Environ Earth Sci
DOI 10.1007/s12665-015-4421-8

ORIGINAL ARTICLE

A data-based model for predicting wildfires in Chapada das Mesas National Park in the State of Maranhão

Fábio Teodoro de Souza[1] · Tarcyzio Cezar Koerner[2] · Rafael Chlad[2]

Received: 12 February 2014 / Accepted: 14 April 2015

Abstract Chapada das Mesas National Park extends over an area of 160,046 ha in the municipalities of Carolina, Riachão, Estreito and Imperatriz in the south central region of the state of Maranhão, northeast Brazil, in a savanna-like biome known as the *Cerrado*. The park has a rich biodiversity, making the need for conservation all the more important. The weather conditions in the region increase the likelihood of wildfires, so that a monitoring and control system for the area is needed to help conservation efforts. This article proposes a methodology that uses data-mining techniques to predict outbreaks of wildfires in the park some hours in advance. Predictive models using wildfire records and a meteorology dataset for 11 months in 2010 were built. Two different classification techniques for predicting wildfires were used: artificial neural networks and classification rules. The two models built with these techniques showed good accuracy when tested with the validation samples and could be used as additional tools for predicting the risk of fires in the area.

✉ Fábio Teodoro de Souza
fabio.teodoro@pucpr.br

Tarcyzio Cezar Koerner
tarcyzio@hotmail.com

Rafael Chlad
rafael_chlad@hotmail.com

[1] Postgraduate Program in Urban Management (PPGTU) - School of Architecture and Design, Course of Civil Engineering - Polytechnic School, Pontifical Catholic University of Paraná (PUCPR), Rua Imaculada Conceição, 1155-Prado Velho, PO BOX: 16210, Curitiba, Paraná 80215-901, Brazil

[2] Civil Engineering Department, Federal University of Paraná (UFPR), Curitiba, Paraná, Brazil

Keywords Wildfires · Meteorological dataset · Data mining · Artificial neural networks · Classification rules

Introduction

Federal Conservation Areas (FCAs) were set up to protect native fauna and flora as well as the processes that govern ecosystems, thereby maintaining regional biodiversity and protecting the customs of indigenous populations. They therefore represent the best strategy for protecting natural attributes and heritage (Arruda 1999).

Chapada das Mesas National Park (CMNP) is a federal reserve that was created by presidential decree on December 12th, 2005 (Fig. 1). It extends over 160,046 hectares of the *Cerrado* (savannah-like grassland) and lies in the municipalities of Carolina, Riachão, Estreito and Imperatriz in the south central region of the state of Maranhão. The reserve was set up in response to historic demands by the local community for a conservation area to protect the Chapada das Mesas region against agricultural expansion and developments that could put its ecological value and ecotourism potential at risk (IBAMA 2013).

CMNP is situated in an area that acts as an ecotone between three biomes—*Cerrado*, *Amazônia* (the Amazon rainforest) and *Caatinga* (semi-arid thorn forest/dryland)—and is therefore an extremely high-priority area for conservation of the biodiversity of the *Cerrado* biome (one of 25 biodiversity hotspots around the world). The region has great potential to support a wide variety and abundance of species of national fauna and flora and is characterized by a humid tropical climate with high temperatures throughout the year. It has two clearly defined seasons: a rainy winter, between November and April, and a dry summer, between May and October, approximately. The months of July,

Published online: 08 May 2015

August and September are the most critical in terms of drought and wildfires. Total annual rainfall is between 1250 and 1500 mm. Mean annual temperature is around 26.1 °C, with minimum temperatures varying from 25.2 °C in January to 27.8 °C in September. Maximum temperatures reach around 36 °C in July and August (IBAMA 2013).

Wildfires are among the main causes of destruction of forest ecosystems. As well as destroying native flora and fauna, they cause pollution and climate changes.

In 2011, Maranhão had around 16,000 hotspots and was classified as one of the three states with the greatest number of wildfires. In addition to the problem of devastation of the biome, pollutants are given off in the form of greenhouse gases when plant biomass is burnt.

Wildfires can be detected by remote sensing, using algorithms and satellite sensors. However, some technological or natural limitations can make it difficult to detect fire fronts and can lead to inaccurate data. These include dense cloud cover, thick forest (where fires do not reach the treetops), fires that occur in the period between images and fires on slopes outside the satellite's field of view.

Despite these disadvantages, satellites are effective tools for controlling and studying wildfires and avoid additional costs associated with terrestrial surveillance.

Understanding wildfire patterns is therefore essential to develop a control and monitoring system and so preserve the biome in the region.

Fires are important phenomena in the Earth system and occur in a wide range of spatial and temporal scales. However, to predict them, probabilistic analyses are used, and despite the complexity involved, predictions are helpful in the decision-making process (Taylor et al. 2013).

Hanson et al. (2000) discussed a proposal for an operational center to monitor and predict wildfires. The prediction of these catastrophic phenomena is essential because they affect negatively the biodiversity and natural environment. This study discussed the implementation of computational models in the mitigation of negative effects, and several different prediction methodologies were proposed by different authors.

Chu et al. (2002) applied a logistic regression model to predict wildfires in growing areas on islands in Hawaii. The model examines atmospheric patterns, such as El Niño and

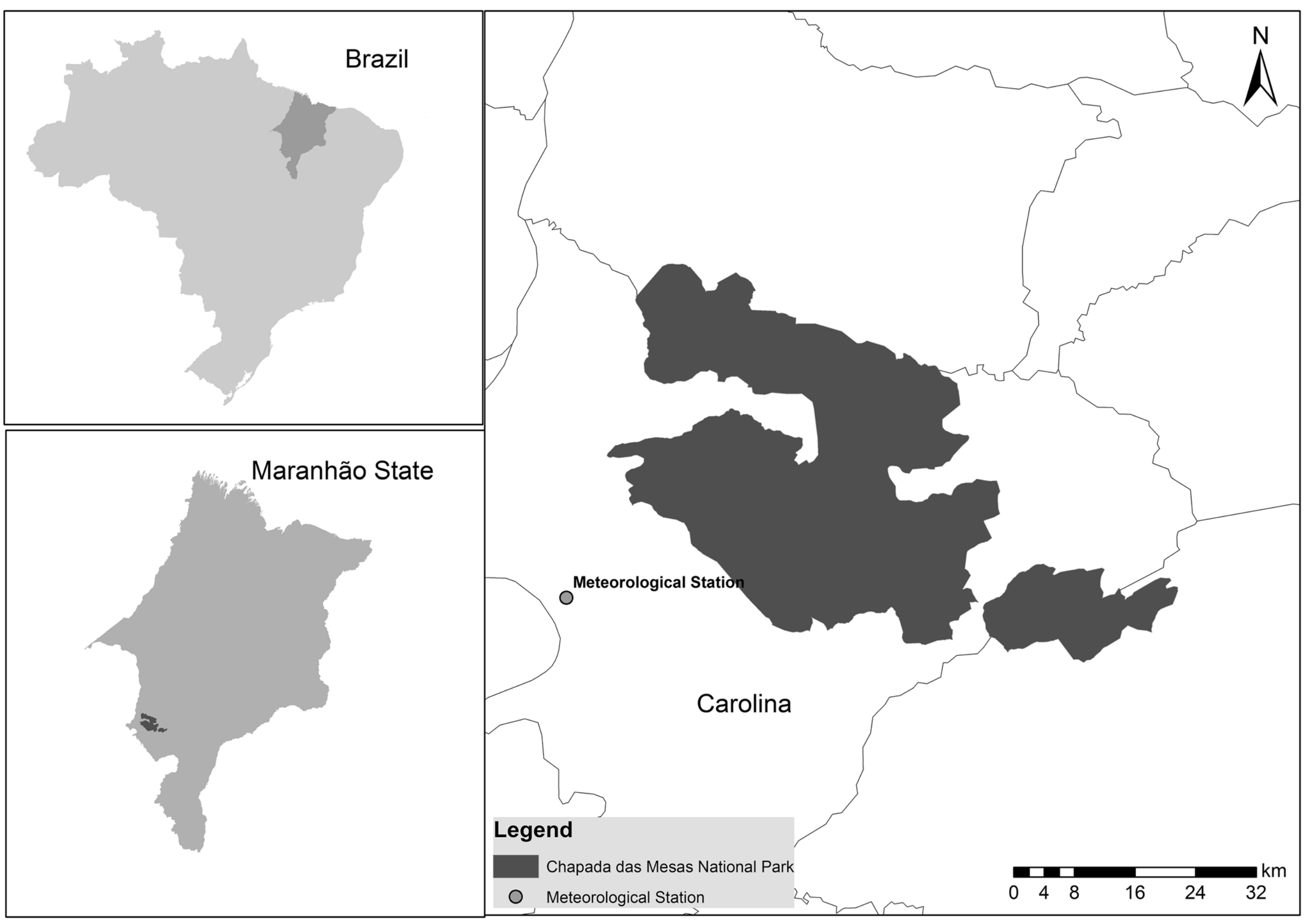


Fig. 1 Geographic location of Chapada das Mesas National Park (source http://mapas.icmbio.gov.br)

Southern Oscillation (ENSO), and enhances the prediction of wildfires a maximum of two seasons in advance.

Beckage and Platt (2003) underlined the problem of wildfires associated with the costs of fire brigades and property losses. They also proposed a methodology with time series models to predict the severity of wildfires in South Florida Everglades National Park. These models can be used as management tools to predict intense wildfires, a month up to a maximum of 1 year in advance.

The study conducted by Peng et al. (2005) considered the use of numerical indices as tools to aid in the management of wildfires and in the evaluation of dangers. Hence, they evaluated the effectiveness of the burning index (BI) in the prediction of wildfires in Los Angeles County, California, using spatial–temporal models.

Goslar (2006) proposed the use of remote sensing to identify indicators of fuel (vegetation) that can assist in the prediction of forest fires in South African rural communities. Therefore, the author analyzed satellite images collected during the dry season (winter), when the quantity of fuel (dry vegetation) was at its highest. For comparative proposes, one image from the rainy season (summer) was also collected. To validate the satellite images sample, the author performed in situ measurements of the aboveground biomass.

Sun et al. (2009) indicated that the complexity of wildfire prediction is because of changes in the interaction of different atmospheric parameters, such as turbulence induced by wind gusts and eddies in the atmospheric boundary layer. The authors proposed simulations using data from fires in pastures and atmospheric parameters. The results have implications in the prediction of fire operational behavior.

The study of Gasull et al. (2011) used computational intelligence to predict wildfires through wireless sensors. The models developed offer two types of information: risk of fire occurrence, and in the case of an active fire, flame direction and propagation speed. These models aid decision-makers in planning escape methods and defining strategies to delay fire occurrence.

Das et al. (2013) described the catastrophic fires that occurred in Australia during the summer, which had a strong negative socio-economic impact. In their paper, they discussed the application of hybrid neural networks to predict locations with higher fire probability using meteorological and hydrological variables in a monthly scale, such as total evaporation, sensible heat flux, precipitation, incoming solar radiation, maximum temperature, and soil moisture.

Imada (2014) discussed the use of neural networks and artificial intelligence in fire prediction. He also contextualized the management of forest resources and the preservation of biodiversity in forest ecosystems.

This article presents a methodology based on data-mining techniques for predicting wildfires in CMNP some hours in advance. The forecasting models were developed with historical wildfire and meteorological data collected between January and November 2010 at a meteorological station close to the park.

The next chapter describes the methodology used to understand the patterns in the association between the variables used to develop the predictive models. The results obtained after validating the wildfire prediction models with meteorological data are given in the "Results" chapter. The conclusions and a discussion of possible further studies using these models can be found in the last chapter.

Materials and methods

Data collection and preparation

The reports used for this study were based on data for the Carolina, MA, meteorological station (coordinates 7°19′48″S, 47°27′36″W, code 82765) and were supplied by the National Institute of Meteorology (INMET). This is the closest station to CMNP, which is 34.41 km away.

The data are for the period January 1st 2010 to November 27th 2010 and cover three daily measurements of rainfall (millimeters accumulated in the last 72 h), maximum temperature and minimum temperature (in degrees Celsius), sunlight (hours per day), relative humidity (%) and wind speed (meters per second) for the periods midnight to noon, noon to 6 pm and 6 pm to midnight (INMET 2012). A new variable (days without rain) was also created. Table 1 lists the meteorological dataset.

Neural networks have high learning ability and great power of generalization. A model to predict landslides induced by heavy rains in Rio de Janeiro, which is a more complex phenomenon, was developed using a history of only 4 years of data (Souza and Ebecken 2012). Another model to locate mass movements induced by earthquakes in Sichuan (China), considered an interval smaller than 4 months to train the neural network (Souza 2014).

This data was complemented with details of wildfires in CMNP based on satellite images generated by the National Institute for Space Research (INPE) for the period January 1st 2010 to November 27th 2010. Figure 2 illustrates 545 wildfires spot occurred within this period. Table 2 illustrates the fire dataset in two coordinate systems (GMS and decimal). The GMT system presents the coordinates according to Greenwich Time, but in this study, we considered Brasília (Brazil) time zone (GMT −2 h). The GMS system presents the coordinates in degrees, minutes, and seconds. Table 2 also lists wildfire date and time, as well as the satellite that detected it.

Wildfire detection was performed through polar orbit and geostationary satellites. For detection, the polar orbit

Table 1 Meteorological dataset

Date	Interval	Rainfall (mm in the last 72 h)	Temp_max (Celsius degrees)	Temp_min (Celsius degrees)	Sunlight (h)	Humidity (%)	Wind speed (m/s)	Rain-free days
13/02/2010	0_to_12 h	8.2	34.5	22.9	9.1	80	2.5	0
13/02/2010	12_to_18 h	8.2	34.5	22.9	9.1	70	2.5	0
13/02/2010	18_to_0 h	8.2	34.5	22.9	9.1	71	2.8	0
14/02/2010	0_to_12 h	7.6	33.5	23.9	4.5	84	1.3	0
14/02/2010	12_to_18 h	7.6	33.5	23.9	4.5	87	0.5	0
13/02/2010	0_to_12 h	8.2	34.5	22.9	9.1	80	2.5	0
13/02/2010	12_to_18 h	8.2	34.5	22.9	9.1	70	2.5	0
13/02/2010	18_to_0 h	8.2	34.5	22.9	9.1	71	2.8	0
14/02/2010	0_to_12 h	7.6	33.5	23.9	4.5	84	1.3	0

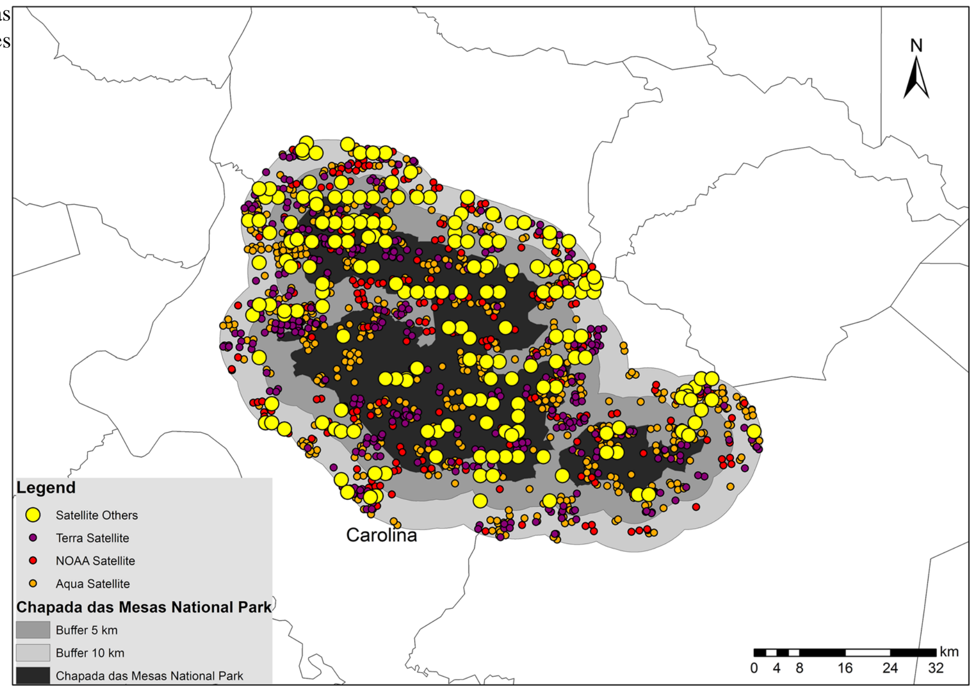


Fig. 2 Satellite image of Chapada das Mesas National Park showing wildfires between January 1st 2010 and November 27th 2010. (INPE 2012, http://www.dpi.inpe.br)—*map scale* 1:2,500,000

satellites (NOAA, located at a distance of 800 km from the Earth's surface, and TERRA and AQUA, both located at approximately 704 km) require wildfire to be a minimum of 30 m long and 1 m wide. In the case of geostationary satellites, which are located at a distance of 25,000 km, wildfire should be twice the size as that required by the polar orbit satellites.

A matrix with 993 lines or records was built, where each line corresponded to a set of measurements of meteorological parameters. The last column in the matrix contained the variable *wildfire* and could take on two values: "wildfire active" and "wildfire not active" according to the information in the wildfire database for the time at which that particular set of meteorological readings was taken. In all, 115 of the 993 records contained the value "wildfire active", and 878 "wildfire not active". The predictive model was trained with two patterns so that in a real situation it could decide between the presence or absence of wildfires.

The variable *rain-free days* was also calculated and added to the database. The value was calculated by working out the number of days when rainfall was equal to zero before a wildfire was observed. Whenever there was any rainfall, this variable was set to zero.

In this research, the following software was used:

- Multivariate analyses (Statistica, developed by StatSoft);
- Artificial neural network models (Statistica Neural Network or SNN, Statistica module);

Table 2 Wildfires dataset

Nr	Lat (Dec)	Long (Dec)	Lat_GMS	Long_GMS	Date-hour, local GMT	Satellite
1	−71.817	−472.000	S 07 10 54	O 47 12 00	13/02/2010, 16:53	AQUA-T
2	−71.137	−473.167	S 07 06 49	O 47 19 00	04/03/2010, 17:00	GOES-12
3	−70.517	−471.183	S 07 03 06	O 47 07 05	15/03/2010, 19:40	NOAA-15
4	−70.533	−471.267	S 07 03 11	O 47 07 36	15/03/2010, 19:40	NOAA-15
5	−70.580	−471.350	S 07 03 28	O 47 08 06	30/03/2010, 16:30	AQUA_M-T
6	−70.610	−471.020	S 07 03 39	O 47 06 07	15/04/2010, 16:30	AQUA_M-T
7	−72.880	−471.290	S 07 17 16	O 47 07 44	20/04/2010, 16:48	AQUA_M-T
8	−71.240	−470.770	S 07 07 26	O 47 04 37	20/04/2010, 16:48	AQUA_M-T
9	−69.800	−471.970	S 06 58 48	O 47 11 49	22/04/2010, 13:41	TERRA_M-T
10	−70.650	−470.767	S 07 03 54	O 47 04 36	22/04/2010, 16:29	AQUA-T

- Classification rules (Rule-based classification or CBA, Singapore University).

Multivariate analysis

Simultaneous analysis of a large number of variables in data mining is a technique involving statistics, databases, data visualization techniques and high-performance computing. Interaction between disciplines is essential to the process of discovering knowledge in databases (Souza 2004).

Four data clustering techniques were used in this study: correlation matrix (CM) clustering, dendrogram clustering, K-means clustering and principal component analysis. Data clustering techniques allow the relationships between the variables associated with the phenomenon being studied to be seen clearly. Elucidation of patterns in the behavior of variables is helpful when constructing a predictive model.

Linear correlation matrix

To determine the degree of linear relationship between the variables, the linear correlation matrix was built. The values in this matrix express the correlation between two variables and can vary from −1 to +1. Values of −1 and +1 represent a perfect negative and positive correlation, respectively, while 0 represents an absence of any correlation.

According to the linear correlation matrix shown in Table 3, all the meteorological variables have a poor correlation with the variable *wildfire*. For example, the highest value is −0.25, for the variable *humidity*, indicating an inversely proportional relationship between these variables. This poor linear correlation means that it is not possible to forecast wildfires with linear models. This study therefore proposes the use of non-linear models. The greatest linear relationship in Table 3 is −0.81, indicating a strong, inversely proportional relationship between the variables *maximum temperature* and *humidity*.

Table 3 Linear correlation matrix

	a1	a2	a3	a4	a5	a6	a7	a8
a1	1.00							
a2	−0.47	1.00						
a3	−0.05	0.07	1.00					
a4	−0.12	0.29	0.49	1.00				
a5	0.39	*−0.81*	0.00	−0.24	1.00			
a6	−0.12	0.34	−0.02	0.10	−0.58	1.00		
a7	−0.22	0.65	0.07	0.18	−0.75	0.46	1.00	
a8	−0.07	0.22	0.07	0.08	*−0.25*	0.21	0.22	1.00

a1 Rainfall, *a2* maximum temperature, *a3* minimum temperature, *a4* sunlight, *a5* humidity, *a6* average wind speed, *a7* rain-free days, *a8* wildfires

Dendrogram

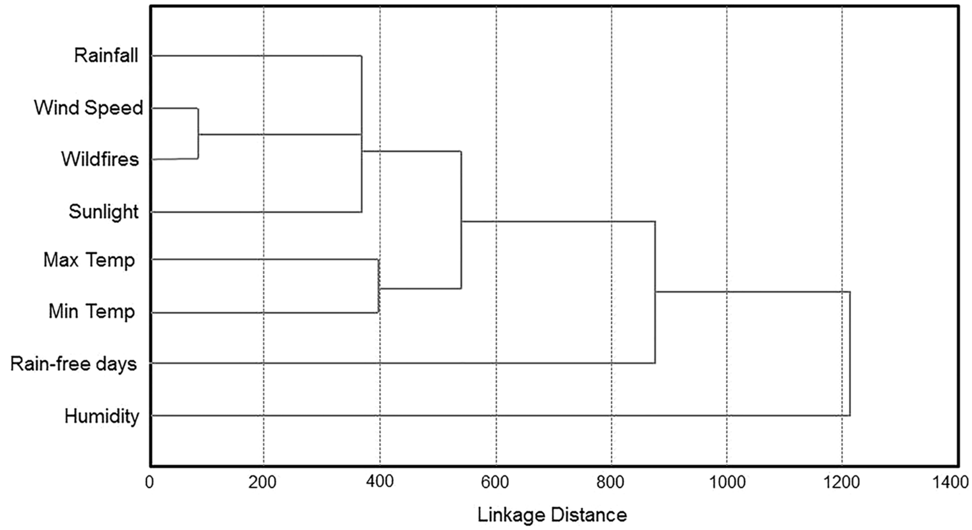
A dendrogram groups similar objects together based on some mathematical distance between the variables being studied. Here we used the Euclidian distance. After the Euclidian distance has been calculated, an algorithm groups the objects together based on their geometric distance using a single-linkage rule, where proximity is calculated based on the closest object (Han and Kamber 2001). Figure 3 shows a dendrogram with the variables involved.

K-means method

One of the statistical methods used for data clustering was the K-means technique, which uses a simple algorithm to divide the objects into clusters so that the sum of the distances from the center of a cluster to each object in the cluster is as small as possible.

In this method k centers are determined randomly, where k is the number of clusters defined by the user. For each of these centers, regions with specific characteristics are generated and similar objects are assigned to these. After this has been done, the positions of the centroids are recalculated and if they change, a new cluster is formed

Fig. 3 Dendrogram



until the positions of the centroids remain the same. Once each position has been defined, the results, i.e., the clusters with their variables and respective distances, are obtained (SANTANA 2008). Using a value of $k = 3$ yielded the results in Table 4.

As in the study by Moori et al. (2002), variables with some kind of physical relationship tended to be closer to each other. *Maximum temperature, minimum temperature* and *rain-free days* clustered together in the first cluster. The variable *humidity* was isolated in the second cluster, as in the dendrogram (Fig. 3). The variable of greatest interest, *wildfires*, clustered with the variables *rainfall*, *sunlight* and *wind speed*. These results agree with those obtained using dendrogram clustering and reinforce the hypothesis that sunlight, the absence of rain and wind speed affect the incidence and spread of wildfire, as proposed by Soares and Paez (1973).

Principal component analysis

This method involves mapping the coordinates of the variables onto a multidimensional coordinate system (Wherry 1984). The original components generate orthogonal principal components through linear combinations.

Table 4 Clusters and their members obtained using the K-means method

Cluster 1		Cluster 2		Cluster 3	
Variables	Distance	Variables	Distance	Variables	Distance
a2	12.66951	a5	0	a1	9.115142
a3	10.01148			a4	9.20507
a7	19.10548			a6	4.139091
				a8	4.988221

a1 Rainfall, *a2* maximum temperature, *a3* minimum temperature, *a4* sunlight, *a5* humidity, *a6* average wind speed, *a7* rain-free days, *a8* wildfires

Each principal component generated has more information than the original component. This method allows the dimensionality of the representative points to be reduced and, normally, more than 90 % of the information to be observed with only two or three principal components (Moita Neto and Moita 1998).

Corroborating the results reported by Moita Neto and Moita (1998), Fig. 4 suggests that low rainfall, high temperatures and wind speed contribute to the incidence and spread of wildfires.

Modeling

Two modeling techniques were used to build the predictive models: artificial neural networks (ANNs) and classification rules (CRs). In this work, the following software were used:

- Multivariate analyses (STATISTICA—developed by StatSoft);
- Artificial neural network models (STATISTICA Neural Network—SNN—STATISTICA module);
- Classification rules (Classification based on rules—CBA—Singapore University).

Artificial neural networks (ANNs)

ANN techniques are based on connectionist systems that simulate biological neural networks and process large quantities of interconnected cells in parallel. These cells correspond to elements or units, the connections correspond to links or synapses and the processing of information is done by neural networks, in allusion to the attempt to imitate the basic characteristics and architecture of brain cells (de Medeiros 1999).

A model for forecasting wildfires was built with ANN techniques using multilayer perceptron (MLP)

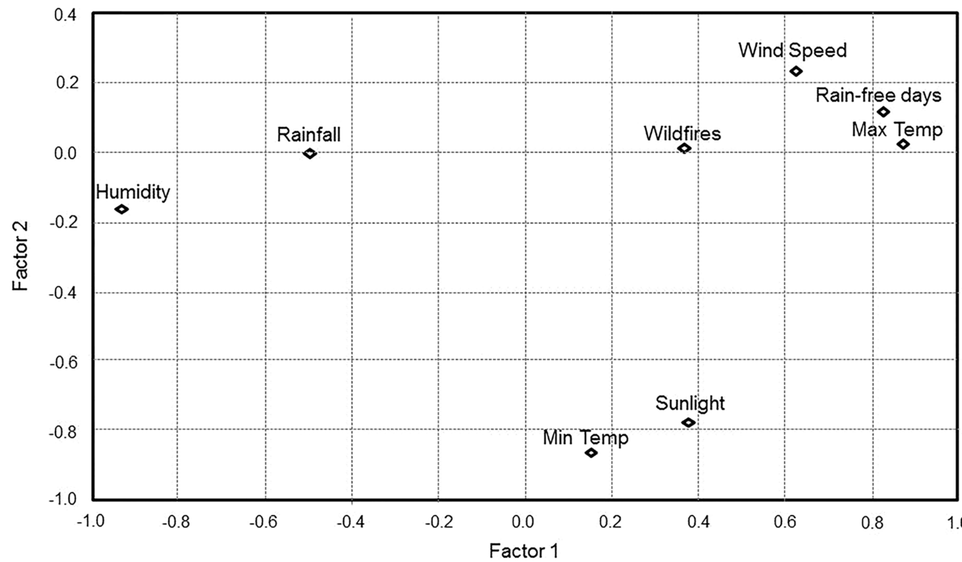


Fig. 4 Principal component analysis

architecture, which consists basically of an input layer made up of input nodes, one or more hidden layers and an output layer made up of computational nodes (Haykin 2001). Backpropagation was used as the training algorithm.

The inputs to the model were *rainfall*, *maximum temperature*, *minimum temperature*, *sunlight*, *relative humidity* and *wind speed*, and the output was *wildfire*.

Classification rules (CRs)

CRs are obtained from a system whose purpose is to identify correlations between variables in a database. The system is divided into two parts: a rule generator based on the algorithm for finding association rules for the variables and a classifier builder, which uses algorithms to build classification rules.

The first algorithm generates all the frequent rule items by making multiple passes over the data. A *rule item* can be an interval for a particular variable calculated to discretize this variable into statistical quartiles. For example, for *relative humidity* values between 70 and 81.5 %, the rule item would be *70_ < _Humidity_ < _81_5*. In the first pass, the algorithm counts the support of each rule item and determines whether it is frequent or not. In each subsequent pass, seeds of rule items found to be frequent in the previous pass are produced. The algorithm uses this seed to generate new, possibly frequent rule items, called candidate rule items.

The actual supports for these candidate rule items are calculated during the pass over the data. At the end of the pass the algorithm determines which of the candidate rule items are actually frequent. From this set of frequent rule items, the classification rules are then produced by the algorithms used by the classifier, and the rule item with the highest confidence is chosen as the possible rule. The final set of classification rules consists of all possible accurate rules (Liu et al. 1998).

Results and discussion

The main results from a classifier are the confusion matrix and correct classification rate (CCR) or accuracy, Souza and Ebecken (2012).

Table 5 shows the confusion matrix after validation of the ANN model. This validation sample is made up of 10 % of all the records, the other 80 % of the records being used for training and 10 % for testing.

The STATISTICA Neural Network software uses five different methodologies, and chooses the ANN model with better performance. The results were obtained with an ANN of the MLP type, with 3 layers and 13 neurons in the hidden layer.

The model classifies the classes "wildfire not active" and "wildfire active" correctly for 76.09 and 8.70 % of the total number of estimates, respectively. The CCR or accuracy of the ANN on the validation sample is therefore 84.79 % (the sum of the two percentages corresponding to the correct classifications) and the ANN precision on the validation sample is 40.01 %. The ANN accuracy on the calibration sample is 78.72 % and the ANN precision on the calibration sample is 33.33 %.

The CR accuracy on the validation sample is equal to 83.71 % and the CR precision is 66.06 %. The CR precision is much better than the ANN because this metric is directly proportional to the sample size. The approach of

Table 5 Confusion matrix for the ANN and CRs on validation sample

	Wildfire notactive	Wildfire active		Wildfire notactive	Wildfire active
Artificial neural network (ANN)			Classification rules (CRs)		
Wildfire not active (%)	76.09	2.17	Wildfire not active (%)	76.06	12.39
Wildfire active (%)	13.04	8.70	Wildfire active (%)	3.93	7.65
Total classification correct rate (ANN) = 84.79 %			Total classification correct rate (CRs) = 83.71 %		

the CR model is different from the ANN because the prediction is carried out by using the generated rules and validating the set of rules in the whole dataset. In order to compare both models, the values are expressed in percentage (the sample sizes are different for the two different models). Therefore, there is no accuracy or precision values for the CR calibration model.

When the sensitivity of the model is analyzed, the variables are organized in order of mathematical importance. The ranking describes the variables *sunlight*, *rain-free days* and *relative humidity* as having the greatest influence on the incidence of wildfires, corroborating the findings of a study that used the Monte Alegre formula (Soares and Paez 1973).

The model used for rule classification was developed with 40 rules (19 of them with 100 % confidence: 5 rules of "active wildfire" and 14 of "non-active wildfire").

The model classifies the classes "wildfire not active" and "wildfire active" correctly for 76.06 and 7.65 % of the total number of estimates, respectively. The CCR for the CR model is 83.71 %.

The rules used to build the predictive classifier have two parts: an antecedent, IF, and a consequent, THEN (or cause and effect). Two quantitative parameters are of particular interest when interpreting a rule: support and confidence. An example of a classification rule used to build the forecasting model is shown below (rule 22).

Rule 22:

IF 70_ < _Humidity_ < _81_5 = Y

THEN − > Class = wildfire_not_active

(24.371 % 99.174 % 242240)

Rule 22 shows that for situations where *relative humidity* is between 70 and 81.5 % wildfires do not occur in 99.174 % of cases. This rule has a confidence of 99.174 % as there were no wildfires 240 of the 242 times that humidity readings were between 70 and 81.5 %. The rule support is therefore 24.371 % as this is the number of times (242) that the rule item 70_ < _*humidity*_ < _81_5 appears in the database expressed as a percentage of the total number of records (993).

Tables 4 and 5 suggest that both methodologies could be useful in forecasting wildfires in CMNP.

Conclusion and future works

CMNP has a great diversity of fauna and flora in a region with a very high incidence of wildfires. This article proposes a methodology for forecasting wildfires using ANNs and CRs.

The predictive models yielded satisfactory results and could be used as additional tools for forecasting the risk of wildfires in CMNP and other national parks.

The accuracy of the models could be improved by expanding the historical data on wildfires in the region and by building new meteorological stations close to the park, as the data were obtained from a station 34.48 km from the reserve.

Further studies may be necessary to characterize other factors affecting wildfires, such as human activity (man-made fires) and natural phenomena in the region.

It should be noted that the methodology described here is based on the acquisition of knowledge from monitoring data; however, the knowledge of specialists could usefully be incorporated to improve the quality of the forecasts. A hybrid model combining more than one data-mining technique, as proposed by Souza (2014), could also be developed.

Highlights

- Wildfires are among the main causes of destruction of forest ecosystems.
- Wildfires destroy native flora and fauna; they cause pollution and climate changes.
- Wildfires data were combined with meteorology data.
- We present a methodology for predicting wildfires some hours in advance.
- This model could be used as an additional tool for biodiversity protection.

References

Gasull VG et al (2011) Computational intelligence applied to wildfire prediction using wireless sensor networks. In: IEEE 2011 Proceedings of the International Conference on Data Communication Networking (DCNET), pp 1–8

Arruda RSV (1999) Populações Tradicionais e a proteção de recursos naturais em unidades de conservação. Ambiente e Sociedade, São Paulo, v. ano II, no 5, pp 79–93

Beckage B, Platt WJ (2003) Predicting severe wildfire years in the Florida Everglades. Front Ecol Environ 1(5):235–239

Chu PS, Yan W, Fujioka F (2002) Fire-climate relationships and long-lead seasonal wildfire prediction for Hawaii. Int J Wildland Fire 11(1):25–31. doi:10.1071/WF01040

Das A, Dutta R, Aryal J (2013) A hybrid neural network based Australian wildfire prediction: a novel approach using environmental and satellite imagery. In: 20th International Congress on Modelling and Simulation (MODSIM2013), p 169. http://ecite.utas.edu.au/86618

de Medeiros JS (1999) Bancos de dados geográficos e redes neurais artificiais: tecnologias de apoio à gestão do território. Tese (Doutorado em Geografia Física)—Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo

Goslar A (2006) Ground vegetation biomass detection for fire prediction from remote sensing data in the lowveld region (Doctoral dissertation, University of Johannesburg)

Han J, Kamber H (2001) Data mining—concepts and techniques. Morgan Kaufmann Publishers, San Francisco

Hanson HP et al (2000) The potential and promise of physics-based wildfire simulation. Environ Sci Policy 3(4):161–172

Haykin S (2001) Redes Neurais—Princípios e Prática, 2nd edn. Bookman, Porto Alegre

IBAMA (2013) Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis (2012). Plano Operativo De Prevenção e Combate aos Incêndios Florestais do Parque Nacional da Chapada das Mesas. Disponível em http://www.ibama.gov.br/phocadownload/prevfogo/plano_operativo_parna_da_chapada_das_mesas.pdf. Acesso em 02 fev. 2013

Imada A (2014) A literature review: forest management with neural network and artificial intelligence. In: Neural networks and artificial intelligence. Springer International Publishing, pp 9–21

INMET (2012) Instituto Nacional de Meteorologia, 2012. Banco de Dados Meteorológicos para Ensino e Pesquisa—BDMEP. Disponível em http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. Acesso em 15 set. 2012

INPE (2012) Instituto Nacional de Pesquisas Espaciais, 2012. Portal do Monitoramento de Queimadas e Incêndios. Disponível em http://queimadas.cptec.inpe.br. Acesso em 22 set. 2012

Liu B, Hsu W, Ma Y (1998) Integrating Classification and Association Rule Mining. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, Plenary Presentation), New York, USA

Moita Neto JM, Moita GC (1998) Uma Introdução à Análise Exploratória de Dados Multivariados. Química Nova, São Paulo, SP, vol 21, n 4, pp 467–469

Moori RG, Marcondes RC, Avila RT (2002) Análise de Agrupamentos como Instrumento de Apoio à Melhoria da Qualidade dos Serviços aos Clientes. RAC. Revista de Administração Contemporânea, Curitiba, PR, vol 6, pp 63–82

Peng RD, Schoenberg FP, Woods JA (2005) A space–time conditional intensity model for evaluating a wildfire hazard index. J Am Stat Assoc 100:26–35

Santana RS (2008) Uma aplicação de CBIR à análise de imagens médicas de imuno-histoquímica utilizando Morfologia Matemática e espectro de padrões. Tese (Conclusão de curso em Engenharia da Computação)—Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife

Soares RV, Paez G (1973) Uma Nova Fórmula Para A Determinação do Grau de Perigo de Incêndios Florestais Na Regiao Centro-Paranaense. FLORESTA, Curitiba, PR, vol 4, n.3, pp 15–25

Souza FT (2004) Predição de Escorregamentos das Encostas do Município do Rio de Janeiro Através de Técnicas de Mineração de Dados. Tese (Doutorado em Engenharia Civil)—Universidade Federal do Rio de Janeiro, Rio de Janeiro

Souza FT (2014) A data-based model to locate mass movements triggered by seismic events in Sichuan, China. Environ Monit Assess 186(1):575–587. doi:10.1007/s10661-013-3400-3

Souza FT, Ebecken N (2012) A data based model to predict landslide induced by Rainfall in Rio de Janeiro City. Geotech Geol Eng 30:85–94

Sun R et al (2009) The importance of fire–atmosphere coupling and boundary-layer turbulence to wildfire spread. Int J Wildland Fire 18(1):50–60

Taylor SW et al (2013) Wildfire prediction to inform fire management: statistical science challenges. Stat Sci 28(4):586–615

Wherry RJ (1984) Contributions to correlational analysis. Academic Press, New York