Predictive model of the outage of transmission lines exposed to wildfires

Modelo previsor de desligamentos de linhas de transmissão expostas a incêndios florestais

Tito Ricardo Vaz da Costa [1]

Ailton Teixeira do Vale [2]

Clarissa Melo Lima [3]

Joaquim Carlos Gonçalez [4]

[1] PhD in Forestry Sciences, Specialist in Regulation, Brazilian National Electric Energy Agency - ANEEL, Brasília, DF, Brazil
E-mail: titovaz@gmail.com

[2] PhD Degree in Energy in Agriculture, Professor, Universidade de Brasília – UnB, Brasília, DF, Brazil
E-mail: ailton.vale@gmail.com

[3] PhD in Forest Sciences, Temporary Professor, Universidade de Brasília – UnB, Brasília, DF, Brazil
E-mail: limaclarissa@yahoo.com.br

[4] PhD in Forestry Sciences, Professor, Universidade de Brasília – UnB, Brasília, DF, Brazil
E-mail: goncalez@unb.br

doi:10.18472/SustDeb.v14n3.2023.50771

Received: 11/09/2023
Accepted: 20/11/2023

ARTICLE- DOSSIER

ABSTRACT

Transmission lines are essential for access to clean and affordable energy sources, Sustainable Development Goal 7. Wildfires are an important factor in the degradation of the quality of public transmission service provision. This work sought to build a model to predict the outage of a transmission line when exposed to a wildfire. The characteristics analysed of the spans exposed to fires of twelve transmission lines at a voltage level of 500 kV in Brazil totalled 3,998 km. The logistic regression technique was used for the study. It was possible to reach a model with a hit rate higher than 73% for the occurrence of transmission line outages. The quantity of fire outbreaks, the climatic variables, and the type of biome of the spans were observed to be the best predictive variables available. The temperature rise can potentially increase the number of outages caused by wildfires.

Keywords: Fires. Interruption. Electric power.

RESUMO

As linhas de transmissão são essenciais para o acesso a fontes de energias limpas e acessíveis, Objetivo de Desenvolvimento Sustentável 7. Os incêndios florestais são um fator importante de degradação da qualidade da prestação do serviço público de transmissão. Este trabalho buscou construir um modelo para previsão de desligamento de uma linha de transmissão quando exposta a um incêndio florestal. Foram analisadas as características dos vãos expostos ao fogo de 12 linhas de transmissão em nível de tensão de 500 kV no Brasil, totalizando 3.998 km. Utilizou-se a técnica de regressão logística para o estudo. Foi possível chegar a um modelo com índice de acerto superior a 73% para a ocorrência de desligamentos de linhas de transmissão. Observou-se que o quantitativo de focos de incêndios, as variáveis climáticas e o tipo de bioma dos vãos são as melhores variáveis previsoras disponíveis. O aumento da temperatura tem potencial para elevar o número de desligamentos por incêndios florestais.

Palavras-chave: *Queimadas. Interrupção. Energia elétrica.*

1 INTRODUCTION

The consequences of wildfires on the electrical grid are quite significant (Operador Nacional do Sistema Elétrico, 2016b), highlighting the reduction of renewable energy generation. However, not every wildfire causes a transmission line outage. The study of this phenomenon involves different factors: climatic variables, land use, operating conditions and technical building characteristics.

The adequate performance of transmission lines greatly influences achieving the Sustainable Development Goal 7. Transmission lines are essential for transporting renewable energy production to large load centres, minimising the effect of intermittency of renewable sources and, thus, lowering energy costs for the consumer.

Transmission line outage can occur due to a short circuit in the presence of fire due to the reduction of the dielectric strength of the air between the phase conductors and between the phase conductors and the ground. Smoke and fly ash from a wildfire can also alter the insulating characteristics of air spaces, as conductive particles drastically lower the dielectric strength of the air (Khan; Ghassemi, 2022). Another factor is the high temperature of a flame, which can decrease the tensile strength of transmission line conductors and accelerate their ageing (Guo *et al.*, 2018).

The operating state of a transmission line subjected to a wildfire varies between on and off. Thus, the operating state of the installation is understood to be a dichotomous variable. Logistic regression is a potential mathematical model for this type of output, corresponding to one variable and with multiple continuous predictive variables (Field, 2017).

Logistic regression is a type of multiple regression with a dichotomous categorical output variable and continuous or categorical predictive variables (Field, 2017). Based on certain information, we can predict which of the two categories a variable belongs to.

In a linear regression, the observed data must have a linear relationship. However, this hypothesis is violated if the output variable is dichotomous (Berry, 1993). One way around this problem is to change the data through a logarithmic transformation (Packard, 2013; Zhang; Wang; Luo, 2015). The logistic regression equation expresses a multiple linear regression equation in logarithmic terms and solves the linearity hypothesis violation problem this way.

In a logistic regression, therefore, we predict the probability of Y occurring when the values of X or Xs are known according to Equation (1).

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n + \varepsilon)}} \tag{1}$$

The probabilistic equation *P(Y)* of the logistic regression has several similarities with the linear regression equation. In reality, the exponent of the natural number *e* contains an expression identical to that of the multiple regression, where b_0 is the Intercept, b_n corresponds to the coefficient of the predictive variable, x_n and ε is the residual term.

The resulting value of the equation is a probability and always varies between 0 and 1 (Heumann; Schomaker; Shalabh, 2016). A value close to 0 means that the occurrence of Y is very unlikely, and a value close to 1 that it is very likely.

Each predictive variable has its own coefficient in the logistic equation. These parameters are estimated by adjusting models based on the observed data. The model chosen will be the one where the values of the predictive variables result in the value of Y closest to the observed value. Specifically, the parameter values are calculated using maximum likelihood estimation (Brandt, 2014). One of the main advantages of this method is that its estimators are consistent, asymptotically normal and efficient (Guera *et al.*, 2018)Pinar del Río, Cuba. Para isso, foram ajustadas dez Funções de Densidade de Probabilidade (FDPs.

As with multiple correlation, it is possible to calculate a more appropriate version of the R-coefficient in logistic regression. This R-coefficient is the partial correlation between the output variable and each of the predictive variables and can range from -1 to 1, where values close to zero indicate no correlation, positive values represent direct correlation, and negative values represent inverse correlation. The R proposed by Cox and Snell (Cox; Snell, 2008), represented by the symbol R^2_{CS}, which is based on the log-likelihood (Brandt, 2014) of the model, the log-likelihood of the original model and the sample size, according to equation (2).

$$R^2_{CS} = 1 - e^{\left[-\frac{2}{n}(VL(Novo) - VL(Básico))\right]} \tag{2}$$

However, this coefficient never reaches its maximum theoretical value of 1. As such, (Nagelkerke, 1991) suggested the following correction (Nagelkerke's R^2), according to equation (3).

$$R^2_N = \frac{R^2_{CS}}{1 - e^{\left[\frac{2(VL(Básico))}{n}\right]}} \tag{3}$$

SPSS (IBM, 2020) uses the R-coefficient proposed by Cox and Snell (Cox; Snell, 2008), considering the correction of Nagelkerke (Nagelkerke, 1991). The terms of the exponent of the natural number of equation (3) come from the log-likelihood (VL) expression described in equation (4).

$$VL = \sum_{i=1}^{N}\{Y_i \ln(P(Y_i)) + (1 - Y_i)ln[1 - P(Y_i)]\} \tag{4}$$

Equation (3) is associated with the probabilities derived from the model and the actual data. The result of the Equation indicates how much unexplained information still exists after the model has been adjusted.

Calculating means does not make sense for dichotomous variables. Thus, the basic value of the likelihood-log (VL(Basic)) of equation (4) corresponds to the category with the highest number of cases.

In logistic regression, a value called Wald presents a special distribution known as chi-square (Hastie; Tibshirani; Friedman, 2009). Wald tells us if the coefficient b_n of each predictor is significantly different from zero (Wald, 1943). If this occurs, we can assume that the predictor x_n is contributing significantly to the prediction of the output variable. Equation (5) shows how Wald is calculated, and it is possible to see that it is equal to t in the linear regression.

$$Wald = \frac{b}{EP_b} \qquad (5)$$

Where b is the regression coefficient and EP_b is its standard error.

Another important variable for the interpretation of the logistic regression is the so-called exponent. This indicator represents the variation of the probability change before and after the inclusion of the analysed variable. When the indicator is greater than 1, it indicates that the increase in the predictor is directly related to the increase in the chance of an increase in the output variable (Field, 2017).

In our specific case, we will use logistic regression to predict whether a transmission line will suffer an outage caused by wildfires, given the characteristics of transmission line spans exposed to the fires.

2 MATERIALS AND METHODS

We analysed the forced outage data of the Brazilian transmission system in 2018 and 2019 that was declared by the transmission utilities to the national network operator (Operador Nacional do Sistema Elétrico, 2016b). The obtained data contained information on the date, time, installation, outage cause and declared location for the defect. Based on this data, the outages caused by wildfires were selected.

We selected six transmission line trunks with asymmetric performance in the period regarding outages caused by fires. The selected circuits are highlighted in Table 1.

Table 1 | Trunks and transmission lines selected for the study.

Trunks	*Installation*	*Length (km)*
1	TL 500 kV COLINAS / RIB.GONÇALVES C 1 TO/PI	379
	TL 500 kV COLINAS / RIB.GONÇALVES C 2 TO/PI	367
2	TL 500 kV IMPERATRIZ / COLINAS C 1 MA/TO	343
	TL 500 kV IMPERATRIZ / COLINAS C 2 MA/TO	343
3	TL 500 kV IMPERATRIZ / P. DUTRA C 1 MA	388
	TL 500 kV IMPERATRIZ / P. DUTRA C 2 MA	388
4	TL 500 kV RIB.GONCALVES / S. JOÃO PIAUÍ C L3 PI	353
	TL 500 kV RIB.GONCALVES / S. JOÃO PIAUÍ C L4 PI	353
5	TL 500 kV TERESINA II / P. DUTRA C C9 PI/MA	210
	TL 500 kV TERESINA II / P. DUTRA C C8 PI/MA	208
6	TL 500 kV TERESINA II / SOBRAL III C V8 PI/CE	334
	TL 500 kV TERESINA II / SOBRAL III C V9 PI/CE	332

Source: Operador Nacional do Sistema Elétrico, 2023.

The six selected trunks cover the states of Piauí, Tocantins, Maranhão and Ceará, as shown in Figure 1, and they correspond to twelve 500 kV transmission lines with a total of 3,998 km.

Figure 1 | Schematic of the analyzed transmission lines (ONS, 2017)

The information in granularity by span was gathered for the six selected trunks, highlighted in Table 2.

Spans are limited by transmission line towers and are analysed according to their area of influence. Each area is delimited by the width of the safety strip established in the environmental license and by the transmission line towers (Figure 2). Span analysis is a major innovation of this work, given that similar articles analysed theoretical models or transmission lines (Guo *et al.*, 2018; Khan; Ghassemi, 2022; Shi *et al.*, 2018).

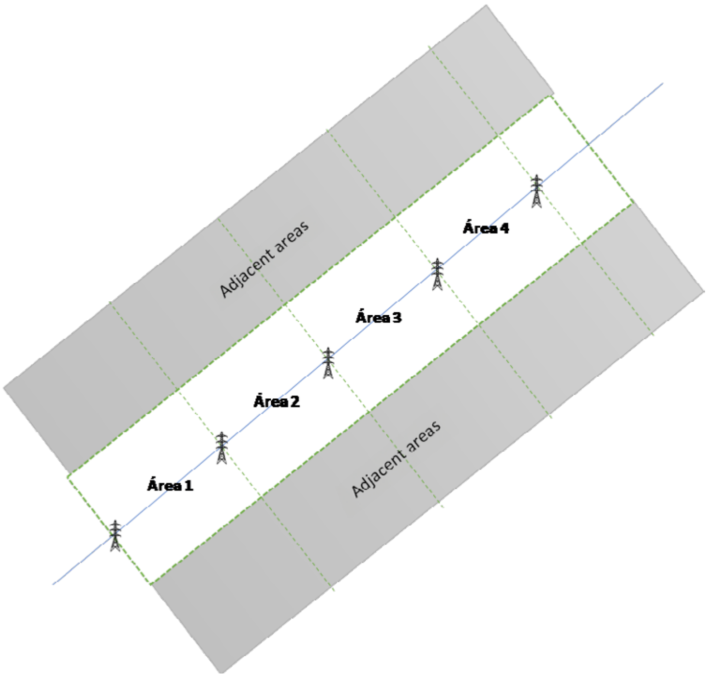

Figure 2 | Schematic representation of the areas of study

Source: Authors.

Table 2 | Analysed variables

Variable	*Measurement Scales*	*Type of characteristic*	*Unit*	*Origin*
Outages	Nominal	Performance	Not applicable	SIPER
Fire outbreaks	Ratio	Performance	Amount	Queimadas program
NDVI	Ratio	Performance	Dimensionless	GGT
Width	Ratio	Constructional	Meters	GGT
Height	Ratio	Constructional	Meters	GGT
Insulators	Ratio	Constructional	Amount	GGT
Days without rain	Ratio	Climatic	Not applicable	Inpe
Humidity	Ratio	Climatic	Percentage	Inpe
Temperature	Ratio	Climatic	Degrees Celsius	Inpe
Wind speed	Ratio	Climatic	Meters per second	Inpe
Biome	Nominal	Terrain	Not applicable	MapBiomas
Type of land	Nominal	Terrain	Not applicable	MapBiomas
Right-of-way clearing	Nominal	Performance	Not applicable	GGT

Source: Authors

The data in Table 2 were classified according to the measurement scales postulated by Stevens (Stevens, 1946). According to this classification, all measurement scales can be classified into nominal, ordinal, interval and ratio. Nominal and ratio data were used in the reported study. Nominal data corresponds to the independent dichotomous output variable. Ratio variables correspond to the continuous predictive variables.

The data of Table 2 was also classified regarding the type of characteristic reported: performance, constructive and climatic.

The performance characteristics are related to the operational dynamics of the transmission line. The forced outage data was obtained from the Integrated System of Disturbances (*Sistema Integrado de Perturbações*, SIPER) (Operador Nacional do Sistema Elétrico, 2016a). This variable represents whether or not forest fires caused a line trip.

The fire outbreak data were obtained from the Queimadas program (Instituto Nacional de Pesquisas Espaciais, 2017). The fire outbreak detection system for the polar orbit satellites can capture a fire front of about 30 m long by 1 m wide or larger. Therefore, the quantity of this variable represents the area affected by forest fire.

The Normalised Difference Vegetation Index (NDVI) index data and the right-of-way clearing data were obtained from the Geospatial Transmission Management System - GGT (Guido JR. *et al.*, 2018). NDVI is calculated by the difference in reflectance between the near infrared and red bands, normalised by the sum of the near infrared and red bands. The index varies on a scale of -1 to +1. The closer to 1, the greater the vegetation cover density; negative values represent bodies of water (Rouse *et al.*, 1973).

The constructive characteristics are those related to the design of the facilities. The width of the right-of-way for the span, the height of the towers at the ends of the span and the number of insulators per chain were obtained from the data of the GGT system (Guido JR. *et al.*, 2018). Although these variables have granularity by span, the available data reflect values by transmission line section.

The climatic data correspond to humidity, temperature, wind speed and number of days without rainfall for each span under analysis. All this information was obtained from the Brazilian Space Research Insititute (*Instituto Nacional de Pesquisas Espaciais*, Inpe). Data is available since 2000 at a resolution (pixel) of 25 km x 25 km. The data are derived from meteorological models of the *Global Forecast System* - GFS (Instituto Nacional de Pesquisas Espaciais, 2020; National Oceanic And Atmospheric Administration, 2020).

It is important to highlight that the temperature information cited in this work refers to the weather conditions and not the flame's temperature or the transmission line conductors. The nominal variables were analysed according to specific classifications, as shown in Table 3.

Table 3 | Categories of biome, land use and right-of-way clearing variables

Variables	*Dichotomous representation*	*Category*
Biome	0	Amazônia
	0	Caatinga
	1	Cerrado
Land use	0	Annual and Perennial Culture
	0	Countryside Training
	0	Forest Formation
	1	Savanna Formation
	0	Urban infrastructure
	0	Agriculture and Grassland Mosaic
	0	Other non-vegetated area
	0	Pasture
	0	River, Lake and Ocean
Right-of-way clearing	1	Authorised
	0	Authorised with restrictions
	0	Prohibited

Source: Authors

It should be stressed that the categories of the variables cited in Table 3 are not exhaustive and are limited to those listed in the database used. The dichotomous representation of the variables is necessary for the use in logistic regression models. The dichotomous representation criterion of Table 3 followed the results of the previous descriptive statistical analyses (Costa, 2021; Costa *et al.*, 2022).

It was found that 71% of the span area that caused outages is related to the land use category 'Savanna Formation'. As such, the value 1 was assigned to this variable for the dichotomous representation of the category 'Savanna Formation' and the value of 0 for the other categories. For the biome variable, the prevalence of span areas that caused outages is in the 'Cerrado' category (59%). Therefore, the dichotomous representation of the category 'Cerrado' gets value 1, and the other categories get value 0. For the right-of-way clearing variable, the 'Authorised' category got a value of 1, and the others got a value of 0.

Based on the data described in Table 2, it was possible to consolidate the data using the identifier code of the transmission line span as an identifying key. The SAS software (SAS, 2020) was used for this crosschecking.

ISSN-e 2179-9067

The data were processed following the flow described in Figure 3.

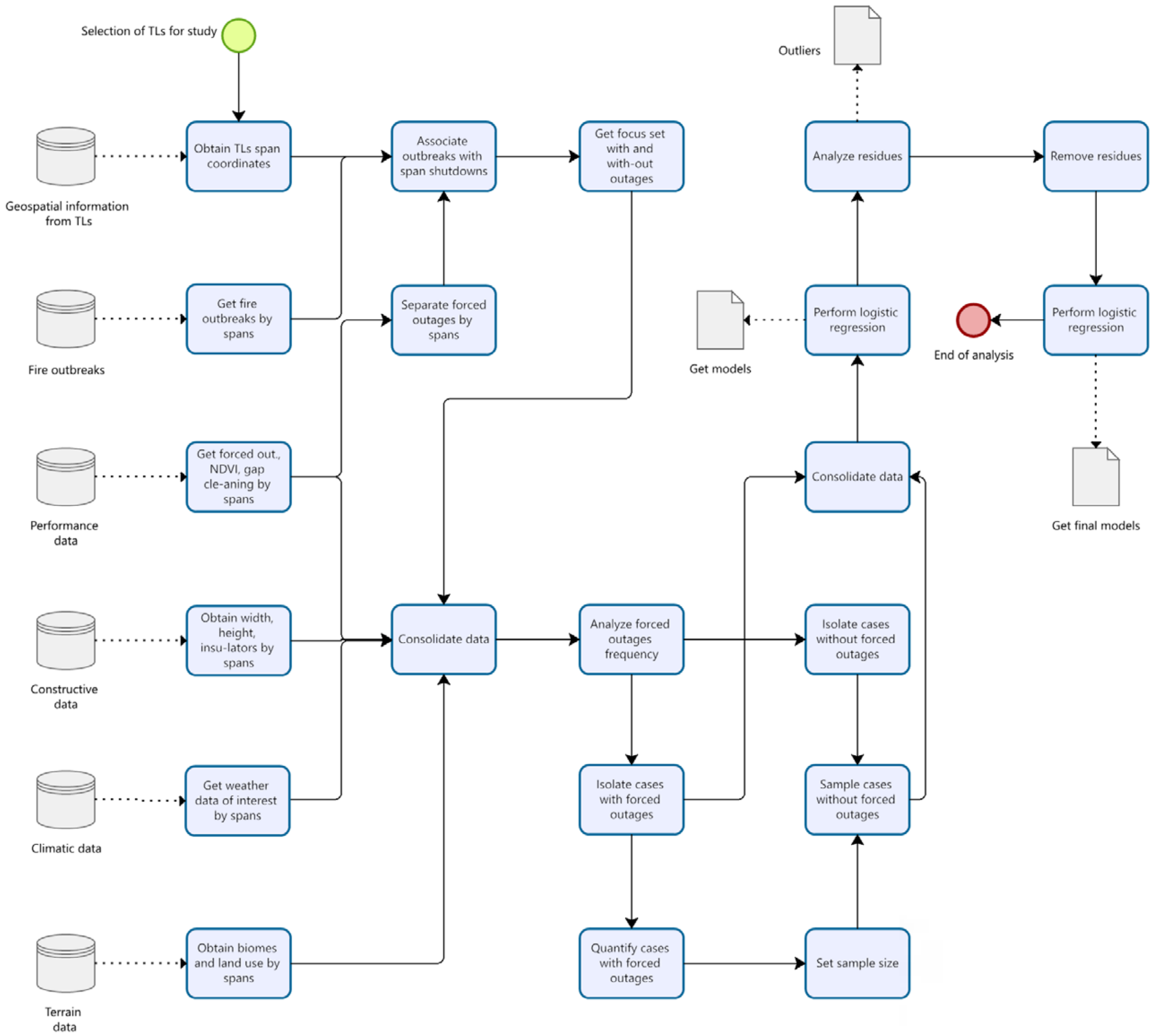

Figure 3 | Summary of the study's methodology

Source: Authors

After consolidation, the data were divided into two sets: those that caused TL outages when exposed to fire outbreaks and those that, under the same condition, did not cause outages.

Subsequently, a frequency analysis was made of the spans associated or not with outages. The spans associated with outages (smaller set) were used as a reference for the sample size of the spans associated with the absence of outages (larger set). The two data sets were once again consolidated, ensuring symmetry between the number of spans with and without outages. This initial analysis was considered a base model without including any of the variables under study.

The Wald value was calculated for the variables not included in the base model. The next step was the inclusion of the significant variable with the highest score in the base model. This model was stored with its R^2 coefficient and hit percentages being calculated.

Next, new Wald values were calculated for the variables not included in the first simulation step. The significant variable with the highest score was again selected for inclusion in the previous model. A new simulation step was generated, with the R2 coefficient and hit percentages calculated. This process was repeated until no significant variables remained outside the model.

At the end of *n* simulation steps, the model with the highest hit percentage for the occurrence of outages was selected. The residuals were also analysed to identify outliers with undue influence on the models. Standardised differences greater than ±2σ (standard deviations) were considered outliers. The outliers were removed from the database, and the logistic regression analysis was repeated, obtaining new probabilistic models.

The SPSS software was used (IBM, 2020) to construct the logistic regression models.

3 RESULTS AND DISCUSSIONS

The initial results of the logistic regression model, when only the constant of Equation (1) is included, indicate that the initial model without predictive variables has a hit probability of 94.8%, always predicting the non-occurrence of outages. However, this output does not have a valid significance. The number of observed fire outbreak events without transmission line outages is much higher than those with outages. As such, the model considers the most frequent event and associates it with the output value.

In order to prevent the difference in the number of observations from causing a bias in the frequencies of observations, the available data was subjected to prior preparation. Basically, the database contains 370 records with outages and 6754 records without outages. For this second set, 370 records were randomly selected.

Based on this consideration, the frequencies of cases with and without outages were forced to be equal. Consequently, the base probability of an outage occurring becomes 50%, and the complementary probability of an outage not occurring also becomes 50%.

Therefore, the first estimate of the model, when only the constant is included, was reviewed and described in Table 4.

Table 4 | Base Model

Observed			*Predicted*		
			Outages		*Correct percentage*
			0	*1*	
Step 0	Outages	0	0	370	0.00
		1	0	370	100.00
	Overall Percentage				50.00

Source: Authors

In Table 4, records with outages are indicated by the number 1, and records without outages are indicated by the number 0. As expected, the model could predict 50% of the occurrences correctly, given that the outage frequencies were the same. In this first step, the constant value used was zero and the results were insignificant ($p>0.05$).

The variables that were not considered in the model for Step 0 are presented in Table 5. The model's general statistic (chi-square) was 156.83 and was considered significant ($p<0.005$). This means that variables that are not in the model are significantly different from zero or, in other words, that adding one or more of these variables to the model will significantly affect its predictive power.

Table 5 | Parameters of the variables outside the model in step 0 of the simulation

Variables	*Score*	*Df*	*Sig.*
Fire outbreaks	64.061	1	0.00
NDVI	6.305	1	0.01
Right-of-way width	3.065	1	0.08
Height	18.662	1	0.00
Insulators	7.830	1	0.01
Days without rain	61.089	1	0.00
Humidity	65.913	1	0.00
Temperature	45.525	1	0.00
Wind speed	1.843	1	0.17
Biome(1)	49.174	1	0.00
Type of land(1)	28.831	1	0.00
Right-of-way clearing(1)	16.520	1	0.00
Overall Statistics	156.830	12	0.00

Source: Authors

The score data of Table 5 represent the Wald values. This is a previous indicator used by SPSS to indicate the order of inclusion of the variables in each simulation step. For the analysed data, the Wald value was not significant ($p>0.05$) for the variables corresponding to the right-of-way width and wind speed. The other variables were considered significant, with the highest score corresponding to the variable humidity.

Table 5 also shows the dichotomous predictor variables: biome, land use and right-of-way clearing authorisation. Based on the results of a previous descriptive statistical analysis, we sought to investigate the Cerrado biome's effect on the performance of transmission lines. As such, the value 1 was assigned to all spans located in Cerrado areas. The value 0 was assigned to the spans located in other biomes. The same procedure was repeated for the land use classified as Savanna Formation, which received the value 1, and the other uses, which received the value 0. In the case of the authorisation variable, the value 1 was assigned to all spans without right-of-way clearing restrictions and 0 for the spans with partial restrictions or prohibition. The three dichotomous variables under analysis were considered significant ($p<0.05$) for the model.

The simulation is done by including one variable at a time, following the order imposed by the Wald value presented in Table 5. For example, in Step 1, the simulation was made considering the variable Humidity and the constant.

At each step of the simulation, the same parameters of Table 5 are recalculated, with the variable with the highest Wald value being included in the model and the non-significant variables ($p > 0.05$) discarded. The statistical summary of the new simulated models is presented in Table 6.

Table 6 | Statistics of the new model

Step	*-2 log-likelihood*	*R2 Cox & Snell*	*R2 Nagelkerke*
1	951.261a	0.096	0.128
2	896.682b	0.160	0.214
3	872.749b	0.187	0.249
4	855.190b	0.206	0.275
5	847.957b	0.214	0.285
6	848.302b	0.213	0.284
7	843.750b	0.218	0.291

a. The estimate was stopped in iteration number 4 because the parameter estimates changed by less than 0.001.

b. The estimate was stopped in iteration number 6 because the parameter estimates changed by less than 0.001.

Source: Authors

As the variables are included, Nagelkerke's R^2 increases so that at the end of the 7 steps, we get the value of 0.291 on a scale of 0 to 1. This result reveals that there are representative random factors that influence the probability of the occurrence of outages.

The results of the hit percentage at each simulation step are shown in Table 7.

Table 7 | Models generated at every step

Observed		*Predicted*		
		Outages		*Correct percentage*
		0	*1*	
Step 1	0	186	184	50.27
	1	118	252	68.11
				59.19
Step 2	0	257	113	69.46
	1	130	240	64.86
				67.16
Step 3	0	271	99	73.24
	1	131	239	64.59
				68.92
Step 4	0	271	99	73.24
	1	111	259	70.00
				71.62
Step 5	0	274	96	74.05
	1	113	257	69.46
				71.76
Step 6	0	279	91	75.41
	1	113	257	69.46
				72.43

Observed		*Predicted*		
		Outages		*Correct percentage*
		0	*1*	
Step 7	0	284	86	76.76
	1	120	250	67.57
				72.16

Table 7 shows that after seven steps, the simulation reached a model that can correctly predict 72.16% of cases. The best average result occurred in the sixth step, with an accuracy of 72.43%. Remembering that the base case (Table 4) reached a 50% hit percentage, we can state that the generated models managed to increase the outage prediction accuracy by up to 44.86%.

However, this work aims to predict the events that will generate transmission line outages with a greater hit probability. According to this objective, the model with the best result was obtained in Step 4, with 70.00% of correct predictions.

After presenting the model's statistical indicators, the equation coefficients' values are presented (1) to Step 4 of the simulation. These results are presented in Table 8.

Table 8 | Coefficients of the generated models

Step	*Variable*	*B*	*Standard Error*	*Wald*	*df*	*Sig.*	*Exp(B)*
4	Fire outbreaks	0.04	0.01	29.69	1.00	0.00	1.04
	Days without rain	0.01	0.00	19.46	1.00	0.00	1.01
	Humidity	-0.04	0.01	13.09	1.00	0.00	0.96
	Biome(1)	0.79	0.19	17.12	1.00	0.00	2.21
	Constant	0.01	0.37	0.00	1.00	0.97	1.01

In addition to the variable's coefficient, Table 8 provides the standard error information associated with each calculated coefficient, the Wald value, the degree of freedom, the significance level and B exponent.

Considering the model with the best fit to the objectives of this work, corresponding to step 4, we realised that it uses the continuous variables fire outbreaks, days without rain and humidity; the dichotomous variable biome; and the constant. That is, the climatic conditions, the terrain conditions, and the size of the wildfire are determinants for the occurrence of transmission line outages.

Considering the B exponent, the variable with the greatest chance of increasing the hit probability of the model is the type of Biome (2.21). With this variable, the influence on the performance of transmission lines with spans located in the Cerrado biome was studied. The results indicate that this type of biome has a higher probability of causing transmission line outages because of wildfires.

The absence of the transmission line's constructive variables in all the generated models stands out. The explanation may be the quality of the available data, which considered average values declared by the transmission concessionaires for the entire transmission line studied, which proved inadequate granularity for the proposed study.

Improving data quality depends on specific and ongoing regulatory action, which consists of building a technical database of transmission assets with a georeferenced basis and submetric accuracy (Agência Nacional de Energia Elétrica, 2019). It is reasonable to infer that a base with these characteristics will be able to reduce the randomness of the models (higher Nagelkerke's R^2) and increase the hit probability of outage events caused by wildfires.

Regulatory incentives for transmission utilities can also be evaluated against results since they get paid only when their system is available. Revenue is discounted in the case of outages, and the goal is to ensure maximum system availability (Agência Nacional de Energia Elétrica, 2016).

Brazilian regulations treat the wildfire phenomenon as an exception to the general rule. Outages caused by these phenomena can be exempt from revenue discounts. This exception reflects an understanding of the Brazilian regulator regarding the limitation of the transmission utility's ability to take preventive actions against wildfires (Agência Nacional de Energia Elétrica, 2016).

The regulator's perspective received validation from the model to some extent. However, apart from the design phase, where the line's layout determines the biomes it traverses, the transmission utilities cannot proactively mitigate outages. The model primarily revealed that climate factors, independent of the transmission utilities' actions, influence outage occurrences.

It is also interesting to note that the environmental restrictions arising from licensing did not prove relevant. In all simulation steps, the right-of-way width variable was statistically insignificant ($p > 0.05$) for the model. The authorisation variable, which represents the existence or not of restrictions to right-of-way clearing along the span, only showed statistical significance ($p<0.05$) in the first step of the simulation. Still, as its Wald value was 5.96, it did not even enter the model corresponding to the simulated step.

It is also worth noting that the variables temperature and land use were considered significant, although they only entered the model after Step 5. Since step 4 obtained the best hit probabilities, these variables did not appear in the final model.

3.1 RESIDUAL ANALYSIS

A standardised residual analysis was performed. These values are the standardised differences between the observed data and the values that the model predicts. Differences greater than $\pm2\sigma$ (standard deviations) were considered discrepancies. Ten records considered atypical were found. These records

represent 1.35% of the database and were excluded because they unduly influenced the model.

3.2 NEW MODEL AFTER THE EXCLUSION OF THE RESIDUALS

The simulations were repeated, considering the remaining 730 records after excluding outliers. New models were obtained from 5 simulation steps. The statistics of the new models are described in Table 9.

Table 9 | Statistics of the new model

Step	*--2 log-likelihood*	*R2 Cox & Snell*	*R2 Nagelkerke*
1	883.116a	0.162	0.216
2	834.389a	0.216	0.288
3	802.387a	0.250	0.333
4	776.089a	0.276	0.368
5	770.014a	0.282	0.376

a. The estimate was stopped in iteration number 6 because the parameter estimates changed by less than 0.001. Source: Authors

ISSN-e 2179-9067

Table 9 shows that the exclusion of outliers raised the values of Nagelkerke's R². That is, the residuals were confusing the models. Considering the best cases before and after removing the residuals, we can state that it was possible to increase Nagelkerke's R2 by 29,2% (0,291 to 0,976).

Regarding the hit percentage, the new results at each simulation step are shown in Table 10.

Table 10 | Hit percentage of the new models generated at each step

Observed		*Predicted*		
		Outages		*Correct percentage*
		0	*1*	
Step 1	0	296	67	81.5
	1	190	177	48.2
				64.8
Step 2	0	272	97	74.9
	1	126	241	65.7
				70.3
Step 3	0	271	92	74.7
	1	140	227	61.9
				68.2
Step 4	0	271	92	74.7
	1	98	269	73.3
				74.0
Step 5	0	276	87	76.0
	1	102	265	72.2
				74,1

Source: Authors

Table 10 shows that a model was reached after five steps of simulation that managed to hit 74.1% of the predictions. This means there was an improvement in the hit percentage of the order of 2.68% compared to simulations without removing residuals. For the base case (Table 4), we can say that the new models generated increased the outage prediction accuracy by up to 48.2%.

However, just as in the previous simulation, the model that achieved the highest hit rate for the occurrence of outages (output value 1) corresponds to step 4. The accuracy observed in this case was 73.3%, an improvement of 4.71% regarding the equivalent model with the presence of residuals.

The values of the coefficients of the equation (1) to step 4 of the simulation are presented below. These results are presented in Table 11.

Table 11 | Coefficients of the newly generated models

Step	*Variables*	*B*	*Standard Error*	*Wald*	*df*	*Sig.*	*Exp(B)*
4	Fire outbreaks	0.08	0.01	47.84	1.00	0.00	1.08
	Days without rain	0.01	0.00	22.82	1.00	0.00	1.01
	Temperature	0.25	0.05	26.66	1.00	0.00	1.29
	Biome(1)	1.02	0.20	25.31	1.00	0.00	2.78
	Constant	-10.46	1.75	35.79	1.00	0.00	0.00

Source: Authors

An important difference observed in the new simulations was removing the humidity variable in Step 4 and replacing it with the temperature variable. After the exclusion of outliers, the humidity variable presented non-significant values ($p>0.05$), which meant it was excluded from the simulations. The increase of the B exponent values for all variables can also be highlighted. In the case of the Biome variable, the value of the B exponent was 2.78, reinforcing the importance of this information for the model.

The model resulting from the reported simulation and corresponding to step 4 can be obtained by replacing the coefficients of Equation (1):

$$P(Y) = \frac{1}{1 + e^{-(-10{,}46+0{,}08X_1+0{,}01X_2+0{,}25X_3+1{,}02X_4)}} \tag{6}$$

The result of Equation (6) represents the probability of an outage occurring because of a wildfire. Numbers 1 to 4 represent the predictive variables corresponding to fire outbreaks, days without rain, temperature, and biome, respectively. The climatic variables of the obtained model coincide with the results of the probabilistic model of (Shi *et al.*, 2018), which studied the influence of forest fires on transmission lines in Hubei Province, China.

Considering Equation (6), it is possible to simulate some scenarios to understand how the performance of the studied transmission lines will be affected. In the first case, we evaluated the variation in the probability of outages due to wildfires in the Cerrado biome and other biomes (Caatinga and Amazon). For this scenario, the variables fire outbreaks (19.2) and days without rain (47.98) are kept constant, and the temperature is varied. Figure 4 shows the results.

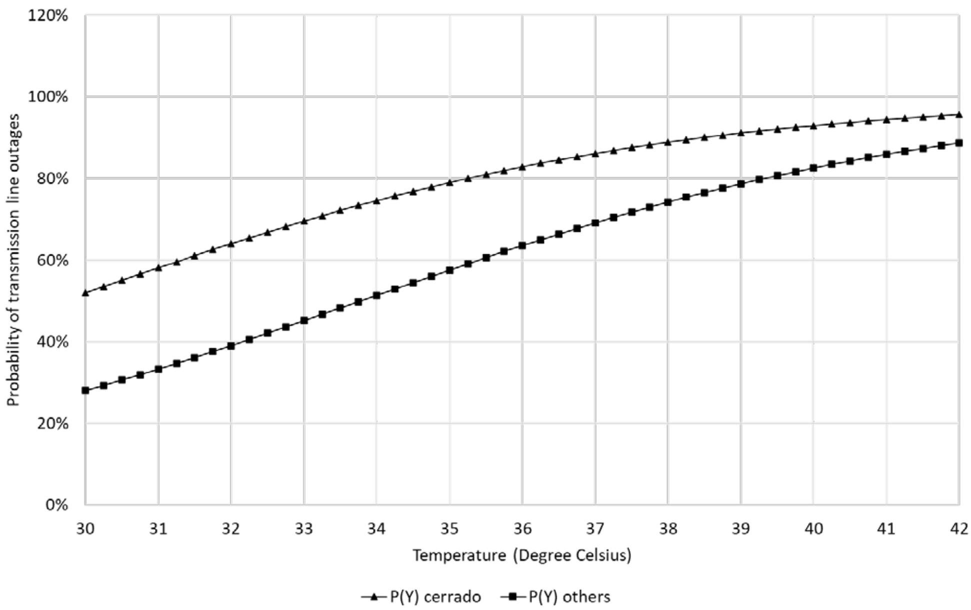

Figure 4 | Probability of outages versus temperature for the Cerrado and other biome scenarios.
Source: Authors.

The simulation demonstrates that the probability of transmission line outages in the Cerrado biome is consistently higher than in other biomes for a temperature range between 30oC and 42oC. It is important to highlight that the indicated temperatures refer to the climate and not the flame's temperature or the transmission lines' conductors.

For the temperature of 36ºC, with the simulated parameters, the risk of outages in the Cerrado biome is 82.9% against 63.5% in other biomes. The probability of power outages on transmission lines is 30% higher in the Cerrado than in other biomes.

With the temperature of 36oC as a reference, Figure 4 shows that 1oC increases the outage probability by 4% in the Cerrado biome and 10% in the other biomes. Even if the UN global warming target below 1.5oC is maintained (Silva *et al.*, 2019), relevant impacts on the performance of the studied transmission lines will be observed.

In a second simulation, we evaluated the variation in the probability of line outages as a function of the number of fire outbreaks. For this scenario, the considered biome (Cerrado) and the days without rain (47.98) were kept constant. The temperature variable was again varied. Figure 4 Figure 5 shows the results.

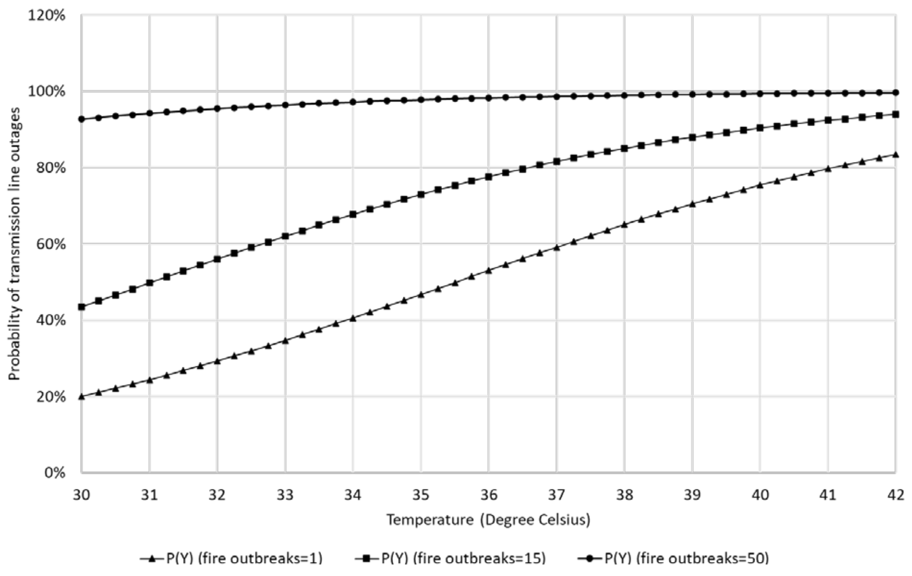

Figure 5 | Probability of outages versus temperature for the different fire outbreak scenarios
Source: Authors

Based on the simulation shown in Figure 5, we can conclude that the number of fire outbreaks, that is, the affected area, greatly influences the probability of transmission line outages. Considering the temperature of 36oC, the probability of outages is 46% higher for the situation with 15 detected outbreaks against 1 outbreak. At the same temperature, for a detection situation of 50 fire outbreaks, the probability of transmission line outages reaches 98.3%.

The last simulation presented in this work evaluates the probability of transmission line outages as a function of the number of days without rain. For this scenario, the biome considered was the Cerrado and the number of fire outbreaks (19,2) was kept constant. Figure 6 shows the results.

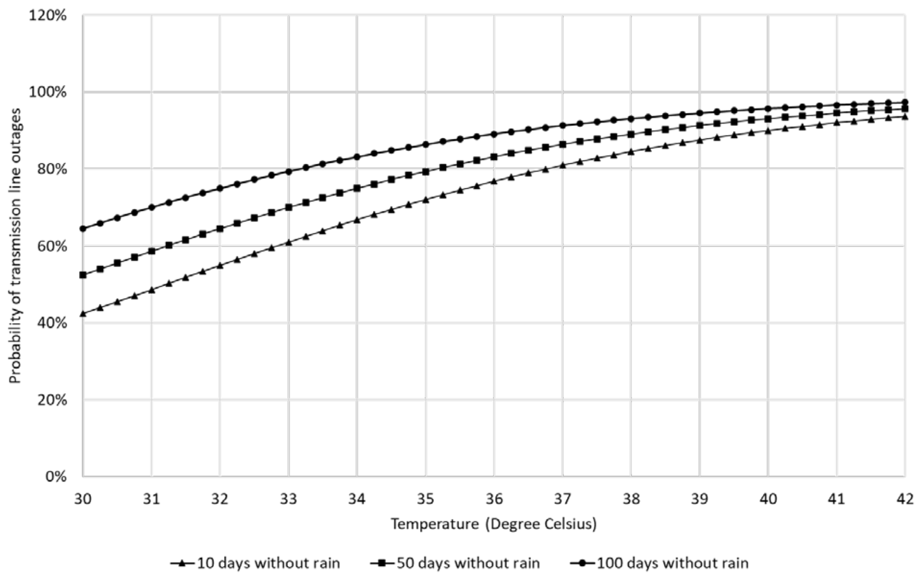

Figure 6 | Probability of outages versus temperature for the different scenarios of days without rain.
Source: Authors

The scenarios presented demonstrate that the influence of days without rain is greater in milder temperatures. With the temperature of 36°C as a reference, the variation in the probability of outages is 8.2% between 10 and 50 days without rain and 7.1% between 50 and 100 days without rain. At a temperature of 40ºC, the probability difference of outages between the 10 and 100-day rain-free scenarios is only 6.3%.

The models found can be applied to improve the accuracy of business plans for new transmission lines, as a tool for choosing layouts for new transmission lines, and to improve the transmission line maintenance process.

4 CONCLUSION

It was possible to build a logistic regression model that calculates the outage probability of a transmission line from the characteristics of the spans exposed to the fire outbreaks. The built model revealed the importance of environmental and terrain characteristics for outages caused by wildfires. The constructive characteristics of the lines and the NDVI index proved inefficient for the proposed application. Simulations also demonstrated the impact of the analysed variables on the probability of transmission line outages. Higher temperatures will invariably cause increased outages of these facilities, with the greatest impacts observed in the biomes Caatinga and Amazon.

The methodology applied in this study can be replicated in other countries sensitive to the phenomenon of wildfires. In this way, the work can contribute significantly to the construction of resilient infrastructures at a global level.

The results of this study can be considered by regulators and planners in the electricity sector in new transmission line projects, reducing the likelihood of shutdowns caused by wildfires. This can be done by seeking to combine the technical needs of the projects with the most favourable climatic and terrain characteristics to ensure the operation of these facilities is not interrupted.

These measures can potentially increase the resilience of these installations, contributing to the fulfilment of SDG 7 in terms of increasing the reliability of electricity supply. Furthermore, there will be a positive economic impact for energy transmission concessionaires, which will avoid discounts on their revenues due to forced shutdowns.

REFERENCES

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Resolução Normativa no 729, de 28 de junho de 2016. Estabelece as disposições relativas à qualidade do serviço público de transmissão de energia elétrica, associada à disponibilidade e à capacidade operativa das instalações sob responsabilidade de concessionária de transmissão integrantes da Rede Básica e das instalações de transmissão de energia elétrica destinadas a interligações internacionais que se conectam à Rede Básica, conforme Resolução Normativa no 442, de 26 de julho de 2011, e dá outras providências. **Diário Oficial da União – DOU**, [*S. l.: s. n*.], 2016. p. 97. Available at: http://www2.aneel.gov.br/cedoc/ren2016729.pdf. Accessed on: 15 out. 2020.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Resolução Normativa no 861, de 26 de novembro de 2019. Dispõe sobre a definição da Base de Dados das Instalações de Transmissão de energia elétrica e dá outras providências. **Diário Oficial da União**, [S. l.: s. n.], 2019. p. 46. Available at: http://www2.aneel.gov.br/cedoc/ren2019861.pdf. Accessed on: 15 out. 2020.

BERRY, W. **Understanding Regression Assumptions**. California: Sage Publications, 1993. Available at: https://doi.org/10.4135/9781412986427

BRANDT, S. **Data Analysis**. Cham: Springer International Publishing, 2014. Available at: https://doi.org/10.1007/978-3-319-03762-2

COSTA, T. R. V. da. **Faixas de Segurança de Linhas de Transmissão e seu Modelo de Desempenho Frente aos Incêndios Florestais**. 130 f. 2021. Universidade de Brasília, [*s. l.*], 2021. Available at: https://repositorio.unb.br/handle/10482/41238

COSTA, T. R. V. da *et al.* Susceptibility Characteristics of 500 kV Transmission Lines to Forced Outages Caused by Wildfires. **Electric Power Systems Research**, [*s. l.*], v. 209, p. 107995, 2022. Available at: https://doi.org/10.1016/j.epsr.2022.107995

COX, D. R.; SNELL, E. J. Analysis of Binary Data. *In*: **The Concise Encyclopedia of Statistics**. New York, NY: Springer New York, 2008. p. 4–5. Available at: https://doi.org/10.1007/978-0-387-32833-1_5

FIELD, A. **Discovering Statistics Using IBM SPSS Statistics**. 5. ed. Thousand Oaks: Sage Publications Inc, 2017. v. 1, E-book.

GUERA, O. G. M. *et al.* Evolução da distribuição diamétrica em plantios de Pinus caribaea Morelet var. caribaea Barrett & Golfari. BIOFIX Scientific Journal, [*s. l.*], v. 3, n. 1, p. 161, 2018. Available at: https://doi.org/10.5380/biofix.v3i1.58132

GUIDO JR. P. *et al.* Sistema de Gestão Geoespacializada da Transmissão – GGT. **Eletro Evolução**, [*s. l.*], v. 6, n. 91, p. 45–52, 2018. Available at: https://drive.google.com/file/d/1l0A6ugM_bKdzQ-1VXRYOT3TBpAaYTolx/view?usp=sharing

GUO, Y. *et al.* Determination of the power transmission line ageing failure probability due to the impact of forest fire. **IET Generation, Transmission & Distribution**, [*s. l.*], v. 12, n. 16, p. 3812–3819, 2018. Available at: https://doi.org/10.1049/iet-gtd.2018.0039

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. 2. ed. New York, NY: Springer New York, 2009. (Springer Series in Statistics). Available at: https://doi.org/10.1007/b94608

HEUMANN, C.; SCHOMAKER, M.; SHALABH. **Introduction to Statistics and Data Analysis**. Cham: Springer International Publishing, 2016. Available at: https://doi.org/10.1007/978-3-319-46162-5

IBM. **Software IBM SPSS**. [S. l.], 2020. Available at: https://www.ibm.com/br-pt/analytics/spss-statistics-software. Accessed on: 4 oct. 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Clima Monitoramento Brasil – Cptec/Inpe**. [*S. l.*], 2020. Available at: http://clima1.cptec.inpe.br/monitoramentobrasil/pt. Accessed on: 28 sept. 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Programa de Monitoramento de Queimadas**. [*S. l.*], 2017. Available at: http://www.inpe.br/queimadas. Accessed on: 21 sept. 2017.

KHAN, I.; GHASSEMI, M. A probabilistic approach for analysis of line outage risk caused by wildfires. **International Journal of Electrical Power & Energy Systems**, [*s. l.*], v. 139, p. 108042, 2022. Available at: https://doi.org/10.1016/j.ijepes.2022.108042

NAGELKERKE, N. J. D. A Note on a General Definition of the Coefficient of Determination. **Biometrika**, [*s. l.*], v. 78, n. 3, p. 691, 1991. Available at: https://doi.org/10.2307/2337038

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION. **National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC)**. [*S. l.*], 2020. Available at: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs. Accessed on: 28 sept. 2020.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **SINDAT**. [*S. l.*], 2023. Available at: http://betasindat.ons.org.br/SINDAT/Home/ControleSistema. Accessed on: 1 sept. 2017.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Sistemas ONS**. [*S. l.*], 2016a. Available at: http://www.ons.org.br/agentes/sistemasons.aspx. Accessed on: 3 aug. 2017.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Submódulo 18.2 - Relação dos sistemas e modelos computacionais**. Brasil: [s. n.], 2016b. Available at: http://ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes.

PACKARD, G. C. Is logarithmic transformation necessary in allometry? **Biological Journal of the Linnean Society**, [s. l.], v. 109, n. 2, p. 476–486, 2013. Available at: https://doi.org/10.1111/bij.12038

ROUSE, J. W. *et al.* **Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation Progress Report RSC 1978-1**. Houston: [*s. n.*], 1973. Available at: https://ntrs.nasa.gov/citations/19730017588.

SAS. **Software de Business Analytics e Business Intelligence – SAS**. [*S. l.*], 2020. Available at: https://www.sas.com/pt_br/home.html. Accessed on: 5 oct. 2020.

SHI, S. *et al.* A Model Design for Risk Assessment of Line Tripping Caused by Wildfires. **Sensors**, [*s. l.*], v. 18, n. 6, p. 1941, 2018. Available at: https://doi.org/10.3390/s18061941

SILVA, P. S. *et al.* Impacts of the 1.5°C global warming target on future burned area in the Brazilian Cerrado. **Forest Ecology and Management**, [*s. l.*], v. 446, p. 193–203, 2019. Available at: https://doi.org/10.1016/j.foreco.2019.05.047

STEVENS, S. S. On the Theory of Scales of Measurement. **Science**, [*s. l.*], v. 103, n. 2684, p. 677–680, 1946. Available at: https://doi.org/10.1126/science.103.2684.677

WALD, A. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. **Transactions of the American Mathematical Society**, [*s. l.*], v. 54, n. 3, p. 426, 1943. Available at: https://doi.org/10.2307/1990256

ZHANG, Y.; WANG, J.; LUO, X. Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel. **Energy Conversion and Management**, [*s. l.*], v. 96, p. 440–451, 2015. Available at: https://doi.org/10.1016/j.enconman.2015.03.012