Data Analytics for Grid Resilience with Early Failures and Wear-out Failures

Ross, Robert; Ypma, Peter A.C.; Koopmans, Gerben

DOI
10.1109/SoutheastCon48659.2022.9764082

Publication date
2022

Document Version
Final published version

Published in
SoutheastCon 2022

Citation (APA)
Ross, R., Ypma, P. A. C., & Koopmans, G. (2022). Data Analytics for Grid Resilience with Early Failures and Wear-out Failures. In *SoutheastCon 2022* (pp. 575-582). (Conference Proceedings - IEEE SOUTHEASTCON; Vol. 2022-March). IEEE. https://doi.org/10.1109/SoutheastCon48659.2022.9764082

Data Analytics for Grid Resilience with Early Failures and Wear-out Failures

Robert Ross [1,2]
[1] IWO – Institute for Science and Development
Ede, Netherlands
[2] Delft University of Technology,
Delft Netherlands
R.Ross@iwo.nl

Peter A.C. Ypma [1,2]
[1] IWO – Institute for Science and Development
Ede, Netherlands
[2] Delft University of Technology,
Delft Netherlands
P.Ypma@iwo.nl

Gerben Koopmans
Performance, Reliability and Data Analytics
IWO – Institute for Science and Development
Ede, Netherlands
G.Koopmans@iwo.nl

SoutheastCon 2022 | 978-1-6654-0652-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/SoutheastCon48659.2022.9764082

Abstract— **The here reported work is part of a project on supporting grid resilience by asset management techniques. The present work focuses on support of decision-making after a few failures occurred that may be the start of many more. Methods are reviewed and new algorithms developed where the present IEEE/IEC standard does not provide. Two cases of early failures and wear-out are analyzed as examples for the data analytics.**

Keywords— early failures, Weibull, bathtub curve, censored data, hazard rate, weighted linear regression, similarity index, failure time prediction, redundancy, performance ratio

I. INTRODUCTION

However reliable a grid may be, failures are likely to occur. Handling failure events adequately makes the grid more resilient. The present study focusses on events where failures start to occur which may or may not be the prelude to many more failures. Most likely, forensic investigations will be carried out to find the cause of failure. Such investigations may be time consuming and may not reveal the probability and rate of next failures. The data analytics that deals with these problems, are subject of the present paper.

The present paper shows results of the Dutch FIND-GO project on asset management policies. The project is carried out by the utilities TenneT TSO and Stedin DSO in cooperation with IWO, TU Delft, GE and led by HAN University of Applied Sciences. Collaboration exists with ENTSOe (EU Network of TSOs). Several analysis methods relevant to grid resilience are reviewed, improved and complemented when deemed necessary. The focus is on developing algorithms to analyze data from failure incidents. An IEEE/IEC standard for dielectric breakdown data [1] forms a foundation for work.

A project aim is to make analytics widely accessible. To that purpose a freeware spreadsheet was developed, named IDA (IWO Data Analyzer). This freeware is relatively light-weight computing and should also work on computing power is enough to be used on mobile phones and tablets should also work on mobile phones and tablets [2]. The main features of IDA are shown in Table I and are reviewed in the present paper.

TABLE I. FEATURES OF THE DATA ANALYZER

Goal	Approach	Features
Data characterization	(W)LR Parameter Estimation	Weibull parameters
	Weibull Plot	Beta + Regression confidence intervals
Probability and time	Converter	Time and/or F into R, f, h
Time of next failure for right cenored case	Analytic	Estimated time of next failure with confidence limits
Comparison of distributions	Quantify similarity	Similarity Index over infinite [0,∞) and finite [0,τ] domain
Life extension	Periodic servicing	Optimize cycle with given OPEX and CAPEX
		Estimate average lifetime, hazard rates, cost rates
		Compare performance of optimized and chosen cycle

It is noteworthy, that the paper relates to specific cases rather than setting up a general asset maintenance strategy for large asset populations like MV cable [3] [4].

The structure of this paper is as follows:

- Hazard rate bathtub models and the various stages of component lifecycles are discussed first. Of special interest are early failures and wear-out. The difference between early failures and teething is addressed as well.
- Redundancy as strategy to warrant system functionality is discussed with its limits. A redundancy performance ratio is defined based on hazard rate, repair rate and configuration
- Two example cases are introduced to illustrate the data analytics. The first case concerns early failures that strike after a circuit has been tested successfully. The second case concerns wear and end-of-life of a connection.
- Next, the failure data analysis with the developed methods is demonstrated: plotting, parameter estimation and a review of the characteristics.
- Then predictions are made about reliability and times of future failures along with redundancy considerations.

The research on data analysis of early failures is funded by the Netherlands Ministry of Economic Affairs and Climate through the RvO agency, Grant ref. nr. TEUE418008, TKI Project FINDGO.

The research into approximation of weights for linear regression received support from the EU H2020 R&I program and RvO under ECSEL grant agreement No 826417 (Project Power2Power).

978-1-6654-0652-9/22/$31.00 ©2022 IEEE

II. Lifecycle Stages

The Weibull distribution applies to situations where the time to breakdown is represented by the one breakdown path out of many that has the lowest time to failure time *t*. The hazard rate $h(t)$ of a single Weibull distribution is:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\beta \cdot t^{\beta-1}}{\alpha^{\beta}} \quad (1)$$

Here f is the distribution density, R the reliability function, α and β are the scale respectively shape parameter. If β < 1, the hazard rate is declining, i.e., becomes less significant over time as in teething processes (also called child disease or mortality). If β = 1, the hazard rate is constant as in random failure. If β > 1, the hazard rate is rising, i.e., becomes increasingly significant over time as in wear-out failure processes.

The lifecycle of products is often discussed in terms of a bathtub curve. This is the plot of the normal hazard rate h_n. Various degradation mechanisms may compete in each product. Usually, the hazard rates of (at least) one teething (h_t), one random failure (h_r) and one wear process (h_w) appear in a bathtub. As these dominate in various periods, they may be regarded as stages. The hazard rates of competing processes add to the total hazard rate h_n:

$$h_n = h_t + h_r + h_w \quad (2)$$

For instance, if a circuit consists of multiple components in series, the degradation mechanisms in all components compete. The total hazard rate bathtub of the circuit is formed by summing all relevant hazard rates in a similar fashion. The competition between processes can also occur in a single component.

Occasionally, a batch may consist of subgroups such as specification compliant and non-compliant products. The non-compliant or defective products can work initially, but may wear fast (h_{fw}) [5]. This leads to a hazard rate h_d for the defective products shaped as a compressed bathtub. If, for convenience, the teething and random processes are assumed to be the same as with the normal batch, h_d becomes:

$$h_d = h_t + h_r + h_{fw} \quad (3)$$

If normal and defective products are mixed in one batch, the resulting bathtub depends on the fractions p_n of normal and p_d of defective products with $p_n + p_d = 1$. With R_n and R_d being the reliability functions of the normal and defective batch, the combined bathtub curve has hazard rate h_c [5]:

$$h_c = \frac{p_n \cdot h_n \cdot R_n + p_d \cdot h_d \cdot R_d}{p_n \cdot R_n + p_d \cdot R_d} \quad (4)$$

Characteristic for a mix is a hump where the hazard rate is temporarily dominated by failure of defective products. The resulting bathtub curve is shown in Fig. 1. When the defective products become extinct, the curve of the mixture returns to the curve of the normal products.

In the following, we discuss the stages teething, random failure, early failure (of fast wear) and wear-out failure. Of particular interest are ‘early failures’, literally meaning failures that occur compared to normal life cycles. Both teething and defect wear cause early failures, but their natures are distinct.

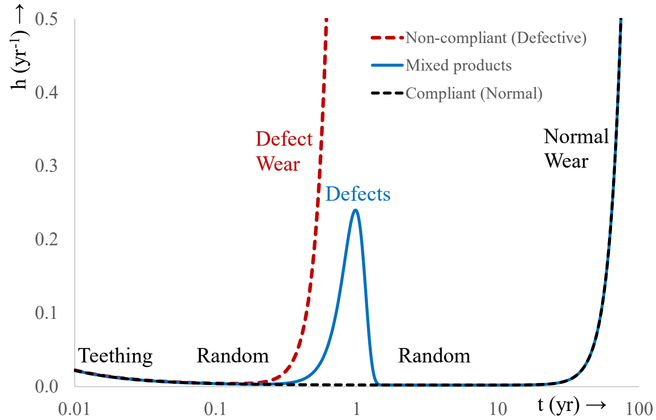

Fig. 1. Example of three bathtub curves: Compliant (Normal) products according to (2); non-compliant (Defective) products according to (3) and a mix of both product types (4).

A. Teething

The hazard rate h_t for teething processes is declining from its (infinite) maximum at t=0. It is reducing in importance from the very beginning. Teething problems are generally tackled with quality control by in factory product screening and testing combined with onsite commissioning tests after installation. The screening process pre-ages all products up to the moment that teething drops below a defined level of acceptance. In the example of Fig.1, that moment might be $t \approx 0.1$ yr. After screening and testing, normally the surviving components are of good quality. An adequate operational stage of random failure with an acceptable, low hazard rate is expected to follow, ultimately ending in a wear-out phase.

If possible, screening is designed such that it specifically accelerates teething and not the other processes. Often however, screening also causes some aging by other processes inevitably. As a consequence the delivered products are already somewhat aged. This can often, but not always, be ignored in data analysis.

B. Random Failure and Servicing of Wear

Random failure is unrelated to age and its hazard rate h_r is constant unless operational circumstances change. Additional preventive measures may lower h_r, while intensified exposure to external threats may enhance h_r. E.g., if a layer of soil above the cable is removed, it becomes more vulnerable to external activities like digging. Random failure at a low hazard rate is desirable in operations and can often be achieved by preventive maintenance. For instance, various types of wear can be (practically) reset by periodic servicing (either time-based or condition-based). It may be noted that periodic renewal is not suitable for teething process.

If wearing products can be periodically renewed by servicing, the hazard rate is repeatedly reset to zero and takes the form of a sawtooth function (section 9.1.1 in [6]). The average sawtooth hazard rate h_{ave} can be viewed as the constant hazard rate that comes with a random failure process. The hazard rate h_{ave} is given by (cf. [7] and section 9.1.1.3 in [6]):

$$h_{ave} = \frac{-\ln\big(R(T)\big)}{T} \quad (5)$$

Here, T is the service cycle and R the reliability function. The shorter T, the lower h_{ave}, but also the higher the operational expenses costs (OPEX). The total expenses (TOTEX) due to replacement (capital expenses, CAPEX) and servicing (OPEX) can be minimized to yield an optimized service cycle T_{opt} [7].

C. *Wear*

As explained above, wear does not only occur at the end of normal life but is also encountered as one of the types of early failures. E.g., wear of defective products causes failures much earlier than specified for normal products. Early failures due to fast wear can be distinguished from teething when installed products are successfully tested and put into operation, but failures start to occur after a (much too short) period [5]. There is a distinct difference between teething and fast wear of defects.

Contrary to teething, the hazard rate of wear (fast or normal) has its minimum ($h = 0$) at $t = 0$ and keeps increasing from then on. The time between failures will shorten from the very beginning. This does not always show in a bathtub curve, but is apparent if the failure data of the defective products can be separated from the mixed data. Then the real hazard rate shows.

If the fast-wearing products constitute only a fraction of the total batch, then this defective group gets extinct at some time. This is the declining part of the hump (Fig. 1). Though it looks similar to teething, the hazard rate of the defective products alone is actually still rising. The decline is not due to a declining hazard rate, but due to extinction of defective products. The hazard rate of a single process is determined as $h = f/R$ (1). In a bathtub curve like Fig. 1, the various distributions are mixed. At the hump, not the R_d of the fast-wearing defective fraction is used, but rather the total surviving fraction $p_n \cdot R_n + p_d \cdot R_d$. Therefore, the righthand side of the hump does not represent the hazard rate of the defective fraction well, but can be very biased.

Fast wear deviates from normal wear by an abnormally low Weibull scale parameter α. The shape parameter, $\beta > 1$. Until forensic investigations shed light on the cause of early failures, a big unknown and often a big challenge is to determine how large the fraction p_d is. Is it worthwhile to repair and continue operation? Or should all components be replaced preventively?

When (normal or fast) wear starts, the worst is yet to come and decisions about repair or replace may become urgent issues. Important related questions are: Are the failures deviant from expectations? Will new failures occur in the near future? At what rate will failures occur?

D. *Single Distribution Approach*

According to (4), the hazard rate can be quite complicated. The statistical analysis may seem overwhelming too. However, if events occur where the hazard rate of a single process dominates, the data analytics can often be approximated by applying a single Weibull distribution [5]. This simplifies the data analytics considerably. This is also the approach of the data analysis discussed here. Although the mentioned methods can deal with teething and random failure as well, the presently reported work particularly aims at dealing with cases of early and normal wear.

III. Configuration Considerations

Grid components may fail, but the prime responsibility of utilities is to supply electric power. Redundancy by multiple distribution alternatives is a common strategy to warrant the grid functionality despite failures.

For instance, if a connection consists of two or more parallel circuits that each can supply the full required power, then failure of a single circuit does not need to interrupt the power supply. During the repair of the failed circuit, the connection will be more vulnerable, because of less redundancy. With the occurrence of a next failure, the connection may fail depending on the degree of redundancy. So, adequate response to failure can shorten the period of elevated vulnerability. Redundancy and speedy repair together decrease the hazard rate of connection failure. Here it is discussed to what extent parallel paths and repair can be sufficient countermeasures.

A. *Redundancy and repair*

Redundancy can be achieved with various alternatives (including emergency generators etc.). Here, we discuss the hazard rate of a system of identical parallel circuits and its relation to the hazard rate of single circuits. It is assumed that the hazard rate is a (semi)-constant for a defined coming period.

Let the hazard rate of a serviced single circuit be $h_1 = h$; of a double circuit h_2 and of a triple circuit be h_3. In this approach, we assume that the ruling distribution is Exponential (which is a special Weibull case) with expected lifetime $\theta = 1/h$. Also, assume one failed circuit can be repaired at a time, with a constant rate μ. Such conditions influence the impact of redundancy [6]. The resulting impact of the present choices is shown in Table II. The bathtub curves of a single and a double are compared in Fig. 2. Table III shows a numerical example.

The impact of redundancy can also be rated with a performance ratio p_r. Here it is defined as the ratio of the single circuit hazard rate divided by hazard rate h_x of an alternative x:

$$p_{r,x} = \frac{h_1}{h_x} \tag{6}$$

TABLE II. IMPACT OF REDUNDANCY ON THE HAZARD RATE AND STEADY STATE AVAILABILITY (CF. CH.8 IN [6]).

Connection configuration	Single circuit	Double circuit	Triple circuit
Hazard rate	$h_1 = h$	$h_2 = \frac{2h^2}{2h + \mu}$	$h_3 = \frac{6h^3}{9h^2 + 4\mu h + \mu^2}$
Availability	$A_1 = \frac{\mu}{h + \mu}$	$A_2 = \frac{\mu^2 + 2\mu h}{\mu^2 + 2\mu h + 2h^2}$	$A_3 = \frac{\mu^3 + 3\mu h + 6\mu h^2}{\mu^3 + 3\mu^2 h + 6\mu h^2 + 6h^3}$
Performance ratio	$p_r = \frac{h_1}{h_1} = 1$	$p_r = \frac{h_1}{h_2} = 1 + \frac{\mu}{2h}$	$p_r = \frac{h_1}{h_3} = 1.5 + \frac{2\mu}{3h} + \frac{\mu^2}{6h^2}$

TABLE III. TYPICAL EXAMPLE WITH $h = 0.01\text{YR}^{-1}$ AND $\mu = 26\text{YR}^{-1}$.

Connection configuration	Single circuit	Double circuit	Triple circuit
Hazard rate	0.01	0.000,008	0.000,000,009
Availability	0.9996	0.999,999,7	0.000,000,000,7
Performance ratio	1	1,300	1,100,000

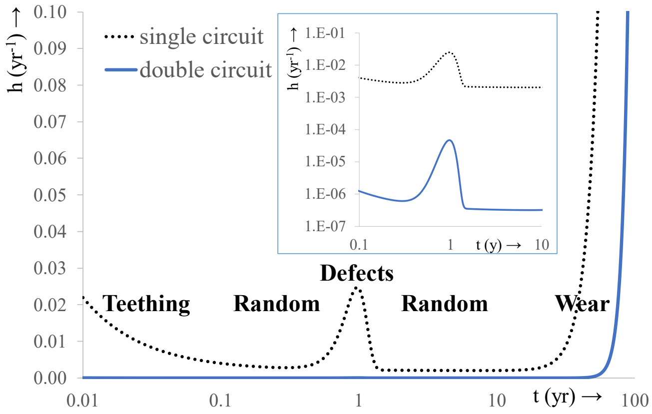

Fig. 2. Comparison of the hazard rates of a single and a double circuit.

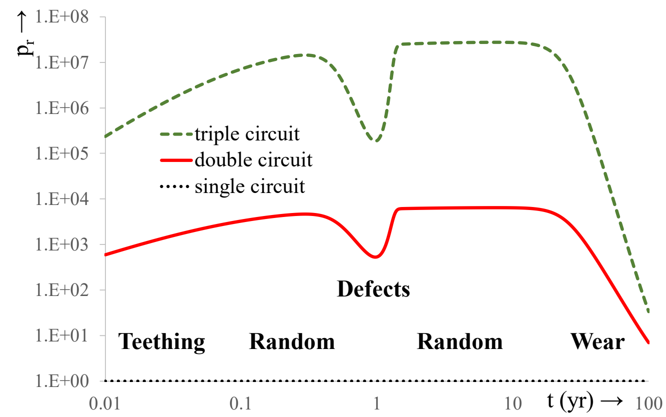

Fig. 3. Comparison of performance ratios.

The higher $p_{r,x}$, the greater the effectiveness of the alternative x. If the hazard rate is a constant, $p_{r,x}$ is the ratio between the MTBF of the alternative θ_x and of the reference θ_1. Fig. 3 shows p_r for the three configurations in Table II.

The assumption of a constant hazard rate will generally not be met, but it provides a convenient approximation. If a forecast is made for a coming year and the hazard rate is not expected to change much during the year, it can be chosen (semi)constant and adapted per year. This is the semi-constant approach.

An illustration of the semi-constant approach is to apply it to the bathtub curve with a defective fraction p_d = 10% (Fig. 1). Fig.2 compares the resulting bathtub curve of a double circuit connection to that of a single circuit. The inset shows the hazard rate at the hump due to the early failures on a double logarithmic scale. If the requirement would be that hazard rate of the connection remains $<10^{-4}$ (meaning that such a connection should have an MTBF of 10,000 yr), then the double circuit appears sufficient to overcome the early failures (Fig. 2). Furthermore, it postpones the moment that the wear-out phase exceeds the high hazard rate levels (e.g., $h=10^{-3}$ or 10^{-4} yr^{-1}).

Comparing Fig. 2 and Fig. 3, it can be seen that redundancy is particularly effective when the hazard rate h_1 is low. In the wear-out phase the hazard rate steeply rises and the performance ratio $p_{r,x}$ collapses. Redundancy is a good strategy during the random failure phase and the onset of the wear-out phase. However, it is not a robust remedy once wear-out is advanced considerably.

B. Redundancy impact on reliability requirements

Depending on their strategic value, various critical hazard rate levels can be defined for components, substations and connections. The reasoning is often based on the acceptable number of incidents within a series of comparable assets. E.g., if a utility has 500 installed transformers and would accept a transformer to fail once per 20 yr maximum, the hazard rate apparently must be smaller than 1/(500·20 yr)=10^{-4} yr^{-1}. It is quite common to demand hazard rates for events below 10^{-4} yr^{-1} or 10^{-5} yr^{-1}.

After the definition, the challenge is to assess the hazard rate. In distribution grids, assets may be employed in large numbers and hazard rates may be readily determined from the number of failures and the number of such assets in operation. However, transmission grids are usually built of much less, but far more strategic, assets. The number of assets in operation multiplied by the acceptable maximum hazard rate may remain <1. A critical hazard rate may then unwittingly be exceeded without failure. As a consequence, the utility may not be aware that the system hazard rate is higher than acceptable [8].

Regular replacement of cable connections and substations can have lead times of several years due to planning, design, permits, production, and installation. The discussion about performance ratio shows that indications of a starting wear-out phase should be taken seriously in order to take timely actions.

C. Example cases

The present paper reports on the data analytical methods as used in the framework of asset management. A widely accessible data analyzer was aimed at. At this point, two example cases are introduced to illustrate the issues to be solved and the available techniques. The examples concern an early failure case and a wear-out case.

The early failure case is taken from Section 9.4.3 in [6]. This concerns two parallel HV cables with each 100 cable joints. The quality control through factory tests and test after installation were passed successfully. The cable system was planned for >40 yr. However, 5 early failures occurred within 4 months (Table IV). The cables were not redundant, and therefore each failure caused an industrial black-out.

The wear-out case is about failures in a HV cable system that had been in operation for more than 30 yr. After more than 28 yr, a phase-to-earth fault caused the voltage on the healthy phases to rise to about the phase-to-phase voltage for about 6 hours. This was believed to trigger electrical treeing in the cable paper insulation. After more than 1.5 yr, a first failure occurred in the cable insulation. The cable system was redundant. Since the pressurized cable system required preconditioning before recommissioning, a repair takes about 2 weeks (or more).

In both cases statistical evaluations were carried out followed by forensic investigations. With repeated faults in both cases, doubts were growing about the integrity of the systems.

TABLE IV. OVERVIEW OF FAILURE DATA FOR BOTH CASES

Early failures	t_1	t_2	t_3	t_4	t_5
Failure time (yr)	0.159	0.208	0.247	0.274	0.293
Wear-out failures	t_1	t_2	t_3	t_4	
Failure time (yr) since installation	30.02	32.02	35.16	35.74	
Failure time (yr) since voltage event	1.595	3.597	6.737	7.315	

Therefore, the ultimate question in both cases became whether to continue repair or to replace.

In order to be prepared for fast response to such incidents, methods and techniques were developed. The algorithms and facilities are combined in a spreadsheet for wide accessibility and is made available as freeware [2].

IV. ISSUES TO BE STUDIED

In the following, both cases are analyzed step by step. The steps are:

- Characterization by plotting and parameter estimation
- Predictions of probabilities and times of failure
- Comparison of observations and references

These steps are discussed in the following sections.

A. Characterization and Data Review

As a first step, the data are reviewed with a Weibull plot and by parameter estimation. The plot follows the IEEE / IEC guidelines [1]. The estimated parameters and the best fit are obtained with ordinary linear regression (LR) or weighted linear regression (WLR). The variables are the plotting positions Z_i and $\log(t_i)$. The developed IDA allows switching between LR and WLR.

The choice for (W)LR and not for the maximum likelihood (ML) method is not fundamental but offers advantages. Firstly, (W)LR provides analytical solutions which allows convenient error calculations in parameters and is fast. Secondly, graphical representation and parameter estimation are fully aligned. Thirdly, both beta and regression confidence intervals can conveniently be drawn. The developed application also runs on light-weight computing devices like mobile phones and might find its way into smart devices.

The WLR method employs weights w_i for processing the data t_i. These w_i are the inverse of the $Z_{i,n}$ variances [9] [10], Computing these weights can be demanding for data sets of tens of failure times. With the development of the analyzer IDA, a weight approximation algorithm was developed [5] [11]:

$$\frac{1}{w_{i,n}} \approx \frac{1}{i-0.5} - \frac{0.1}{(i-0.3445)^3} + \frac{0.125 \cdot (i-1)^{1.4}}{(n+0.343)^{1.656} \cdot (n-i+0.8)^{0.75}} \qquad (7)$$

This algorithm has an error ≤ 1% for each individual weight w_i for sets with a sample size up to n=500 and < 2.8% for sets with size up to n=2000. Compared to the maximum weight in the set (per n), the error of weights is ≤ 0.34% (for $n = 2$) and decreasing with increasing n. It is also suitable for non-integer indices i that can occur in adjusted ranking with censoring [1]. Equation (7) enables WLR with light-weight computing.

As for the present failure data (Table IV) of the early failures as well as in the wear-out case, there is a set of r observed failure times t_i ($r = 5$ and $r = 4$, respectively). The sample size n (i.e., the total number) of observed plus future faults is unknown in both cases. However, next failure times will be larger than the already observed data, i.e.: $t_i > t_r$ for $i > r$. Such cases are called (singly) right censored cases and adjusted ranking is not needed. In other cases, the condition that future failure times $t_i > t_r$ may not hold, for instance, if some assets were put into operation later and consequently may live shorter than presently observed failure times. Equation (7) is easier and more precise than the recommended practice in the present standard [1].

For plotting and parameter estimation an assumption must be made about n. As an exercise, often a sequence of future failure numbers $k = n - r$ is explored. A typical sequence is $k = 0, 1, 3, 10, 30$. These k values represent the respective qualifications of: 'no more failures of this kind to follow', 'only one more failure are to follow', 'a few more failures are to follow', 'considerably more failures are to follow', 'many more failures are to follow'. A next challenge is to estimate which situation is most applicable given the already observed failures. This is discussed below.

For review of the two example cases, the estimations a_k and b_k for the Weibull parameters α and β are shown in Table V. The failure times and times since the previous time are shown in Table VI. The data are plotted in Fig. 4 with the best fits and 90% regression confidence intervals. WLR was used for plotting and parameter estimation. A review of data could be as follows.

As for the early failures, $b_k > 1$ for all k, which makes it a clear wear mechanism. The time between failures is in the order of days (0.01yr ≡ 3.65d). The associated high hazard rate is in the order of $10 - 100\ yr^{-1}$. The time-between-failures show very little scatter and keep even consistently shortening with increasing i (Table VI). This means in a fixed time interval the number of failures increases, which means the distribution density f is on the rise. Weibull density functions with $\beta > 2$ are bell shaped and the region for curving upward is at the very start of the bell meaning most failures are still to come. The hazard rate rises even more than f ($h = f/R$ and $R < 1$). Therefore, the hazard rate is not only high, but still consistently rising. The situation bears a great resemblance to the region $t \approx$ 0.5 yr in Fig.1 at the onset of the fast wear curve and of the hump (which ever is applicable).

TABLE V. ESTIMATED WEIBULL PARAMETERS FOR VARIOUS k

Early failures	$k = 0$	$k = 1$	$k = 3$	$k = 10$	$k = 30$
a_k (yr)	0.261	0.280	0.313	0.393	0.523
b_k	4.90	4.36	3.96	3.60	3.42
Wear-out	$k = 0$	$k = 1$	$k = 3$	$k = 10$	$k = 30$
a_k (yr)	5.81	7.38	10.49	20.9	48.2
b_k	1.74	1.53	1.39	1.269	1.214

TABLE VI. FAILURE TIMES AND TIMES BETWEEN FAILURES

Early failures	t_1	t_2	t_3	t_4	t_5
t_i (yr)	0.159	0.208	0.247	0.274	0.293
$t_i - t_{i-1}$ (yr)	0.159	0.049	0.038	0.027	0.019
Wear-out	t_1	t_2	t_3	t_4	
t_i (yr)	1.595	3.597	6.737	7.315	
$t_i - t_{i-1}$ (yr)	1.595	2.002	3.140	0.578	

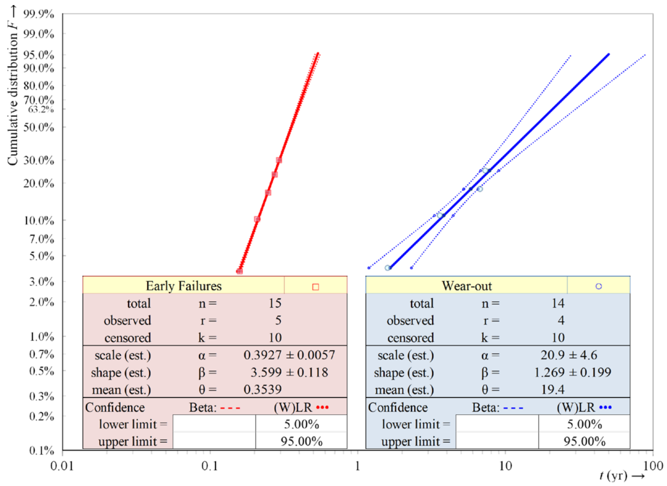

Fig. 4. Weibull plot of the early failures and the Wear-out data. In both cases a number of future failures $k = 10$ is used as an example.

As a first conclusion for the early failures, the problems seem to have just started and the worst is yet to come.

As for reviewing the wear situation, the failure times scatter more about the best fit (Fig. 4). With a time between failures in the order of 2 yr the influence of seasons may have an effect. This may cause a bigger scatter in data and the wider confidence intervals. The times between failures on average are also considerably longer than with the early failures. Unlike the early failure case, the development of intervals is not very consistent. In the beginning the intervals lengthen, but then drastically shorten (Table VI). The question might rather be whether the present average h is acceptable or signaling an end-of-life.

B. Predictions of Probabilities and Failure Times

After characterization and review, parameters are estimated and a plot is available. A Weibull distribution is defined and that can be used for further evaluations like time to next failure and probability of failure in a given period.

As mentioned above the estimated Weibull parameters depend on the assumed sample size n. Various methods are being explored to estimate n from the data. Such a method exploits the shape of the probability density $f(t)$. That method was qualitatively addressed above, but a quantitative exercise employing a similarity index is discussed in [12] and section 9.4 of [6]. The present early failures case shows a great similarity to cases with large sample sizes. However, for data that scatter much, the method remains less conclusive. Positive is that the method also indicates that a high similarity is not going to be reached which prevents drawing wrong conclusions [6]. The statistical significance is still being studied. As for the early failures case with discussed findings of a high hazard rate, the absence of redundancy and the expectation that the problems only just began, preventive replacement is recommended. Still, next failure times can be estimated as explained below.

As for the wear-out case, the data scatter makes this case less clear cut at this stage. The matter of estimating total sample sizes after a relatively small number of failures is an ongoing study subject as yet. However, predictions of probabilities and next failure times can provide deeper insights.

TABLE VII. PREDICTED NEXT FAILURE TIMES WITH CONFIDENCE LIMITS

Early failures	$t_r = 0.293$	$k = 1$	$k = 3$	$k = 10$	$k = 30$
$\langle t_{r+1}\rangle - t_r$ (yr)	-	0.039	0.026	0.020	0.019
$t_{r+1,10\%} - t_r$ (yr)	-	0.006	0.003	0.002	0.002
$t_{r+1,50\%} - t_r$ (yr)	-	0.032	0.020	0.015	0.014
$t_{r+1,90\%} - t_r$ (yr)	-	0.081	0.056	0.044	0.041
Wear-out	$t_r = 7.315$	**k = 1**	**k = 3**	**k = 10**	**k = 30**
$\langle t_{r+1}\rangle - t_r$ (yr)	-	3.99	2.61	2.06	2.06
$t_{r+1,10\%} - t_r$ (yr)	-	0.51	0.31	0.23	0.23
$t_{r+1,50\%} - t_r$ (yr)	-	3.04	1.91	1.48	1.35
$t_{r+1,90\%} - t_r$ (yr)	-	8.79	5.89	4.69	4.29

In the following, predictions are made firstly about the next failure times and associated confidence limits. Next the influence of the degree of redundancy on the predictions is considered. As may be noticed from Table V, the estimated shape parameters b_k vary with k, but the variation is modest in the range $k = 1,..,30$. As a result, the times do vary with k, but also modestly.

The distribution of the remaining failures after t_r can be estimated employing the Beta function [5]. Let $\theta \geq t_r$ be a time after which the remaining $n - r$ last failures will occur. If the complete set of n failures is distributed according to a given Weibull distribution $F(t)$, then θ can be associated with probability value $F_\theta = F(t = \theta)$. In the present approach, the estimated parameters in Table V, define such distributions. For each $k > 0$, we can calculate an F_θ with given θ.

The probabilities p that correspond to the future failure times are uniformly distributed over $[F_\theta, 1]$. Next, a coordinate q can be defined as:

$$q = \frac{(p - F_\theta)}{(1 - F_\theta)} \tag{8}$$

The expected next failure time $\langle t_{r+1}\rangle$ and its A% confidence limit $t_{r+1,A\%}$ can then be found by the following expressions:

$$\langle t_{r+1}\rangle = \int_0^1 t\big(F_\theta + q\cdot(1 - F_\theta)\big)\cdot f_B(q;1,n-r)\mathrm{d}q \tag{9}$$

$$t_{r+1,A\%} = t\left(F_\theta + (1 - F_\theta)\cdot B^{inv}(A\%;1;n-r)\right) \tag{10}$$

B^{inv} is the inverse Beta distribution and f_B the Beta density function. In our analysis, we chose $\theta = t_r$. Table VII shows the expected time to the next failure with the 10%, 50% and 90% confidence limits for the early failures and wear-out cases. The facility is built into the freeware data analyzer.

Although n is not known, in both cases the moment of the next failure changes only modestly with n c.q. k. The biggest change in $\langle t_{r+1}\rangle$ and the confidence limits $t_{r+1,A\%}$ is with small k, but all estimates are still well in the same range. This holds for both cases.

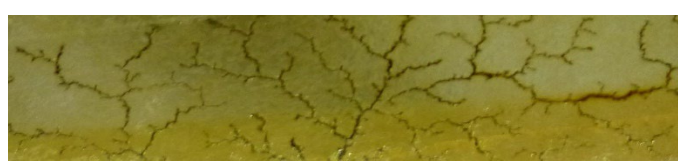

Fig. 5. Treeing on paper insulation. This is not a breakdown, but it disturbs the electric field and weakens the insulation. It will cause a fault ultimately.

As for the early failure case, the estimates confirm the dramatic situation. The next failure can be expected in the order of a few days. And this industrial cable is not redundant.

As for the wear-out case, forensic studies meanwhile showed that the failures were due to extensive electrical trees growing in the paper insulation (cf. Fig. 5). Based on the visual inspections, k is rather in the range of tens or hundreds than only one. The next failure is expected in about 2 yr (Table VII). The system is redundant. Is frequent repair workable or unjustifiable then?

C. *Redundancy revisited*

With redundancy, a failure of one circuit does not down the power supply by the system. However, as explained above, redundancy loses effectiveness with increasing hazard rate. In the wear-out case, the hazard rate is at such a level that about every 2 yr a repair is necessary. Apart from the burden of personnel planning and high OPEX, the redundancy may not be adequate anymore in preventing blackouts. This is investigated by estimating the hazard rate of the connection.

The hazard rate for electrical tree failures can be determined with (1) and the estimated parameters (Table V). However, it should be noted that the distribution based on a_k and b_k defines the failure behavior of the electrical trees. The hazard rate of the electrical trees h_e is adapted from (1):

$$h_e(t) = \frac{b_k \cdot t^{b_k - 1}}{{a_k}^{b_k}} \tag{11}$$

Originally, the circuit contained n growing electrical trees that are competing for failure. The circuit behaves as a series system of electrical trees (section 7.3 in [6]) and the hazard rates of the trees must be summed. Ergo, the circuit hazard rate h_1 equals:

$$h_1 = n \cdot h_e \tag{12}$$

The connection hazard rate h_2 is given in Table II and can be rewritten as:

$$h_2 = \frac{2h_1^2}{2h_1 + \mu} = \frac{2 \cdot (n \cdot h_e)^2}{2 \cdot n \cdot h_e + \mu} \tag{13}$$

The repair rate $\mu = 26.1\ \text{yr}^{-1}$ as mentioned above. This is the Markow chain approach with (semi)-constant hazard rates and a single repair crew. Other assumptions will lead to somewhat different equations. E.g., having multiple repair crews will increase μ and lower h_2. So, the precise formulation of (13) can vary if alternative case specific assumptions are made.

Table VIII shows the results for h_1 and h_2 at t_r and at the 2 consecutive years. Having the forensic findings that k is rather in the range of tens or hundreds, the hazard rates are determined for $k = 10, 30, 100$ which corresponds to the total number of electrical trees (i.e., failures to be) per circuit $n = 14, 34, 104$.

The results show that the connection hazard rate is in the range of about 0.3-0.25, which means that such a connection will suffer a black-out every 30 to 40 yr. If the requirement is that a connection hazard rate $h_2 < 10^{-4}$ yr typically, then the present case needs urgent attention.

TABLE VIII. HAZARD RATES OF INDIVIDUAL CIRCUITS AND OF THE REDUNDANT CONNECTION

Circuit fault	***at time (yr)***	**k = 10**	**k = 30**	**k = 100**
h_1 (yr^{-1})	t_r	0.640	0.572	0.546
h_1 (yr^{-1})	t_r +1	0.662	0.588	0.559
h_1 (yr^{-1})	t_r +2	0.683	0.603	0.571
Connection fault	***at time (yr)***	**k = 10**	**k = 30**	**k = 100**
h_2 (yr^{-1})	t_r	0.030	0.024	0.022
h_2 (yr^{-1})	t_r +1	0.032	0.025	0.023
h_2 (yr^{-1})	t_r +2	0.034	0.027	0.024

V. DISCUSSION AND CONCLUSION

The present work is part of a project by Dutch utilities, research organizations and industry. The broad aim is to align views on resilience in the grid and asset management methods. Here, we review the data analytics that are presently applied in case of a (usually small) set of failures, that potentially is the start of an escalating issue.

The foundation for the analysis are IEEE/IEC guidelines [1]. However, the decision-making in various cases requires answers to questions beyond the guidelines. A research program is ongoing aiming at decision-making support in case of scarce data and emergency situations. Examples of questions concern the number of failures to follow, when next failures can be expected, etc. Often the issue behind the questions is whether assets should continue to be repaired or all should be replaced. Another prime question is often to quantify to what extent two distributions are the same, meaning whether the observed failure behavior matches a reference (such as specifications).

A range of techniques has been and still is developed. The present paper illustrated the mainstream of failure data analytics to support grid resilience. The focus in the present paper was to recognize wear-out and to interpret early failures properly. As shown, early failures after a seemingly successful testing are often not teething but fast wear of weak products. These can be discriminated by the Weibull shape parameter.

Another subject beyond the standard is evaluation of redundant systems. From various cases in the project, it is evident that redundancy is usually effective to prevent interruption by random failures which normally have a low hazard rate. The effectiveness of redundancy is strongly reduced with increasing hazard rate. A performance ratio is defined to quantify the approximate quality of the redundancy.

The paper discussed two cases: a typical early failures case and a wear-out case. Both examples concern cases that called for drastic measures. The paper shows how such conclusions can be reached.

Work on sample size estimation and similarity of distributions were mentioned, but briefly discussed here.

A secondary aim in the resilience project is to disclose the data analytics to a wide audience. To that purpose software is made available that is accessible in utility asset management departments and small businesses. Many techniques including an interactive Weibull plot and weighted linear regression are cast in algorithms that can be handled by typical, widely accessible, office software like a spreadsheet. This is made available as freeware [2] for non-commercial purposes and for education.

The present analyzer bundles facilities for plotting Weibull data with both Beta and regression confidence limits. It allows to estimate parameters by weighted or ordinary linear regression and it provides basic calculators to convert time into distribution functions and vice versa. With the analyzer it is possible to optimize the cycle for periodic maintenance based on OPEX and CAPEX, and to predict the time of the next failure.

References

[1] IEC TC112, “IEC 62539(E):2007 Guide for the statistical analysis of electrical insulation breakdown data,” International Electrotechnical Committee, Geneva, 2007.

[2] IWO, "IWO Data Analyzer Freeware," IWO, February 2022. [Online]. Freeware available at: www.iwo.nl.

[3] IEC WP Asset Management, “Strategic Asset Management of Power Networks”, White paper, International Electrotechnical Committee, Geneva, 2015.

[4] T. Reske, N. Tichelkamp, N. Hopfer, and M. Zdrallek, “Determination of the Ageing Behaviour of Medium Voltage Cables for Optimised Asset Management”, ETG-FB 162 VDE Hochspannungstechnik, paper 53, online, 2020.

[5] R. Ross, P.A.C. Ypma, and G. Koopmans, “Weighted Linear Regression based Data Analysis for Decision Making after Early Failures,” Proceedings IEEE PES ISGT-Asia 2021, paper 148, Brisbane / online (2021).

[6] R. Ross, Reliability Analysis for Asset Management of Electric Power Grids, 1st Ed, Hoboken, NJ: Wiley-IEEE Press, 2019.

[7] R. Ross, “Health Index methodologies for decision-making on asset maintenance and replacement,” CIGRÉ Winnipeg 2017 Colloquium, Paper CIGRÉ D1-95, 2017.

[8] R. Ross, and P.A.C. Ypma, “Data analytics for decision-making about preventive replacement and incident evaluation with ageing assets,” Cigré Colloquium SC A1/B2/D1, New Delhi, India, 2019.

[9] J. S. White, "The Moments of Log-Weibull Order Statistics," Technometrics, vol. 11, no. 2, pp. 373-386, May 1969.

[10] J. M. van Zyl, “Exact expressions for the weights used in least-squares regression estimation for the log-logistic and Weibull distribution,” Communications in Statistics - Theory and Methods, vol. 46, no. 4, pp. 1720–1730, Feb. 2017, doi: 10.1080/03610926.2015.1026995.

[11] R. Ross, “Power Function Algorithm for Linear Regression Weights with Weibull Data Analysis,” Ch.40 in Advanced Computing, Vol. 1528 of the Communications in Computer and Information Science series, 2022.

[12] P.A.C. Ypma, and R. Ross, "Determining the Similarity between Observed and Expected Ageing Behavior," in Proceedings ICEMPE, Xi'an, 2017.