This is the published version of a paper presented at *2018 Power Systems Computation Conference (PSCC)*.

Assessment of Explanatory Variables on the Failure Rate of Circuit Breakers Using the Proportional Hazard Model

Jan Henning Jürgensen
Lars Nordström
Patrik Hilber
School of Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden
{jhjur, larsno, hilber}@kth.se

Elin Andreasson
Anna Lilly Brodersson
Vattenfall Eldistribution AB
Stockholm, Sweden
{elin.andreasson, annalilly.brodersson}@vattenfall.com

Abstract—This paper utilises the proportional hazard model to understand and quantify the impact of explanatory variables on the failure rate of circuit breakers (CB). Particularly, 4496 work orders with 2622 high voltage CBs are investigated with an occurrence of 281 major failures. Different explanatory variables such as CB type, manufacturer, preventive maintenance (PM), and others are gathered to quantify their significance and magnitude of their effect. The results present that PM has a positive impact, the number of operations within the last year a negative impact, and age has a small but negative impact on the failure rate. The CB type is not significant in all analyses which can be explained by examining the PM and age of these CB types. This paper contributes to the understanding of how explanatory variables impact the failure rate which is essential for power system asset management.

Index Terms—Asset management, circuit breaker reliability, failure rate, preventive maintenance, proportional hazard model.

I. INTRODUCTION

Power system asset management (AM) is a system of strategies, decision-making, and prioritization with the overall aim of achieving a "lifetime optimum" under the constraints of acquisition, use, maintenance, and disposal of assets [1]. Generally, it is a business-driven paradigm but practically a data-driven process that requires data to execute planning and operation decisions [1]. For this data-driven process, component reliability data is essential for reliability calculations, risk analysis [2], network reliability modelling and optimal maintenance planning [3] which are the base for the AM strategies.

One important component reliability measure is the failure rate which describes the probability of failure in the next predefined interval. Failure rate estimation is a part of statistical data-driven processes [4] that require failure data and the respective survival time of the components. Due to the long lifetime of equipment, the capital intensive data gathering and documentation, and the poor quality of existing failure data,

an average failure rate is traditionally assigned to all components [5]. However, this approach neglects the heterogeneity of components, such as component specific characteristics, and therefore does not distinguish between population and individual failure rates [6] which causes over- or underestimation. Thus, [5] argues that every equipment type's failure rate should be a function of critical parameters such as age, manufacturer, voltage, size, and maintenance. One statistical model which investigates the impact of explanatory variables on the failure rate is the relative risk or cox model [7]. It has gained high popularity, especially, in medical sciences [8] and is mostly known as proportional hazard model (PHM) when the investigated covariates are constant over time. The high data input requirements of the PHM such as occurred equipment failures, survival time, and explanatory variables has led to only a few applications in the power system domain [6]. Explanatory variables such as *season*, *weather*, and *time trend* or *weather*, *cause*, and *voltage* on failure and repair data of transmission statistics in the United Kingdom are assessed in [9] and [10], respectively. Whereas [9] investigates transmission failures in general, [10] focuses on the analysis of component specific failures of lines, circuit breakers (CBs), protection equipment, transformer, and isolators. Here, the explanatory variables *cause* and *voltage* are not significant but *weather*, *season*, and *time trend* are for some equipment types. Only time is a significant explanatory variable for CB failures in the North Wales dataset. Another improvement towards a failure rate based on critical parameters is [11] where early cable failures are analysed with the a non-parametric PHM. Reference [11] assess the impact of *installer*, *joint manufacturer*, and *cable manufacturer* to essentially identify two manufacturer who have a significant effect on the cable joint failures with 7.711 and 34.609 compared to a reference manufacturer. An in-depth study on a disconnector population is conducted in [12] where the authors show that *remote control availability* has a negative impact but *preventive maintenance* a positive impact on the failure rate. Moreover, it is

shown that single pole disconnectors have a 9.37 times higher failure rate compared to double side break disconnectors in this study. Some additional PHM applications in the power system domain are summarized in [6]. The aforementioned papers deliver results to model the failure rate as a function of critical component parameters, however, as more data is available the PHM offers the opportunity to gain additional insights into more equipment types.

This article continues with the exploratory failure data analysis and focuses on a CB population in Sweden. Particularly, the explanatory variables *preventive maintenance*, *manufacturer*, *operating mechanism*, *geographical area*, and *number of operations* are investigated towards their impact on the failure rate.

The next section briefly discusses the exploratory data analysis approach with the proportional hazard model. Section III describes the CB dataset and the assumptions made to conduct the study. In the following, section IV and V present the results and discussion of the study and section VI the concluding remarks.

II. EXPLORATORY FAILURE DATA ANALYSIS WITH THE PROPORTIONAL HAZARD MODEL

The PHM is a relative risk model which belongs to the class of regression models in survival analysis. This model assesses explanatory variables, also known as covariates, which might have an impact on the baseline failure rate of component i. Lets assume that T is the survival time and x_i is the vector of covariates, the general form of the relative risk model is

$$
\begin{aligned}
\lambda(t; x_i) \quad &= \lim_{\Delta t \longrightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t, x_i)}{\Delta t} \\
&= \lambda_0(t)\, r(t, x_i), \quad t > 0
\end{aligned}
\tag{1}
$$

with $\lambda_0(t)$ as the baseline failure rate. The most common approach is to specify the relative risk function with an exponential form when defining the PHM. Then, the PHM is given with

$$
\lambda(t; x_i) = \lambda_0(t)\, e^{(Z_i(t)\beta)}
\tag{2}
$$

where β is the vector of regression parameters and $Z_i(t)$ is a vector of covariates derived from the basic covariate vector x_i. The PHM is a technique which has been treated extensively in the literature and thus is well described and documented. Examples of these literature are [8] for an extensive theoretical treatment, [13] for a self-learning text with straightforward illustrations, or [14] which focuses on censored and truncated survival data. Therefore, the process of the exploratory data analysis is described briefly here from data requirements to interpretation of the results. The process can be divided into: data acquisition, covariate coding, cox regression, and the interpretation of the results.

A. Data Acquisition

The PHM requires a data triplet of the form $(T_i, \delta_i, Z_i(t)), i = 1, \ldots, n$ where δ_i is the failure indicator and n the population size. The basic covariates x_i are initially selected based on their potential risk on component i. In this case, knowledge of the impact of certain characteristics, measurements, and environmental factors on components is necessary to do the first selection of covariates.

B. Covariate Coding

Covariates which might be investigated are either continuous, categorical, or binary. They need to be known from the start of the study and are time-independent if they do not change during the study period and time-dependent if they do. This crucial distinction is very important because the methods to analyse time-dependent covariates are different. Generally, the coding of the covariates is an essential part because this is the base for the later interpretation. Since the PHM assumption must be fulfilled which describes that the hazard ratio (HR) must be constant over time, certain continuous covariates might be transformed. Reference [13] discusses the coding of the three covariate types in more detail.

C. Cox Regression

This part is the actual investigation of the impact and the significance of the covariates which is similar to a linear regression model [13]. In this setting, the survival time and failure indicator would be the dependent variables whereas the covariates are independent variables. The null hypothesis is stated with $\beta = 0$ which describes that the covariate has no impact on the failure rate. To test the significance of the covariates, the p-value test is commonly used. Here, the null hypothesis can be rejected if the significance level is below $\alpha = 0.05$. The assessment can be conducted in a univariate and multivariate analysis. The latter requires a covariate selection procedure such as forward, backward, or stepwise regression which is discussed in [15].

D. Interpretation of the Results

In the PHM, the covariates are constant over time and a possible form of interpretation of β is the use of the ratio between two individuals 1 and 2 defined as the hazard ratio (HR)

$$
HR = \frac{\lambda_0(t)\, e^{(Z_2(t)\beta)}}{\lambda_0(t)\, e^{(Z_1(t)\beta)}} = e^{((Z_2(t) - Z_1(t))\beta)}
\tag{3}
$$

If we assume $Z_1(t) = 0$, the failure rate of i-th individual is proportional to $\lambda_0(t)$ with $e^{(Z_i(t)\beta)}$. Depending on the form of the covariate, the interpretation differs. In general, if HR $= e^{(Z_i(t)\beta)} = 1$ the covariate has no impact, HR < 1 the covariate has a positive impact, and HR > 1 has a negative impact on the failure rate. In addition, the confidence interval (CI) of the HR should always be considered while interpreting the results.

III. CIRCUIT BREAKER DATASET

CBs are mechanical switching devises to control and protect the power grid. Since they protect the power grid from overloading, short-circuit faults, and other faults, CBs are essential for operating the power grid. The importance of these
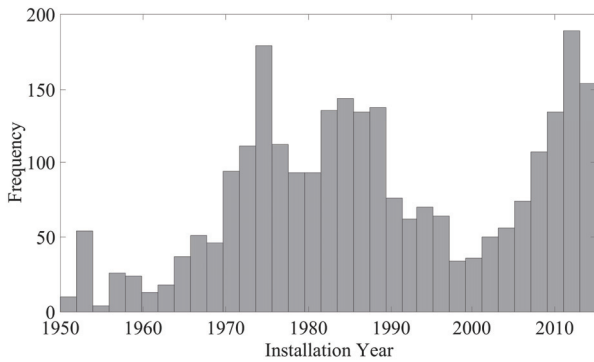
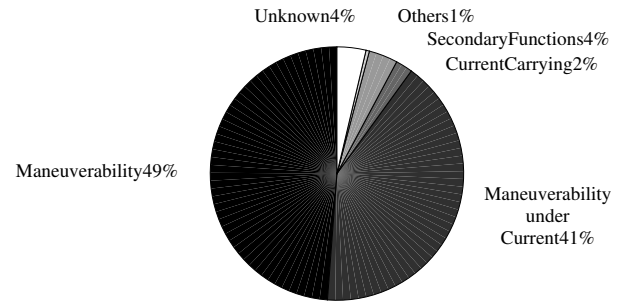Fig. 1. Distribution of CBs installation years

Fig. 2. Frequency of failure modes of 461 major failures in CB population

Fig. 3. Frequency of failure locations of 461 major failures in CB population

switching devises makes the requirement for understanding which covariates have an effect on the failure rate even more important. In this paper, we examine a larger CB population in Sweden.

A. Dataset Description

The population includes 2622 CBs with a total of 4496 work orders which have been recorded from 2008 to 2015. General information of the CBs and work orders are extracted from the internal asset management system of the utility. The work orders include information such as work order type, completed task, date, CB type, CB manufacturer and specific manufacturer type. The study is limited to CBs on voltage levels from 40 kV to 400 kV which are produced by 18 different manufacturers. The population can be classified into 3 different CB types: Oil CBs (52.9 %), SF6 CBs (45.2 %), and vacuum CBS (1.9 %). The installation years of the population are from 1950 to 2015 and illustrated in Fig. 1. These CBs are installed in four different regions in Sweden where the climate varies considerably.

B. Selected Data and Assumptions

1) Data Requirements: The dataset needs to be consistent and complete to be investigated with the PHM. Thus, assumptions are necessary to code the data and missing or incorrect data must be estimated. Thus, this section states all assumptions made to conduct the study. Moreover, missing data are estimated with the multiple imputation using regression switching imputation with predictive mean matching method (MI-MICE-PMM). The MI-MICE-PMM approach is described in [16], [17].

2) Failures: A total of 1658 work orders are identified which describe CB failures. With the definition of major and minor failures according to [18], the failures can be divided into 461 major and 1197 minor failures. The major failures are categorized after failure modes in Fig. 2 and failure locations in Fig. 3. In this study, only major failures are assumed to be equipment failures. Moreover, this study is limited to single failures which means that a CB is removed from the study after a major failure. Neglecting recurrent failures, 281 major
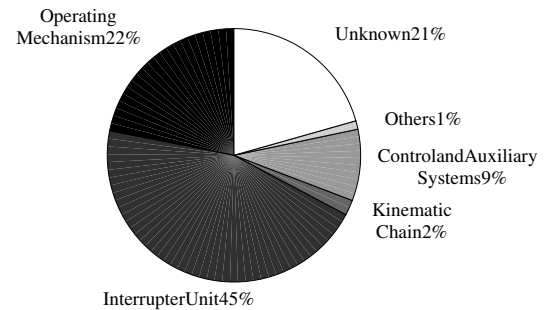
failures are considered in this study which results in a dataset with 89.3 % censoring.

3) Time to Failure: Two approaches are commonly used as time to failure, depending on when an individual is exposed to risks. Firstly, the time to failure can be measured from the beginning of a study to a failure or the end of the study (censored). Secondly, the time to failure is measured from the point the component is exposed to risks such as the point it is installed in the power grid. Consequently, age of a CB seems to be the reasonable choice as time to failure. Reference [12] mainly used the time on study as time to failure since 31 % of the installation years are missing and no further knowledge about the disconnectors is available before study entry. This study faces the same challenge. Since there is no information regarding the CB before 2008, the data is known to be left truncated. Thus, this must be considered when age is used as time to failure.

4) Covariates: A set of covariates is gathered to investigate their impact on the failure rate. These covariates are manufacturer, CB type, operating mechanism (OM) geographical area (GA), preventive maintenance (PM) as maintenance intensity (MI), voltage level (VL), number of operations (average per year (M#O), within the last year (#OLY)), remote control availability, and age at admission. The covariate PM is defined as MI which is number of maintenance tasks per time in study [12]. It is defined as intensity since as longer as a

TABLE I
CODING OF CATEGORICAL COVARIATES

Covariate	Abbreviation	Value	Category	Coding		
Circuit Breaker Type	CB Type	Oil	1	1	0	
		SF6	2†	0	0	
		Vaccum	3	0	1	
Voltage Level in $[kV]$	VL	$(x \le 80)$	1†	0	0	
		$(80 < x \le 134)$	2	1	0	
		$(134 < x)$	3	0	1	
Operating Mechanism	OM	Hydraulic	1	1	0	
		Mechanical Spring	2†	0	0	
		Watch Spring	3	0	1	
Geographical Area	GA	Region 1	1	1	0	0
		Region 2	2	0	1	0
		Region 3	3	0	0	1
		Region 4	4†	0	0	0
Maintenance Intensity in $\left[\frac{\text{\# maintenance tasks}}{\text{time-on-study}}\right]$	MI	$x = 0$	0†	0	0	0
		$(0 < x \le 0.2)$	1	1	0	0
		$(0.2 < x \le 0.4)$	2	0	1	0
		$(0.4 < x)$	3	0	0	1
Number of Operations within the last Year	#OLY	$x \le 10$	0†	0 0 0 0		
		$(10 < x \le 20)$	1	1 0 0 0		
		$(20 < x \le 40)$	2	0 1 0 0		
		$(40 < x \le 60)$	3	0 0 1 0		
		$(60 < x)$	4	0 0 0 1		

† reference group

CB is in the study as more maintenance tasks it can receive. Generally, maintenance is defined as a task to prevent failure and degradation of the equipment [19]. Thus, minor failures, which according to the definition of [18] do not lead to major failures, are also counted as maintenance tasks.

IV. RESULTS

The dataset is explored and analysed with the PHREG function of the software tool SAS University Edition 3.6 (SAS Institute, Cary NC). The exploratory approach of the failure data analysis assesses the covariates in a univariate and multivariate setting with varying failure and time to failure definitions. This approach enables a wider understanding of the dataset and prevents misinterpretation.

A. Data Preparation and Covariate Coding

The PHM requires that all input parameters $(T_i, \delta_i, Z_i(t))$ are consistent and complete. Fig. 2 and 3 show that also unknown or incomplete data exist in this dataset. Whereas unclassified failure data can be tolerated in the PHM when competing risks are not considered, covariates and time to failure data must be complete. Thus, the MI-MICE-PMM algorithm is applied to estimate 2.05 % of the missing installation years, 2.05 % of the remote control availability, 10.64 % of the OM, and 10.5 % of the CB types. The two aforementioned time to failure T_i approaches are used in this analysis. Generally, time-on-study in days is used and age is adjusted as a linear continuous covariate because there is no information about the CB before 2008 and using age as time to failure would bias the results. However, the second approach is age as time to failure and adjusting for left truncation in the dataset. In this case, age is recorded in years.
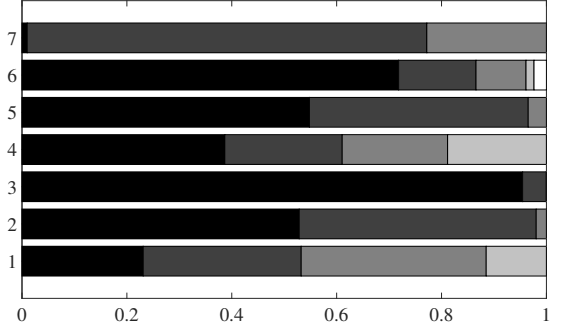

Fig. 4. Relative Group size of categorical sub classes of covariates with 1 - GA, 2 - CB Type, 3 - remote control availability, 4 - MI, 5 - VL, 6 - #OLY, and 7 - OM
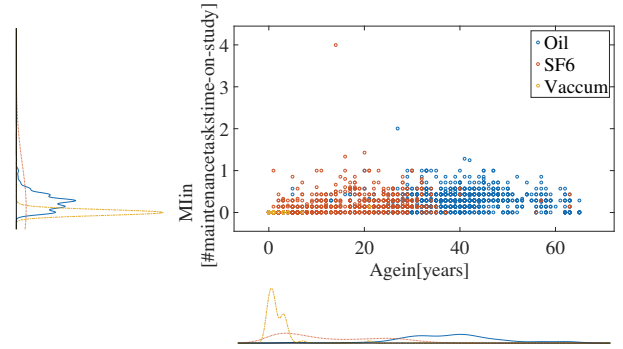

Fig. 5. Relationship between Maintenance Intensity and Age at Event depending on CB type

The covariates investigated in this study are all categorical covariates except age at admission and M#O. The abbreviation, the categories, and the coding are displayed in Table I and the frequency of the sub-classes is illustrated in Fig. 4. The covariates maintenance intensity (MI), which is PM as number of maintenance tasks per time-on-study, and number of operations last year (#OLY) are continuous but transformed into categorical covariates because the functional form is not linear and thus the proportional assumption of the PHM would be violated. The functional form of age at admission is linear which can be seen by assessing the martingale residuals.

The pearson correlation coefficients of days in study (DiSt), age, and other continuous covariates are shown in Table II. A moderate correlation can be noted between age and MI and M#O and #OLY whereas others are generally rather weakly correlated. The relationship between MI, age, and CB type is illustrated in Fig. 5. Table III presents the results of the one-way ANOVA to test whether the means of the MI conducted for each CB type are different. The results show that the means are not equal, hence, there is a difference in conducted MI for each CB type which is graphically illustrated in Fig. 6. Conducting the same test for GA instead of CB type resulted in no significant difference.

TABLE II
PEARSON CORRELATION COEFFICIENTS, N=2622

	DiSt	Age	MI	#OLY	VL	M#O
Days in Study (DiSt)	1.000	-0.011	-0.001	-0.120	-0.082	-0.052
Age	-0.011	1.000	0.390	0.023	-0.126	0.000
MI	-0.001	0.390	1.000	0.011	-0.006	-0.005
#OLY	-0.121	0.023	0.011	1.000	0.021	0.437
VL	-0.082	-0.126	-0.006	0.021	1.000	0.001
M#O	-0.052	0.000	-0.005	0.437	0.001	1.000

TABLE III
ANOVA TABLE TO TEST WHETHER CB TYPES HAVE A SIGNIFICANT EFFECT ON MI CONDUCTED

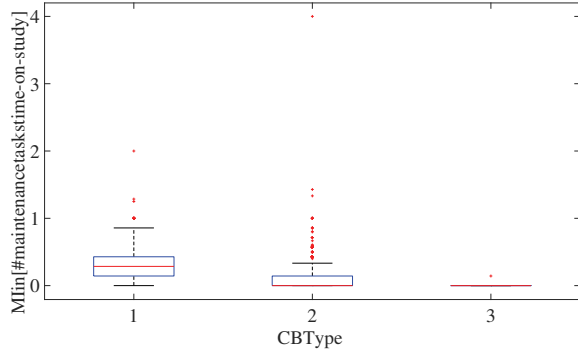| Source	DF	Sum of Squares	Mean Square	F Value	Prob >F
Model	2	14.792	7.396	160.79	<.0001
Error	2619	120.466	0.046		
Corrected Total	2621	135.257			

Fig. 6. Conducted MI specified by CB type

B. Covariate Assessment

The significance level for rejecting the null hypothesis $\beta = 0$ is set to $\alpha = 0.05$. Table IV presents the results of the univariate analysis. Age at admission is significant with a HR of 1.014. This strengthens the approach of also using age as time to failure in the later multivariate analysis. Compared to CB type 2, the other CB types are not significantly correlated. All MI categories have a positive effect on the failure rate compared to no maintenance. Moreover, the more the CBs are operated within the last year, the higher the HR from 1.992 to 6.066. The assessment if one of the manufacturers is significantly correlated shows no relevant results.

The results of the multivariate model are presented in Table V and Table VI for time-on-study and age as time to failure, respectively. The covariates are selected with the stepwise regression procedure and ties are handled with the Breslow method. As in the univariate analysis, $\alpha = 0.05$ is set as entry requirement. The Tables show the maximum likelihood estimates, 3 tests to test the global null hypothesis, and the model fit statistics.

The covariates which have entered the model in Table V are CB Type, GA, MI, OM, VL, #OLY, and age at admission.

	Maximum Likelihood Estimates						
Covariate	DF	β	SE	Chi-Square	Pr>ChiSq	HR	95% CI	
GA 1	1	-0.325	0.190	2.914	0.0878	0.723	0.498	1.049
GA 2	1	-0.293	0.181	2.634	0.105	0.746	0.524	1.063
GA 3	1	-0.691	0.188	13.459	0.0002	0.501	0.347	0.725
CB Type 1	1	-0.126	0.120	1.112	0.292	0.881	0.697	1.115
CB Type 3	1	-1.830	1.003	3.326	0.068	0.160	0.022	1.146
MI 1	1	-1.773	0.230	59.540	<.0001	0.170	0.108	0.266
MI 2	1	-1.760	0.241	53.570	<.0001	0.172	0.107	0.276
MI 3	1	-0.740	0.163	20.670	<.0001	0.477	0.347	0.657
VL 2	1	0.225	0.124	3.280	0.0701	1.252	0.982	1.597
VL 3	1	0.947	0.240	15.598	<.0001	2.577	1.611	4.123
OM 1	1	0.564	0.452	1.5546	0.2125	1.757	0.724	4.263
OM 3	1	-0.317	0.156	4.0951	0.0430	0.729	0.536	0.990
#OLY 1	1	0.689	0.155	19.742	<.0001	1.992	1.470	2.699
#OLY 2	1	0.778	0.177	19.390	<.0001	2.177	1.540	3.077
#OLY 3	1	1.225	0.327	14.048	0.0002	3.404	1.794	6.459
#OLY 4	1	1.803	0.220	66.963	<.0001	6.066	3.939	9.341
M#O	1	0.001	0.000	8.676	0.0032	1.001	1.000	1.001
Age at admission	1	0.014	0.004	15.206	<.0001	1.014	1.007	1.021
Remote Control 2	1	-0.354	0.339	1.088	0.2968	0.702	0.361	1.364
Remote Control 3	1	1.536	1.001	2.356	0.1248	4.646	0.654	33.021

The covariate effects and their significance are similar to the univariate analysis but differ in the magnitude of the effects. The HR of age at admission is 1.038 which means that with one year difference the failure rate would increase by this magnitude. The relative risk between two CB with ten years age difference is then $e^{10*0.038} = 1.46$. An interesting observation is that CB type 1 is significant and has a positive impact compared to CB type 2. This unexpected effect will be further discussed in section V.

In Table VI only the covariates GA, MI, and #OLY have entered the model. The reference curves for the survival probability in both settings are depicted in Fig. 7. These reference curves are plotted based on the reference level of all categorical covariates and the mean of all continuous covariates.

The results from this study also strengthen the general conclusions of [20] who found that the average failure frequency increases with the voltage level and number of operations of the CBs.

V. DISCUSSION

The results underline the importance of applying the PHM in an exploratory approach. The correct interpretation requires in-depth knowledge about the dataset and should consider the whole range of results and data such as sample size of the population, categorical sub-class size, p-value, HR, and CI. For example, the category GA 1 in the univariate analysis has a p-value of 0.087 which is close to $\alpha = 0.05$. One might

TABLE V

RESULTS OF THE MULTIVARIATE ANALYSIS WITH THE PROPORTIONAL HAZARD MODEL WITH STEPWISE REGRESSION AND TIME-ON-STUDY IN DAYS AS TIME TO FAILURE

Maximum Likelihood Estimates										Testing Global Null Hypothesis: $\beta = 0$			
Covariates	DF	β	SE	Chi-Square	Pr>ChiSq	HR	95% CI			Test	Chi-Square	DF	Pr>ChiSq
CB Type 1	1	-0.735	0.193	14.554	0.0001	0.480	0.329	0.700		Likelihood Ratio	368.604	15	<.0001
CB Type 3	1	-1.745	1.009	2.988	0.0839	0.175	0.024	1.263		Score	447.514	15	<.0001
GA 1	1	-0.466	0.194	5.775	0.0163	0.627	0.429	0.918		Wald	370.169	15	<.0001
GA 2	1	-0.237	0.183	1.685	0.1943	0.789	0.551	1.129					
GA 3	1	-0.708	0.194	13.389	0.0003	0.493	0.337	0.720					
MI 1	1	-2.328	0.235	98.127	<.0001	0.097	0.061	0.155		Model Fit Statistics			
MI 2	1	-2.415	0.249	93.987	<.0001	0.089	0.055	0.146		Criterion	Without Covariates	With Covariates	
MI 3	1	-1.401	0.175	63.992	<.0001	0.246	0.175	0.347		-2 LOG L	4361.740	3993.136	
VL 2	1	-0.011	0.132	0.007	0.9357	0.989	0.764	1.281		AIC	4361.740	4023.136	
VL 3	1	0.689	0.247	7.772	0.0053	1.992	1.227	3.233					
#OLY 1	1	0.936	0.158	35.177	<.0001	2.551	1.872	3.476					
#OLY 2	1	1.345	0.186	52.307	<.0001	3.837	2.665	5.524					
#OLY 3	1	1.279	0.330	15.014	0.0001	3.594	1.882	6.866					
#OLY 4	1	1.844	0.224	67.901	<.0001	6.322	4.077	9.802					
OM 1	1	0.179	0.459	0.152	0.6967	1.196	0.486	2.940					
OM 3	1	-0.506	0.167	9.219	0.002	0.603	0.435	0.836					
Age at admission	1	0.038	0.005	53.280	<.0001	1.038	1.028	1.049					

TABLE VI

RESULTS OF THE MULTIVARIATE ANALYSIS WITH THE PROPORTIONAL HAZARD MODEL WITH STEPWISE REGRESSION AND AGE AS TIME TO FAILURE DATA CONSIDERING LEFT-TRUNCATION

| Maximum Likelihood Estimates										Testing Global Null Hypothesis: $\beta = 0$			
Covariates	DF	β	SE	Chi-Square	Pr>ChiSq	HR	95% CI			Test	Chi-Square	DF	Pr>ChiSq
GA 1	1	-0.421	0.205	4.227	0.0398	0.657	0.440	0.981		Likelihood Ratio	324.340	10	<.0001
GA 2	1	-0.306	0.192	2.544	0.111	0.737	0.506	1.072		Score	420.978	10	<.0001
GA 3	1	-0.758	0.214	12.495	0.0004	0.469	0.308	0.713		Wald	309.084	10	<.0001
MI 1	1	-2.482	0.237	109.475	<.0001	0.084	0.053	0.133					
MI 2	1	-2.465	0.251	96.809	<.0001	0.085	0.052	0.139		Model Fit Statistics			
MI 3	1	-1.694	0.191	79.060	<.0001	0.184	0.127	0.267		Criterion	Without Covariates	With Covariates	
#OLY 1	1	0.731	0.169	18.753	<.0001	2.077	1.492	2.891		-2 LOG L	2711.323	2386.983	
#OLY 2	1	1.028	0.193	28.477	<.0001	2.796	1.917	4.079		AIC	2711.323	2406.983	
#OLY 3	1	0.965	0.338	8.140	0.0043	2.625	1.353	5.093					
#OLY 4	1	1.468	0.245	35.925	<.0001	4.338	2.685	7.010					

interpret it as significant but considering the calculated CI, the effect could be either positive or negative which makes an interpretation impossible. Moreover, the size of the sub-classes of the categorical covariates has an effect on the results. The covariate remote control has a high frequency for the actual subcategory remote control (95.5 %) but a low frequency for manual operation with 4.5 %, see Fig. 4. This unbalance bias the results and thus the subcategories of a categorical covariate should always be similar in size.

The results in Table V show that CB type 1 compared to CB type 2, has a positive significant impact. However, this contradicts with the results of the univariate analysis in Table IV and the multivariate analysis in Table VI. One reason might be that the CB type 1 are generally older than CB type 2 and 3 which can be seen in Fig. 5. If age as time to failure is not adjusted for left-truncation, CB type 1 also has a positive

significant impact which seems logical since survival time and failure are the response variables. However, the positive significant effect between CB types might be explained by the significant difference of conducted MI for both types, see Table III and Fig. 6.

VI. CONCLUSION

This paper continues to characterise the failure rate of CBs with critical parameters such as age, manufacturer, size, and maintenance. Particularly, the impact of explanatory variables or risk factors is investigated by testing their significance and quantifying the magnitude of their effect. To do so, a CB population of 1622 with 281 major failures is analysed. The results show that age is a relevant factor, PM has a positive impact, and that the number of operations within the last year have a negative impact on the failure rate in the multivariate analysis. The study also addresses the importance of using
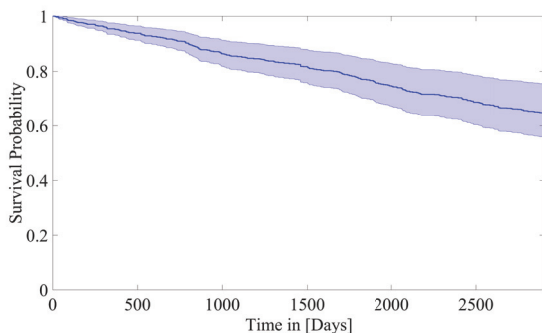
(a) Reference survival function for CB at age 19.37 placed in GA 4, CB type 2, OM 2, MI 0, and VL 1 with time-on-study in days as time to failure
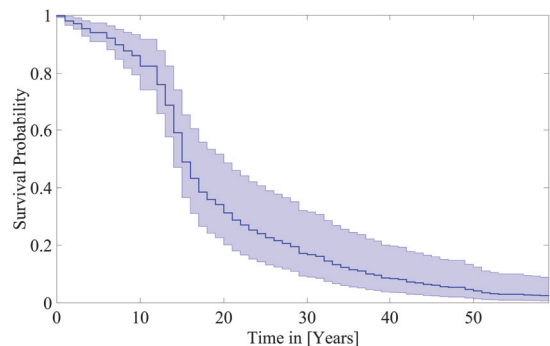
(b) Reference survival function for CB at age 19.26 placed in GA 4, CB type 2, OM 2, MI 0, and VL 1 with age in years as time to failure

Fig. 7. Survival probability functions for reference settings with 95% confidence intervals

an exploratory approach when analysing failure data with the PHM. For example, the impact of the CB type is different in the univariate and multivariate analysis and could therefore be misinterpreted if the dataset is not fully understood. The study should be continued while considering competing risks and the recurrence of major failures to widen the understanding of the dataset. Moreover, the CB service role functions in the system such as line, busbar, reactor, etc or the number of operations with fault current have not been investigated which could be done in future studies.

References

[1] R. E. Brown and H. L. Willis, "Substation asset management," in *Electric Power Substations Engineering, Third Edition*. CRC press, 2012, pp. 1–38.

[2] J. Schneider, A. J. Gaul, C. Neumann, J. Hogrfer, W. Wellow, M. Schwan, and A. Schnettler, "Asset management techniques," *International Journal of Electrical Power and Energy Systems*, vol. 28, no. 9, pp. 643 – 654, 2006, selection of Papers from 15th Power Systems Computation Conference, 2005.

[3] A. Moradkhani, M. R. Haghifam, and M. Mohammadzadeh, "Bayesian estimation of overhead lines failure rate in electrical distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 56, pp. 220 – 227, 2014.

[4] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation a review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1 – 14, 2011.

[5] R. E. Brown, G. Frimpong, and H. L. Willis, "Failure rate modeling using equipment inspection data," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 782–787, May 2004.

[6] J. H. Jürgensen, L. Nordström, and P. Hilber, "A review and discussion of failure rate heterogeneity in power system reliability assessment," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct 2016, pp. 1–8.

[7] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220, 1972.

[8] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*. John Wiley & Sons, 2011, vol. 360.

[9] S. J. Argent, A. Bendell, P. T. Manning, J. Marshall, S. G. Ryan, and D. Wightman, *Proportional hazard modelling in the analysis of transmission failure statistics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 624–633.

[10] M. Baxter, A. Bendell, P. Manning, and S. Ryan, "Proportional hazards modelling of transmission equipment failures," *Reliability Engineering & System Safety*, vol. 21, no. 2, pp. 129 – 144, 1988.

[11] Z. Tang, C. Zhou, W. Jiang, W. Zhou, X. Jing, J. Yu, B. Alkali, and B. Sheng, "Analysis of significant factors on cable failure using the cox proportional hazard model," *Power Delivery, IEEE Transactions on*, vol. 29, no. 2, pp. 951–957, April 2014.

[12] J. H. Jürgensen, A. Brodersson, L. Nordström, and P. Hilber, "Impact assessment of remote control and preventive maintenance on the failure rate of a disconnector population," *IEEE Transactions on Power Delivery*, vol. PP, no. 99, pp. 1–1, 2017.

[13] D. G. Kleinbaum and M. Klein, *Survival analysis: a self-learning text*. Springer Science & Business Media, 2006.

[14] J. P. Klein and M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.

[15] D. Collett, *Modelling survival data in medical research*. CRC press, 2015.

[16] L. R. Landerman, K. C. Land, and C. F. Pieper, "An empirical evaluation of the predictive mean matching method for imputing missing values," *Sociological Methods & Research*, vol. 26, no. 1, pp. 3–33, 1997.

[17] D. B. Rubin, *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, 2004, vol. 81.

[18] IEC 62271-1:2007:, "IEC 62271-1:2007: High-voltage switchgear and controlgear - part 1: Common specifications." 2007.

[19] A. Høyland and M. Rausand, *System reliability theory: models and statistical methods*. John Wiley & Sons, 2009, vol. 420.

[20] T. M. Lindquist, L. Bertling, and R. Eriksson, "Circuit breaker failure data and reliability modelling," *IET Generation, Transmission Distribution*, vol. 2, no. 6, pp. 813–820, November 2008.

[21] E. Shayesteh and P. Hilber, "Reliability-centered asset management using component reliability importance," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct 2016, pp. 1–6.

[22] A. Abiri-Jahromi, M. Parvania, F. Bouffard, and M. Fotuhi-Firuzabad, "A two-stage framework for power transformer asset maintenance management - part i: Models and formulations," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1395–1403, May 2013.