class of sensitivity helps to measure the ability to minimize false negatives, which is recognized as the ability to correctly identify observations that the model also correctly identified.

However, the statistical class of specificity has the purpose of measuring the negative values in the set of values returned by the model, which are also preestablished as negative. The values of each class range between zero (0) and one (1).

Table 10 – Variable "observed fire risk" combined with the calculated output.

ObservedFireRisk (real output)	ObservedFireRisk (calculated output)	Difference (module)
0.84	0.88	0.04
0.85	0.89	0.04
0.94	0.89	0.05
0.95	0.89	0.06
1	0.89	0.11

Table 11 – Classification of levels of fire risk.

Level of Classification	Intervals
Minimum	[0; 0.15]
Low	(0.15; 0.40]
Medium	(0.40; 0.70]
High	(0.70; 0.95]
Critical	(0.95; 1]

Source: adapted from INPE (2019).

Table 13 shows the values correlated between the level of classification and the statistical class.

Through the values shown in Table 13, an analysis to examine the values returned from the execution of the fuzzy matrix in relation to the critical level of classification can be carried out.

In this analysis, it is shown that that specificity was very close to one (1), while sensitivity reached the value of one (1). In general, the mean value for sensitivity and specificity was approximately 0.38 and 0.86, respectively. These values were calculated right after the fuzzy matrix was built.

The set of values returned by these statistical classes belonging to each level of classification was taken into consideration. Thus, the sum of all values and their subsequent division was carried out according to the levels of classification, in order to use them to build the fuzzy matrix.

The results obtained through the confusion matrix show that Model 3 demonstrates a good ability to be able to specifically identify a value that was assigned to a classification but which actually should belong to another.

The ability to correctly identify the observations that the hybrid model recognized as true can be considered correct in relation to its sensitivity to the "high" and "critical" levels of classification, based on the data set used to build the fuzzy matrix.

Specificity also represented the low existence of negative values for the "minimum", "low", "medium", and "critical" levels of classification, which present values close to one (1).

Table 12 – Fuzzy matrix.

	Minimum	Low	Medium	High	Critical
Minimum	0	0	0	0	0
Low	0	0	0	0	0
Medium	0	0	0	0	0
High	0	6	21	313	0
Critical	0	2	6	28	10

Table 13 – Results of the statistical classes of sensitivity and specificity.

	Minimum	Low	Medium	High	Critical
Sensitivity	0	0	0	0.92	1
Specificity	1	1	1	0.40	0.90

In short, Model 3 is the most suitable for use as a hybrid model to build a classification index of levels of fire risk in Brazil, reaching the goals proposed by this paper.

Conclusion

According to the models presented in this study, ways of assessing the presence of fire risk were identified for the intervals of 1 or 3 days applied to the municipality of Sorocaba, which can also be reproduced in other locations around the world. The comprehension of behaviors, rules, and conditions that can cause the occurrence of natural phenomena, especially fires was predominant to establish a form of initial analysis of the data set selected for the study.

The database preparation enabled a clearer definition of the scenario to be worked on, that is, after data was processed, standardized, and ready to be recognized as observations, the path of application of the ANFIS hybrid model was defined.

The results found become a valuable source of knowledge offered to the society in general, companies, and teaching or research institutions regarding the definition of programs and public policies for the prevention, conservation, and combat of fires.

The proposed model is prepared to learn based on its level of assertiveness and all the factors applied to it, which allow its reproduction in other studies and as a support tool to be applied in new public policies, together with expert systems, or simply as a new source of knowledge.

This study was limited to the use of spatial data provided by INPE related to fires and fire risk, which were combined with rainfall data provided by INMET, both related to the municipality of Sorocaba, to establish the prediction of the presence of fire for the maximum period of three days.

Contribution of authors:

GALVÃO JÚNIOR, P. A.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing — Original Draft. ROVEDA, S. R. M. M.: Conceptualization; Investigation; Methodology; Supervision; Project Administration – Revision and Editing. VIEIRA, H. E. M.: Formal analysis; Investigation; Methodology; Validation; Writing – Revision and Editing.

References

Azad, A.; Manoochehri, M.; Kashi, H.; Farzin, S.; Karami, H.; Nourani, V.; Shiri, J., 2019. Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling, v. 571, 214-224. https://doi.org/10.1016/j.jhydrol.2019.01.062.

Cuevas, E.; Díaz, P.; Zaldívar, D.; Cisnero, P.M., 2018. Nonlinear system identification based on ANFIS-Hammerstein model using Gravitational search algorithm, v. 48, (1), 182-203. https://doi.org/10.1007/s10489-017-0969-1.

Duarte, M.L.; Silva, T.A., 2019. Avaliação do desempenho de três algoritmos na classificação de uso do solo a partir de geotecnologias gratuitas. Revista de Estudos Ambientais, v. 21, (1), 6-16. https://doi.org/10.7867/1983-1501.2019v21n1p6-16.

Instituto Nacional de Meteorologia (INMET), 2021. Banco de dados meteorológicos. Mapa das Estações (Accessed Apr., 2021) at:. https://mapas.inmet.gov.br/.

Instituto Nacional de Pesquisas Espaciais (INPE), 2019. INPE program fire risk calculation method – version 11. INPE, 2019. (Accessed Apr., 2021) at:. https://queimadas.dgi.inpe.br/~rqueimadas/documentos/RiscoFogo_Sucinto.pdf.

Instituto Brasileiro de Geografia e Estatística (IBGE), 2021. Estimativa Populacional. IBGE (Accessed May 9, 2022) at:. https://ibge.gov.br/cidades-e-estados/sp/sorocaba.html.

Jallal, M.A.; González-Vidal, A.; Skarmeta, A.F.; Chabaa, S.; Zeroual, A., 2020. A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction. Applied Energy, v. 268, 114977. https://doi.org/10.1016/j.apenergy.2020.114977.

Lemos, C.F.; Justino, F.B.; Costa, L.C.; Maddock, J.E.L., 2012. Distribuição espacial do índice de Haines para Minas Gerais por análise da média atmosfera. Revista Brasileira de Agropecuária Sustentável, v. 2, (1), 132-143. https://doi.org/10.21206/rbas.v2i1.68.

Ling, H.; Qian, C.; Kang, W.; Liang, C.; Chen, H., 2019. Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment. Construction and Building Materials, v. 206, 355-363. https://doi.org/10.1016/j.conbuildmat.2019.02.071.

Lourenço, R.W.; Silva, D.C.C.; Sales, J.C.A. 2014. Elaboração de uma metodologia de avaliação de fragmentos de remanescentes florestais como ferramenta de gestão e planejamento ambienta. Revista Eletrônica Capital Científico, v. 10, (3), 685-698. https://doi.org/10.5935/ambiencia.2014.03.03.

Malmström, A. 2010. The importance of measuring fire severity – Evidence from microarthropod studies. Forest Ecology and Management, v. 260, (1), 62-70. https://doi.org/10.1016/j.foreco.2010.04.001

Milan, S.G.; Roozbahani, A.; Azar, A.N.; Javadi, S., 2021. Development of adaptive neuro fuzzy inference system – Evolutionary algorithms hybrid models (ANFIS-EA) for prediction of optimal groundwater exploitation. Journal of Hidrology, v. 598, 126285. https://doi.org/10.1016/j.jhydrol.2021.126258.

Mubarak, S.M.J.; Crampton, A.; Carter, J.; Parkinson, S., 2021. Robust data expansion for optimized modelling using adaptive neuro-fuzzy inference systems. Expert Systems with Applications, v. 189, 116138. https://doi.org/10.1016/j.eswa.2021.116138.

Organização das Nações Unidas (ONU), 2015. Transformando Nosso Mundo: A Agenda 2030 para o Desenvolvimento Sustentável. Nova York: ONU (Accessed Apr., 2021) at:. https://nacoesunidas.org/wp-content/uploads/2015/10/agenda2030-pt-br.pdf.

Setzer, A.W.; Sismanoglu, R.A.; Santos, J.G.M., 2019. Método do cálculo do risco de fogo do programa do INPE. Brasil: Ministério da Ciência, Tecnologia, Inovações e Comunicações, v. 11. (Accessed Jun., 2019) at:. http://queimadas.dgi.inpe.br/~rqueimadas/documentos/RiscoFogo_Sucinto.pdf.

Silva, I.D.B.; Pontes, A.C.F., 2011. Development of a Forest Fire Risk Factor using Fuzzy Logic. Biomatemática, v. 21, 113-128.

Soares, R.V.; Batista, A.C., 2007. Incêndios florestais: controle, efeito e usos do fogo. Universidade Federal do Paraná, Curitiba, 250 pp.

Wong, T.T., 2017. Parametric methods for comparing the evaluated by k-fold cross validation on multiple data sets, v. 65, 97-107. https://doi.org/10.1016/j.patcog.2016.12.018.