School of Electrical Engineering, Computing and
Mathematical Sciences

Centre for Transforming Maintenance Through Data Science

Bayesian Hierarchical Modelling of Equipment Reliability in Mining: A Pragmatic Approach

Ryan K. Leadbetter

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

November 2023

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Ryan K. Leadbetter

"*The Quote*"

— AUTHOR

Acknowledgements

Write your acknowledgements here

Abstract

From pit to port, the consistent and efficient operation of iron ore machinery is essential for maximizing profits. To this end, reliability modelling is an invaluable tool for improving the design and execution of the maintenance strategies that ensure the reliable operation of mining machinery. There are well established reliability models in the literature, but there is a discrepancy between this literature and what is actually done by practitioners in the mining industry; a theory-practice gap. This gap exists because of the imperfect reality of collecting data in the field–data sets that are small, incomplete, noisy, or all three–and the lack of methods for expanding reliability modelling to account for these imperfections. My industry-linked PhD has aimed to reduce this gap by demonstrating how Bayesian statistical modelling framework can address some of the common problems faced when fitting models to such reliability data in mining applications.

In the first part of the work, I adapt and evaluate a method for constructing an informative joint prior distribution for the parameters of weibull lifetime analysis to combat bias introduced through heavily censored lifetime data. I first illustrate the bias caused by heavy censoring and then show how encoding domain information into a joint prior for the two Weibull parameters constrains the bias. Secondly, I evaluate the proposed method through a simulation study. Finally, I provide recommendations on applying the method in practice and demonstrate on an industry data set from an overland iron ore conveyor.

In the second part of the work I focus on degradation modelling. Particularly, how the Bayesian hierarchical framework can extend the gamma stochastic

degradation process to noisy observations and then to the degradation of surfaces. In doing so, I simplify some of the literature on noisy gamma processes by demonstrating how separating the observation-degradation process into two separate conditional models removes the need for complicated inferential algorithms. Furthermore, I show the hierarchical models implementation using flexible tools that are accessible to a wider reliability audience. I also show how reparametrisation can make the gamma process more interpretable and therefore simplify prior specification and further expansions of the model. Taking this one step further, I expand the noisy gamma process to functional data analysis in order to model the degrading surface of conveyor belting.

Throughout the work I emphasise how complicated reliability processes found in practice can be broken down into manageable sub-models and how these models can be fit, evaluated, expanded, and compared using Bayesian workflow considered to be good statistical practice. In doing so hope to contribute at a larger level by providing an applied case study of the the Bayesian workflow in a reliability setting that can be used by other applied reliability practitioners to develop solutions of their own for new problems.

Contents

List of Figures

Chapter 1

Introduction

The work presented in this thesis is part of an industry linked PhD under the Center for Transforming Maintenance Through Data Science (CTMTDS), a center comprised of both academic and industry partners. One of the centre's goals is to develop methods that support reliability engineers in managing uncertainty during the maintenance decision making process–i.e. how and when they should maintain an asset based on their understanding of the asset and the available data. As part of the industry linked PhD, I spent a reasonable period working on industry placement projects (900 hours in total between two different industry partners) to outlining research topics that are not only novel in an academic sense but also facilitate a greater use of robust statistical modelling by reliability practitioners when making maintenance decisions in the mining and mineral processing industry.

It was apparent from my placement time that there is a disconnect between the reliability modelling literature and the methods used by reliability practitioners in the mining and mineral processing industry, often referred to as a theory-practice gap. There are many well-established models for reliability data such as the Weibull distribution for lifetime data, or stochastic process models for degradation data. There is also a desire by mine, processing plant, and refinery operators to use these models since once an asset is put into service, completely

new information becomes available, which can be utilised so that operators can make better maintenance decisions based on the specific reliability characteristics of their assets, rather than estimates from the manufacturer (Jardine & Tsang, 2013). But this new information is collected in the field (where there are many sources of noise) and within large companies (where it is difficult to regulate data collection practice and where there is a constant trade-off between cost, safety, and time versus the quality and amount of data collected). This results in unique observational processes that need their own sophisticated modelling approaches before reliability practitioners in the mining and mineral processing industry can take advantage of the well-established reliability models in the literature. Two examples of issues confronting reliability engineers that I tackle in this thesis are 1) obtaining reasonable estimates of lifetime distributions when lifetime data are heavily censored due to the pre-emptive replacement of assets or limitations of data recording and 2) forecasting complex degradation processes with noisy and sparsely observed condition monitoring data. In this thesis, I show novel expansions of some of the well-established reliability methods through the Bayesian model building approach (Gelman, Vehtari, et al., 2020) and demonstrate how they can be applied to observational industry data sets (from overland iron ore conveyors) provided by the centre's industry partners.

The Bayesian paradigm became a strong backbone of this thesis because Bayesian methods provide a formal structure to build complicated models and incorporate multiple sources of information, such as domain expert knowledge (Meeker, Escobar, & Pascual, 2022). Furthermore, the resulting full posterior distribution obtained through Bayesian analysis allows us to easily produce estimates and uncertainty intervals for complicated functions of the model parameters (Meeker et al., 2022), which is extremely useful for propagating uncertainty through a decision-making process. While there is a well-developed subfield of Bayesian analysis in the reliability literature (Hamada, Wilson, Reese, & Martz, 2008; Meeker et al., 2022), the Bayesian framework is underutilised in industry.

This underutilisation is most likely because, for most cases, inference must be obtained through Monte Carlo simulation, and in the past, this has meant constructing Markov Chain Monte Carlo (MCMC) algorithms by hand. However, the recent increase in popularity of Bayesian methods has led to the development of flexible and accessible probabilistic programming languages such as BUGS (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013), JAGS (Plummer, 2003) and Stan (Stan Development Team, 2022), which in many cases alleviate the analyst form the need to construct bespoke MCMC algorithms. The result is a newfound ability to fit and explore complex models relatively quickly and simply.

To harness these new aspects of statistical modelling more effectively, the applied Bayesian statistical community has started to develop a more rigorous workflow for building, fitting, checking, and comparing Bayesian models. Throughout this thesis, I clearly emphasise the components of this workflow and demonstrate them in a reliability setting. In doing so, I hope this thesis may also be used as a template for other maintenance decision-making problems in the field.

The rest of this chapter provides a general background for the rest of the thesis. First, in *section* 1.1, I provide some context around maintenance decision-making in the mining and mineral processing industry. Then, in *section* 1.2, I give a high-level overview of reliability modelling and how it informs maintenance decisions. *Section* 1.4 outlines Bayesian methods and the key components of the Bayesian model building workflow, which will be a strong thematic thread throughout the remainder of the thesis. Finally, in *section* 1.5, I lay out the structure of the thesis.

1.1 Maintenance decision making

The maintenance of an asset can be considered as *"all activities aimed at keeping an {asset} in, or restoring it to, the physical state considered necessary for the fulfilment of its production function"* (Geraerds, 1985). In other words, the main objective of maintenance actions is to fix/replace an asset's components to ensure

that the asset can perform its desired duty at an acceptable level of performance. In this context, the only consideration when deciding when to maintain the asset is whether or not the asset is performing its duty at an acceptable level. However, in reality, the maintenance of any single asset exists in the much larger context of a company (Jardine & Tsang, 2013). There are finite resources, budget, and time that can be allocated to the maintenance of any specific asset, and some assets are more critical to production than others. This "big picture" management of an asset's maintenance is what we refer to as asset health management. It is in this bigger context that reliability engineers and planners must make their decisions about how and when to maintain an asset. Asset health management requires foresight, planning, and–most importantly–risk management.

Maintenance strategies help to roughly allocate resources and plan maintenance schedules ahead of time. There are three general strategies: reactive, preventative, and predictive maintenance (Jardine & Tsang, 2013). We provide a more detailed overview of these strategies a little later. An asset can have different strategies for its different components, and typically, the choice of strategy is dictated by how critical the component is, how expensive it is, and what type of data we can collect. But even with a maintenance strategy, once an asset is put into service, we start to gather new data that can be used to refine/inform the maintenance strategy. For instance, *Chapter* 4 uses failure time data to inform the timing of a bulk-replacement strategy.

Reactive vs Preventative vs Condition-based maintenance strategy

The simplest replacement strategy is a reactive maintenance strategy, whereby components are only replaced once they fail (Heng, Zhang, Tan, & Mathew, 2009). Reactive strategies are used mostly for non-critical components. They are not typically used for mechanical components in mining because the cost due to lost production when an asset fails unexpectedly is orders of magnitude greater than the cost of planned maintenance. On the other hand, a preventative replace-

ment strategy is when components are replaced pre-emptively after a designated period of time or operation. This proactive approach to maintenance is suitable for cheap components whose reliability decreases with time. i.e. components that wear out (most mechanical components). A downfall of preventative maintenance is that it can result in overmanning assets, which is a waste of money and resources. If a component is costly and critical, and it is possible to monitor its condition somehow, then a condition-based maintenance strategy should be used. Condition-based maintenance balances using as much of the component's useful life as possible with the reduced risk of lost production by monitoring the degradation of a component and replacing it when it gets to a predetermined, unacceptable level.

There are obvious ways in which statistical modelling can inform preventative and condition-based strategies. Implementing a preventative replacement strategy requires choosing a pre-emptive time to replace the component. The better the choice, the better the strategy will perform. A component's specific environmental and operating conditions affect its reliability (Meeker et al., 2022). So, if it is possible to use data to "tune" the replacement time to the component's reliability under the specific operating conditions, then the preventative policy will be more successful. Condition-based strategies, on the other hand, are more useful if we can forecast the degradation through time to predict the failure time (useful life) of the component. More detailed and accurate forecasts will result in better maintenance plans and reduce the risk of an unexpected failure. In both cases, the more accurately we can estimate the reliability quantities, the better the strategies will perform. We can estimate these quantities and manage uncertainty around the estimates by fitting reliability models to data.

1.2 Reliability modelling

In the engineering context, reliability is the "ability of an item to perform a required function under given conditions for a given time interval" (ISO, 2016).

This definition is very closely tied to the definition of maintenance actions in *Section* 1.1; maintenance actions are to ensure reliability. In the reliability modelling context, the definition of reliability is slightly different. It is the "probability for an item to perform a required function under given conditions over a given time interval $(0, t)$" (ISO, 2013). In other words, it quantifies the engineering definition of reliability as a probability that a unit will not fail before t, that is, $P(T > t)$. Here, time t can be calendar time, operating time, or some other exposure, such as loading cycles, distance travelled, or throughput (Lee & Whitmore, 2006).

The modelling definition of reliability focuses on binary outcomes (i.e., success/failure data) for a set time interval (Hamada et al., 2008). But typically, we have more detail in data and instead want to estimate the reliability at all values of $t = [0, \infty)$. This representation is the reliability function, $R(t)$. The reliability function can alternatively be represented as its complement, $P(T \leq t)$, the cumulative failure time distribution, $F(t)$ (Meeker et al., 2022). Reliability analysis aims to estimate these functions from data. Two general approaches are taken, depending on the type of observations available: Lifetime modelling (also referred to as failure time models) and degradation models (sometimes referred to as repeat repeated-measures degradation models).

Lifetime modelling The most common form of reliability data are lifetime data. These are the recorded installation and failure times of units in operation. Lifetime modelling, therefore, aims to estimate the failure time distribution from lifetime data. These lifetime data can come from repeated failures of an asset or the lifetimes of a population of assets. The estimated failure time distribution from lifetime analysis allows the analyst to make general statements about the reliability of a population conditional on some exposure time t and sometimes on covariates Moore (2016). However, it is common for reliability datasets to be limited in size or the number of observed failures (Meeker et al., 2022). For example, a particular asset may only have a small number of failures, or in a population of highly reliable assets, only a few may fail over the period of obser-

vation. In these cases, the data are not very informative. To combat a lack of information, the analyst can either supplement the analysis with other sources of information (which we elaborate on in *Part* I) or use a degradation model if they have access to measurements of the degradation process that drives failure (the focus of *Part* II).

Degradation modelling Using an example from Meeker et al. (2022), consider the case that in a lifetime dataset, only two out of one hundred units fail. In this case, the ninety-eight units that did not fail provide no information about how close they were to failure. If, in addition, there are repeated measurements of the level of the degradation mechanism that drives the failure, then degradation analysis allows us to look inside the other ninety-eight units and, more precisely, estimate the failure time distribution. In fact, degradation modelling can be used to derive failure time distributions if there are no failures or even for a single unit that has not yet failed (*Part* II). The connection between degradation models and failure time distribution is well explored (Bae, Kuo, & Kvam, 2007; Lawless & Crowder, 2004; Lu, Meeker, & Escobar, 1996; Meeker et al., 2022), and the connection is typically made using soft failure.

Soft failure is defined by a predetermined threshold of degradation, compared to hard failures, which are when the component can no longer operate Hamada et al. (2008). Hard failures are more stochastic in nature, making them more risky, hence why soft failure is often used. However, regardless of whether a soft or hard failure definition is used, when the failure time distribution is derived from a degradation model, it is important to note that the distribution is conditional on the particular degradation failure mode we are modelling.

Statistical degradation models can be divided into general path models or stochastic process models (Pandey & Yuan, 2006; Si, Wang, Hu, & Zhou, 2011). General path models assign a functional form to the degradation path of a unit, typically a theoretically motivated function such as in (Robinson & Crowder, 2000). This method assumes that the functional form can sufficiently represent

the underlying degradation and that measurement error can completely account for any deviations in measurements from this fixed path. Heterogeneity between units can be added through regression/random effects (Robinson & Crowder, 2000). On the other hand, stochastic processes do not assume a fixed path (Pandey & Yuan, 2006). In Stochastic processes, the jumps in degradation are modelled as random variables, meaning that they account for random variation in the degradation process over time. The stochastic process model we focus on is the Gamma process, like in (Lawless & Crowder, 2004). A rough comparison of the two methods is made in (Ye & Xie, 2015); however, no comprehensive comparisons have been explored. This is an open-ended research question, but we will not tackle it in this thesis.

1.3 Industry Examples

In this thesis, I show two examples of industry problems. Both problems relate to the components of an overland iron ore conveyor. One example is a preventative maintenance problem, and the other is a condition-based one. In *Part* I of the thesis, I look at the preventative replacement of idlers. Whereas in *Part* II I focus on forecasting the degradation of the conveyor belting to inform condition-based decisions. The two components are shown in *Figure* 1.1.

Idlers Idlers (sometimes called rollers) support the weight of the belt and ore. They are relatively cheap components, and there can be hundreds or thousands of them on a single conveyor. Idlers are organised in frames, usually consisting of three idlers: one central idler directly under the belt and two wing idlers supporting the sides of the belt to create the cupped shape. The idlers and idler frames are shown in *Figure* 1.1. Idlers are mechanical components and, therefore, wear out with operation. When an idler fails, it does not cause a direct impact on production; however, failed idlers can damage the belt, and damage to the belt results in major downtime. Reliability engineers need to manage the replacement

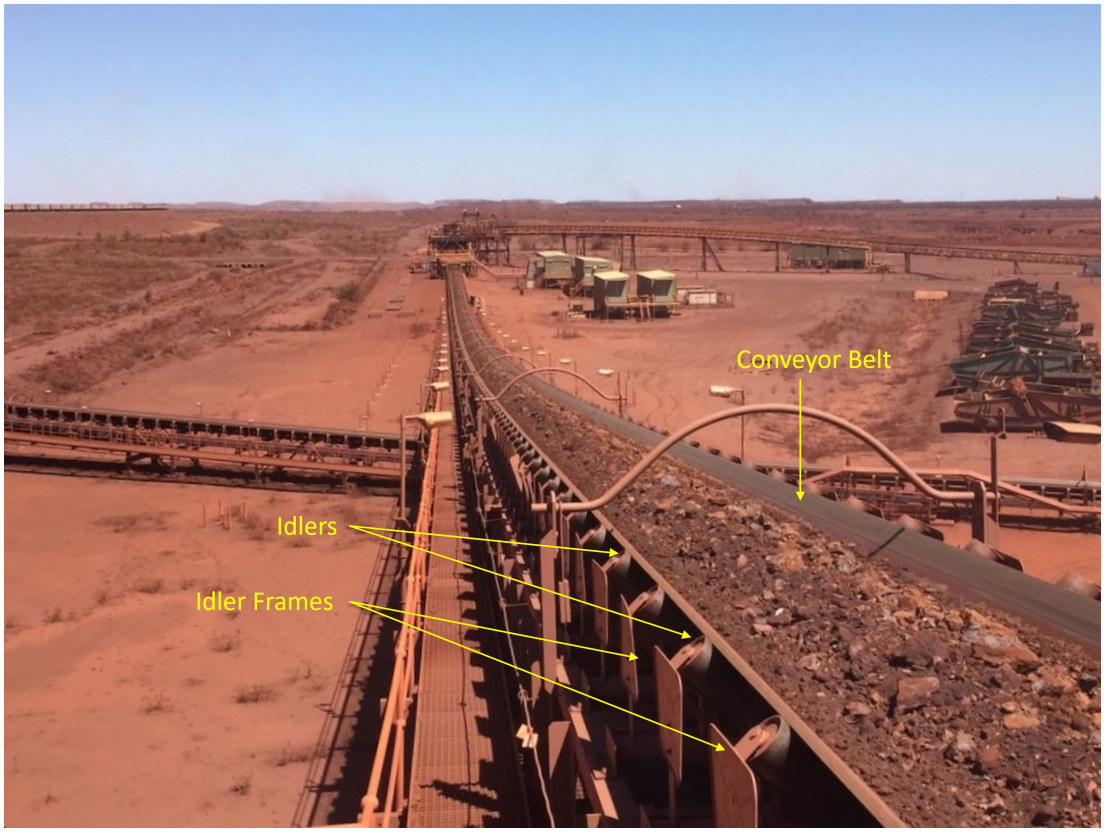

Figure 1.1: An annotated image of an overland iron ore conveyor (Creagh, 2020) showing the belting, idlers, and idler frames.

of the idlers to minimise the risk of them failing and damaging the belt while simultaneously minimising the maintenance cost. It is not yet financially viable to monitor the condition of all idlers on a single conveyor, let alone all conveyors on a mine site. Therefore, a preventative maintenance strategy is used.

Survival data of the idlers–derived from installation and replacement times–can be used to inform the preventative maintenance strategy for the idlers. Unfortunately, the failure data is usually only reliably recorded down to the frame number level, not the position of the idler in the frame. However, when one of the idlers in a frame fails, usually all the idlers in that frame are replaced, meaning that we can model the reliability of the idler frames to inform the preventative maintenance strategy. An example data set is shown in *Figure* 1.2 *a.* The figure shows the frame lifetimes for a single overland conveyor over six years. Because

idlers are long-lasting components and also because they are preventatively replaced, there are many cases where we do not observe the entire lifetime of an idler either because it had not failed by the time we stopped observing it or it was pre-emptively replaced. These partially observed lifetimes are censored lifetimes and are shown in blue in *Figure* 1.2 *a*. We discuss censoring in more detail in *Part* I. *Figure* 1.2 *b* shows the non-censored and censored lifetimes plotted cumulatively. In the ordered plot, we can see that many of the shorter lifetimes are fully observed, while most long lifetimes are obscured by censoring.

Belt The belt of an overland conveyor is much more costly to replace (both in terms of time and money). Furthermore, if the belt fails, then major downtime is inevitable. Over time, the constant loading of ore onto the belt wears away a protective topcoat of rubber, exposing the structural components of the belt to the risk of being damaged by the ore. To ensure the structural integrity of the belt, Engineers stop the belt occasionally to take ultrasonic-thickness (UT) measurements of the topcoat and make sure that it is thick enough to provide adequate protection. An example of the UT data is shown in *Figure* 1.3.

Engineers use these UT measurements to estimate the soft failure time of the belt and plan its replacement, i.e. forecast when the top coat will no longer be thick enough to protect the belt. However, forecasting the wear of the belt to inform decisions requires robust statistical modelling to properly quantify the many different sources of uncertainty, which can be done in the Bayesian framework.

1.4 Bayesian reliability modelling

In the Bayesian statistical framework, probabilities are subjective statements about our uncertainty. In a Bayesian analysis, we aim to construct probability statements about parameters of interest, θ, conditioned on the data y and (implicitly) on the values of covariates (Gelman, Carlin, et al., 2020). That is, we wish to obtain $P(\theta|y)$, known as the posterior distribution. To do this, Bayesian

inference relies on Bayes rule,

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}, \tag{1.1}$$

where $P(y|\theta)$ is the likelihood of the data conditional on θ, $P(y)$ is the marginal distribution of the data, and $P(\theta)$ is the prior distribution of the parameters. The prior distribution encodes our belief about the parameters before observing the data and, therefore, encodes any additional information that we may have about the phenomenon we are modelling, either from historical data or domain expertise. Alternatively, we can write the un-normalised posterior as

$$P(\theta|y) \propto p(y|\theta)P(\theta). \tag{1.2}$$

In practice, however, the posterior distribution is rarely available in closed form, and we need to simulate draws from the posterior distribution using Markov chain Monte-Carlo (MCMC) methods to perform inference. There are several powerful and flexible probabilistic programming languages, such as STAN, which allow us to easily implement MCMC algorithms for complex model structures.

Bayesian workflow Surrounding Bayesian inference is a larger workflow of good statistical practice. Just as there is good practice for statistical modelling, the applied Bayesian community has developed its own workflow tailored to the specifics of Bayesian analysis (Gelman, Vehtari, et al., 2020), the main components of which focus on model construction, drawing from the posterior using MCMC methods and diagnosing issues with computation, sense checking the model with simulated data, evaluating and using the posterior distribution, and comparing and expanding models. Here, I give a very high-level overview of the components in the workflow most relevant to this thesis. See (Gelman, Vehtari, et al., 2020) for an in-depth description of a Bayesian workflow. The components of this workflow that I use within this thesis are; conditional modelling, prior

predictive checking, sampling and diagnostics, posterior predictive checking and posterior inference, and model comparison.

Model specification The first step of any Bayesian analysis is to postulate a joint probability distribution for the model. For complicated processes, this first step can be simplified by using a Bayesian Hierarchical (multi-level) Modelling approach, which uses the fundamental notion of the *the law of total probability*, $P(A, B, C) = P(A|B, C)P(B|C)P(C)$, to decompose a complicated joint probability into a string of simpler conditional probabilities (Wikle, Zammit-Mangion, & Cressie, 2019, p. 13). Berliner (1996) proposed Bayesian Hierarchical Modelling (BHM) as a way of studying an underlying latent process by breaking the joint probability of the data, process, and parameters down into three sub-models;

$$\begin{aligned} p(\text{data}, \text{process}, \text{parameter}) = & p(\text{data}|\text{process}, \text{parameter}) && \text{data model} \\ & p(\text{process}|\text{parameter}) && \text{process model} \\ & p(\text{parameter}) && \text{parameter model} \end{aligned}$$

The first level is the data model, $p(\text{data}|\text{process}, \text{parameter})$, which describes the observation process. The second level in the hierarchy is the process model, $p(\text{process}|\text{parameter})$. It describes the underlying process that is of scientific interest. The third level in the hierarchy, $p(\text{parameter})$, is the parameter model, and in a Bayesian setting refers to the prior distribution. Each of these different levels in the hierarchy can also be made up of smaller constituent conditional models. Cressie and Wikle (2011) advocate using the BHM approach for studying underlying latent spatial and spatio-temporal processes, but the same general approach is used to break down models for nested data structures under the term multi-level modelling (Gelman, Carlin, et al., 2020).

In the last level of this hierarchical structure, the prior distribution summarizes any a priori beliefs the analyst has about the process they are trying to study before having observed the data. There are two different ways in which

this information is encoded into the parameter model: the choice of distribution, and the values of the hyperparameters. Before the advent of contemporary sampling algorithms, Bayesian analysis relied on conjugate prior distributions, or convenient prior distributions that facilitated the use of Gibbs samplers or conventional Metropolis-Hastings algorithms (Gilks, Richardson, & Spiegelhalter, 1996). However, with the development of more efficient sampling algorithms such as Hamiltonian Monte Carlo (Betancourt, 2017), we are no longer limited by such requirements and can select priors that reflect our state of knowledge, facilitate efficient computation, and that can be justified and evaluated in a principled way. A useful tool for choosing the parameter model and understanding how it interacts with the process and data models is to simulate data from the full Bayesian model.

Simulation for model checking Bayesian analysis generally uses a fully generative model, so long as the prior is proper. When using a generative model, the model can not only be run "backwards" to perform inference but also "forward" to simulate fictitious data. For example, likelihood methods require a distribution for the data given the parameters, $P(y|\theta)$, but since there is no distribution for the parameters, there is no way of simulating data from this model unless we supply some reasonable values of the parameters. Bayesian analysis, on the other hand, specifies a distribution for both the data, y, and the parameters, θ; $P(y, \theta) = P(y|\theta)P(\theta)$. Using this generative characteristic of Bayesian models to simulate data is useful for understanding unfamiliar or complicated models.

Prior predictive simulation is when we simulate data from the model before conditioning on the observed data and is one of the key steps in the 'Bayesian workflow' (**?**, Figure 1). Prior predictive simulation can be used to understand the plausibility of a parameter model in the context of the likelihood (Gelman, Simpson, & Betancourt, 2017). Prior predictive simulations can also be a useful tool to elicit domain expert knowledge on the measurable outcome in order to develop an informative prior, rather than specifying domain knowledge directly

on the parameters of the model. In the terminology of **?**, priors that when combined with the likelihood lead to simulated data that could be plausibly observed are known as *weakly informative priors*. According to **?**, such weakly informative priors should, for the most part, lead to plausible simulated data but may have some mass around extreme, but not completely implausible, realizations. Nevertheless, when using prior predictive checks to evaluate priors and to find sensible ones, the idea is *not* to try different values of the hyperparameters until the realizations are concentrated around the data that we are analyzing; instead, as **?** write, the analyst "should have enough familiarity with the subject matter to look at prior predictive simulations ... without needing to make direct comparisons with the data that will be used for model fitting." They go on to say that "a *reasonable* [our emphasis] prior is a prior that yields a reasonable prior data-generating process, not that the researcher should tailor the prior to suit the particular observations in hand."

Furthermore, simulating data from the model and then re-fitting the model to the simulated data is another useful way in which prior predictive simulation can help us better understand our Bayesian model. By fitting the model to simulated data for which we know the true parameter values, we gauge an understanding of what our model is capable of learning from the data... (example)

HMC and diagnostics Throughout this thesis, I use the No-U-Turn sampler (**?**) implemented in the probabilistic programming language Stan (Stan Development Team, 2022) to draw samples from the posterior distributions of Bayesian models. The No-U-Turn sampler is an adaptive variant of the successful Hamiltonian Monte Carlo (HMC) algorithm (**?**). HMC borrows the idea of hamiltonian dynamics from physics to improve the random walk behaviour of traditional MCMC methods in order to move much more rapidly through the target distribution (Gelman, Carlin, et al., 2020). The No-U-Turn sampler improves the HMC algorithm by alleviating the user from the difficult task of choosing the step size and number of steps used to approximate the hamiltonian trajectories (**?**).

The theoretical foundations of HMC are formulated in differential geometry, an advanced field of mathematics, and so I do not discuss the details of HMC in this thesis. Betancourt (2017) provides a very nice conceptual introduction to HMC and a more rigorous overview is given in Gelman, Carlin, et al. (2020, p. 300), any reader interested in the specifics should look to **?**.

One added advantage of using a variant of HMC is the useful within chain diagnostics. For most general MCMC methods, we can check that chains have mixed using numerical summaries such as the potential scale reduction factor, $\hat{R}$, and follow up with trace plots of the individual chains, and we can check for inefficient exploration of the posterior using auto-correlation functions of each chain. However, it is difficult to diagnose the reasons that sampling is poorly behaved. Alternatively, when using HMC or one of its variants, one of the requirements for the algorithm to work efficiently is that the geometry of the set that contains the bulk of the target distribution is fairly smooth (**?**). While it is most often not possible to check for this condition mathematically, it can be checked numerically. When this set is not smooth, the leapfrog algorithm used to approximate the hamiltonian trajectories diverges from the energy conserving trajectory in the areas of high curvature (non-smooth areas), and race off to infinity. Using a threshold energy above which trajectories are considered divergent trajectories, we can diagnose problematic areas in the posterior(**?**), referred to by some as degeneracies(Betancourt, 2020). Sometimes these degeneracies and the poor sampling that results can be resolved by re-parametrising the model (**?**) while in other cases it cannot. In the latter, the degenerate behaviour may indicate an issue with the model.

If divergent transitions are present, then visually plotting the divergent trajectories alongside the non-divergent trajectories highlights the troublesome areas of high curvature in the posterior that obstructs exploration, since the true divergent transitions will be clustered around the problematic areas of the parameter space (**?**). Two useful visual diagnostics are the bivariate scatter plot, also called

a pairs plot, and the parallel coordinate plot. *Figure* **??** shows an example of the two plots. Since divergent traditions are flagged using a threshold, it is possible that some divergences are false positives. If this is the case, then their distributions should match that of the non-divergent samples. However, if there are in fact areas of high curvature in the posterior, then the divergent transitions should be spatially correlated with these areas. Throughout this thesis, I predominately use pairs plot for checking sampling, however, in more complicated cases (Chapter 4) I also use parallel coordinate plots.

Evaluating and using the posterior Once we have postulated the model, generated samples from the posterior, and are confident that the samples sufficiently represent the posterior, we can then use the posterior samples to perform inference and inform decisions. The result of fitting the model with MCMC methods is that we obtain S simulations of the parameters θ from their posterior distribution,

$$\theta_s \sim {}_{\theta|y}\pi. \tag{1.3}$$

Using the posterior draws of the parameters, we can not only find the estimated expected values of the parameters and credible intervals but also posterior predictive distributions for new data, and uncertainty estimates for new functions of the parameters, such as failure time distributions.

Posterior predictive checking ...

Comparing models Once we have fit a series of suitable models for a data set, we next want to evaluate how well they describe the true data generating process and to compare them. To do so, we evaluate their ability to predict new observations. In the absence of an independent, external test set, it is conventional to use *information criteria* to compare models. These criteria, such as AIC, DIC, and others, are used to seek a compromise between goodness-of-fit and model complexity and to assess out-of-sample prediction accuracy. AIC

and DIC are easy to calculate, but they are not fully Bayesian; hence, criteria such as WAIC (Watanabe-Akaike Information Criterion) and leave-one-out cross-validation (LOO-CV) are to be preferred (**?**).

To compare models, we use LOO-CV, where the measure of distributional predictive accuracy is the *log score*. The log score is the log likelihood of a new observation $\tilde{y}_i$ given the posterior distribution of the parameters. It is also the probability of the new observation under the posterior predictive density, and it can be written as,

$$\text{lpd} = \log \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta = \log p(\tilde{y}_i|y), \tag{1.4}$$

where θ is the set of parameters and y is the observed data. The parameters can also include unobserved latent variables. The measure in eq. (5.1) is called the *log posterior density (lpd)*. If we observe multiple new data points $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_I)$, this can be dealt with in a point-wise fashion using the *log point-wise posterior density (lppd)*,

$$\text{lppd} = \sum_{i=1}^{I} \log p(\tilde{y}_i|y). \tag{1.5}$$

In cases where each of the new observations are independent of one another given the parameters and latent variables, which is the case for the majority of cases in this thesis, the point-wise predictive density is equal to the joint predictive density of the set of new observations; $\log p(\tilde{y}|y) = \sum_{i=1}^{I} \log p(\tilde{y}_i|y)$. **??** and **??** are defined for a given set of new observations, but the new unobserved data points $\tilde{y}_i$ arises from the true data generating process and so are a random variable with distribution

$$\tilde{Y}_i = f(\tilde{y}_i). \tag{1.6}$$

Hence, a better measure of predicted accuracy is the expectation of the lppd, the

elppd, which is obtained by integrating over $\tilde{Y}_i$

$$\text{elppd} = \sum_{i=1}^{I} \int \log p(\tilde{y}_i|y) f(\tilde{y}_i) d\tilde{y}_i. \tag{1.7}$$

In the context of Bayesian models fit with MCMC, the computed elppd can be calculated by averaging over the $s = 1, ..., S$ MCMC draws from the posterior,

$$\text{computed elppd} = \sum_{i=1}^{I} \int \log \frac{1}{S} \sum_{s=1}^{S} p(\tilde{y}_i|\theta^s) f(\tilde{y}_i) d\tilde{y}_i. \tag{1.8}$$

Although this would be the best measure of predictive accuracy for our Bayesian models, we obviously do not know the true data generating process and so cannot define $f(\tilde{y}_i)$. We can, however, approximate the expectation above by using cross-validation whereby we iteratively withhold a portion of the observed data, sample from the posterior conditioned on the wrest of the data, and then calculate the log likelihood of the withheld portion of the data given the samples from the posterior. The simplest form of cross-validation is leave-one-out (LOO), where we withhold each observation,

$$\text{elppd}_{loo} = \sum_{i=1}^{I} \log \frac{1}{S} \sum_{s=1}^{S} p(y_i|\, [\theta]^s_{-[i]}). \tag{1.9}$$

$[\theta]^s_{-[i]}$ is the posterior draws for the set of parameters and latent variables conditioned on all the observed data except the withheld observation y_i.

In hierarchical models, the definition of a new observation and the likelihood of those observations depends on what aspect of the model's predictive performance we are trying to assess. For example, in a degradation dataset with multiple units and multiple observations per unit, we could obtain new observations for the same units at the same observation times, $\tilde{y}_{n,i}|z_{n,i}, \sigma$ (although this case is a bit unrealistic), new observations for an observed unit at some time in the future, $\tilde{y}_{n,I+1}|\tilde{z}_{n,I+1}, \sigma$, or we could observe an entirely new unit, $\tilde{y}_{n+1}|\tilde{z}_{n+1}, \sigma$, where $\tilde{y}_{n+1} = [\tilde{y}_{n+1,1}, ..., \tilde{y}_{n+1,I}]$. In both all three cases, the likelihood of the ob-

servations conditional on the draws from the posterior predictive distribution are much the same, since for the most part we assume in our data models that observations are independent given the underlying degradation parameters. However, the method used for constructing the predictive distributions of the parameters and intermediate quantities will be different depending on the model's hierarchical structure.

1.5 Structure of this thesis

This chapter has introduced the industry-derived motivation for the work in this thesis and provided a high-level introduction to the strong threads that flow through the body of work: reliability analysis and Bayesian model building. The remaining body of the thesis is broken into two parts and unified at the end by a general discussion/concluding chapter. The two parts of the thesis separate the works into those that address lifetime analysis and those that address degradation modelling. At the beginning of each part, I've included a preamble that provides a background on the industry placement project/s that motivated the work in the part and points out which chapters have been published or submitted for publication. I hope these short sections of metadiscourse provide a glimpse into the extra work that has gone into defining novel research problems whose solutions are truly useful to reliability practitioners in the industry.

The first part of the thesis focuses on lifetime analysis. Specifically, how we can obtain reliable inference from Weibull analysis of data that are heavily censored by thoughtfully constructing an informative joint prior for the shape and scale parameters of the Weibull distribution. The part comprises three chapters: chapters two, three, and four. *Chapter Two* starts with the introduction of Weibull lifetime analysis and censoring of lifetime data and then proceeds to demonstrate how, when lifetime data are heavily censored, fitting the model with maximum likelihood or Bayesian methods with commonly used priors results in biased parameter estimates. The chapter concludes by demonstrating a method

for constructing an informative joint prior, which encodes information about how the parameters covary with one another and shows how encoding information into the model in this way reduces the effects of the bias caused by heavy censoring, allowing us to obtain usable estimates of the lifetime parameters. *Chapter Three* then presents a simulation study demonstrating the systematic reduction in bias provided by the informative joint prior and explores the method's limitations. From the findings in the simulations study, we consolidate our recommendations to practitioners when analysing heavily censored lifetime data. In *Chapter Four*, we apply the methods and recommendations from chapters two and three to an industry dataset. The case study analyses the censored lifetime data of idlers from an overland iron ore conveyor shown in *Figure* 1.2 and demonstrates how the posterior can be propagated through functions used in the decision-making process-such as cost functions-to better incorporate uncertainty and risk into the decision making process.

Part two of the thesis is more loosely structured. The three chapters–five, six, and seven–show an iterative model-building process centred around a Gamma stochastic process for degradation. *Chapter Five* demonstrates how the Gamma process can be extended through the Bayesian Hierarchical modelling (BHM) framework to account for noisy observations. *Chapter Six* then shows the expansion of the noisy gamma process model through the same BHM structure and the use of functional data analysis to model the wearing surface of an overland conveyor's belt. *Chapter Seven* then explores possible ways to incorporate a spatial random effect in the belt wear model.

The concluding chapter, *Chapter Eight*, ties the two parts of work in the thesis back to the overarching topics of reliability and maintenance. At this higher level, the thesis concludes with a discussion of the strengths and limitations of the work, areas of future work, and the implications of this work for industry practitioners.

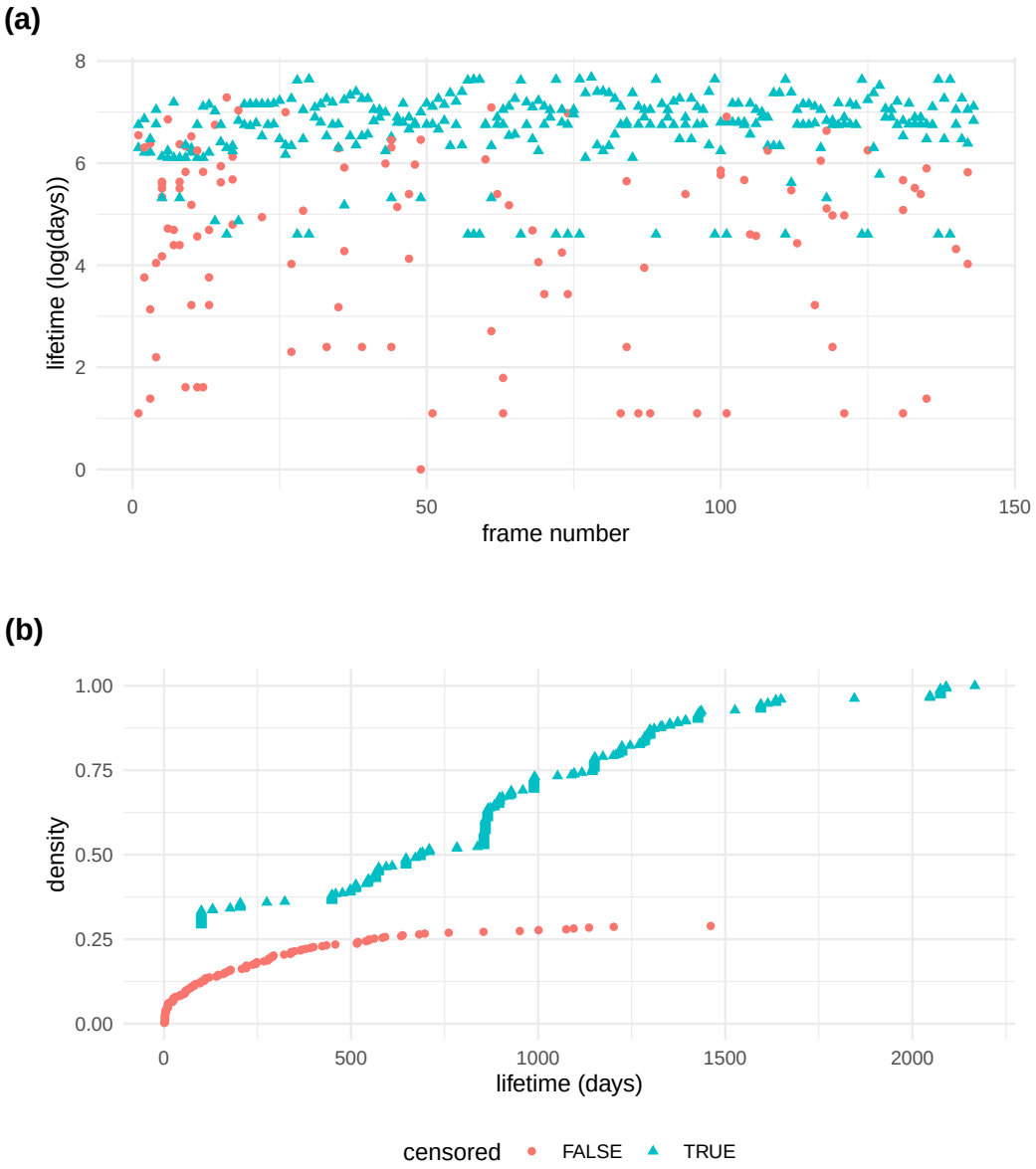

Figure 1.2: (a) frame lifetimes. (b) eCDF of lifetimes

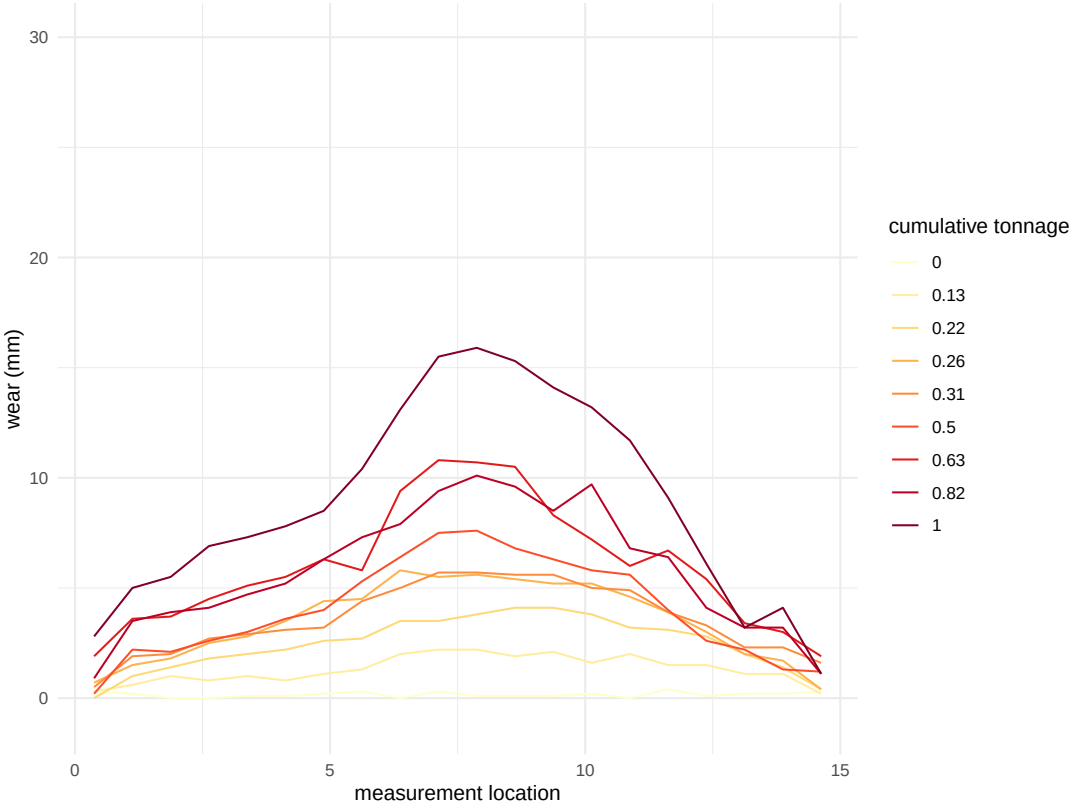

Figure 1.3: Belt UT measurement data.

Part I

Part one: lifetime analysis

A preamble about which chapters have been published and which chapters came from industry placements.

"Since the initial approach was published in Leadbetter et al. (2021), I have taken an alternative approach that does..."

Chapter 2

Heavily censored lifetime data

Computerised maintenance management systems (CMMS) such as SAP (SAP SE, 2023) are now embedded in companies maintenance procedures, meaning that these companies now posses large scale datasets of component installation and replacement times. A natural use of these personalised failure time data sets is for tailoring replacement strategies for the companies specific operating environments (Meeker et al., 2022, p. 13), rather than solely relying on the manufacturers recommendations. One problem however, is that these large observational datasets collected through CMMS are much messier that the experimental ones used by manufacturers in traditional reliability/warranty analysis. This messiness comes about because of reporting issues, incomplete historic records, and the fact that most components are pre-emptively replaced before they fail because of the risk to production and employee safety. The result is that many of the valuable data sets stored in CMMS systems are incomplete in the form of censoring and left-truncation. Censoring is when the true lifetime of a failed component is not know but either an upper bound, lower bound, or both are known. Whereas left-truncation arises when only units that have lasted more than some truncation time are observed.

The censored and left-truncated nature of such data make what would otherwise be a very strait forward analysis far more complicated. Worse yet, the

incompleteness of the data is not always obvious and mistreatment during analysis can lead to biased or miss informed results. In this part of the thesis I focus on the lifetime analysis of the idler-frame dataset shown in 1.3. In this case, right censoring arises due to the set of idlers in a frame either being preventatively replace or still being in operation when the data were analysed. A portion of the lifetimes are also left-truncated, since any idlers that were installed and failed before t_{start}, when failures started being recorded in the CMMS, are not captured in the dataset. Treatment of right censored and left-truncated data was addressed by Hong, Meeker, and McCalley (2009). A further complicating factor of the idler frame dataset is that the installation time of idler-frames that were already in operation when data started being captured in the CMMS are unknown, meaning that the left-truncated lifetimes are also censored and have unknown truncation times. This issue is sometime referred to as unknown initial conditions () or unknown exposure history (). In this chapter, I propose a method for handling such cases in a Bayesian framework by imputing censored lifetimes and along with them, the truncation times.

Incompleteness of a dataset due to censoring and truncation increase uncertainty in lifetime analysis, particularly with respect to the upper tail of the lifetime distribution, since the data only contains partial information about the longer lifetimes. To improve analysis, domain knowledge can be used to fill the gap left by censoring, however only if the prior is constructed properly to do so. Kaminskiy and Krivtsov (2005) propose a method for constructing a joint prior for the two parameters of the Weibull distribution by eliciting domain expert knowledge on the CDF...

In this chapter of the thesis, I show how right censored and left-truncated lifetimes with unknown exposure history can be handled using fully a Bayesian treatment, and how the methods of Kaminskiy and Krivtsov (2005) for developing a joint prior for the two parameters of the Weibull distribution can be incorporated into the model to supplement the analysis with additional sources

of information and hence "fill in the gaps" left by the incomplete lifetime data. In this chapter I demonstrate the proposed methods using fictitious lifetime data simulated from known parameter values in a way that emulates the idler frame lifetime data to show that the method is capable of reclaiming the true parameter values. In the next chapter, I apply the method to the real idler frame dataset, and show how the fully Bayesian treatment has advantages for... (Follow Meekers paper on the transformers.)

In the remainder of this chapter is structured...

- Background on Weibull lifetime, censoring, left-truncation, left-truncation with unknown exposure history, constructing a joint prior.
- Demonstrate through simulation that the method accounts for the incompleteness of the data without introducing much bias.
- Extend the prior proposed in Kaminskiy and Krivtsov (2005) and demonstrate it's usefulness in this context.

2.1 Background

Lifetime analysis, also called survival analysis, is the analysis of failure time data from a population of particular components/assets to derive the risk of failure of a component dependent on it's level of exposure (usually some form of time) and sometimes other covariates (Moore, 2016). From here on I will use the general term unit/s to refer to individual/groups of the same asset or component. Lifetime analysis of a population of units typically takes place by first specifying a sampling distribution for the lifetimes by choosing some parametric lifetime distribution for the units and incorporating any observational characteristics of the data–for example censoring–then, estimating the parters of the distribution from failure time data using an appropriate inferential mechanism, and finally using the fitted model to derive useful reliability measures about the population

which can be used to inform asset management plans. When done in a Bayesian context, the first step of this process also includes specifying a prior distribution. From the resulting inference, we can devise optimal replacement strategies that minimise the risk of unplanned failures, and hence the risk of lost production, and also the cost of the maintenance strategy.

2.1.1 Lifetime distribution

The lifetimes of the units are modelled as a random variable defined in terms of t, the exposure time, on $[0, \infty)$. t is some continuous or discrete exposure time from a clearly defined origin, the installation of the component, to a well defined event, the failure of the component. In reliability analysis, the exposure is typically absolute time, the operating time of the unit, or cycles of operation. In this analysis I use absolute time since operating time was not available. Next, a specific parametric lifetime distribution is chosen for the random variable t, $p(t|\theta)$, expressed as the probability density function (PDF) and the parameters of the lifetime distribution are estimated from the data. Once the estimates are obtained, different specifications of the lifetime distribution can be used to draw useful insights in order to inform decisions:

- **Cumulative distribution function** (CDF), $F(t)$, it the probability that a unit will have failed by time T, i.e. $P(t <= T)$. It is also sometimes called the cumulative risk function.
- **Survival function**, $S(t)$, is the complement of the CDF and defines the probability of a unit surviving up to an exposure time t.
- **Hazard function**, $h(t)$, which is the instantaneous failure rate, i.e. the probability that given a unit has survived up to time T it will fail in the next small interval.

For example, the CDF quantifies the risk of unplanned failures given a chosen preventative maintenance interval, and the Hazard function identifies if a units

risk of failure increases as it ages and therefore if a preventative maintenance strategy is even suitable at all.

2.1.2 The Weibull distribution

In the analysis that follows, I use the Weibull distribution to model the idler frame lifetimes, that is

$$y|\beta,\eta \sim \text{Weibull}(\beta,\eta), \tag{2.1}$$

where β is the shape parameter and η is the scale. The Weibull distribution is a commonly used lifetime distribution because of its ability to capture an increasing, constant, or decreasing risk of failure. In addition, the weibull distribution is the limiting distribution for the minimum value in a sample when the sample space is lower bounded; such as lifetimes, which must be greater than zero. This characteristic of the weibull distribution gives it a convenient interpretation in component reliability; the lifetime of a unit is the time of the first occurring catastrophic failure mode of the unit. In my analysis that follows, I use the coupled parameterization of the two-parameter Weibull distribution, which has PDF

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp^{-\left(\frac{t}{\eta}\right)^{\beta}}, \tag{2.2}$$

CDF

$$F(t) = 1 - \exp^{-\left(\frac{t}{\eta}\right)^{\beta}}, \tag{2.3}$$

and hazard function

$$h(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}. \tag{2.4}$$

The shape parameter β dictates whether the hazard increases, $\beta > 1$, de-

creases, $\beta < 1$, or stays constant, $\beta = 1$. The effect of the shape parameter on the hazard function is demonstrated in *Figure* 2.1. Practically speaking, if the hazard function increases with exposure, then this corresponds to a wear out failure mechanism, where as if it decreases then this corresponds to infant mortality. This is important from a maintenance perspective because if the component does not wear out, then preventative replacement policy is not suitable (Jardine & Tsang, 2013). In other words, we want to be sure that $\beta > 1$ before implementing a preventative policy.

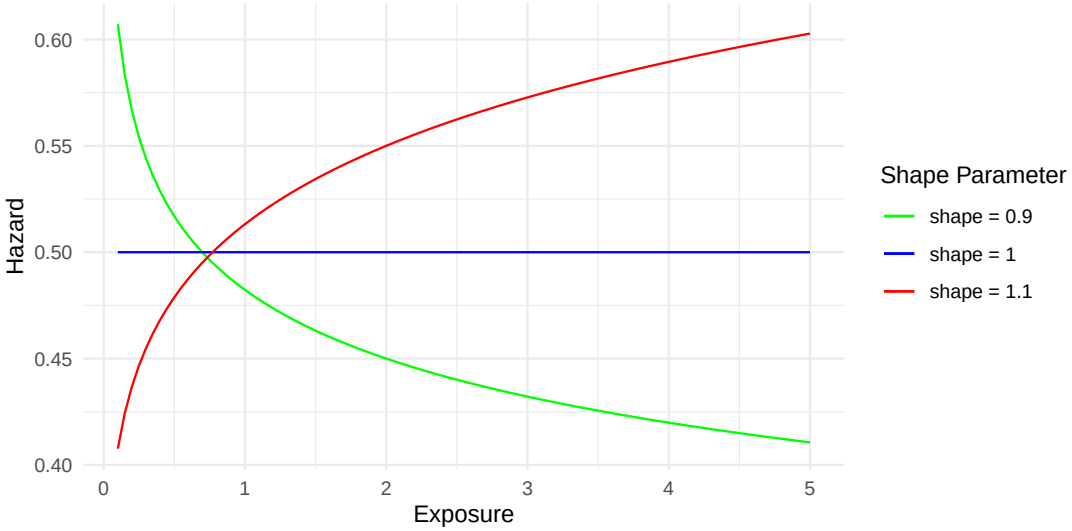

Figure 2.1: Hazard function for $\beta = 0.9$, $\beta = 1$, and $\beta = 1.1$.

2.1.3 Censoring

It is very common for lifetime data to be censored. Censoring occurs when we only partly observe the lifetime of a unit, or in other words we only observe upper and lower bounds for the lifetime. There are three types of censoring: left, interval, and right censoring, but all are treated in much the same why. *Figure* 2.2 demonstrates these three types of censoring. In the figure, three units are installed at time t_0 and then two follow up inspections are performed at times t_1 and t_2. The true, partially observed, failure times of each unit are shown as

crosses in the figure. Left censoring occurs when we only observe an upper bound of the lifetime, an example is if we know the install time of a unit and observe it as failed at t_1 and so we only know that the true value of the lifetime must be less than t_1. Interval censoring occurs if we only know an upper and lower bound for the lifetime, i.e. if a unit fails between inspection times then we know that the lifetime must be greater than t_1 but less than t_2. Right censoring occurs when we only know a lower bound for the lifetime. For example, if a unit is still in operation when we perform our analysis, e.g. it has lasted longer than t_2, then we only know that the true value of the lifetime must be greater than t_2. Right and left censoring are special cases of interval censoring where the upper or lower bound of the lifetime are infinity or zero respectively. Left censoring is fairly uncommon and so In the discussions that follow, I focus on right and interval censoring but all of the methods can be easily extended to accommodate left censored data as well.

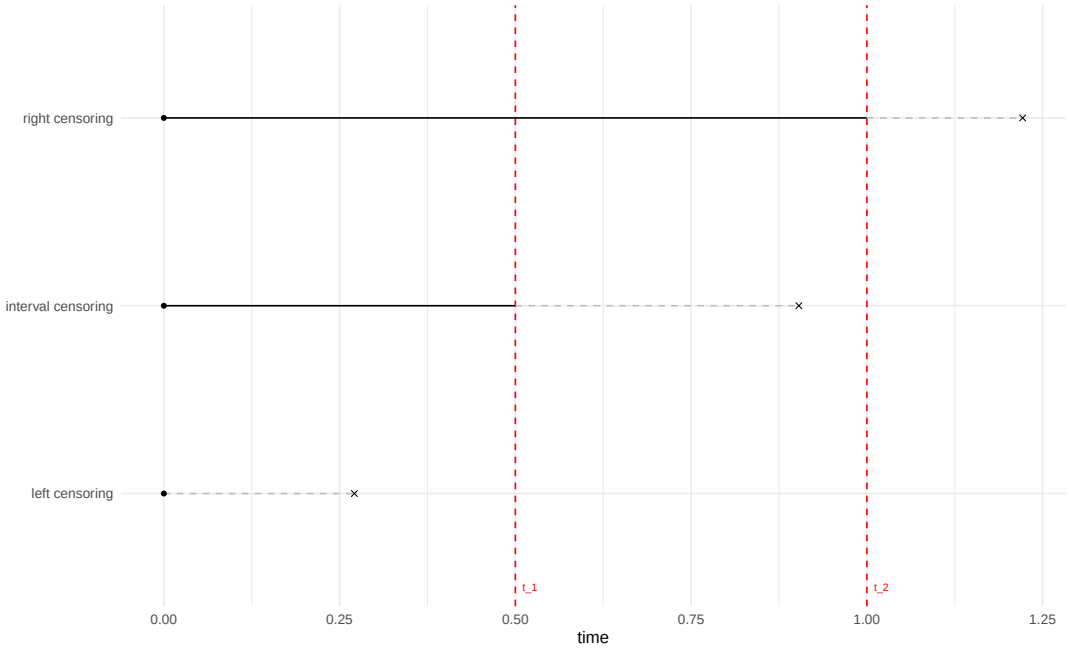

Figure 2.2

One way of handling censored data is to treat the censored lifetimes as missing data, which in a Bayesian framework is to treat them as a random variable in the

model (Reich & Ghosh, 2019, p.211), and constrain their values to fall within the upper and lower censoring bounds (Stan Development Team, 2024). This is easily done in probabilistic programming languages like Stan. Treating the missing data as random variables in the model requires us to specify a distribution for the censored lifetimes, which we assume is the same distribution as the rest of the population,

$$y_i^C|\beta,\eta \sim \text{Weibull}_{t_i^{\text{Lower}}}^{t_i^{\text{Upper}}}(\beta,\eta). \tag{2.5}$$

Here y_i^C is an unobserved censored lifetimes, and the superscript t^{Upper} and subscript t^{Lower} indicate that the distribution is constrained by the upper and lower censoring times. Once the missing lifetimes have been imputed, the likelihood of the observed and imputed lifetimes can be calculated in the same way as a typical lifetime dataset with no censoring. This approach of imputing the censored lifetimes is not unique to Bayesian methods. The same can be done using an Expectation Maximisation algorithm and maximum likelihood (Mitra, 2013). However, using the Bayesian approach, along with MCMC methods, it is very simple to derive uncertainty intervals for the parameters, imputed values, and useful quantities.

An alternative approach is to simply integrate out the censored observations. The probability that a censored observation falls between the upper and lower censoring times is

$$Pr\left[t^{\text{Lower}} < y_i^C \leq t^{\text{Upper}}\right] = \int_{t^{\text{Lower}}}^{t^{\text{Upper}}} f\left(y_i^C\right) dy_i^C = F\left(t^{\text{Upper}}\right) - F\left(t^{\text{Upper}}\right). \tag{2.6}$$

By integrating out the censored observations, the likelihood can be written as

$$L\left(\theta|y^O, t^{\text{Upper}}, t^{\text{Lower}}\right) = \prod_{i=1}^{N^O} f(y_i^O) \prod_{j=1}^{N^C} \left[F(t_j^{\text{Upper}}) - F(t_j^{\text{Lower}})\right], \tag{2.7}$$

where θ are the parameters of the lifetime distribution, N^O and N^C are the number of fully observed and censored observations respectively, and y_i^O are the fully observed lifetime. This second approach is much more commonly used, particularly in the reliability literature (). However, as I show later, for the particular problem when data are also left-truncated with unknown installation times, it is convenient to frame the model in the form of the first imputation approach.

2.1.4 Left-truncation

Truncation arises when a sample comes from an incomplete population, or in other words, there is some criteria that part of the population must appease in order to be observable (Guo, 1993). Left-truncation, for example, arises when some units must survive up to a certain time in order to be included in the dataset. It is also possible for data to be right- or doubly-truncated, but left-truncation is the most common in lifetime data. The definition of left-truncation and left-censoring may seem very similar, however, they are distinctly different (Mitra, 2013). Censoring is a characteristic of the sample, i.e. we know the number of left-censored observations but not the exact values of their lifetimes, whereas truncation is a characteristic of the population because we do not know how many units were not included in the dataset because they did not survive past the truncation time(the time from the installation of the unit to the start of observation) and hence our sample is not representative of the true population. An example of a left-truncated dataset is shown in *Figure* 2.3.

In *Figure* 2.3, units that were installed prior to t_{start} come from a left-truncated distribution since any other unit that was installed at the same time but did not last until t_{start} are not included in the sample. The left-truncated cases caused by the start of the observation period (units three, four, five, and seven in *Figure* 2.3) tend to over-represent low-risk cases since any high risk case installed at the same time is absent from the sample (Guo, 1993). Observations

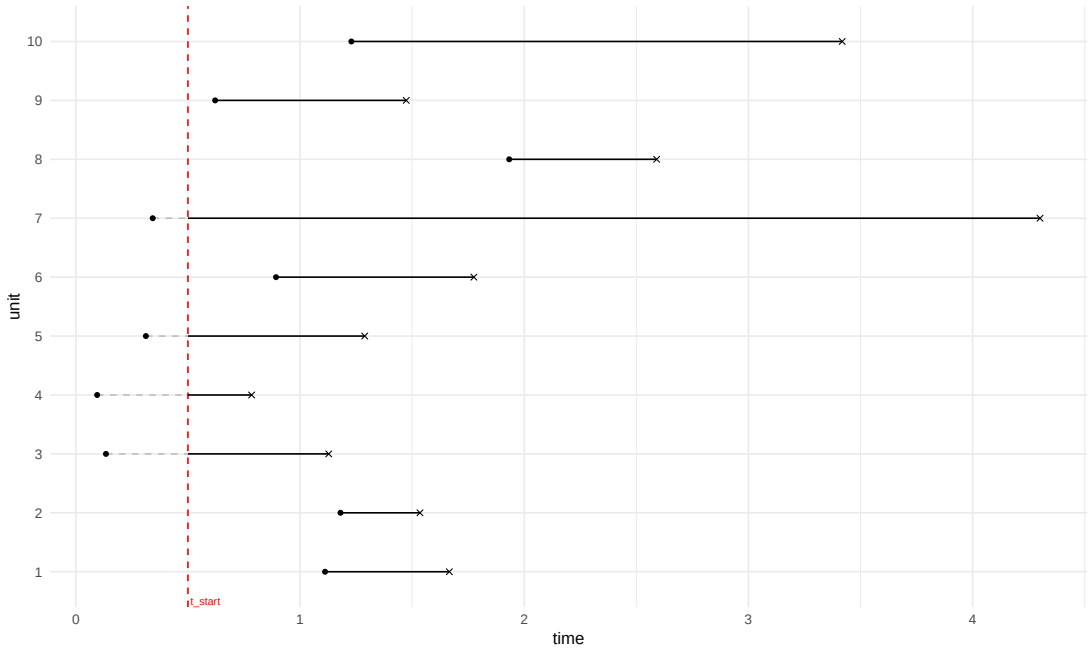

Figure 2.3

that arise from a left-truncated distribution can be included in the analysis by re-normalising their lifetime distribution by dividing by the probability of surviving past the truncation time;

$$L\left(\theta|y_i^T\right) = \frac{f\left(y_i^T\right)}{1 - F\left(\tau_i^{\text{Left}}\right)}, \tag{2.8}$$

where y_i^T is the left-truncated lifetime, and τ_i^{Left} is the truncation time.

2.1.5 Left-truncation and right-censoring

A common scenario in reliability datasets is a combination of both left-truncation and right-censoring. This case naturally arises in historic observational datasets, such as those found in CMMS, where units are repeatedly replaced once they fail and any units that were installed and failed before the start of the observation process (which might be the data a new CMMS was adopted) are absent in the dataset. *Figure* 2.4 shows a toy example of this case, where three units are repeatedly replaced when they fail and we start to observe their failures at

t_{start} and stop at t_{end}. Any lifetimes that fail before t_{start} are unobserved and so are greyed out, resulting in the first observed lifetime of each unit being a left-truncated sample. Lifetimes that surpass t_{end} are only partially observed (right censored), hence the portion of these lifetimes that sits to the right of t_{end} is also greyed out. Hong et al. (2009); Kundu and Mitra (2016); Mitra (2013) analyse a dataset of electrical transformer failures that follows this general structure and Mittman (2018) looks at a similar case for computer hard drives. The idler frame failure data is also an example of this type of dataset.

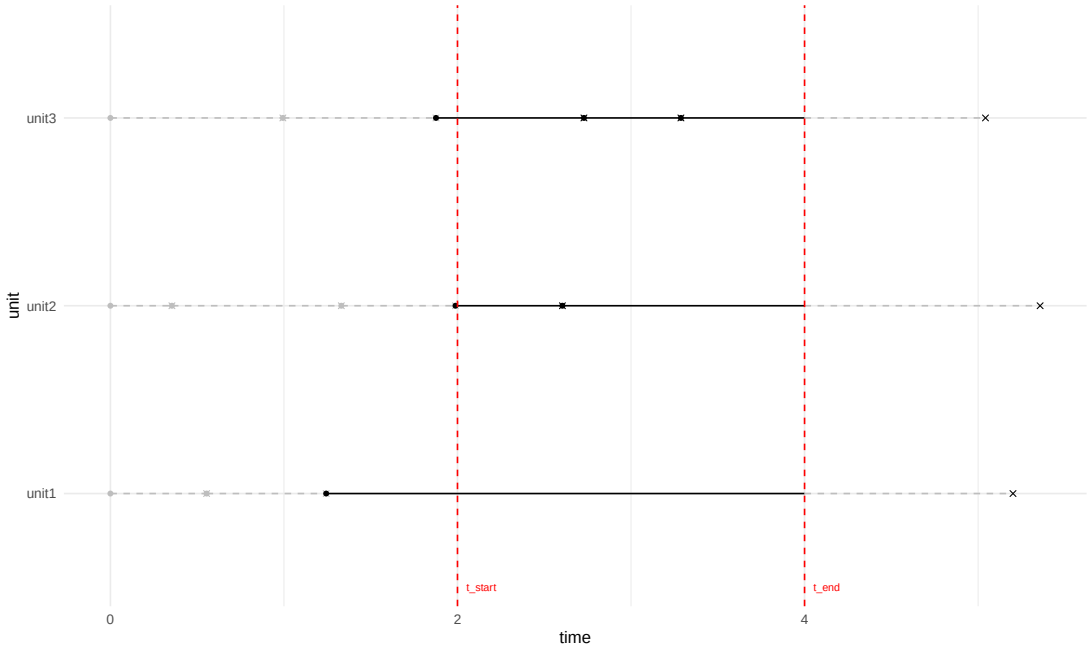

Figure 2.4

Hong et al. (2009) shows how a likelihood for data that is left-truncated and right-censored can be constructed using the general approach of integrating out the censored observations,

$$L\left(\theta|y^O, t^{\text{Lower}}, \tau_i^L\right) = \prod_{i=1}^{N^O}\left[\frac{f(y_i^O)}{1-F(\tau_i^L)}\right]\prod_{j=1}^{N^C}\left[\frac{1-F(t_j^{\text{Lower}})}{1-F(\tau_j^L)}\right], \tag{2.9}$$

where τ^L are the truncation times and if the lifetime is not truncated $\tau^L = 0$. Kundu and Mitra (2016) then implements the same approach in a Bayesian

framework using a gibbs sampling algorithm to draw samples from the posterior. Mitra (2013) takes the alternative approach of imputing the censored lifetimes using an expectation maximisation algorithm and the complete data likelihood

$$L\left(\theta|y^O,\hat{y}^C,\tau_i^L\right)=\prod_{i=1}^{N^O}\left[\frac{f(y_i^O)}{1-F(\tau_i^L)}\right]\prod_{j=1}^{N^C}\left[\frac{f(\hat{y}_j^C)}{1-F(\tau_j^L)}\right], \tag{2.10}$$

where the $\hat{y}_j^C$ are the imputed values of the censored observations. I express the same approach as 2.10 in a Bayesian framework as

$$\begin{aligned} y_i^O|\beta,\eta &\sim \text{Weibull}(\beta,\eta) \quad T[\tau_i^L,] \\ y_j^C|\beta,\eta &\sim \text{Weibull}_{t_j^{\text{Lower}}}(\beta,\eta) \quad T[\tau_j^L,] \\ \beta,\eta &\sim \pi(\theta_{\beta,\eta}), \end{aligned}$$

where $\text{Weibull}_{t_j^{\text{Lower}}}$ indicates that the random variable y_j^C has a Weibull distribution and is constrained to be greater than the censoring time, and $T[\tau^L,]$ indicates that the distributions are re-normalised by the probability $P(y>\tau^L)$ (Note: I stole this notation from the stan code). For the moment, I express a joint prior for the parameters in its most general form.

Unknown truncation time A problem arises when the installation time of the left-truncated lifetimes is unknown, since to normalise the truncated lifetime distribution of the left-truncated observation τ^{Left} must be known. For example, if for the dataset shown in *Figure* 2.4 there was no information at all prior to t_{start}, then we could not use the likelihood in 2.9 or 2.10. This is the case for the idler frame data in. This problem is referred to either as unknown exposure history () or initial conditions (). In these case, there are two approaches that can be taken (Guo, 1993). The first, discard all of the left-truncated samples (), in which case the parameter estimates are still unbiased (). However, in doing so we throw away a large amount of information. In most cases of left-truncation and right

censoring, the right censoring masks any information about longer lifetimes and so the left-truncated samples are the only source of information about the upper tail of the lifetime distribution. The second approach is to assume a constant hazard, i.e. $\beta = 1$, since in this case the Weibull distribution reduces to the exponential and, no matter the age of a unit, the probability of it surviving a given period is constant (this is the memoryless trait of the exponential distribution). However, assuming a constant hazard is very restrictive and often one of the aims of performing lifetime analysis in the first place is to determine if $\beta > 1$. Furthermore, assuming an exponential distribution when the data do not have a constant hazard may lead to severe bias in the parameter estimates (**?**). In *Section* **??** I show how treating the unknown installation times as a case of censoring and using the method of imputing the censored data, the missing truncation times can also be imputed and reasonable parameter estimates can be obtained.

2.2 Imputing truncation times

Using the toy example in *Figure* 2.4, say we do not observe any of the installation or failure times to the left of t_{start}. In this case, we know that the first, partially, observed lifetime from each unit started some time between $t = 0$ and $t = t_{\text{start}}$. This is a case of interval censoring, where the lower censoring bound is the time from the beginning of observation to the failure time and the upper bound is from $t = 0$ to the failure time. If we did not know the origin time $t = 0$ with respect to $t = t_{\text{start}}$ then it would be a case of right censoring, but the following logic would still apply. Let t_i^{failure} be the failure time of the i^{th} observation. Treating the observations as interval censored, the left-truncated lifetime can be imputed as in *Section* **??** by sampling from

$$\hat{y}_i^L | \beta, \eta \text{ Weibull}^{t_i^{\text{failure}}}_{t_i^{\text{failure}} - t_{\text{start}}}(\beta, \eta). \tag{2.11}$$

Using the imputed values of the lifetime it is then possible to calculate the truncation time by

$$\tau_i^L = \hat{y}_i^L - \left(t_i^{\text{failure}} - t_{\text{start}}\right). \tag{2.12}$$

A complication arises when the lifetime is both interval censored by the start of the observation period and right censored by the end, such as unit one in *Figure* 2.4. In this case the truncation times value is unknown but is between $\tau^L = 0$ and $\tau^L = \min\left(t_{\text{start}}, \hat{y}_i^L - (t_{\text{end}} - t_{\text{start}})\right)$. Since we have no reason to expect that the truncation time is not uniform, we can impute the truncation time using the uniform distribution

$$\tau_i^L \text{ Uniform}(0, \min\left(t_{\text{start}}, \hat{y}_i^L - (t_{\text{end}} - t_{\text{start}})\right)). \tag{2.13}$$

Sampling the value of τ_i^L in this way incorporates the extra uncertainty in our Bayesian estimates

2.3 Analysis of simulated data

2.3.1 Simulation method

In order to simulate the data generating process of the idler frames, I sample $N \times M$ draws from a Weibull distribution with known shape parameter β, and scale parameter η. I then assign these lifetimes to M units. To calculate failure times rather than lifetimes, I take the cumulative sum of the N lifetimes assigned to each unit. Installation times are calculated by taking the lag of the failure times. I then define a start, t_{start}, and end, t_{end}, time for the observation window. Any lifetimes where both the install and failure times sit either before t_{start} or after t_{end} are discarded. Of the remaining lifetimes, if the install time is less than t_{start} then t_{start} is substituted for the install time and the lifetime is marked as interval censored, while if the failure time is greater than t_{end} then t_{end} is

Table 2.1: Example of simulated censored lifetime data.

unit	true lifetime	install time	failure time	censoring	observed lifetime
2	2397.8	1500.0	2397.8	interval	897.8
2	822.5	2397.8	2800.0	right	402.2
3	1461.5	1500.0	2237.7	interval	737.7
3	31.7	2237.7	2269.4	NA	31.7
3	51.5	2269.4	2320.9	NA	51.5
3	2970.0	2320.9	2800.0	right	479.1
4	1648.5	1500.0	1756.4	interval	256.4
4	1958.0	1756.4	2800.0	right	1043.6
5	741.7	1500.0	1876.5	interval	376.5
5	14.4	1876.5	1890.8	NA	14.4

substituted as the failure time and the lifetime is marked as right censored. If a lifetime is both interval censored at the beginning of the lifetime and right censored at the end, then the install time is set as t_{start}, the failure time as t_{end} and it is marked as right censored. *Figure* 2.5 shows a simulated idler lifetime data set when $\beta =$, $\eta =$, $t_{start} =$, and $t_{end} =$. *Table* 2.1 shows the first ten rows of the simulated censored data shown in *Figure* 2.5. Next we fit the censored weibull model to the simulated data by maximizing the likelihood in **??** and also by fitting the Bayesian model in **??** with noninformative priors using Hamiltonian Monte Carlo in STAN.

2.3.2 Bias in results

Do demonstrate the bias caused by heavy censoring of the lifetime data, I first demonstrate with the empirical kaplan-mayer approach and fitting the censored weibull model by maximizing the likelihood in **??**. When fitting the model via MLE I look at two cases; the first where all censored data are treated as right censored (i.e. if I didn't specify an upper bound of the interval censored data); and the second where the intercal censored data is treated properly by defining an upper bound for the lifetimes. This comparison is made purely for demonstration. I then fit the Bayesian equivalent to the MLE, which is the Weibull model with

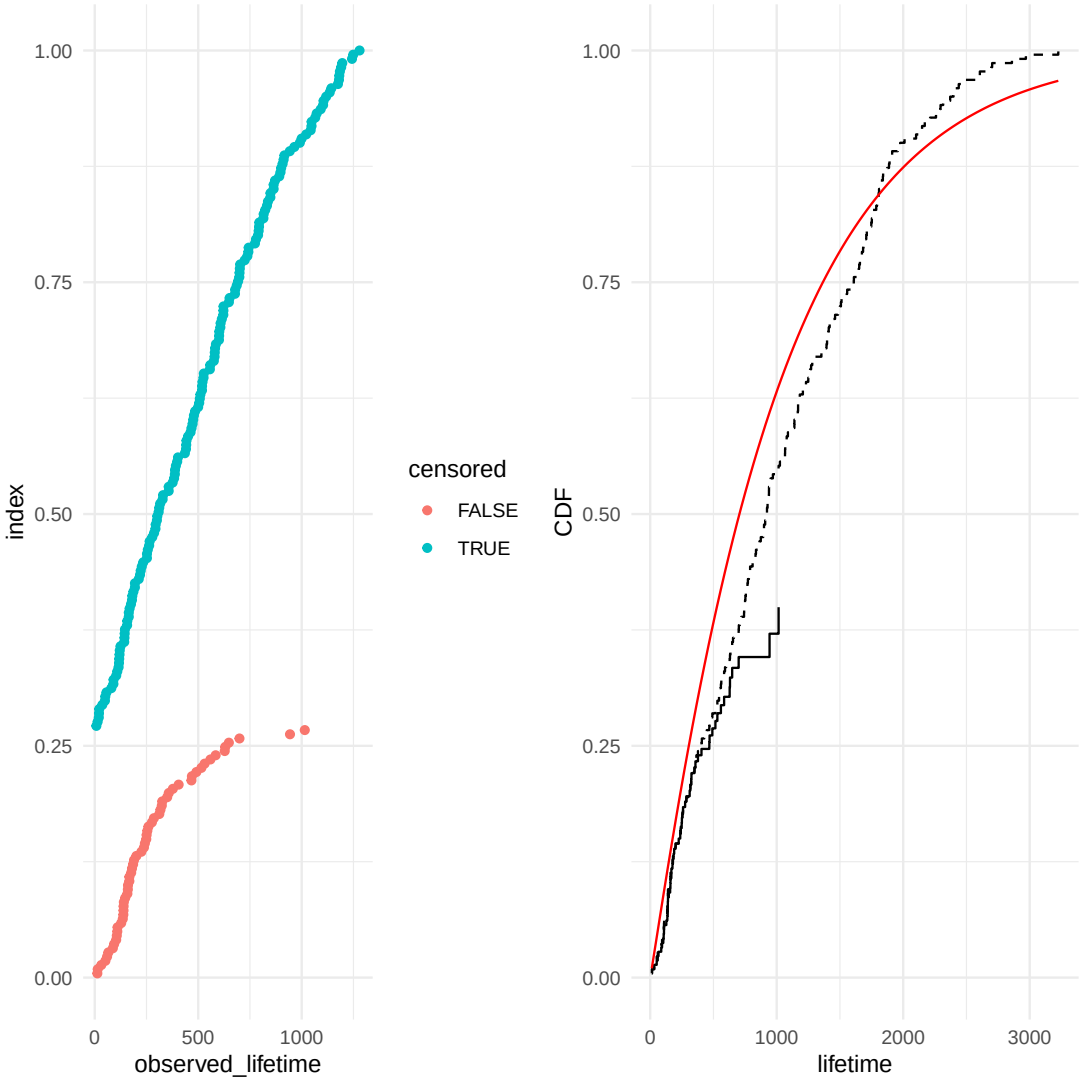

Figure 2.5: The simulated lifetimes separated by censored(both right and interval) vs noncensored, and ordered by length of observed lifetime.

non-informative independent priors on the parameters β and η. I fit both versions of the Bayesian Weibull model.

MLE - Say we are unsure about defining an upper bound for the interval censored lifetimes and we treat them as right censored; the upper bound is infinite. Th

- Fit the Weibull model w/ MLE and Bayesian non-informative independent priors.

- Plot the joint density of the parameters from the Bayesian posterior and overlay the true values and MLE estimate.

- Show the estimated CDF's with the true CDF and the "true" sample values (i.e. uncensored observations).

- Add small simulation to show that this is not a one off...

Noninformative Bayesian model The priors are sufficiently vague that they encode vary little information into the model. Therefore making the results roughly equivalent to the MLE.

2.4 Informative Bayesian analysis

How can informative priors help us in this case?

Independent Construction of independent priors.

Joint There is some work on bias correction:

But most focus on finding a way of correcting for bias (which is convoluted and difficult to communicate to reliability engineers), also the foundations of these approaches were laid in the late 90s/early 2000 and so the simulations are difficult to follow (Not the right wording but I mean in terms of presenting results and visualisations due to computational limitations).

One paper (Kaminskiy & Krivtsov, 2005) investigated using encoding an informative joint prior in a Bayesian framework in order to suplement analysis when data have small sample sizes and heavy censoring. This method is easy to communicate to reliability engineers and practitioners and reduces the bias demonstrated in **??**. Through the informative prior, informations can be encoded into the model to inform areas of the posterior where the data do not contain information. However, that is not to say that we are exploring a fictional scenario where we know the truth, but engineers can work closely with manufacturers to understand the upper tail of the distribution of the lifetimes and then encode it into models for their specific population of units in order to get more accurate reliability estimates.

Construction of the joint prior.

2.4.1 Effect of informative priors

Compare the estimated CDF for the three different models with the truth.

2.5 Discussion

...

Chapter 3

Simulation study

Outline on structure of chapter and what I wish to achieve in the simulation experiment.

3.1 Simulation structure

How do we structure the simulation experiment? What are we testing?

3.2 Results

Visualise/summarise the results of the simulation experiment.

3.3 Discussion

Typical results section.

3.4 Recommendations

What are our recommendations for handling heavily censored data? (which we will implement in the next chapter)

Part II

Part two: Degradation modelling

A preamble about which chapters have been published and which chapters came from industry placements.

Contribution statements for Noisy GP paper and Belt wear paper.

Chapter 4

Noisy gamma process for modelling degradation measurements with uncertainty

If there are very few or no failures observed for a particular component or asset then the lifetime methods that we have looked at in Chapters 2 and 3 are not very useful in reliability decision making. If there is some measure of the degradation process that drives failure, then degradation modelling can be used to forecast the degradation of units and inform reliability decision making. Gamma stochastic processes are a widely used degradation model for degradation that evolves monotonically (**?**). However, most degradation data collected in industrial settings is contaminated by noise, or error. This noise can be attributed to different sources, including measurement error, instrument noise, placement of sensors, and other environmental factors (**?**). Consequently, models for gamma processes must be extended to account for such noise.

In this chapter I

1. Show how the BHM framework can be used to simply extend the GP to incorporated noisy observations.

2. Show a reparameterisetion of the GP that simplifies prior elicitation and also makes it more obvious where covariates or radome effects can be included in the model.

3. Demonstrate through simulated data some identifiability issues that arise when the GP is extended to noisy observations and show present some methods for how to resolve these issues.

4.1 Gamma process

The gamma process is a type of stochastic jump process. It was introduced to the reliability domain by **?**, and since then has been used in many applications including the modelling of the corrosion of steel coatings, wear of brake pads, erosion of breakwaters, thinning of pressure vessels, and degradation of LED lights (**?**).

Consider a sequence $\{z_i\}$ of noise-free measurements of the degradation of a unit observed at times t_i, $i = 0, 1, 2 \ldots, I$. Without loss of generality, I assume that $z_0 = 0$ at $t_0 = 0$. A gamma process (**?**) models the jumps in degradation between measurements, $\Delta z_i = z_i - z_{i-1}$, as independent samples from a gamma distribution. Thus, we can write that

$$\Delta z_i | \eta(\cdot), \xi \sim \text{Ga}\left\{\eta(t_i) - \eta(t_{i-1}), \xi\right\}, \tag{4.1}$$

with rate ξ and shape $\eta(t_i) - \eta(t_{i-1})$, where $\eta(\cdot)$ is a given monotone increasing shape function. The simplest gamma process for modelling degradation is a stationary gamma process, which has a linear shape function (**?**), for example, $\eta(t_i) = \beta t_i$. Of course, nonlinear shape functions can be used; however, even when the degradation trace appears to be nonlinear, a time transformation can often be applied so that a stationary gamma process can be fitted. Therefore, in what follows we consider only the stationary gamma process. When using a

linear shape function, we can write eq. 4.1 more simply as

$$\Delta z_i|\beta,\xi \sim \text{Ga}\left(\beta\Delta t_i,\xi\right), \tag{4.2}$$

where $\Delta t_i = t_i - t_{i-1}$.

The gamma process described in eqs. 4.1 and 4.2 can be extended to describe situations commonly encountered in practice, namely, the need to account for measurement error and/or unit-to-unit variability when the degradation of several identical or similar units is being measured. I discuss measurement error next and defer discussing unit-to-unit variability until Chapter 5.

4.2 A Noisy Gamma process

In this section I give a background to noisy gamma process and describe how it's implementation can be simplified using the BHM framework introduced in 1.4. In an early paper, **?** fit a single parameter gamma process to noisy data by using the additive model $y_i = x_i + \epsilon_i$, where y_i represents the noisy observations, x_i represents the underlying gamma process, and ϵ_i is independent and identically distributed Gaussian noise. The gamma process is parameterised in terms of the mean wear rate (β/ξ). They then use the differences of the measured (noisy) jumps, $\Delta y_i = y_i - y_{i-1}$, to formulate the likelihood; consequently, the likelihood is determined by a convolution because the random variable $\Delta Y_i = \Delta X_i + \Delta E_i$ is the sum of the two random variables $\Delta X_i = X_i - X_{i-1}$ and $\Delta E_i = E_i - E_{i-1}$. In addition, calculating the difference of the errors leads to a dependence structure between the $\Delta\epsilon_i$. To carry out inference, **?** use simulation to approximate the likelihood. **?** extended their work by developing a faster method for approximating the likelihood using the Genz transform and a quasi-Monte Carlo method. Their method also allows both of the parameters of the gamma process, β and ξ, in 4.2 to be estimated.

Building on the work of **?** and **?**, **?** proposed a way to include degradation-

dependent measurement error. Other researchers focused on improving computational efficiency by alternative methods such as deconvolution (**?**) or by using faster algorithms to approximate the likelihood, for example, approximate Bayesian computing (**??**). Common to all of these works, however, is a convolution-based likelihood based on a *marginal* model that requires the evaluation of, or approximations to, a complicated multidimensional integral. By contrast, hierarchical modelling based on *conditional* models provides a more straightforward, tractable, and flexible alternative when it is combined with an efficient inferential method. We describe hierarchical modelling in a Bayesian framework in the next section, but first note in passing that **?** and **?** also formulate a conditional likelihood to model a complex noisy gamma process and use maximum likelihood estimation combined with an EM algorithm and particle filtering for estimation and inference.

To demonstrate how a noisy GP can be postulated under the BHM framework described in Section **??**, consider the noisy degradation trace in *Figure* **??**. Figure **??** shows a degradation trace generated from an stationary gamma process (the solid line) and the noisy observations of this degradation trace (the red points). Let, y_i refer to the measured, noisy, degradation data at time t_i, $i = 0, 1, 2, \ldots, I$ and, using the same notation as in Section 4.1, $\{z_i\}$ refer to the values of the underlying gamma degradation path at times t_i.

To specify the Bayesian hierarchical model, in the data model, I assume that *given the value of the underlying gamma process*, the noisy observations are normally distributed and independent of each other; in other words, the y_i are *conditionally* independent. That is

$$y_i|z_i, \sigma \sim \mathrm{N}(z_i, \sigma) \qquad \text{data model}$$

where σ is the standard deviation of the Gaussian distribution. I then assume in the next level of the model that the underlying degradation, the z_i, follow a gamma degradation process. As a consequence of the independence of the incre-

ments and eq. 4.2, $z_i = \sum_{j=0}^{i} \Delta z_j$ has a gamma distribution given by $\text{Ga}(\beta t_i, \xi)$. Therefore, we wright the process model as

$$z_i = \sum_{j=0}^{i} \Delta z_j$$

$$\Delta z_i | \beta, \xi \sim \text{Ga}(\beta \Delta t_i, \xi) \qquad \text{process model}$$

In the final level of the hierarchy I specify a distribution for the parameters β, ξ and σ, but for the moment, I write the distribution in its most general form, as the joint distribution

$$\beta, \xi, \sigma | \theta \sim \pi(\theta) \qquad \text{parameter model}$$

where $\pi(\theta)$ represents the parameters of the joint distribution. In the next section I show how a reparametrisation of the process model results in more interpretable parameters than the shape and rate and how this simplifies this last step of specifying the parameter model. Then in Section **??** I use simulation to choose suitable distributions for these parameters.

4.3 Reparametrisation

The gamma process described in eq. 4.2 has density function

$$f(z_j; \beta t_i, \xi) = \frac{\xi^{\beta t_i}}{\Gamma(\beta)} e^{-\xi x} z^{\beta t_i - 1}, \tag{4.3}$$

and the mean and variance, which I denote by μ and σ^2, are given by

$$\mu = \frac{\beta}{\xi} t_i \quad \text{and} \quad \sigma^2 = \frac{\beta}{\xi^2} t_i. \tag{4.4}$$

Both the average degradation rate and the variability of the gamma process depend on the parameters β and ξ. Hence, it is challenging to specify prior distri-

butions of β and ξ so as to separate their effects on the stochastic process. From the perspective of the user, it is desirable to reparameterise the gamma process so that the new parameters have clear interpretations and effects. In addition, if they are *orthogonal* (**?**), there are several desirable statistical consequences for estimation, inference, and computation.

One such parameterisation is in terms of the mean μ and coefficient of variation $\nu = \sigma/\mu = 1/\sqrt{\beta}$: the mean represents the average degradation rate per unit time, whereas the coefficient of variation describes the volatility of the degradation process; or how much heterogeneity there is in the wear rate over time. For the user, therefore, μ and ν have a more intuitive interpretation than the shape and the rate. Furthermore, using a result due to **?**, it is straightforward to show that these parameters are also orthogonal. (We note in passing that orthogonal parameterisations are not unique; the mean μ and shape β are also orthogonal (**?**).)

Substituting μ and ν in the expression for the distribution of the increments in the process model in Section 4.1 yields

$$\Delta z_i | \mu, \nu \sim \text{Ga}\left(\frac{\Delta t_i}{\nu^2}, \frac{1}{\mu\nu^2}\right). \tag{4.5}$$

I use this reparameterisation in the remainder of this thesis. **?** also use the shape and coefficient of variation, pointing out that it can be easier for a plant engineer to interpret them. They do not, however, exploit their orthogonality, preferring to fix the value of ν in their analysis instead of estimating it.

4.4 Constructing the prior

The prior distribution in the parameter model summarizes our beliefs about the parameters. There are two different ways in which this information is encoded: the choice of distribution, and the values of the hyperparameters. Before the advent of contemporary sampling algorithms, Bayesian analysis relied on conju-

gate prior distributions, or convenient prior distributions that facilitated the use of Gibbs samplers or conventional Metropolis-Hastings algorithms (**?**). However, with the development of more efficient sampling algorithms such as Hamiltonian Monte Carlo (**?**), we are no longer limited by such requirements and can select priors that reflect our state of knowledge, facilitate efficient computation, and that can be justified and evaluated in a principled way. In this section I...

In the degradation modelling literature, a gamma distribution is often used as the prior distribution for the rate parameter ξ of the gamma process (**?**) and also for the shape parameter (**?**). It is well known that a gamma prior on the rate parameter is conditionally conjugate(**?**), and its use leads to analytically tractable results, as **?** show. Nevertheless, little work has been done to assess whether other prior distributions might be more appropriate. The gamma distribution has a heavy tail, and its use can lead to MCMC chains that converge very slowly or that are highly autocorrelated; moreover, it can lead to physically implausible realizations of the gamma process, as we demonstrate below. Using the new parameterisation of the GP in terms of μ and ν, conditional conjugacy no longer exists and so there is even less motivation for a gamma prior.

In *Figure* **??**, I illustrate prior predictive checking of a noise-free gamma process using three sets of priors for its parameters: first, 'conventional' priors, $\mathrm{Ga}(1, 0.001)$ and $\mathrm{Ga}(0.001, 0.001)$, that are widely used in the literature for the both shape and rate parameters of the usual parameterization of a noise-free gamma process in eq. (4.2), and second, priors on μ and ν in the alternative parameterization of eq. (4.5) with carefully thought out weekly-informative priors. All three sets of priors yield an average degradation rate of 1 unit per unit time.

The distribution $\mathrm{Ga}(\epsilon, \tilde{\epsilon})$, where $\epsilon, \tilde{\epsilon} \longrightarrow 0$, is often used as a noninformative prior distribution, especially in mixed linear models, where it is a conditionally conjugate prior for the precision (**?**, p. 33). In addition, as we pointed out above, the gamma distribution is conditionally conjugate for the rate parameter: if $\{z_i\}$, $i = 1, 2, \ldots, n$, represents an independent sample from $\mathrm{Ga}(\beta, \xi)$, then the con-

ditional distribution of ξ given β and the data is $\text{Ga}(n\beta + \epsilon, \sum_{i=1}^{n} z_i + \tilde{\epsilon})$ when the prior distribution of ξ is $\text{Ga}(\epsilon, \tilde{\epsilon})$. Hence, when ϵ and $\tilde{\epsilon}$ are both small, the prior adds very little information, but it is noninformative *with respect to the rate parameter only*; furthermore, inferences about ξ may be sensitive to the values of ϵ and $\tilde{\epsilon}$ in data sets where small values of ξ may be possible (**?**, p. 130). When $\text{Ga}(\epsilon, \tilde{\epsilon})$ is used for *both* parameters, we can no longer assume that the joint prior will be noninformative and therefore must evaluate it to determine whether it is indeed diffuse. For further discussion on the consequences of using $\text{Ga}(\epsilon, \tilde{\epsilon})$ as a prior distribution and guidance on using more sensible alternatives, see **?** and **?**.

Figure **??**(a) and (b) show 100 draws from the prior predictive distribution of a noise-free gamma process when both the shape and rate parameters are assigned the prior distribution $\text{Ga}(1, 0.001)$ (Fig. **??**(a)) or $\text{Ga}(0.001, 0.001)$ (Fig. **??**(b)). In Fig. **??**(a), we can clearly see that the degradation traces resulting from a $\text{Ga}(1, 0.001)$ prior distribution are all nearly linear, without the jumps expected of gamma processes; furthermore, many of the rates of degradation are unrealistically high and unrealistically low. In Fig. **??**(b), where a $\text{Ga}(0.001, 0.001)$ prior is used, most of the prior predictive distribution has mass around implausibly low values of the average rate, and there is one unrealistically steep degradation trace. As we pointed out earlier, the gamma distribution is highly skewed and has heavy tails; consequently, depending on the values of the shape and rate, the prior can place mass on high, low, or both high and low values, resulting in simulated data that simply could not be observed in practice. By contrast, prior simulations generated according to the weekly-informative priors constructed with respect to the alternative parameters μ and ν in Fig. **??**(b) look much more plausible.

To specify independent prior distributions of the parameters μ and ν in the GP model in (4.5), we adopt the approach introduced by **?**: design priors that favour simpler models over more complex ones and that are consistent with domain knowledge. The mean μ controls the average degradation rate, similar to the action of the slope parameter in a linear degradation path model. We have no

reason to believe that the variability about the mean degradation rate would be asymmetric, so a Gaussian distribution with a small standard deviation is both appropriate and convenient

$$\mu \sim \mathrm{N}(1, 0.5).$$

The coefficient of variation ν is a measure of the volatility of the degradation process, and although we might expect some heterogeneity in the wear rate as degradation progresses, we do not expect the wear rate to be extremely volatile. Hence, we use a truncated Student t-distribution with 3 degrees of freedom as a prior for ν;

$$\nu \sim t_3^+(0, 1),$$

where the superscript $(+)$ denotes a truncated distribution whose lower bound is zero; furthermore, I use the location-scale form of a t distribution with n degrees of freedom, written as $t_n(\text{location}, \text{scale})$. This prior places a large mass near zero but still allows the posterior distribution to move away from zero. In addition, it has lighter tails than a gamma distribution, and consequently does not give too much weight to extremely volatile degradation paths. Figure **??**(c) shows 100 draws from this parameter model for the reparameterised gamma process. The degradation traces have the appearance of paths expected from a gamma process, that is, there are discrete jumps between time points, in contrast to Fig. **??**(a), where all the traces are straight lines. Furthermore, more than half the degradation values at the end (eleventh time point) are between 6 and 16, as would be expected when the degradation varies around one. Finally, although there are some extreme realizations, there are only one or two that are completely implausible.

To fully specify the model, I also need to specify a prior for the standard deviation of the measurement error, σ. Following the recommendations of **?**, Chapter 17, I use a vague Uniform$(0, A)$ prior for σ, where A is chosen to be large relative to the expected scale of σ. We use such a vague prior for demonstration

purposes. However, in practice, an analyst should have a reasonable grasp of the scale of the measurement error and should be able to specify a weakly informative prior, we do exactly this in Section 4.5.4. Because our initial prior on σ is so vague, we do not include the measurement error in the prior predictive checking in Fig. **??**.

...

4.5 Fitting the noisy GP

To improve our understanding of the noisy GP model I fit the Bayesian hierarchical model outlined in Section **??** with the priors defined in Section **??** to the single simulated degradation trace in *Figure* **??** as well as to a subset of the simulated degradation measurements. For clarity, the full model is

$$
\begin{aligned}
y_i | z_i, \sigma &\sim \mathrm{N}(z_i, \sigma) && \text{data model} \\
z_i &= \sum_{j=0}^{i} \Delta z_j && \text{process model} \\
\Delta z_i | \mu, \nu &\sim \mathrm{Ga}\left(\frac{\Delta t_i}{\nu^2}, \frac{1}{\mu \nu^2}\right) \\
\mu &\sim \mathrm{N}^{+}(10, 10) && \text{parameter model} \\
\nu &\sim t_2^{+}(0, 1) \\
\sigma &\sim \mathrm{Unif}(0, 100).
\end{aligned}
$$

The single path example shows that the noisy GP is more difficult to fit than a noise-free GP model: when the sample size is small, the model struggles to separate the parameters describing the variance of the measurement error and the volatility of the underlying gamma process because there is not enough information in the data to do so. To demonstrate this problem with identifiability, I fit the BHM of two data sets: one 'large' data set consisting all 20 simulated noisy degradation measurements in *Figure* **??**, and another 'small' data set that

is a subset of 10 points. I fit the BHM of the noisy GP outlined in Sections **??** and **??** to these two data sets and evaluate how well the true parameter values and underlying degradation path is reclaimed in the two resulting posterior distributions. I also investigate the efficiency of the No-U-Turn sampler for the two cases.

4.5.1 Data simulation

To generate the degradation trace in *Figure* **??**, which I coin the 'large' dataset, by sampling 20 time increments from a Unif$(0.8, 1.3)$ distribution. Then next, I sampled 20 jumps in degradation from $\text{Ga}(\Delta t_i/\nu^2, 1/\mu\nu^2)$, using $\mu = 10$ and $\nu = 1.119$. I then calculated the cumulative sum of the jumps to obtain the underlying, noise-free degradation trace z_i, where $z_0 = 0$ at $t_0 = 0$. Finlay I add Gaussian noise with standard deviation $\sigma = 4$ to the underlying degradation path to get the noisy observations. To create the second smaller data set, I randomly select ten of the twenty noisy observations; The degradation observations selected for the small data set are shown in red in Fig. **??**.

Tables of big and small data sets.

4.5.2 Computation

To sample from the posteriors of the noisy GP model conditioned on the two different datasets I use the No-U-Turn sampler implemented in the probabilistic programming language *Stan* (**?**). I generated $88,000$ samples from each posterior distribution using four chains of $25,000$ iterations each, with a burn-in of $3,000$ iterations and no thinning. To ensure a detailed exploration of the posterior, I also change the sampling parameters *adapt delta* and *maximum tree depth* to 0.99 and 13 respectively. Raising *adapt delta* results in a more aggressive (smaller) choice of the step size used for the leapfrog algorithm that approximates the hamiltonian trajectories and raising the *maximum tree depth* allows each leapfrog algorithm to run for longer. Increasing these two sampling parameters results in a slower

sampler but ensures a more detailed exploration of the posterior. All of the code to define the model in stan, simulate the data in R, and sample from the posterior using RStan is available in XXXX.

During sampling–despite increasing *adapt delta* and *maximum tree depth*–80 divergent transitions occur when fitting the model to the small data set, whereas only four occur when sampling from the larger data set. Further more, there are signs that the sampler for the small data is having trouble efficiently exploring the posterior. *Figure* **??** shows the chain energies for the ...

4.5.3 Results and diagnostics

My objective here is to investigate how the size of the dataset affects inference from the BHM for a noisy GP. To do so I assess how well the parameters and underlying degradation trace are reclaimed in the two posterior distributions. Visualising the two posteriors shows that when the model is fitted to all twenty degradation observation it is able to recover the parameter values and underlying degradation path; when only a subset of ten noisy observations is used, the model fails to do so because it is unable to disentangle the observation noise from the volatility of the gamma process. I come to this conclusion by exploring the degenerate behaviours in the posterior that are flagged by divergent trajectories that occur during sampling.

marginal densities *Figure* **??** shows the marginal distributions of the parameters μ, ν, and σ conditioned on the small and big data sets as well as the true values of the parameters. For each marginal density the median and 66% and 95% credible intervals are shown. It is clear that when the model is fit to the small data set it fails to reclaim the true parameter values, but when fit to the bigger dataset, it successfully reclaims the true values. The marginal posterior densities of the the parameters conditioned on all twenty degradation observations are centred around the true values of the parameters whereas the marginal

posteriors of σ and ν conditioned on the subset of ten observations are centred around fifteen and zero respectively. Furthermore, the marginal posterior of σ conditioned on the smaller subset of the data appear to have some multimodality. To understand the implication of these parameter estimates as well as the effect of how they covary with one another in the posterior I look at their joint effect on the outcome variables, which in this case is the predictive distribution of the filtered degradation path.

posterior predictive density *Figure* **??** shows the posterior predictive distribution of the underlying degradation trace, the z_i, for the two posteriors with the true degradation trace and noisy observations overlaid. The thick grey line in each plot is the median of the posterior predictive distribution; additional quantiles are shown in different shades of blue. Clearly in *Figure* **??** A the model has been able to reclaim the underlying degradation from the noisy degradation observations when fit to all twenty observations: the median path follows the actual path almost exactly, with uncertainty bands that are narrow enough to be useful. However, as was the case with parameter values, the median path derived from the posterior distribution conditioned on the subset of the data has not recovered the true path (*Figure* **??** B). In *Figure* **??** B, the median path is a nearly straight line through the data points. In addition, the uncertainty intervals are much wider.

pairs Clearly there are some issues occurring in the posterior distribution of the model conditioned on the smaller subset of the data. This was pre-eluded to by the divergent trajectories that occurred during sampling(section **??**). In the case of fitting the model to simulated data, where we can be sure that the model is properly specified and implemented, poorly behaved sampling is often a sign of a deeper issue with the model. As discussed in Section 1.4, the divergent transitions can point to the problematic areas in the posterior. *Figure* **??** shows a pairs plot of the parameters μ, ν, and σ and the first degradation jump, Δz_1. The

divergent trajectories are shown in red. In the bivariate scatter plots there are strong funnel shapes between μ and $log(\nu)$ as well as between $log(\nu)$ and the first degradation jump. The divergent trajectories are concentrated at the entrance to these funnel, suggesting that these are the cause of the sampling issues. The funnel shapes occur because as ν shrinks towards zero μ and Δz_1 approach very particular values; $\mu = 10$ and $\Delta z_1 = 10 \times \Delta t_1$.

parallel coordinate plot The divergent trajectories can help to further explore how the degenerate behaviour manifests in the multidimensional posterior. *Figure* **??** shows a parallel coordinate plot of the posterior draws for all of the parameters in the model. The divergent trajectories are highlighted red. The divergences draw a clear structure through parameter space. They all pass through the values $\mu = 10$, $\nu = 0$, and $\Delta z_i = 10 \times \Delta t_i$ and have an inflated value of σ that is much larger than the true value $\sigma = 4$. This structure equates to a linear degradation trace with large uncertainty. To emphasise this point, in *Figure* **??** I plot the posterior predictive distribution of the underlying degradation path from the small data set and overlay the divergent trajectories. Showing that the areas of tight curvature in the posterior occur around the models where the degradation trace is effectively linear.

comparison of the two posteriors In comparison, this degenerate behaviour in the posterior is almost completely washed out by extra information in the large dataset. *Figure* **??** shows the joint distributions of the intermediate quantities Δz_{15} from the model conditioned on the big dataset and Δz_9 for the small with $\log(\nu)$ and σ. In the two datasets, Δz_{15} and Δz_9 are the same jump in degradation. In the joint posterior of Δz_9 and $\log(\nu)$ there is the deep funnel shape around $\Delta z_9 = 9$ and $\log(\nu) = -\infty$ and there is a second mode in the joint distribution of Δz_9 and σ. However, in the joint distribution of Δz_{15} and $\log(\nu)$ there is very little mass around $\Delta z_{15} = 9$ and no second mode in the joint distribution of Δz_{15} and σ. Leading to the conclusion that the nonidentifiability only exists

when there is few observations.

4.5.4 Solutions to computational issues

The identifiability issue for the small data set can also be solved by injecting more information into the analysis that helps to disentangle σ and ν. This can be in the form of supplementary data or prior information that inform one of the nonidentifiable parameters. It would typically be much easier to get extra information about the measurement error. This information could be in the form of domain expertise or supplementary experimental data. Here I show that adding this a small amount of supplementary information using either of these two approaches helps to identify σ and therefore ν resulting in much smoother geometries in the posterior and therefore much more efficient sampling. The results on the inference is arguably better than if we fit the model to all twenty degradation observations.

Prior information In Section 4.5.3, I have used an effectively non-informative prior for the standard deviation of the measurement error. Typically a technician would have some understanding of the variability in the measurement process. To emulate this, I place a Gaussian prior on the standard deviation of the measurement error

$$\sigma \sim \mathrm{N}(4, 1).$$

This prior is centred around the true value of σ and places 95% of the mass between $\sigma = 2$ and $\sigma = 6$. Sampling from the posterior of this model with this stronger prior conditioned on the small dataset is much quicker and no divergent transitions occur. *Figure* **??** shows the chain energies for the sampler...

The pairs plots of the MCMC samples in *Figure* **??** look much smoother and there is little evidence of the funnel shaped degeneracies between $\log \nu$ and μ and Δz_1. In *Figure* **??** I compare the marginal distribution for σ, μ, and ν with the true values and the models fit in section 4.5.3. The marginal distributions for

the model with the stronger prior are much smoother and there is now no mass around zero in the posterior of ν.

Supplementary data As an alternative, I sample five supplementary observations of the measurement noise from the distribution

$$y_{\text{sup}} \sim \text{N}(0, 4).$$

Extra supplementary observations such as these could be obtained by taking multiple measurements at time $t = 0$, when the degradation is known to be zero; just before decommissioning the component, after which detailed non-noisy measurements can be obtained; or performing some kind of small experiment. The supplementary observations can be easily incorporated into the Hierarchical model through the data model

$$\begin{aligned} y_i|z_i, \sigma &\sim \text{N}(z_i, \sigma) && \text{data model} \\ y_{\text{sup}} &\sim \text{N}(0, \sigma). \end{aligned}$$

Similar to the more informative prior, the sampler is much more efficient, no divervgencies occur during sampling, and the bivariate posterior distributions in the pairs plots look much smoother that. The resulting marginal posterior distributions of σ, μ, and ν for the supplementary data are also compared in *Figure* **??**. The marginal distributions of the parameters are very much the same as when a stronger prior is used and the model clearly successfully reclaims the true parameter values.

4.6 Discussion

The main objective of this chapter has been to show that using the Bayesian hierarchical formalism allows us to frame a model for a noisy gamma stochastic

process in a tractable and transparent manner. Decomposing the noisy gamma process into a sequence of conditional models—the data, process, and parameter models—removes the need for complex deconvolutions that require the evaluation of, or approximations to, multidimensional integrals. Making this connection that allows the simple extension of the GP to noisy degradation observation and showing its implementation in the contemporary Bayesian computational environment Stan, is a step towards making stochastic degradation models applicable to use in industry and also accessible to practitioners. Bellow, I summarise the main element of this chapter and highlight the contributions.

Reparameterising the gamma process in terms of the mean μ and coefficient of variation ν results in more interpretable parameters than the shape β and rate ξ: μ is the mean wear rate, and ν is the inverse of the 'signal-to-noise' ratio, and hence is a measure of the volatility of the gamma process. The interpretability of μ and ν simplifies specifying prior distributions because they are easier to elicit domain information about. Additionally, reparameterising the gamma process in this way also helps clarify how extensions of the model, which I show in the next chapter, such as unit-to-unit variability or covariates, can be incorporated into the model. Finally, the parameters μ and ν are orthogonal, which has desirable computational benefits.

I've re-assessed good choice of priors under this new parameterisation and shown a pricipled way of assessing these priors.

Clear pathological behaviour in the posterior of the small dataset.

Some signs of this in the poor behaviors in the bigger dataset but much better behaved.

Discussions of pre-asymptotic identifiability.

When more extra information is added into the analysis that informs one of the poor behaving parameters the computational and inferential problems are solved.

Lead into next chapter on using data from multiple units.

The from the small dataset; this 'oversmoothing' leads us to conclude that the two modes of the marginal distributions of σ and ν are associated—when ν approaches zero and the path from the gamma process becomes nearly linear with only very small jumps, σ is overestimated to compensate for the unaccounted-for volatility of the degradation path.

Chapter 5

Noisy gamma process with unit-to-unit variability

Unit-to-unit variability description.

Introduce crack growth data and adding noise.

Prior predictive checks for crack growth.

5.1 Models for multiple units

Commentary on random effects...

complete pooling model

varying mu model

varying nu model

varying mu and nu model

5.2 Computation and posteriors

5.2.1 Model comparison

Having fitted several models to the crack-growth data in Fig. **??**, the next step is to check how well they describe the data generating process and to compare them. To do so, we evaluate their ability to predict new observations. In the absence of an independent, external test set, it is conventional to use *information criteria* to compare models. These criteria, such as AIC, DIC, and others, are used to seek a compromise between goodness-of-fit and model complexity and to assess out-of-sample prediction accuracy. AIC and DIC are easy to calculate, but they are not fully Bayesian; hence, criteria such as WAIC (Watanabe-Akaike Information Criterion) and leave-one-out cross-validation (LOO-CV) are to be preferred (**?**).

In comparing the models below, we use LOO-CV, where the measure of distributional predictive accuracy is the *log score*. The log score is the log likelihood of a new observation $\tilde{y}_i$ given the posterior distribution of the parameters. It is also the posterior predictive density, and it can be written as,

$$\text{lpd} = \log \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta = \log p(\tilde{y}_i|y), \tag{5.1}$$

where θ is the set of parameters and y is the observed data. The parameters can also include unobserved latent variables; in models for the noisy gamma process, the latent variables represent the underlying gamma process. This measure in eq. (5.1) is called the log posterior density (lpd). If we observe multiple new data points $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_I)$, this can be dealt with in a point-wise fashion using the log point-wise posterior density (lppd),

$$\text{lppd} = \sum_{i=1}^{I} \log p(\tilde{y}_i|y). \tag{5.2}$$

In cases where each of the new observations are independent of one another

given the parameters and latent variables, which is the case here, the pointwise predictive density is equal to the joint predictive density of the set of new observations; $\log p(\tilde{y}|y) = \sum_{i=1}^{I} \log p(\tilde{y}_i|y)$. (**??**) and (**??**) are defined for a given set of new observations, but the new unobserved data points $\tilde{y}_i$ arises from the true data generating process and so are a random variable with distribution

$$\tilde{Y}_i = f(\tilde{y}_i). \tag{5.3}$$

Hence, a better measure of predicted accuracy is the expectation of the lppd, the elppd, which is obtained by integrating over $\tilde{Y}_i$

$$\text{elppd} = \sum_{i=1}^{I} \int \log p(\tilde{y}_i|y) f(\tilde{y}_i) d\tilde{y}_i. \tag{5.4}$$

In the context of Bayesian models fit with MCMC, the computed elppd can be calculated by averaging over the $s = 1, ..., S$ MCMC draws from the posterior,

$$\text{computed elppd} = \sum_{i=1}^{I} \int \log \frac{1}{S} \sum_{s=1}^{S} p(\tilde{y}_i|\theta^s) f(\tilde{y}_i) d\tilde{y}_i. \tag{5.5}$$

Although this would be the best measure of predictive accuracy for our Bayesian models, we obviously do not know the true data generating process and so cannot define $f(\tilde{y}_i)$. We can, however, approximate the expectation above by using cross-validation whereby we iteratively withhold a portion of the observed data, sample from the posterior conditioned on the wrest of the data, and then calculate the log likelihood of the withheld portion of the data given the samples from the posterior. The simplest form of cross-validation is leave one out (loo), where we withhold each observation,

$$\text{elppd}_{loo} = \sum_{i=1}^{I} \log \frac{1}{S} \sum_{s=1}^{S} p(y_i| \, [\theta]^s_{-[i]}). \tag{5.6}$$

$[\theta]^s_{-[i]}$ is the posterior draws for the set of parameters and latent variables conditioned on all the observed data except the withheld observation y_i.

In hierarchical models, the definition of a new observation and the likelihood of those observations depends on what aspect of the model's predictive performance we are trying to assess. For example, in the crack growth problem, we could obtain new observations for the same units at the same observation times, $\tilde{y}_{n,i}|z_{n,i},\sigma$ (although this case is a bit unrealistic), new observations for an observed unit at some time in the future, $\tilde{y}_{n,I+1}|\tilde{z}_{n,I+1},\sigma$, or we could observe an entirely new unit, $\tilde{y}_{n+1}|\tilde{z}_{n+1},\sigma$, where $\tilde{y}_{n+1} = [\tilde{y}_{n+1,1},...,\tilde{y}_{n+1,I}]$. Here we focus on the last two of these cases; $\tilde{y}_{n,i+1}|\tilde{z}_{n,i+1},\sigma$, since we would like to compare the models on their ability to forecast the future degradation for each of the units being tested (step ahead prediction); and $\tilde{y}_{n+1}|\tilde{z}_{n+1},\sigma$ to compare the models on their ability to predict completely new, unseen, units (leave one unit out). In both cases, the likelihood of the observations conditional on the draws from the posterior predictive distribution are the same, since we assume in our data model that any noisy observations are independent and normally distributed given the underlying filtered degradation and the standard deviation of the measurement error, that is,

$$p(\tilde{y}_{n,i}|\tilde{z}_{n,i},\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{\tilde{y}_{n,i} - \tilde{z}_{n,i}}{\sigma}\right)^2, \tag{5.7}$$

however, the predictive distributions of the $\tilde{z}_{n,i}$ may be different in the two cases depending on the model's hierarchical structure.

Step ahead prediction For the case of step ahead prediction, the $\text{elppd}_{\text{step ahead}}$ can be estimated using leave one out,

$$\text{elppd}_{\text{step ahead}} = \sum_{n=1}^{N} log \frac{1}{S} \sum_{s=1}^{S} p(y_{n,I}|\,[\tilde{z}_{n,I},\sigma]^s_{-[n,I]}), \tag{5.8}$$

where $[\tilde{z}_{n,I},\sigma]^s_{-[n,I]}$ are the draws from the posterior distribution of σ and posterior predictive distribution of $\tilde{z}_{n,I}$ conditioned on all data except the withheld observation $y_{n,I}$. The draw from the posterior predictive distribution of $\tilde{z}_{n,I}$ is calculated by sampling a value of predicted jump in degradation $\Delta\tilde{z}^s_{n,I}|\,[\mu_n,\nu_n]^s_{-[n,I]}\,,(t_I-$

t_{I-1}) according to the process model (Note that this will be different depending on which hierarchical structure we are using... for the complete pooling model we would use μ and ν whereas for the partial pooling models we wold use μ_n or ν_n when applicable). Then, we add the sampled jump to the draw of the previous filtered degradation measurement for unit n to get a draw from the posterior predictive distribution of $\tilde{z}_{n,I}$;

$$\tilde{z}^s_{n,I} = z^s_{n,I-1} + \Delta\tilde{z}^s_{n,I}. \tag{5.9}$$

The $\text{elppd}_{\text{step ahead}}$ for the several different models are displayed in Table-

! Add Table of results.

Leave on unit out The same can be done for an entirely new unit, except the likelihood for all the observations from a new unit is

$$p(\tilde{y}_{N+1}|\tilde{z}_{N+1}, \sigma) = \prod_{i=1}^{I} p(\tilde{y}_{N+1,i}|\tilde{z}_{N+1,i}, \sigma). \tag{5.10}$$

Using cross validation, iteratively withholding each unit, we can estimate the $\text{elppd}_{\text{new unit}}$ via

$$\text{elppd}_{\text{new unit}} = \sum_{n=1}^{N}\sum_{i=1}^{I} log\frac{1}{S}\sum_{s=1}^{S} p(y_{n,i}|\,[\tilde{z}_{n,i}, \sigma]^s_{-[n]}). \tag{5.11}$$

Note that for each iteration of the leave one unit out the posterior is conditioned on the dataset with **all** the observations from unit n being withheld. In this case, draws from the posterior predictive distribution of $\tilde{z}_{n,i}$ are obtained by sampling I jumps in degradation from the data model conditional on the hyper parameters, $\Delta\tilde{z}_n|\theta^s_{-[n]}$, and then taking the cumulative sum of the jumps to get the $\tilde{z}_n$; $\tilde{z}_n = cumsum(\Delta\tilde{z}_n)$. The draws of the hyper parameters for the new unit, $\theta^s_{-[n]}$, are either sampled from the hierarchical prior using the draws of the hyperparameters, i.e. $\tilde{\mu}^s_n|[\mu_\mu, \sigma_\mu]^s_{-[n]}$ and $\tilde{\nu}^s_n|[\nu_\nu, \sigma_\nu]^s_{-[n]}$, or taken to be the posterior draws of the parameters $\mu^s_{-[n]}$ and $\nu^s_{-[n]}$ depending on whether or not the

parameter is partially pooled (the former case) or completely pooled (the latter case). Table- shows the estimated $\text{elppd}_{\text{new unit}}$ for the several different models using the leave one out procedure.

! Add Table of results.

Both of the comparisons above involve repeatedly re-fitting the model to different subsets of the data, which is computationally inefficient. A much more efficient method is to approximate the elppd such as in **?**, however, in hierarchical modeling cases, where the "left out" portions of the data are nested and as the size of the nested portions increases, the approximations are less likely to work well (**?**). Hence, why we choose to use the full cross validation scheme and incur the computational overhead.

$\hat{(}$WAIC) is not an approximation like PSIS-LOO is ...

5.3 Failure time distributions

5.4 Discussion

- Discuss how the sampling gets stuck around the smaller model, just like in the single path with very few measurements.

Chapter 6

Conveyor belt wear forecasting

6.1 Background

6.1.1 FDA

6.2 Process models

6.2.1 Linear

6.2.2 Noisy GP

Chapter 7

Conveyor belt wear forecast with spatial random effect

Chapter 8

Discussion

8.1 Tie together discussion

8.2 Strengths and limitations

8.3 Future directions

8.4 Industry practitioner implications

Appendices

Appendix A

Appendix Title

Appendix B

Copyright Information

References

Bae, S. J., Kuo, W., & Kvam, P. H. (2007). Degradation models and implied lifetime distributions. *Reliability Engineering & System Safety*, *92*(5), 601–608. Retrieved from https://www.sciencedirect.com/science/article/pii/S0951832006000536 doi: https://doi.org/10.1016/j.ress.2006.02.002

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In K. M. Hanson & R. N. Silver (Eds.), *Maximum entropy and bayesian methods* (pp. 15–22). Dordrecht: Springer Netherlands.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv*. Retrieved from https://arxiv.org/abs/1701.02434 doi: 10.48550/ARXIV.1701.02434

Betancourt, M. (2020, June). *Identity Crisis.* Retrieved from https://betanalpha.github.io/assets/case_studies/identifiability.html (Accessed: 31-05-2023)

Creagh, B. (2020). *Bridgestone: A partner in products, service and technology.* https://www.australianmining.com.au/wp-content/uploads/2020/04/bridgestone-Material-size-web-1024x768.jpg.

Cressie, N. A. C., & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data.* Hoboken, N.J: Wiley.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2020). *Bayesian data analysis, third edition.* Boca Raton, FL: Taylor & Francis.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be

understood in the context of the likelihood. *Entropy*, *19*(10). Retrieved from https://www.mdpi.com/1099-4300/19/10/555 doi: 10.3390/e19100555

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ... Modrák, M. (2020, November). Bayesian Workflow. *arXiv:2011.01808 [stat]*. Retrieved 2022-03-02, from http://arxiv.org/abs/2011.01808 (arXiv: 2011.01808)

Geraerds, W. (1985). The cost of downtime for maintenance: preliminary considerations. *Maint. Manage. Int.*, *5*(1), 13–21.

Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). *Markov chain monte carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC.

Guo, G. (1993). Event-history analysis for left-truncated data. *Sociological Methodology*, *23*, 217–243. Retrieved 2024-05-01, from http://www.jstor.org/stable/271011

Hamada, M. S., Wilson, A. G., Reese, C. S., & Martz, H. F. (2008). *Bayesian reliability*. Springer New York, NY.

Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art , challenges and opportunities. *Mechanical Systems and Signal Processing*, *23*, 724–739. doi: 10.1016/j.ymssp.2008.06.009

Hong, Y., Meeker, W. Q., & McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *The Annals of Applied Statistics*, *3*(2), 857–879. Retrieved from https://doi.org/10.1214/00-AOAS231 doi: 10.1214/00-AOAS231

ISO. (2013). *ISO 12489:2013 petroleum, petrochemical and natural gas industries — reliability modelling and calculation of safety systems*. ISO International Standard. Retrieved from https://www.iso.org/obp/ui/en/#iso:std:iso:tr:12489:ed-1:v1:en:sec:3.1.8

ISO. (2016). *ISO 14224:2016 petroleum, petrochemical and natural gas industries — collection and exchange of reliability and maintenance data for equip-*

ment. ISO International Standard. Retrieved from https://www.iso.org/obp/ui/en/#iso:std:iso:14224:ed-3:v2:en

Jardine, A. K., & Tsang, A. H. (2013). *Maintenance, replacement, and reliability: Theory and applications (2nd edition).* CRC Press.

Kaminskiy, M., & Krivtsov, V. (2005, 01). A simple procedure for bayesian estimation of the weibull distribution. *IEEE Transactions on Reliability*, *54*, 612-616.

Kundu, D., & Mitra, D. (2016). Bayesian inference of weibull distribution based on left truncated and right censored data. *Computational Statistics & Data Analysis*, *99*. doi: https://doi.org/10.1016/j.csda.2016.01.001

Lawless, J., & Crowder, M. (2004, 09). Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, *10*, 213–227.

Lee, M.-L. T., & Whitmore, G. (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary [Article]. *Statistical Science*, *21*(4), 501 - 513. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-34249306163&doi=10.1214%2f088342306000000330&partnerID=40&md5=b05c2d1629e569781a4083581c9ebe7e (Cited by: 215; All Open Access, Bronze Open Access, Green Open Access) doi: 10.1214/088342306000000330

Lu, C., Meeker, W., & Escobar, L. (1996, 01). A comparison of degradation and failure-time methods for estimating a time-to-failure distribution. *Statistica Sinica*, *6*.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The bugs book: A practical introduction to bayesian analysis.* Boca Raton, FL: CRC Press.

Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2022). *Statistical methods for reliability data.* John Wiley & Sons.

Mitra, D. (2013). *Likelihood inference for left truncated and right censored lifetime data* (PhD thesis). McMaster University, Hamilton, ON, Canada. (Available at https://macsphere.mcmaster.ca/handle/11375/12738)

Mittman, E. T. (2018). *Applications of bayesian hierarchical models in gene expression and product reliability* (PhD thesis). Iowa State University, Ames, IA, United States. (Available at https://dr.lib.iastate.edu/entities/publication/7c167cf8-a345-476f-a77d-4fc829e4b2c9)

Moore, D. F. (2016). *Applied survival analysis using r* (R. Gentleman, K. Hornik, & G. Parmigiani, Eds.). Springer Cham.

Pandey, M., & Yuan, X.-X. (2006, January). A comparison of probabilistic models of deterioration for life cycle management of structures. In (pp. 735–746). doi: 10.1007/1-4020-4891-2_62

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing (dsc 2003)* (pp. 1–10). Vienna.

Reich, B., & Ghosh, S. (2019). *Bayesian statistical methods*. Boca Raton, FL: CRC Press.

Robinson, M. E., & Crowder, M. J. (2000, 12). Bayesian methods for a growth-curve degradation model with repeated measures. *Lifetime Data Analysis*, *6*.

SAP SE. (2023). *Sap software suite*. https://www.sap.com.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation - a review on the statistical data driven approaches. *European Journal of Operational Research*, *213*(1), 1-14. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221710007903 doi: https://doi.org/10.1016/j.ejor.2010.11.018

Stan Development Team. (2022). Stan Modeling Language Users Guide and Reference Manual [Computer software manual]. https://mc-stan.org.

Stan Development Team. (2024). Stan User's Guide, Version 2.34 [Computer software manual]. Retrieved from https://mc-stan.org/docs/2_34/stan-users-guide/truncation-censoring.html#censored-data (Chapter: Censored Data)

Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019). *Spatio-Temporal Statistics with R*. New York: Chapman and Hall/CRC. doi: 10.1201/9781351769723

Ye, Z.-S., & Xie, M. (2015). Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry*, *31*(1), 16-32. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2063 doi: https://doi.org/10.1002/asmb.2063