Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P066-cd

Analysis of the Use of Field Data Under Variable Conditions to Develop Lifetime Models for Electrical Distribution Devices

Roman Mukin

Corporate Research Center, ABB Switzerland Ltd, Baden-Dättwil, CH-5405.
E-mail: roman.mukin@ch.abb.com

Kai Hencken

Corporate Research Center, ABB Switzerland Ltd, Baden-Dättwil, CH-5405.
E-mail: kai.hencken@ch.abb.com

Lorenzo Privitera

Corporate Research Center, ABB Switzerland Ltd, Baden-Dättwil, CH-5405.
E-mail: lorenzo_privitera@hotmail.com

Lifetime models have been predominantly developed using constant but accelerated conditions to assess their base lifetime and the acceleration factor under different conditions. This approach is expensive and time-consuming, especially for highly reliable devices, as found in electrical distribution systems. On the other hand, online monitoring provides a large amount of data on the conditions and failures of the fleet of devices. However, constant conditions are not generally present. Therefore, developing efficient methods to estimate parameters from field data is of interest.

Proportional hazard (PH) and accelerated failure time (AFT) models are commonly used to describe the failure of devices under time-varying stress factors. This work analyses how these can be used efficiently to estimate reliability models' parameters, focusing on real-world electrical distribution devices.

The reliability function of a highly reliable device is challenging to acquire, as failure will generally only happen after a long time, and most of the time, devices are not run until failure. In addition, the dependency of the failure rate on environmental conditions the device is operating in requires to make a series of experiments to infer the acceleration factors in the classical setting. Therefore for such devices, accurate reliability curves or hazard rates are often not known, which limits the application of lifetime models, e.g., for maintenance or service planning. Up to now, mainly the "average" reliability of a type of device was used, meaning that the environmental conditions were often unknown. For this, most often, field or fleet data was already used. Where even this was not possible, the reliabilities of whole classes of devices were studied. Overall the effect of an aggregation of failure data over a diverse population will lead to a spread of the reliability curve compared to the one using a specific device type or specific environmental conditions, hindering a precise prediction of its failure. It is therefore of interest to find ways to make use of all available information to improve this. We explore this in this study for two different models and using simulated failure data coming from real environmental conditions.

Keywords: Proportional hazard model, Accelerated failure time model, Field data, Variable conditions, Electrical distribution, Capacitors, Breakers.

1. Introduction

Whereas in the past, there was no other approach apart from a general assessment of the environment, e.g., in terms of application or general installation conditions, to incorporate effects in lifetime models, more and more devices are now equipped with additional sensors, which have the capability to capture them in details. This capability is made possible by enhanced connectivity, exemplified by the Internet of Things (IoT). Whereas this helps with understanding the usage conditions, it can also be combined with failure information, e.g., from service calls. Digitalization also allows for the automatic and reliable tracking of individual devices in the field, of full devices, or even components using printed serial numbers, unique bar codes, or component lists that can be determined given the device's serial number.

We expect that in the near future, this can be used to improve our knowledge of device reliability, especially concerning its dependency on external influence factors. In this work, we explore methods of how this can be used in cases that are typical for the electric device industry. Despite us having this application in mind, we still see this topic as being also of interest to other areas.

Specifically, we consider two general survival analysis methods: the Cox proportional hazard (PH) and the accelerated failure time (AFT) model. The PH model is a popular choice for analyzing failure data in a large number of applications. It has some advantages with respect to its flexibility with respect to the modeling. The AFT model, on the other hand, is popular in engineering applications, where it was originally used to capture changes in the usage of a device.

2. Censoring in the case of field data

One of the specialties of reliability analysis is the presence of censored data. For dedicated experimental data, the most common scheme is right censoring, meaning that devices are still functional at the end of the test, often having one common censoring time horizon. For field data, more complex censoring schemes need to be considered. For typical devices in high voltage installation, one has quite strong right censoring, as they are not run to failure but preventively maintained or replaced. Left truncation, that is, devices failing without being recorded, is expected to be less of a concern for online monitored devices, as the dates of production, installations, and the start of usage are available. For devices that serve as protection devices and are therefore only operated rarely, failures are often only observed after some time, either during a scheduled inspection or during a normal operation. If the time between such operations is long, one needs to treat them as interval-censored. Whereas normal devices' removal from operation might be unnoticed, this is usually not an issue due to the assumed presence of online monitoring. For detailed modeling, the type of failure might also not be available. Additional loss of data is possible for the environment data, e.g., if data from the device is unavailable or lost for some time. Here one would expect these gaps to be filled using historical data recorded before or after the missing interval. Devices that have their online monitoring removed completely, e.g., due to an end of a contract, can be treated as normal right-censored data.

Other situations can be present but are only noted here for completeness: Devices might record their conditions, but data is only made available after a failure, e.g., when devices are repaired, and data is only copied from the device then. A fleet of devices, some with and some without online monitoring, might be available, with failure data recorded for both types. Finally, the choice of online monitoring installed might be linked to, e.g., difficult conditions, which can disturb the analysis, as it assumes the monitored devices represent the general fleet. In the following, we restrict ourselves to a situation that we expect to be typical for a first investigation, taking only the right censoring into account.

3. The Weibull Model

Two different approaches to survival analysis are found in the literature: the reliability curve or its related functions are assumed to be of a parametric form, or a non-parametric approach is taken. Whereas in an electrical system, the number of devices deployed, and therefore available for the analysis, is large, we do not expect to have a large number of failures within the observation time interval. If, in addition, devices are typically exploring very different environmental conditions, a non-parametric approach will, in general, be rather uncertain.

An alternative is using a parametric form, restricting its shape substantially but reducing the uncertainty. The most commonly used model is the Weibull distribution, Abernethy (2005), according to it, the failure probability density function $f(t)$, the reliability function $R(t)$, and the failure probability distribution function $F(t)$ in the following form:

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \quad (1)$$

$$R(t) = 1 - F(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \quad (2)$$

where α is the scale and β the shape parameter. From these the hazard rate $h(t)$ and cumulative hazard rate $H(t)$ can be derived as

$$h(t) = -\frac{d\ln(R(t))}{dt} = \frac{dH(t)}{dt} = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$$

$$H(t) = -\ln(R(t)) = \left(\frac{t}{\alpha}\right)^{\beta}$$

4. Reliability models under variable conditions

The conditions will generally be variable for devices in the field, meaning varying with time. This is in contrast to classical accelerated testing strategies, where mostly constant conditions are used Meeker et al. (2022); Kalbfleisch and Prentice (2002). We first discuss the two most commonly used models for time-constant conditions and their generalization to time-varying ones. These are the proportional hazard (PH) and the accelerated failure time model (AFT). Whereas the PH model is typically used in medical models, the AFT model has been applied extensively for industrial devices.

In the PH model, one assumes a reference hazard rate, which increases or decreases depending on the conditions e:

$$h^{PH}(t) = h_0(t)\,\mathrm{A}(e)$$

where $h_0(t)$ is the base hazard rate as a function of time for some reference conditions e_0, and $\mathrm{A}(e)$ is the acceleration factor depending on the real value of e. If the conditions are constant over time, the reliability function is of the form:

$$R^{PH}(t) = \exp\left(-\,\mathrm{A}(e)\int_0^t h_0(t')dt'\right)$$

In the case of variable conditions $e(t)$, one gets:

$$h^{PH}(t) = h_0(t)\,\mathrm{A}(e(t))$$

leading to

$$R^{PH}(t) = \exp\left(-\int_0^t h_0(t')\,\mathrm{A}(e(t'))dt'\right)$$

In the AFT model, the reliability depends on an "effective time" τ, which depends on e as

$$\tau = t\,\mathrm{B}(e)$$

where $\mathrm{B}(e)$ is the equivalent acceleration factor for the AFT model. The reliability in this model is therefore given by

$$R^{AFT}(t) = R_0(\tau) = R_0(t\,\mathrm{B}(e))$$

and correspondingly

$$h^{AFT}(t) = \mathrm{B}(e)h_0(\mathrm{B}(e)t)$$

Under variable conditions $e(t)$ one gets

$$\tau^{AFT} = \int_0^t \mathrm{B}(e(t'))dt'$$

and

$$R^{AFT}(t) = R_0\left(\tau^{AFT}\right)$$

Both models can be applied independently of the form of the base model of the hazard rate or reliability. Especially for the PH model, non-parametric approaches, e.g., the Kaplan-Meier estimator, exist; this is less common for the AFT model. We are making use here of the Weibull distribution. For constant conditions, one gets the reliability in both models as:

$$R^{PH}(t) = \exp\left(-(t/\alpha)^{\beta}\,\mathrm{A}(e)\right)$$
$$R^{AFT}(t) = \exp\left(-(\mathrm{B}(e)t/\alpha)^{\beta}\right)$$

The two are of the same form, only requiring a redefinition of the acceleration factors according to

$$\mathrm{A}(e) = \mathrm{B}(e)^{\beta}.$$

The Weibull distribution is known to be the only model where PH and AFT models lead to the same form Bagdonavicius and Nikulin (2001). But this property is only valid for constant conditions, whereas for variable ones, one gets for the PH model:

$$R^{PH}(t) = \exp\left(-\int_0^t \mathrm{A}(e(t'))\frac{\beta}{\alpha}\left(\frac{t'}{\alpha}\right)^{\beta-1} dt'\right)$$

and for the AFT model:

$$R^{AFT}(t) = \exp\left(-\left(\frac{\int_0^t \mathrm{B}(e(t'))dt'}{\alpha}\right)^{\beta}\right) \quad (3)$$

Using the relation between the two acceleration factors for constant conditions, one can express the acceleration factor in the PH model $\mathrm{A}(e(t'))$ in terms of the one in the AFT model $\mathrm{B}(e(t'))$ to get a compatible form as

$$R^{PH}(t) = \exp\left(-\int_0^t \frac{\beta}{\alpha^\beta}\left(\mathrm{B}(e(t'))^{\beta}(t')^{\beta-1}dt'\right)\right), \quad (4)$$

showing that the difference is coming from how the integration over the conditions is done.

Note that within the AFT model, the effect of variable conditions is independent of when-in-time they apply, meaning that severe conditions towards the end of life have the same effect as at the beginning. This is in contrast to the PH model, where the increase of the base hazard rate makes severe conditions later more damaging and reduces the lifetime stronger.

In almost all cases, the acceleration factor $\mathrm{A}(e)$ or $\mathrm{B}(e)$ is in the form of a power law, potentially after a reparameterization of e (e.g., $e = \exp(-1/T)$ in the case of thermal effects), leading to

$$B(e) = \left(\frac{e}{e_0}\right)^{\theta} = (e)^{\theta},$$

where e_0 defines some reference conditions, which we will set to one in the following, and θ is the parameter describing the severity of the effect. The power-law form is also true for the acceleration factor $A(e)$ in the PH model, with a simple redefinition of the exponent of θ to $\beta\theta$.

5. Efficient calculation of effective age and cumulative hazard rate

In order to determine the parameters of the acceleration model from field data, one obstacle is the long lifetime and a large amount of data. As the effect is nonlinear, a recalculation is necessary when parameters are changed, making optimization with respect to parameters computationally demanding. It is, therefore, of interest to develop methods that can summarize the varying conditions over a longer time scale while allowing for a change of the model parameters within them.

In the AFT model, one is looking for an efficient way to approximate

$$\tau = \int_0^t (e(t'))^{\theta}\, dt' \quad (5)$$

One solution to this is proposed in Caramia et al. (2000). The integral is replaced by an expectation value over the distribution $p(e)$ of $e(t_i)$ from the sampling done, approximating the rate distribution de/dt over a suitable time interval. This leads to the alternative form of the integral in (5):

$$\tau = t\int (e)^{\theta}\, p(e)de$$

or for increments $\Delta\tau$ over a time interval. This approach is suitable for a single acceleration factor but becomes more cumbersome if more than one is considered. The acceleration factors are multiplied, and it is not expected that they are generally statistically independently distributed. However, this is a good assumption if the influencing factors have independent origins. In other cases, suitable multidimensional distributions need to be considered. In this work, we restrict ourselves to assumed statistically independent factors. In Caramia et al. (2000), the distribution is not further specified.

Here we propose the use of parameterized forms allowing for analytical calculations as an additional step. Two cases are discussed:

If e is positive but unbounded, a Gamma distribution is a suitable model. If e is additionally bounded, especially $e' \in [0, 1]$, the Beta distribution is convenient to be used.

The Gamma distribution is given by

$$p(e|a, b) = \frac{b^a}{\Gamma(a)} e^{a-1} \exp(-be) \qquad (6)$$

where the two parameters a and b can be estimated by the "method of moments" from the mean μ and variance σ^2 of the measurements e_i. In the simplest approach, this is given by

$$a = \frac{\mu^2(e)}{\sigma^2(e)}, \; b = \frac{\mu(e)}{\sigma^2(e)}$$

The expectation value over the distribution is known analytically as

$$E[e^\theta] = \frac{\Gamma(a + \theta)}{\Gamma(a) b^\theta} \qquad (7)$$

Changes in the distribution over time, e.g., due to seasonal effects, can be handled by aggregating mean and variance only over some fixed time periods instead of the whole time series.

The second example follows in the same way: The Beta distribution is given by

$$p(e|a, b) = \frac{1}{B(a, b)} e^{a-1} (1 - e)^{b-1}$$

with the method of moments estimating the two-parameter a and b as

$$a = \mu(e) \left(\frac{\mu(e)(1 - \mu(e))}{\sigma^2(e)} - 1 \right)$$

$$b = (1 - \mu(e)) \left(\frac{\mu(e)(1 - \mu(e))}{\sigma^2(e)} - 1 \right)$$

with the restriction that $\sigma^2(e) < \mu(e)(1 - \mu(e))$. As before, the expectation value over e is known analytically as

$$E[e^\theta] = \frac{B(a + \theta, b)}{B(a, b)} = \frac{\Gamma(a + \theta)\Gamma(a + b)}{\Gamma(a + \theta + b)\Gamma(a)} \qquad (8)$$

In the PH model case, we need to calculate

$$H(t) = \int_0^t \frac{\beta \; (t')^{\beta-1}}{\alpha^\beta} (e(t'))^{\theta\beta} dt'$$

In contrast to the AFT model, we can not separate the averaging over time from the time evolution of the hazard rate. Nevertheless, we expect the same procedure to be applicable as long as we consider time intervals, where $h_0(t)$ changes minimally compared to variations of e. This means an average over days to months is probably still reasonable.

6. Maximum likelihood estimation formulation

Maximum likelihood estimation (MLE) is one of the most common approaches for estimating the parameters of a model using the data. Alternatively, some background knowledge can be incorporated, leading to the maximum a-posterior (MAP) or equivalent to a penalized maximization approach. For simplicity, we consider here only the MLE.

The log-likelihood function $l(\Psi) = \sum_i l_i(\Psi)$ over all devices, including potential censoring, is required. Assuming only potentially right-censored data $c_i \in \{0, 1\}$ one can express it in terms of the hazard and cumulative hazard function using $f(t_i) = h(t_i)R(t_i)$ as

$$l_i(\Psi) = \sum_i (1 - c_i) \ln(h(t_i|\Psi)) - H(t_i|\Psi)$$

where Ψ comprises the Weibull parameters α, β, as well as the one of the acceleration factor(s) θ. Using the Weibull model, we get for the PH model:

$$l_i^{PH}(\Psi) = \textstyle\sum_i (1 - c_i) \ln \left(\frac{\beta t_i^{\beta-1} e(t)^{\beta\theta}}{\alpha^\beta} \right)$$
$$- \int_0^t \frac{\beta}{\alpha} \left(e(t')\right)^{\theta\beta} (t')^{\beta-1} dt' \qquad (9)$$

and for the AFT model:

$$l_i^{AFT}(\Psi) = \textstyle\sum_i (1 - c_i) \cdot$$
$$\ln \left(\frac{\beta e(t)^\theta \left(\tau^{AFT}\right)^{\beta-1}}{\alpha^\beta} \right) - \left(\frac{\tau^{AFT}}{\alpha} \right)^\beta \qquad (10)$$

where

$$\tau^{AFT} = \int_0^t \left(e(t')\right)^\theta dt'.$$

The MLE is then determined by numerical optimization. Please note, that in both models,

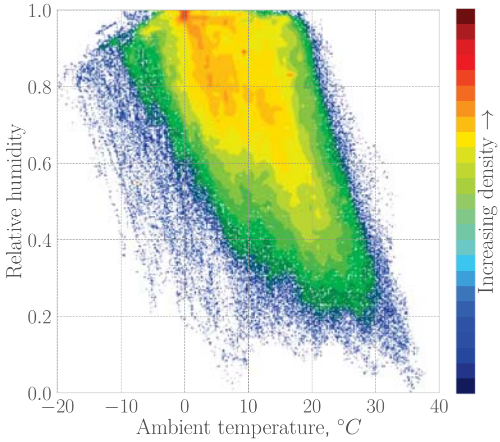

Fig. 1. Density distribution of ambient temperature and relative humidity overall times and weather stations used in the analysis.

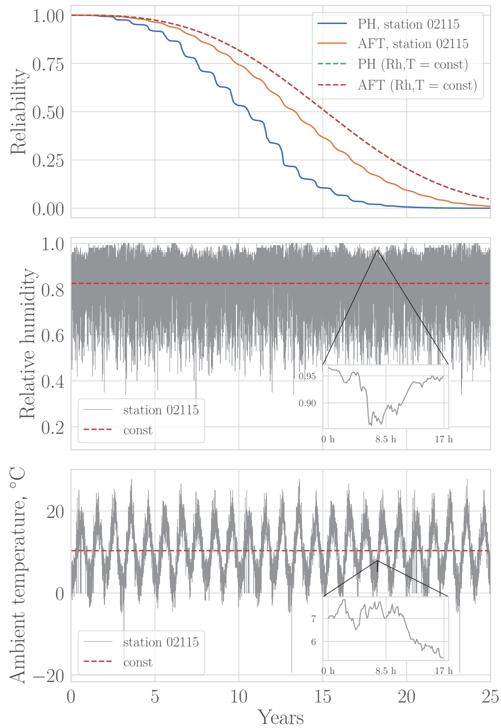

Fig. 2. Top: Reliability curve comparing PH and AFT models for temperature and humidity variability at one weather station, including results using average values. Middle: Relative humidity. Bottom: Ambient temperature over 25 years at weather station 02115.

the derivative with respect to the parameters Ψ can be done analytically, so both gradient- and hessian-based approaches can be used.

7. Application to realistic environmental data and simulated failures

We have applied our approach to simulated failure times using realistic environmental data. Temperature and relative humidity were used as environmental factors, taken from an open database DWD Climate Data Center (CDC) (2019). This database consists of 10-minute measured data from a large number of meteorological weather stations in Germany for a period of ten years, where we have picked a random sample of 107 stations and extended data to 25 years. The distribution of relative humidity and ambient temperature is very broad, see Figure 1; however, it is mostly concentrated between 0°C and 20°C for the temperature and above 0.6 for the relative humidity.

For the failure data, we have used Weibull parameters $\alpha = 5.5$(years) and $\beta = 3$, which are typical values used in a wide range of applications, e.g., for electronics or mechanical components relevant for breakers. Finally, for the acceleration model, we have assumed a Peck-Hallberg model with $M = 2.6$ for the humidity acceleration factor AF_H with relative humidity $RH_{ref} = 0.8$, and $E_a = 0.7$ eV as the activation energy for the temperature acceleration factor AF_T with relative temperature

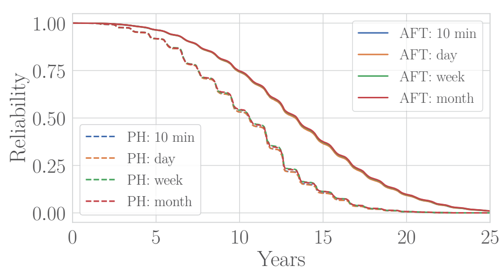

Fig. 3. Comparison of the reliability calculated using the full calculation compared to the binning approach proposed in Sec. 5 with bindings over one day, one week, and one month. Calculations are done for one selected station 02115.

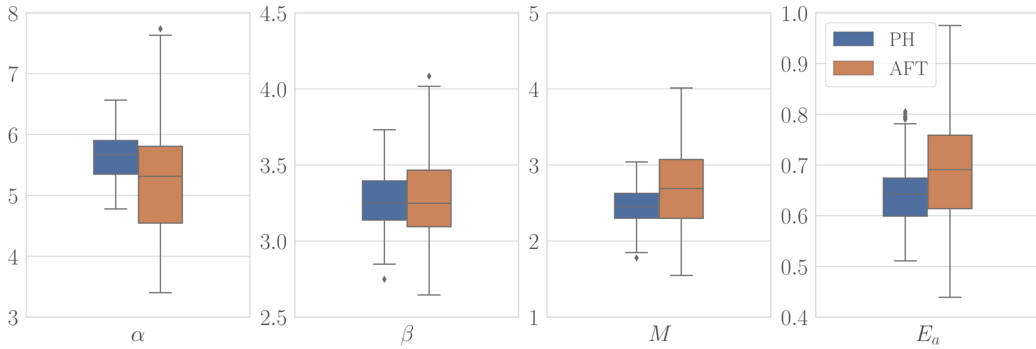

Fig. 4. Box plots of the distribution of α, β, M, and E_a results for the MLE for PH (left box) and AFT (right box), respectively.

$T_{ref} = 25°\mathrm{C}$, according to Feinberg (2016); Hallberg and Peck (1991):

$$AF_H = \left(\frac{RH_{use}}{RH_{ref}}\right)^M$$
$$AF_T = \exp\left(\frac{E_a}{k_B}\left[\frac{1}{T_{ref}} - \frac{1}{T_{use}}\right]\right)$$
$$A(e) = AF_H \cdot AF_T$$

We will call this typical set of α, β, M, and E_a parameters a *ground truth* set of parameters.

An example of the calculated reliability for AFT (Eq. (3)) and PH (Eq. (4)) model, respectively, is presented in Figure 2. As expected, both models give identical results for constant (average) environmental conditions. Reliability under variable conditions has a wavy form due to seasonal variations, mainly due to the ambient temperature variation. It can also be seen that variable conditions have a larger effect on the PH model than on the AFT model. This result can be intuitively explained by the form of the integration in both models, where severe environmental factors within the variation contribute stronger in the integral in Eq. (4) compared to Eq. (3).

As the integration over the variable conditions can go up to 25 years and with the high frequency of data the calculation is rather computationally expensive. We have also explored the accuracy of the approach in Sec. 5. An approximation using the Gamma distribution was used as the temperature effect is positive and unbounded. For the relative humidity, bounded between 0 and 100%, the Beta distribution is appropriate. Binning was done over a day, week, and month (28 days); see Fig. 3. Results show that keeping only mean and standard deviation over the averaged time period agrees well with the original high-frequency detailed 10-minute data. Averaging over a longer period (e.g., over an entire season), the Gamma and Beta assumptions for the distribution of environmental data are no longer valid.

The failure times are sampled from the calculated cumulative distribution function (CDF) of each meteo-station with randomly selected time-horizon between $[2, 10]$ years. Time-horizon corresponds to the assumed current age of the device to censor the data. We chose 10 devices per environmental data set (107 stations), i.e., the total set of failure data consists of 1080 censored and failure times. On average, from the total failure data, the failures take in the range from ∼13% to ∼15% for the PH model and from ∼5 to ∼7% of failure data points for the AFT model. Note that the number of failures is higher in the PH case, as confirmed by the lower reliability curve in Fig. 2.

Using both a numerical optimization of the log-likelihood (Eqs. 9 and 10), we have determined the MLE of the four parameters α, β, M and E_a. The results of a larger number of

Table 1. Final parameters estimations from MLE for PH and AFT models.

	PH		AFT		Ground
	Value	StdDev	Value	StdDev	truth
α	5.635	0.3533	5.289	0.9338	5.5
β	3.269	0.1861	3.288	0.2939	3.0
M	2.467	0.2447	2.718	0.5845	2.66
E_a	0.645	0.0602	0.694	0.1183	0.7

repetitions of the failure time generation and analysis are summarized in Figure 4 and Table 1, giving a hint on the expected accuracy of the approach.

Comparing the loglikelihoods in Eqs. (9) and (10) using failure data generated by the PH model lead to a significant difference between models, which allows us to conclude that the approach is able to distinguish between the two models in the case of variable environmental conditions.

8. Summary and Conclusion

This paper explores a possible approach to parameter estimation of lifetime models using field-deployed devices recording variable environmental conditions and failures for the case of a Weibull distribution and a Peck-Hallberg form to describe the influence of temperature and relative humidity. The aspects studied were the formulation of a model for the two most commonly used approaches to describe acceleration under variable conditions (PH and AFT model). The MLE approach for parameter estimation was used on a set of data using realistic environmental conditions together with simulated failure data.

This work is the first step in using field data (environmental measurements and failure information) to improve the knowledge of the lifetime of devices. In the current framework of this work, a sensitivity analysis of the parameters in both models for generating synthetic failure data has not been conducted. Performing such a sensitivity analysis would necessitate additional computational resources. However, it is important to note that the main objective of this study is to emphasize the significance of accounting for variations observed in the field data. In the case of highly reliable products, where accelerated lifetime testing is time-consuming and costly, this seems to be a viable option, especially with the deployment of connected devices that can collect the data.

References

Abernethy, R. B. (2005). *The New Weibull Handbook: Reliability & Statistical Analysis for Predicting Life, Safety, Survivability, Risk, Cost and Warranty Claims.* Abernethy.

Bagdonavicius, V. and M. Nikulin (2001). *Accelerated Life Models: Modeling and Statistical Analysis (1st ed.).* Chapman and Hall/CRC.

Caramia, P., G. Carpinelli, P. Verde, G. Mazzanti, A. Cavallini, and G. Montanari (2000). An approach to life estimation of electrical plant components in the presence of harmonic distortion. In *Ninth International Conference on Harmonics and Quality of Power. Proceedings*, Volume 3, pp. 887–892.

DWD Climate Data Center (CDC) (2019). Dwd climate data center (cdc): Historical 10-minute station observations of pressure, air temperature (at 5cm and 2m height), humidity and dew point for germany, version v1, 2019. https://opendata.dwd.de/.

Feinberg, A. (2016). *Thermodynamic Degradation Science: Physics of Failure, Accelerated Testing, Fatigue, and Reliability Applications.* John Wiley & Sons.

Hallberg, Ö. and D. S. Peck (1991). Recent humidity accelerations, a base for testing standards. *Quality and Reliability Engineering International* 7(3), 169–180.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data* (2nd ed ed.). Wiley Series in Probability and Statistics. J. Wiley.

Meeker, W., L. Escobar, and F. Pascual (2022). *Statistical Methods for Reliability Data.* Wiley Series in Probability and Statistics. Wiley.