

Master Thesis in Statistics and Data Mining

Survival analysis of gas turbine components

Alessandro Olivi



Division of Statistics
Department of Computer and Information Science
Linköping University

Supervisors

Bertil Wegmann, Daniel Dagnelund, Davood Naderi

Examiner

Linda Wänström

Contents

| | |
|---|-----------|
| Abstract | 1 |
| Acknowledgments | 3 |
| 1. Introduction | 5 |
| 1.1. Background | 5 |
| 1.2. Objective | 6 |
| 2. Data | 9 |
| 2.1. Data sources | 9 |
| 2.2. Data description | 10 |
| 2.2.1. Dataset 1: Analysis of failure modes | 10 |
| 2.2.2. Dataset 2: Multivariate analysis with environmental attributes | 10 |
| 3. Methods | 15 |
| 3.1. Parametric analysis | 15 |
| 3.1.1. Bayesian Weibull AFT Model | 19 |
| 3.1.2. Bayesian Generalized Weibull Model | 20 |
| 3.2. Non-parametric analysis | 21 |
| 3.3. Multiple failure modes | 22 |
| 3.4. Optimal replacement time | 23 |
| 4. Results | 25 |
| 4.1. Failure mode analysis - Dataset 1 | 25 |
| 4.2. Multivariate survival analysis - Dataset 2 | 29 |
| 4.2.1. Categorical variables | 29 |
| 4.2.2. Bayesian Weibull AFT model | 30 |
| 4.2.3. Bayesian Generalized Weibull model | 31 |
| 5. Discussion | 37 |
| 6. Conclusions | 41 |
| A. Failure mode analysis | 43 |
| B. Density and MCMC diagnostics plots for the Bayesian Weibull AFT Model | 45 |

| | |
|---|-----------|
| C. Stan code for the Bayesian Generalized Weibull Model | 47 |
| D. Density and MCMC diagnostics plots for the Bayesian Generalized Weibull Model | 49 |
| Bibliography | 51 |

Abstract

Survival analysis is applied on mechanical components installed in gas turbines. We use field experience data collected from repair inspection reports. These data are highly censored since the exact time-to-event is unknown. We only know that it lies before or after the repair inspection time. As event we consider irreparability level of the mechanical components. The aim is to estimate survival functions that depend on the different environmental attributes of the sites where the gas turbines operate. Then, the goal is to use this information to obtain optimal time points for preventive maintenance. Optimal times are calculated with respect to the minimization of a cost function which considers expected costs of preventive and corrective maintenance. Another aim is the investigation of the effect of five different failure modes on the component lifetime. The methods used are based on the Weibull distribution, in particular we apply the Bayesian Weibull AFT model and the Bayesian Generalized Weibull model. The latter is preferable for its greater flexibility and better performance. Results reveal that components from gas turbines located in a heavy industrial environment at a higher distance from sea tend to have shorter lifetime. Then, failure mode A seems to be the most harmful for the component lifetime. The model used is capable of predicting customer-specific optimal replacement times based on the effect of environmental attributes. Predictions can be also extended for new components installed at new customer sites.

Keywords: Bayesian Weibull regression, failure modes, optimal replacement time, reliability, survival analysis.

Acknowledgments

I would like to thank Siemens Industrial Turbomachinery AB for giving me the opportunity to work with them, as well as for providing the data for this thesis project.

In particular, I would like to express my deep gratitude to my supervisors Daniel Dagnelund and Davood Naderi for the support and guidance during the planning and development of this research work.

I would also like to thank the manager Erik Ärlebäck and the expert Pontus Slottner for the valuable suggestions during my thesis project.

Then, my special thanks are extended to my supervisor from Linköping University Bertil Wegmann for the support during the project and for the time spent to review my thesis.

Thanks also to my opponent Jillahoma musa Wemmao for the useful comments provided at the revision meeting.

I am then particularly grateful to all the students of Statistics and Data Mining of Linköping University for the unforgettable moments spent together during the Master's studies.

Finally, I would like to express my very great appreciation to my family and Andrea. This thesis and these Master's studies would not have been possible without their continuous encouragement and support.

1. Introduction

1.1. Background

Survival analysis involves the modelling of time-to-event data. Events are usually referred to death in biological systems and failure in mechanical systems, but can also be divorce, relapse of a disease, insurance claim, and so on. The topic is also called reliability analysis in engineering. One of the first contributor on the topic of statistical survival analysis was the Swedish engineer, scientist, and mathematician Waloddi Weibull (1887-1979). He invented the Weibull distribution which is the most commonly used distribution to model lifetimes. Weibull (2013) discussed fundamental concepts of fatigue testing as well as statistical methods to predict reliability of various components. A general description of Weibull survival analysis can be found in Abernethy (2006). The book can be considered as a practical guide to reliability life-data analysis since it gathers methods to predict failure times as well as optimal parts replacement intervals for minimizing costs.

In our study the event corresponds to irreparability of mechanical components from gas turbines manufactured by Siemens Industrial Turbomachinery AB (SIT AB). This kind of components have a limited lifetime and must be replaced at a certain time. Irreparability occurs when the damage exceeds some predefined criteria and so it must be rejected and replaced. Regular inspections are carried out to check whether the component exceeds or not the predefined criteria. The idea is to prevent catastrophic failures and unplanned replacements since it would be much more costly for the company. Preventive replacement is the maintenance philosophy for Siemens, and the aim is to prevent the damage before it actually occurs. Currently, replacement is regularly done according to the application of deterministic models developed by the engineering department. The main idea in this thesis is to turn these models into a more probabilistic approach to include elements of randomness.

Censored data is a form of incomplete data where the exact failure time is unknown. In our analysis we only know that irreparability can happen before or after the time of inspection. Standard methods from Weibull (2013) cannot deal with such data. However, non parametric methods like the Turnbull estimator (Turnbull, 1976) can handle this type of censored data. Parametric methods based on the Weibull distribution need the computation of the full likelihood to accommodate all kinds of censoring as in Nánási (2014). The Weibull parameters can then be estimated by directly maximizing the likelihood, or by following a Bayesian estimation approach. Bayesian methods on interval-censored data for the Weibull

distribution are described in Mohammed Ahmed (2014) where gamma priors are selected for the Weibull coefficients. The paper compares Bayesian and Maximum Likelihood estimation methods and shows that Bayesian estimation performs better in terms of prediction accuracy. In our analysis we follow the Bayesian philosophy and derive the full posterior distribution of the parameters. This makes it easier to derive credibility intervals for the estimates to then propagate the uncertainty to the associated cost function. More specific studies about lifetime estimation of gas turbine components can be found in Li et al. (2015), where the lifetime of gas turbine blades is estimated with Bayesian models. Analysis of field experience data on aircraft turbine blades is carried out in Zaretsky et al. (2013), but it is assumed that the last failure has happened exactly at the time of inspection, and this seems to be quite a strong assumption for our analysis.

1.2. Objective

In this study we apply statistical methods to predict the lifetime, or generally the survival distribution, of mechanical components installed in gas turbines located in different sites and environments. A set of attributes describes the different features of the environments where the gas turbines operate. We try to quantify the impact of these attributes on the lifetime. Based on the survival distributions we then estimate the optimal times where it is more convenient to carry out preventive maintenance. A cost function is thus considered for the derivation of optimal replacement times. Matter of interest is also the analysis of the impact of five different failure modes in the component lifetime. We follow a Bayesian approach for the parametric model estimation so we are able to first infer our prior belief, and second derive the whole posterior distribution of the parameters. It is then easy to derive credibility intervals on the estimates and propagate them to the cost function.

To summarize, this thesis aims to:

- Estimate the lifetime of mechanical components from gas turbines located at different sites, and evaluate the optimal inspection/replacement time which minimizes losses with respect to a predefined cost function.
- Quantify the influence of environmental attributes on the component lifetime. Attributes can be of various kinds (ambient temperature, altitude, pollution classification, . . .) and forms (categorical, continuous). The idea is to investigate how these attributes affect the component lifetime.
- Investigate how different failure modes affect the component lifetime, and find which are the most harmful ones.

The thesis is structured as follows: Chapter 2 includes a general description of field and censored data as well as a specific description of the datasets used in this analysis. Chapter 3 shows the methodology applied in this work. Results of the

analysis are reported in Chapter 4 with the help of plots and tables. It follows then a discussion of the most relevant results in Chapter 5. Finally, we try to draw some conclusions as well as suggestions for future work in Chapter 6.

2. Data

2.1. Data sources

This thesis work is based on the statistical analysis of field data collected by Siemens Industrial Turbomachinery AB (SIT AB) located in Finspång, Sweden. SIT AB is part of the Siemens Energy Sector that supplies products and services for the generation, transmission and distribution of power, and for the extraction, conversion and transport of oil and gas. SIT AB delivers gas turbines as well as maintenance services. SIT AB continuously gathers a large amount of field experience data in form of various reports from repairs, inspections and service. All the information from field data is stored and organized in a common SQL-database called CompoSIT, which makes data extraction easy to apply for analysis purposes.

The analysis of field data is usually complicated by the fact that they are contaminated by customer behaviors. Lab data instead are collected under well-controlled testing conditions and are thus homogeneous (Ye and Ng, 2014). Gas turbines operate in heterogeneous conditions because they are used in different geographical locations and environmental conditions (e.g. temperature, humidity, pollution, etc.). In addition, different customers have different usage behaviors (e.g. workload, type of fuel or lube oil used, etc.). Such conditions should therefore be considered in making survival analysis. Otherwise the time measure would be biased and not fully appropriate as a measure of lifetime (Barabadi, 2013). For example, a turbine running 2 hours at the highest load is expected to have a lower lifetime compared to another turbine running 2 hours at idle conditions. To deal with this discrepancy a new multi-dimensional time measure was developed at SIT AB: the Equivalent Operating Hours (EOH). EOH is calculated according to a deterministic formula that combines the usual operating hours (OH) with the effect of internal temperature, cycles, fuel type, and so on. In this way, the bias caused by the heterogeneous conditions should be reduced leading to a more reliable lifetime measure. In the previous example the EOH of the turbine operating at full load would be higher than the EOH of the turbine running at idle conditions.

One challenge in this analysis is determined by the highly censored nature of the time data. Censoring means that the time-to-event is only partially known. We do not know the exact time when the event occurred on the failed component. We only know that it occurred before the inspection time. In addition, the dataset contains a lot of right censored observations, also called suspensions, that are those components that have not yet experienced a failure event at the time of inspection.

In this case we only know that the event must occur after the time of inspection. In general, different types of censoring exist:

- *Right censoring*: the event occurs after the time of inspection but it is unknown by how much.
- *Interval censoring*: the event occurred somewhere on an interval between two time inspections.
- *Left censoring*: the event occurred before the time of inspection but it is unknown by how much. This kind of censoring could be considered as a special case of interval censoring when the interval is $(0, t]$, where t is the time of inspection.

In our analysis we only dispose of left and right censored data. Standard methods in survival analysis can accommodate only exact and right-censored time-to-event. To deal with interval and left censored data special methods or extensions must be considered (Singh and Totawattage, 2013). In this study we analyze reliability of mechanical components installed in Siemens gas turbines. These components have a limited lifetime since they are subjected to extreme conditions like high temperature and high flow. We base our analysis on two datasets. The first one includes information about which failure mode caused the component to become irreparable. The second one includes a set of covariates related to the environment where the turbine operates. The response variable is the time expressed in EOH, that is censored since we do not know the exact failure time. We discard the use of OH because this time measure is potentially biased by different working conditions. Most of the observations are right censored, the rest are left censored.

2.2. Data description

2.2.1. Dataset 1: Analysis of failure modes

The first dataset contains information about 1937 components where 77 (4%) of them encountered irreparability whereas the other ones were found still in good conditions. We know the failure mode that made the component to become irreparable. In particular, 5 different failure modes can cause irreparability, and we call them A, B, C, D, E. Each component cannot fail for multiple reasons, but only for one. Table 2.1 shows how dataset 1 is structured.

2.2.2. Dataset 2: Multivariate analysis with environmental attributes

Dataset 2 includes data about 40 sets of components where each set is generally composed of 52 components. For each set we know the number of components

| Inspection time (EOH) | Failure mode | | | | |
|-----------------------|--------------|-----|-----|-----|-----|
| | A | B | C | D | E |
| 18161 | 1 | 0 | 0 | 0 | 0 |
| 17161 | 0 | 0 | 0 | 0 | 0 |
| 13442 | 0 | 0 | 0 | 0 | 0 |
| 20707 | 0 | 1 | 0 | 0 | 0 |
| 21823 | 0 | 0 | 0 | 0 | 1 |
| 19669 | 0 | 0 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... |

Table 2.1.: Extract of dataset 1: *Practical example*: the first line says that a component is found failed due to failure mode A at the inspection done at 18161 EOH. The second line shows that a component is still in good conditions at 17161 EOH.

that are found failed and those that are still in good conditions at the time of inspection. Therefore, the components found failed are left-censored observations since we only know they failed before the time of inspection. The components still in good conditions are instead right-censored observations since we can only say that they will fail after the time of inspection. A set of covariates describes the environmental conditions where the turbines are located. The covariates are listed in Table 2.2 with summary statistics like minimum, mean, and maximum value for the continuous variables, and frequencies for the categorical variables. The summary statistics of the continuous variables are displayed in the original scale. However, continuous covariates are scaled to mean 0 and standard deviation 1 before applying statistical models in order to improve the efficiency of the estimation and to better interpret the model coefficients. Also the time response variable is divided by 100000 to avoid too large numbers in the estimation.

The mean temperatures and relative humidities are raw estimates derived by the mean between the minimum and the maximum value recorded both in January and July. We think that considering both minimum and maximum values would be redundant, therefore we opt for the mean only. The month of January can be representative for the winter periods, and the month of July for the summer periods. Notice that we invert these attributes for those observations related to the southern hemisphere where summer is in January and winter in July. For the same reason we consider absolute values of the latitude. The pollution classification categorizes different kinds of environment according to some predefined criteria. From the least to the most polluted category, the environment is classified in: clean, urban, industrial, heavy industrial. Then, a binary variable describes the presence or not of salinity in the air.

The censored inspection times of dataset 2 are clearly visualized in Figure 2.1. Horizontal lines represent the sets of components. Red solid intervals are the graphical representations of left censored observations (equivalently, interval censored times

between time of installation $t = 0$ and time of inspection). Gray dashed intervals describe right censored observations. The number of left and right censored times is reported in the plot near each inspection time. Notice that the red solid intervals are quite large due to the left-censored nature of the data. In addition, they overlap for most of the time. These conditions potentially lead to increase the amount of uncertainty in the analysis results.

| Covariate | Type | Unit | Summary statistics / Frequencies |
|--------------------------------|-------------|---------|--|
| Mean temperature January | Continuous | °C | Min: -19.5, Mean: 2.562, Max: 19 |
| Mean temperature July | Continuous | °C | Min: 2, Mean: 19.61, Max: 27 |
| Mean relative humidity January | Continuous | % | Min: 64, Mean: 78.38, Max: 87.5 |
| Mean relative humidity July | Continuous | % | Min: 56.5, Mean: 72.56, Max: 84.5 |
| Distance from sea | Continuous | km | Min: 5, Mean: 152.3, Max: 621.5 |
| Altitude | Continuous | m | Min: 0, Mean: 112.79, Max: 378 |
| Latitude | Continuous | ° [deg] | Min: 26.67, Mean: 46.67, Max: 61.13 |
| Pollution Classification | Categorical | | Clean: 3, Urban: 5, Industrial: 15, Heavy industrial: 17 |
| Salt Exposure | Binary | | No: 34, Yes: 6 |

Table 2.2.: List of covariates followed by the type, the unit measure, and some summary statistics (for the continuous variables) or level frequencies (for the categorical variables).

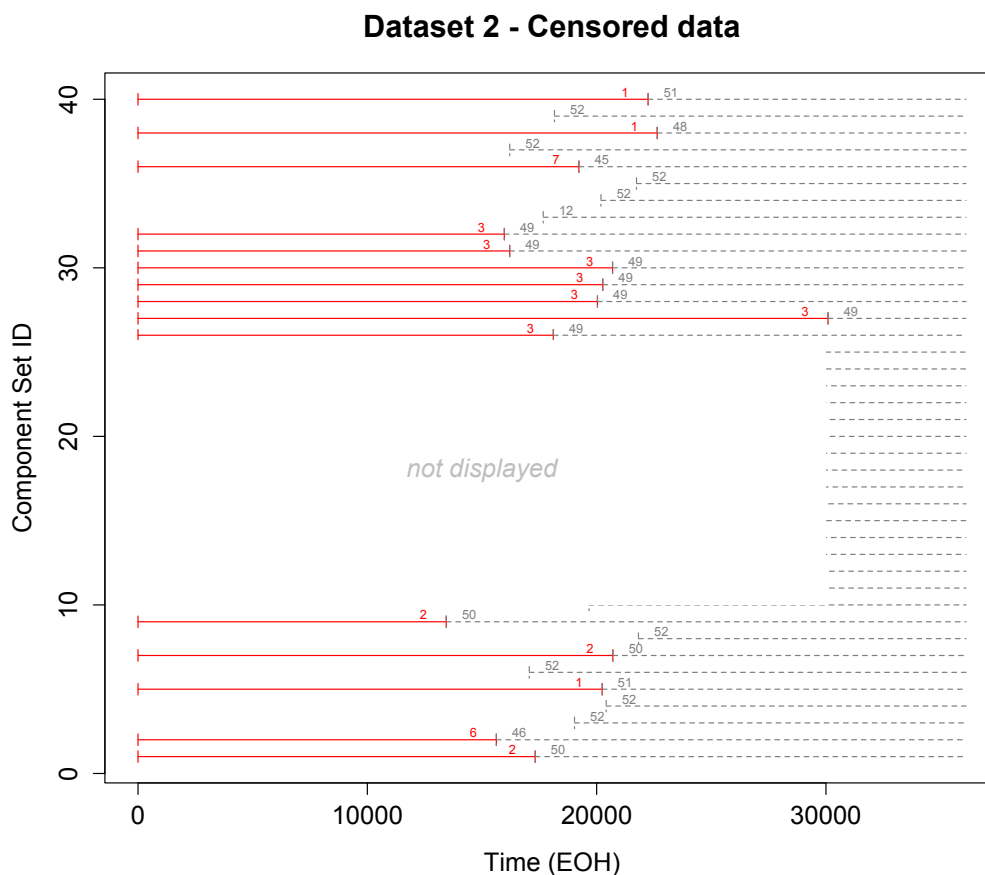


Figure 2.1.: Visualization of censored times from dataset 2. Horizontal lines represent the sets of components. Red solid intervals are the graphical representations of left censored observations (equivalently, interval censored times between 0 and inspection times). Gray dashed intervals describe right censored observations. The number of left and right censored times is reported in the plot near each inspection time. *A practical example:* Component Set ID 40 was inspected at around 21000 EOH and at that time 1 component was found beyond reparability limits, whereas 51 components were still in healthy conditions.

3. Methods

This chapter describes the statistical methods used in this analysis. We start in Section 3.1 with the description of the Weibull distribution as well as parametric models like the Bayesian Weibull AFT model and the Bayesian Generalized Weibull model. Non parametric methods like the Turnbull estimator are described in Section 3.2. We then report the main concepts of multiple failure mode analysis in Section 3.3. Finally, in Section 3.4 we describe the preventive maintenance philosophy and we propose the use of a cost function to derive the optimal replacement time.

3.1. Parametric analysis

Parametric survival analysis assumes that time-to-event data follow a predefined distribution based on a set of parameters. The most commonly used distribution in survival analysis is the Weibull distribution and it has been successfully used in the past to model the lifetimes of electronic components, relays, and ball bearings (Weibull, 2013). The Weibull density function $f(t)$ is as follows:

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\eta}\right)^\beta\right\} \quad (3.1)$$

The Weibull cumulative distribution function $F(t)$ describes the association between the event probability and the lifetime. More formally, it corresponds to the probability that the event occurs before the time point t . It takes the following form:

$$F(t) = P(T < t) = 1 - \exp\left\{-\left(\frac{t}{\eta}\right)^\beta\right\} \quad \text{for } t \geq 0 \quad (3.2)$$

This form can be linearized by applying logarithmic transformations twice (Abernethy, 2006). This is useful for visualization purposes. The transformation is as follows:

$$\ln(-\ln(1 - F(t))) = \beta \ln(t) - \beta \ln(\eta) \quad (3.3)$$

The complementary form $1 - F(t)$ associates the probability to survive longer than t , and is called the survival function $S(t)$ or the reliability function $R(t)$:

$$S(t) = 1 - F(t) = P(T > t) = \exp \left\{ - \left(\frac{t}{\eta} \right)^\beta \right\} \quad \text{for } t \geq 0 \quad (3.4)$$

The hazard function $h(t)$ describes the instantaneous failure rate at any point in time t . The hazard function is equal to the ratio between the density and the survival function, and assuming the Weibull distribution for time-to-event leads to:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \quad (3.5)$$

The Weibull distribution is defined by two positive parameters: the shape β and the scale η . The high flexibility of the Weibull distribution allows to model hazard functions with decreasing, increasing, and constant failure rate, according to the value of β . In particular:

- $\beta < 1$ implies infant failures: failure rate decreases as time increases. This can be due to inadequate burn-in tests, misassembly, quality control problems, or bad raw materials from the suppliers (Abernethy, 2006).
- $\beta = 1$ implies random failures with a constant failure rate. Thus, no relationships between time and event. In this case the Weibull distribution corresponds to the exponential distribution. Random failures are more typical for electronic components (Abernethy, 2006).
- $\beta > 1$ means wear-out life: failure rate increases as time increases. In mechanical components this is mainly due to thermal-mechanical fatigue, stress, corrosion, or degradation process (Abernethy, 2006).

The scale parameter η is also called characteristic life (Abernethy, 2006) and it corresponds to the time t when the event probability is approximately 0.632.

The Weibull distribution is often visually assessed by using the so called Weibull probability plot, where the axes are scaled according to Equation 3.3 so that $F(t)$ takes the form of a straight line. In this representation, β is also called slope since it gives an indication of the steepness of the straight line (Abernethy, 2006).

The parameters can be estimated by maximizing the log-likelihood function. Abernethy (2006) recommends the median rank regression (MRR) with the ordinary least-squares estimation method, and he states that the maximum likelihood estimation (MLE) is optimistically biased when the number of observations is less than 500 with fewer than 100 failures. However, more recent simulation studies carried out by Genschel and Meeker (2010) found that MLE is a more appropriate method even with small samples. In addition, the full Weibull log-likelihood function can easily handle all kinds of censoring in the data, whereas MRR cannot do. The full

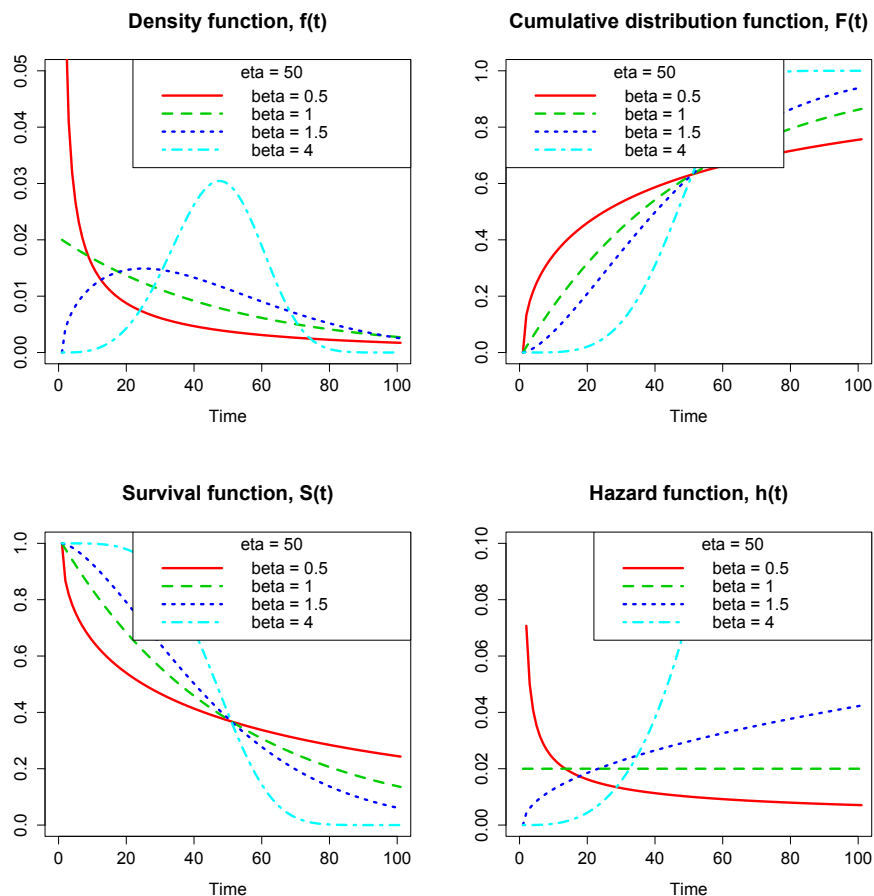


Figure 3.1.: Some examples of the Weibull density, distribution, survival, and hazard functions for different values of β .

Weibull log-likelihood function is composed of three main parts and takes this form (Abernethy, 2006):

$$l = \sum_{i=1}^F \ln \left[\frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} e^{-\left(\frac{t_i}{\eta} \right)^\beta} \right] - \sum_{i=1}^S \left(\frac{t_i}{\eta} \right)^\beta + \sum_{i=1}^I \ln \left[e^{-\left(\frac{t_{Li}}{\eta} \right)^\beta} - e^{-\left(\frac{t_{Ri}}{\eta} \right)^\beta} \right] \quad (3.6)$$

The first part handles exact failure times t_i for $i = 1, \dots, F$, the second part includes suspensions (right censored observations) t_i for $i = 1, \dots, S$, and the third part handles interval data $[t_{Li}, t_{Ri}]$ for $i = 1, \dots, I$. Notice that left-censored data is a special kind of interval-censored data when $t_{Li} = 0$, leading to $e^{-\left(\frac{t_{Li}}{\eta} \right)^\beta} = 1$ in Equation 3.6. Since we only have right and left censored data, the log-likelihood reduces to:

$$l = - \sum_{i=1}^S \left(\frac{t_i}{\eta} \right)^\beta + \sum_{i=1}^I \ln \left[1 - e^{-\left(\frac{t_{Ri}}{\eta} \right)^\beta} \right] \quad (3.7)$$

In our parametric analysis we assume that the lifetime of the components are independent and Weibull distributed. We apply the Kolmogorov-Smirnov test and the Anderson-Darling test to check this assumption (Stephens, 1974). Both test results do not reject the Weibull assumption. We follow a Bayesian approach for the estimation of the parameters β and η . Compared to the frequentist approach, it has the advantage to combine our prior belief with the observed data, to then get the full posterior distribution of the parameters. Knowing the posterior distribution it is then easier to calculate credible intervals on the parameters, and propagate them to the cost function and to the system survival function. It would be hard to derive uncertainty measures on the cost function if we only rely on point estimates calculated through frequentist methods.

For the analysis of dataset 1, we select normal priors on the logarithms of β and η :

$$p(\ln(\beta)) = \mathcal{N}(0, 3) \quad (3.8)$$

$$p(\ln(\eta)) = \mathcal{N}(0, 3) \quad (3.9)$$

Logarithmic transformation is used to map the parameters to the whole real space, and so to make the application of a normal prior reasonable. In fact, β and η must be positive in the original linear scale. The prior mean is set to 0 to center the parameters to 1, since $\exp(0) = 1$. We then select quite a high value (=3) for the prior standard deviation to make the prior more vague and thus less informative since we have little prior knowledge on the parameters.

Based on this prior and on the observed data we sample from the posterior distribution through Monte Carlo Markov Chain (MCMC) iterations by using the software RStan (Stan Development Team, 2014). A closed form of the posterior distribution is not available when both parameters are treated as random variable. Some conjugacy exists when one of the parameters are known (Rinne, 2009), but this is not our case since we want to estimate both parameters. We derive posterior median values as point estimates, as well as 95% credible intervals. We prefer posterior medians instead of means because some posterior distributions show asymmetric shapes. The mean is potentially biased by extreme values at the tail of the posterior distribution.

The parametric models are trained and estimated according to a cross-validation scheme. Due to the small sample size (40 obs.) we opt for a leave-one-out cross-validation (LOOCV) which give more reliable estimates despite being more computationally expensive. The models are compared according to measures like AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria):

$$AIC = 2p - 2\hat{l} \quad (3.10)$$

$$BIC = -2\hat{l} + p \ln(n) \quad (3.11)$$

where p is the number of variables, \hat{l} is the maximized value of the log-likelihood function l (Equation 3.7), and n is the number of observations. In terms of model comparison, the model with the lowest value of AIC or BIC is preferred. It is relevant to note that the BIC criterion may not be reliable for our data because n is not much larger than p (Giraud, 2014).

3.1.1. Bayesian Weibull AFT Model

The Weibull Accelerated Failure Time (AFT) model allows to include and quantify the effect of covariates on the survival time. The underlying assumption of an AFT model is that the effect of covariates is multiplicative with respect to the survival time. Consequently, the AFT model describes the acceleration (or contraction) of the survival time as a function of covariates. To get the idea of the AFT assumption we can take the lifespan of dogs as an example (Kleinbaum and Klein, 2005). In this scenario it is likely to say that dogs accelerate through the lifetime 7 times faster than humans. More formally, the probability of a dog surviving 10 years is equal to the probability of a human surviving 70 years, and $70/10 = 7$ is the multiplicative factor, also called acceleration factor, which is constant over the time.

In the Weibull AFT model the parameter η is reparametrized as:

$$\eta = \exp(b_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \dots + b_p\mathbf{x}_p) \quad (3.12)$$

where $\mathbf{b} = b_0, \dots, b_p$ is the set of coefficients that measures the impact of the p covariates \mathbf{x}_i on the lifetime. According to the AFT assumption, the parameter β does not vary with the covariates, so it is simply reparametrized as $\exp(a_0)$ for inference purposes:

$$\beta = \exp(a_0) \quad (3.13)$$

We follow a Bayesian approach for the parameter estimation, thus we select a prior distribution for the coefficient a_0 and \mathbf{b} and we make inference on the posterior distribution. We select a normal prior on a_0 and a multivariate normal g-prior on \mathbf{b} :

$$p(a_0) = \mathcal{N}(0, 3) \quad (3.14)$$

$$p(\mathbf{b}) = \mathcal{N}_{p+1}(\mathbf{0}, g(\mathbf{X}^T\mathbf{X})^{-1}) \quad (3.15)$$

where g is a positive scalar, and \mathbf{X} is the matrix of explanatory variables (intercept included) of dimension $n \times (p + 1)$.

The hyperparameters for the normal prior in Equation 3.14 leads to a quite diffuse and uninformative prior, as explained in Section 3.1. The g-prior developed by

Zellner (1986) is a popular way to introduce prior information in a multivariate regression model. We set g to be equal to n which leads to the unit informative prior (Kass and Wasserman, 1996). This is a weakly informative prior where the amount of information in the prior corresponds to the amount of information in only one observation.

A closed form for the posterior distribution is not available, so we draw parameter values from the posterior distribution by using MCMC methods with the software RStan. We can see from Figure B.1 that all the posterior distributions from the sampled coefficients are symmetric and bell-shaped. Posterior means are considered as best point estimates and 95% credible intervals are also derived. Interpretation of the coefficients is quite straightforward in this model: a positive coefficient indicates that the covariate is positively correlated with the survival, whereas a negative coefficient means that the covariate is negatively correlated to the survival. In other words, a positive coefficient is beneficial for the survival, a negative coefficient is harmful for the survival, and a coefficient close to zero indicates no relationship between the covariate and the survival.

3.1.2. Bayesian Generalized Weibull Model

The Weibull AFT Model relies on the AFT assumption that may not hold. We use a more general Weibull model where we let both parameters β and η to depend on covariates. Therefore, both β and η are reparametrized to vary with covariates. The parametrization of η is the same as in Equation 3.12, whereas β is reparametrized as:

$$\beta = \exp(a_0 + a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p) \quad (3.16)$$

where $\mathbf{a} = a_0, \dots, a_p$ is the set of coefficients that measures the impact of the p covariates \mathbf{x}_i on β .

There are not so many studies in the literature regarding this kind of model, but in Kleinbaum and Klein (2005) it is mentioned as an alternative Weibull model where the AFT assumption is violated. In Zuehlke (2013) it is called the “*Non-proportional Weibull Hazard model*”. We select multivariate normal g-priors (Zellner, 1986) on \mathbf{a} and \mathbf{b} with $g_1 = g_2 = n$ to make them little informative (Kass and Wasserman, 1996):

$$p(\mathbf{a}) = \mathcal{N}_{p+1}(\mathbf{0}, g_1(\mathbf{X}^T\mathbf{X})^{-1}) \quad (3.17)$$

$$p(\mathbf{b}) = \mathcal{N}_{p+1}(\mathbf{0}, g_2(\mathbf{X}^T\mathbf{X})^{-1}) \quad (3.18)$$

Again, we sample from the posterior distributions through MCMC methods in RStan. Code can be found in Appendix C. All the posterior distributions from

the sampled coefficients show bell shapes and so uni-modality (see Appendix D), however, some of them reveal asymmetry. For this reason we select posterior medians instead of means as the most representative point estimates. The mean is potentially biased by the extreme values at the tail of the distributions. We also derive 95% credible intervals.

3.2. Non-parametric analysis

Parametric analysis has the advantage to be highly powerful, but it relies on the assumption of a survival distribution that may not be appropriate. For this reason non-parametric estimates can be considered since they are based on ranking methods that are distribution-free.

One common non-parametric method, especially in Weibull analysis, is the median rank. The median rank is calculated using the incomplete Beta distribution, but can be simplified by the Benard's approximation:

$$\frac{i - 0.3}{n + 0.4} \tag{3.19}$$

where i is the rank order and n is the total number of failures. The rank i must be adjusted to accommodate suspensions (Abernethy, 2006).

Another common non-parametric tool for survival analysis is the Kaplan-Meier estimator (Kaplan and Meier, 1958). It is a step function with discontinuities or jumps at the observed event times:

$$\hat{F}(t) = 1 - \prod_{t_i < t} 1 - \frac{d_i}{n_i} \tag{3.20}$$

where d_i is the number of events happened at time $t_{(i)}$, and n_i is the number alive just before $t_{(i)}$.

Both median rank and Kaplan-Meier estimates can deal with suspensions, but cannot handle with left-censored or interval data. With interval data common practice was to calculate estimates on both the left and the right time, or alternatively consider the mid-points among the interval time. However, according to Giolo (2004) and Singh and Totawattage (2013) this can lead to invalid inferences. In particular, Law and Brookmeyer (1992) showed that the statistical properties of midpoint imputation strongly depend on the width of the time interval. If the interval is wide, mid-point estimation would lead to biased outcomes. According to Kim (2003), also the standard error would be underestimated since the mid-point estimation assumes that the event occurs exactly at the midpoint without considering that it lies among the whole interval. Using a non-parametric maximum likelihood estimator (NPMLE) is a more appropriate choice to handle interval censored data where the

intervals are large and overlapping like in our case. Peto (1973) was the first scientist who developed a NPMLE estimator for interval data, but it seems that the method tends to concentrate much of the probability mass at the endpoints of the intervals (Ng, 2002). In this analysis we apply the estimator derived by Turnbull (1976). It is similar to the method developed by Peto (1973) but it uses an iterative self-consistent algorithm to estimate the event-probability at time t . The survival curve can only jump on a set of disjoint sub-intervals, also called equivalence sets, defined by the potentially overlapping intervals in which the events are known to have occurred (Singh and Totawatattage, 2013).

3.3. Multiple failure modes

Mechanical components can become irreparable for different causes. In our data we have information about 5 different failure modes and for each failed component we know which failure mode made the component to become irreparable. We assumed independence between the failure modes. We checked the assumption of independence both by computing correlations between the failure modes, and by visually inspecting scatterplot pairs. In this scenario we can apply single reliability analysis for each sub-population and then calculate the reliability function of the whole system that includes the effect of all the failure modes. Formally, we model the distribution of the minimum time-to-event $T \sim \min\{T_1, \dots, T_n\}$, where T_i are the times related to each sub-population from failure mode i . When there is dependence between failure modes, conditional probabilities must be taken into account to calculate the reliability of a series system:

$$T \sim P(T_1 \cap T_2 \cap \dots \cap T_n) = P(T_1)P(T_2 | T_1)P(T_3 | T_1 T_2) \cdots P(T_n | T_1 T_2 \dots T_{n-1}) \quad (3.21)$$

In our case T_i are assumed to be independent and Weibull distributed. When independence is assumed, the distribution of the minimum T is simply the product of the single reliability functions (Crowder, 2012):

$$T \sim \prod_{i=1}^n P(T_i > t) = \prod_{i=1}^n R_i(t) = \prod_{i=1}^n (1 - F_i(t)) \quad (3.22)$$

We apply these concepts on dataset 1 to derive the resulting survival function that includes the effects of all the failure modes. First, we run a single Weibull analysis on each failure mode to find the survival function which describes how that specific failure mode affects the component lifetime. Then, we calculate the system reliability function as the product of each single survival functions related to each failure mode as in Equation 3.22.

3.4. Optimal replacement time

The distribution function describes the probability that the event occurs before a specific time point. However, for this function we would like to find the optimal time when component replacement is more convenient (or less risky). One way is to arbitrarily choose a level of risk among the distribution function. For example, we could choose the time when the event-probability is 5%. In other words, that would be the time when the probability that the component becomes irreparable is 5%. However, selecting the most appropriate level of risk is not straightforward: 5% is an arbitrary choice and may not be the optimal one.

A more suitable way to estimate the optimal replacement time for a specific gas turbine is to define a cost function which take both corrective and preventive replacement costs into account. Corrective replacement occurs when the turbine stops to work suddenly, and the customer has to call the maintenance service to get some help to fix the problem. It is an unplanned kind of maintenance and it is undesirable by Siemens because it is much more costly compared to a planned maintenance. On the other hand, preventive maintenance occurs regularly according to planned replacement programs. The primary goal of preventive maintenance is to prevent the failure of components before it actually occurs. The ideal preventive maintenance program would prevent all corrective maintenance actions. Long-term benefits of preventive maintenance include: improved system reliability, decreased cost of replacement, decreased system downtime, and better spares inventory management (McCool, 2012).

One possible model to associate costs and risks with the lifetime is described in Abernethy (2006) and McCool (2012) and takes this form:

$$c(t) = \frac{c_1 F(t) + c_2(1 - F(t))}{\int_0^t 1 - F(\tau) d\tau} \quad (3.23)$$

where c_1 is the cost of a corrective replacement, and c_2 is the cost of a preventive replacement. $F(t)$ can be assumed to be the distribution function of the Weibull distribution, or any other distribution. The denominator acts as a normalization variable to derive the cost per unit of time. Two conditions must hold to make preventive maintenance philosophy reasonable to apply:

1. The corrective replacement cost c_1 must be greater or much greater than the preventive replacement cost c_2 .
2. The hazard rate must increase with time, thus $\beta > 1$ if $F(t)$ follows a Weibull distribution.

If these conditions do not hold, the cost function $c(t)$ does not have a minimum. In our case the cost for corrective maintenance is 20 times that of preventive maintenance: $c_1 = 20 \cdot c_2$. The choice of the cost ratio equal to 20 is based on expert's

knowledge from SIT AB. In addition, this ratio seems to be commonly used in repairability analysis of mechanical components (Abernethy, 2006).

Notice that c_1 and c_2 are constants and do not vary during the lifetime. The variation is completely modeled by the distribution function $F(t)$, therefore $c_1F(t)$ is the expected cost function for corrective replacement, and $c_2(1 - F(t))$ is the expected cost function for preventive replacement. There is a trade-off between the two expected cost functions. The expected cost function for corrective replacement increases with time, because the more we wait to apply preventive maintenance, the higher is the chance that the components fail at such a level that corrective maintenance is needed. On the other hand, the expected cost function for preventive replacement decreases with time. It is high in the beginning of the life, because if the components are replaced too early, then it would be too costly, and the advantages compared to corrective maintenance would be negligible. This is the reason why we look for an optimal time. The sum of the two normalized expected costs leads to the convex cost function $c(t)$ with the minimum at the optimal time. Figure 3.2 shows an example of this optimal replacement policy.

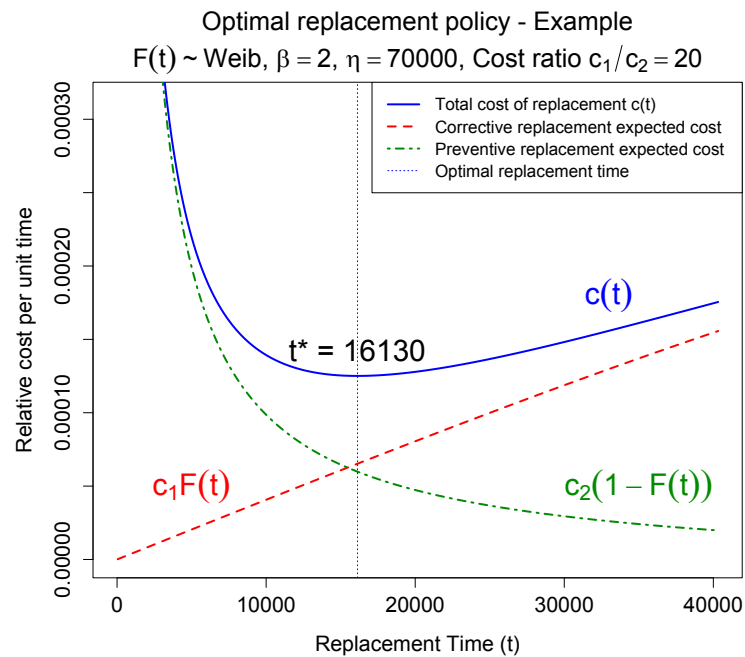


Figure 3.2.: Example of optimal replacement policy: the calculated optimal replacement time is 16130, when $F(t) \sim \text{Weib}(2, 70000)$ and the cost ratio c_1/c_2 is 20.

4. Results

We show in this chapter the most relevant results of our analysis. We first display the outcomes from the failure mode analysis performed on dataset 1. Then, we report the main findings on the multivariate survival analysis applied on dataset 2.

4.1. Failure mode analysis - Dataset 1

Dataset 1 contains information about 5 failure modes that occurred on some components before the inspection times (see Section 2.2.1). We have several reasons to assume independence between the failure modes. First, there are several physical reasons which suggest that some failure modes cannot interact between each other. In addition, we found no relevant correlation between failure modes after visualizing scatterplots as well as correlation indexes. In fact, as shown in Figure 4.1 scatterplots do not show any regular pattern in favor of correlation between failure modes.

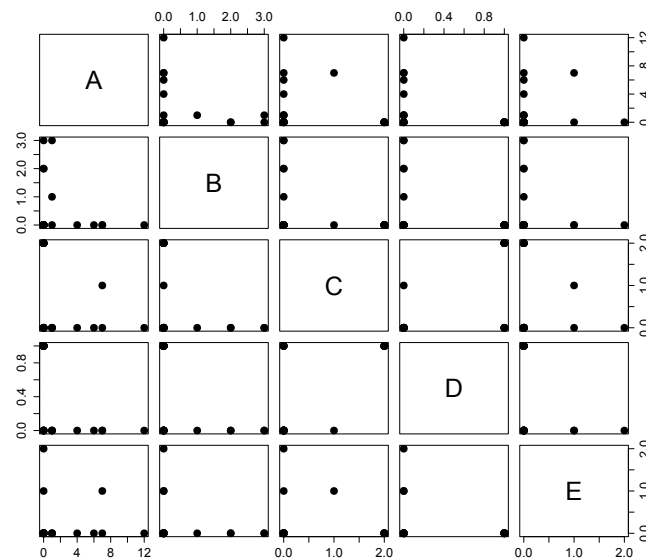


Figure 4.1.: Scatterplot pairs of the 5 failure modes considered from dataset 1. The absence of a regular pattern between observation pairs confirms that assuming independence is reasonable.

Since independence is assumed, we can carry out a single survival analysis on each failure mode and then calculate the system survival function. We apply the Turnbull non-parametric estimator as well as parametric analysis based on the Weibull distribution. Figure 4.2 shows the results on two Weibull plots from both methods respectively. In the Weibull plot the distribution function is transformed as in Equation 3.3 such that the Weibull curves are displayed as straight lines to facilitate interpretation of the results. Note that the y-axis represents probability of failure (or unreliability), and the time expressed in EOH is in the x-axis. Appendix A shows a version of the picture where Turnbull and Weibull curves are overlaid in the same plot. Results from Weibull analysis can be visualized more conveniently in the Weibull plot in Figure 4.3 where the time range in the x-axis is extended to 100000 EOH and some credible intervals are not shown for simplicity. Both Turnbull and Weibull curves reveal that the most dominant failure modes are A and B. In particular, failure mode A prevails at early times: before 27000 EOH according to the Turnbull curves, and before 42000 EOH according to Weibull analysis. This difference lies inside the related 95% confidence bands. After these times failure mode B seems to be the most relevant on causing irreparability. In contrast, failure mode E barely affect the component lifetime.

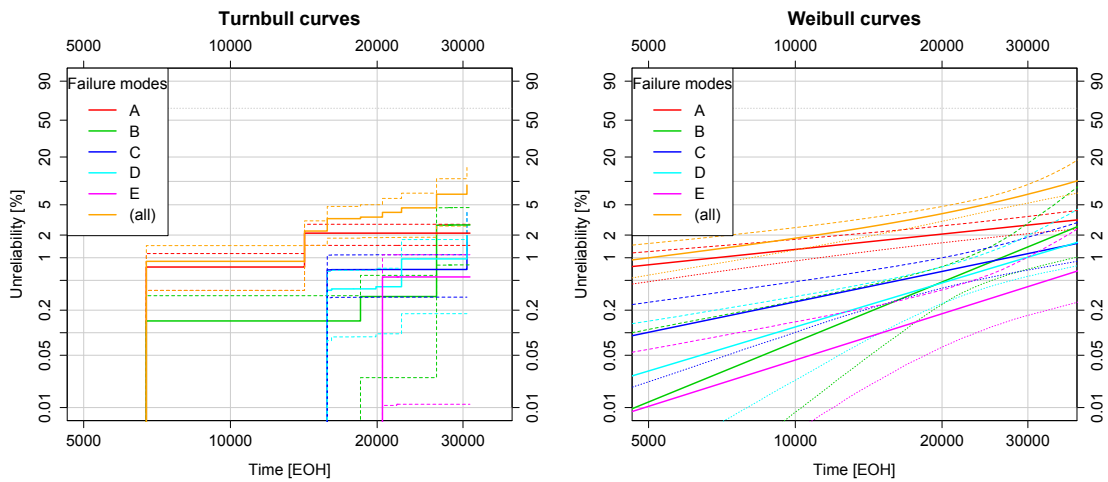


Figure 4.2.: *Left:* Non-parametric survival curves and 95% confidence intervals calculated from the Turnbull estimator. *Right:* Bayesian Weibull survival functions based on posterior median parameter estimation with related 95% credible intervals.

Table 4.1 reports the estimated parameters β and η for each failure mode. We derive posterior median estimates because the posterior distributions of some parameters revealed an asymmetric shape. In this scenario posterior means may not be reliable point estimates due to the bias introduced by extreme samples at the tails of the posterior distribution. This fact is more evident among the parameters η rather than β . We also calculated 95% credible intervals on the parameters. It can be clearly

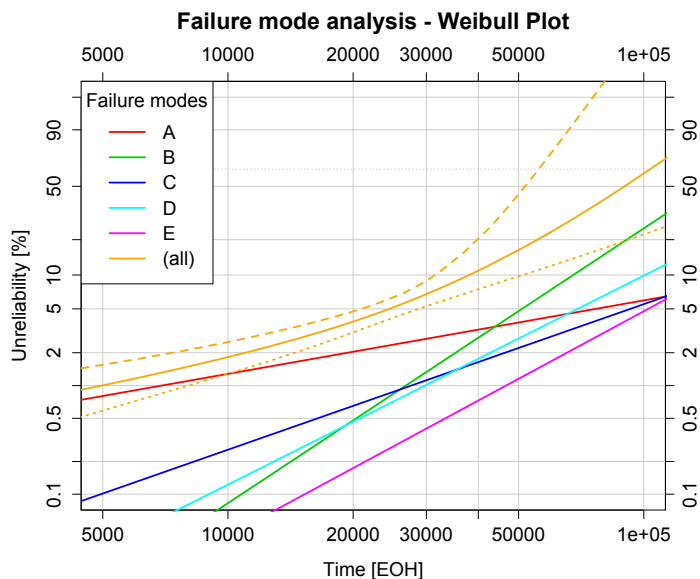


Figure 4.3.: Bayesian Weibull survival functions displayed in the Weibull probability plot. The x-axis is log-transformed and the y-axis is log-log transformed to better visualize and interpret the Weibull distribution.

seen from Table 4.1 that failure mode A has a shape parameter equal to 0.682 in a credible region between 0.451 and 1.053, thus mostly lower than 1 indicating infant failures. Consequently, parameter η tends to be high when β is small because the survival function is more flat. The survival function of failure mode B is instead steeper with $\beta = 2.663$ indicating wear-out life. Also failure modes C and D generally have parameter β greater 1, although the credibility regions cover quite wide ranges that includes 1.

| Failure mode | Parameter β : Post.med. [95% CI] | Parameter η : Post.med. [95% CI] |
|--------------|--|---------------------------------------|
| A | 0.682 [0.451 , 1.053] | 5 843 000 [803 000 , 99 298 000] |
| B | 2.663 [1.029 , 5.281] | 148 000 [58 000 , 2 978 000] |
| C | 1.351 [0.759 , 2.553] | 822 000 [149 000 , 14 422 000] |
| D | 1.954 [0.905 , 4.058] | 312 000 [79 000 , 6 524 000] |
| E | 2.055 [1.019 , 4.794] | 433 000 [81 000 , 8 497 000] |

Table 4.1.: Posterior inference of Weibull parameters from each failure mode sub-population: posterior median estimates and 95% credible intervals are reported.

The choice of the optimal replacement time is based on the minimization of the cost function described in Section 3.4. The Weibull system survival function (Equation 3.22), visualized as an orange line in Figure 4.3, is considered as the lifetime distribution function $F(t)$ to describe the variation of the risk at different times. We use $F(t)$ as an input for the cost function to derive the optimal replacement time.

We set the ratio of the corrective and preventive replacement costs c_1/c_2 equal to 20. The optimal replacement time is then equal to 32100 EOH with credible interval [26000, 73900] as shown in Figure 4.4. One could consider the minimum of the left credible interval 26000 EOH as the optimal replacement time. In fact, in risk analysis it is common to consider the worst-case scenario for investment planning. Since these models are probabilistic, the worst scenario does not exist because there is always a probability, although little, that the event does not occur. However, the left credible interval could be considered as a boundary between safe and risky decisions. The left credible interval defines a 97.5% one-sided credibility region where the optimal replacement time is greater than 26000. On the other hand, there is a 2.5% probability that the optimal replacement time is below 26000.

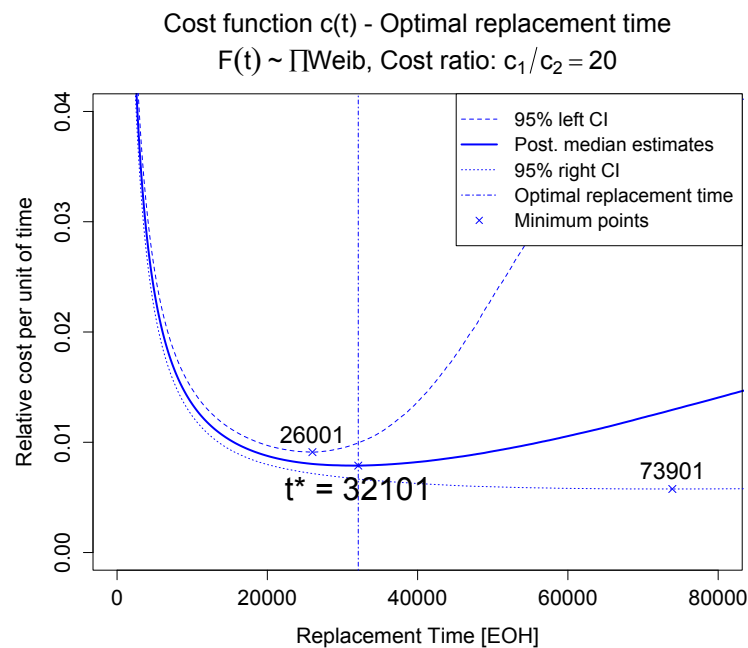


Figure 4.4.: Optimal replacement time based on the minimization of the cost function $c(t)$ (solid line), where $F(t)$ is the system survival function, c_1/c_2 is the ratio between corrective and preventive replacement cost. Left and right credible intervals are displayed in dashed and dotted lines respectively. The crosses indicate the points of minima of the cost function on the posterior median and on the 95% CI estimates.

In the analysis we consider the cost ratio $c_1/c_2 = 20$. The line plot in Figure 4.5 shows how the optimal replacement time varies with respect to the cost ratio. Reasonably, the optimal time tends to decrease as the cost ratio increases, however, the decreasing speed goes slower and slower until converging. This means that even if the cost ratio is more or much more than 20, the optimal replacement time is not dramatically affected.

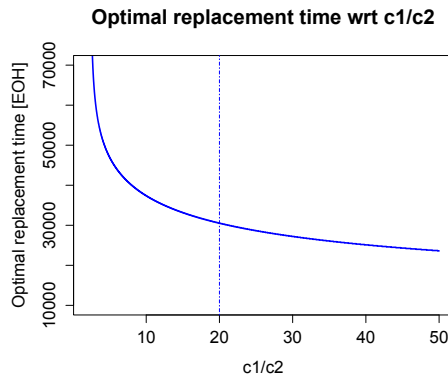


Figure 4.5.: Line plot showing how the optimal replacement time varies with respect to the cost ratio c_1/c_2 . The ratio is set to 20 in our analysis.

4.2. Multivariate survival analysis - Dataset 2

As previously reported in Section 2.2.2, dataset 2 contains attributes related to the environmental conditions where the turbines operate. We first explore survival plots showing Turnbull curves about the two categorical variables. We then apply the Bayesian Weibull AFT model and a more generalized version of the Weibull model to predict specific survival functions for each turbine, and to investigate which attributes are more harmful or beneficial for the component lifetime. The adequacy of the Weibull distribution has been confirmed by the application of tests like the Kolmogorov-Smirnov and the Anderson-Darling.

4.2.1. Categorical variables

The survival plots in Figure 4.6 display Turnbull and Weibull curves calculated on the two categorical variables Salt Exposure and Pollution Classification. It seems that components exposed to salt tend to have a slightly longer life compared to those that are not exposed to. However, dataset 2 includes only 6 observations with components exposed to salt. Therefore, these results may be biased by the small sample size. In addition, turbines operating in sites exposed to salt are usually equipped with better air filters. Thus, the salt exposure parameter might measure the combined effect of the air salinity and the improvement from air filters. Regarding the pollution classification, components located in a urban, industrial and heavy industrial environment show similar Weibull curves, a part from a deviation of the curve related to the heavy industrial environment. In fact, components in a heavy industrial environment tend to have a shorter life expectation compared to those located in a urban and industrial environment. The survival curve of the clean environment seems to differ more from the rest, revealing lower lifetime expectation. The sample size of the clean environment is quite small since it is composed only

of 3 observations where 2 of them belong to a turbine which encountered a particular high number of failures. Therefore, the shorter lifetime estimation on the clean environment may be biased by this fact. Generally, it is relevant to note that the event probabilities where the curves lie vary in a small range between 0% and 10% or 15%. This is typical of our highly censored data: we have few failures and many suspensions.

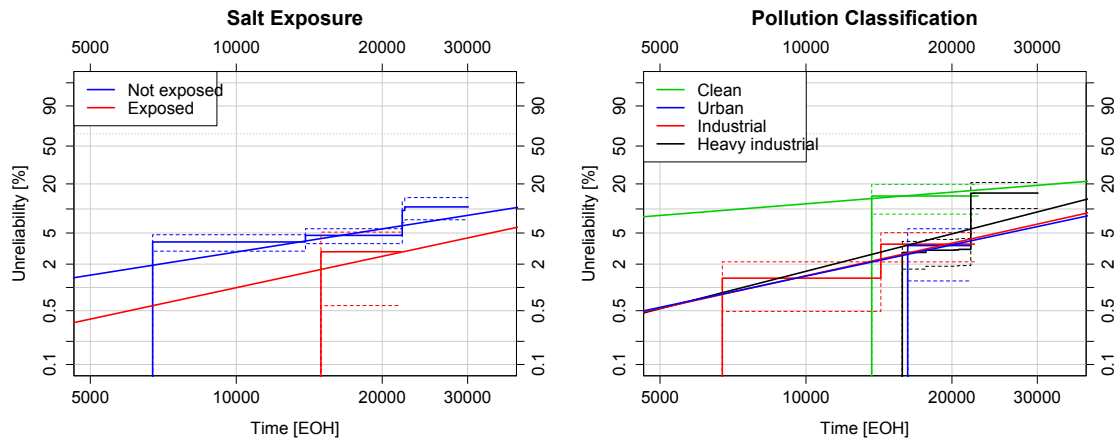


Figure 4.6.: Weibull plots showing Turnbull and Weibull curves about the two categorical attributes: Salt Exposure (on the left) and Pollution Classification (on the right).

4.2.2. Bayesian Weibull AFT model

When fitting the Bayesian Weibull AFT model we include all the attributes, both categorical and continuous, to see the global effect of each covariate. The MCMC chains converge fast to the region of the target distribution. The posterior distributions of the estimates show a symmetric bell shape, thus we select posterior means as point estimates and 95% credible intervals to measure uncertainty. Results are reported in Table 4.2. According to the AFT assumption the parameter β is fixed and for this particular case is estimated to 1.51 [1.21, 1.90], thus greater than 1 and confirming wear-out life. The coefficients \mathbf{b} from Equation 3.12 express the variation of η for each covariate. A positive coefficient means that the covariate is positively correlated with η , whereas a negative coefficient means that the covariate is negatively correlated with η . In other words, as the value of the covariate increases, a positive coefficient means longer survival, whereas a negative coefficient means shorter survival. According to Table 4.2, higher temperature and higher humidity in July decrease the component lifetime, as well as higher latitude. In contrast, more polluted environment seems to be beneficial for the lifetime as well as salt exposure. The credible intervals for the relative humidity in January, the altitude, and the

distance from sea include zero, which means that these attributes are not correlated with the lifetime.

Figure 4.7 reports a Weibull plot of the predicted survival functions from the Bayesian Weibull AFT model applied on dataset 2. The survival curves are parallel according to the AFT assumption that keeps the slope β as fixed. The more the line is on the left, the shorter is the lifetime expectation. Turbine 3 is the most exposed at risk. On the other hand, the model estimates the longest survival for turbine 2. The AIC calculated on this model is 791.87, the BIC is 812.14.

| | Post. means | 2.5% | 97.5% | |
|--------------------------------------|-------------|-------|-------|---|
| beta | 1.51 | 1.21 | 1.90 | * |
| (Intercept) | -0.44 | -1.00 | 0.21 | |
| MeanTempJanuary | -0.63 | -0.94 | -0.34 | * |
| MeanTempJuly | -0.82 | -1.32 | -0.36 | * |
| MeanRelHumidJanuary | -0.22 | -0.58 | 0.14 | |
| MeanRelHumidJuly | -0.40 | -0.70 | -0.10 | * |
| DistanceFromSea | -0.29 | -0.60 | 0.03 | |
| Altitude | 0.13 | -0.13 | 0.40 | |
| Latitude | -0.46 | -0.96 | -0.01 | * |
| Poll.Classif.: Heavy industrial | 0.87 | 0.33 | 1.46 | * |
| Pollution Classification: Industrial | 0.98 | 0.31 | 1.71 | * |
| Pollution Classification: Urban | 1.02 | 0.27 | 1.86 | * |
| Salt Exposed | 0.64 | 0.03 | 1.30 | * |

Table 4.2.: Estimated coefficients from the Bayesian Weibull AFT model: posterior means and 95% credible interval boundaries are reported. Significant coefficients are labelled with a star (*).

4.2.3. Bayesian Generalized Weibull model

The Bayesian Weibull AFT model is based on the AFT assumption where β does not vary with covariates. We now let β depend on the covariates and apply the Bayesian Generalized Weibull model from Section 3.1.2. The MCMC chains converge a bit slower compared to the previous model because of the higher complexity, however convergence is generally quite fast. The posterior distributions of the coefficients are bell-shaped, but in some cases they show asymmetry emphasized by the presence of long tails (see Appendix D). We therefore choose posterior medians as best point estimates, and 95% credible intervals. The estimated values are visualized in Table 4.3. The coefficients \mathbf{a} from Equation 3.16 express now the variation of β for each covariate, whereas the coefficients \mathbf{b} express the variation of η as before. It is relevant to note that some coefficients a_1, \dots, a_p significantly deviate from zero, indicating that the AFT assumption is violated. In particular, the AFT assumption is violated for all covariates having significant a_1, \dots, a_p coefficients like Distance

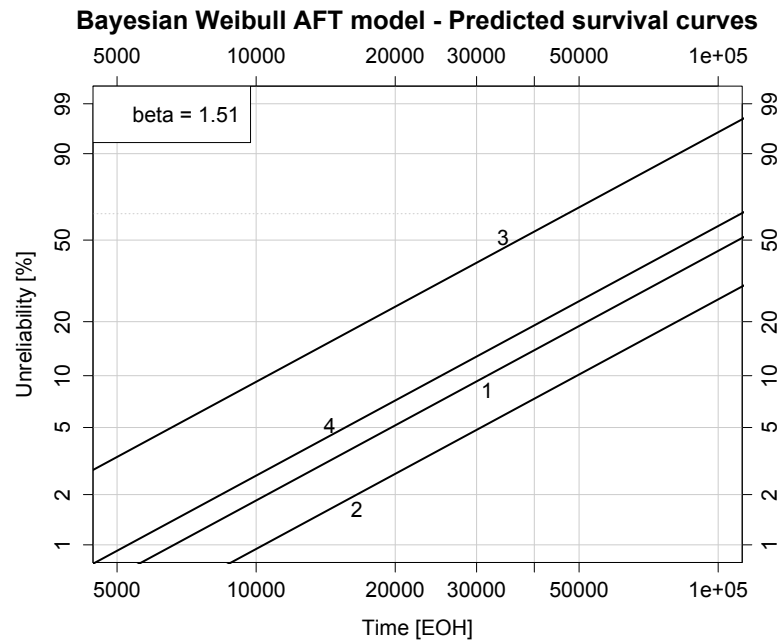


Figure 4.7.: Most relevant predicted survival functions from the Bayesian Weibull AFT model applied on dataset 2.

From Sea, Urban and Heavy industrial pollution classification. We can thus say that the AFT assumption cannot hold for some attributes, and assuming it may lead to biased results. If the AFT assumption holds, the coefficients a_1, \dots, a_p should be closed to zero indicating no variation of β . We can say that the AFT assumption holds for all the attributes except the three ones mentioned previously. When the AFT assumption is met we just need to interpret the coefficients \mathbf{b} as we have done in the previous model. However, none of the b_1, \dots, b_p coefficients seem to be significative, a part two ones that violate the AFT assumption: Distance From Sea and Pollution Classification=Heavy industrial. When the AFT assumption does not hold, the interpretation of the covariate effect is less straightforward because it depends on both coefficients \mathbf{a} and \mathbf{b} . A visual exploration of the most relevant survival curves can help to interpret the impact of the significative coefficients on the component lifetime.

Figure 4.8 shows a Weibull plot of the representative predicted survival functions from the Bayesian Generalized Weibull model applied on dataset 2. The lines are not parallel since the slope β is now free to vary. Comparing lines with different slopes is more tricky, however we can still identify a short-life pattern for the components installed in turbine 3. This survival function is quite flat and the risk of early failures is quite high. Components installed in other turbines reveal instead steeper survival functions suggesting more rapid wear-out life. In particular, turbine 4 shows quite a steep survival function. Turbines 1 and 2 reveal longer survivability because at same

| | \hat{a} | 2.5% | 97.5% | | \hat{b} | 2.5% | 97.5% | |
|---------------------------------|-----------|-------|-------|---|-----------|-------|-------|---|
| (Intercept) | -0.91 | -2.11 | 0.15 | | 1.94 | 0.06 | 3.94 | * |
| MeanTempJanuary | 0.06 | -0.40 | 0.58 | | -0.39 | -1.18 | 0.44 | |
| MeanTempJuly | 0.26 | -0.41 | 1.03 | | -0.64 | -1.60 | 0.32 | |
| MeanRelHumidJanuary | 0.16 | -0.55 | 1.00 | | -0.29 | -1.18 | 0.60 | |
| MeanRelHumidJuly | 0.29 | -0.27 | 0.88 | | -0.55 | -1.19 | 0.33 | |
| DistanceFromSea | 0.57 | 0.01 | 1.42 | * | -0.78 | -1.31 | -0.24 | * |
| Altitude | -0.11 | -0.53 | 0.25 | | 0.09 | -0.21 | 0.71 | |
| Latitude | 0.13 | -0.84 | 0.76 | | -0.65 | -1.69 | 0.96 | |
| Poll.Classif.: Heavy industrial | 1.85 | 0.78 | 2.95 | * | -2.13 | -3.97 | -0.23 | * |
| Pollution Classif.: Industrial | 1.24 | -0.23 | 2.91 | | -1.02 | -3.24 | 0.90 | |
| Pollution Classif.: Urban | 2.35 | 0.07 | 3.50 | * | -2.84 | -4.73 | 1.88 | |
| Salt Exposed | 0.04 | -0.74 | 1.70 | | 0.52 | -2.40 | 1.99 | |

Table 4.3.: Estimated coefficients from the Bayesian Generalized Weibull model: Posterior median estimates and 95% credible interval boundaries are reported. Significant coefficients are labelled with a star (*).

times the unreliability is usually lower compared to the other turbines. Turbine 4 is situated in a heavy industrial environment far from the sea. Turbine 1 is also located in a heavy industrial environment, but at a shorter distance from the sea. Turbine 2 is instead situated in an industrial environment close to the sea, and this may explain the reason why these components has a longer lifetime. Generally, according to the results from Table 4.3 and to the Weibull plot in Figure 4.8, it seems that components from turbines located in heavy industrial environment at a higher distance from the sea tend to have a shorter lifetime. Turbine 3 is located in a clean environment not so close to the sea. The flat survival function suggests that these components experienced infant mortality problems. We cannot derive optimal maintenance times on components with decreasing failure rate.

The AIC calculated on this model is equal to 774.49, the BIC is 839.02. The AIC and BIC measures of the AFT and Generalized models are reported in Table 4.4 to be easily compared. Results show that the AIC criterion selects the Generalized model as the best, whereas the BIC measure prefers the AFT model.

| Model | AIC | BIC |
|-------------|---------------|---------------|
| AFT | 791.87 | 812.14 |
| Generalized | 774.49 | 839.02 |

Table 4.4.: AIC and BIC criteria evaluated on the Bayesian Weibull AFT model and Bayesian Weibull Generalized model. Lower values of AIC and BIC select the best model, and are reported in bold.

Each of the estimated survival functions are then considered as input $F(t)$ for the cost function $c(t)$, because the main aim is to find the optimal time to replace the

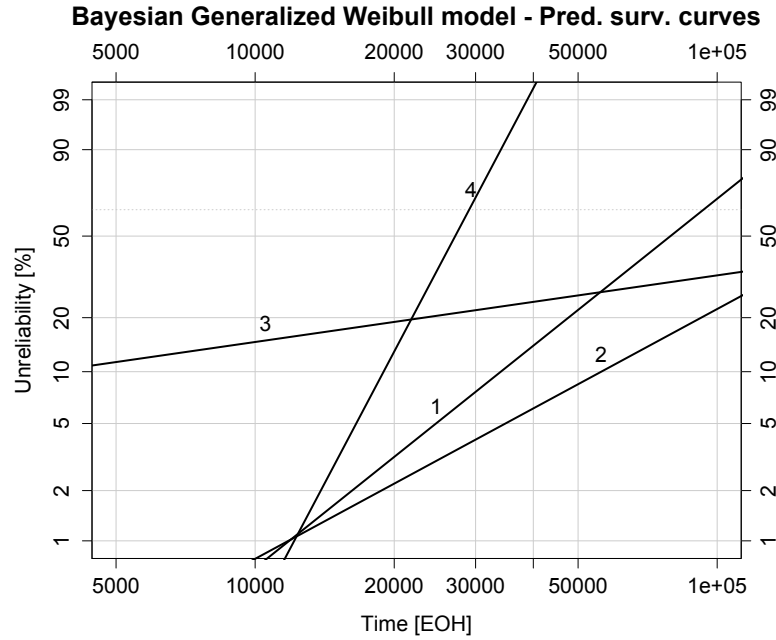


Figure 4.8.: Representative predicted survival functions from the Bayesian Generalized Weibull model applied to four different turbines in dataset 2.

components. Since we estimated a specific survival function for each observation, then each turbine will have its own replacement time. Figure 4.9 shows two examples of optimal replacement time calculated in turbines 1 (left) and 4 (right). In this case, the model suggests to apply preventive maintenance at 22786 [13641, 65653] and 12629 [6598, 29696] EOH respectively to minimize the costs. Note that when β is high the cost function is more convex and so the minimum is better defined. But the more β is closed to 1, the more the curve is flat until no minimum can be derived when $\beta \leq 1$. This is reasonable because preventive maintenance applies only with increasing hazard functions. In this analysis we apply the cost function to calculate optimal replacement times only when $\beta > 1$ to assure reasonable outcomes. This choice excludes turbine 3 from the analysis, because its β is too low to derive an optimal time.

The optimal replacement times calculated from dataset 2 mainly vary from 10000 to 35000 EOH with some cases around 40000 and 55000 EOH, as shown in the histogram in Figure 4.10 on the left. The corresponding event-probability (unreliability) is then calculated at each optimal time to have an idea of the level of risk taken at that time. Figure 4.10 on the right displays how the risk is distributed: most of the probability lies between 1% and 5% with some cases at 8% and 10%.

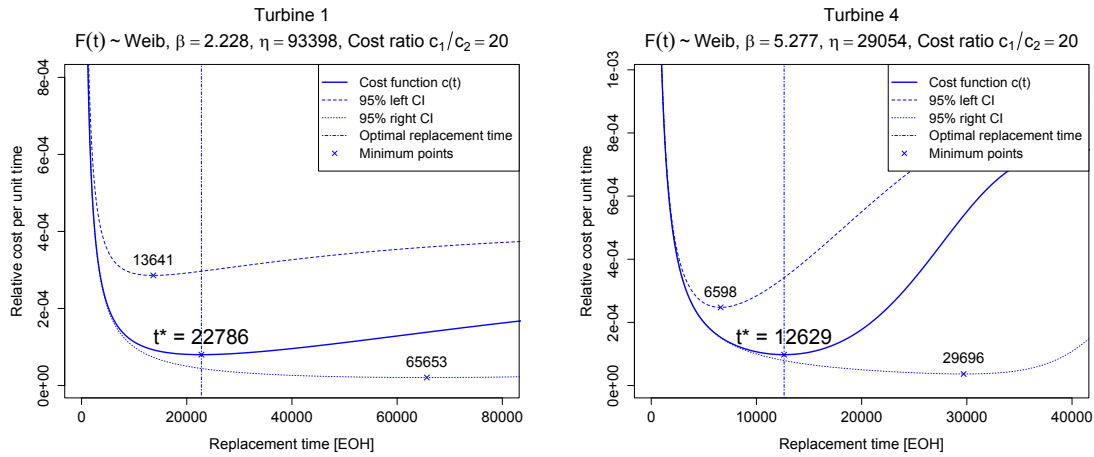


Figure 4.9.: Two examples of optimal replacement times estimated on turbines 1 (on the left) and 4 (on the right).

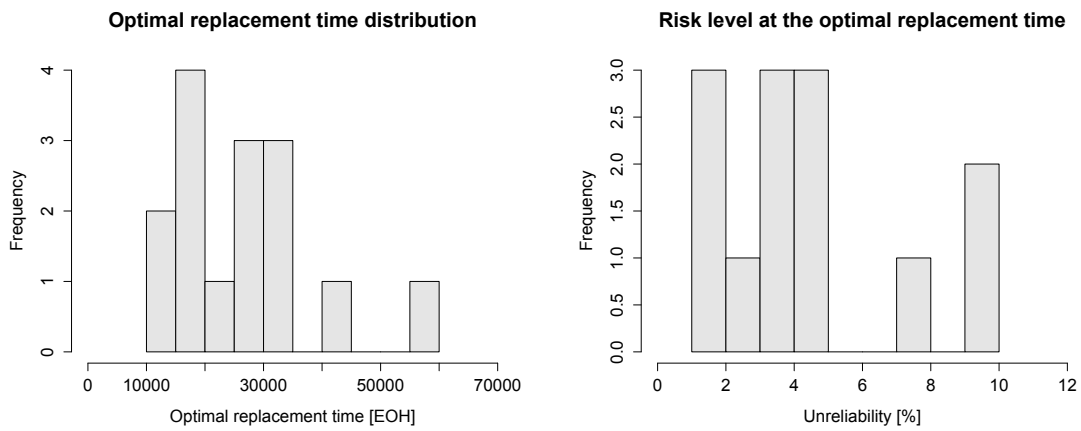


Figure 4.10.: *Left:* Histogram showing the frequency distribution of the estimated optimal replacement times from dataset 2. Optimal times are mainly distributed between 10000 and 35000 EOH with some cases around 40000 and 55000 EOH. *Right:* Histogram of the corresponding risk values evaluated at the optimal times. The most common risk values are between 1% and 5% with some cases at 8% and 10%.

5. Discussion

This thesis applies survival analysis on mechanical components installed in gas turbines developed by Siemens Industrial Turbomachinery AB. The aim is both to estimate survival functions that depend on the properties of the gas turbines, and to use this information to obtain optimal time points for preventive maintenance. Another aim is the investigation of the effect of different failure modes on the component lifetime.

The failure mode analysis in Section 4.1 revealed that the 5 failure modes have a different impact on the component lifetime. Before 42000 EOH failure mode A has the highest risk to occur followed by failure mode C. After that time failure mode B has the highest risk to occur. The analysis clearly suggests that acting on failure mode A is the most effective choice to decrease the risk of early failures and extend the component lifetime. Failure mode C, D, and E play a secondary role on the lifetime since they generally have a lower probability to occur compared to failure mode A and B. The effect of failure mode B is more evident at late times, however preventive maintenance is usually applied before these times. In other words, the components usually do not have enough time to experience the full effect of failure mode B, because they are replaced earlier.

The optimal replacement time calculated on dataset 1 is 32100 [26000, 73900]. Notice that the credible interval is quite large because the system survival function is not steep enough to clearly define a minimum in the cost function. This is probably due to the mixture of infant and wear-out failures that contributes in different ways to add uncertainty in the model. Then, the highly censored nature of the data can be a relevant source of bias for the model. It is also important to note that this optimal time may be biased by the fact that effects from specific covariates are not taken into account. We can consider this optimal time as a general indication for optimal replacement time based on the effect of multiple failure modes. For observation-specific optimal times we apply the multivariate methods described in Section 3.1.1 and Section 3.1.2. In this way, we can investigate how optimal replacement times vary in different turbines located at different sites, and thus operating at different conditions.

Section 4.2 shows results from the multivariate survival analysis carried out on dataset 2 where covariate effects are included. The analysis applied on the two categorical attributes Salt Exposure and Pollution Classification (Figure 4.6) reveals shorter lifetime for the components not exposed to salt, and for components located on clean or heavy industrial environment. According to our previous knowledge we

would expect decreasing life expectancy for the components exposed to salt, but this is not confirmed by the model. We would also expect a more beneficial effect of the clean environment compared to the polluted ones. We think that these results may be biased by low sample size of the groups `SaltExposure=1` and `PollutionClassification=Clean`. These small groups may be more exposed to the bias from outliers or from the effect of other hidden attributes not considered in the model like air filter type. In fact, turbines running in sites exposed to salt are usually equipped with better air filters to prevent corrosion derived from salinity in the air. Thus, the salt exposure parameter might be biased by the effect of different air filter types.

In the parametric models we include the effect of a larger set of covariates which describe the environmental conditions where the turbines operate. Both the AFT and Generalized Weibull models reveal that the altitude and the average relative humidity in January show no correlation with the lifetime. The non-significativeness of the altitude might be explained by the fact that the observed data related to the altitude lie in a short range between 0 and 378 meters. Probably, the lifetime is not affected in this short range of altitude.

In terms of model comparison we suggest to apply parametric methods if the assumption of a distribution is reasonable. In our work we assumed the Weibull distribution after checking its adequacy through tests like the Kolmogorov-Smirnov and the Anderson-Darling. Non-parametric estimates are useful if we are not sure of the distributional assumption, alternatively they can be used to double check the results obtained from the parametric analysis. Generally, parametric analysis is more preferred for its higher statistical power. The Bayesian Weibull AFT model has a lower BIC value since it includes half the parameters of the Generalized model. However, the AFT assumption seems to be violated for some coefficients, potentially leading to biased outcomes. In addition, the BIC criterion can be biased when the number of observations is not much larger than the number of covariates, as explained in Section 3.1. We would prefer the AIC criterion for model selection in our data. According to this measure, the Bayesian Weibull Generalized model has lower AIC. The Bayesian Weibull Generalized model is a flexible solution to model lifetime data where both Weibull parameters are free to vary with covariates. This makes it possible to predict observation-specific survival functions with own shape and scale parameters. The Bayesian philosophy behind the model allows to infer prior belief based on previous knowledge. We use quite diffuse priors with large variance to make them less informative. However, one might use more informative priors to exploit previous experience from the parameter β , for example. In our analysis we do not include a so large number of attributes to avoid overfitting issues given the small sample size. However, inclusion of additional parameters is possible if it is subjected by some kind of variable selection technique.

The cost function allows to estimate the optimal time when applying preventive maintenance. The accuracy of the cost function depends on the adequacy of the survival distribution and on the estimated costs. In our analysis the calculated optimal times can have a large credible interval, that is direct consequence of the

large credible interval from the estimated distribution $F(t)$. The cost ratio c_1/c_2 is constant over the time. One may consider it as a variable $c_1(t)/c_2(t)$, to build a more advanced cost function which includes for example variations in the electricity price. However, we think that this simpler approach is a good approximation to derive the optimal cost.

The models used in this analysis allow to predict survival distributions and optimal replacement times not only for components owned by existing customers, but also for components installed at new sites. For example, we may derive a specific replacement time for the components installed in a gas turbine located at a new customer from e.g. Mexico. Knowing the environmental features of the new site, it is then easy to derive that specific optimal replacement time.

Nevertheless, it is important to note that the model output could be potentially affected by various kinds of bias. First, the model is trained in a quite small sample size that may not be representative for all the world sites. For example, sites with extreme cold or warm temperatures or high altitude may be wrongly estimated. Generally, models cannot guarantee accurate predictions for new observations that lie outside the range of the trained data.

Second, failures could be driven by other environmental factors not considered in the model, for example presence of sand in the air. Some of the environmental attributes we considered are quite raw approximations: for example, the temperature is a raw mean between two values and it is calculated only in the months of January and July. It would be reasonable here to consider not only the mean temperature but also the variation. Implementing the temperature as a time-varying covariate could be a possible way to make the analysis more accurate. We mainly focused on environmental factors, but the component lifetime could be affected also by other reasons like customer behaviors. The customer might use the gas turbine in a wrong configuration, thus increasing the chance of damages. Attributes related to the turbine configuration like air filter type and lube oil type may have an impact on the component lifetime.

Finally, the left-censored nature of our lifetime data can be considered as a source of uncertainty for our results. We only know that the failure has happened somewhere before the inspection time. Before means between the installation time and the inspection time, and this time interval can be quite large in some cases. As shown in Figure 2.1 our analysis is based on quite large and overlapping intervals, and this fact significantly increases the uncertainty on the fitted distributions. Information from intermediate inspections would help to narrow the intervals and thus to reduce the bias introduced by the uncertainty. Simulation studies done during this analysis revealed that narrower intervals lead to more precise estimates with smaller credible intervals. SIT AB regularly checks the components with intermediate inspections, but these inspection procedures simply consist on visual examinations through boroscope tubes. This method is not able to identify all the failures, and it potentially leads to false negatives. However, some useful information could be extracted by

these intermediate inspections. The suggestion here is to improve the efficacy of the intermediate inspections, or alternatively try to gather some additional information from the repair inspections. For example, it would be good to derive whether the component has failed in the first or in the last part of its life.

6. Conclusions

Survival analysis based on field experience data is applied on mechanical components installed in gas turbines manufactured by SIT AB. We investigated the impact of multiple failure modes on the component lifetime, and we found that failure mode A is the most harmful for the component. The findings revealed that this failure mode is characterized by infant failure problems, probably due to inadequate burn-in tests, quality control problems, or bad raw materials from the suppliers. We then examined if different environmental features can have a significant impact on the component lifetime. We considered the response variable as Weibull distributed, and we applied the Bayesian Weibull AFT model and a more generalized version that we called Bayesian Generalized Weibull model. The AFT model is based on assumptions that are in some cases violated for our specific data. On the other hand, the flexibility of the generalized model allows to avoid strict assumptions and to estimate specific survival functions with own shape and scale parameters. Therefore, we considered results from the latter model as the most representative for our analysis. According to these findings, it seems that components installed in gas turbines located in a heavy industrial environment and far from the sea have a shorter lifetime. Based on the estimated survival functions, we calculated optimal times for preventive maintenance with respect to the minimization of a cost function. The model is capable to predict customer-specific optimal replacement times, not only for existing customers but also for new incoming customers. The credible intervals on the optimal times can be quite large and this can be consequence of low sample size and censored nature of the input data.

For future work, it is encouraged to add more observations in the analysis. Results will be more reliable when based on bigger sample size. Additional information could then be extracted from intermediate inspections in order to reduce the uncertainty from censoring. Then, additional environmental attributes could be considered in the analysis, for example the presence of sand in the air may affect the component reliability. Besides environmental features, the impact of attributes related to the turbine configuration could be also investigated, for example air filter type and lube oil type. The effect of interaction terms can then be taken into account for future analyses. These additions should be followed by some kind of variable selection technique in order to select the most relevant attributes and avoid problems of overfitting.

A. Failure mode analysis

Turnbull and Weibull curves are visualized in Figure A.1. Generally, the Turnbull curves lie inside the confidence bands of Weibull curves. Some deviations are due to small sample size or to uncertainty derived by left-censored data.

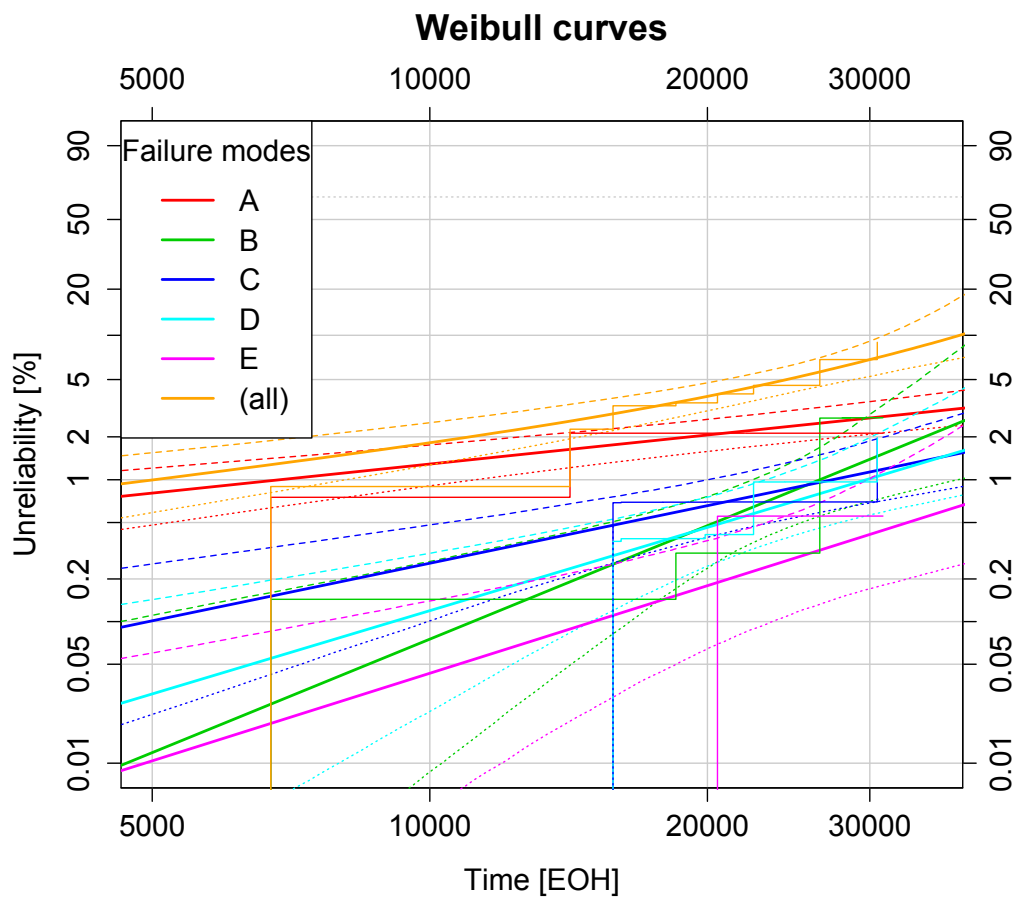


Figure A.1.: Weibull curves with 95% credible intervals. Turnbull curves are also overlaid to facilitate comparison to the Weibull distribution.

B. Density and MCMC diagnostics plots for the Bayesian Weibull AFT Model

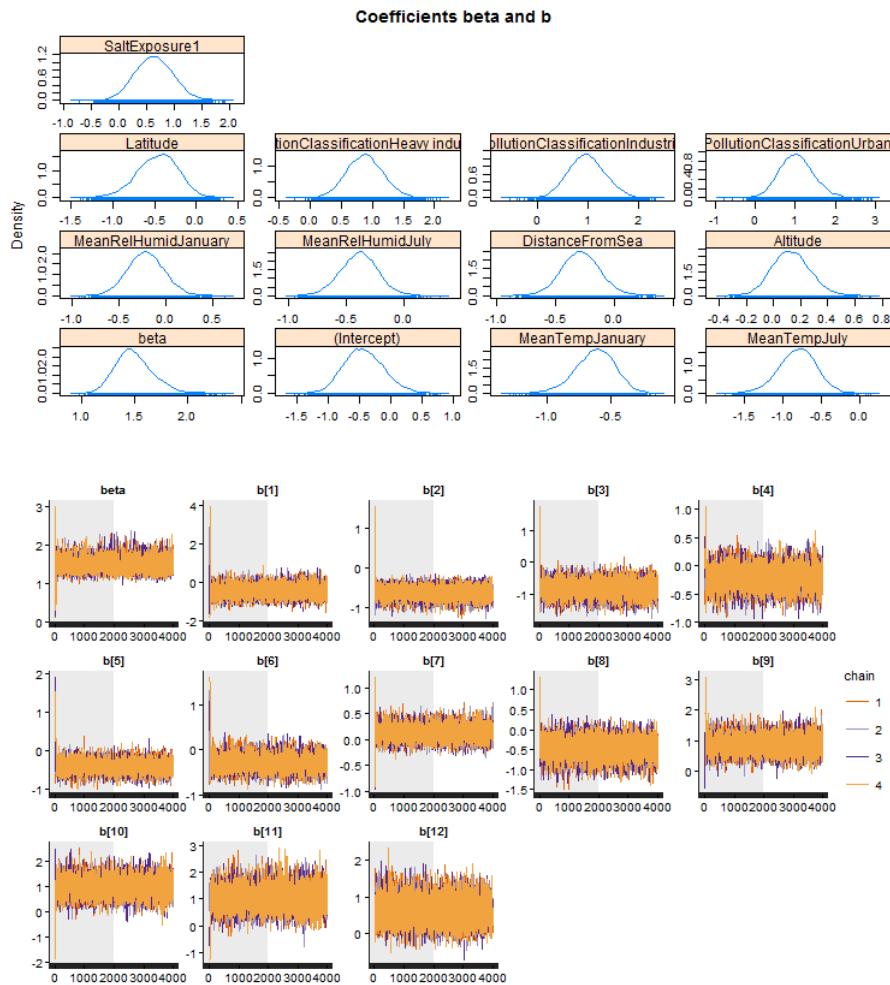


Figure B.1.: *Top:* Posterior distributions of coefficients from the Bayesian Weibull AFT model. All distributions show bell and symmetric shape. *Bottom:* Traceplots of sampled coefficients from the Bayesian Weibull AFT model. MCMC converge quickly to the target distribution.

C. Stan code for the Bayesian Generalized Weibull Model

```
bwmgrp='
data{
  int<lower=0> N; # Number of sets
  int<lower=0> f[N]; # Number of scrapped components in a set
  int<lower=0> s[N]; # Number of not scrapped components in a set
  real<lower=0> t[N]; # Time response
  int<lower=2> p; # Number of covariates (intercept included)
  row_vector[p] X[N]; # Matrix [N,p] of covariates (intercept included)
  vector[p] a0; # Mean vector g-prior for a
  vector[p] b0; # Mean vector g-prior for b
  cov_matrix[p] V1; # Cov. matrix g-prior for a
  cov_matrix[p] V2; # Cov. matrix g-prior for b
}
parameters {
  vector[p] a; # parameter vector a
  vector[p] b; # parameter vector b
}
model{
  a ~ multi_normal(a0, V1); # g-prior for a
  b ~ multi_normal(b0, V2); # g-prior for b
  for (i in 1:N){ # log-likelihood (for right and left censored data)
    increment_log_prob(s[i]*weibull_ccdf_log(t[i], exp(X[i]*a),
      exp(X[i]*b)));
    increment_log_prob(f[i]*log(weibull_cdf(t[i], exp(X[i]*a),
      exp(X[i]*b))));
  }
}
generated quantities{
  vector[N] beta; # shape parameter for each observation
  vector[N] eta; # scale parameter for each observation
  for (i in 1:N){
    beta[i] <- exp(X[i]*a);
    eta[i] <- exp(X[i]*b);
  }
}
,
```

D. Density and MCMC diagnostics plots for the Bayesian Generalized Weibull Model

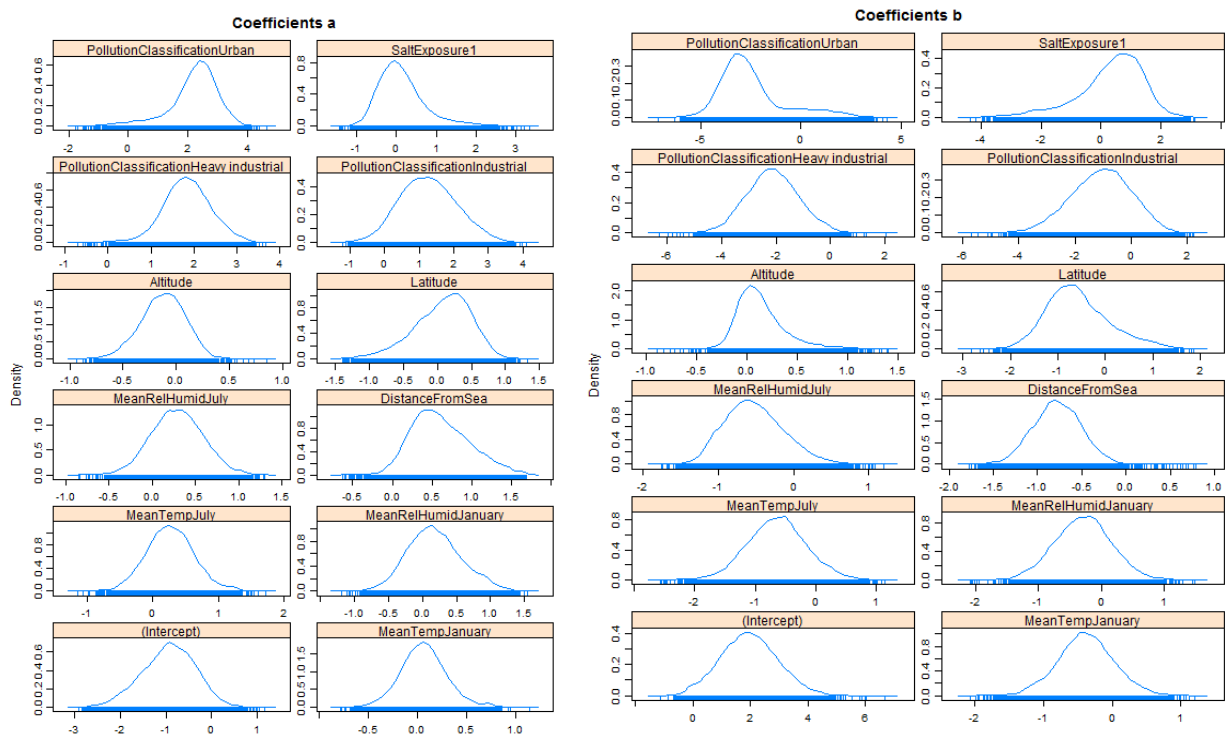


Figure D.1.: Posterior distributions of coefficients from the Bayesian Generalized Weibull model. Some distributions reveal asymmetric shape.

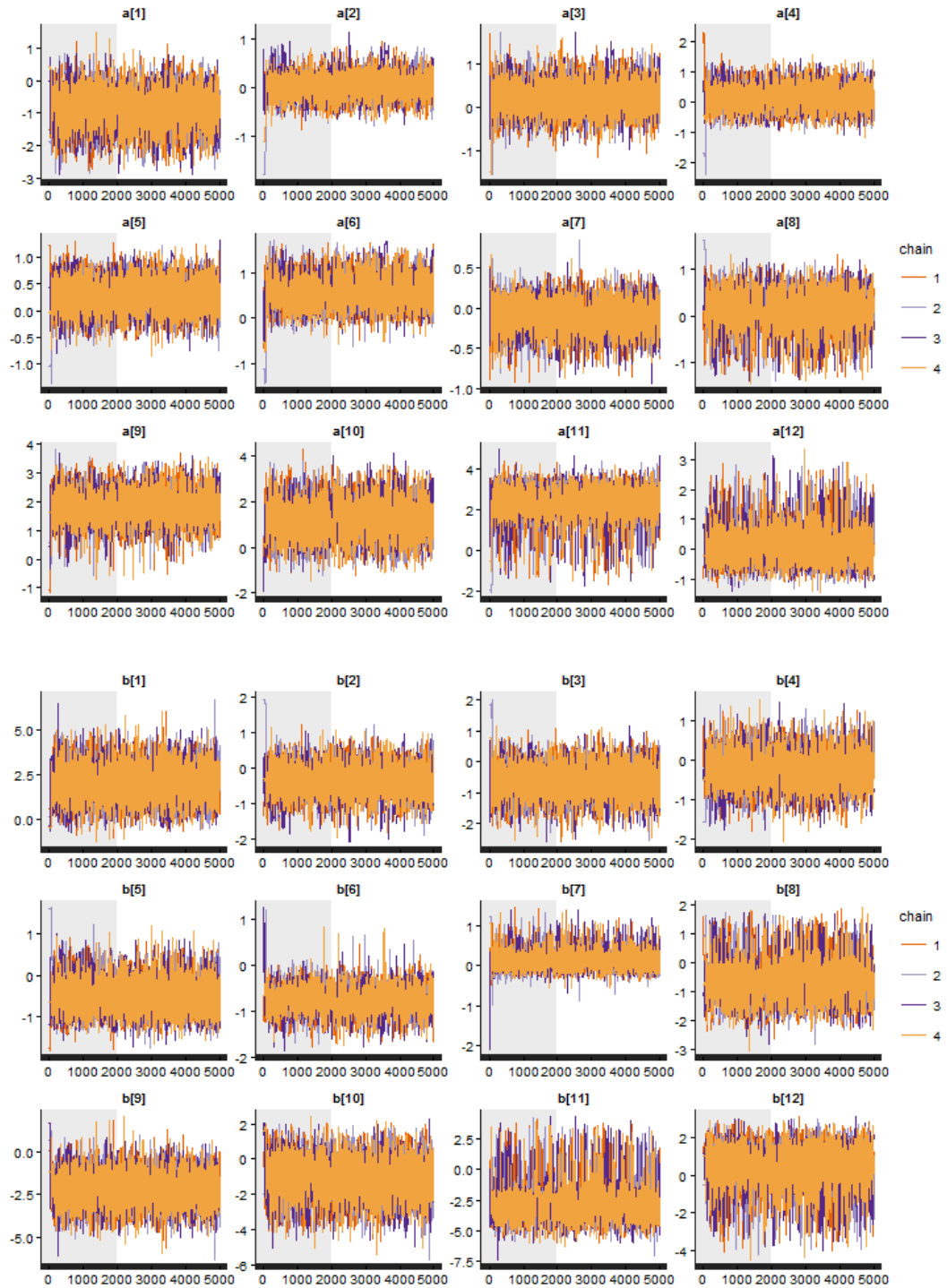


Figure D.2.: Traceplots of sampled coefficients from the Bayesian Generalized Weibull model. MCMC convergence is fast. Asymmetry is evident in $a_8, a_{11}, a_{12}, b_{11}, b_{12}$.

Bibliography

- Abernethy, R. B. (2006). *The New Weibull Handbook - Reliability and Statistical Analysis for Predicting Life, Safety, Supportability, Risk, Cost and Warranty Claims*. Abernethy, Robert B., 5th edition.
- Barabadi, A. (2013). Reliability model selection and validation using Weibull probability plot - A case study. *Electric Power Systems Research*, 101:96–101.
- Crowder, M. (2012). *Multivariate Survival Analysis and Competing Risks*. Chapman and Hall/CRC.
- Genschel, U. and Meeker, W. Q. (2010). A Comparison of Maximum Likelihood and Median-Rank Regression for Weibull Estimation. *Quality Engineering*, 22(4):236–255.
- Giolo, S. R. (2004). Turnbull's nonparametric estimator for interval-censored data'. Technical report, Department of Statistics, Federal University of Paraná, Paraná, Brazil.
- Giraud, C. (2014). *Introduction to high-dimensional statistics*, volume 138. CRC Press.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282):457–481.
- Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association*, 91(435):1343–1370.
- Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(2):489–502.
- Kleinbaum, D. G. and Klein, M. (2005). *Survival Analysis - A Self-Learning Text*. Springer, 2nd edition.
- Law, C. G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*, 11(12):1569–1578.
- Li, Y.-F., Zhu, S.-P., Li, J., Peng, W., and Huang, H.-Z. (2015). Uncertainty Analysis in Fatigue Life Prediction of Gas Turbine Blades Using Bayesian Inference. *International Journal of Turbo & Jet-Engines*, 32(4):319–324.
- McCool, J. I. (2012). *Using the Weibull Distribution: Reliability, Modeling and Inference*. Wiley.

- Mohammed Ahmed, A. O. (2014). Comparison of the Bayesian Methods on Interval-Censored Data for Weibull Distribution. *Open Journal of Statistics*, 4:570–577.
- Nánási, T. (2014). Interval Censored Data Analysis with Weibull and Exponential Distribution. *Applied Mechanics and Materials*, 693:74–79.
- Ng, M. P. (2002). A modification of Peto’s nonparametric estimation of survival curves for interval-censored data. *Biometrics*, 58(June):439–442.
- Peto, R. (1973). Experimental Survival Curves for Interval-Censored Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 22(1):86–91.
- Rinne, H. (2009). *The Weibull Distribution: A Handbook*. Chapman and Hall/CRC.
- Singh, R. S. and Totawattage, D. P. (2013). The Statistical Analysis of Interval-Censored Failure Time Data with Applications. *Open Journal of Statistics*, 3:155–166.
- Stan Development Team (2014). RStan: the R interface to Stan, Version 2.5.0.
- Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737.
- Turnbull, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):290–295.
- Weibull, W. (2013). *Fatigue Testing and Analysis of Results*. Elsevier Science.
- Ye, Z. and Ng, H. K. T. (2014). On analysis of incomplete field failure data. *The Annals of Applied Statistics*, 8(3):1713–1727.
- Zaretsky, E. V., Litt, J. S., and Hendricks, R. C. (2013). Determination of Turbine Blade Life From Engine Field Data. Technical report, National Aeronautics and Space Administration - John H. Glenn Research Center at Lewis Field, Cleveland, Ohio 44135-3191.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier, New York, NY.
- Zuehlke, T. W. (2013). Estimation and testing of nonproportional Weibull hazard models. *Applied Economics*, 45(15):2059–2066.

LIU-IDA/STAT-A-16/007-SE