

## Methodology for determining classes of forest fire risk using the modified Monte Alegre Formula

Metodologia para determinação das classes de risco de incêndio florestal usando a Fórmula de Monte Alegre modificada

Fernando Coelho Eugenio<sup>I</sup>, Alexandre Rosa Santos<sup>II</sup>,  
Beatriz Duguy Pedra<sup>III</sup>, José Eduardo Macedo Pezzopane<sup>IV</sup>,  
Cássio Carlette Thiengo<sup>V</sup>, Nathália Suemi Saito<sup>VI</sup>

### Abstract

The objective of this study was to calculate the Modified Monte Alegre Formula (FMA<sup>+</sup>) and adjust the wildfire risk classes of areas of forests planted in the northern-central coast of the state of Espírito Santo and the southern coast of Bahia state. The methodology used included six stages: risk calculation; spreadsheet development according to the occurrence, or not, of forest fires; spreadsheet development according to the wildfire season in the study region; risk class definition; result analysis of the determined classes; best fit selection. It was observed that the class definition methodology for the FMA<sup>+</sup> system obtained excellent results. It increased by 19.3, 21.53, and 31.3% for the percentage of success for subzones 1, 2, and 3, respectively, which implies this an important study for the successful implementation of FMA<sup>+</sup> in other areas.

**Keywords:** FMA<sup>+</sup>; Wildfire; Statistics

### Resumo

O objetivo deste estudo foi calcular a Fórmula Modificada de Monte Alegre (FMA<sup>+</sup>) e ajustar as classes de risco de incêndio em áreas de florestas plantadas na costa norte-central do estado do Espírito Santo e na costa sul do estado da Bahia. A metodologia utilizada incluiu seis etapas: cálculo de risco; desenvolvimento de planilhas de acordo com a ocorrência, ou não, de incêndios florestais; desenvolvimento de planilhas de acordo com a estação de incêndios florestais na região estudada; definição de classe de risco; análise de resultados das classes determinadas; seleção de melhor ajuste. Observou-se que a metodologia de definição de classe para o sistema FMA<sup>+</sup> obteve excelentes resultados. Aumentou em 19,3; 21,53 e 31,3% para o percentual de sucesso nas subzonas 1, 2 e 3, respectivamente, o que traduz em resultados este importante estudo para a implementação bem-sucedida da FMA<sup>+</sup> em outras áreas.

**Palavras-chave:** FMA<sup>+</sup>; Incêndios florestais; Estatística

<sup>I</sup> Engenheiro Florestal, Dr., Professor do Programa de Pós-graduação em Engenharia Florestal, Universidade Federal de Santa Maria, Cidade Universitária, Av. Roraima, 1000, CEP 97105-900, Santa Maria (RS), Brasil. fernando.eugenio@ufsm.br (ORCID: 0000-0002-1148-1167)

<sup>II</sup> Agrônomo, Dr., Professor do Programa de Pós-Graduação em Ciências Florestais, Universidade Federal do Espírito Santo, Av. Governador Lindemberg, 316, CEP 29550-000, Jerônimo Monteiro (ES), Brasil. mundogeomatica@yahoo.com.br (ORCID: 0000-0003-2617-9451)

<sup>III</sup> Bióloga, Professora agregada da Universidade de Barcelona, Diagonal, 643, CEP 08028, Barcelona, Espanha. bduguy@ub.edu (ORCID: 0000-0002-2903-1981)

<sup>IV</sup> Engenheiro Florestal, Dr., Professor do Programa de Pós-Graduação em Ciências Florestais, Universidade Federal do Espírito Santo, Av. Governador Lindemberg, 316, CEP 29550-000, Jerônimo Monteiro (ES), Brasil. pezzopane@pq.cnpq.br (ORCID: 0000-0003-0024-4016)

<sup>V</sup> Engenheiro Florestal, Mestrando do Programa de Pós-Graduação Solos e Nutrição de Plantas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Avenida Pádua Dias, 11, CEP 13418-900, Piracicaba (SP), Brasil. cassiocarlette@hotmail.com (ORCID: 0000-0002-3564-7440)

<sup>VI</sup> Engenheira Florestal, Dr., Pesquisadora Autônoma, Av. Governador Lindemberg, 316, CEP 29550-000, Jerônimo Monteiro (ES), Brasil. nssaito@gmail.com (ORCID: 0000-0002-3329-7870)



## Introduction

Forest fires are the result of complex interactions among vegetation, climate, topography, and anthropogenic activities over time. However, local climatic conditions have a direct influence on their occurrence and spread, given that the intensity of a fire and the speed by which it advances are directly linked to relative air humidity, air temperature, and wind speed because these affect the moisture content of the fuel; the amount of biomass, which is the main controller of principal fire characteristics; and the vegetation type (CHANG *et al.*, 2015).

This relationship between meteorological variables (air temperature, precipitation, relative humidity, wind speed, etc.) and forest fires (occurrence, propagation, burned area, etc.) has always been the object of research conducted by many researchers throughout the world. However, the study of climatic variables to understand forest fires is not recent. According to Eugenio *et al.* (2019), during the last several decades this relationship has been studied by several researchers, among which are the notable works of Flannigan and Harrington (1988); Viegas, Viegas and Ferreira (1992); Vázquez and Moreno (1993); Viegas and Viegas (1994); Piñol. Terradas anLloret (1998), Skinner *et al.* (1999); Diaz-Delgado (2000); Vázquez and Moreno (2001); Viegas *et al.* (2001); Duguy (2003); Viegas *et al.* (2004); Pereira *et al.* (2005); Krawchuket *et al.* (2009); Pausas and Keeley (2009); Bediaet *et al.* (2012); Duguyet *et al.* (2013); and San-Miguel-Ayanz, Moreno and Camia (2013).

Therefore, because of their abilities to provide quantitative estimates of the possibility of forest fire occurrence, hazard indices based on meteorological data have become important tools to evaluate the potential risk of regional fires (HOLSTEN *et al.*, 2013).

In 1972, Dr. Ronaldo Viana Soares developed the first fire risk index for Brazilian conditions, the Monte Alegre Formula (FMA). From a simplistic view, it can be said that the FMA is a cumulative index that uses as meteorological variables relative air humidity and precipitation (NUNES; SOARES; BATISTA, 2006). José Renato Soares Nunes in 2005 added wind speed, a factor of great importance for prevention and fighting forest fires, thus developing the Modified Monte Alegre Formula (FMA<sup>+</sup>). It was tested and approved for use in the region of Telêmaco Borba, state of Paraná (NUNES; SOARES; BATISTA, 2006).

Since its creation, FMA<sup>+</sup> has been used by several researchers and companies in Brazil and is among the most used indices throughout the entire nation. Notable relevant studies are those of Nunes, Soares and Batista (2006; 2009), Nunes *et al.* (2010), Borges *et al.* (2011), Pereira, Batista and Soares (2012), Rodríguez *et al.* (2012), Souza, Casavecchia and Stangerlin (2012), White *et al.* (2013), Soriano, Daniel and Santos (2015), and White, White and Ribeiro (2015). However, several of these studies have a low percentage of success; the studies by White *et al.* (2013) and White, White and Ribeiro (2015) obtained 38.64 and 36%, respectively, in areas with eucalyptus plantations in the north coast of Bahia state. Borges *et al.* (2011) found values of success that ranged between 51.54 and 56.47% in the same area of the present study.

It is important to stress that to use of FMA<sup>+</sup> in the entire state of Paraná is feasible, Nunes, Soares and Batista (2009) developed an FMA<sup>+</sup> adjustment methodology, which is based on the evaluation of the index performance in several regions of the state. However, the use of this methodology in different states of Brazil has been occurring over the years without any adjustment of its classes, which has led to the difficulty of applying the methodology proposed by Nunes, Soares and Batista (2009). This has led to low percentages of success in several studies using FMA and; therefore, the use of other indices in areas outside the state of Paraná by companies in the forest sector has been common.

Based on this assumption and considering the need to adjust the fire risk classes provided by the FMA<sup>+</sup> for any area other than that in which it was developed, the objective of this study was to calculate the FMA<sup>+</sup> and adjust the wildfire risk classes to areas of forests planted in the

northern-central coast of the state of Espírito Santo and the southern coast of Bahia state.

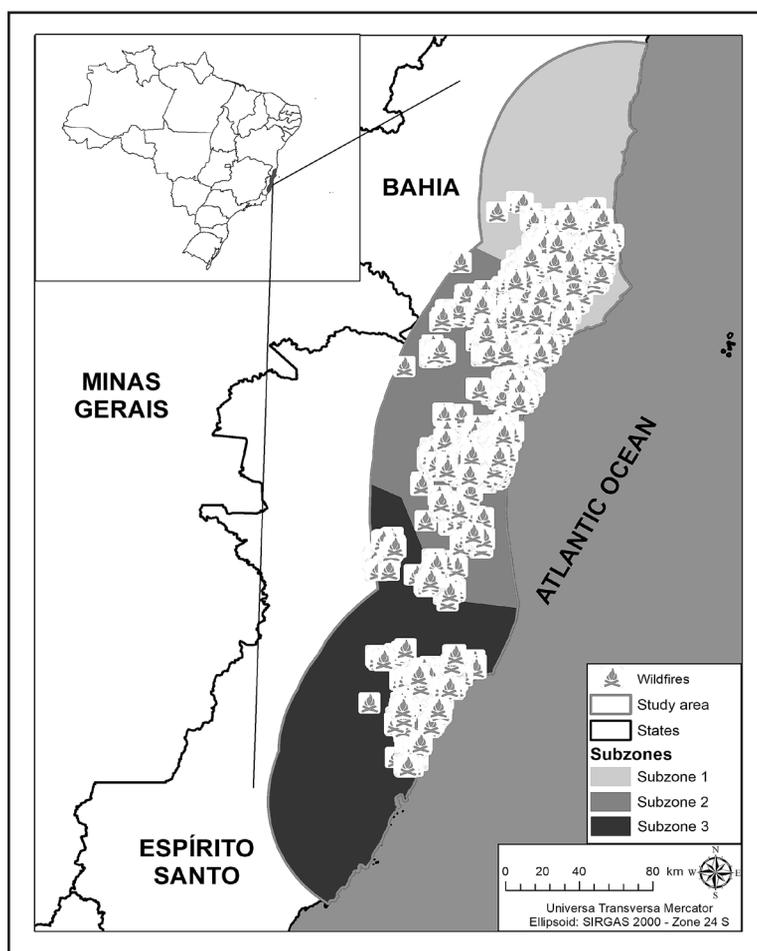
## Materials and methods

### Study area

The study area extends from the northern-central coast of the state of Espírito Santo to the southern coast of Bahia state, as described by Eugenio (2019) for the delimitation of the area that was used a buffer of 70 km from the coast, because, within this buffer, is the region that has the largest number of planted eucalyptus forests and all of the meteorological stations of the forestry companies. The study area was divided into three climatic subzones delimited according by Eugenio (2019). The occurrences of forest fires and their climatic subzones are shown in Figure 1.

**Figure 1 – Study area and its climatic delimitations for forest fires**

Figura 1 – Área de estudo e a delimitação climática para os incêndios florestais



Source: Eugenio (2019)

For a better understanding of the methodology used, it was divided into six stages as follows: stage 1 – risk calculation; stage 2 – spreadsheet development according to the occurrence,

or not, of forest fires; stage 3 –spreadsheet development according to the wildfire season in the study region; stage 4 – risk class definition; stage 5 – result analysis of the determined classes; stage 6 – best fit selection.

### Stage 1 – Risk Calculation

The FMA<sup>+</sup> was used for the daily calculation of the risk of occurrence of forest fires, according to equation 1, conceived by Nunes, Soares and Batista (2009), as follows in Equation (1):

$$FMA^+ = \sum_{i=1}^n \left( \frac{100}{H_i} \right) e^{0,04V} \tag{1}$$

in which:

FMA<sup>+</sup>: Monte Alegre Altered Formula;

H: air relative humidity (%) measured at 13 o'clock;

n: number of days without rainfall greater than or equal to 13.0 mm;

V: wind speed in m/s, measured at 13 o'clock.

Because FMA<sup>+</sup> is a cumulative index, regarding relative humidity, it is subject to precipitation restrictions (Table 1).

**Table 1 – FMA<sup>+</sup> restrictions according to the amount of rain in the day and its modifications due to the air relative humidity (H) at 1 p.m**

Tabela 1 – Restrições da FMA<sup>+</sup> de acordo com a quantidade de chuva no dia e suas modificações devido à umidade relativa do ar (H) às 13h

Rain in the day (mm)	Modifications
< 2.5	None
2.5 + 5.0	Decrease 30% in the value of FMA + calculated on the day before and add up (100 / H) of the day
5.0 + 10	Decrease 60% in the value of FMA + calculated on the day before and add (100 / H) of the day
10 + 13	Decrease 80% in FMA + calculated on the day before and add up (100 / H) of the day
≥ 13	Stop calculation (FMA + = 0), starting the next day.

Source: Nunes, Soares and Batista (2006), adapted by the author

### Stage 2 – Spreadsheet development according to the occurrence, or not, of forest fires

The area of coverage of each meteorological station was delimited in the work of Eugenio (2019). Given this, it was possible to identify the dates of wildlife occurrence during each season. From the database of each meteorological station, separation of the days with and without forest fires was completed to obtain two new spreadsheets for each: those with the FMA<sup>+</sup> values of the days with and without wildfire occurrence.

It is important to note that the data of with or without wildlife occurrence originates from a forestry company spreadsheet that contains the date, time, area, cause, and geographical

coordinates of each forest fire occurrence within the study area during all the years studied; therefore, they are the true field data record of forest fires.

### Stage 3 – Spreadsheet development according to the wildfire season in the study region

Eugenio (2019) described that the study area has three climatic subzones (subzones 1, 2, and 3). The author also states that each subzone has different wildfire occurrence times as follows: for subzone 1, season 1: December to March and season 2: August to October; for subzone 2, season 1: January to March and season 2: August to October; and for subzone 3, season 1: January and February and season 2: August to October.

Taking advantage of the spreadsheets generated during stage 2, the subzones were also separated by the occurrence time to calculate the accuracy percentage and skill score test. This culminated in the generation of nine spreadsheets, arranged as follows:

- Spreadsheet 1:* subzone 1 + days of season 1 WITH fires + days of season 1 WITHOUT fires;
- Spreadsheet 2:* subzone 1 + days of season 2 WITH fires + days of season 1 season 2 WITHOUT fires;
- Spreadsheet 3:* subzone 1 + total days WITH fires + total days WITHOUT fires;
- Spreadsheet 4:* subzone 2 + days of season 1 WITH fires + days of season 1 WITHOUT fires;
- Spreadsheet 5:* subzone 2 + days of season 2 WITH fires + days of season 2 WITHOUT fires;
- Spreadsheet 6:* subzone 2 + total days WITH fires + total days WITHOUT fires;
- Spreadsheet 7:* subzone 3 + days of season 1 WITH fires + days of season 1 WITHOUT fires;
- Spreadsheet 8:* subzone 3 + days of season 2 WITH fires + days of season 2 WITHOUT fires;
- Spreadsheet 9:* subzone 3 + total days WITH fires + total days WITHOUT fires.

### Stage 4 – Risk class definition

The first class methodology used was that conceived by the creator of the model (NUNES; SOARES; BATISTA, 2006) as follows: Null= $\leq 3.0$ ; Small= $3.1-8.0$ ; Average= $8.1-14.0$ ; High= $14.1-24.0$ ; and Very High= $\geq 24.0$ .

With the predefined number of classes, the stage for defining the limits of classes of the FMA<sup>+</sup> followed the subzones of the present study, there were two types of analyses conducted as follows: definition by percentiles and logistic regression.

The analysis based on the definition by percentiles was obtaining the FMA<sup>+</sup> values in the percentiles of 20, 40, 60, 80, and 90 of the days during which forest fires occurred. The class values derived from the percentile analysis were described by the identifiers (id.) A, C, E, G, I, K, M, O, and Q.

The class definition by means of logistic regression was performed based on the regression model described by Hosmer and Lemeshow (2000) as presented in Equation (2):

$$P(y_1 = 1) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2)$$

in which:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i$$

$P(y_1 = 1)$ : probability of occurrence of wildfire;

$x_i$ : independent variable (FWI value);

$\beta_i$ : coefficients expected by the maximum likelihood method.

The following probability values were used to determine the risk classes: 0.20, 0.40, 0.60,

0.80, and 0.90. The class values derived from the logistic regression were described by the ids B, D, F, H, J, L, N, P, and R.

### Stage 5 – Results analysis of the determined classes

After performing the calculations and determining the values for each class, the fire risk behavior for the study area was obtained using the different methodologies.

To analyze the risk behavior based on the value of FMA<sup>+</sup> in the different classes, a cross-reference was conducted of the calculated risk and the dates on which forest fires occurred, or not. Skill Score (SS) and Percent Success (PS) methods were also used in the works of Sampaio (1999), Nunes, Soares and Batista (2006), Borges *et al.* (2011), and Dimitrikoupolos, Bemmerzouk and Mitsoupoloulos (2011).

The cross-risk analysis provided an estimate of the confidence of each calculated risk because it is based on the ratio of the difference between the correctness of the prediction and the expected number of hits, as well as the difference between the number of days observed and the number of days (Tables 2 and 3).

**Table 2 – Contingency table**

Tabela 2 – Tabela de contingência

Event	Forest fires		Expected total	
	Observed	Not observed		
<b>Wildfire</b>	Expected	a	b	$N_2 = a + b$
	Not expected	c	d	$N_4 = c + d$
<b>Total observed</b>		$N_1 = a + c$	$N_3 = b + d$	$N = a+b+c+d$

Source: Sampaio (1999), adapted by the author

**Table 3 – Calculations of the contingency table**

Tabela 3 – Cálculos da tabela de contingência

Event	Forest fires		Expected total	
	Observed	Not observed		
<b>Wildfire</b>	Expected	$a / (a+c)$	$b / (b+d)$	1
	Not expected	$c / (a+c)$	$d / (b+d)$	1
<b>Total observed</b>		1	1	2

Source: Sampaio (1999), adapted by the author.

The equations for performing the calculations were, in Equations (3), (4), (5), (6) and (7):

$$N = a + b + c + d \tag{3}$$

in which:

N: total number of observations;

a: number of days with occurrences of expected and observed fires;

*b*: number of days with occurrences of expected fires and not observed;  
*c*: number of days with occurrences of fires not expected and observed; and  
*d*: number of days with occurrences of fires not expected and not observed.

$$G_a = a + d \quad (4)$$

in which:

$G_a$ : number of hits in the forecast.

$$H_a = N(1 - p)(1 - q) + Npq \quad (5)$$

in which:

$H_a$ : expected number of hits;

$p = N_1 / N$ : number of days with occurrences of expected and observed fires;

$q = N_2 / N$ : number of days with occurrences of fires anticipated and not observed;

$N_1$ : number of days with occurrences of fires planned and observed plus number of days with occurrences of fires not expected and observed;

$N_2$ : number of days with occurrences of fires anticipated and observed plus number of days with occurrences of fires anticipated and not observed;

$N_3$ : number of days with occurrences of fires expected and not observed plus number of fires days with occurrences of unexpected and unobserved fires;

$N_4$ : number of days with occurrences of fires not expected and observed plus number of days with occurrences of fires not expected and not observed;

$p$ : number of days with occurrences of fires expected and observed plus number of occurrences of unexpected and observed fires occurring, divided by the total number of comments; and

$q$ : number of days with occurrences of expected fires and observed plus number of days with occurrences of expected and unobserved fires, divided by the total number of observations.

$$SS = (G_a - H_a) / (N - H_a) \quad (6)$$

in which:

SS: skill score.

$$PS = (G_a / N) 100 \quad (7)$$

in which:

PS: percentage of success.

## Stage 6 - Best fit selection

Class selection is a crucial stage for the correct use of a risk index. The first selection was made according to the original model classes. If the values of success and skill score do not agree with those expected, one can choose from among the other classes tested in the ids from A to L.

The methods used in the present study consisted of the validation and choice of the presented results via the percentages of success and the skill score value for each subzone. Initially, normalization of the data of the percentage of success with and without fires was completed, in general. In addition, the skill score values were normalized within a scale ranging from 0 to 100.

After value normalization, a hypothesis test was conducted using the Shapiro–Wilk test to determine if the sample is or is not originating from a normal distribution. If  $H_0$ , the sample originates from a normal distribution and if  $H_1$ , the sample does not originate from a normal distribution.

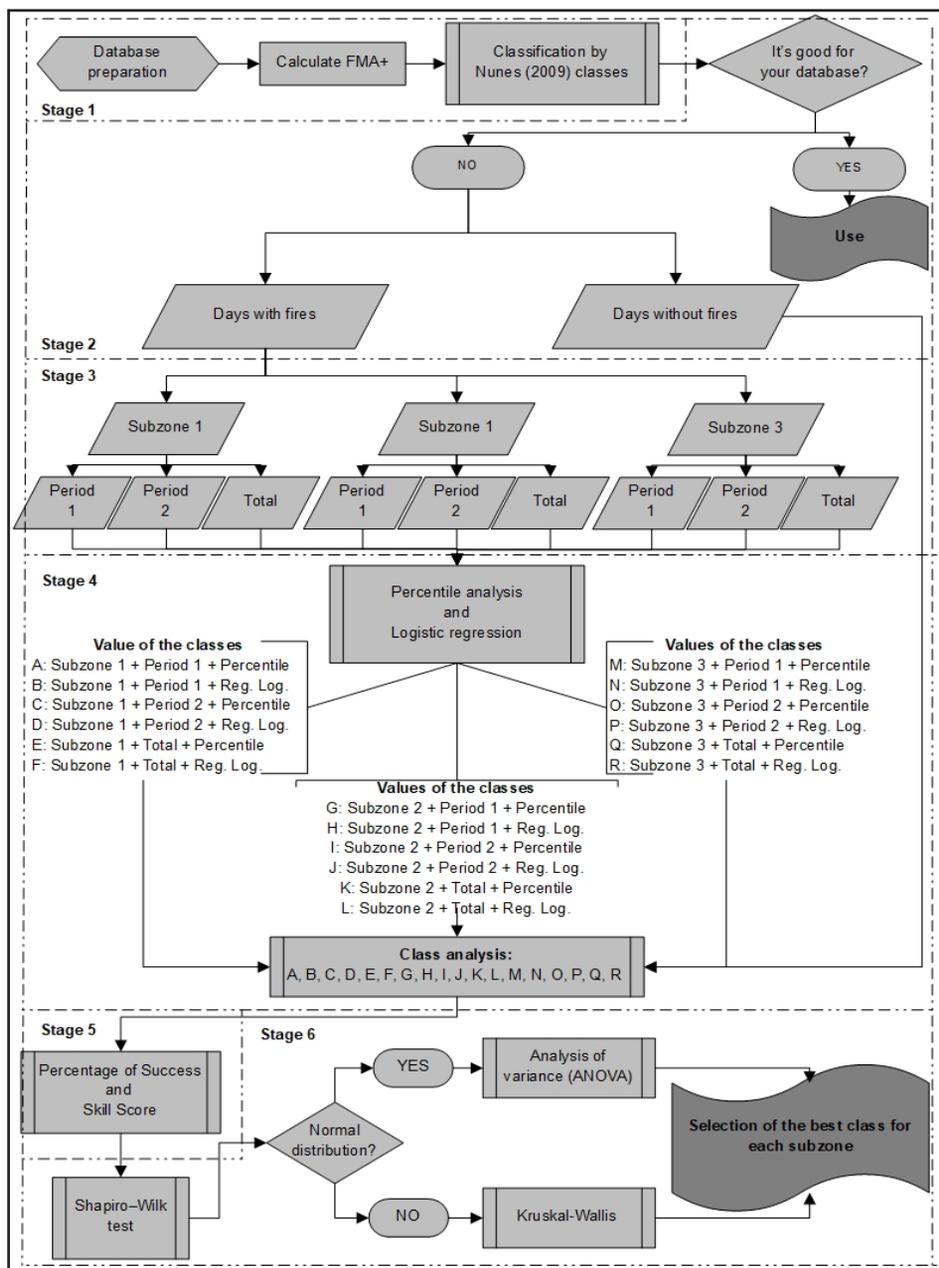
To determine if the sample originate from a normal distribution, the Analysis of Variance (ANOVA) was performed, which possesses as a hypothesis the equality between the averages of two or more populations. If the F test is significant, the Tukey–Kramer *post-hoc* test (significance level = 0.05) was completed to compare all the ids among each other. Given this, the highest average value is the chosen id.

However, if the sample did not present a normal distribution, Kruskal–Wallis nonparametric analysis of variance was used and the null hypothesis was determined is the equality between the categories; the ids in each subzone were assessed to a level of significance equal to 0.05 and 95% reliability intervals.

This methodology attempted to solve a difficult situation experienced in areas of forests planted in the northern-central coast of Espírito Santo and the southern coast of Bahia, because, of all the years used to define this methodology, in this region approximately 92% of days were without the occurrence of forest fires and only 8% of days had forest fires. In addition to this factor, as reported by Eugenio (2019), the area has a history of illegally started fires, thus reducing the relationship between the variables and their responses in the model. All six stages used to conduct the present methodology are shown in Figure 2.

**Figure 2 – Flowchart of the six steps required to perform the methodology**

Figura 2 – Fluxograma das seis etapas necessárias para executar a metodologia



Source: Authors (2019)

## Results and discussion

After the database preparation, the risks were calculated for all subzones and classified using the classes proposed by Nunes, Soares and Batista (2009). The results found for the success percentage of the days with and without fire and the general and skill score tests are presented in Table 4.

**Table 4 – Results obtained by the success percentage test and skill score test for the class values proposed by Nunes, Soares and Batista (2009)**

Tabela 4 – Resultados obtidos pelo teste de porcentagem de sucesso e teste de *skill score* para os valores de classe propostos por Nunes, Soares and Batista (2009)

Subzone	Sucesspercentage (%)			Skill score
	With wildfire	Without wildfire	Geral	
1	87.36	47.08	51.88	0.1306
2	90.24	34.98	41.11	0.0779
3	92.41	37.26	40.64	0.0545

Source: Authors (2019)

As can be observed in Table 4, there was a good correlation for the days on which fire occurred, reaching 92.41% in subzone 3; however, when analyzing the days on which fire did not, good results were not obtained, the reason for which the overall success percentage was reduced. The analysis of the values found by the skill score test reflects this reduced accuracy because the very low values - except in subzone 1. Therefore, there is a need to continue the methodology because it is believed that better classes can be determined, according to the local database, and, consequently, a better accuracy is obtained in the results of the percentage of general success and skill score.

Table 5 portrays the results obtained using the different approaches described in the methodology. Notably, all the classes originated using the percentile method have the same percentages of analysis and the classes of those originated using the logistic regression following their equation.

Analyzing Table 5, it is possible to identify a large difference between the limit values of the originating classes using the percentile method versus the classes from the logistic regression. When analyzing the limit values of the "extreme" class obtained by percentiles, the highest value was found in id G (FMA=78). This same value was surpassed by all the lower limits of the "average" class obtained via logistic regression, with its lower value in id F (FMA=86).

Therefore, the lowest limit value of the percentile classes is less than the value considered as the risk divisor of forest fires of the classes defined by the logistic regression for the FMA<sup>+</sup>. The analysis of the results found using the percentage of success and skill score test is shown in Table 4.

**Table 5 – Results found for each id followed by the percentage or equation used, with the limit FMA<sup>+</sup> values for the classes**

Tabela 5 – Resultados encontrados para cada ID seguido da porcentagem ou equação utilizada, com os valores limite de FMA<sup>+</sup> para as classes

Subzone	id	Percentage / Equation	FWI limit values for the classes				
			Low	Moderate	High	Very high	Extrem
1	A	20%. 40%. 60% and 90%	0 - 10	10.1 - 17	17.1 - 26	26.1 - 56	>56
	B	$1/[1+\exp(-2.462+0.024 \times \text{FMA}^+)]$	0 - 28	28.1 - 94	94.1 - 148	148.1 - 267	>267
	C	20%. 40%. 60% and 90%	0 - 13	13.1 - 20	20.1 - 30	30.1 - 65	>65
	D	$1/[1+\exp(-1.811+0.015 \times \text{FMA}^+)]$	0 - 40	40.1 - 89	89.1 - 129	129.1 - 219	>219
	E	20%. 40%. 60% and 90%	0 - 10	10.1 - 17	17.1 - 25	25.1 - 56	>56
	F	$1/[1+\exp(-2.180+0.020 \times \text{FMA}^+)]$	0 - 45	45.1 - 86	86.1 - 120	120.1 - 194	>194
2	G	20%. 40%. 60% and 90%	0 - 14	14.1 - 23	23.1 - 35	35.1 - 78	>78
	H	$1/[1+\exp(-2.398+0.013 \times \text{FMA}^+)]$	0 - 44	44.1 - 101	101.1 - 149	149.1 - 254	>254
	I	20%. 40%. 60% and 90%	0 - 14	14.1 - 22	22.1 - 32	32.1 - 59	>59
	J	$1/[1+\exp(-2.125+0.017 \times \text{FMA}^+)]$	0 - 126	126.1 - 323	323.1 - 485	485.1 - 843	>843
	K	20%. 40%. 60% and 90%	0 - 12	12.1 - 20	20.1 - 30	30.1 - 64	>64
	L	$1/[1+\exp(-2.018+0.005 \times \text{FMA}^+)]$	0 - 78	78.1 - 153	153.1 - 216	216.1 - 354	>354
3	M	20%. 40%. 60% and 90%	0 - 14	14.1 - 23	23.1 - 35	35.1 - 63	>63
	N	$1/[1+\exp(-3.168+0.020 \times \text{FMA}^+)]$	0 - 66	66.1 - 105	105.1 - 138	138.1 - 209	>209
	O	20%. 40%. 60% and 90%	0 - 14	14.1 - 23	23.1 - 34	34.1 - 60	>60
	P	$1/[1+\exp(-3.037+0.025 \times \text{FMA}^+)]$	0 - 88	88.1 - 154	154.1 - 207	207.1 - 327	>327
	Q	20%. 40%. 60% and 90%	0 - 14	14.1 - 22	22.1 - 31	31.1 - 59	>59
	R	$1/[1+\exp(-2.712+0.015 \times \text{FMA}^+)]$	0 - 89	89.1 - 138	138.1 - 179	179.1 - 268	>268

Source: Authors (2019)

It is possible to predict—with the limit values of the classes—that the risk analysis through logistic regression will encompass the highest number of FWI values in the low and moderate classes, as opposed to the values obtained via percentile analysis. It can be said that for the conditions of the present study, the logistic regression analysis is less sensitive to the lower values of wildfire risk and may overestimate low risk conditions, which may lead to a situation in which the qualification of the risk presented is low, but it actually is a very high risk. This fact is proven by the analysis of the results found using the percentage of success test and skill score test, both demonstrated in Table 6.

**Table 6 – Results obtained by the percentage of success test and skill score test for the**

**values of class**

Tabela 6 – Resultados obtidos pelo teste de porcentagem de sucesso e teste de *skill score* para os valores da classe

Subzone	Id.	Percentage of success (%)			Skill score
		With wild fire	With out wild fire	General	
1	A	60.29	65.30	64.35	0.1802
	B	0.91	97.92	79.61	0.0000
	C	58.31	71.17	69.18	0.1998
	D	1.81	97.52	82.75	0.0000
	E	61.39	72.50	71.18	0.1979
	F	1.54	98.66	87.09	0.0032
2	G	59.83	63.17	62.64	0.1408
	H	3.16	98.66	83.56	0.0286
	I	61.32	58.25	58.66	0.0979
	J	0.00	100.00	86.80	0.0000
	K	61.17	64.18	63.85	0.1214
	L	0.14	99.73	88.69	0.0000
3	M	58.29	73.05	71.94	0.1346
	N	0.00	99.74	92.23	0.0000
	O	59.59	64.29	63.87	0.0970
	P	0.59	99.80	90.95	0.0069
	Q	58.90	70.74	70.01	0.1023
	R	0.42	99.73	93.63	0.0028

Source: Authors (2019)

Table 6 can be analyzed in a number of ways: by the percentage of success, that is, percentage of correctness of the model for the days on which fires occurred; for the days on which fires did not occur; a general analysis encompassing every day without fires; or even by the values obtained using the skill score test.

The ids originating from the logistic regression have the highest value overall success percentages. It can be observed that the lowest value of assertiveness was obtained via logistic regression was id B which was 79.61%; it exceeds by 7 percentage points the highest value obtained using the percentile method.

Regarding the days with fire, the days with the lowest correct value for the ids referring to the percentile classes is always much higher than those of the logistic regression class, reaching up to a 55 percentage point difference - id M with 58.29% and id H with 3.16%. Notably, logistic regression ids occurred that do not correctly identify any day with fire, as in the case of ids N and J, with a 0.00% accuracy.

There is also a difference between the values determined in the skill score test when comparing the different methods for the class boundaries, the lowest percentile value being id O, with 0.0970. These values were greater than the highest value found in the logistic regression

classes, with its highest value for id H of 0.0286. There are ids with values equal to zero originating from logistic regression, as in the case of ids B, D, J, L, and N, which reinforces the importance of class delimitation according to the results found in the study area, in which it was applied using the formula to calculate the probability of wildfire risk.

This preliminary analysis is a fundamental part in understanding the behavior of the classes and their limits against a population of values of wildfire risk because, in a simplistic analysis, the percentage of overall success would lead to failure because, as previously reported, there are cases in which a high accuracy in the general percentage does not show an accuracy for those days on which fire occurred.

In the case of id J, in which 100% of the days without fire were identified and it had an overall success rate of 86.80%, 0% of days with wildfire were accurately projected and it had a skill score of 0.0000, though. If a researcher has used this model because it had a high overall percentage of correctness, he/she could also conclude that no model of fire risk would be necessary because of an assumption that there will be no fire on any day, which translates into an error of 12.60% of the days; however, the researcher would be accurate 87.40% of the time for the number of days on which there was no fire.

Nevertheless, the opposite—high accuracy on days with fire—leads to the understanding that it a wildfire will occur every day. In this case, the error rate would be approximately 90% of the days. Considering such issues, it is understood there is a need for a statistical evaluation with data standardization which is incontestable to infer which model to use.

After analyzing the results in a visual manner, statistical tests were performed to verify the existence, or not, of differences between the results obtained using the classes of different ids.

Initially, the normalization of the success percentage data with and without fire and the general and skill score values were normalized to a scale that ranges between 0 and 100. Following value normalization, a hypothesis test was performed using the Shapiro-Wilk test, in which  $H_0$  is a sample originating from a normal distribution and  $H_1$  is a sample that does not originate from a normal distribution (Table 7).

As can be observed in Table 7, the ids presented values higher than 0.05; thus, it is acceptable to have a null hypothesis.

Together with the data that presented a normal distribution for both subzones, a parametric test of ANOVA was conducted, which has as a hypothesis the equality between the averages of two or more populations; in the present case, the equality between the normalized values for each id. As the F test was significant, the Tukey-Kramer *post-hoc* test was used to compare all the ids. The test results are shown in Table 5.

The Tukey-Kramer test was performed at a significance level of 0.05 and a 95% confidence interval. For both subzones, the null hypothesis was not rejected; that is, the distribution of the normalized values is the same among the different identifiers for each subzone. Therefore, the choice of the best identifier for each subzone was based on the highest average found. As can be seen in Table 5, the highest average value for subzone 1 was id E, with an average value equal to 24.06; for subzone 2, it was chosen id G, with an average value of 23.90; and for subzone 3, it was id M with an average value equal to 25.32.

**Table 7 – Results found using the Shapiro-Wilk and Tukey-Kramer test for the different ids**

Tabela 7 – Resultados encontrados no teste *Shapiro-Wilke Tukey-Kramer* para os diferentes ids

Subzone	Season	Method	Id.	Shapiro-Wilk	Tukey-Kramer
---------	--------	--------	-----	--------------	--------------

1	1	Percentile	A	0.095	22.72
		Logistic regression	B	0.077	9.37
	2	Percentile	C	0.129	23.85
		Logistic regression	D	0.070	9.64
	Total	Percentile	E	0.064	24.36
		Logistic regression	F	0.055	8.45
2	1	Percentile	G	0.155	23.90
		Logistic regression	H	0.383	12.06
	2	Percentile	I	0.379	20.87
		Logistic regression	J	0.051	10.03
	Total	Percentile	K	0.092	22.96
		Logistic regression	L	0.054	10.05
3	1	Percentile	M	0.200	25.32
		Logistic regression	N	0.062	8.34
	2	Percentile	O	0.207	21.91
		Logistic regression	P	0.081	10.21
	Total	Percentile	Q	0.137	22.84
		Logistic regression	R	0.057	9.23

Source: Authors (2019)

When analyzing the ids, the same result originates from the percentile analysis of all the days on which forest fires occurred in subzone 1; their accuracy for the days on which forest fires occurred was 61.39%, and for the days on which forest fires did not occur, it was 72.50%. Its success percentage was equal to 71.18% and its skill score was 0.1979. For subzones 2 and 3, ids G and M were chosen, both from season 1, with 59.83 and 58.29% accuracy of days on which fire occurred; 63.17 and 73.05% accuracy on days on which fires did not occur; 62.24 and 71.94% values for success percentage; and skill score test values of 0.1408 and 0.1366, respectively.

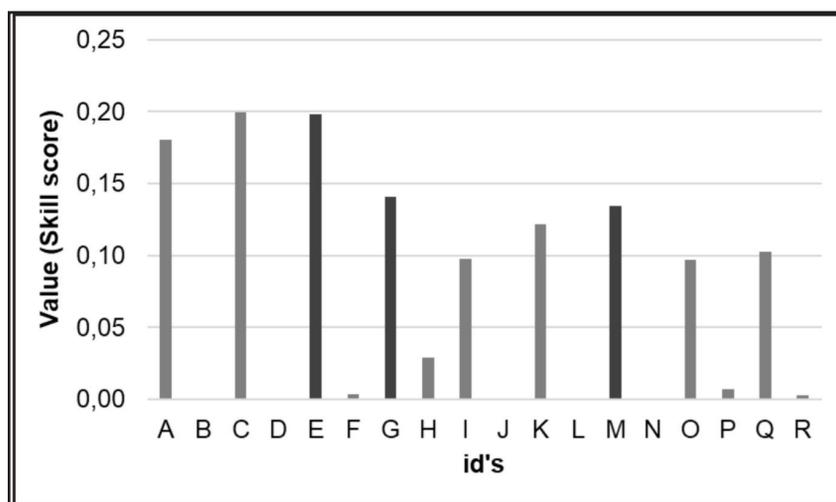
The variation in values of the skill score test, among the different identifiers tested, are shown in Figure 3.

There is a visual difference between the ids: they can be divided into two groups for each subzone. The groups derived from the percentile methodology—A, C, and E; G, I, and K; and M, O, and Q—with values higher than those found via logistic regression—and B, D, and F; H, J, and L; and N, P, and R—which have the lowest values.

In relation to the skill score test, in both subzones the results found in the present study are higher than those found by Nunes, Soares and Batista (2006) and Nunes *et al.* (2010) for the forest district of Monte Alegre in the municipality of Telêmaco Borba, state of Paraná, Brazil; a value of 0.11165 for the skill score test was obtained.

**Figure 3 – Skill score values between the different identifiers (ids) tested, highlighting the ids chosen for subzones 1, 2 and 3**

Figura 3 – Valores de *skill-score* entre os diferentes identificadores (ids) testados, destacando os ids escolhidos para as subzonas 1, 2 e 3



Source: Authors (2019)

White *et al.* (2013), when evaluating FMA<sup>+</sup> in areas of eucalyptus plantations on the northern coast of Bahia state from January 1st, 2002, to December 31st, 2009, obtained a skill score equal to 0.059. White, White and Ribeiro (2015), for the period between January 1st, 2002, and December 31st, 2012, obtained a skill score equal to 0.05. Rodríguez *et al.* (2012) evaluated the performance of wildfire risk indices for areas of the Macujire forestry company, in Cuba, during the period between January, 2006, and December, 2011, and obtained using FMA<sup>+</sup> as skill score test of 0.0737.

The results of the present study were slightly lower when compared to those found by Borges *et al.* (2011), who conducted a study to verify the performance of some fire risk indices in eucalyptus plantations in the northern region of Espírito Santo state between the years 2003 and 2004 and obtained skill score values ranging from 0.1626 to 0.2055.

The variations in the values of the percentage of success among the different identifiers tested are shown in Figure 4.

A visual difference between the identifiers is apparent: they can be grouped into two groups for each subzone. The groups derived from the percentile methodology (A, C, and E; G, I, and K; and M, O, and Q), with values higher than those found via logistic regression, (B, D, and F; H, J, and L; and N, P, and R), which have the lowest values for the percentage of global success.

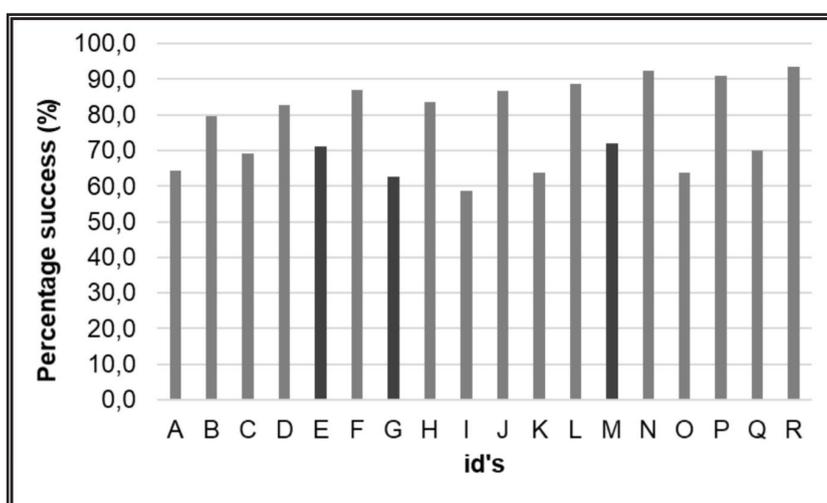
In relation to the percentage of success obtained, for both subzones the results found in the present study are superior to those found by Nunes, Soares and Batista (2006) and Nunes *et al.* (2010), which obtained a value of 55.64%. Souza (2014) obtained 63.53% in a study of the percentage of success in the municipality of Lages, in Santa Catarina state, higher only than that found in subzone 2 of the present study. Rodríguez *et al.* (2012), when working with FMA<sup>+</sup> for areas of a forest company in Cuba, obtained 57.10%, which is a lower value than that found in both subzones of the present study.

Borges *et al.* (2011) obtained values of success percentage that ranged between 51.54 and 56.47%, less than the value found in the present study of 68.45% on average.

However, for fire occurrences in the Serra de Itabaiana National Park, a value of 48% was obtained, which was less than that found in the subzones of the present study. Borges *et al.* (2011) obtained values of success percentage that ranged between 51.54 and 56.47%, less than those in the present study, which had an average of 68.45%. White *et al.* (2013) and White, White and Ribeiro (2015) obtained 38.64% and 36%, respectively, values that were much lower than those found in the present study for the percentage of FMA<sup>+</sup> success in areas with eucalyptus plantations in the northern coast of Bahia.

**Figure 4 – Percentage success values between the different identifiers (ids) tested, highlighting the chosen ids for subzone 1 (id E), for subzone 2 (id G), and for subzone 3 (id M)**

Figura 4 – Valores da porcentagem de sucesso entre os diferentes identificadores (ids) testados, destacando os ids escolhidos para a subzona 1 (id E), para a subzona 2 (id G) e para a subzona 3 (id M)



Source: Authors (2019)

The average success rate found in the present study was higher than most of the previous studies analyzed; however, to also verify the methodology described in Nunes, Soares and Batista (2009), the decreasing curve in relation to the risks was analyzed.

Nunes, Soares and Batista (2009) reported that the number of days expected in each risk class should have an inverse relationship with the risk class in such a manner that the higher the risk class, the lower the number of days expected for it.

Therefore, the analysis completed with the data obtained by means of the id classes E, G, and M for subzones 1, 2, and 3, respectively. For both subzones, there is an inverse relationship between the number of days expected and the risk classes, with the class of higher risk having the least expected number of days. The  $R^2$  values for the subzones were 0.87, 0.75, and 0.90, respectively, originating from a decreasing exponential function, meeting that proposed by the FMA<sup>+</sup> author.

As reported by Nunes *et al.* (2010), the observed mismatch in wildfire risk is a factor that interferes with its performance; this mismatch is probably caused by the change in rainfall regimes and, consequently, of relative humidity over time. Therefore, to avoid the use of an unsafe risk index, which may lead to mistaken decisions regarding wildfire prevention and control procedures, a prior adjustment is necessary.

Hence, the adjustment of the limit values of the classes performed by the presented methodology is equivalent to a substantial gain when compared to the classes of the origin of the index presented in Table 2, which has an average overall success percentage of 44.54%; it increases to 68 and 59%. As observed in the present study, higher values were obtained for the success percentage and skill score test when compared to the original values found in the index development.

It is believed that numerous works could or can be improved by a simple study of the correction of the limit values of classes in previous use of the FMA<sup>+</sup> in study areas other than that of the origin of risk.

## Conclusions

The percentage of success increased by 19.3, 21.53, and 31.3% for subzones 1, 2, and 3, respectively, when compared to the original method.

Logistic regression analysis is less sensitive to lower values and can overestimate low-risk conditions, which may lead to a situation where the risk rating is low, while the risk is actually very high.

The methodology proposed for the test of classes was efficient and allowed an analysis of the values obtained for the classes, making it possible to perform the analysis of times with a greater wildfire occurrence and the total data set.

It was observed that the application of the percentiles for the development of limits for new classes resulted in a greater index accuracy for the study subzones.

## References

- BEDIA, J. *et al.* Sensitivity of Fire Weather Index to different reanalysis products in the Iberian Peninsula. **Natural Hazards and Earth System Sciences**, [s. l.], v. 12, p. 699-708, 2012.
- BORGES, T. E. *et al.* Desempenho de alguns índices de risco de incêndios em plantios de eucalipto no norte do Espírito Santo. **Floresta e Ambiente**, Seropédica, v. 18, n. 2, p. 153-159, 2011.
- CHANG, Y. *et al.* Environmental controls on the characteristics of mean number of forest fires and mean forest area burned (1987-2007) in China. **Forest Ecology and Management**, [s. l.], v. 356, p. 13-21, 2015.
- DÍAZ-DELGADO, R. **Caracterización mediante teledetección del régimen de incendios forestales em Cataluña (periodo 1975-98) y su influencia em los procesos de regeneración**. 2000. Thesis (Doctorate) - University of Barcelona, Barcelona, 2000.
- DIMITRAKOPOULOS, A.; BEMMERZOUK, A.; MITSOUPOLOULOS, I. Evaluation of the Canadian fire weather index system in a eastern Mediterranean environment. **Meteorological Applications**, Reading, v. 18, p. 83-93, 2011. DOI: <https://doi.org/10.1002/met.214>
- DUGUY, B. *et al.* Effects of climate and extreme events on wild fire regime and their ecological impacts. In: NAVARRA, A.; TUBIANA, L. (ed.). **Regional assessment of climate change in the Mediterranean**. Netherlands: Springer, 2013. v. 2. (Advances in Global Change Research, 51).
- DUGUY, B. **Interacción de la historia de usos del suelo y el fuego en condiciones mediterráneas: respuesta de los ecosistemas y estructura del paisaje**. 2003. Thesis (Doctorate) - Universidad de Alicante, Alicante, 2003.
- EUGENIO, F. C. *et al.* Causal, temporal and spatial statistics of forest fires in areas of planted forests in Brazil. **Agricultural and Forest Meteorology**, Amsterdam, v. 266, p. 157-172, 2019.
- FLANNIGAN, M. D.; HARRINGTON, J. B. A study of the relation of meteorological variables to monthly provincial area burned by wild fire in Canada (1953-80). **Journal of Applied Meteorology**, Boston, v. 27, p. 441-452, 1988.
- HOLSTEN, A. *et al.* Evaluation of the performance of meteorological forest fire indices for German federal states. **Forest Ecology and Management**, Amsterdam, v. 287, p. 123-131, 2013.
- HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2nd ed. Wiley: New York, 2000.
- KRAWCHUK, M. A. *et al.* Global pyrogeography: The current and future distribution of wildfire.

**PLoS ONE**, [s. l.], v. 4, n. 4, p. e5102, 2009.

NUNES, J. R. S. *et al.* Análise da fórmula de monte alegre (fma) e da fórmula de monte alegre alterada (FMA+) no distrito florestal de monte alegre. **Floresta**, Curitiba, v. 40, n. 2, p. 319-326, 2010.

NUNES, J. R. S.; SOARES, R. V.; BATISTA, A. C. Análise da fórmula de monte alegre (FMA+) para o estado do Paraná. **Floresta**, Curitiba, v. 39, n. 3, p. 473-484, 2009.

NUNES, J. R. S.; SOARES, R. V.; BATISTA, A. C. FMA+ - Um Novo Índice de Perigo de Incêndios Florestais para o Estado do Paraná - Brasil. **Floresta**, Curitiba, v. 36, n. 1, p. 178-179, 2006.

PAUSAS, J. G.; KEELEY, J. E. A. Burning story: the role of fire in the history of life. **Bioscience**, Washington, v. 59, n. 7, p. 593-601, 2009.

PEREIRA, J. F.; BATISTA, A. C.; SOARES, R. V. Variação da umidade combustível florestal em função de índices de perigo de incêndios. **CERNE**, Lavras, v. 18, n. 3, p. 371-376, 2012.

PEREIRA, M. G. *et al.* Synoptic patterns associated with large summer forest fires in Portugal. **Agricultural and Forest Meteorology**, [s. l.], v. 129, p. 11-25, 2005.

PIÑOL, J.; TERRADAS, J.; LLORET, F. Climate warming, wild fire hazard, and wild fire occurrence in coastal eastern Spain. **Climatic Change**, [s. l.], v. 38, n. 3, p. 345-357, 1998.

RODRÍGUEZ, M. P. R. *et al.* Ajuste e desempenho dos índices de perigo de incêndios Nesterov, FMA e FMA+ na empresa florestal Macujire, Cuba, **Floresta**, Curitiba, v. 42, n. 4, p. 651-660, 2012.

SAMPAIO, O. B. **Análise da eficiência de quatro índices na previsão de incêndios florestais para a região de Agudos - SP**. 1999. Tese (Doutorado) - Universidade Federal do Paraná, Curitiba, 1999. 157 p.

SAN-MIGUEL-AYANZ, J.; MORENO, J. M.; CAMIA, A. Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. **Forest Ecology and Management**, Amsterdam, v. 294, p. 11-22, 2013.

SKINNER, W. R. *et al.* The association between circulation anomalies in the mid-troposphere and area burned by wildland fire in Canada. **Theoretical and Applied Climatology**, Wien, v. 63, p. 89-105, 1999.

SORIANO, B. M. A.; DANIEL, O.; SANTOS, S. A. Eficiência de índices de risco de incêndios para o pantanal Sul-Mato-Grossense. **Ciência Florestal**, Santa Maria, v. 25, n. 4, p. 809-816, 2015.

SOUZA, A. P.; CASAVECCHIA, B. H.; STANGERLIN, D. M. Avaliação dos riscos de ocorrência de incêndios Florestais nas regiões Norte e Noroeste da Amazônia Matogrossense. **Scientia Plena**, [s. l.], v. 8, n. 5, p. 1-14, 2012.

SOUZA, J. I. **Comportamento e ajustes dos índices de perigo de incêndio - FMA e FMA+ para o município de Lages-SC**. 2014. Dissertação (Mestrado em Engenharia Florestal) - Universidade Estadual de Santa Catarina, Lages, 2014.

VÁZQUEZ, A.; MORENO, J. M. Sensitivity of fire occurrence to meteorological variables in Mediterranean and Atlantic areas of Spain. **Landscape and Urban Planning**, Amsterdam, v. 24, p. 129-142, 1993.

VÁZQUEZ, A.; MORENO, J. M. Spatial distribution of forest fires in Sierra de Gredos Central Spain. **Forest Ecology and Management**, Amsterdam, v. 147, p. 55-65, 2001.

VIEGAS, D. X. *et al.* Calibração do sistema canadiano de perigo de incêndio para aplicação em Portugal. **Silva Lusitana**, Oeiras, v. 12, n. 1, p. 77-93, 2004.

VIEGAS, D. X. *et al.* Estimating live fine fuels moisture content using meteorologically based