



Lifetime Study of Electrical Power Distribution Systems Failures

Matheus S. S. Fogliatto¹  · Luiz Desuó N.¹ · Rafael R. M. Ribeiro¹ · José Roberto B. A. Monteiro¹ · João B. A. London¹ · Michel Bessani² · Carlos D. Maciel¹

Received: 25 March 2021 / Revised: 8 August 2021 / Accepted: 10 December 2021 / Published online: 7 January 2022
© Brazilian Society for Automatics–SBA 2022

Abstract

Electricity is a fundamental resource for modern society. However, some threats are faced by electrical power distribution systems, which are responsible for delivering electricity to end consumers. Analysing how much time these hazards will threaten these systems, causing failure events, is an essential area of study. Through statistical methods, it is possible to study this behaviour from time until failure, as well as to observe the influence of variables at this time, providing models to predict when a failure event will occur. In this study, reliability analysis regression techniques are used on real data, constructing a model for all failures and for different groups of failures, using nonparametric and parametric methods to estimate the reliability and cumulative hazard curves. An analysis of the failure causes directly linked to weather events, using six weather variables, is also made.

Keywords Causes of failure · Distribution system · Failure event · Reliability/survival regression

1 Introduction

Electrical power distribution system (DS) is the part of the power system responsible for delivering electricity to final customers, such as houses, hospitals and industries. Therefore, maintaining their operation in normal condition is a concern of modern society (of Economic Advisers et al. 2014)(Brem 2015). However, DSs generally occupy a large area and are exposed to external environment, ending up being subject to various threats that can cause interruptions in the supply of electricity. As a consequence, the study of DSs failures is a current concern of researchers, public agencies and society to enhance the reliability and resilience of DSs (Zio 2009).

The leading causes of failures in DSs (Złotecka and Sroka 2018)(Sroka and Złotecka 2019) are accidents with animals

(Sahai and Pahwa 2006) or vegetation (Radmer et al. 2002), weather events (Konal et al. 2018)(Pahwa 2007), load transfer (Rodriguez-Garcia et al. 2019), device failures (Li et al. 2019) and rare events like terrorism, vandalism and cyber-attacks (Ni and Li 2019). Therefore, several factors may be associated with these failures, from equipment failure to vehicle accidents colliding with network elements. For a better understanding of them, a separation of these failures into groups (bringing together similar threats or causes) may be interesting.

Two factors are often recurrent in failure studies: how long until it happens (time until failure or lifetime) and how long until the system recover itself (time until repair or duration) (Bessani et al. 2016). For these types of studies, reliability analysis regression (Colosimo and Giolo 2006) (Cox and Oakes 2018), also known as survival analysis regression, provides models to estimate the reliability and hazard using time as a variable, making it possible to determine the group of failure that is more recurrent, unlike an analysis considering only the number of failure events for each group. This approach is used in areas such as medicine (Moolgavkar et al. 2018) and engineering, in general problems (Dantas et al. 2010) or more specifically, as in reliability engineering (Murthy et al. 2004). For these models (Mishra et al. 2019), three major techniques can be found in the literature: the parametric (Zhang 2016), the nonparametric (Rink et al.

An early version of paper was presented at XXIII Congresso Brasileiro de Automática (CBA 2020). This work was partially support FAPESP: 2014/50851-0, CNPq: 465755/2014-3 and BPE Fapesp 2018/19150-6.

✉ Matheus S. S. Fogliatto
matheusfogliatto@usp.br

Carlos D. Maciel
carlos.maciel@usp.br

¹ University of São Paulo, São Carlos, Brazil

² Federal University of Minas Gerais, Belo Horizonte, Brazil

2013) and the semi-parametric (Shauly et al. 2011). Parametric statistics need a family of probability distribution and can be more accurate than nonparametric techniques (Zhang 2016). From previous works, the Weibull distribution proved a good choice for problems involving DS failures (Fogliatto et al. 2019).

For some failures groups, weather events are directly linked with the occurrence of a failure event. Regression including weather events as covariates can be used to evaluate the impact of a particular weather event in a failure group, by analysing the value of the coefficients for these variables with the values for the model constructed considering all failure events. An understanding of the characteristics of the different groups of failure events that threaten DSs is useful for improvements in the reliability of the system, and real data are useful because the results achieved can be used in locations with similar characteristics.

In this paper, the different causes of failure events in a real DS from a Brazilian city, from a period of almost two years, are separated into six groups, according to the characteristics of these events. A group considering all failures events (all groups together) are used. The main outputs of reliability analysis regression, that are the reliability and cumulative hazard functions, are plotted by nonparametric (Kaplan–Meier and Nelson–Aalen) and parametric (Weibull) techniques for all groups, to estimate the dangers of each failure group in terms of the lifetime of the system (time intervals between failures). For all groups, the nonparametric estimators and the Weibull univariate were used considering only the lifetime values, and to include weather events as covariates the accelerated failure time (AFT) model was used. For the failure groups that have some relation with weather events and for the “All Failures” group, an analysis using the Weibull AFT model was used to include six covariates (from data of maximum daily values of the number of atmospheric discharges, wind speed, maximum and minimum temperature, precipitation and relative humidity of the analysed city) to measure the impact of weather in the lifetime of the system. From the models including weather covariates, the reliability function graphics varying the weather values was shown to demonstrate how these covariates modify the lifetime of the DS. Also, tables of predictions in terms of a median and an expected time for the survival of the system from the three groups that weather covariates were included are presented.

The paper is organized into four sections. Sect. 2 presents a brief description of the analysed real DS; the failure data set, with examples of the failure groups that were used; an example table of the weather data; an analysis of the issue count of failures for each group, showing the threat order of each group in percentage terms; the statistical theory for the univariate and AFT regression models, for the Kaplan–Meier/Nelson–Aalen methods for the nonparametric and the Weibull distribution for the parametric analyses. Section 3

summarizes the results for the constructed models, with equations and tables showing the coefficients for the regressions, and graphics of the reliability and cumulative hazard functions, the main outputs for these methods. Section 3 presents the conclusions and final remarks.

2 Material and Methods

The real Brazilian DS used in this research is illustrated in Fig. 1. This DS (as the failure and weather data presented below) is a Brazilian city located in the southern part of the country. This city features a relatively new electrical network. It has a population of about 500,000 inhabitants and its economy generates mainly around agriculture, industry and services. A utility operates this system and is responsible for its proper functioning. In general, the components of a DS can be defined in a more general way, through substations, feeders, sectors and buses. Since the substations are the elements that receive electrical energy with high voltage values coming from the transmission system, the feeders are branches of the substations that carry electrical energy under conditions called medium voltage to the final customers. Feeders are independent in the sense that they are not connected to each other and are composed of smaller elements such as bars (depending on the context, they can represent switches, transformers or poles).

Failure data (any case of interruption in the supply of electricity) from this system were used together with atmospheric event data from this same geographic region to provide a model to predict the lifetime of the system. In Fig. 2, considering “start” as the starting point for analysing the operation of a distribution system, normal operation (DS operating in its best configuration and serving all customers) can be defined

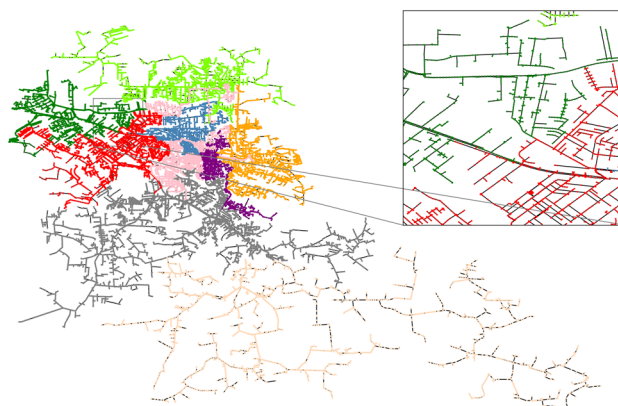


Fig. 1 DS of a midsize Brazilian city with 500.000 people, and about 85 years old, composed of 8 substations, 65 feeders, 40143 buses (distribution transformers) and a network with length of 1660.17 kilometres. A subfigure is plotted for more details. Each colour represents the feeders of the same substation (Color figure online)

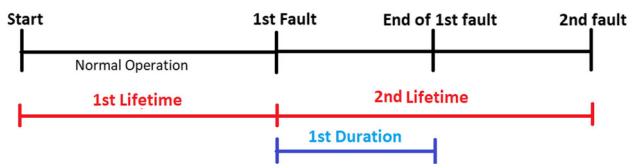


Fig. 2 Example to define concepts covered throughout the article, such as interruption occurrence, interruption duration and lifetime

as the period between the “start” of the analysis until the first recorded failure event (interruption). Lifetime is this period until the occurrence of an interruption, or, in other words, it is the time spent between failures, and is different from duration, which is the time period from the initiation of an interruption to a customer until service has been restored to that customer (IEEE Std 2004). Considering that failure or interruption events are independent since the distribution system is structured in several sections without direct connection to each other (in general, the consequences of an interruption only influence the operation of the feeder to which some section needs to be interrupted, not causing influence on the other feeders that make up the system), it was assumed that a new lifetime starts from the outage event record point and lasts until a new event occurs. (Using the example in the figure, the second lifetime would be the period between the first and second recorded failures.)

Table 1 presents the failure event data set from the DS of the city analysed. “Lifetime”, presented in minutes, was calculated using the starting date and hour of the each interruption event and represents the amount of time the system has been operating under normal conditions until a failure has occurred (time to event/failure). It is considered that the system fully recovers after a failure, thus starting a new survival period. From the analysed period, 12028 failure events were registered (failures of any duration). “Substation” and “Feeder” columns serve to filter the region where the interrupt event occurred and were used to filter out failure events in the city of interest in this study. “Type” and “CauseDesc.” provide information about the characteristics of the event that occurred, such as whether the utility operating the DS expected this event or not and what caused it.

Table 2 shows the atmospheric data set, composed of the maximum daily values for six weather events in the period from February/2013 to December/2014 (689 days): the number of atmospheric discharges, wind speed in kilometre/hour,

amount of precipitation in millimetres, the maximum and minimum temperature in °Celsius and relative humidity in percentage. The climate of the city of Londrina is the subtropical humid mesothermal, characterized by rainfall in all months of the year, but in greater quantities in the summer (December to March).

From all failure events, the maximum lifetime was 6419 minutes (almost 107 hours, about 4 and a half days). The maximum number of atmospheric discharges was 2267, the maximum wind speed was 97.2 km/h, the maximum amount of precipitation was 80.8 mm, the maximum relative humidity was 98% and the maximum and minimum temperature was 38.3 and 10.2 °C, respectively, for the analysed period. None of the peak values for atmospheric events occurred on the same day. A histogram of the lifetime is presented in Fig. 3 and a concentration of the number of occurrences from 0 to 150 minutes is observed. The mean lifetime value was 87.5 minutes.

Table 3 details the information shown in the “Type” column present in the failure data set, where three groups can be found: accidental, volunteer and scheduled. This classification refers to the system utility’s planning for the occurrence of that interruption. Accidental would be unplanned interruptions, caused by external elements, volunteer refers to events caused by the utility, but in an “emergency” regime, and scheduled the events planned and scheduled by the utility (planned interruption), generally for network expansion, load manoeuvring or maintenance preventive. From the used data, 21.82% of failures are scheduled, 23.00% are volunteers, and 55.17% are accidental failures. Considering that volunteer events end up having consequences for customers, it is possible to say that more than 75% of the failures events analysed in this research are not planned.

The information in the column “CauseDesc.”, detailed in Table 4, refers to the causes of the failure pointed out by the utility or by the maintenance team that went to the event location. For a better analysis of the causes that cause failures in distribution systems, six groups were created according to the similarities found in these records. The “Not Identified” group gathers all outage events where the cause was not found and represents 16.94% of the failure data set. “Equipment Failure” and “Operational” are the events more directly linked to the utility’s system operation, where the first specifically representing interruptions caused by fail-

Table 1 Example for the failure events data set used in this research

Substation	Feeder	Lifetime	Type	CauseDesc.
Sub.A	Feeder1	136	Scheduled	Third party request
Sub.E	Feeder37	144	Accidental	Atmospheric discharge
...
Sub.B	Feeder12	168	Accidental	Capacitor bank

Table 2 Example for the atmospheric/weather events data set used in this research

Date	At.Dis.	Wind	Tmax	Tmin	Precip.	HR
01/02/2013	0	24	28.8	21.5	6.4	77
02/02/2013	195	37	32.6	21.2	0.7	77
...
31/12/2014	530	44	31	21.3	24.9	40

. “At.Dis.” are the number of atmospheric discharges, “Wind” are the wind speed in kilometre/hour, “Tmax” and “Tmin” are, respectively, the maximum and minimum temperature in °Celsius and “HR” are the relative humidity in %

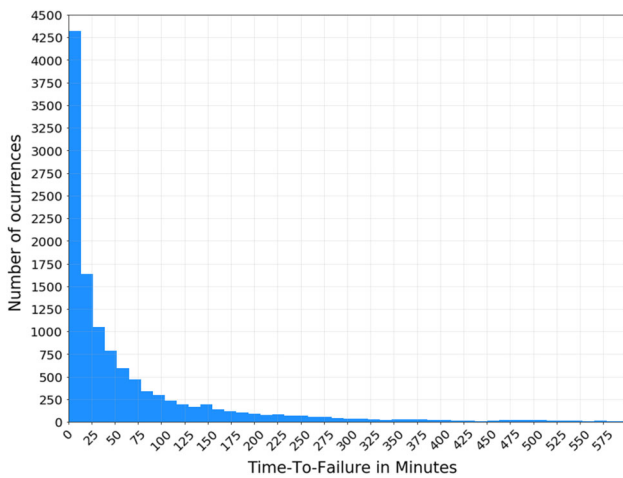


Fig. 3 Histogram for the most concentration of survival time of the DS analysed

ures in system components or the need for maintenance and the second aggregating events more linked to the operation itself, such as shutdown requests. The percentages of these groups in total failures are 13.77% and 45.9%, respectively. “Urban” is linked to situations involving human beings, such as accidents involving motor vehicles or vandalism and represents 7.71% of recorded interruptions. “Atmospheric” and “Environmental”, with the participation of 6.46% and 9.22%, respectively, are linked with situations involving external forces such as nature, where the first would be the events most linked to atmospheric events and their consequences, such as heavy rains that can cause erosion.

Table 3 The “Type” of failures observed in the failure data set

	Ex.1	Ex.2
Accidental	Animals/insects/birds	Branches touching the network (pruning)
Volunteer	Manoeuvres	Operational opening of another switch
Scheduled	Improvements or/and expansion	load transfer

2.1 Statistics

Reliability analysis was used with nonparametric and parametric techniques for each one of the proposed groups, and for considering all failures, to analyse the lifetime of a distribution system. Besides that, climatic events are associated with atmospheric and environmental failure groups, to observe the influence of different values for these variables in the survival of the analysed DS.

Reliability analysis has two main outputs (Davidson-Pilon et al. 2019) (Cox and Oakes 2018): reliability (survival) function, $R(t)$, and cumulative hazard function, $H(t)$. Reliability function is the probability that a system survives longer than time t , and cumulative hazard function is the accumulation of the hazard. (Hazard is the event rate at time t conditional on survival until time t or later ($R(t) = P(T \geq t)$) over time.) First, for reliability function, the nonparametric technique (Chowdhury et al. 2015) is called Kaplan–Meier (Kaplan and Meier 1958), and for cumulative hazard, Nelson–Aalen (Aalen 1978). In Eqs. 1 and 2, the general form of Kaplan–Meier and Nelson–Aalen estimators, respectively, is presented. The variable n_j is the number of samples at risk in the time t_j , and d_j is the number of occurred events at time t_j .

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) \tag{1}$$

$$\tilde{\Lambda}(t) = \sum_{j:t_j < t} \left(\frac{d_j}{n_j} \right) \tag{2}$$

Weibull model, in the form of the univariate model and the accelerated failure time (AFT), was used as the parametric technique. The univariate model does not include covariates and was used in conjunction with nonparametric techniques to analyse all failure groups. Equations 3 and 4 present the reliability and cumulative hazard functions, respectively. The shape parameter and the scale parameter are, respectively, ρ and λ , and t is the variable for lifetime.

$$R(t) = \exp \left(- \left(\frac{t}{\lambda} \right)^\rho \right), \lambda > 0, \rho > 0 \tag{3}$$

$$H(t) = \left(\frac{t}{\lambda} \right)^\rho \tag{4}$$

Table 4 The groups of failures observed in the failure data set

	Ex.1	Ex.2
Not identified	–	–
Equipment failure	Capacitor bank	Oil switch
Urban	Vandalism/theft	Bump
Operational	Third party request	Energy rationing
Environmental	Burnt/fire	Branches touching the network (pruning)
Atmospheric	Erosion	Atmospheric discharge

The AFT model was applied for three groups, all failures, atmospheric and environmental because the used covariates were weather events. Equations 5, 6 and 7 present the equations for the Weibull AFT model output functions. The coefficients of the regression model are represented by β_i variables, and the real values for the weather events are the x_i variables. The y variable in equations 6 and 7 represents that values for ρ are “independent”, meaning that it can include covariates or not (in this research, not included).

$$\lambda(x) = \exp(\beta_0 + \beta_1x_1 + \dots + \beta_nx_n) \tag{5}$$

$$R(t; x, y) = \exp\left(-\left(\frac{t}{\lambda(x)}\right)^{\rho(y)}\right) \tag{6}$$

$$H(t; x, y) = \left(\frac{t}{\lambda(x)}\right)^{\rho(y)}. \tag{7}$$

AFT models do the addition of covariates to regression models. The analysis of the values and some statistical criteria can make the relevance for the model of these covariates. The classical approach when evaluating covariates coefficients significance from regression models is the *pvalue* (Wang et al. 2019). *Pvalues* equal or lower than 0.05 are considered statistically significant, but this value has been discussed (Amrhein et al. 2019), so there is no need to discard covariates which have a higher value. Another statistical criteria generally presented are the standard error, that is the standard deviation of its sampling distribution, and the confidence interval, which is a range of values that can contain the real value of an unknown parameter. In this research, the 95% confidence interval is presented.

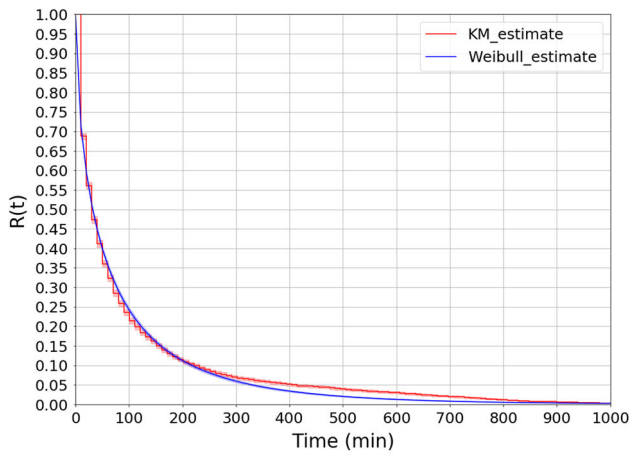
3 Results

Considering all interruptions from the data set exemplified in Table 1 as group called “All”, the nonparametric techniques of Kaplan–Meier estimator (reliability function $R(t)$) and Nelson–Aalen estimator (cumulative hazard function $H(t)$) were plotted together using the “All” group data together with the same data fitting a parametric model using Weibull distribution as shown in Fig. 4. When comparing nonparametric curves with parametric curves, an assessment is made of the

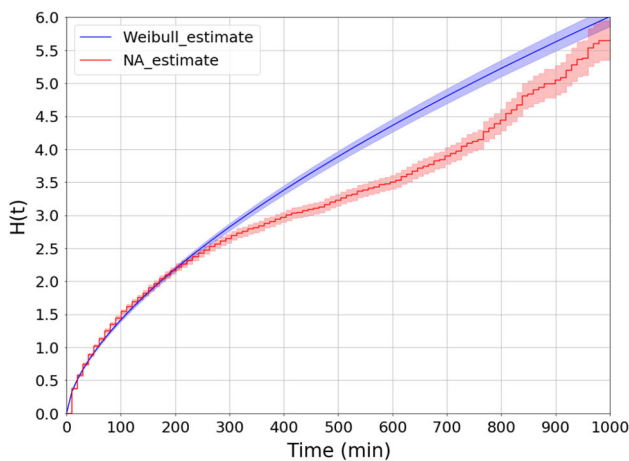
good fitting of the data to the parametric function in the sense that the closer the curves are, the better the fitting. Figure 4 shows a great proximity between the two curves for the two functions, but mainly for the reliability function, which is of greatest interest for this work. As the proximity between the curves can be observed, we can conclude that the Weibull model presents a good fit for the data used. The nonparametric techniques represent the real data, as some “downhill” can be observed, and the Weibull Model has a “smoothed” curve, being useful for any desired value, and being a more reliable model for forecasting.

The same approach was used for the six groups of failure events, comparing the nonparametric and parametric functions. Figures 6 and 7 show, respectively, the curves of reliability and cumulative hazard functions, first with the nonparametric estimators and in a second moment with the parametric technique using the Weibull distribution. Table 5 presents the Weibull distribution parameters for each group. In the matter of comparing the curves, the behaviour of the groups was similar for the two techniques. Evaluating the results for the groups themselves, a different risk order is observed than when considering only the failure quantity. According to Fig. 5, in terms of risk order regarding the number of failures, we have the following order: operational, not identified, equipment failure, environmental, urban and atmospheric. As the nature of the faults present in the group of unidentified faults is unknown, this group will be removed from the discussions.

But from Fig. 6, considering the period until 250 minutes, the risk order is: operational, atmospheric, not identified, equipment failure, urban and environmental. Considering both the reliability curves and the failure count for the groups, we have that operational interruptions are the ones that happen the most and happen frequently. On the other hand, despite the small number of faults related to atmospheric events, when they happen the tendency is to happen several faults in a short period of time. Unlike operational failures, which are largely planned and measures to reduce the impact on customers can be taken in advance, atmospheric failures have a greater unpredictability. In addition to greater difficulty in the cause of the failure and in what would be necessary to resolve it, critical atmospheric conditions also



(a)



(b)

Fig. 4 **a** Reliability function for the Weibull fitter, considering all causes of failures. **b** Cumulative hazard function for the Weibull fitter, considering all causes of failures

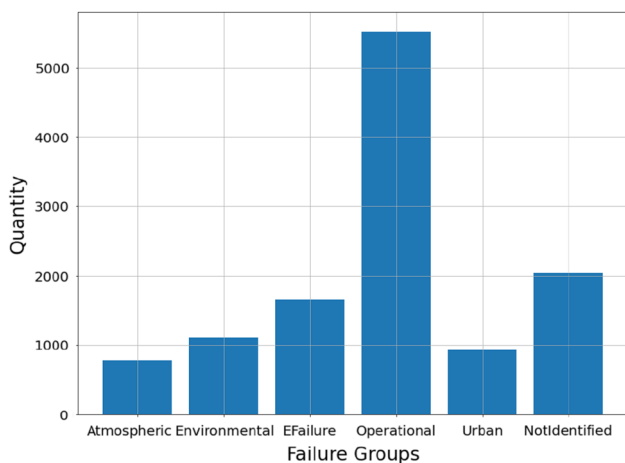
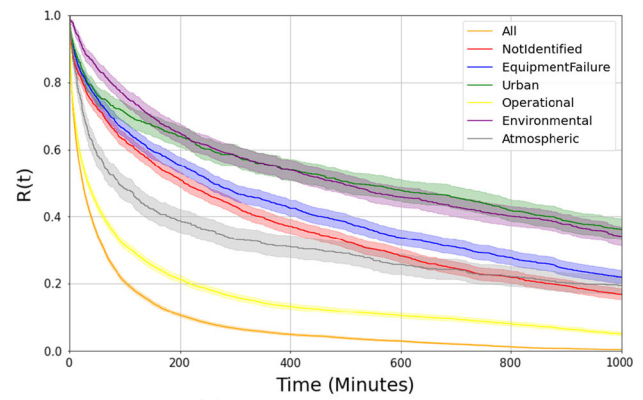
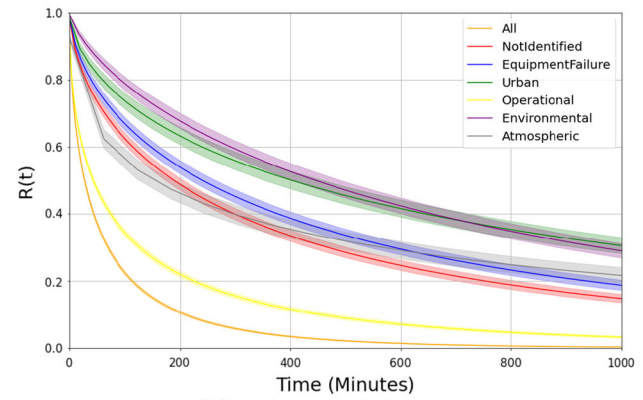


Fig. 5 Count of failures for each one of the proposed groups. EFailure represents the equipment failures group



(a) $R(t)$ for Kaplan-Meier Fitter.

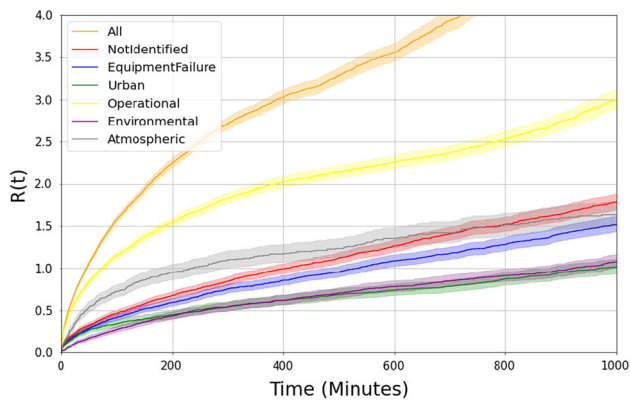


(b) $R(t)$ for Weibull Fitter.

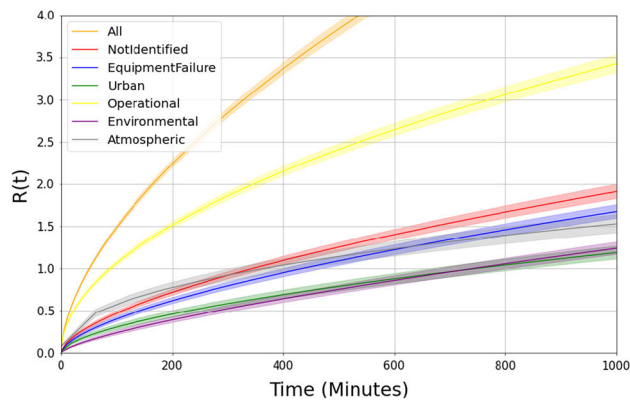
Fig. 6 **a** Reliability function graphic using nonparametric technique of Kaplan–Meier fitter and **b** using parametric technique of Weibull fitter, for the groups of failure causes

tend to make it difficult for maintenance crews to move and delay their service speed. As for the failures of urban and environmental groups, the behaviour is very similar both in quantity and in frequency of occurrence.

Considering that a large number of failures in a short period are as harmful as a large number of failures, studying the nature of failures involving atmospheric events is relevant. Using the accelerated failure time model with the Weibull distribution to include some climatic covariates, an assessment is aimed at whether such analysed atmospheric events are relevant to the occurrence of failures (in general, evaluating through the “All” group, such as more specifically, in the “Atmospheric” and “Environmental” groups, which in theory would be the most linked to the influence of these natural phenomena). The Weibull AFT model equations considering all failures are presented in Eqs. 8 and 9. “AD” is the number of atmospheric discharges, “W” is the wind speed, “Tx” is the maximum temperature, “Tn” is the minimum temperature, “P” is the amount of precipitation, and “HR” is the relative humidity. In Fig. 8, the $R(t)$ function is plotted, considering different values (chosen from random days of the used data set) for the three main covariates: wind



(a) $H(t)$ for Nelson-Aalen Fitter for each failure group.



(b) $H(t)$ for Weibull Fitter for each failure group.

Fig. 7 **a** Cumulative hazard function graphic using nonparametric technique of Nelson–Aalen fitter and **b** using parametric technique of Weibull fitter, for the groups of failure causes

speed, number of atmospheric discharges and relative humidity. For these curves, all other covariates were considered at their mean values. The baseline curve is considering all covariates at their mean values.

$$\lambda = \exp(5.18783 - 0.00049AD - 0.02995W + 0.00831Tx - 0.00302Tn - 0.00174P - 0.00308HR) \tag{8}$$

$$R(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^{-0.42509}\right) \tag{9}$$

In Eq. 8, a positive value, as in the intrinsic and maximum temperature coefficients, represents a positive contribution to the lifetime of the system. Therefore, a negative value, as for atmospheric discharges, wind speed, minimum temperature, amount of precipitation and relative humidity, decreases the survival time. (They are a threat for the system.)

From the Weibull AFT model, it is possible to predict in terms of the median (percentile) of survival and the expectation ($E[T|x]$). First, percentiles are measurements that divide

Table 5 Parameters of Weibull distribution for a reliability analysis regression technique for each of the proposed failure groups

Groups	λ	ρ
All	50.70832	0.58815
Not Identified	345.22319	0.60926
Equipment Failure	435.51710	0.62013
Urban	749.00735	0.59067
Operational	88.30617	0.50774
Environmental	742.92210	0.71807
Atmospheric	365.67945	0.42223

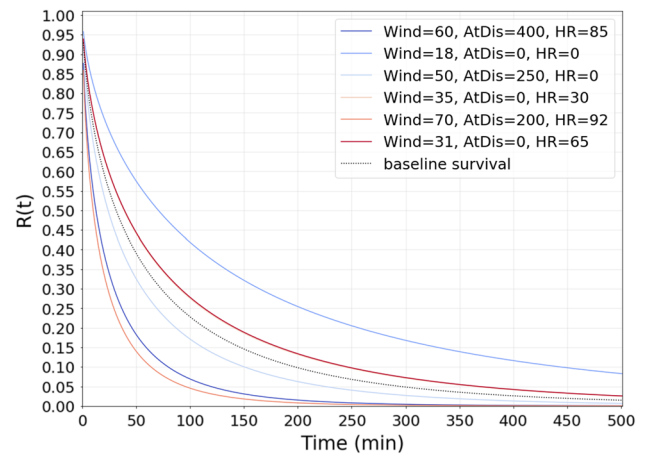


Fig. 8 Reliability function plot for Weibull AFT regression model, varying the wind speed, the number of atmospheric discharges and the value for relative humidity, and keeping the mean value for the other variables. The baseline survival is considering all covariates as their mean value. AtDis is the atmospheric discharges, and HR is the relative humidity

the sample in ascending order of data into 100 parts, each with an approximately equal percentage of data, and the median is the 50^o percentile. Second, the expectation of a random variable is the sum of the product of each probability of leaving the analysis by its respective value, representing the “expected” average value of an experiment if repeated many times. In Table 6 is presented these predictions for real values of the weather data.

The behaviour of climate covariates was made fitting the Weibull AFT regression model for the atmospheric and environmental failure groups, separately. The coefficients for the covariates are presented in Tables 8 and 10. The values of the coefficients of the covariates can be interpreted multiplying them by the maximum and average values that the variables (weather events) can assume. In Table 7, this operation was made using 2000 and 90 for atmospheric discharges (AD), 100 and 40 for wind speed (W), 40 and 25 for maximum temperature (Tx), 25 and 15 for minimum temperature (Tn), 80 and 10 for precipitation (P) and 100 and 60 for relative humidity (HR). These values are near the maximum and the mean

Table 6 Prediction of the percentile and the expectation of the lifetime of the analysed distribution system for real values of the weather events, considering all failure events

Percentile	Expectation	AD	W	Tx	Tn	P	HR
46.2876	110.0950	0.0	24.0	28.8	21.5	6.4	77.0
29.7015	70.6451	195.0	37.0	32.6	21.2	0.7	77.0
7.9735	18.9651	108.0	78.0	23.2	19.0	59.2	95.0
44.3196	105.4141	0.0	25.0	29.2	16.0	41.2	68.0
33.57057	79.8474	0.0	37.0	29.8	17.4	0.0	65.0

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

Table 7 Product between the coefficient values for the three presented AFT models and the maximum and mean values of the weather events

	All		Atmospheric		Environmental	
	Max	Mean	Max	Mean	Max	Mean
AD	0.9800	0.0441	2.2800	0.1026	0.0800	0.0036
W	2.9950	1.1980	3.3560	1.3424	1.4820	0.5928
Tx	0.3324	0.20775	1.5528	0.9705	0.3504	0.2190
Tn	0.0755	0.04530	0.06170	0.0371	1.6323	0.9794
P	0.1392	0.0174	1.2496	0.1562	0.7256	0.0907
HR	0.3080	0.1848	0.4140	0.2484	0.4040	0.2424

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

values for the weather events considering all analysed days. This operation was done for the coefficients of the three presented models (all, atmospheric and environmental failures groups), and the resulting values are presented in “module”.

The percentile and the expectation for these two groups are shown in Tables 13 and 12, with real values for the weather events. (The weather values differ from each other because they are from events associated with the failure group. So, the values of weather events presented for each table are from days that occurred a failure of the analysed group, either an atmospheric failure or an environmental failure.) To maintain the same prediction logic, the weather events that presented a higher *pvalue* in the atmospheric and failure groups regression models were kept.

In Tables 9 and 11, some statistical criteria for the covariates coefficients are presented. For the atmospheric group,

Table 8 Coefficient values for a Weibull AFT regression for atmospheric failure events in the analysed distribution system

	Intercept	AD	W	Tx	Tn	P	HR
λ	7.18672	-0.00114	-0.03356	0.03882	-0.00247	-0.01562	-0.00414
ρ	-0.79611						

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

Table 9 *P* values, standard errors (*se*) and confidence interval of the covariates of the Weibull AFT regression parameters for the atmospheric failure events

	<i>p</i>	<i>se</i>	Lower 0.95	Upper 0.95
AD	$< 5e^{-6}$	0.00025	-0.00163	-0.00065
W	$< 5e^{-6}$	0.00552	-0.04439	-0.02274
Tx	0.13892	0.02623	-0.01259	0.09023
Tn	0.93892	0.03228	-0.06574	0.06079
P	0.01499	0.00642	-0.02821	-0.00303
HR	0.15167	0.00289	-0.00981	0.00152

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

atmospheric discharges, wind speed and precipitation were the covariates with a *pvalue* lower than 0.05. For the environmental failure events, minimum temperature, precipitation and relative humidity are the statistically significant covariates according to their *pvalues*. It is possible to observe that with the confidence interval cross zero, the *pvalue* ends up returning a value above 0.05, that is not desirable. However, if the largest portion of the confidence interval focuses on a positive or negative value, it turns out to be odd to disregard the covariate just because of the value of *pvalue*.

The modelling of “ ρ ” for all covariates is not done due to the characteristics of the weather data, that are independent events. (One climatic event value is not directly related to the value of another weather event.) The intercept value increased for both atmospheric and environmental groups, as expected. In terms of the covariates, in the atmospheric group, for all threats (coefficients with a negative value) the value increased considerably, but for the environmental group, some weather variables decreased (as wind speed and atmospheric discharges) and others presented a drastic increase (precipitation and minimum temperature). The prediction of the lifetimes has higher values when compared with the failure model for all groups, with extreme weather events impacting more in the lifetime considering the atmospheric group than in the environmental group.

From Table 7, considering that for the three models only maximum temperature (Tx) presented a positive coefficient that has the meaning of a positive contribution for the lifetime of the system, it can be observed that wind (W) is the most dangerous threat, due to a product bigger than 1 for almost all

Table 10 Coefficient values for a Weibull AFT regression parameters for environmental failure events in the analysed distribution system

	Intercept	AD	W	Tx	Tn	P	HR
λ	8.36078	-0.00004	-0.01482	0.00876	-0.06529	-0.00907	-0.00404
ρ	-0.27167						

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

Table 11 *P* values, standard errors (*se*) and confidence interval of the covariates of the Weibull AFT regression for the environmental failure events

	<i>p</i>	<i>se</i>	Lower 0.95	Upper 0.95
AD	0.86452	0.00025	-0.00054	0.00045
W	0.00001	0.00341	-0.02150	-0.00814
Tx	0.51777	0.01355	-0.01779	0.03531
Tn	0.00005	0.01604	-0.09674	-0.03385
P	0.01541	0.00374	-0.01641	-0.00173
HR	0.01646	0.00168	-0.00733	-0.00074

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

Table 12 Prediction of the percentile and the expectation of the lifetime of the analysed distribution system for real values of the climatic events, considering atmospheric failure events

Percentile	Expectation	AD	W	Tx	Tn	P	HR
328.0820	1822.7876	195.0	37.0	32.6	21.2	0.7	77.0
20.1203	111.7864	108.0	78.0	23.2	19.0	59.2	95.0
299.9608	1666.5494	0.0	25.0	29.2	16.0	41.2	68.0
234.9093	1305.1307	165.0	40.0	26.9	19.4	0.0	88.0
126.1716	700.9959	138.0	44.0	25.4	19.0	27.8	93.0

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

Table 13 Prediction of the percentile and the expectation of the lifetime of the analysed distribution system for real values of the weather events, considering environmental failure events

Percentile	Expectation	AD	W	Tx	Tn	P	HR
162.4786	308.8923	108.0	78.0	23.2	19.0	59.2	95.0
433.7354	824.5855	0.0	25.0	29.2	16.0	41.2	68.0
489.8989	931.3593	0.0	37.0	29.8	17.4	0.0	65.0
407.7476	775.1792	6.0	44.0	23.7	17.8	0.0	65.0
374.5182	712.0061	165.0	40.0	26.9	19.4	0.0	88.0

Acronyms: AD for atmospheric discharges, W for wind speed, Tx for maximum temperature, Tn for minimum temperature, P for precipitation and HR for relative humidity

situations. Wind also presented a higher contribution for the decrease in the lifetime, considering the “maximum” value on the atmospheric failure group model. Assuming a crite-

rium that considering the product of the coefficient with the maximum value of the variable must be greater than 0.1, the minimum temperature for all failures and atmospheric failures group and atmospheric discharges for the environmental failure group presented a “null” impact in these situations.

4 Conclusion

Through reliability analysis regression models, an evaluation of the lifetime of a real Brazilian DS due to different factors was made in this research. First, utilizing nonparametric (Kaplan–Meier and Nelson–Aalen) and parametric (Weibull) techniques, the curves for reliability and cumulative hazard function were presented, considering all registered failures from a period of almost 2 years. As both curves presented a similar behaviour, especially in the time period where most of the survival times of the analysed DS are concentrated, the parametric model represents the data well.

The two techniques were applied for each of the different failure groups: not identified, equipment failure, urban, operational, environmental and atmospheric. The threat level has a different order than the number of occurrences order for the failure groups, which ends up highlighting the group of atmospheric failures. However, for the context analysed, operational failures are the principal threat in terms of both the number of failures and the time frame in which they occur.

Using the Weibull AFT regression model, an analysis of the atmospheric and environmental failure groups, which are most linked with weather events, was performed using six weather events with their maximum daily values. From the model of the all failures groups, reliability function was plotted using different real values of the significant weather events, demonstrating the impact on survival time given extreme weather events. A discussion was made about the different coefficients values of the weather events covariates, from the all, atmospheric and environmental failure groups. For the atmospheric group, all coefficients of negative variables (threats) increased. For the environmental failure group, some coefficients decreased, but other presented an increase, and specifically, precipitation presented a great increase.

Predictions for the median (percentile) and the expectation of the lifetime of the system were made for all failures, atmospheric and environmental groups, considering some

real values of the weather data. The principal difference was observed in the atmospheric group predictions, where the expected values can be much higher than the percentile values.

Considering the most generic scenario of any distribution network to more specific cases, such as interruptions involving atmospheric events, the models proposed in this work aimed to present a new technique regarding the issue of failures in electrical energy distribution systems. Instead of the classic use of a random number of failures for simulations in areas such as maintenance, operation and expansion of distribution systems, providing a temporal approach can bring several benefits, especially when considering the possibility of adding external variables to the model. The fact that the model was created using real data of interruptions in a distribution system seeks to represent more faithfully to what is found in the real operation of electrical systems. In future works, the time-between-failure model presented in this paper can be integrated with failure duration models, using the same approach of using an overview or getting into the question of the causes of failures for more specific problems, a tool that generates a complete scenario of outages in electricity distribution systems and that reflects its operation in the real world.

References

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4), 701–726. <https://doi.org/10.1214/aos/1176344247>.
- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>.
- Bessani, M., Fanucchi, R., Achcar, J., Maciel, C. (2016). A statistical analysis and modeling of repair data from a Brazilian power distribution system. In: Proceedings of international conference on harmonics and quality of power, ICHQP 2016-December, pp. 473–477. <https://doi.org/10.1109/ICHQP.2016.7783446>
- Brem, S. (2015). Critical infrastructure protection from a national perspective. *European Journal of Risk Regulation*, 6(2), 191–199.
- Chowdhury, F., Gulshan, J., & Hossain, S. (2015). A comparison of semi-parametric and nonparametric methods for estimating mean time to event for randomly left censored data. *Journal of Modern Applied Statistical Methods*, 14(1), 196–207. <https://doi.org/10.22237/jmasm/1430453760>.
- Colosimo, E., & Giolo, S. (2006). Análise de sobrevivência aplicada. Edgard Blücher.
- Cox, D., & Oakes, D. (2018). *Analysis of survival data*. <https://doi.org/10.1201/9781315137438>.
- Dantas, M., Valença, D., da Silva, Platiny, Freire, M., Medeiros, P., Da Silva, D., & Aloise, D. (2010). Weibull-regression models to study failure data in oil pumps. *Produção*, 20, 127–134.
- Davidson-Pilon, C., Kalderstam, J., Zivich, P., Kuhn, B., Fiore-Gartland, A., Moneda, L., Wilson, D., Parij, A., Stark, K., Anton, S., Besson, L., Gadgil, H., Golland, D., Hussey, S., Kumar, R., Noorbakhsh, J., Klintberg, A., Ochoa, E., Albrecht, D., Medvinsky, D., Zgonjanin, D., Katz, D.S., Chen, D., Ahern, C., Fournier, C., Rendeiro, A.F. (2019). Camdavidsonpilon/lifelines: v0.22.3 (late). <https://doi.org/10.5281/zenodo.3364087>.
- Economic Advisers PC, the US Department of Energy's Office of Electricity Delivery, Reliability E (2014) Economic benefits of increasing electric grid resilience to weather outages, vol 2
- Fogliatto MSS, Santos TMO, Bessani M, Maciel CD (2019) Survival analysis of electrical power distribution systems using weibull regression. Simpósio Brasileiro de Automação Inteligente.
- IEEE Std. (2004). Ieee guide for electric power distribution reliability indices. IEEE Std 1366-2003 (Revision of IEEE Std 1366-1998) pp. 1–50. <https://doi.org/10.1109/IEEESTD.2004.94548>.
- Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.1080/01621459.1958.10501452>.
- Konal M, Öz I, Uzunoğlu C, & Kaçar F (2018) Electrical distribution network's failure analysis based on weather conditions. 2018 5th International Conference on Electrical and Electronics Engineering, ICEEE 2018, pp. 269–272. <https://doi.org/10.1109/ICEEE2.2018.8391344>
- Li, Q., Gao, J., & Flowers, G. (2019). Analysis of electromagnetic behaviors induced by contact failure in electrical connectors. *Microwave and Optical Technology Letters*, 61(11), 2579–2585. <https://doi.org/10.1002/mop.31925>.
- Mishra, P., Pandey, C., Singh, U., Keshri, A., & Sabaretnam, M. (2019). Selection of appropriate statistical methods for data analysis. *Annals of Cardiac Anaesthesia*, 22(3), 297–301. https://doi.org/10.4103/aca.ACA_248_18.
- Moolgavkar, S., Chang, E., Watson, H., & Lau, E. (2018). An assessment of the cox proportional hazards regression model for epidemiologic studies. *Risk Analysis*, 38(4), 777–794.
- Murthy, D. P., Bulmer, M., & Eccleston, J. A. (2004). Weibull model selection for reliability modelling. *Reliability Engineering and System Safety*, 86(3), 257–267.
- Ni, M., & Li, M. (2019). Reliability assessment of cyber physical power system considering communication failure in monitoring function. In 2018 international conference on power system technology, POWERCON 2018 - Proceedings, pp. 3010–3015. <https://doi.org/10.1109/POWERCON.2018.8601964>.
- Pahwa, A. (2007). Modeling weather-related failures of overhead distribution lines. (2007). IEEE Power Engineering Society General Meeting. *PES*. <https://doi.org/10.1109/PES.2007.386167>.
- Radmer, D., Kuntz, P., Christie, R., Venkata, S., & Fletcher, R. (2002). Predicting vegetation-related failure rates for overhead distribution feeders. *IEEE Transactions on Power Delivery*, 17(4), 1170–1175. <https://doi.org/10.1109/TPWRD.2002.804006>.
- Rink, M., Kluth, L., Shariat, S., Fisch, M., Dahlem, R., & Dahm, P. (2013). Kaplan-meier analysis in urological practice [kaplan-meier-analysen in der urologischen praxis]. *Urologe - Ausgabe A*, 52(6), 838–841. <https://doi.org/10.1007/s00120-013-3150-4>.
- Rodriguez-Garcia, L., Perez-Londono, S., & Mora-Florez, J. (2019). An optimization-based approach for load modelling dependent voltage stability analysis. *Electric Power Systems Research*. <https://doi.org/10.1016/j.epr.2019.105960>.
- Sahai, S., & Pahwa, A. (2006). A probabilistic approach for animal-caused outages in overhead distribution systems. In 2006 9th international conference on probabilistic methods applied to power systems, PMAPS <https://doi.org/10.1109/PMAPS.2006.360321>
- Shauly, M., Rabinowitz, G., Gilutz, H., & Parmet, Y. (2011). Combined survival analysis of cardiac patients by a cox ph model and a markov chain. *Lifetime Data Analysis*, 17(4), 496–513. <https://doi.org/10.1007/s10985-011-9196-y>.
- Sroka, K., & Złotecka, D. (2019). The risk of large blackout failures in power systems. *Archives of Electrical Engineering*, 68(2), 411–426. <https://doi.org/10.24425/aee.2019.128277>.

- Wang, B., Zhou, Z., Wang, H., Tu, X., & Feng, C. (2019). The p-value and model specification in statistics. *General Psychiatry*. <https://doi.org/10.1136/gpsych-2019-100081>.
- Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine* 4(24).
- Zio, E. (2009). Reliability engineering: Old problems and new challenges. *Reliability Engineering and System Safety*, 94(2), 125–141. <https://doi.org/10.1016/j.res.2008.06.002>.
- Złotecka, D., & Sroka, K. (2018). The characteristics and main causes of power system failures basing on the analysis of previous blackouts in the world. 2018 International Interdisciplinary PhD Workshop. *IIPhDW, 2018*, 257–262. <https://doi.org/10.1109/IIPHDW.2018.8388369>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.