

**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

August 2024

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

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.

The first part of the works at methods for analysing failure time data that is subject to right censoring and left truncation with unknown exposure history. I show how this incompleteness of lifetime data can arise from the repeated replacements of a set of units. I propose a method for imputing partially observed left-truncated lifetimes through a Bayesian analysis. I also show how an informative joint prior for the two Weibull parameters can be carefully constructed to supplement the analysis in the case of censoring and truncation. I evaluate the methods using simulation and demonstrate on industry dataset of idler-frame replacements. Using the output of the Bayesian analysis, I show how the failure times of idler-frames still in operation as well as the expected number of fail-

ures are easily obtained, and how inference about the parameters of the Weibull distribution can be used to inform the design of a fixed time replacement strategy.

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 models for multiple units 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 clarify how the model can incorporate unit-to-unit variability (i.e. random effects). 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.

Thesis Publications

1. **Leadbetter, R.**, Phatak, A., Polpo, A. & Hodkiewicz, M. (2021) *Informative Bayesian Survival Analysis to Handle Heavy Censoring in Lifetime Data*, International Conference on Maintenance and Intelligent Asset Management (ICMIAM), Ballarat, Australia, pp. 1-6,
doi: 10.1109/ICMIAM54662.2021.9715184.
2. **Leadbetter, R.**, González Cáceres, G., & Phatak, A. (2024). *Bayesian hierarchical modelling of noisy gamma processes: Model formulation, identifiability, model fitting, and extensions to unit-to-unit variability*.
3. **Leadbetter, R.** & Phatak, A. (2024). *Functional degradation modelling of the wearing surface of conveyor belting using Bayesian hierarchical modelling and Gamma processes*.

Statement of Contributors

The candidate, Ryan Leadbetter, was responsible for all aspects of the research presented within this thesis, conceptualisation, data analysis, coding, interpretation, reporting of results and submission of manuscripts. The supervisory team and co-authors also contributed to the research conceptualisation, and some aspects of analysis, interpretation, writing and editing (Appendix A).

Ryan k. Leadbetter

Date

I, as co-author, endorse that this level of contribution indicated by the candidate indicated above is appropriate.

| | | | |
|---|-----------------------------------|-----------|------|
| 1 | Dr. Aloke Phatak | signature | Date |
| 2 | Professor Melinda Hodkiewicz | signature | Date |
| 3 | Associate Professor Adriano Polpo | signature | Date |
| 4 | Gabriel González Cáceres | signature | Date |

Acknowledgements

- First and foremost Aroke!
- Co-supervisory team Michael Small, Chris Aldrich, and Darfiana Nur (and my informal co-supervisors Melinda and Adriano)
- Very importantly Natasha!
- The people in industry who helped me (If I can put their names); Chris Rowlands and his team, Julian Bradley, Mitrin & Batina, and Saf & Hutch.
- The other students in the centre (such a great culture to be a part of), particularly the 201 office (Sandy, Gabriel, YingYing, Ponpot, and Srimali).
- The small Bayesian community at curtin Arash, Shih Ching, Suman, and Zhanglong
- My family.
- Jody.

Acknowledgement of Country

In the spirit of reconciliation, I acknowledge and pay my deepest respect to the Traditional Owners of the land on which this research was conducted, the Whadjuk people of the Noongar Nation, whose ancestral lands Curtin University is built upon. I also wish to acknowledge the Traditional Owners of the lands I have been fortunate enough to travel on during my PhD, and their enduring connections to land, sea, and community. I pay my respects to their Elders, past and present. This always was, and always will be, Aboriginal land.

Copyright Statement

I have obtained permission from the copyright owners to use any of my own published work (e.g., journal articles) in which the copyright is held by another party (i.e., publisher, co-author).

Contents

| | |
|---|-------------|
| Abstract | vii |
| Thesis Publications | ix |
| Statement of Contributors | xi |
| Acknowledgements | xiii |
| Acknowledgement of Country | xv |
| Copyright Statement | xvii |
| 1 Introduction | 1 |
| 1.1 Maintenance decision making | 3 |
| 1.2 Reliability modelling | 5 |
| 1.3 Industry Examples | 8 |
| 1.4 Bayesian reliability modelling | 13 |
| 1.5 Structure of this thesis | 23 |
| I Part one: lifetime analysis | 25 |
| 2 Weibull analysis of partially observed lifetime data | 29 |
| 2.1 Background | 32 |

| | | |
|----------|---|-----------|
| 2.1.1 | Lifetime distribution | 32 |
| 2.1.2 | The Weibull distribution | 34 |
| 2.1.3 | Censoring | 35 |
| 2.1.4 | Left-truncation | 39 |
| 2.1.5 | Left-truncation and right-censoring | 40 |
| | Unknown exposure history | 43 |
| 2.2 | Imputing unknown exposure histories | 44 |
| 2.3 | Informative joint prior | 47 |
| 2.4 | Analysis of simulated data | 50 |
| 2.4.1 | Simulation method | 52 |
| 2.4.2 | Weakly informative prior | 54 |
| 2.4.3 | Strongly informative prior | 57 |
| 2.5 | Simulation experiments | 59 |
| 2.5.1 | Factor levels | 60 |
| 2.5.2 | Accuracy measures | 61 |
| 2.5.3 | Results | 62 |
| 2.6 | Concluding remarks | 66 |
| 3 | Idler-frame case study | 71 |
| 3.1 | Idler-frame lifetime data | 72 |
| 3.2 | An informative prior | 74 |
| 3.3 | Posterior draws | 75 |
| 3.4 | Using the posterior | 77 |
| 3.4.1 | Failure time of units in operation | 78 |
| 3.4.2 | Expected number of failures | 79 |
| 3.4.3 | Cost functions and preventative policy design | 81 |

| | |
|--|------------|
| 3.5 Discussion and conclusions | 90 |
| II Part two: Degradation modelling | 93 |
| 4 A noisy gamma process for modelling degradation measurements with uncertainty | 97 |
| 4.1 The gamma process | 98 |
| 4.2 A noisy gamma process | 99 |
| 4.3 Reparametrisation | 102 |
| 4.4 Constructing the prior | 103 |
| 4.5 Fitting the noisy gamma process | 108 |
| 4.5.1 Data simulation | 109 |
| 4.5.2 Computation | 110 |
| 4.5.3 Results and diagnostics | 114 |
| 4.5.4 Solutions to computational issues | 122 |
| 4.6 Discussion | 125 |
| 5 Noisy gamma process with unit-to-unit variability | 129 |
| 5.1 Models for multiple units | 132 |
| 5.1.1 The complete pooling model | 134 |
| 5.1.2 Partial pooling models | 136 |
| varying μ model | 137 |
| varying ν model | 137 |
| varying μ and ν model | 137 |
| 5.2 Computation, posteriors, and predictive distributions | 138 |
| Complete pooling | 139 |
| Varying μ | 140 |

| | |
|---|------------|
| Varying ν | 143 |
| varying μ and ν | 146 |
| Posterior predictive checks | 147 |
| 5.2.1 Model comparison | 150 |
| Leave-one-unit-out cross-validation | 150 |
| Step-ahead cross-validation | 151 |
| 5.3 Failure time distributions | 153 |
| 5.4 Discussions | 155 |
| 6 Conveyor belt wear forecasting | 159 |
| 6.1 Functional data analysis (the data model) | 164 |
| B-splines | 165 |
| Functional time series | 165 |
| 6.2 Process models | 167 |
| 6.2.1 Noisy gamma process | 168 |
| 6.2.2 Linear model | 168 |
| 6.3 Parameter model | 169 |
| 6.4 Posterior sampling and inference | 173 |
| Marginal posteriors of parameters | 174 |
| Intermediate quantities | 177 |
| Posterior predictive distributions | 178 |
| 6.5 Forecasting degradation curves | 181 |
| 6.6 Comparison of methods | 182 |
| 6.7 Failure time distributions | 187 |
| 6.8 Discussions | 190 |

| | |
|---|------------|
| 7 Thesis discussion | 195 |
| 7.1 Overview and main discussions | 196 |
| 7.2 Future directions | 199 |
| 7.3 Implications for industry practitioners | 201 |
| Appendices | 203 |
| A Authorship Agreement | 205 |
| References | 207 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | An annotated image of an overland iron ore conveyor (Creagh, 2020) showing the belting, idlers, and idler frames. | 9 |
| 1.2 | (a) frame lifetimes. (b) empirical cumulative distribution function (eCDF) of lifetimes | 11 |
| 1.3 | Belt UT measurement data. | 12 |
| 2.1 | The Weibull hazard function when $\beta = 0.9$, $\beta = 1$, or $\beta = 1.1$ and $\eta = 1$ | 36 |
| 2.2 | An example of the difference between left censoring in a dataset (a) and left-truncation (b). In (a) the lifetimes that fail before the red line at age = 0.5 are left censored since their existence is known but not their exact failure time. In (b) any lifetimes that fail before the red line at age = 0.5 are absent from the dataset completely and therefore the likelihood of the observed data needs to be adjusted to reflect this. | 37 |
| 2.3 | An example of a dataset with left-truncated samples. Units are put under test at random times, and their failure times are recorded. However, the unit is only captured in the dataset if it fails after t_{start} ; therefore, any units that were installed and failed before t_{start} are not included in the sample. | 40 |

| | |
|--|----|
| 2.7 The draws from the joint posteriors conditioned on the simulated dataset when the left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed and a weakly informative prior is used. (d), (e), and (f) show the corresponding uncertainty around the CDF (in the form of the 0.5 and 0.8 uncertain intervals) that result from (a), (b), and (c), respectively. The true parameter values and CDF are shown in red. | 57 |
| 2.8 The joint posteriors conditioned on the simulated dataset when the left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed and an informative joint prior is used that encodes information into the upper tail of the lifetime distribution. (d), (e), and (f) show the corresponding uncertainty around the CDF (in the form of the 0.5 and 0.8 uncertain intervals) that result from (a), (b), and (c), respectively. The true parameter values and CDF are shown in red. | 59 |
| 2.9 Comparison of the marginal prior and posterior for F_{t_1} (top) and F_{t_2} (bottom) when left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed to show how both the elicited distributions have been updated in the posterior. | 60 |
| 2.10 The simulation results are summarised by the mean of the one hundred repetitions for each factor combination. The columns show the three levels of $t_{end} - t_{start}$, and the rows show the levels of t_{start} and N . The Bayesian p -values of β and η are reported on the horizontal and vertical axis, respectively, and the colours of the points show the treatment of the left-truncated samples, and the shape shows the prior. The size of the points correspond the the average elpdd_{100} calculated according to eq. (2.5.2) | 64 |

| | |
|--|----|
| 2.11 The elppd ₁₀₀ simulation results summarised by the mean of the one hundred repetitions for each factor combination. The columns show the different levels of $t_{end} - t_{start}$ and rows the different levels of t_{start} . The number of units N is shown on the horizontal axis, and the average elppd ₁₀₀ score is shown on the vertical axis. The colours of the points show the treatment of the left-truncated samples and the shape shows the prior. | 67 |
| 3.1 The idler-frame lifetimes plotted along the length of the conveyor. On the horizontal axis is the frame number that the lifetimes belong to, and on the vertical is the log value of the lifetime in log-days. The fully observed lifetimes are red points, while the partially observed (censored) lifetimes are blue. The censored lifetimes that are left-truncated by the start of the observation period are shown as triangular points. | 75 |
| 3.2 The joint informative prior from eliciting information at $t_1 = 1825$ and $t_2 = 2920$. (a) shows 3000 draws from the informative joint prior, and (b) shows the resulting uncertainty surrounding the CDF. | 76 |
| 3.3 The joint draws of the Weibull parameters β and η from the posterior distribution. (a) shows the plain draws, and (b) compares the draws with the contours of the informative joint prior. | 77 |
| 3.4 The resulting posterior uncertainty surrounding the Weibull CDF. | 78 |
| 3.5 The remaining useful life distributions for the current lifetimes (the lifetimes right censored by the end of observation). The left plot shows frames 1 to 72, and the right shows frames 73 through 143. | 80 |
| 3.6 Ten draws of the cumulative failures proceeding the end of the observation period. The horizontal axis is the days since the end of the observation period, and the vertical axis is the cumulative number of failures. | 81 |

| | |
|--|-----|
| 3.7 The predictive distributions for the number of failures in the one (a), three (b), and six (c) month period after the end of the observation period. | 82 |
| 3.8 The cost per unit time of preventative maintenance policies for idlers. On the horizontal axis is the cost per unit time, and on the vertical is the fixed time replacement interval. The point and intervals show the median and 0.5 and 0.9 intervals. The grey densities show the full distributions. | 88 |
| 4.1 A simulated degradation trace (black) generated from a gamma process with parameters $\mu = 10$ and $\nu = 1.119$ and the simulated noisy observations (red) of the trace. | 104 |
| 4.2 One hundred realizations from prior predictive distributions of a noise-free gamma process with mean degradation rate of 1 unit per unit time and different prior distributions for parameters: (a) shape, rate $\sim \text{Ga}(1, 0.001)$, (b) shape, rate $\sim \text{Ga}(0.001, 0.001)$, and (c) parameterization mean/coefficient of variation—see text for details. | 105 |
| 4.3 The simulated degradation trace with the subset of observations selected for the small dataset highlighted in Red. | 110 |
| 4.4 The cain energy diagnostics for each chain when sampling from the posterior of the big data set (left) and small data set. Each plot compares the marginal energy distribution, π_E , with the first differenced distribution, $\pi_{\Delta E}$ | 113 |

| | |
|--|-----|
| 4.5 The marginal posterior distributions of the parameters σ , μ , and ν when the BHM for the noisy gamma process is fitted to the ‘large’ and ‘small’ simulated data in Fig. 4.1. The points and intervals shown in each distribution represent, respectively, the median and 95% and 66% credible intervals. The values used to simulate the data are shown in the top row. | 115 |
| 4.6 The posterior predictive distribution of the filtered degradation path compared to the true degradation path when (a) the model is fitted using the big and (b) small data set. If the distribution of π_E is much wider than that of $\pi_{\Delta E}$, then sampling is inefficient. . . | 116 |
| 4.7 Pairs plot showing the MCMC draws from the posterior distribution of the parameters σ , μ , $\log \nu$, and the filtered value Δz_1 when the BHM of the noisy gamma process is fitted to the small dataset. The red points indicate divergences, which congregate at the end of the funnel in the pairwise plots of $\log \nu/\Delta z_1$ and $\log \nu/\mu$. . . | 118 |
| 4.8 The parallel coordinate plot of the draws from the posterior conditioned on the small dataset with the divergent traces plotted in red. | 119 |
| 4.9 The divergent traces overlaid (in red) on the posterior predictive distribution of the degradation path for the small data set. . . . | 120 |
| 4.10 The contours of the posterior joint distribution of Δz_{15} and $\log(\nu)$ (a) and Δz_{15} and σ (c) from the posterior when fit to all twenty observations and Δz_9 and $\log(\nu)$ (a) and Δz_9 and σ (c) from the posterior when fit to only ten observations. The divergent transitions are overlaid in red and show the areas in the posterior that are difficult to sample from. | 121 |
| 4.11 The chain energy diagnostics when the model is fit with a stronger prior on σ | 123 |

| | |
|---|-----|
| 4.12 The pairs plot of the MCMC draws when using a stronger prior on σ | 124 |
| 4.13 The marginal posterior distributions of the parameters σ , μ , and ν when information is added into the analysis either through a stronger prior or supplementary data compared to the original two posteriors. | 125 |
| 5.1 Crack-growth propagation data of Rodríguez-Picón, Rodríguez-Picón, Méndez-González, Rodríguez-Borbón, and Alvarado-Iniesta (2018): original data (solid lines); noisy data (dashed lines). | 131 |
| 5.2 Four non-noisy prior predictive simulations from the complete pooling model. Each simulation contains the same number of units and observations as the crack growth data in Fig. 5.1. | 135 |
| 5.3 Marginal posterior distributions of the underlying gamma process from a BHM where all parameters are completely pooled. | 140 |
| 5.4 Marginal posterior distributions of parameters from the model where mean degradation rate μ_j varies from unit-to-unit. | 142 |
| 5.5 A pairs plot of the posterior samples of μ_1 , μ_2 , μ_μ , and $\log(\sigma_\mu)$ from the varying μ model. In each bivariate plot, divergencies are plotted in red. | 144 |
| 5.6 Parallel coordinate plot for the parameters and hyper parameters of the varying μ model. The divergent traces are plotted in red. | 145 |
| 5.7 Marginal posterior distributions of the underlying gamma process from a BHM where mean degradation rate μ_j varies from unit-to-unit. | 145 |
| 5.8 Parallel coordinate plot for the parameters and hyper parameters of the varying ν model. The divergent traces are plotted in red. | 147 |

| | |
|--|-----|
| 5.9 Posterior predictive simulations from each of the four models. Each simulated dataset, is indicated by a different colour and line type. Plot (a) shows three simulated datasets from the fitted complete pooling model, (b) from the model with varying μ , (c) from the varying ν model, and (d) from the model where both μ and ν vary from unit-to-unit. | 149 |
| 5.10 Posterior distributions of the failure time distributions from the complete pooling model (left) and the varying μ model (right). . . | 155 |
| 5.11 Posterior distributions of the predictive failure time distributions from the complete pooling model (left) and the varying μ model (right) for Unit 3. | 156 |
| 6.1 The growing wear profile of a conveyor's belt over time. Ultrasonic thickness measurements are taken at $N = 20$ measurement locations across the width of the belt at repeated times. The vertical axis is wear in mm , calculated by subtracting the measured thickness from the original thickness of the belt. The horizontal dashed line at $25mm$ of wear indicates the maximum allowable wear before the belt needs to be replaced, i.e. the soft failure threshold. The colour gradient indicates the time the belt has been in operation. Measurements taken at the beginning of the belt's life are shown in light yellow, and the most recent set of measurements are plotted in dark red | 161 |

| | |
|--|-----|
| 6.2 A demonstration of some typical approaches used to trend belt wear. (a) and (b) demonstrate the method of (Webb, Sikorska, Khan, & Hodkiewicz, 2020), where the measurements at the fastest wearing location which are circled in (a) are trended into the future in figure (b). (c) and (d) demonstrate an alternative but similar approach used by some reliability software, where the maximum wear measurement in each profile is trended instead. | 162 |
| 6.3 (a) shows a fitted wear profile for the fifth observation; (b) shows the weighted set of basis functions that make up the profile in (a); and (c) shows the unweighted set of basis functions | 166 |
| 6.4 Two historic belt wear data sets from the same conveyor | 170 |
| 6.5 The estimated average wear rate for each spline coefficients calculated from the datasets in Fig. 6.4. The colour of each fit corresponds to the basis functions in Fig. 6.3 (c). The horizontal scale of each plots is in tonnes and the vertical scale is the value of the spline coefficient. | 171 |
| 6.6 Sixteen prior predictive simulations of the average wear profile generated from the linear path (dashed line) and gamma process (solid line) process models with the priors defined in Sec. 6.3. The values of the hyper parameters for ν in the gamma process that were used to generate each simulation are displayed on each plot. | 172 |
| 6.7 Marginal posterior distributions of the gamma process model parameters. | 174 |
| 6.8 Marginal posterior distributions of the linear general path model parameters. | 176 |
| 6.9 The posterior draws of the filtered spline coefficients for (a) the gamma process model and (b) the linear general path model. | 178 |

Chapter 1

Introduction

The work presented in this thesis is part of an industry-linked PhD under the Centre for Transforming Maintenance Through Data Science (CTMTDS), a centre 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—how and when they should maintain an asset based on their understanding of the asset, the available data, and what this data tells them about the current condition of the asset. As part of the industry-linked PhD, I spent 900 hours working on industry placement projects between two different industry partners to define research topics that are not only novel in an academic sense but also useful for applied problems in reliability that are faced by practitioners when making maintenance decisions in the mining and mineral processing industry.

From my placement time, it was apparent that there was a disconnect between the reliability modelling literature and the methods used by reliability practitioners in the mining and mineral processing industry—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 avail-

able in the form of infield reliability data, 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). However, these in-field data are typically noisy, messy, incomplete and uncertain as a result of 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 sensible estimates of lifetime distributions when lifetime data are heavily censored and truncated due to the pre-emptive replacements of assets and the way in which data is recorded, and 2) forecasting complex degradation processes with noisy and sparsely observed condition monitoring data. In this thesis, I develop novel extensions 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 the basis 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 from Bayesian analysis allows us to easily produce estimates and uncertainty intervals of 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, 2008a; 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 manually. However, the recent increase in popularity of Bayesian methods is due 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 from 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 (Gelman, Vehtari, et al., 2020). Throughout this thesis, I 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 how to carry out and report the results of Bayesian analyses for reliability and maintenance problems in the field.

The remainder of this chapter provides a background to the rest of the thesis. First, in Sec. 1.1, I provide some context around maintenance decision-making in the mining and mineral processing industry. Then, in Sec. 1.2, I give a high-level overview of reliability modelling and how it informs maintenance decisions. Sec. 1.4 outlines Bayesian methods and the key components of the Bayesian model-building workflow, which will be a strong thread throughout the remainder of the thesis. Finally, in Sec. 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 it 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 allocate resources and plan maintenance schedules ahead of time. There are three general strategies: reactive, preventative, and predictive maintenance (Jardine & Tsang, 2013). I provide a more detailed overview of these strategies below, but first, note that 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 reliability or condition monitoring data that can be used to refine/inform the maintenance strategy. For instance, Chap. 3 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 replacement strategy is when components are replaced pre-emptively after a designated period of time or operation. This proactive approach to maintenance is suit-

able for cheap components whose reliability decreases with time, i.e., components that wear out (most mechanical components). A disadvantage of preventative maintenance is that it can result in overmanning assets (replacing components too frequently when they still have remaining useful life), which is a waste of money and resources. If a component is costly and critical, and it is possible to monitor its condition, 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 that the choice balances the cost of maintenance with the cost of unplanned failure, 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 than one where the manufacturer’s default recommendations are used. Condition-based strategies, on the other hand, are more useful if we can forecast the degradation through time to predict the failure time (or remaining 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.

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, $\Pr [T > t]$, where T is the time of failure. 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 given time interval (Hamada et al., 2008a). 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, $R(t)$. Reliability can alternatively be expressed as the complement of the reliability function, $\Pr [T \leq t]$, which is 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 observation. In these cases, the data are not very informative of the parameters in a lifetime model. 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 To use 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 that drives the failure, then degradation analysis allows us to incorporate how close the other ninety-eight units were to failing into the analysis 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; C. Lu, Meeker, & Escobar, 1996; Meeker et al., 2022), and the connection is typically made using, not catastrophic failure, but soft failure.

Soft failure is defined by a predetermined threshold of degradation, compared to hard failures, which occur when the component can no longer operate (Hamada et al., 2008a). Hard failures are more uncertain in nature, since the level of degradation is only indicative of when the component will catastrophically fail—e.g. the catastrophic failure of a steel beam with a crack in it is both a function of the cracks length and the random load applied to the beam—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 and stochastic process models (Pandey & Yuan, 2006; Si, Wang, Hu, & Zhou, 2011). General path models assign a functional and/or empirical form to the degradation path of a unit, typically a theoretically motivated function such as in Robinson and 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, as in Lawless and Crowder (2004). A vague comparison of the two methods is made in Ye and Xie (2015); however, no comprehensive comparisons have been explored.

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 Fig. 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



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

frames are shown in Fig. 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 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 on all conveyors on a mine site. Therefore, a preventative maintenance strategy is used.

The survival data of idler frames I use in this thesis, derived from their installation and replacement times, can be used to inform the preventative maintenance strategy. Unfortunately, the failure data is 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. The data set is shown in Fig. 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, it was pre-emptively replaced, or it was in operation before we started reliably capturing failure and installation data. These partially observed lifetimes are censored lifetimes and are shown in blue in Fig. 1.2 (a). I discuss censoring in more detail in Part I of the thesis. 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

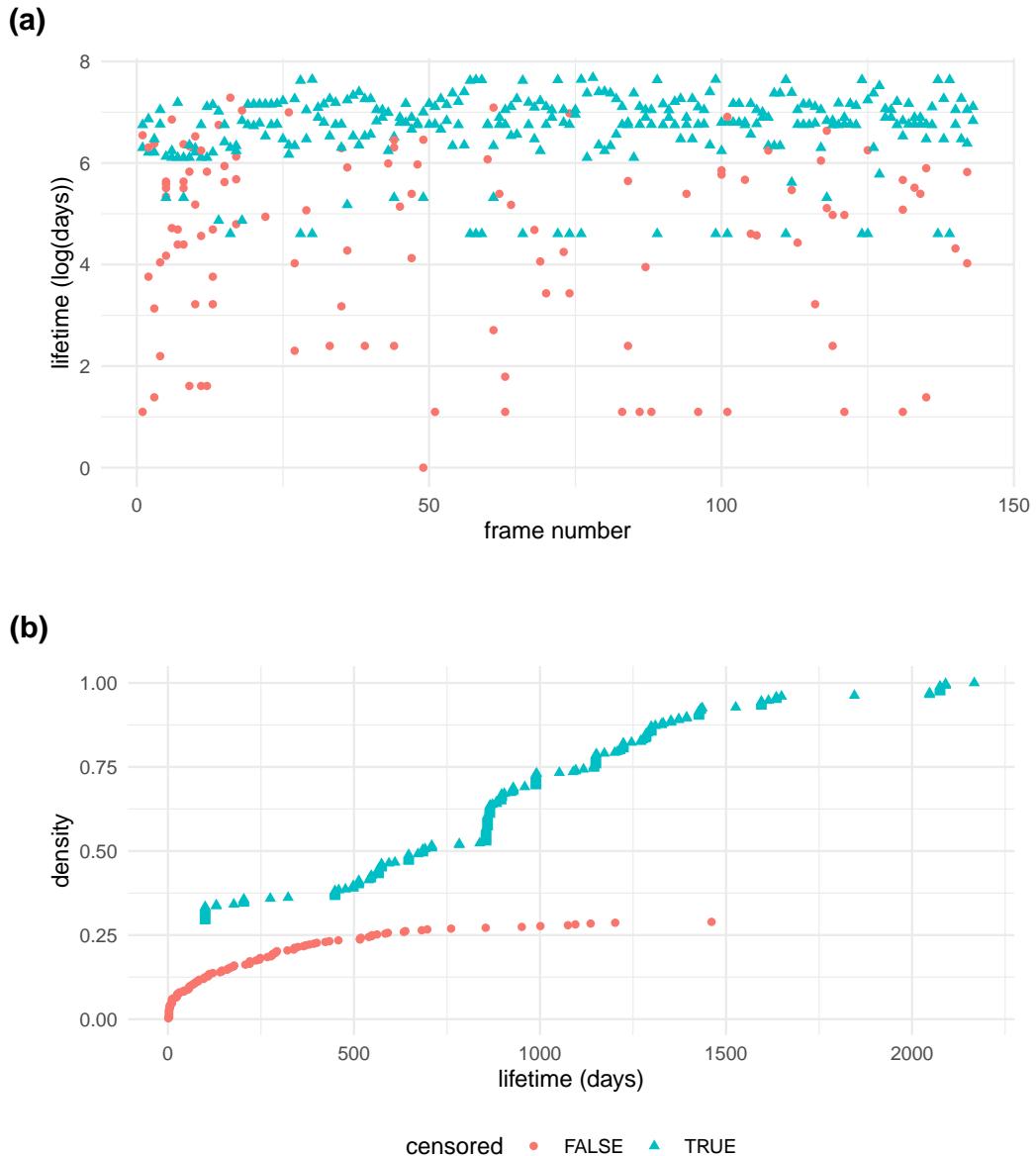


Figure 1.2: (a) frame lifetimes. (b) empirical cumulative distribution function (eCDF) of lifetimes

to the risk of being damaged by the ore. To assess 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 Fig. 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 replacement decisions requires principled statistical modelling to properly quantify the many different sources of uncertainty, which can be done in the Bayesian framework.

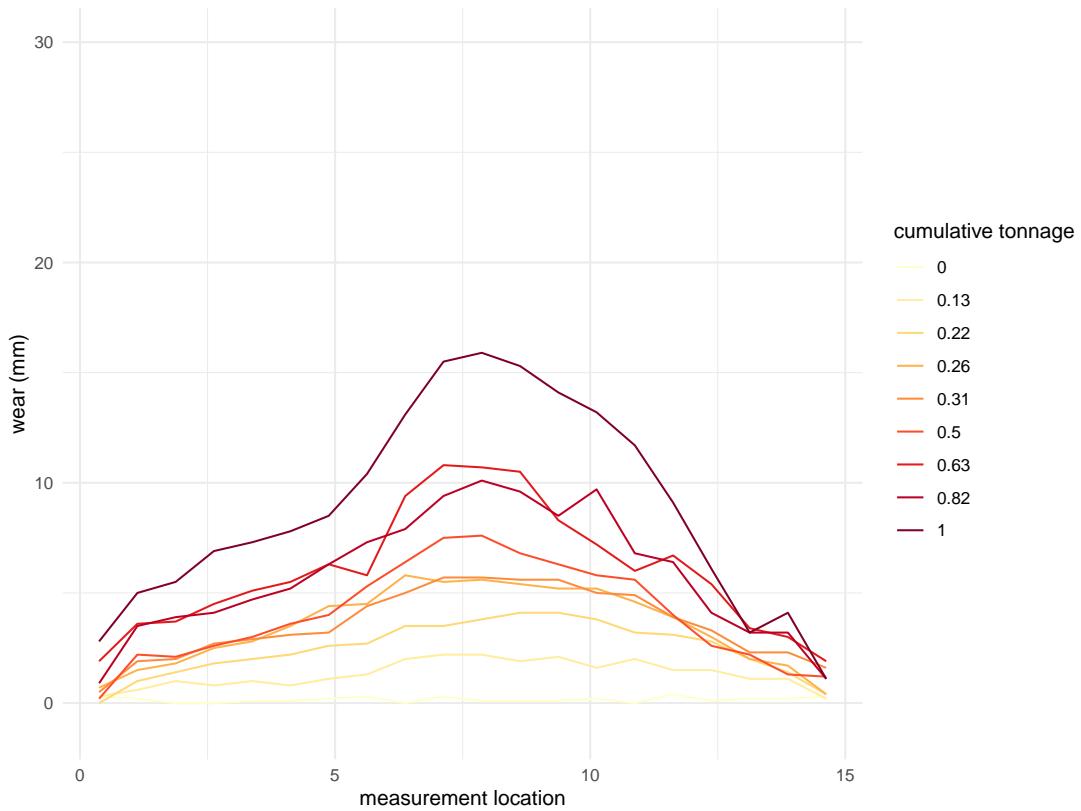


Figure 1.3: Belt UT measurement data.

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,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad (1.1)$$

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

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (1.2)$$

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 methods to perform inference. There are several powerful and flexible probabilistic programming languages, such as Stan (Stan Development Team, 2022), 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 compo-

nents 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 of the model components. 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 sequence 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;

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

The first level is the data model, $p(\text{data}|\text{process, parameter})$, 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 comprise 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 formulate 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 a model with a proper prior before conditioning on the observed data and is one of the key steps in the ‘Bayesian workflow’ (Gelman, Vehtari, et al., 2020, 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 Gabry, Simpson, Vehtari, Betancourt, and Gelman (2019b), priors that when combined with the likelihood lead to simulated data that could be plausibly observed are known as *weakly informative priors*. According to Gabry et al. (2019b), 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 Gabry, Simpson, Vehtari, Betancourt, and Gelman (2019a) 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* [my 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 can understand what our model is capable of learning from the data. For example, in Chapter 4 we explore

the limitation of a noisy gamma process model for noisy degradation data when only a few degradation measurements are observed.

HMC and diagnostics Throughout this thesis, I use the No-U-Turn sampler (Hoffman & Gelman, 2014) 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 (Neal, 2011). 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 sparing the user from the difficult task of choosing the step size and number of steps used to approximate the Hamiltonian trajectories (Hoffman & Gelman, 2014). 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); the reader interested in the specifics should refer to Betancourt (2015).

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} (Gelman & Rubin, 1992), 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 why 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 (Gabry et al., 2019b). While it is most often not possible to check for this condition mathematically, it can be checked numerically during sampling. 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 races off to infinity. Using a threshold energy, above which trajectories are considered divergent trajectories, we can diagnose problematic areas in the posterior (Gabry et al., 2019b), referred to by some as degeneracies (Betancourt, 2020). Sometimes, degeneracies—and the poor sampling that results—can be resolved by re-parametrising the model (Betancourt & Girolami, 2015) while in other cases they 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 (Gabry et al., 2019b). Two useful visual diagnostics are the bivariate scatter plot, also called a pairs plot, and the parallel coordinate plot, see (Gabry et al., 2019b) for examples and descriptions. Since divergent transitions are flagged using a threshold, some reported divergences may be false positives. If this is the case, their distributions should match that of the non-divergent samples in either type of plot. 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 use both pairs plots and parallel coordinate plots for checking sampling but only present them if they add to the discussion of the thesis (Chapters 4 and 5).

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 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,

$$\theta_s \sim p(\theta|y). \quad (1.3)$$

Using the posterior draws of the parameters, we can find not only 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 (Gelman, Carlin, et al., 2020) is a visual method for evaluating if the model has fit the data. After obtaining the posterior distribution, we can simulate replications of the data conditioned on the observed data. The predictive distribution for which these replicated datasets arise is the posterior predictive distribution. If the model fits the data, then the replications under the fitted model should look similar to the observed data (Gelman, Carlin, et al., 2020, pp. 143). Put another way, the observed data should be plausible under the posterior predictive distribution. The posterior predictive distribution can be expressed as

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \quad (1.4)$$

where θ is the set of parameters and latent variables in the model. Computationally, we obtain draws from $p(\tilde{y}|y)$ by simulating from the data model conditional on the posterior draws of the parameters,

$$\tilde{y}^s|y, \theta^s \sim p(\tilde{y}|\theta^s) \quad (1.5)$$

where the superscript s signifies a particular draw from the posterior obtained through MCMC simulation. For example, if we had assumed that data were independent and normally distributed conditional on the underlying parameters, then we could simulate draws of \tilde{y} by simulating from $N(\mu^s, \sigma^s)$. In the same way, we can generate predictive distributions for any latent variables in a hierarchical model. If draws from these predictive distributions are consistent with the observed data, then this is a sign that the model has reasonably fit the data.

In a similar fashion, estimates of functions of the parameters and uncertainty intervals for these ‘utility’ functions can be obtained. Two examples of useful

utility functions are failure time distributions and cost functions. To calculate the distribution of the function, we calculate the value of the function conditional on the MCMC draws of the parameters

$$F(x|\theta^s) \sim F_{X|y}. \quad (1.6)$$

Comparing models Once we have fitted 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 (Vehtari, Gelman, & Gabry, 2017).

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,

$$\text{lpd} = \log \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta = \log p(\tilde{y}_i|y), \quad (1.7)$$

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. (1.7) is called the *log posterior density* (lpd). If we observe multiple new data points $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_I)$, this can be dealt with in a point-wise fashion using the *log point-wise posterior*

density (lppd),

$$\text{lppd} = \sum_{i=1}^I \log p(\tilde{y}_i|y). \quad (1.8)$$

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; $\log p(\tilde{y}|y) = \sum_{i=1}^I \log p(\tilde{y}_i|y)$. Equations (1.7) and (1.8) are defined for a given set of new observations, but the new unobserved data points \tilde{y}_i arise from the true data generating process and so are a random variable with distribution

$$\tilde{Y}_i = f(\tilde{y}_i). \quad (1.9)$$

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

$$\text{elppd} = \sum_{i=1}^I \int \log p(\tilde{y}_i|y) f(\tilde{y}_i) d\tilde{y}_i. \quad (1.10)$$

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

$$\text{computed elppd} = \sum_{i=1}^I \int \log \frac{1}{S} \sum_{s=1}^S p(\tilde{y}_i|\theta^s) f(\tilde{y}_i) d\tilde{y}_i. \quad (1.11)$$

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 rest of the data, and then calculate the log-likelihood of the withheld portion of the data given the samples from the posterior. In leave-one-out cross-validation (LOO-CV) each observation is

withheld according to

$$\text{elppd}_{\text{LOO-CV}} = \sum_{i=1}^I \log \frac{1}{S} \sum_{s=1}^S p(y_i | [\theta]_{-[i]}^s). \quad (1.12)$$

$[\theta]_{-[i]}^s$ 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 $\tilde{y}_{n+1} = \tilde{y}_{n+1,1}, \dots, \tilde{y}_{n+1,I}$. In all three cases, the likelihoods of the observations 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 differ depending on the model's hierarchical structure.

Using cross-validation to estimate elppd involves repeatedly re-fitting the model to different subsets of the data, which is computationally inefficient. A much more efficient method is to approximate the $\text{elppd}_{\text{LOO-CV}}$ (Vehtari et al., 2017); however, in hierarchical modelling cases where the 'left out' portions of the data are nested, and as the size of these nested portions increases, the approximations are less likely to work well (Vehtari et al., 2017). All of the cases in this thesis where $\text{elppd}_{\text{LOO-CV}}$ is used to compare models is in the context of such nested data structures. Furthermore, most of the datasets are small. Therefore, I use the full cross-validation scheme and incur the computational overheads.

1.5 Structure of this thesis

This chapter has introduced the industry-derived motivations for the work in this thesis and provided a high-level introduction to the threads that flow through the body of work: reliability analysis and Bayesian model building. The body of the thesis is divided into two parts and unified at the end by a general discussion/-concluding chapter. Part I addresses lifetime analysis, whereas Part II addresses degradation modelling. At the beginning of each part, I have included a preamble that provides a background on the industry placement project/s that motivated the works 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 may be useful to reliability practitioners in the industry.

Part I of the thesis focuses on lifetime analysis, specifically, how to obtain reliable inference from Weibull analysis of data that can be both left-truncated with unknown exposure history and right-censored; a problem that naturally arises when in failure time records of repeatedly replaced components whose failures are only reliably recorded after a particular date, which is the case for the idler-frame dataset. Incorrectly accounting for left truncation in a sample can result in biased inference. Chapter 2 demonstrates that if partially observed lifetimes that are left truncated are treated as missing values in a Bayesian analysis, their values can be imputed, and the corresponding truncation times can be calculated. In the chapter, I also build upon a method for constructing a joint prior for the Weibull parameters, implementing the method in models for right-censored and left-truncated data and demonstrating its usefulness in the context of partially observed data. Chapter 3 applies the methods from the previous chapter to the idler-frame dataset and shows how predictive distributions for the RUL of idler-frames in operation and the cumulative number of failures can be constructed from the posterior draws. I also show how the draws of the Weibull parameters can be used to propagate uncertainty in the analysis through a cost function to

decide a fixed time replacement interval for the idler frame.

Part II of the thesis is more loosely structured. Chapters 4–6 show an iterative model-building process centred around a Gamma stochastic process for degradation. Chapter 4 demonstrates how the Gamma process can be extended through the Bayesian hierarchical modelling (BHM) framework to account for noisy observations. Chapter 5 then shows the extension of the noisy gamma process model through the same BHM structure to simultaneously model the degradation traces of multiple units and incorporate unit-to-unit variability. Chapter 6 looks at the applied problem of modelling the wearing surface of an overland conveyor’s belt. In the chapter, I show how functional data analysis can be used alongside degradation modelling to model the degradation of a wearing surface. In the sixth chapter, I apply the methods from Chap. 4 and 5 to model the degradation of the belt. I also compare the stochastic process method with a simpler general path model.

The concluding chapter, Chap. 7, ties the two parts of work in the thesis back to the overarching topics of reliability and maintenance in the iron ore and mineral processing industry. 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.

Part I

Part one: lifetime analysis

In this first part of the thesis, I look at the problem of analysing the partially observed lifetime data of repeatedly replaced components. That is, when a component has been repeatedly replaced for many years, but the failure records are only available from a particular date up to the present. In these cases, the first observed failure of each component is the end of a lifetime with an unknown start time, and the most recent replacement marks the beginning of a lifetime, which we have not yet observed the end of.

The problem was motivated by an industry placement where I collated a dataset of idler failures for all of the conveyors at several of the industry partners' mines to inform the design of fixed-time replacement strategies for the idlers—that is, determine what the best time interval is to replace the idler so that the reliability of the conveyor is maximised and the cost of preventative maintenance is minimised. Constructing such a dataset required me to collate information from multiple systems (planned and unplanned replacement records and operation records). Understanding the process of idler replacements and wrangling and cleaning the dataset was a substantial piece of work. The idler-frame dataset that I present in this thesis is an example conveyor from this larger dataset.

The importance of maintenance and the culture around it at the company has changed a lot over the decades that the mines have been operating. It has only been in recent years that replacement records have reliably been captured in the company's records. This gives rise to the problem of partially observed lifetime data, which I will address in this part of the thesis.

My initial approach to modelling such data was published as a conference paper:

- Leadbetter, R., Phatak, A., Polpo, A. & Hodkiewicz, M. (2021) *Informative Bayesian Survival Analysis to Handle Heavy Censoring in Lifetime Data*, International Conference on Maintenance and Intelligent Asset Management (ICMIAM), Ballarat, Australia, pp. 1-6,
doi: 10.1109/ICMIAM54662.2021.9715184.

I was the first author of this publication, performed all of the analysis (under the guidance of the co-authors), and produced the first version of the manuscript. All authors revised and edited the paper. However, since the initial approach was published, I have taken an alternative approach to the problem; this is the approach that I present in Chap. 2. In Chap. 3, I apply the newly proposed approach to the industry idler-frame data.

Chapter 2

Weibull analysis of partially observed 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 possess large datasets of component installation and replacement times. A natural use of these personalised failure time data sets is to tailor replacement strategies for the company's specific operating environments (Meeker et al., 2022, p. 13), rather than solely relying on the manufacturer's recommendations. One problem, however, is that these large observational datasets collected through CMMS are much messier than the experimental ones used by manufacturers in traditional reliability/warranty analysis. This messiness comes about because of reporting issues, incomplete historical 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 known, but either an upper bound, lower bound, or both are known. On the other hand, left-truncation arises when only units that have lasted more than some truncation time are observable. The incomplete—censored and truncated—nature of

such datasets means they are only not strongly informative of the lifetime model.

The censored and left-truncated nature of such data makes what would otherwise be a very straightforward analysis far more complicated. Worse yet, the incompleteness of the data is not always obvious, and mistreatment during analysis can lead to biased results and misinformed decisions. The idler-frame dataset introduced in Sec. 1.3 is one such case where data are both left truncated and right censored. In this case, right censoring arises due to the set of idlers in a frame either being preventatively replaced or still being in operation when the data were analysed, and left-truncation arises since any idlers that were installed and failed before the time failures started being recorded in the CMMS are not present in the dataset but any that were install before this time and failed after are. A further complicating factor of the idler frame dataset is that the installation times 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 sometimes referred to as unknown initial conditions or unknown exposure history (Guo, 1993). Treatment of right-censored and left-truncated data was addressed by Hong, Meeker, and McCalley (2009); however, not for cases with unknown exposure history of left-truncated samples. In this chapter, I propose a method for handling such cases in a Bayesian framework by imputing the unobserved portion of the left-truncated lifetimes with unknown exposure history and, along with them, the truncation times. I demonstrate the method using simulated data that mimics the observation process of the idler-frame data.

The incompleteness of a dataset that results from censoring and truncation reduces the information in the dataset, meaning that the data are only weakly informative of the model. Particularly when a large proportion of the dataset is right censored, there is little information in the dataset about longer lifetimes—the upper tail of the lifetime distribution. To reduce uncertainty in the analysis of incomplete lifetime data, domain knowledge can be used to inform the model

where the data cannot, but only if the prior is constructed properly. I show how to do this in the simulation example using a development of a method for eliciting a joint prior for the parameters of a Weibull distribution, originally proposed by Kaminskiy and Krivtsov (2005). In their original proposal of the method only briefly demonstrate the method; I show how to properly implement the prior in a model for censored and truncated lifetime data and demonstrate how elicitation can be performed to encode information in different parts of the lifetime distribution. I do this to complement the analysis of the heavily censored and truncated simulation data.

In the next section, Sec. 2.1, I provide a background of lifetime analysis, the Weibull distribution, and how censoring and truncation can be included in the likelihood of the data. I then describe my proposed approach for modelling left-truncated data with unknown exposure history by imputing the unobserved portions of the left truncated samples and their truncation time in Sec. 2.2. The result of Sec. 2.2 is a likelihood that can account for data that is left truncated with unknown exposure history and right censored. The partial information in the data that this likelihood accounts for does not strongly inform the parameters. In such a case, using weakly informative priors can, at best, leave areas of the posterior unusually diffuse and, at worst, place mass in spurious parts of parameter space. Therefore, in Sec. 2.3, I demonstrate how to construct an informative joint prior for the Weibull parameters. I introduce the method proposed by Kaminskiy and Krivtsov (2005) for constructing the joint prior, point out its limitations, and introduce my developments of the method. In Sec. 2.4, I demonstrate, using simulated data, my method for imputing partially observed left-truncated lifetimes and incorporating an informative joint prior. In this demonstration, I compare the imputation method alongside the case where we simply discard the partially observed left truncated lifetimes and a case where we fully observe them (if we know their installation times). Sec. 2.5 presents a small simulation experiment that explores the limitations of the imputation approach. The chapter concludes

with Sec. 2.6, where I summarise the key contributions and findings from the chapter and provide recommendations for analysing lifetime datasets with right censoring and left truncated observations with unknown exposure histories, such as the idler-frame lifetime data. The recommendations distilled in this final section are applied to the idler-frame data in Chap. 3. Table 2.1 provides a reference for the notation used in this chapter.

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 its 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 individuals/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 parameters 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, as well as 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 Y , the exposure time, on $[0, \infty)$. Y is some continuous or discrete exposure time from

Table 2.1: Nomenclature for the chapter.

| | |
|----------------|--|
| y | The true value of a lifetime. |
| \tilde{y} | The imputed value of a missing lifetime. |
| y^O | A lifetime that is fully observed; not left-truncated by the begining of observation or right-censored by the end. |
| y^C | The true (unobservable) value of a lifetime that is censored at its. |
| y^T | The value of a lifetime that is truncated at the beginning of its life. |
| y^{TC} | The value of a lifetime that is truncated at the beginning of its life and censored at its end. |
| c^{Lower} | The lower censoring time of a censored observation. If $c^{Lower} = 0$ the lifetime is left censored. |
| c^{Upper} | The upper censoring time of a censored observation. If $c^{Upper} = \infty$ the lifetime is right censored. |
| $c^{C;Lower}$ | The lower bound of a lifetime that is censored at the end of the lifetime. |
| $c^{TC;Lower}$ | The lower bound of a lifetime that is censored at the end of the lifetime and truncated at the beginning. |
| $c^{T;Lower}$ | The lower bound of a lifetime that is truncated at the beginning of the lifetime when the exposure history is unkown. |
| $c^{T;Upper}$ | The upper bound of a lifetime that is censored at the end of the lifetime and truncated at the beginning. |
| τ | The left truncation time. |
| $\tilde{\tau}$ | The imputed left truncation time. |
| τ^T | The left truncation time of a lifetime truncated at its beginning but not censored at its end. |
| τ^{TC} | The left truncation time of a lifetime truncated at its beginning and censored at its end. |
| n^O | The number of fully observed lifetimes; i.e. not truncated by the beginning of observation or right censored by the end. |
| n^C | The number of observations that are censored by the end of observation but not truncated by the beginning. |
| n^T | The number of observation that are truncated by the beginning of observation but not right censored by its end. |
| n^{TC} | The number of observation that are both truncated by the beginning of observation and right censored by its end. |

a clearly defined origin, the installation of the component, to a well-defined event, the component's failure. In reliability analysis, the exposure is typically absolute time, the operating time of the unit, or cycles of operation. For example, the idler-frame failures are recorded in absolute time since operating time is unavailable. Next, a specific parametric lifetime distribution is chosen for the random variable Y , $p(Y|\theta)$, and the parameters θ 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), $F(y|\theta)$, is the probability that a unit will have failed by age Y , i.e. $\Pr[Y \leq y|\theta]$. It is also sometimes called the cumulative risk function.
- **Survival function**, $S(y|\theta)$, is the complement of the CDF, $S(y|\theta) = \Pr[Y > y|\theta] = 1 - F(y|\theta)$, and defines the probability of a unit surviving up to an exposure time Y .
- **Hazard function**, $h(y|\theta)$, which is the instantaneous failure rate conditioned on the age of the unit and can be calculated as $h(y|\theta) = p(y|\theta)/S(y|\theta)$.

For example, the CDF quantifies the risk of unplanned failures given a chosen preventative maintenance interval, and the hazard function identifies if a unit's 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 component lifetimes, that is

$$Y|\beta, \eta \sim \text{Weibull}(\beta, \eta), \quad (2.1)$$

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 the analysis that follows, I use the coupled parameterization of the two-parameter Weibull distribution, which has PDF

$$p_W(y|\beta, \eta) = \frac{\beta}{\eta} \left(\frac{y}{\eta} \right)^{\beta-1} \exp^{-\left(\frac{y}{\eta}\right)^\beta}, \quad (2.2)$$

CDF

$$F_W(y|\beta, \eta) = 1 - \exp^{-\left(\frac{y}{\eta}\right)^\beta}, \quad (2.3)$$

and hazard function

$$h_W(y|\beta, \eta) = \frac{\beta}{\eta} \left(\frac{y}{\eta} \right)^{\beta-1}. \quad (2.4)$$

The shape parameter β dictates whether the hazard increases, $\beta > 1$, decreases, $\beta < 1$, or stays constant, $\beta = 1$. The effect of the shape parameter on the hazard function is demonstrated in Fig. 2.1. Practically speaking, if the hazard function increases with exposure, this corresponds to a wear-out failure mechanism, whereas if it decreases, it corresponds to infant mortality. This distinction is important from a maintenance perspective because if the component does not wear out, a 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.

2.1.3 Censoring

It is very common for lifetime data to be censored (Tian, Lewis-Beck, Niemi, & Meeker, 2024). 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 three

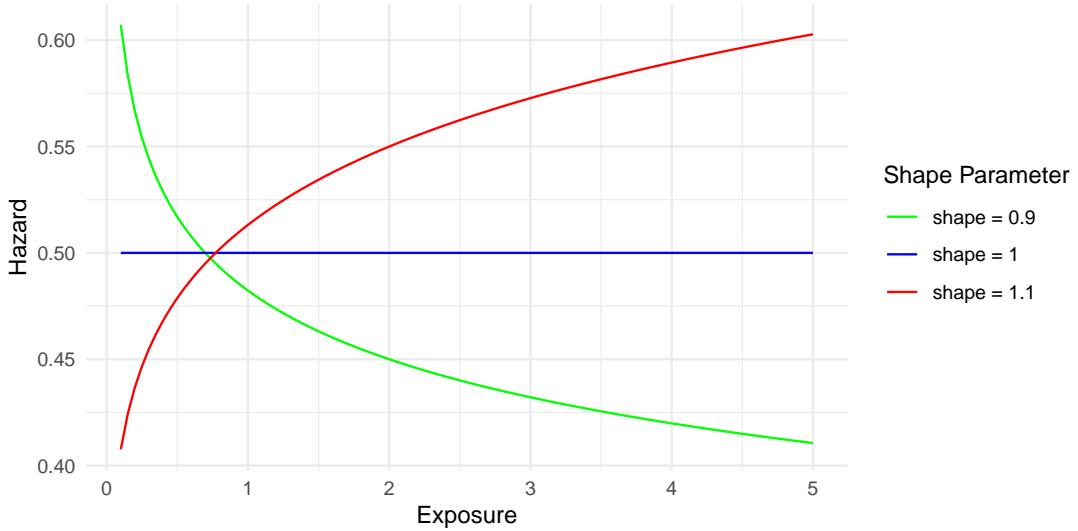


Figure 2.1: The Weibull hazard function when $\beta = 0.9$, $\beta = 1$, or $\beta = 1.1$ and $\eta = 1$.

are treated in much the same way. Figure 2.2 demonstrates these three types of censoring. For demonstration, say that you want to know the average lifetime of a light bulb to decide how many to buy for your house and Fig. 2.2 shows an experiment with three bulbs. In the figure, the three bulbs are installed at time t_0 , and then you check if they are still operating at $t_1 = 0.5 \times 1000$ hours and again at $t_2 = 1 \times 1000$ hours. When you check at time t_1 , one of the bulbs has failed and by the time you check again at t_2 , so has another. Each unit's true—unobserved—failure times are shown as crosses in the figure. The bulb that fails before t_1 is left censored because you only observe an upper bound of the lifetime and so its true lifetime must be between $(0, t_1)$. The second bulb to fail is interval censored since you observed it operating at t_1 but failed at t_2 and so its true value must be between the upper and lower bounds (t_1, t_2) . The third bulb, which has yet to fail, is right censored since you only know that it has lasted longer than t_2 and so its true failure time must be between (t_2, ∞) . Right and left censoring are special cases of interval censoring where the upper or lower bound of the lifetime is infinity or zero, respectively. Left censoring is uncommon in reliability, so in the discussions that follow, I focus on right and interval censoring. However, all

of the methods can be easily extended to accommodate left-censored data.

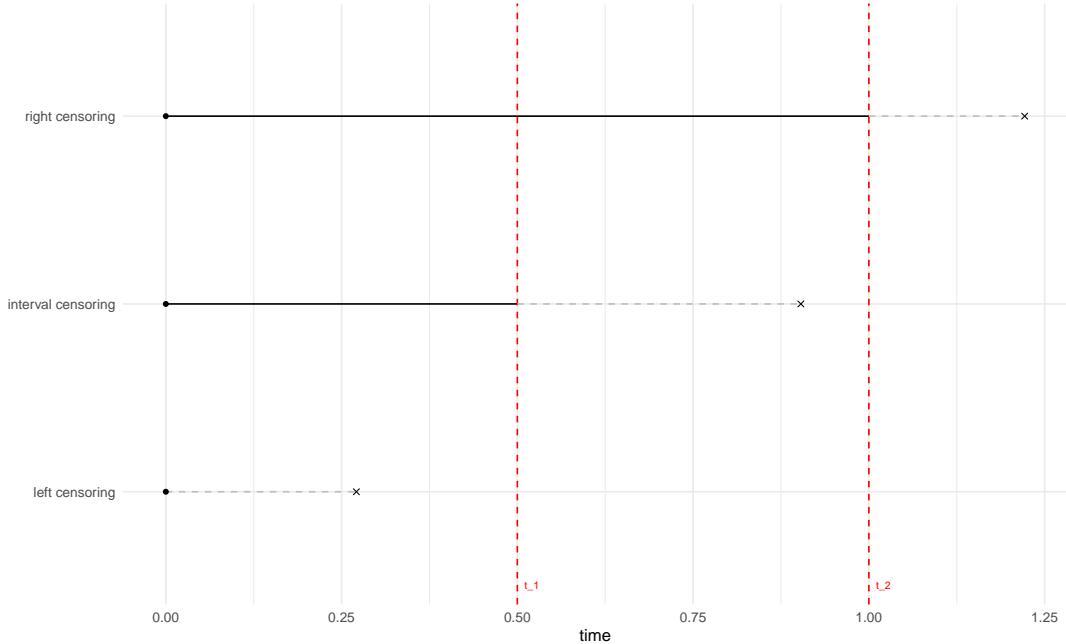


Figure 2.2: An example of the different between left censoring in a dataset (a) and left-truncation (b). In (a) the lifetimes that fail before the red line at age = 0.5 are left censored since their existence is known but not their exact failure time. In (b) any lifetimes that fail before the red line at age = 0.5 are absent from the dataset completely and therefore the likelihood of the observed data needs to be adjusted to reflect this.

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 (parameter) 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)¹. This is easily done during MCMC routines since, at each step, the imputed values can be sampled in the same way as the other parameters in the model. The distribution of the imputed censored lifetimes is assumed to be the same distribution of the rest of the population

$$\tilde{Y}_i^C | \beta, \eta \sim \text{Weibull}_{c_i^{Lower}}^{c_i^{Upper}}(\beta, \eta). \quad (2.5)$$

¹The Stan user guide shows how to impute censored observation in [Truncated or Censored Data](#).

Here \tilde{Y}_i^C is the imputed value of the censored lifetime, and the superscript c^{Upper} and subscript c^{Lower} indicate that the distribution is constrained by the upper and lower censoring times. The imputed missing lifetimes are used to evaluate the full data likelihood 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 and MCMC methods, it is very simple to derive uncertainty intervals for the parameters, imputed values, and useful quantities, which I show in Chap. 3.

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

$$\begin{aligned}\Pr [c^{Lower} < Y_i^C \leq c^{Upper}] &= \int_{c^{Lower}}^{c^{Upper}} p_W(y_i^C) dy_i^C \\ &= F_W(c^{Upper}) - F_W(c^{Lower}).\end{aligned}$$

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

$$L(\theta|\text{DATA}) = \prod_{i=1}^{n^O} p_W(y_i^O) \prod_{j=1}^{n^C} [F_W(c_j^{Upper}) - F_W(c_j^{Lower})], \quad (2.6)$$

where $\theta = \{\beta, \eta\}$ is the set of parameters of the Weibull lifetime distribution, n^O and n^C are the number of fully observed and censored observations, respectively, and $\text{DATA} = \{y_1^O, \dots, y_{n^O}^O, c_1^{Upper}, \dots, c_{n^C}^{Upper}, c_1^{Lower}, \dots, c_{n^C}^{Lower}\}$ is the observed lifetimes and censoring times. This second approach is much more commonly used, particularly in the reliability literature (for example Hong et al. (2009); Meeker et al. (2022); Mittman (2018); Tian et al. (2024)). However, as I show later, it is convenient to frame the model using the first approach—where censored lifetimes are imputed—for the particular problem when data are also left-truncated with unknown installation times.

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 satisfy in order to be observable (Guo, 1993). Left truncation, for example, arises when some units must survive up to a certain time to be observed. It is also possible for data to be right- or doubly-truncated, but left-truncation is the most common in lifetime data, particularly in observational reliability datasets (Emura & Michimae, 2022). The term ‘left truncated observation’ is often used to mean an observation that arises from a truncated population. 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. In contrast, truncation is a characteristic of the population because we do not know how many lifetimes were not included in the dataset because they did not survive past the truncation time, and hence, our sample is not representative of the true population. It is a difference between known-unknowns (censoring) and unknown-unknowns (truncation). Left-truncated samples tend to over-represent longer lifetimes (Guo, 1993). An example of a left-truncated dataset is shown in Fig. 2.3.

Figure 2.3 shows an example of the difference between a data set with left censoring and one that is left truncated. Using the example of lightbulbs once again, consider that we test ten bulbs but only monitor their failures after 0.5×1000 hours. Figure 2.3 (a) shows a sample of ten lifetimes with the portion of lifetimes that sits to the left of the start of observation (the red line) greyed out to show that they are unobserved. In this case, lifetimes 2, 3, 4, and 8 are left censored because we know that they were installed and that they failed sometime before 0.5×1000 hours. However, now consider that someone else runs the bulbs for the first 0.5×1000 hours and then gives us the six un-failed bulbs to perform our experiment (Fig. 2.3 (b)). In this case, we do not know that four bulbs

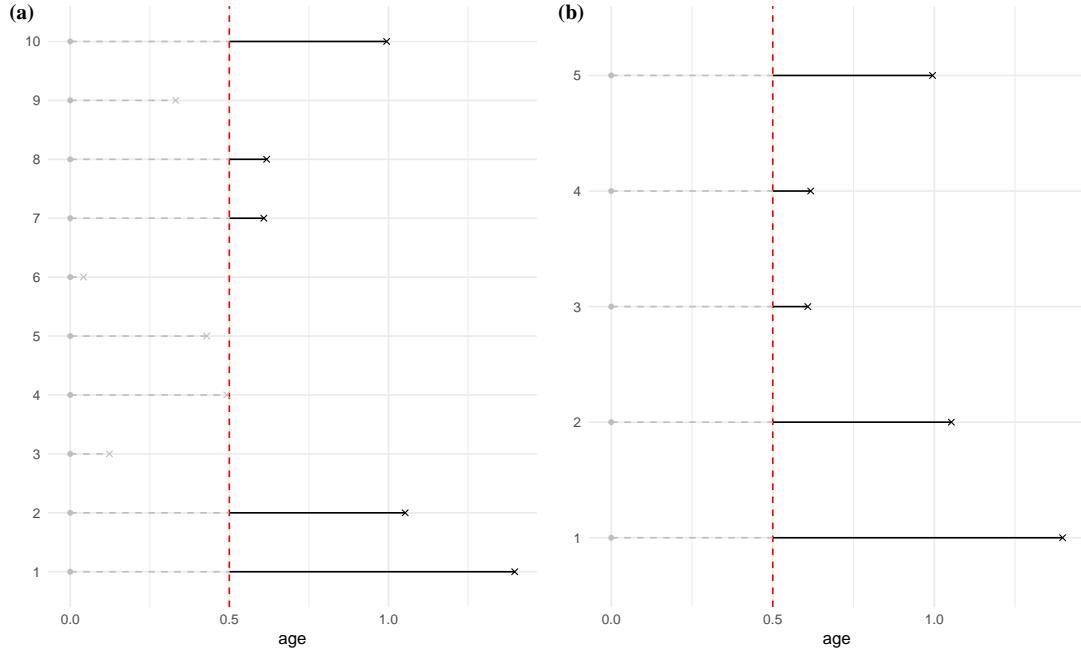


Figure 2.3: An example of a dataset with left-truncated samples. Units are put under test at random times, and their failure times are recorded. However, the unit is only captured in the dataset if it fails after t_{start} ; therefore, any units that were installed and failed before t_{start} are not included in the sample.

have already failed, only that the bulbs we have are already 0.5×1000 hours old (the truncation time). The probability of an observation that arises from a left truncated distribution is $\Pr [Y^T | Y^T > \tau^T]$ where τ^T is the truncation time. Therefore, the contribution of a left-truncated observation to the likelihood is

$$L(\theta | y_i^T) = \frac{p_W(y_i^T)}{1 - F_W(\tau_i^T)}, \quad (2.7)$$

where the denominator re-normalized the density by the probability of surviving past τ^T .

2.1.5 Left-truncation and right-censoring

Observational reliability datasets commonly contain both left-truncated and right-censored observations. This combination naturally arises in 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 date a new CMMS was adopted) are absent in the dataset. Figure 2.4 shows a toy example, 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 greyed out in the figure—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 in the figure. Returning to the example of light bulbs, say you recently moved into a new home and started recording the failures of bulbs in your home; the lifetimes of any bulbs that were already installed when you moved in would be left truncated, and the lifetimes of any bulbs yet to fail are right censored. Hong et al. (2009), Mitra (2013), and Kundu and Mitra (2016) 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 that I analyse in Chap. 3 is also an example of this type of dataset since the idlers in the frames are repeatedly replaced when they fail, and the replacement records of the idlers are only available from the end of 2014 onwards, but the conveyor has been in operation for much longer and is still in operation today.

Hong et al. (2009) shows that a likelihood for data that is left-truncated and right-censored can be written as,

$$L(\theta|\text{DATA}) = \prod_{i=1}^{n^O} [p_W(y_i^O)] \times \prod_{j=1}^{n^T} \left[\frac{p_W(y_j^T)}{1 - F_W(\tau_j^T)} \right] \times \prod_{k=1}^{n^C} \left[1 - F_W(c_k^{C;Lower}) \right] \times \prod_{l=1}^{n^{TC}} \left[\frac{1 - F_W(c_l^{TC;Lower})}{1 - F(\tau_l^{TC})} \right], \quad (2.8)$$

where the censored lifetimes are integrated out as in eq. (2.6). n^O , n^T , n^C , and n^{TC} is the number of fully observed lifetimes, left truncated lifetimes, right censored lifetimes, and lifetimes that are both left truncated and right cen-

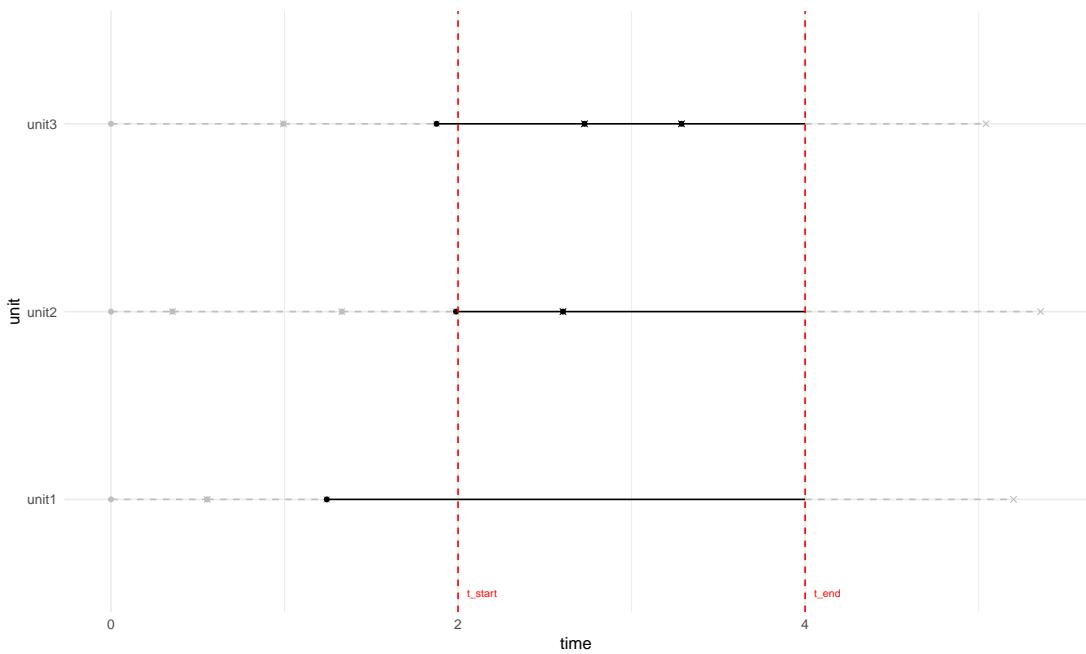


Figure 2.4: An example of how left truncation and right censoring arise from the repeated replacement of three units. Three units are shown on the vertical axis, and time on the horizontal. The failure times of the units are shown as black crosses, and every time a unit fails, it is instantaneously replaced. If we start observing these failures at t_{start} and stop at t_{end} , then any lifetime that began before t_{start} and failed after is left truncated, and any lifetimes that began before t_{end} and failed after are right censored.

sored, respectively. Now, DATA = $\{y^O, y^T, \tau^T, c^{C;Lower}, c^{TC;Lower}, \tau^{TC}\}$ where $y^O = \{y_1^O, \dots, y_{n^O}^O\}$ is the set of fully observed lifetimes; $y^T = \{y_1^T, \dots, y_{n^T}^T\}$ the set of fully observed left truncated lifetimes and $\tau^T = \{\tau_1^T, \dots, \tau_{n^T}^T\}$ their truncations times; $c^{C;Lower} = \{c_1^{C;Lower}, \dots, c_{n^C}^{C;Lower}\}$ the set of right censoring times of the right censored observations; and $c^{TC;Lower} = \{c_1^{TC;Lower}, \dots, c_{n^{TC}}^{TC;Lower}\}$ the set of right censoring times of the lifetimes that are both right censored and left truncated and $\tau^{TC} = \{\tau_1^{TC}, \dots, \tau_{n^{TC}}^{TC}\}$ their truncation times. I take a moment to note here that other publications, such as Hong et al. (2009), use indicator variables to identify if lifetimes are observed, right censored, and/or left truncated, however, I am explicitly splitting up the population because it makes it clear which lifetimes are observed and which are missing. 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, formulated as

$$L(\theta|\text{DATA}) = \prod_{i=1}^{n^O} [p_W(y_i^O)] \times \prod_{j=1}^{n^T} \left[\frac{p_W(y_j^T)}{1 - F_W(\tau_j^T)} \right] \\ \times \prod_{k=1}^{n^C} [p_W(\tilde{y}_k^C)] \times \prod_{l=1}^{n^{TC}} \left[\frac{p_W(\tilde{y}_l^{TC})}{1 - F_W(\tau_l^{TC})} \right], \quad (2.9)$$

where $\tilde{y}^C = \{\tilde{y}_1^C, \dots, \tilde{y}_{n^C}^C\}$ and $\tilde{y}^{TC} = \{\tilde{y}_1^{TC}, \dots, \tilde{y}_{n^{TC}}^{TC}\}$ are the imputed values of the censored and censored/truncated observations and are now included in the set of parameters $\theta = \{\beta, \eta, \tilde{y}^C, \tilde{y}^{TC}\}$. To simplify notation, I express the model for the data in eq. (2.9) in a Bayesian framework as

$$\begin{aligned} Y_i^O | \beta, \eta &\sim \text{Weibull}(\beta, \eta) \\ Y_j^T | \beta, \eta &\sim \text{Weibull}(\beta, \eta) & T[\tau_j^T,] \\ Y_k^C | \beta, \eta &\sim \text{Weibull}_{c_k^{C;Lower}}(\beta, \eta) \\ Y_l^{TC} | \beta, \eta &\sim \text{Weibull}_{c_l^{TC;Lower}}(\beta, \eta) & T[\tau_l^{TC},] \\ \beta, \eta &\sim \pi(\theta_{\beta, \eta}), \end{aligned}$$

where the sequence of the first four expressions corresponds to the sequence of products of the likelihood in eq. (2.9), $\text{Weibull}_{c^{Lower}}$ indicates that the random variable has a Weibull distribution and is constrained to be greater than the censoring time, and $T[\tau,]$ indicates that the distributions are re-normalised by the probability $\Pr[Y > \tau]$ (Stan Development Team, 2022)². For the moment, I express the joint prior for the Weibull parameters in its most general form.

Unknown exposure history A problem arises when the installation time of the left-truncated lifetimes (lifetimes that begin before the start of observation)

²This is the notation used in the Stan user manual in *Truncated or Censored Data*.

is unknown since, to re-normalise the truncated lifetime distribution of these observations, we must know how old they are at the beginning of observation τ^T . For example, you probably would not know the age of the lightbulbs in your house when you first move in since you would not know when the previous owner installed them. The left truncated observations in the idler frames dataset also have unknown exposure history since replacement records before 2014 are unreliable and not recorded at the frame level.

This problem is known as unknown exposure history or initial conditions (Guo, 1993). In these cases, two approaches can be taken. The first is to discard all 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, 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 (Heckman & Singer, 1986). In Sec. 2.2, I show how, by treating the unknown installation times as a case of censoring and imputing the censored data, the missing truncation times can also be imputed, and reasonable parameter estimates can be obtained.

2.2 Imputing unknown exposure histories

Using the toy example in Fig. 2.4, say we do not observe any installation or failure times to the left of t_{start} . In this case, we know that each unit's first, partially

observed lifetime started sometime between $t = 0$ and $t = t_{\text{start}}$. This is a case of interval censoring, where the lower censoring bound $c^{T;Lower}$ is the time from the beginning of observation to the failure time, and the upper bound is from $t = 0$ to the failure time $c^{T;Upper}$.

In many practical applications, there will be no clear origin $t = 0$ with which to calculate an upper bound for an interval-censored lifetime, and the analyst will need to make an informed decision. For example, in the light bulb example, $t = 0$ may be when the house was first built or when the particular type of bulb started being manufactured, whichever is more recent. For the case of the idler-frames, $t = 0$ is the date that the conveyor was first commissioned. If there is no sensible choice of $t = 0$, then the lifetimes with unknown exposure histories could be treated as right censored, in which case $c^{T;Upper} = \infty$, but the following logic would still apply. In this case, the right censoring time is relative to the end of the lifetime, rather than the beginning, like typical right censored observations.

Treating the left truncated lifetimes with unknown exposure history as interval-censored, their distribution is

$$Y_j^T | \beta, \eta \sim \text{Weibull}_{c_j^{T;Lower}}^{c_j^{T;Upper}}(\beta, \eta) \quad T[\tilde{\tau}_j^T,], \quad (2.10)$$

where the missing truncation time $\tilde{\tau}_j^T$ —needed to renormalise the contribution to the likelihood—is calculated from the imputed value of the lifetime as

$$\tilde{\tau}_i^T = \tilde{y}_i^T - c_j^{T;Lower}. \quad (2.11)$$

In the lightbulb example, if you moved into a house a year ago and when you moved in the house was five years old, a left truncated lifetime with unknown exposure history would be the life of a bulb that was installed when you moved in and failed within the year that you have been living in the house. Say a bulb that was installed when you moved in failed after six months. Using the imputation method, you impute the true value of the lifetime for this bulb, which you know

is between six months and five years and six months. Say the imputed age is four years, according to eq. (2.11), the imputed truncation time is three years and six months.

When the lifetime is both interval censored by the start of the observation period and right censored by the end, such as unit 1 in Fig. 2.4 it is not possible to calculate the missing truncation time according to eq. (2.11). For example, a light bulb that was installed when you moved into your house and is still operating after the one year that you have lived there. The value of such lifetimes can be imputed like any right censored observation—where the censoring time is the period that we observed its operation $c^{TC;Lower} = t_{end} - t_{start}$ —from the distribution

$$Y_l^{TC} | \beta, \eta \sim \text{Weibull}_{c_l^{TC;Lower}}(\beta, \eta) \quad T[\tilde{\tau}_l^{TC},]. \quad (2.12)$$

This is the same as right censoring because for a bulb that was installed when you moved in and is yet to fail, you only know that the lifetime is longer than the one year you have lived in the house—which is right censoring. However, because there is no observed installation or failure time for the lifetime, $\tilde{\tau}^{TC}$ can take on any value between zero and $\min(t_{start}, \tilde{y}_l^{TC} - c^{TC;Lower})$. I impute the truncation time from

$$\tau_l^{TC} \sim \text{Unif}\left(0, \min\left(t_{start}, \tilde{y}_l^{TC} - c_l^{TC;Lower}\right)\right), \quad (2.13)$$

where $c^{TC;Lower}$ is the right censoring time—in otherworlds the time that we observed the lifetime for. Sampling the value of $\tilde{\tau}^{TC}$ in this way should incorporate the extra uncertainty in the posterior. For example, for a light bulb that has been in operation the entire time that you have lived in your house (one year) you impute a value for the true value. If the imputed value is two years, then the missing truncation time could be anywhere from zero (if the bulb was installed just before you moved in) to one year (if the bulb fails immediately after you perform your analysis). However if the imputed value is seven years, then the truncation period can only be between zero and five years because it is not

possible that the bulb was installed before the house was built.

The method to impute the missing lifetimes and left truncation times of the left-truncated lifetimes with unknown exposure history allows the information in the these incomplete observations to be retained in the analysis. However, incomplete data such as these lifetimes and the right-censored ones are typically only weekly informative of the parameters (Tian et al., 2024). When this is the case, the model for the data should be paired with a well constructed informative prior that ensures that the posterior does not have mass in implausible ares of parameter space.

2.3 Informative joint prior

In cases where data only provide partial information—such as lifetime data where some lifetimes are masked by censoring—an informative prior can help to ‘fill in the gaps’ and inform areas of the lifetime distribution that the incomplete data cannot. For example, right censoring masks the upper tail of the lifetime distribution, and if there are no observed failures beyond some censoring time, then the data do not contain any information that informs this upper tail. This will become clear in Sec. 2.4.2 when I fit a Weibull distribution to censored lifetimes after discarding any truncated lifetimes. An added motivation for using an informative prior in these cases is that non-informative priors can place mass in implausible parts of the parameter space, and when combined with a weak likelihood—i.e. the data are not strongly informative of the parameters—, this leads to spurious parameter estimates (Tian et al., 2024).

Kaminskiy and Krivtsov (2005) proposed a method for encoding an informative joint prior for the two parameters of the Weibull distribution by eliciting information about cross sections of the Weibull CDF. In their method, the analyst provides an informed estimate of the expected value of the CDF at two exposure times, t_1 and t_2 , and the level of uncertainty around these estimates. The pair of estimate and uncertainty level at each exposure time ($\mu_{\hat{F}_{t_i}}, \sigma_{\hat{F}_{t_i}}$), $i = \{1, 2\}$,

are encoded as the mean and standard deviation of a beta distribution. Then, by sampling realisations of the CDF at each exposure time from the two beta distributions, ensuring that $\hat{F}_{t_1} < \hat{F}_{t_2}$, the parameters of the Weibull CDF that passes through the two realisations can be calculated to obtain a draw from the informative joint prior. Kaminskiy and Krivtsov (2005) then show how to obtain a Bayesian point estimate of the CDF at some new exposure time where binomial failure data are available. To do so, they calculating the parameters of a new beta distribution that describes the CDF at the new exposure time according to the joint prior draws, they then updating the parameter estimates of the beta distribution at t_3 with the data, taking advantage of the beta-binomial conjugacy.

The method for obtaining joint draws from an informative prior for the two Weibull parameters is consistent with the recommendations of Gelman, Vehtari, et al. (2020), who suggest eliciting information on the outcome space—which is more familiar to domain experts—and then translate this information into an informative prior in the parameter space that also indirectly describes how the parameters should be allowed to covary with one another. An alternative method proposed by Meeker et al. (2022) to elicit an informative prior for the Weibull parameters is to reparameterise the distribution in terms of the shape β and some quantile, q_r , of the distribution and then specify independent priors for β and q_r . Both methods have proven useful in practice and have been implemented in reliability software (Krivtsov & Frankstein, 2017). However, it can be difficult to elicit information about β for practitioners unfamiliar with Weibull analysis. Therefore, I develop a variation of Kaminskiy and Krivtsov’s method for obtaining draws from a joint informative prior for the Weibull parameters. I build upon their method by elaborating on what it means to elicit information at different values of t_1 and t_2 and how this translates into covariance in the joint distribution and implement the joint prior in a lifetime model so that the draws from the prior are properly filtered through the likelihood.

The method of Kaminskiy and Krivtsov (2005) provides only point estimates

of the updated parameter values of the beta distribution at t_3 . To obtain corresponding values of the Weibull parameters, the analyst must once again sample pairs of realisations along the CDF—now at either t_3 and t_1 or t_3 and t_2 —and calculate the parameters of the Weibull CDF that passes through the two points. In doing so, he/she ‘reuse’ the prior distribution at either t_1 or t_2 to generate the posterior, and so the ‘prior belief’ about the CDF at whichever time is used is not updated. Furthermore, the resulting joint draws will be sensitive to whether t_1 or t_2 is used to generate the joint posterior distribution of the Weibull parameters. By instead implementing the method for obtaining draws from the joint prior within the HMC sampling algorithm, I show how to properly filter the prior through the likelihood to obtain the full posterior. Doing so also means that the prior can be updated with lifetime data (i.e., failure times rather than binomial trial data), censoring and truncation information can be included in the likelihood, and consequently a proper Bayesian joint posterior is obtained for the two Weibull parameters.

Instead of beta distributions, I use normal distribution truncated between zero and one to encode prior belief about the CDF at t_1 and t_2 . When specifying prior information close to the upper and lower tails of the Weibull CDF a beta distribution places a lot of mass very close to zero or one, which can cause numerical issues during sampling. Truncated normal distributions are much better behaved. The prior can therefore be expressed as

$$\begin{aligned}\hat{F}_{t_1} &\sim N_0^1(\mu_{\hat{F}_{t_1}}, \sigma_{\hat{F}_{t_1}}) \\ \hat{F}_{t_2} &\sim N_{\hat{F}_{t_1}}^1(\mu_{\hat{F}_{t_2}}, \sigma_{\hat{F}_{t_2}}) \\ \beta &= \frac{g(\hat{F}_{t_2}) - g(\hat{F}_{t_1})}{\log(t_1/t_2)} \\ \eta &= \exp \left[\log(t_1) - \frac{\hat{F}_{t_1}}{\beta} \right]\end{aligned}$$

where $g(\hat{F}) = \log(-\log(1-\hat{F}))$ and the superscript and subscripts on the normal

distribution indicate the upper and lower constraints of the support.

Depending on the choice of t_1 and t_2 , the analyst encodes different information into the joint prior, which is reflected in how the two parameters covary with one another. For demonstration, plots (a), (b), and (c) in Fig. 2.5 show the draws from three different joint priors constructed using this elicitation method. To construct the prior, I specify some ‘true’ values of the parameters ($\beta = 1.1; \eta = 1$) and then specify the priors to reflect the true value of the CDF at different quantiles $t_i = q_r$ —where r specifies the probability—so that all of the priors contain the ‘true’ parameter values. Figure 2.5 (a) shows a prior where information is encoded at early exposure times ($t_1 = q_{0.05}$ and $t_2 = q_{0.20}$), (b) shows one where information is encoded around the median age ($t_1 = q_{0.40}$ and $t_2 = q_{0.60}$), and (c) shows one where the elicitation times are in the upper tail of the lifetime distribution($t_1 = q_{0.90}$ and $t_2 = q_{0.99}$). Plots (d), (e), and (f) in Fig. 2.5 show the corresponding uncertainty around the Weibull CDF that results from the different priors in (a), (b), and (c), respectively. We see that the joint prior allows us to express our prior belief in one area of the PDF while allowing the prior to be vague in others. This characteristic has useful application in the context of heavy censoring since censoring masks longer lifetimes and hence increases uncertainty in the upper tail of the distribution but still contains useful information in the lower tail.

In sections 2.4-2.5, I compare the informative joint prior shown in Fig. 2.5 (c) and (f), which encodes information in the upper tail, with the vague independent priors $\beta \sim N^+(1.1, 1)$ and $\eta \sim N^+(1, 1)$, where the superscript (+) indicates that the prior is truncated to be positive.

2.4 Analysis of simulated data

In section, I demonstrate the proposed method for imputing the left-truncated samples with unknown exposure time using data simulated in a way that emulates the idler-frame observation process. Repeated replacements of a set of units are simulated from a Weibull(1.1, 1) distribution and left truncation and right cen-

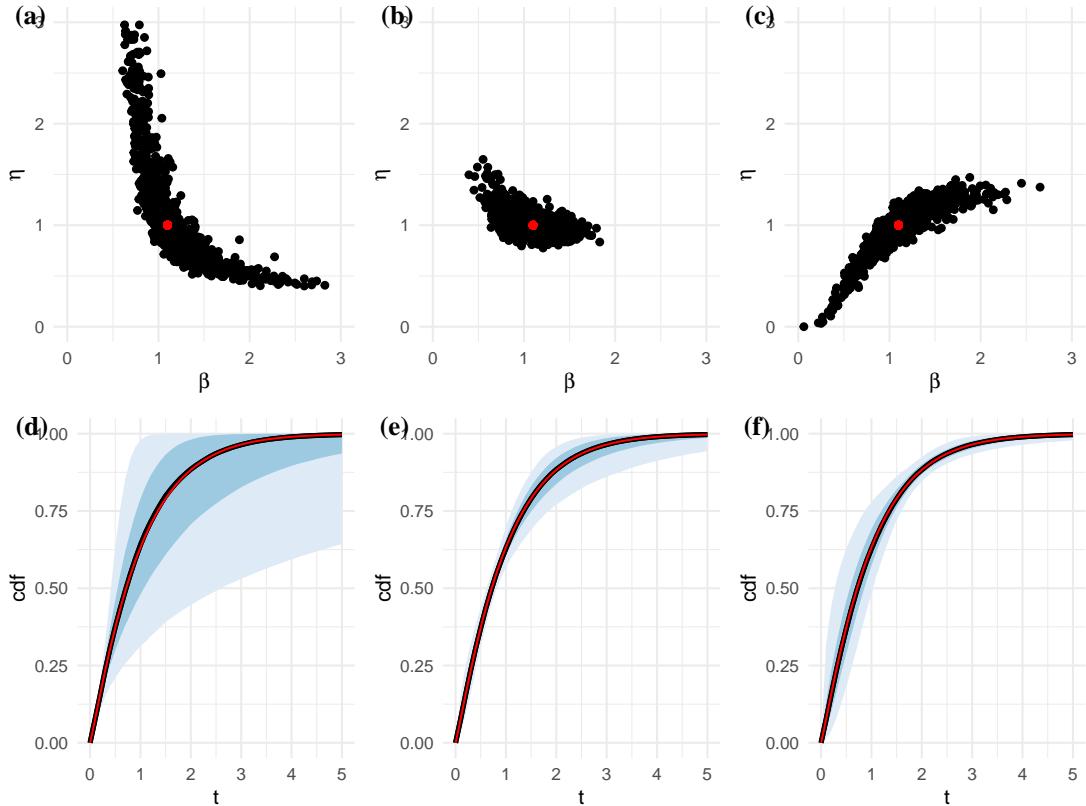


Figure 2.5: Three different informative joint priors constructed using the method of Kaminskiy and Krivtsov (2005). The joint draws of the parameters are shown in the top row—(a), (b), and (c)—and the corresponding uncertainty in the CDF is shown in the bottom—(d), (e), and (f). (a) and (d) show a prior where information is elicited around $t_1 = 0.07$ and $t_2 = 0.26$ (the 0.05 and 0.20 quantiles of the true distribution $\text{Weibull}(1.1, 1)$), (b) and (e) show a prior where information is elicited around $t_1 = 0.46$ and $t_2 = 1.04$ (the 0.35 and 0.65 quantiles), and (c) and (f) show a prior elicited at $t_1 = 1.54$ and $t_2 = 2.71$ (the 0.80 and 0.95 quantiles).

soring educed in the simulated dataset by only retaining the failures that occur within an observation period. I then fit a Bayesian model, which imputes any partially observed lifetimes in the dataset and compare the results with two alternative analyses: one where the left-truncated samples are discarded, and only the fully observed and right-censored observation are used to estimate the Weibull distribution, and another where the left-truncated lifetimes are fully observed. The latter is the ideal scenario, since if the left-truncated lifetimes are fully observed then the likelihoods in either eq. (2.8) or (2.9) can be used to model the data. I compare the three different cases using both vague independent priors for the Weibull parameters and an informative joint prior. The six model-prior treatment combinations of the simulated data are laid out in Tab. ??.

The simulation experiments in this section identify that, for small samples, the imputation approach gives almost the same inference as the case where we know the true installation times. When we throw out all of the left truncated observations, there is a large increase in uncertainty, particularly around the upper tail of the lifetime distribution. However, if an informative joint prior is carefully constructed to inform the upper part of the distribution, then the results of the three different analyses are very similar. In Sec. 2.4.1, I explain how I simulate data that emulates the data-generating process of the idler-frames. I then analyse the simulated data in Sec. 2.4.2 using the three different approaches and a vague prior and compare the posteriors. In Sec. 2.4.3, I repeat the three analyses using an informative joint prior and compare the posterior distributions again.

2.4.1 Simulation method

To simulate data that emulates the idler-frame lifetime data set,

1. I sample $N \times M$ draws from a Weibull distribution with known shape parameter $\beta = 1.1$ and scale parameter $\eta = 1$.
2. These lifetimes are then assign to $M = 10$ units.

Model treatments

Including left-truncation

The samples that are left-truncated by the beginning of observation have known exposure history and are fully accounted for using the likelihood in eq. (2.8) or (2.9).

Discarding left-truncation

The samples that are left-truncated by the beginning of observation have unknown exposure history and are discarded, leaving only the fully observed and lifetime data to estimate the lifetime distribution (Guo, 1993).

Imputing left-truncation

The samples that are left-truncated by the beginning of observation have unknown exposure history and the method proposed in Sec. 2.2 is used to impute their true values and truncation times.

Prior treatments

Weakly informative

Normal distributions that place mass over a wide but plausible range of parameter space:

$$\begin{aligned}\beta &\sim N^+(1.1, 1) \\ \eta &\sim N^+(1, 1).\end{aligned}$$

Strongly informative

A well constructed prior that encodes information about the CDF at $t_1 = 1.54$ and $t_2 = 2.71$ to encode information about the upper tail of the lifetime distribution:

$$\begin{aligned}\hat{F}_{t_1} &\sim N_0^1(0.8, 0.1) \\ \hat{F}_{t_2} &\sim N_{\hat{F}_{t_1}}^1(0.95, 0.05).\end{aligned}$$

Table 2.2: The different model-prior combinations fit to the simulated data.

3. To calculate failure times rather than lifetimes, I take the cumulative sum of the N lifetimes assigned to each unit. The installation times are calculated by taking the lag of the failure times.
4. Defining a start, $t_{start} = 5$, and end time, $t_{end} = 6$, for the observation window, any lifetimes where both the install and failure times sit either before t_{start} or after t_{end} are discarded.
5. Of the remaining lifetimes, if the install time is less than t_{start} , then t_{start} is set as the observed start of the lifetime and the observation is marked as left truncated. At the same time, if the failure time is greater than t_{end} , then t_{end} is substituted as the observed failure time, and the observation is marked as right censored.

Table 2.3 presents a simulated dataset using this method and parameter values. The simulated observations are also plotted in Fig. 2.6, where the start and end of observation are shown as red vertical lines, and the observed portions of the simulated data are shown in black and the unobserved ‘incomplete’ part is shown in grey. Notice that if we were to throw away the left truncated samples, then we would be discarding more than 50% of the sample.

2.4.2 Weakly informative prior

I now fit the imputation model to the simulated data in Table 2.3 and compare the resulting posterior with the case where the installation times of the left truncated samples are known and also with the alternative treatment of the unknown exposure histories, which is to discard all of the left truncated samples. I fit these models with the weakly informative, independent prior Gaussian priors for β and η described in the right column of Tab. 2.2 and previously mentioned in Sec. 2.3. The three Stan models are available on the GitHub repository. To sample from the posteriors, I use four chains, each 1000 iterations long with a burn-in of 500. Figure 2.7 shows the three posteriors. Plots (a), (b), and (c) in the figure show

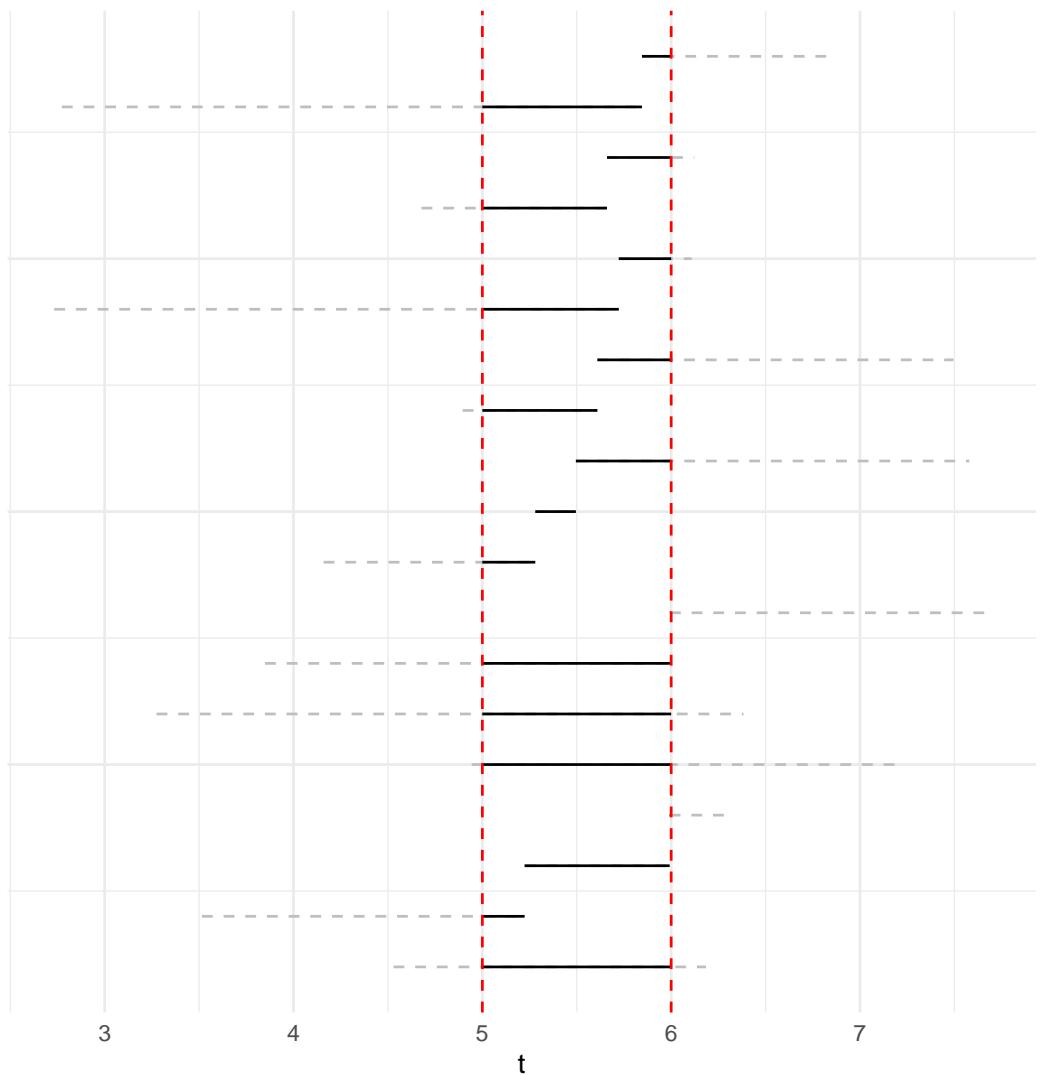


Figure 2.6: The set of simulated lifetimes data. The exposure is shown on the horizontal axis, and the beginning and end of the observation period are shown as red vertical lines. The observed portions of the lifetimes are shown as solid black lines, while the unobserved portions are shown as dashed grey lines.

Table 2.3: The simulated CMMS data for ten units.

| Unit | Lifetime | True failure | True install | Observed install | Observed failure |
|------|----------|--------------|--------------|------------------|------------------|
| 1 | 1.66 | 6.18 | 4.53 | 5.00 | 6.00 |
| 2 | 1.71 | 5.22 | 3.51 | 5.00 | 5.22 |
| 2 | 0.77 | 5.99 | 5.22 | 5.22 | 5.99 |
| 2 | 0.31 | 6.31 | 5.99 | 5.99 | 6.00 |
| 3 | 2.28 | 7.23 | 4.94 | 5.00 | 6.00 |
| 4 | 3.11 | 6.38 | 3.27 | 5.00 | 6.00 |
| 5 | 2.14 | 5.99 | 3.85 | 5.00 | 5.99 |
| 5 | 1.69 | 7.68 | 5.99 | 5.99 | 6.00 |
| 6 | 1.12 | 5.28 | 4.16 | 5.00 | 5.28 |
| 6 | 0.21 | 5.50 | 5.28 | 5.28 | 5.50 |
| 6 | 2.08 | 7.58 | 5.50 | 5.50 | 6.00 |
| 7 | 0.71 | 5.61 | 4.90 | 5.00 | 5.61 |
| 7 | 1.89 | 7.50 | 5.61 | 5.61 | 6.00 |
| 8 | 2.99 | 5.72 | 2.73 | 5.00 | 5.72 |
| 8 | 0.39 | 6.11 | 5.72 | 5.72 | 6.00 |
| 9 | 0.98 | 5.66 | 4.68 | 5.00 | 5.66 |
| 9 | 0.46 | 6.12 | 5.66 | 5.66 | 6.00 |
| 10 | 3.07 | 5.85 | 2.77 | 5.00 | 5.85 |
| 10 | 1.03 | 6.88 | 5.85 | 5.85 | 6.00 |

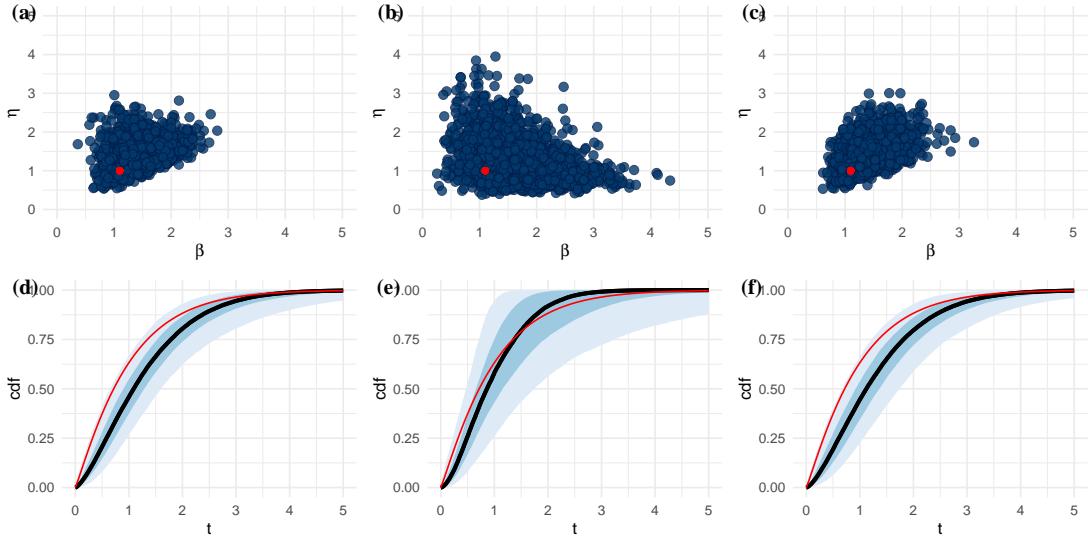


Figure 2.7: The draws from the joint posteriors conditioned on the simulated dataset when the left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed and a weakly informative prior is used. (d), (e), and (f) show the corresponding uncertainty around the CDF (in the form of the 0.5 and 0.8 uncertain intervals) that result from (a), (b), and (c), respectively. The true parameter values and CDF are shown in red.

the joint draws of the two Weibull parameters for the fully observed, discarded, and imputed treatment of the left truncated samples, respectively. The corresponding CDFs with uncertainty are shown in (d), (e), and (f). The true values of the parameters and the true CDF are plotted in red in the plots. The resulting inference from imputing the installation times of the left truncated samples is almost the same as if we had fully observed the left truncated samples. Discarding the left truncated samples results in a much more diffuse posterior and, consequently, more uncertainty around the CDF, particularly for larger exposure times. Next, I refit the models using a the version of the informative joint prior shown in Fig. 2.5 (c) and (f), which informs the upper tail of the distribution.

2.4.3 Strongly informative prior

The prior in Fig. 2.5 (c) and (f), which elicits information about the CDF at $t_1 = q_{0.80} = 1.54$ and $t_2 = q_{0.95} = 2.71$ strongly informs the Weibull model

in the upper tail of the distribution but is sufficiently vague in the lower tail, where the data are strongly informative. Here, I refit the Weibull models with a slightly more diffuse version of the prior, i.e., I increase the uncertainty around the estimates of the CDF at each elicitation time by using a larger value of the standard deviation. The informative joint prior is now

$$\begin{aligned}\hat{F}_{t_1} &\sim N_0^1(0.8, 0.1) \\ \hat{F}_{t_2} &\sim N_{\hat{F}_{t_1}}^1(0.95, 0.05).\end{aligned}$$

Using the informative prior, I refit the three models from Sec. 2.4.2. Once again, I use four chains, each 1000 iterations long, with a burn-in of 500 to sample from the three posteriors. The resulting draws are plotted in (a), (b), and (c) of Fig. 2.8 for the fully observed, discarded, and imputed treatment of the left truncated samples, respectively. Comparing the posterior of the imputation method (plots (c)) with the case where the left truncated samples are fully observed (plots (a)), the resulting inference is again very similar. However, using the informative prior, the posterior of the case where I have discarded the left truncated samples, plot (b), is much closer to the other two posteriors than when a vague prior is used. The corresponding posterior CDFs are shown in (d), (e), and (f) of Fig. 2.8. The uncertainty in all cases has been refined, but most drastically in the upper right part of the CDF. In Fig. 2.8 (e), the effect of combining a likelihood that informs the CDF at lower exposure times with a prior that informs the CDF at higher exposure times has resulted in a more precise estimate of the CDF.

In the implementation of the joint prior that I have used, the prior is properly updated through the MCMC routine. i.e. the observed data have slightly updated our belief about the value of the CDF at both t_1 and t_2 . Figure 2.9 compares the distribution of the posterior draws of $\hat{F}(t_i)$ (grey densities) with the prior (red curves). In the three plots in Fig. 2.9, the posterior distributions are clearly narrower than the prior, more so at t_1 . Fitting the Weibull models to the data in Tab. 2.3 has shown that, for small sample sizes, in cases where we do not know the

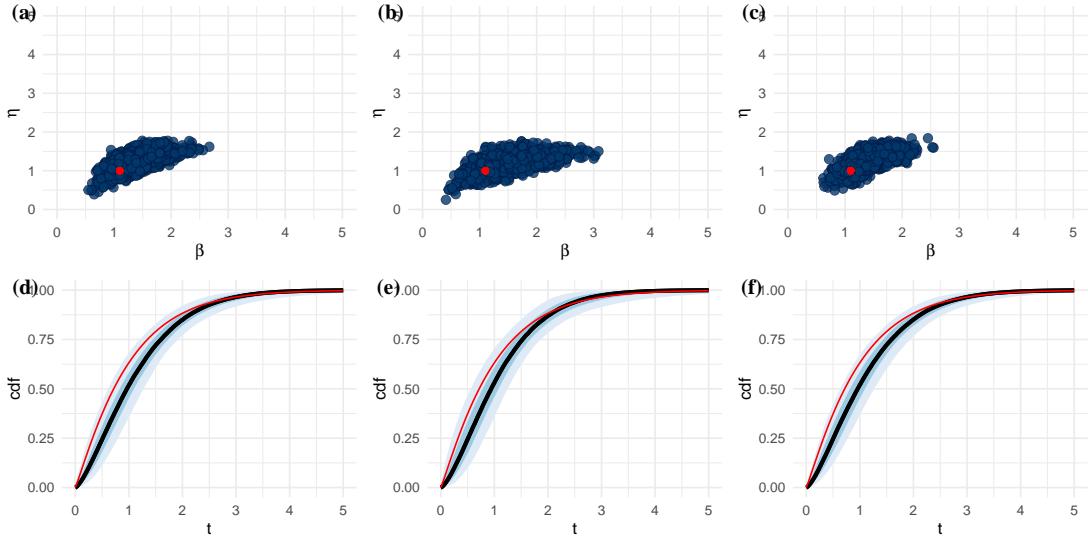


Figure 2.8: The joint posteriors conditioned on the simulated dataset when the left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed and an informative joint prior is used that encodes information into the upper tail of the lifetime distribution. (d), (e), and (f) show the corresponding uncertainty around the CDF (in the form of the 0.5 and 0.8 uncertain intervals) that result from (a), (b), and (c), respectively. The true parameter values and CDF are shown in red.

exposure history of the left truncated samples the imputation method provides almost equivalent inference to if we had fully observed the left-truncated samples. Furthermore, if we alternatively ‘throw-away’ the left truncated observations, this results in a large loss of information; however, this loss can be compensated for if there is prior information about longer lifetimes and this information is properly encoded into a joint prior. Here, I have shown one case for a small sample. Next, I devise a small simulation experiment to more rigorously evaluate the imputation method for the left truncated samples with unknown exposure history.

2.5 Simulation experiments

To explore the behaviour of the imputation method for left truncated lifetimes with unknown exposure history, I repeat the simulation and model fitting process of Sec. 2.4 for a number of different combinations of the simulation parameters:

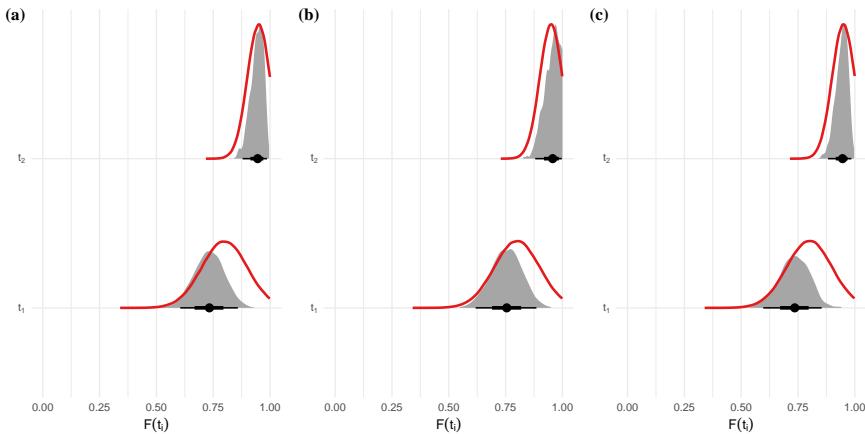


Figure 2.9: Comparison of the marginal prior and posterior for F_{t_1} (top) and F_{t_2} (bottom) when left-truncated lifetimes are (a) fully observed, (b) discarded, or (c) imputed to show how both the elicited distributions have been updated in the posterior.

number of units, start of observation time, and length of observation time, once again comparing the imputation method—under both a vague and informative prior—with the alternative method of discarding the left truncated observations as well as with the case where the left truncated samples are fully observed (as ground truth). In this section Sec. 2.5.1 defines the factor levels of each simulation parameter that make up the factor combinations in the simulation experiments; Sec. 2.5.2 describes the measures of model accuracy that I use to compare the three models; and Sec. 2.5.3 presents the results of the simulation experiments.

2.5.1 Factor levels

I vary three factors when simulating the datasets, each with three levels:

N : 10, 100, 500 units

t_{start} : 1, 5, 15 mean lifetimes

$(t_{end} - t_{start})$: 1, 3, 6 mean lifetimes.

In total, the three three-level factors, result twenty seven factor combinations of the simulation parameters. Increasing the number of units, N , increases the

sample size; increasing t_{start} increases the range of possible values the left truncated samples with unobserved truncation time could take (and therefore reduces the information they contribute); and increasing the window size, $t_{end} - t_{start}$, increases the number of fully observed lifetimes, the maximum length of the fully observed lifetimes and the number of samples that are both left truncated and right censored. For each factor combination, I perform one hundred simulations, each time fitting all three model treatments for the left-truncated observations with both prior treatments—the model and prior treatments are the same as the ones I used in Sec. 2.4 and described in Tab. 2.2. The flow of the simulation experiments is described in Alg. 1. In total, there are one-hundred and sixty-two scenarios of simulation factor-combinations and model-prior treatments each repeated one-hundred times.

Algorithm 1 Structure of the simulation experiments.

- 1: **for** each factor combination of the simulation parameters, of with there are $(3 \times 3) = 27$. **do**
 - 2: Simulate 100 datasets using the simulation parameters.
 - 3: Fit each of the model-prior treatments—three model treatments and two prior treatments—to each of the datasets.
 - 4: Calculate the Accuracy measures for each model fit.
 - 5: **end for**
-

2.5.2 Accuracy measures

To compare how well the Bayesian models recover the true data generating mechanism, I calculate the Bayesian p -values of the true parameter values, $\beta = 1.1$ and $\eta = 1$, from the fitted posteriors and the elppd of a dataset of one hundred fully observed lifetimes generated from the true Weibull distribution. The Bayesian p -value for beta is easily calculated using the posterior draws as (Gelman, Carlin, et al., 2020)

$$\Pr [\beta \geq \beta_{true} | y] = p_\beta = \frac{\sum_{s=1}^S \beta_{true} \leq \beta_s}{S},$$

where β_{true} is the true value of β used to simulate the data. The p -value of η can be calculated using the same expression with η in place of β . If the model posterior is well calibrated, then the p -values from repeating the same simulation should have a uniform distribution (Stan Development Team, 2024; ?)³. If the parameter estimates are unbiased, then the average p -values of the one-thousand repeated experiments for a factor combination will sit at 0.5. If the posterior uncertainty is overly diffuse, then the distribution of the one-thousand p -values will be concentrated around 0.5, and if the uncertainty is too narrow, then the p -values will be pushed towards the boundaries (zero or one). The p -value provides an indication that the model posterior is well calibrated and if the parameter values are under or over-predicting, but does not express to what degree.

To determine the scale of any discrepancy, I also calculate the elppd of a new sample of one hundred fully observed lifetimes under the different posteriors. I do this by simulating one hundred datasets, each with one hundred observations and calculating the expected log-likelihood of each simulated dataset and then taking the average:

$$\text{elppd}_{100} = \frac{\sum_{n=1}^{100} \frac{1}{S} \sum_{s=1}^S \sum_{i=1}^{100} p(\tilde{y}_{n,i} | \beta_s, \eta_s)}{100}.$$

2.5.3 Results

Figures 2.10 and 2.11 show the results of the simulation experiments. The simulation experiment results show that under some circumstances, there is a slight bias in the posterior estimates of the parameters when the partially observed left-truncated lifetimes are imputed. There are two separate causes of the bias: the first comes from the improper treatment of lifetimes that are both left truncated by the beginning of the observation period and right censored

³This manuscript and section *Simulation-Based Calibration* from the Stan user guide discuss the concept of simulation-based calibration. What I do is not quite full simulation-based calibration, since I do not simulate the true values of the parameters from the prior, but many of the ideas should still hold.

by its end, and so only impacts inference when the observation period is short ($t_{end} - t_{start} = 1$ mean lifetime = 0.95); the second occurs because the uncertainty expressed in the posterior is an expression of *our* uncertainty and so is not equivalent to frequentist confidence intervals, and so the p -values appear biased. The scale of this second source of bias appears less extreme. Despite these two sources of bias, elppd results suggest that in some cases, it is better to use the slightly biased treatment in favour of discarding the left-truncated lifetimes because the precision of the posterior places more mass around the true data-generating mechanism, particularly if a weak prior is used.

Figure 2.10 summarises the p -values of the 100 simulation runs under each factor combination by their means. The different treatments of the left-truncated lifetimes—fully observed, imputed, or discarded—are shown as different colours and the type of prior—informative or vague—is shown as different point types, for example, fully observed left-truncated lifetimes and a strongly informative prior is shown as a blue dot. If the model’s posterior is unbiased, then the point should sit in the centre of each plot, i.e. an average p -value of 0.5 for both parameters. If the average of the p -values is at one or zero, then there is typically very little posterior mass around the true parameter value, and hence, we can conclude that the inference is biased. When $N = 10$, the bias in parameter estimates is negligible compared to the uncertainty in the posterior, so all of the different treatment-prior combinations perform similarly. However, as the number of units increases and the sample size gets larger, the posteriors become more precise, and so the bias becomes more obvious.

The leftmost column in Fig. 2.10 shows the simulations where the window size is equal to one mean lifetime, $t_{end} - t_{start} = 0.95$. In this case, there is a high proportion of lifetimes that begin before the start of the observation period and end after the observation period—i.e. they are both left-truncated at the beginning and right-censored at the end. For these cases, the Bayesian p -value of the shape parameter, β , is greater than 0.5 and depending on the value of t_{start}

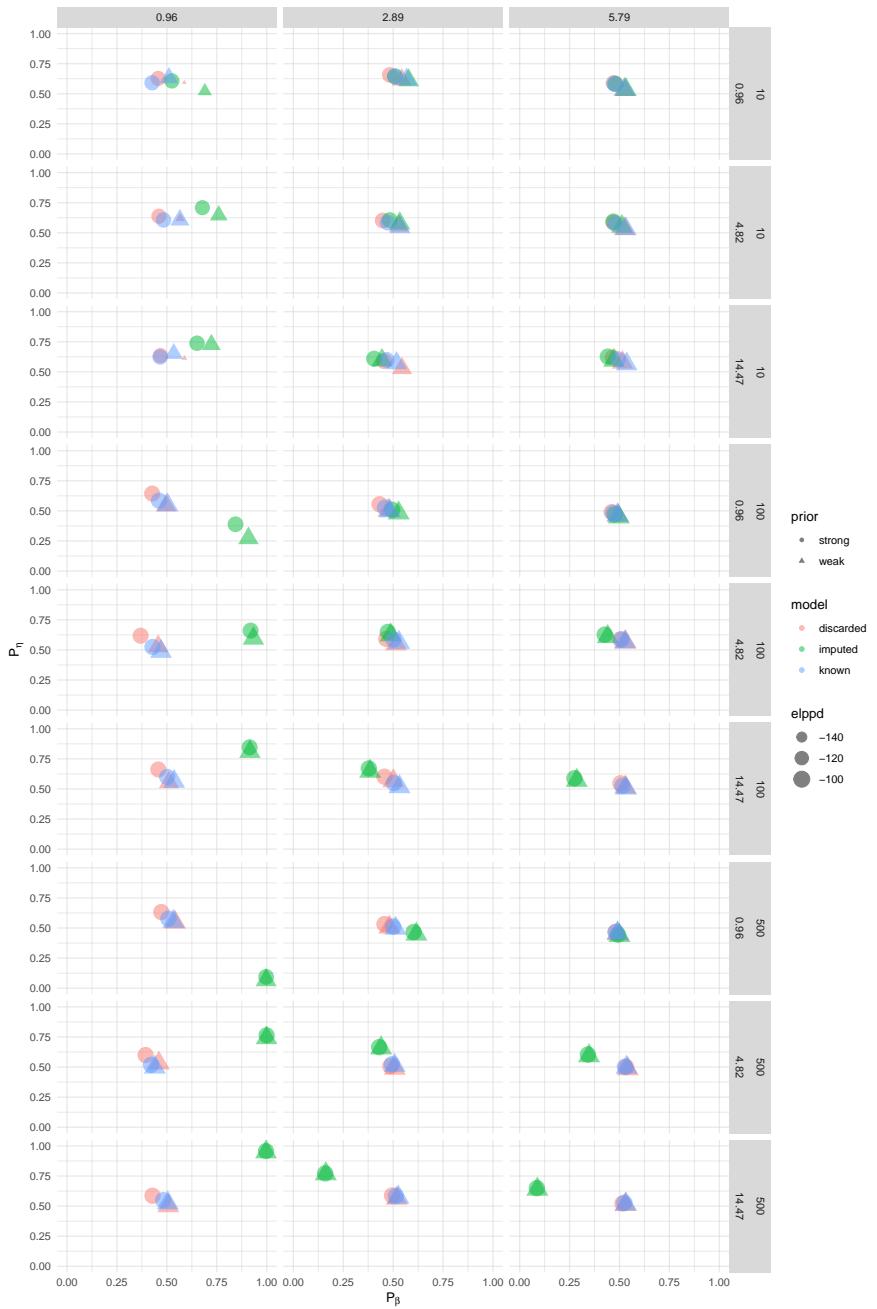


Figure 2.10: The simulation results are summarised by the mean of the one hundred repetitions for each factor combination. The columns show the three levels of $t_{end} - t_{start}$, and the rows show the levels of t_{start} and N . The Bayesian p -values of β and η are reported on the horizontal and vertical axis, respectively, and the colours of the points show the treatment of the left-truncated samples, and the shape shows the prior. The size of the points correspond the the average elppd_{100} calculated according to eq. (2.5.2)

the scale parameter η is either over ($p_\eta > 0.5$) or under ($p_\eta < 0.5$) estimated. Bias when t_{start} is small is expected since the assumption of uniform entry I make in the model in (2.13) is not valid since there is a higher chance that lifetimes were installed at t_0 . However, the bias in the cases where $t_{start} \ll 0$ is unexpected. Inspecting the posteriors more closely shows that the HMC algorithm is updating the random variable τ^L in eq. (2.13), where it should instead be treated as a nuisance parameter and excluded from the updating procedure. This occurs because of the way the HMC algorithm in Stan is implemented: it is not possible to cut the flow of information⁴. For the cases where the $t_{end} - t_{start} > 1$ mean lifetime, the proportion of lifetimes that are both left truncated and right censored is low, and hence, so is the bias that results from the treatment of these observations.

In Fig. 2.10, when $t_{end} - t_{start} > 1$ (the two right-most columns), the posterior where the left-truncated lifetimes are imputed is more similar to the posterior from the other two treatments. However, as the number of units is increased, a slight underestimation of the shape parameter β becomes apparent when $t_{start} \gg 0$. However, even for the factor combinations where $N = 500$ and $t_{end} - t_{start} = 6$ mean lifetimes—the combination that results in the largest sample size and hence the narrowest posterior—the average p -value of β is still not zero, showing that even when the posterior is extremely precise it still generally contains the true parameter value. Hence, this second apparent bias is more a result of the uncertainty that we have encoded into the model, which does not align with the frequentist interpretation of uncertainty intervals that p -values measure. When t_{start} is large, the upper bound of the partially observed left-truncated lifetimes is increased. Extending the support of these random variables means that the upper tails become heavier, and so too does the posterior of the Weibull shape parameter. This second form of bias is only noticeable when the number of units is very large ($N = 500$), and so, in most cases, it would be unnoticeable over the

⁴The Stan user forum has some commentary on the matter [here](#).

uncertainty in the posterior.

Despite both sources of bias, the elppd of one hundred new observations generated from the true Weibull model is very close for all treatment-prior combinations. In Fig. 2.10, the size of the points indicates the average values of elppd_{100} calculated according to eq. (1.11) and from this we can see that the only noticeable difference between the three treatments of the left truncated data is that when there is a short period of observation, $t_{end} - t_{start} = 0.95$, then when there are few units, $N = 10$, discarding the left-truncated observations has a large impact on the elppd if analysis is not supplemented with an informative joint prior. This has already been shown in Fig. 2.7 (b) and (e). Figure 2.11 compares the average elppd_{100} scores under the different factor combinations more closely. In the figure, columns show the different levels of $t_{end} - t_{start}$ and rows show the different levels of t_{start} ; the levels of N are plotted on the horizontal axis, and the treatment of the left truncated samples and the type of prior are shown by the colour and shape of the point respectively. The plotted elppd values show that when there are few units or a short period of observation, the method of imputing the partially observed values of the left-truncated samples results in an elppd_{100} score that is closer to the fully observed treatment than if we were to discard the partially observed left-truncated lifetimes, except when $t_{start} = 1$ mean lifetime. The fact that it is hard to distinguish between the approaches based on the elppd_{100} score, despite the bias, shows that the scale of the bias is typically inconsequential.

2.6 Concluding remarks

In this chapter, I have shown that when censored observations are treated as missing and their values imputed—rather than integrated out—observations that are left truncated with unknown exposure history can be treated as a case of interval censoring and as a result, their partially observed value and truncation time imputed through Bayesian analysis. Doing so retains the information in the

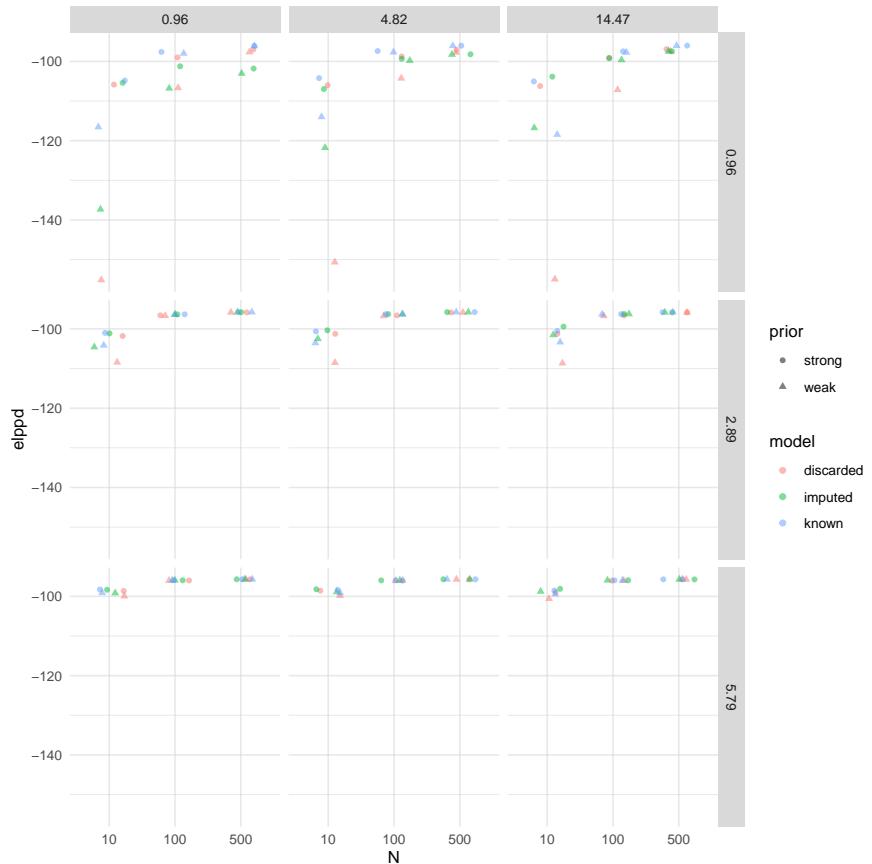


Figure 2.11: The elppd₁₀₀ simulation results summarised by the mean of the one hundred repetitions for each factor combination. The columns show the different levels of $t_{end} - t_{start}$ and rows the different levels of t_{start} . The number of units N is shown on the horizontal axis, and the average elppd₁₀₀ score is shown on the vertical axis. The colours of the points show the treatment of the left-truncated samples and the shape shows the prior.

left truncated samples, resulting in much more precise estimates than the alternative option, which is simply to discard the left truncated lifetimes. I also extend on the method proposed by Kaminskiy and Krivtsov (2005) for constructing a joint prior for the Weibull parameters by demonstrating how the choice of the two elicitation times, t_1 and t_2 , dictates where in the lifetime distribution information is encoded—i.e., in the lower tail or upper tail—and implement the prior in a fully Bayesian model for left truncated and right censored lifetime data in a way that ensures the joint draws from the informative prior are properly filtered through the likelihood, and hence our prior belief around the CDF at both t_1 and t_2 is updated in the posterior.

I used a simulation example to demonstrate the method to impute left truncated lifetimes with unknown exposure history and their corresponding truncation times alongside the alternative case where the left truncated samples are discarded and the case where the left truncated samples are fully observed. I fit all three cases with both a vague and informative joint prior. When a vague prior is used, imputing the left truncated lifetimes with unknown exposure history retains the extra information in these samples, resulting in a much more precise posterior than when the left truncated samples are discarded and roughly the same inference as when the lifetimes are fully observed. When a carefully constructed joint prior is used, which encodes information in the upper tail of the lifetime distribution, the posterior where the left truncated samples have been discarded is much closer to the other two posteriors—where the left truncated lifetimes are fully observed or imputed.

Finally, I performed a small simulation experiment to explore the two approaches of imputing or discarding the left truncated samples with unknown exposure times—with and without an informative prior—under different simulation scenarios. The simulation experiments show that imputing the left truncated lifetimes and truncation times results in a slight bias in the parameter estimates, but in most cases this bias is small relative to the uncertainty in the posterior.

There are two sources of bias. The first comes about because the treatment of observations that are both left truncated/interval censored by the start of the observation period and right censored by its end. For these lifetimes, I make the assumption that the entry time is uniformly distributed by assigning a uniform prior to the parameter τ^L . However, this parameter is included in the updating procedure of the HMC algorithm. To remove the bias, the model should be re-implemented either in an alternative probabilistic programming language such as BUGS, where nuisance parameters that are not updated can be included in the MCMC procedure, or using a custom MCMC algorithm for the model. However, this is only needed if the observation period is small enough to result in a large proportion of lifetimes that are left truncated and right censored.

The second source of bias arises because imputing the left truncated lifetimes and their truncation times results in uncertainty intervals that are not equivalent to frequentist confidence intervals but rather encode *our* uncertainty. This bias is small, and the true value of the parameters is still contained in the posterior. However, in all cases, the elppd of a new set of one hundred fully observed lifetimes from the true data-generating mechanism are very close, showing that the scale of the bias is relatively negligible.

Recommendations Based on the learnings from this chapter, it should be suitable to use the proposed approach for data that are left truncated with unknown installation times, so long as the observation period is sufficiently long. However, suppose there is a strong likelihood—i.e., a very large sample size—and the analyst requires a very accurate estimation of the parameter values. In such a case, it is better to discard partially observed left truncated lifetimes and, if information is available to do so, construct an informative joint prior that informs the upper tail of the distribution. When the likelihood is weaker, using the imputation method results in a posterior that is almost equivalent to the case where the installation times of the left-truncated lifetimes are known.

In the next chapter, I use the methods proposed here and the learnings of

the simulation experiments to analyse the idler-frame dataset and show that when lifetime analysis is performed in the Bayesian framework and the censored observations are imputed, MCMC sampling naturally provides estimates and uncertainty intervals for the failure times of units still in operation, the expected number of failures in the next short time interval, and the cost per unit time of a preventative replacement strategy.

Chapter 3

Idler-frame case study

In this chapter, I apply the methods and learnings from Chap. 2 to an industry dataset of the failure times of idler-frames on an overland iron ore conveyor. The idler-frames were described briefly in Sec. 1.3. For a reliability engineer tasked with maintaining the conveyor, it can be useful to quantify the expected failure times of the idler-frames currently in operation and the expected number of failures in the next short time interval, such as Hong et al. (2009) do for power transformers. I show how these quantities are already naturally contained in the full posterior of the Bayesian model when the censored lifetimes are imputed as in Sec. 2.1.3. It is also useful to propagate uncertainty in the posterior estimates of the parameters through any decision criteria to understand risk in long-term maintenance plans, such as the design of a fixed-time replacement strategy. I demonstrate how the joint posterior draws of the Weibull parameters can be propagated through a cost function to make an informed decision about a fixed-time replacement strategy for the idlers in the frames.

The chapter is structured as follows. Section 3.1 describes the data in detail. In Sec. 3.2, I construct an informative prior for the idler-frame analysis based on prior knowledge supplied by the idler manufacturer and a conveyor engineer. Section 3.3 describes the model fitting process and posterior inference on the parameters. Section 3.4 goes on to demonstrate how the draws from the posterior

can be used. Specifically, Sec. 3.4.1 and 3.4.2 show how to obtain the estimated remaining useful life of the currently installed idler-frames and the expected number of failures in a time interval following the end of observation, respectively. Sec 3.4.3 then shows how to propagate the posterior uncertainty about the lifetime distribution through utility functions—specifically a cost function—to inform the choice of a preventative replacement interval for the idler-frames. Finally, Sec. 3.5 summarises and concludes the chapter.

3.1 Idler-frame lifetime data

The data is a synthesis of preventative and reactive replacement records of idlers on a single overland iron ore conveyor. It is one of many such datasets for similar conveyors on the mine. This specific conveyor has one hundred and forty-three frames of idlers, each with a three-idler configuration. When an idler in the frame fails, all three of the idlers are typically replaced, so each idler frame can be viewed as one unit. The replacements of the idlers in a frame are typically captured in the CMMS, and if an idler in the frame has been observed as failed and scheduled to be replaced, then this information is included in the replacement record. However, if an idler in the frame has failed in a way that threatens to damage the belt immediately, then this is raised in a different system, the belt is shut down, and the idlers in that frame are replaced. During an industry placement, I cleaned and collated these different sources of replacement records into a single dataset. The dataset spans just over six years, but the conveyor has been in operation for twenty.

From the replacement records, the lifetimes of the idler-frames can be calculated as the time between the replacements. However, since the records do not go back to the commissioning of the conveyor, the first observed lifetime for each frame is left truncated and has unknown installation time. Because the sets of idlers in some frames are preventatively replaced or because some were still in operation when I constructed the dataset, there are many right censored lifetimes.

Table 3.1: Summary of the idler frame data set.

| | |
|--|------------|
| Maximum lifetime | 2167 days |
| Minimum lifetime | 1 days |
| Maximum fully observed lifetime | 1461 days |
| Beginning of observation | 2014-12-10 |
| End of observation | 2020-11-15 |
| Number of observations | 402 |
| Number of unique frames | 143 |
| Number of left truncated observations | 143 |
| Number of right censored observations | 144 |
| Number of left truncated and right censored observations | 1 |

Table 3.1 gives an overview of the dataset of idler frame lifetimes, and Fig. 3.1 shows the lifetimes of each frame along the conveyor. In Fig. 3.1, the fully observed lifetimes—i.e. we observed the failure of an idler in the frame as well as the previous failure for that frame—are shown as orange points, while the partially observed lifetimes are shown in blue. The partially observed lifetimes that are left-truncated by the beginning of the observation period are shown as triangular points, while the remainder of the blue points are right-censored observations.

Table 3.1 shows that roughly thirty-six per cent of the lifetimes in the dataset are left-truncated with unknown installation times. Therefore, if we were to discard these observations, we would throw away a third of the data. It is better to retain the information in these left-truncated lifetimes.

Based on our understanding of how the idler-frames fail and from the characteristics of the dataset described in Tab. 3.1, it is appropriate to use the methods that I developed in Chap. 2 to analyse the industry dataset and impute the partially observed truncated lifetimes. The Weibull distribution appears appropriate to model the idler-frame lifetimes since the idlers in the frame should fail via a wear-out failure mechanism. Furthermore, the frame’s failure time is when the first idler in the frame fails, which aligns with the extreme value distribution characteristic of the Weibull (discussed in Sec. 2.1.2). The observation period spans just over six years, and it starts roughly 14 years after the conveyor was

commissioned. According to domain knowledge, the expected lifetime of an idler is roughly five years, and since the lifetime of a frame is the first of the three idlers to fail, the expected lifetime of a frame should be a little under this value. Hence, the observation window is greater than the expected lifetime, and the time from the commissioning of the conveyor ($t = 0$) to the start of the observation period (t_{start}) is roughly three times greater. Furthermore, only one lifetime is left-truncated/interval-censored by the beginning of observation and right-censored by its end. Under these circumstances, it is acceptable to use the methods proposed in Chap. 2 to account for the lifetimes that are left-truncated with unknown exposure history.

In the industry dataset, there are some very short lifetimes—twenty-five that are less than three weeks—that most likely arise from manufacturing defects or incorrect installation. In this analysis, I aim to model the wear-out failure mechanism of the idler frames, and therefore, I treat any lifetimes shorter than three weeks as right censoring events; this is done by Hong et al. (2009) in their analysis of power transformers—i.e. the failure due to wear was right censored by the early failure due to another cause.

3.2 An informative prior

Based on engineering knowledge and information provided by the manufacturer, I construct an informative joint prior for the idler-frames to supplement the analysis. According to the manufacturer, the expected lifetime of an idler is five years, and according to conveyor engineers, it is unlikely that they will last longer than eight. I express this information as the expectation of the CDF at $t_1 = 5 \times 365 = 1825$ and $t_2 = 8 \times 365 = 2920$ encoded as normal distributions, as in Sec. 2.3. The expected value of the CDF at $t_1 = 1825$ is 0.50, and the standard deviation is 0.15, while the expected value of the CDF at $t_2 = 2920$ is 0.95 with a standard deviation of 0.05. I set the standard deviations relatively large since the expected lifetime of a frame of idlers is actually the expected smallest value

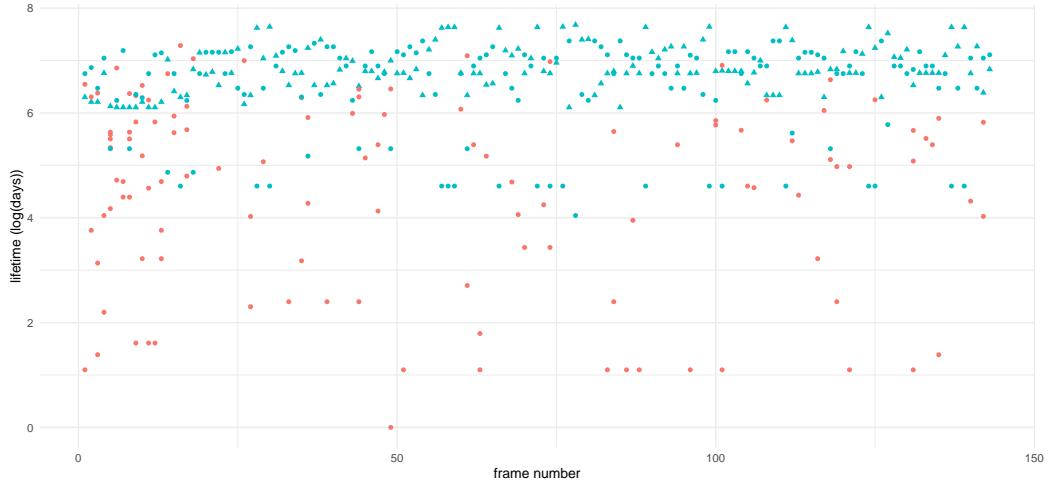


Figure 3.1: The idler-frame lifetimes plotted along the length of the conveyor. On the horizontal axis is the frame number that the lifetimes belong to, and on the vertical is the log value of the lifetime in log-days. The fully observed lifetimes are red points, while the partially observed (censored) lifetimes are blue. The censored lifetimes that are left-truncated by the start of the observation period are shown as triangular points.

of the three idlers in the frame. The resulting informative joint prior is shown if Fig. 3.2, plot (a) shows three thousand joint draws of the shape β and scale η from the informative prior, and plot (b) shows the resulting prior uncertainty in the CDF.

3.3 Posterior draws

To perform inference, I draw 6000 samples from the posterior using four chains, each 2000 iterations long and with a burn-in of 500 iterations and no thinning using Stan (Stan Development Team, 2022). The Stan output summarising the joint posterior draws of β and η is shown in Tab. 3.2 and the joint draws are plotted in Fig. 3.3 (a). Sampling is efficient with no divergences, the chains mix well—indicated by \hat{R} values of ≈ 1 —and both parameters have a large number of effective samples. The posterior mean of the shape is just above one (1.10), but there is a small amount of mass just below one, and the posterior mean of

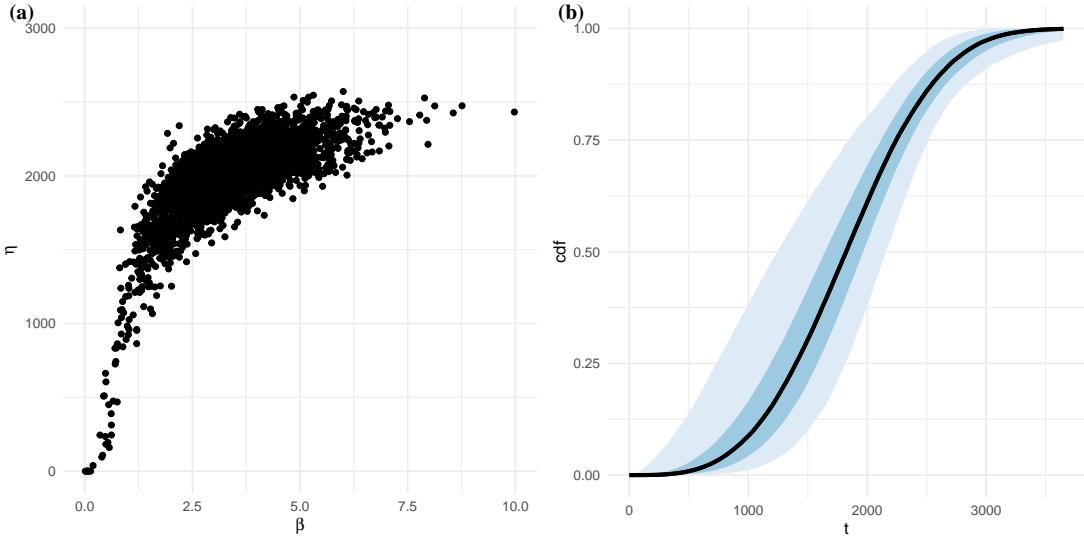


Figure 3.2: The joint informative prior from eliciting information at $t_1 = 1825$ and $t_2 = 2920$. (a) shows 3000 draws from the informative joint prior, and (b) shows the resulting uncertainty surrounding the CDF.

Table 3.2: Summary of sampling for β and η .

| Parameter | Mean | 2.5% | 50% | 97.5% | n_{eff} | \hat{R} |
|-----------|---------|---------|---------|---------|------------------|-----------|
| β | 1.10 | 1.01 | 1.10 | 1.20 | 5337 | 1.0001 |
| η | 1362.56 | 1200.58 | 1361.17 | 1535.58 | 4386 | 1.0003 |

the scale is 1363. These values of the parameters yield an average frame lifetime of 1315 days or 3.6 years with 90% lower and upper bounds of 3.2 years and 3.9 years, which is significantly smaller than the recommended average lifetime of an idler provided by the manufacturer—which was five years.

Figure 3.3 (b) compares the draws from the posterior with the informative joint prior. The posterior draws sit far in the tail of the joint prior but are still contained in the prior. The likelihood is strong enough that inference about the shape and scale are fairly invariant to changes in the prior, although I do not show this here. Figure 3.4 shows the refined uncertainty around the CDF that results from the posterior. The uncertainty surrounding the CDF of the lifetime distribution in the figures is much more refined compared with the prior (Fig. 3.2 (b)). The posterior mean of the CDF at our elicitation times— $t_1 =$

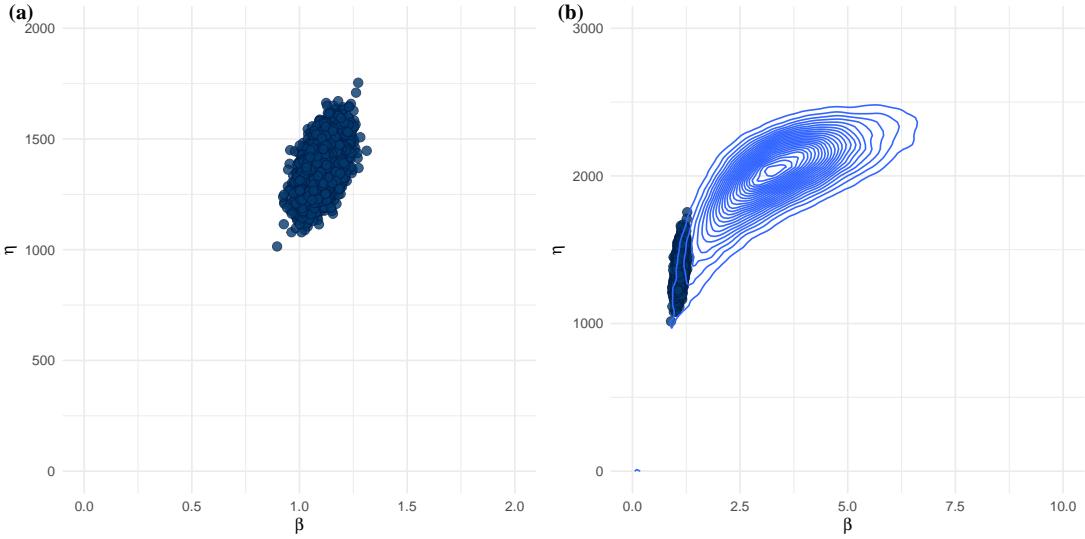


Figure 3.3: The joint draws of the Weibull parameters β and η from the posterior distribution. (a) shows the plain draws, and (b) compares the draws with the contours of the informative joint prior.

1825 days and $t_2 = 2920$ days—are 0.75 and 0.90, respectively. These updated estimates sit in the tails of the distributions I specified to construct the prior. The discrepancy between the prior and the posterior indicates a slight prior-likelihood conflict; however, the likelihood is strong enough, in this case, to not be too influenced by the prior, and the prior does appear to have some mass around the final model.

3.4 Using the posterior

While the posterior estimates of the parameters are useful in understanding the lifetime distribution, the real value in a Bayesian analysis comes from the uncertainty quantification expressed in the full posterior. Using the draws from the full posterior, which includes the draws of the latent parameters in the model, such as the imputed censored values, we can quantify risk and inform decisions. For example, by imputing the underlying values of the partially observed lifetimes during the MCMC sampling routine, we naturally obtain a distribution for their predicted failure times and consequently can also derive the expected number of

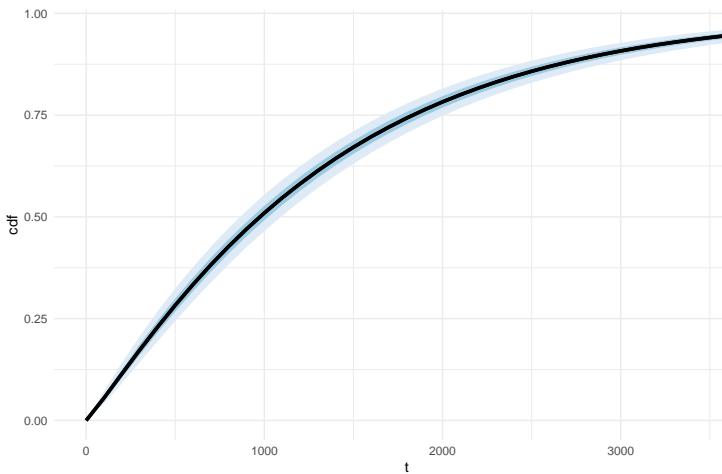


Figure 3.4: The resulting posterior uncertainty surrounding the Weibull CDF.

failures in the next short time interval. We can also pass the joint draws of the parameters through ‘utility’ functions (Gelman, Carlin, et al., 2020, Chap. 9)—for example, a cost function—to incorporate the uncertainty in the analysis into long-term decisions. In this section, Sec. 3.4.1 shows how to obtain predictive distributions of remaining useful life (RUL) for each idler-frame still in operation at the end of the observation period. I then show how these RUL distributions can be used to construct a predictive distribution for the cumulative number of failures going forward from the end of the observation period in Sec. 3.4.2. Finally, Sec. 3.4.3, demonstrates how the joint draws of the parameters can be pushed through a cost function to propagate the uncertainty from the analysis and inform the choice of a fixed-time replacement strategy.

3.4.1 Failure time of units in operation

In the model for the idler-frame lifetimes, the unobserved values of the censored lifetimes are treated as missing data and their values are imputed, which in a Bayesian framework is to treat them like a parameter in the model. Since I impute the values during the HMC routine, I consequently obtain draws of the imputed values of the censored lifetimes. The distributions of these draws form a predictive distribution missing lifetimes of the right censored units conditioned

on their age. This can also be calculated if we were to integrate out the censored lifetimes as discussed in Sec. 2.1.3 and also if we use maximum likelihood to perform inference, such as Hong et al. (2009) do. But in these cases, we would need to calculate the distribution conditional on each censored component’s age separately, and if we were to use maximum likelihood, we would need to calculate uncertainty intervals using an appropriate method. It is much more convenient to impute the values in the Bayesian analysis since the estimates and uncertainty for each censored observation are already naturally contained in the posterior.

Using the posterior draws of the imputed lifetimes, I calculate draws of the predicted remaining useful life of each unit still under test by subtracting the censoring time from the imputed value of the lifetime according to

$$\text{RUL}_i^s = \tilde{y}_i^s - c_i^{\text{Lower}}, \quad (3.1)$$

where s indicates a particular draw from the posterior. The remaining useful life distributions for each of the one-hundred and forty-three frames are shown in Fig. 3.5. These distributions can not only be used to obtain point estimates and uncertainty bounds for the RUL, but also order the frames in terms of the highest risk of failure in a particular number of days or, as I show in the next section, determine the expected number of failures within the next short time interval.

3.4.2 Expected number of failures

With each draw from the posterior, I obtain a set of imputed values for the lifetimes of the idler frames still in operation and, with this, a prediction of the remaining useful life. Using the draws of RUL, I can calculate the cumulative failures going forward from the end of the observation period. Using an adaptation of the notation of Hong et al. (2009, sec. 6), the number of failures at t days from the end of the observation period is $K = \sum_{i=1}^n I_i$, where $I_i = 1$ if $\text{RUL}_i < t$ and $I_i = 0$ if $\text{RUL}_i > t$, and the RUL_i are calculated as in eq. (3.1). I express this

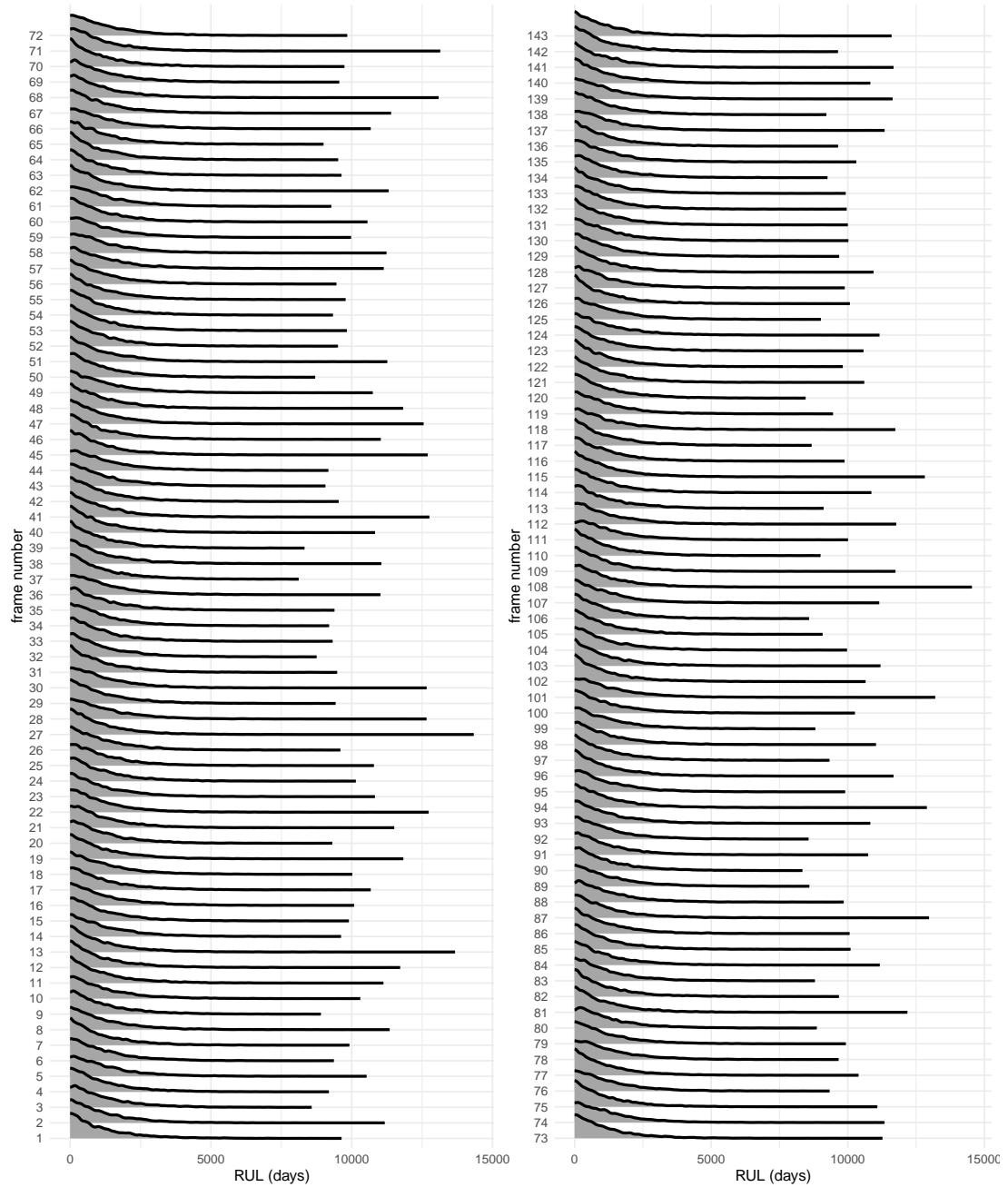


Figure 3.5: The remaining useful life distributions for the current lifetimes (the lifetimes right censored by the end of observation). The left plot shows frames 1 to 72, and the right shows frames 73 through 143.

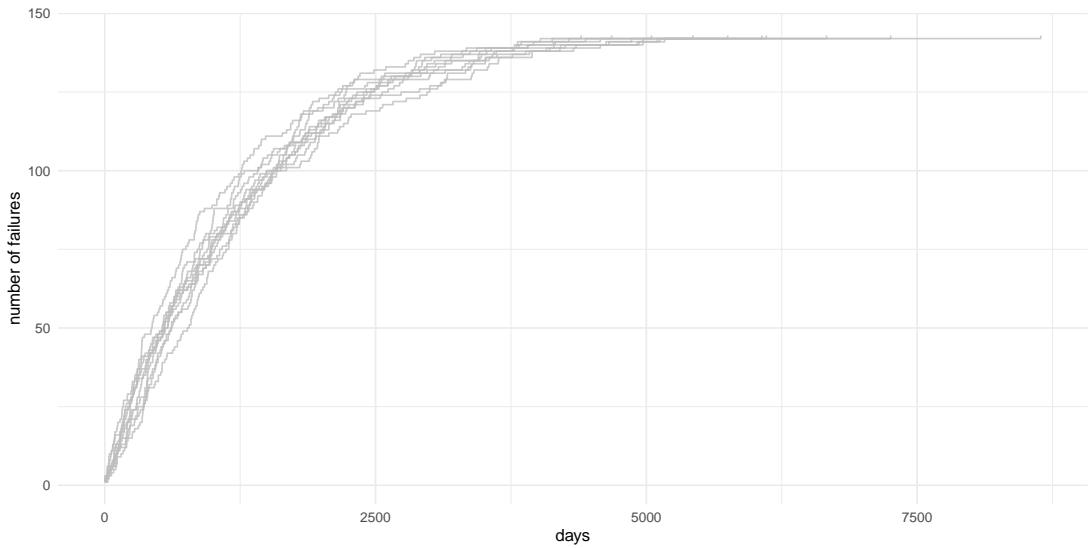


Figure 3.6: Ten draws of the cumulative failures proceeding the end of the observation period. The horizontal axis is the days since the end of the observation period, and the vertical axis is the cumulative number of failures.

step function as $F_K(t|\text{RUL}^s, \theta^s)$ where $\text{RUL} = \{\text{RUL}_1, \dots, \text{RUL}_n\}$. Figure 3.6 shows ten draws of $F_K(t|\text{RUL}, \theta)$.

Using the resulting distribution of $F_K(t|\text{RUL}, \theta)$, predictive distributions for the number of failures at any time t in the future can be obtained. Figure 3.7 shows the predictive densities for the number of failures within the next one (Fig. 3.7 (a)), three (Fig. 3.7 (b)), and six (Fig. 3.7 (c)) months. From the draws that make up the distributions shown in the figure, it is straightforward to obtain the posterior estimates and uncertainty intervals. Such distributions are useful for managing risk in the short term and planning, for example, managing the inventory of replacement idlers kept on site.

3.4.3 Cost functions and preventative policy design

The posterior draws are not only useful for making short-term decisions, functions of the draws of the parameters can inform long-term planning. As an example, in this section, I incorporate the posterior uncertainty in the parameters into the choice of fixed time replacement interval for a preventative maintenance policy.

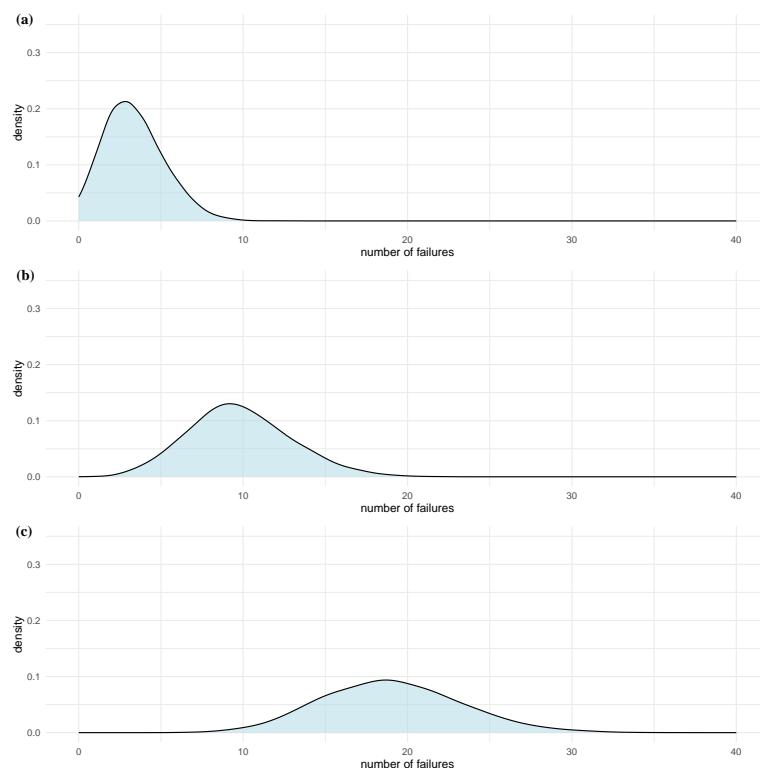


Figure 3.7: The predictive distributions for the number of failures in the one (a), three (b), and six (c) month period after the end of the observation period.

To implement a preventative replacement strategy, a reliability practitioner must choose a fixed time interval at which components are replaced. This choice aims to minimise the expected cost of maintenance and lost production but must also align with what is practically possible. Jardine and Tsang (2013) propose a rough structure for quantifying the cost of a particular preventative maintenance policy according to

$$\text{Cost per unit time} = \frac{C_p N + C_r E[K|\Delta t_p, \theta]}{\Delta t_p},$$

which balances the cost of planned replacements C_p against the cost of unplanned (reactive) replacements C_r in terms of cost per unit time. N is the number of units covered by the scheme, Δt_p is the proposed fixed time interval, and $E[K|\Delta t_p, \theta]$ is the expected number of reactive maintenance events in the interval given the parameters of the lifetime distribution. Since the cost due to reactive maintenance depends on the lifetime distribution, we should incorporate the uncertainty around the parameters of this distribution.

Here, I give a general example of propagating posterior uncertainty through a cost function for the idler-frames bulk replacement. Let Δt_p denote the time interval between bulk replacements of the idler-frames analysed in Sec. 3.3. The cost to replace an idler in a bulk replacement during a shutdown period is $\approx \$250$ per idler, that is $\approx \$750$ per frame. However, if an idler needs to be replaced during a period of operation, this costs $\approx \$2,000$ in parts and labour and typically takes around two hours. Critical conveyors in the circuit will transport around 15000 tonnes/hour. At a conservative iron ore price of \$100 a tonne, a two-hour stoppage equates to \$3,000,000 in lost production. The cost function for the bulk replacements of the idlers is therefore

$$\text{Cost per unit time} = \frac{\$750 \times 143 + (\$3,000,000 + \$2,000)E[K|\Delta t_p, \beta, \eta]}{\Delta t_p}.$$

Here, because the expected number of unplanned replacements in the inter-

val, $E[K|\Delta t_p, \beta, \eta]$, depends on the posterior values of the parameters β and η , $E[K|\Delta t_p, \beta, \eta]$ is a random variable, and as a result, so too is the *cost per unit time*. Samples from the distribution of *cost per unit time* are obtained by calculating $E[K|\Delta t_p, \beta^s, \eta^s]$, conditioning on each draw from the posterior. However, It is slightly more complicated to calculate this value for the idlers than in the examples provided in Jardine and Tsang (2013).

Expected number of reactive maintenance events The expected number of reactive maintenance in an interval is non-trivial for two reasons. First, the idlers are repeatedly replaced when they fail, and secondly, not all idler-frame failures will need to be replaced immediately, only those that fail in a way that poses an immediate threat to the belt. The expected number of failures, therefore, needs to be approximated numerically. Section 2.4.3 of Jardine and Tsang (2013) discusses some methods to determine the expected number of repeat failures in a time interval for more straightforward cases and here I describe a method for the idler frames when not all unplanned failures result in reactive maintenance and inflated costs.

In the idler-frame problem, planned and non-urgent maintenance is only performed during shutdowns every six weeks. I therefore confine my simulations to discrete values of Δt_p , which are multiples of six weeks, ranging from one to forty-four shutdowns (five years). For each draw of β and η from the posterior, I run one thousand simulations, iteratively simulating the failures in the periods between shutdowns, going up to forty-four. This process is described in Alg. 2. Each simulation run reflects the typical workflow on the mine; frames fail in the period between shutdowns and, unless they are an immediate threat to the belt, are flagged and maintained in the next shutdown. However, if an idler in the frame fails in a way that immediately threatens to damage the belt, then the conveyor is stopped and the idlers in that frame replaced, which is when the high reactive maintenance cost is incurred. Therefore, in the simulation runs, starting with all frame ages of zero, I simulate the number of failures in the interval

between shutdowns by summing N independent Bernoulli trials

$$I_n | \Delta t_{shuts}, t_n, \beta^s, \eta^s \sim \text{Bernoulli}(P_n)$$

$$K_{\text{UF}} = \sum_{i=1}^N I_i$$

where $P_{\text{UF}n}$ is calculated according to the Weibull lifetime distribution

$$P_{\text{UF}n} | \Delta t_{shuts}, t_n, \beta^s, \eta^s = \frac{F_{\text{Weibull}}(t_n + \Delta t_{shuts} | \beta^s, \eta^s) - F_{\text{Weibull}}(t_n | \beta^s, \eta^s)}{1 - F_{\text{Weibull}}(t_n | \beta^s, \eta^s)}.$$

t_n is the age of the frame in days at the beginning of the interval, and $\Delta t_{shuts} = 6 * 7$ days is the period between shutdowns. During the operational periods, I assign a probability $P_{\text{RM}} = 0.05$ that an unplanned frame failure must be replaced immediately. The number of reactive maintenance events in a period between two shutdowns is, therefore, a binomial trial with sample size K_{UF}

$$K_{\text{RM}} | \Delta t_{shuts}, t_n, \beta^s, \eta^s \sim \text{Binomial}(K_{\text{UF}}, P_{\text{RM}}).$$

At the end of each step (the period between shutdowns), I reset the age of any failed frames to zero, progress the surviving frames ages by t_{shuts} and then move to the next step and repeat the same process.

In each run, at the end of the forty-four shutdowns, I calculate the cumulative number of reactive maintenance events, and at the end of the one-thousand repeat simulations for a draw, I calculate the average for each of the forty-four intervals across the simulations. The result is an approximation of the function for the average number of reactive replacements conditioned on the draw from the posterior by times $\Delta t_p = \{6, 12, \dots, 264\}$ weeks.

Algorithm 2 is computationally intensive, and so I thin the draws from the posterior, using only every fifth draw. The resulting draws of $E[K_{\text{RM}} | \Delta t_p, \theta^s]$ approximate the distribution of the average cost per unit time for the deferent choices of fixed time replacement intervals.

Algorithm 2 Numerical procedure to calculate the expected cumulative number of reactive maintenance events. For each draw from the posterior, 1000 simulations are run, each 44 shutdowns long.

```

1: Assign simulation parameters
2:  $N_{\text{frames}} \leftarrow 143$                                 ▷ Number of frames
3:  $N_{\text{shuts}} \leftarrow 100$                                ▷ Number of shutdowns to run each simulation for
4:  $N_{\text{runs}} \leftarrow 1000$                              ▷ Number of runs of the simulation
5:  $\Delta t_{\text{shuts}} \leftarrow (6 \times 7)$                 ▷ Operating days between shutdowns
6:  $P_{\text{RM}} \leftarrow 0.05$                             ▷ Operating days between shutdowns
7:
8: for every fifth posterior draw  $s = 1 : 5 : S$  do
9:   for  $i = 1$  to  $N_{\text{runs}}$  do
10:     $t_1, \dots, t_N \leftarrow 0$                            ▷ Set age of frames to zero
11:    for  $j = 1$  to  $N_{\text{shuts}}$  do
12:      Calculate probability of frame failures in period between shutdowns.
13:       $P_{\text{UF}n} \leftarrow \frac{F_{\text{Weibull}}(t_n + \Delta t_{\text{shuts}} | \theta^s) - F_{\text{Weibull}}(t_n | \theta^s)}{1 - F_{\text{Weibull}}(t_n | \theta^s)}$ 
14:       $I_n \sim \text{Bernoulli}(P_{\text{UF}n})$                   ▷ Simulate frame failures
15:       $K_{\text{UF}} \leftarrow \sum_{n=1}^N I_n$                    ▷ Number of frame failures
16:       $K_{\text{RM}|\Delta t_{\text{shuts}}}[j] \sim \text{Binom}(K_{\text{UF}}, P_{\text{RM}})$  ▷ Number of reactive maintenance
events
17:      for  $n = 1$  to  $N$  do
18:        if  $I_n = 1$  then
19:           $t_n \leftarrow 0$                                  ▷ Reset age of failed frames to 0
20:        else
21:           $t_n \leftarrow t_n + \Delta t_{\text{shuts}}$            ▷ Progress age of unfailed frames
22:        end if
23:      end for
24:    end for
25:    Take Cumulative sum of the reactive maintenance events over the  $N_{\text{shuts}}$ 
periods between shutdowns.
26:     $K_{\text{RM}|t} \leftarrow \text{cumulative sum of } K_{\text{RM}|\Delta t_{\text{shuts}}}$ 
27:  end for
28:  Average across the simulation runs
29:   $E[K_{\text{RM}|t}]$ 
30: end for

```

Choosing a fixed time replacement interval Figure 3.8 shows the predictive distribution of *cost per unit time* for the different choices of fixed time replacement intervals going up to five years along with the median and 50% and 90% intervals of the distributions. Importantly, the distributions provide the range from the ‘best case scenario’ to the ‘worst case scenario’ of the average cost for a chosen fixed time replacement interval based on the posterior uncertainty. In the figure, when the fixed time interval is greater than 150 weeks, bulk replacements stop having much of an impact on the *cost per unit time*, which makes sense considering that the posterior expectation of an idler-frame lifetime is around three and a half years (185 weeks). The biggest potential cost savings are for the fixed time replacement intervals less ≤ 60 weeks.

Table 3.3 summarises the predictive distributions for the ten shortest fixed time intervals. The fixed time interval with the smallest potential cost per day is $\Delta t_p = 12$ weeks (every two shutdowns) since the lower 90% uncertainty bound is only \$9,818 per day—in other words

$$\Pr [\text{cost per unit time} < \$9,818 | \Delta t_p = 12 \text{ weeks}, \text{DATA}] = 0.05.$$

However, the uncertainty around the lifetime distribution also means that replacing the idlers in the frames every twelve weeks could be over-maintaining the idler-frames since the upper bound of the 90% uncertainty interval at $\Delta t_p = 12$ weeks is the second largest, after $\Delta t_p = 6$ weeks.

The safer choices are the fixed time interval with the lowest risk of over-maintaining the idler-frames. For example the upper bound of the 90% intervals for $\Delta t_p = 30, 36,$ and 42 weeks are \$16,480, \$16,480, and \$16,498 per day, respectively. These choices offer reduced costs—although possibly not the largest reduction—and the reasonably large replacement intervals are desirable because they provide flexibility when planning the replacement work around the maintenance of other assets—i.e. they would not take up less of the collective labour resources.

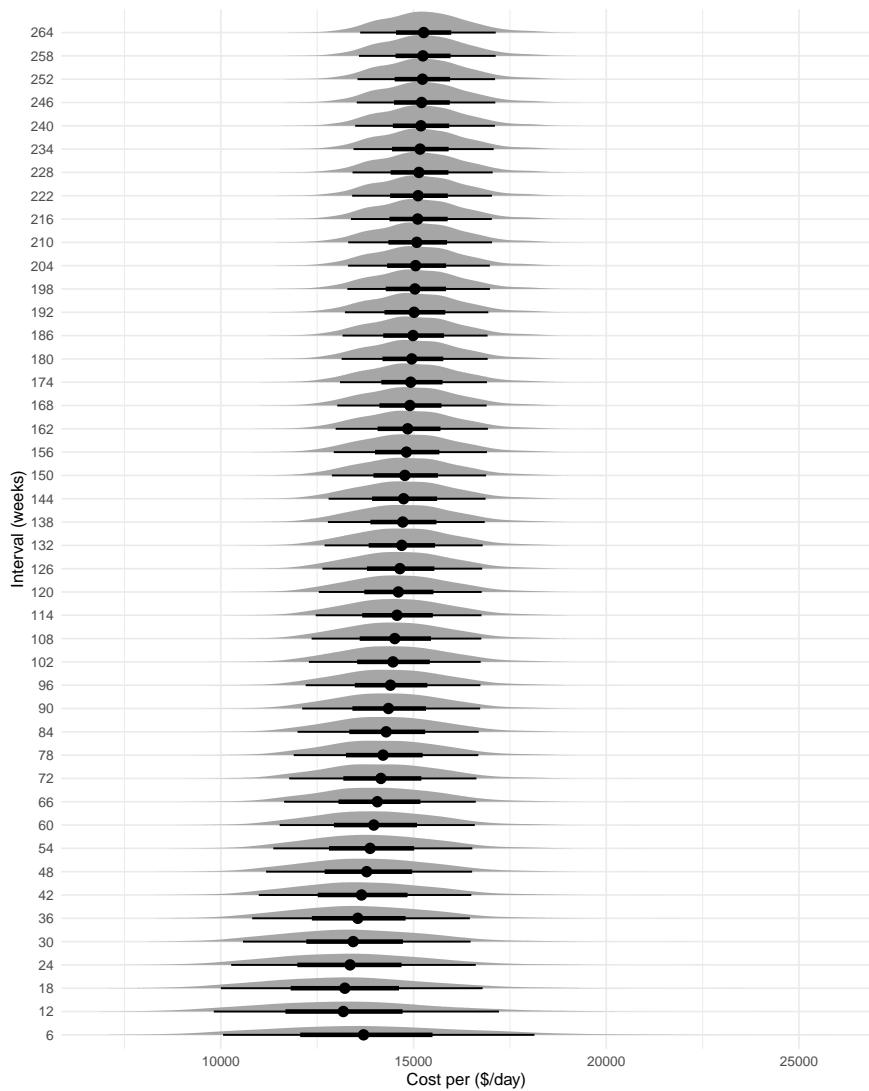


Figure 3.8: The cost per unit time of preventative maintenance policies for idlers. On the horizontal axis is the cost per unit time, and on the vertical is the fixed time replacement interval. The point and intervals show the median and 0.5 and 0.9 intervals. The grey densities show the full distributions.

Table 3.3: Summary of predictive distributions of cost per unit time (in \$/day).

| Interval (weeks) | Mean | 5% | 25% | 50% | 75% | 95% | Pr[< \$15000] |
|------------------|-------|-------|-------|-------|-------|-------|---------------|
| 6 | 13857 | 10059 | 12060 | 13704 | 15491 | 18135 | 0.69 |
| 12 | 13273 | 9818 | 11677 | 13178 | 14714 | 17216 | 0.79 |
| 18 | 13267 | 10000 | 11811 | 13217 | 14622 | 16792 | 0.80 |
| 24 | 13369 | 10270 | 11985 | 13352 | 14688 | 16613 | 0.80 |
| 30 | 13498 | 10575 | 12219 | 13434 | 14724 | 16480 | 0.79 |
| 36 | 13615 | 10813 | 12362 | 13553 | 14792 | 16461 | 0.78 |
| 42 | 13725 | 10984 | 12516 | 13649 | 14844 | 16498 | 0.77 |
| 48 | 13833 | 11175 | 12694 | 13784 | 14965 | 16519 | 0.76 |
| 54 | 13930 | 11362 | 12806 | 13872 | 15016 | 16525 | 0.75 |
| 60 | 14019 | 11526 | 12933 | 13968 | 15088 | 16588 | 0.73 |

Another way that the fixed time interval could be chosen is based on some criteria, e.g. the highest probability that the cost will be less than \$15,000 a day, in other words, the minimum $\Pr[\text{cost per unit time} < \$15,000 | \Delta t_p, \text{DATA}]$. The rightmost column in Tab. 3.3 shows this probability for each fixed time interval. Based on this criterion and the posterior uncertainty around the lifetime distribution, the most suitable fixed time interval is either 18 or 24 weeks. However, 12-48 weeks all have very similar probabilities and so are all reasonable choices, and the final decision depends on the wider context of maintenance planning at the mine. Note that these conclusions would differ for different conveyors since their reliability, the cost of maintaining them, and their impact on production may differ.

Lastly, in the cost function I have used, I have looked only at the expectation of the number of reactive maintenance events; I do not include the variability. It may also be important to investigate the variability in the cost over a short period. For example, the cost of preventative maintenance over a short period should always be roughly the same, whereas the cost due to reactive maintenance may vary a lot. Therefore, it may be preferable to favour shorter fixed-time replacement intervals if it reduces the risk of large financial costs incurred over a short period of time.

3.5 Discussion and conclusions

In this chapter, I have demonstrated my proposed method for modelling lifetime data that are right censored and left truncated with unknown exposure time from Chap. 2 on an industry dataset of idler-frame failure times. From the output of the Bayesian model that imputes the partially observed lifetimes, I showed how it is easy to obtain predictive distributions (and therefore point estimates and uncertainty intervals) for the remaining useful life of the units still in operation at the end of the observation period as well as the cumulative number of failures going forward from the end of the observation period. I also showed how to pass the posterior draws of the Weibull parameters through a cost function to choose a fixed time replacement interval for a preventative maintenance policy in a way that accounts for the posterior uncertainty in the analysis. In this last section, I discuss the main conclusions and outcomes from the chapter and the potential for future work.

According to the analysis in Sec. 3.3, the an idler-frame is 3.6 years (with 90% lower and upper bounds of 3.2 and 3.9 years). This estimate is shorter than the prior belief about the expected lifetime of an idler provided by the manufacturer, which is a sign of conflict between the prior and the data. Possible reasons for this conflict are 1) because the analysis I perform is at the frame level since failures are not reliably reported at the idler-level, and 2) the analysis is performed with respect to calendar date. A better exposure would be throughput or travel of the belt since these more strongly indicate usage; however, only calendar time was available.

I demonstrated, using the posterior draws, that imputing the partially observed lifetimes—in particular the right-censored lifetimes—in a Bayesian approach provides a very natural and straightforward way of obtaining estimates and uncertainty intervals for the RUL of the frames still in operation and the predicted cumulative number of failures. The Bayesian approach, therefore, alleviates the analyst from choosing a suitable method for generating uncertainty

intervals. The predictive distributions of the RUL of the idlers can be used to prioritise replacements, and the predictive distribution of the cumulative number of failures—derived from the distributions of RUL—can help with logistics; for example, many mines are remote, so they can only hold a particular amount of stock on site, knowing a range for the possible number of failures in the next month would help decide how many replacement idlers should be kept on site.

I have shown that analysis of the partially observed idler-frame lifetimes is useful for short and long-term decisions. However, this could be improved further by including extra information. For example, there is extra information for the idlers, such as manufacturer type—which is available for the recent idler failures—or the location on the belt—idlers in the impact zone where ore is loaded onto the belt or idlers near the head or tail of the belt where it transitions onto the pulleys. This information could be included through covariates or used to divide the population into groups of similar but different idler-frames, and these groups modelled hierarchically in the Bayesian model. Furthermore, in my analysis, I treated short lifetimes where $t < 21$ days as right censoring events. These early failures can still result in downtime of the belt. Methods could be explored to include these ‘infant mortalities’ in the model, e.g. using a mixture distribution such as in (Mittman, 2018). However, such an extension would need to investigate the impacts on the imputation of the partially observed lifetimes and missing truncation times.

In this case study analysis of the idler-frame failures, I have emphasised how the posterior can inform decisions. This extension of the discussion of Bayesian analysis of lifetime data is useful for people applying these methods in practice. Rather than taking the point estimates of the quantities of interest and ‘plugging’ them into some utility function, one of the biggest benefits of the Bayesian framework is how easy it is to propagate uncertainty through these types of functions. It is important to emphasise this in work that targets reliability practitioners.

Part II

Part two: Degradation modelling

In this second part of the thesis, I focus on degradation modelling—primarily using the gamma stochastic process. The more foundational work in this second half of the thesis—Chap. 4 and 5—arose out of necessity in order to apply the gamma process to industry data from the Centre’s industry partners. Particularly, the work was motivated by two applied problems: modelling the degradation of many pipe spools in a bauxite processing plant and modelling the wear of an overland iron ore conveyor’s belt—the latter was in one of my industry placements. The pipe degradation project, undertaken by Gabriel González Caceres and Aloke Phatak, looked at the degradation in wall thickness of pipe spools. The gamma process should theoretically be a good model for the erosion of the inside of the spools. However, the degradation signals for each spool are very noisy and consist of only a few measurements; therefore, there was a need to accommodate noise in the gamma process model and to borrow information between similar units in order to estimate the parameters of the model. The belt wear project was an industry placement where I was asked to analyse belt wear of all conveyors at an industry partner mine, processing facility, and port and to devise a new method for modelling the wear. To analyse the many wear profiles, I used functional data analysis techniques. This led to taking the same approach when modelling the belts’ degradation. The gamma process should also be a good model for the degradation of the belt since the wear occurs in small increments as ore is loaded onto the belt. However, similar to the pipe spool problem, the signals of the spline coefficients in the functional analysis were noisy and sparsely observed.

In the first chapter of Part II, Chap. 4, I demonstrate how the gamma process can be extended to incorporate noisy degradation measurements through Bayesian hierarchical modelling. Then, in Chap. 5, I show how unit-to-unit variability (which facilitates the borrowing of information between units) can be included in the proposed noisy gamma process model. The work in these two chapters has been submitted for publication as the manuscript *Bayesian hier-*

archical modelling of noisy gamma processes: model formulation, identifiability, model fitting, and extensions to unit-to-unit variability to Applied Models in Business and Industry. The chapters present an expansion of the manuscript by tying the supplementary material submitted alongside the manuscript into the main analysis and discussion and providing a deeper discussion of the two main topics. There is also a larger focus on the computational details of the analysis.

In Chap. 6, I present the functional belt wear model and apply it to wear data from a single overland conveyor belt. The belt wear model utilises the noisy gamma process and extensions from Chap. 4 and 5. The work in Chap. 6 has been submitted for publication to the Annals of Applied Statistics as the manuscript *Functional degradation modelling of the wearing surface of conveyor belting using Bayesian hierarchical modelling and Gamma processes*. In addition to the work presented in the manuscript, Chap. 6 provides a side-by-side comparison of the gamma process-based belt wear model from the manuscript with a linear general path-based model.

The e-prints for the two papers have been uploaded to arXiv.org. The full references for the manuscripts are

1. Leadbetter, R., González Cáceres, G., & Phatak, A. (2024). *Bayesian hierarchical modelling of noisy gamma processes: Model formulation, identifiability, model fitting, and extensions to unit-to-unit variability*.
2. Leadbetter, R. & Phatak, A. (2024). *Functional degradation modelling of the wearing surface of conveyor belting using Bayesian hierarchical modelling and Gamma processes*.

Chapter 4

A 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 Chaps. 2 and 3 are not useful because the uncertainty in the parameter estimates is so large. 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 (Lawless & Crowder, 2004). 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 (Ye & Xie, 2015). Consequently, models for gamma processes must be extended to account for such noise.

Add a line

In this chapter, I show that this necessary extension can be facilitated in a straightforward and tractable way using the Bayesian hierarchical modelling framework. I also demonstrate, through simulation, that this noisy gamma

to say that

I'm abusing

good

notation

and using

lowercase

letters for

random

process model is more difficult to fit than a standard gamma process due to a pre-asymptotic identifiability issue. In the remainder of the chapter starts with Sec. 4.1, which gives a brief introduction to the gamma process as it is used for degradation modelling. In Sec. 4.2, I provide a short overview of current works on modelling noisy gamma processes in the reliability literature and introduce how the model can be implemented using the Bayesian hierarchical modelling framework. Section 4.3 discusses the merits of reparameterising the gamma distribution in terms of orthogonal parameters that are interpretable in terms of the average degradation rate and volatility of the gamma process; these parameters allow us to more easily think about how to specify prior distributions for the parameters of a gamma process. I then discuss the equally important step of justifying the prior and performing prior predictive checking in Sec. 4.4, which allows us to specify sensible prior distributions. In Sec. 4.5, I fit the noisy gamma process model to simulated data and show that when there are only a few noisy observations, MCMC sampling and posterior inference are poorly behaved. I diagnose that these issues results from the challenges of separating the measurement error from the inherent volatility of the gamma processes using the useful diagnostics of HMC. After identifying and explaining the issue, the section concludes with a demonstration of how this poor behaviour of sampling and inference can be resolved by adding a small amount of supplementary information into the analysis: either through a more informative prior or supplementary data. Section 4.6 summarises the main results and points the way to future work.

4.1 The gamma process

The gamma process is a type of stochastic jump process. It was introduced to the reliability domain by Abdel-Hameed (1975), 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 (van Noortwijk, 2009).

Consider a sequence $\{z_i\}$ of noise-free measurements of the degradation of a unit observed at times t_i , $i = 0, 1, 2, \dots, I$. Without loss of generality, I assume that $z_0 = 0$ at $t_0 = 0$. A gamma process (Lawless & Crowder, 2004) 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

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

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 (Frenk & Nicolai, 2007), 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, I consider only the stationary gamma process. When using a linear shape function, we can write eq. (4.1) more simply as

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

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 Chap. 5.

4.2 A noisy gamma process

In this section, I give a background to the noisy gamma process and describe how its implementation can be simplified using the BHM framework introduced

in Sec. 1.4. In an early paper, Kallen and van Noortwijk (2005) 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, Kallen and van Noortwijk (2005) use simulation to approximate the likelihood. D. Lu, Pandey, and Xie (2013) 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 eq. (4.2) to be estimated.

Building on the work of Kallen and van Noortwijk (2005) and D. Lu et al. (2013), Pulcini (2016) proposed a way to include degradation-dependent measurement error. Other researchers focused on improving computational efficiency by alternative methods such as deconvolution (Rodríguez-Picón et al., 2021) or by using faster algorithms to approximate the likelihood, for example, approximate Bayesian computing (Hazra, Bhadra, & Pandey, 2022; Hazra, Pandey, & Manzana, 2020). 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 Giorgio, Mele, and Pulcini (2019) and Esposito, Mele, Castanier, and Giorgio (2022) also formulate a conditional like-

lihood 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 Sec. 1.4, consider the noisy degradation trace in Fig. 4.1. Figure 4.1 shows a degradation trace generated from a 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 , $i = 0, 1, 2, \dots, I$ and, using the same notation as in Sec. 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

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

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

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

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

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

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

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 I explain how this simplifies the last step of specifying the parameter model. Then, in Sec. 4.4, I use simulation to choose suitable distributions for these parameters.

4.3 Reparametrisation

The gamma process described in eq. (4.2) has density function

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

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

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

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 distributions 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* (Cox & Reid, 1987), 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 Huzurbazar (1956), 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 (Huzurbazar, 1956).)

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

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

I use this reparameterisation in the remainder of this thesis. Kallen and van Noortwijk (2005) 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 summarises our beliefs about the parameters. There are two ways 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 conjugate prior distributions, or convenient prior distributions that facilitated the use of Gibbs samplers or conventional Metropolis-Hastings algorithms (Gilks et al., 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. In this section, I compare some commonly used ‘default’ priors for the GP from the reliability literature with a set of well-thought-out priors for the new parameters through prior

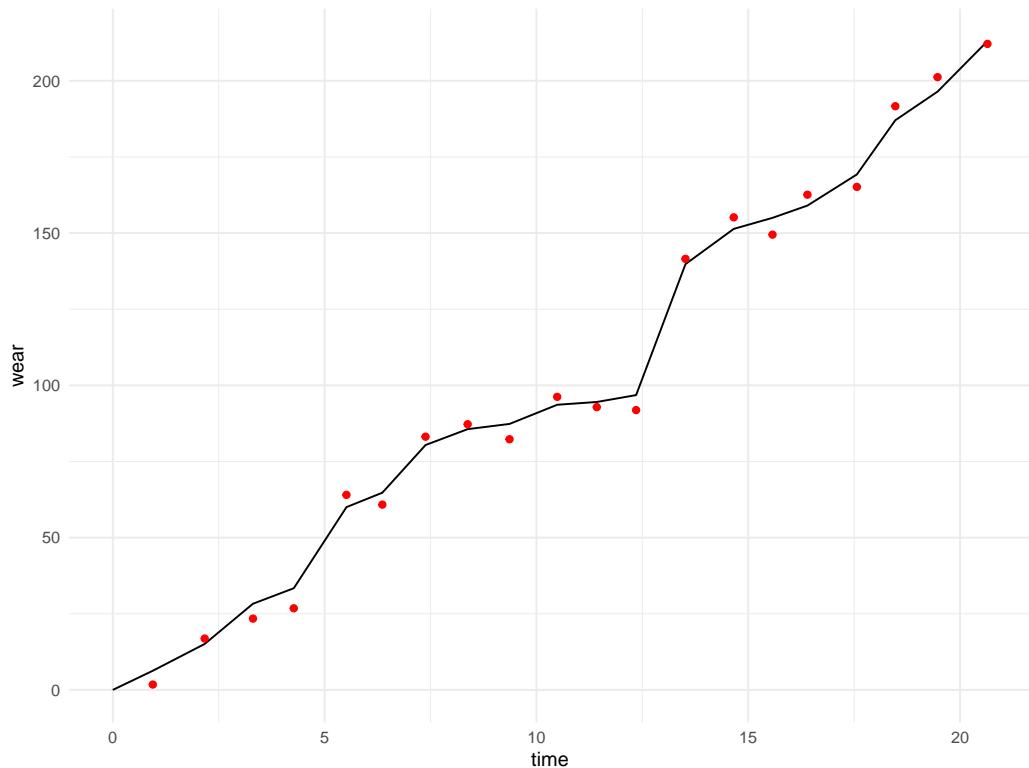


Figure 4.1: A simulated degradation trace (black) generated from a gamma process with parameters $\mu = 10$ and $\nu = 1.119$ and the simulated noisy observations (red) of the trace.

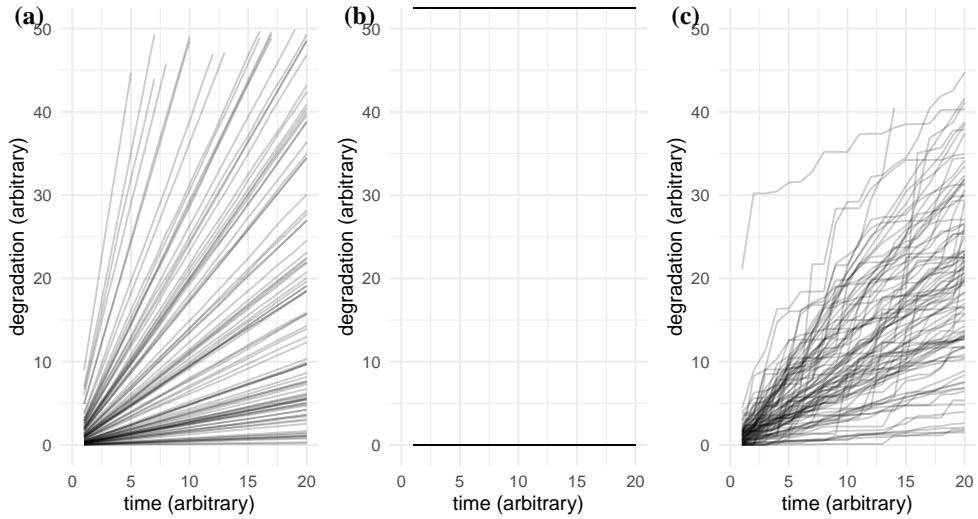


Figure 4.2: One hundred realizations from prior predictive distributions of a noise-free gamma process with mean degradation rate of 1 unit per unit time and different prior distributions for parameters: (a) shape, rate $\sim \text{Ga}(1, 0.001)$, (b) shape, rate $\sim \text{Ga}(0.001, 0.001)$, and (c) parameterization mean/coefficient of variation—see text for details.

predictive simulation: along the way, I provide justification for the new choice of priors for the alternative parameters μ and ν .

In the degradation modelling literature, a gamma distribution is often used as the prior distribution for the rate parameter ξ of the gamma process (Lawless & Crowder, 2004) and also for the shape parameter (Rodríguez-Picón et al., 2018). It is well known that a gamma prior on the rate parameter is conditionally conjugate (Pradhan & Kundu, 2011), and its use leads to analytically tractable results, as Lawless and Crowder (2004) 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 realisations 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 4.2, I illustrate prior predictive checking of a noise-free gamma process using three sets of priors for its parameters: first, ‘conventional’ pri-

ors, $\text{Ga}(1, 0.001)$ and $\text{Ga}(0.001, 0.001)$, that are widely used in the literature for both the shape and rate parameters of the usual parameterisation of a noise-free gamma process in eq. (4.2), and second, priors on μ and ν in the alternative parameterisation 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 $\text{Ga}(\epsilon, \tilde{\epsilon})$, where $\epsilon, \tilde{\epsilon} \rightarrow 0$, is often used as a noninformative prior distribution, especially in mixed linear models, where it is a conditionally conjugate prior for the precision (Hodges, 2014, p. 33). In addition, as we pointed out above, the gamma distribution is conditionally conjugate for the rate parameter: if $\{z_i\}$, $i = 1, 2, \dots, n$, represents an independent sample from $\text{Ga}(\beta, \xi)$, then the conditional 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 (Gelman, Vehtari, et al., 2020, 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 Hodges (2014) and Gelman, Vehtari, et al. (2020).

Figure 4.2 (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. 4.2 (a)) or $\text{Ga}(0.001, 0.001)$ (Fig. 4.2 (b)). In Fig. 4.2 (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. 4.2 (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. 4.2 (b) look much more plausible.

To specify independent prior distributions of the parameters μ and ν in the GP model in eq. (4.5), we adopt the approach introduced by Simpson, Rue, Riebler, Martins, and Sørbye (2017): 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

$$\mu \sim N(1, 0.5).$$

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 4.2 (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. 4.2 (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 realisations, 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 Gelman, Vehtari, et al. (2020, Chap. 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 Sec. 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. 4.2.

Of course, had I used different values of the hyperparameters for the N and t_3^+ priors, the appearance of the degradation traces in Fig. 4.2 (c) would have been different. For example, if we had specified a much more diffuse prior for the mean wear rate, then no doubt unrealistically fast and slow wearing degradation traces would have been generated. It is only through the use of prior predictive checking to choose sensible values of the hyper-parameters in conjunction with suitable distributional forms of the priors that a well-justified prior is obtained.

4.5 Fitting the noisy gamma process

To improve our understanding of the noisy GP model, I fit the Bayesian hierarchical model outlined in Sec. 4.2 with the priors defined in Sec. 4.4 to the single simulated degradation trace in Fig. 4.1 as well as to a subset of the simulated

degradation measurements. For clarity, the full model is

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

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 of all 20 simulated noisy degradation measurements in Fig. 4.1, and another ‘small’ data set that is a subset of 10 points. I fit the BHM of the noisy GP outlined in Sections 4.2 and 4.4 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

I generate the degradation trace in Fig. 4.1—which I coin the ‘large’ dataset—by first sampling twenty time increments from a $\text{Unif}(0.8, 1.3)$ distribution. Next, twenty jumps in degradation are sampled from $Ga(\Delta t_i/\nu^2, 1/\mu\nu^2)$, using $\mu = 10$ and $\nu = 1.119$. I then calculate the cumulative sum of the jumps to obtain the

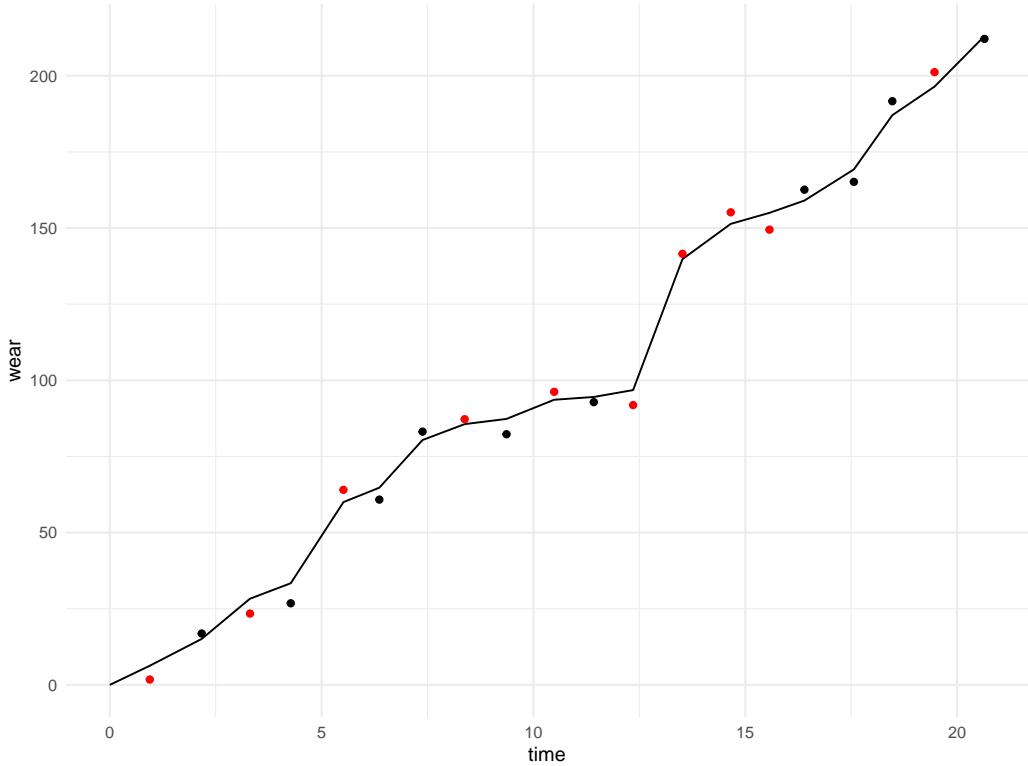


Figure 4.3: The simulated degradation trace with the subset of observations selected for the small dataset highlighted in Red.

underlying, noise-free degradation trace z_i , where $z_0 = 0$ at $t_0 = 0$. Finally, I add Gaussian noise with standard deviation $\sigma = 4$ to the underlying degradation path to get the noisy observations. The big dataset is described in Tab. 4.1. To create the second smaller data set, ten of the twenty noisy observations are randomly selected. The selected degradation observations that make up the small data set are highlighted in red in Fig. 4.3 and displayed in Tab. 4.2.

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* (Stan Development Team, 2022). I generated 88,000 samples from each posterior distribution using four chains of 25,000 iterations each, with a burn-in of 3,000 and no thinning. To ensure a detailed

Table 4.1: The twenty simulated noisy degradation observation that make up the ‘big’ data set.

| t | Δz | z | y |
|-------|------------|--------|--------|
| 0.00 | NA | 0.00 | NA |
| 0.95 | 6.29 | 6.29 | 1.77 |
| 2.17 | 8.81 | 15.10 | 16.87 |
| 3.31 | 13.20 | 28.30 | 23.42 |
| 4.27 | 5.09 | 33.39 | 26.80 |
| 5.52 | 26.63 | 60.02 | 64.04 |
| 6.36 | 4.73 | 64.75 | 60.84 |
| 7.38 | 15.66 | 80.41 | 83.15 |
| 8.38 | 5.21 | 85.62 | 87.24 |
| 9.36 | 1.71 | 87.33 | 82.31 |
| 10.49 | 6.30 | 93.63 | 96.23 |
| 11.43 | 0.92 | 94.54 | 92.85 |
| 12.35 | 2.25 | 96.79 | 91.88 |
| 13.52 | 43.01 | 139.80 | 141.53 |
| 14.66 | 11.57 | 151.37 | 155.16 |
| 15.57 | 3.63 | 155.00 | 149.48 |
| 16.40 | 4.02 | 159.02 | 162.60 |
| 17.56 | 10.23 | 169.26 | 165.16 |
| 18.47 | 17.79 | 187.05 | 191.65 |
| 19.47 | 9.40 | 196.44 | 201.20 |
| 20.65 | 16.57 | 213.01 | 212.11 |

Table 4.2: The subset of ten simulated noisy degradation observation from the ‘big’ data set which make up the ‘small’ data set.

| t | Δz | z | y |
|-------|------------|--------|--------|
| 0.00 | NA | 0.00 | NA |
| 0.95 | 6.29 | 6.29 | 1.77 |
| 3.31 | 13.20 | 28.30 | 23.42 |
| 5.52 | 26.63 | 60.02 | 64.04 |
| 8.38 | 5.21 | 85.62 | 87.24 |
| 10.49 | 6.30 | 93.63 | 96.23 |
| 12.35 | 2.25 | 96.79 | 91.88 |
| 13.52 | 43.01 | 139.80 | 141.53 |
| 14.66 | 11.57 | 151.37 | 155.16 |
| 15.57 | 3.63 | 155.00 | 149.48 |
| 19.47 | 9.40 | 196.44 | 201.20 |

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 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 when there are areas of tight curvature. All of the code to define the model in Stan, simulate the data in R, and sample from the posterior using RStan is available on a GitHub repository.

During sampling—despite increasing *adapt delta* and *maximum tree depth*—eighty divergent transitions occur while fitting the model to the small data set, whereas only four occur when sampling from the posterior conditioned on the larger data set. The divergencies signify incomplete exploration of the target distribution, which I explore in the next section. In addition to evaluating sampling through divergent transitions, energy diagnostics quantify the heaviness of the tails of the posterior distribution and thus can identify inefficient sampling (Gabry & Mahr, 2024). The chain energy plots in Fig. 4.4 compare the marginal energy distribution, π_E , with the first differenced distribution, $\pi_{\Delta E}$, for

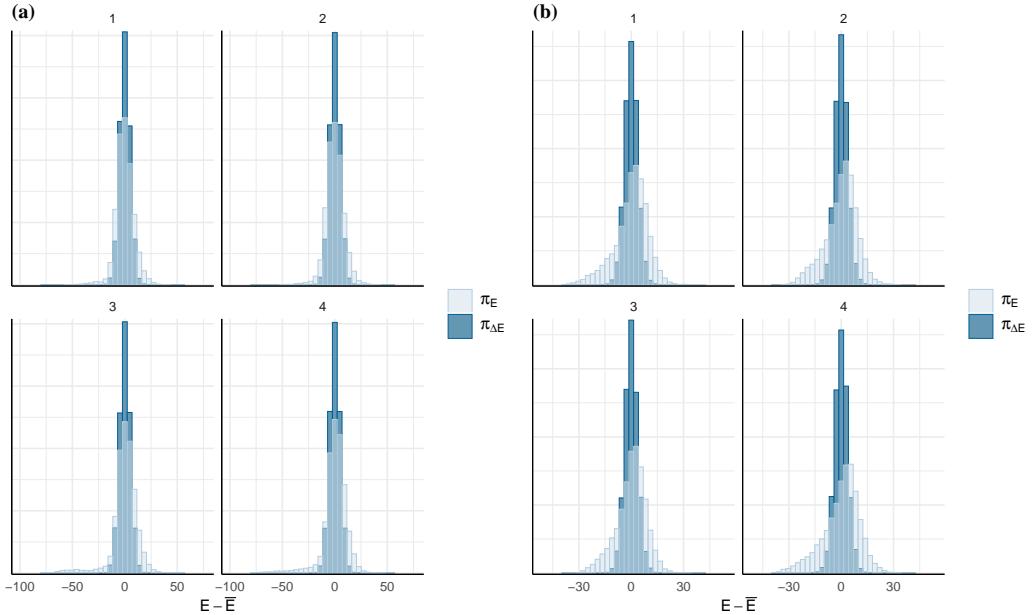


Figure 4.4: The cain energy diagnostics for each chain when sampling from the posterior of the big data set (left) and small data set. Each plot compares the marginal energy distribution, π_E , with the first differenced distribution, $\pi_{\Delta E}$.

each chain. These plots are similar to those in Betancourt (2017) that compare the energy transition distribution (equivalent to $\pi_{\Delta E}$) with the marginal energy distribution (equivalent to π_E). Ideally, these two overlaid distributions should look the same. However, if the distribution of $\pi_{\Delta E}$ is much narrower than that of π_E , like it is in Fig 4.4 (b), then this indicates slow exploration of the target distribution. The energy diagnostics from fitting the model to the small data set in Fig. 4.4 (b) show that exploration of the posterior is very inefficient, whereas when fitting the model to the big dataset, Fig. 4.4 (a), sampling is much more efficient—although it is not perfect. The number of flagged divergent transitions and the energy plots are warning signs are warning signs of an issue with the model, they do not diagnose the issue. Section 4.5.3 looks more closely at the posterior draws and divergent transitions to properly identify the issues with the model.

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 observations, 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 measurement error from the volatility of the underlying gamma process. I come to this conclusion by exploring the degenerate behaviours in the posterior that are flagged by divergent trajectories that occurred during sampling.

Marginal densities Figure 4.5 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 it is fit to the bigger dataset, it successfully reclaims the true values. The marginal posterior densities of 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 appears 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 4.6 shows the posterior predictive distribution of the underlying degradation trace, the z_i , for the two posteriors with

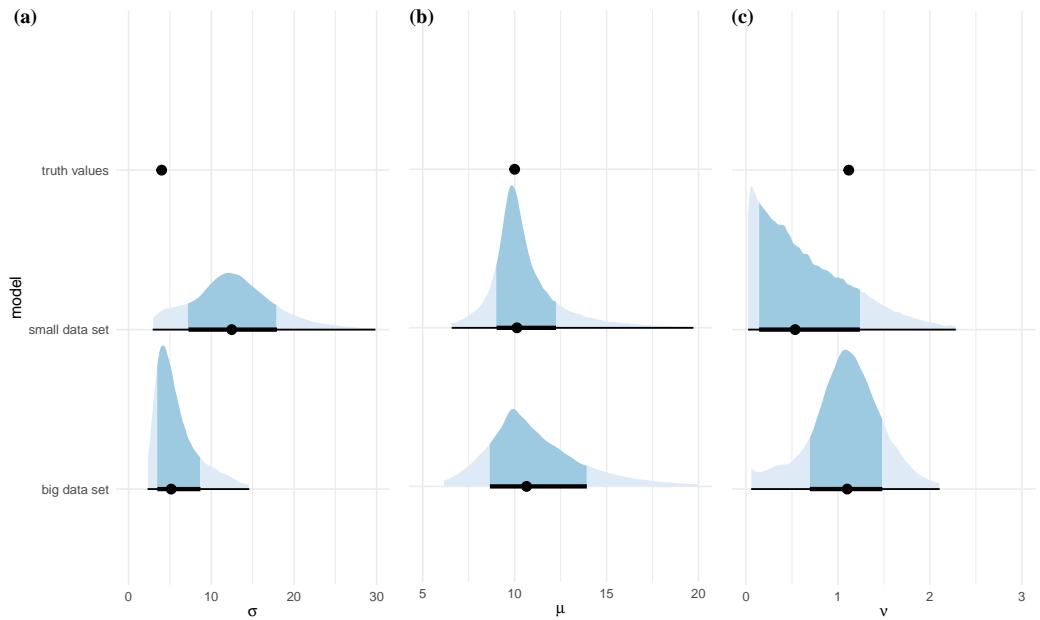


Figure 4.5: The marginal posterior distributions of the parameters σ , μ , and ν when the BHM for the noisy gamma process is fitted to the ‘large’ and ‘small’ simulated data in Fig. 4.1. The points and intervals shown in each distribution represent, respectively, the median and 95% and 66% credible intervals. The values used to simulate the data are shown in the top row.

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 Fig. 4.6 (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 (Fig. 4.6 (b)). In Fig. 4.6 (b), the median path is a nearly straight line through the data points. In addition, the uncertainty intervals are much wider.

Pairs plots Clearly, some issues are occurring in the posterior distribution of the model conditioned on the smaller subset of the data. These issues were pre-

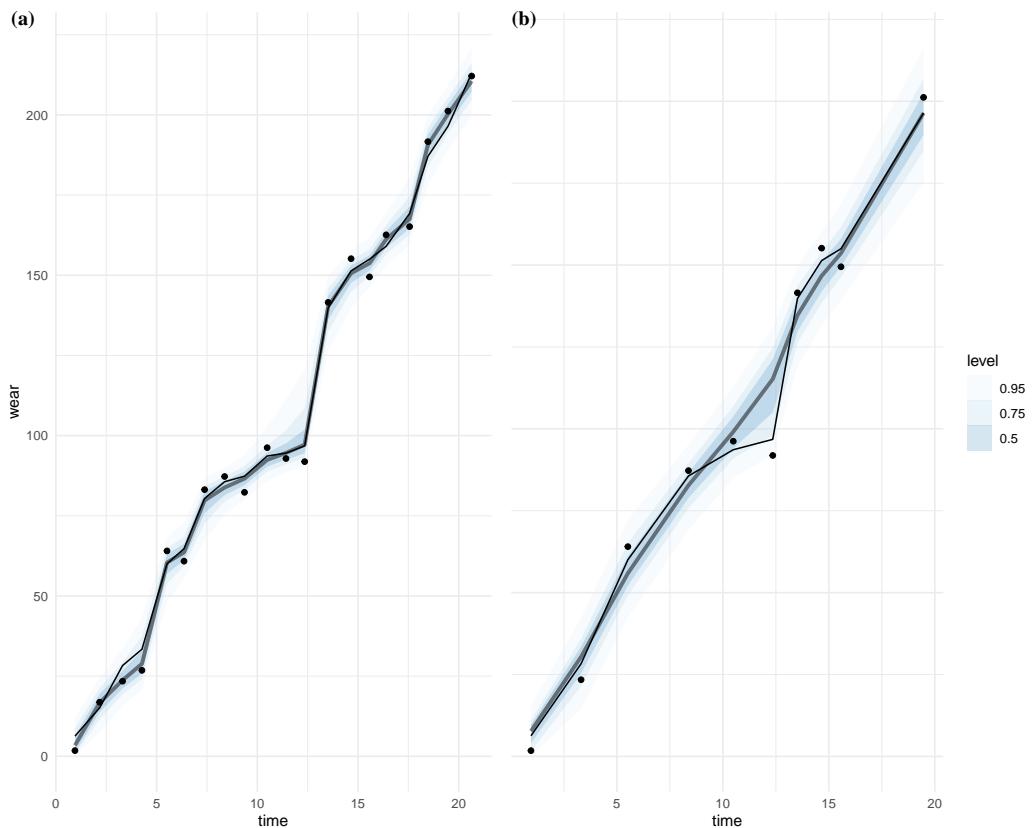


Figure 4.6: The posterior predictive distribution of the filtered degradation path compared to the true degradation path when (a) the model is fitted using the big and (b) small data set. If the distribution of π_E is much wider than that of $\pi_{\Delta E}$, then sampling is inefficient.

eluded to by the divergent trajectories that occurred during sampling (Sec. 4.7). 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 Sec. 1.4, the divergent transitions can point to the problematic areas in the posterior. Figure 4.7 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 $\log(\nu)$ and between $\log(\nu)$ and the first degradation jump. The divergent trajectories are concentrated at the entrance to these funnels, suggesting that they 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 4.8 shows a parallel coordinate plot of the posterior draws for all of the parameters in the model. The divergent trajectories are highlighted in red. The divergencies 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 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 measurement uncertainty. To emphasise this point, in Fig. 4.9, I plot the posterior predictive distribution of the underlying degradation path from the small data set and overlay the divergent paths. From this, it is clear 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 4.10 shows the joint distributions of the intermediate quantities Δz_{15} ((a) and (c) in Fig. 4.10) from the model conditioned on the big dataset

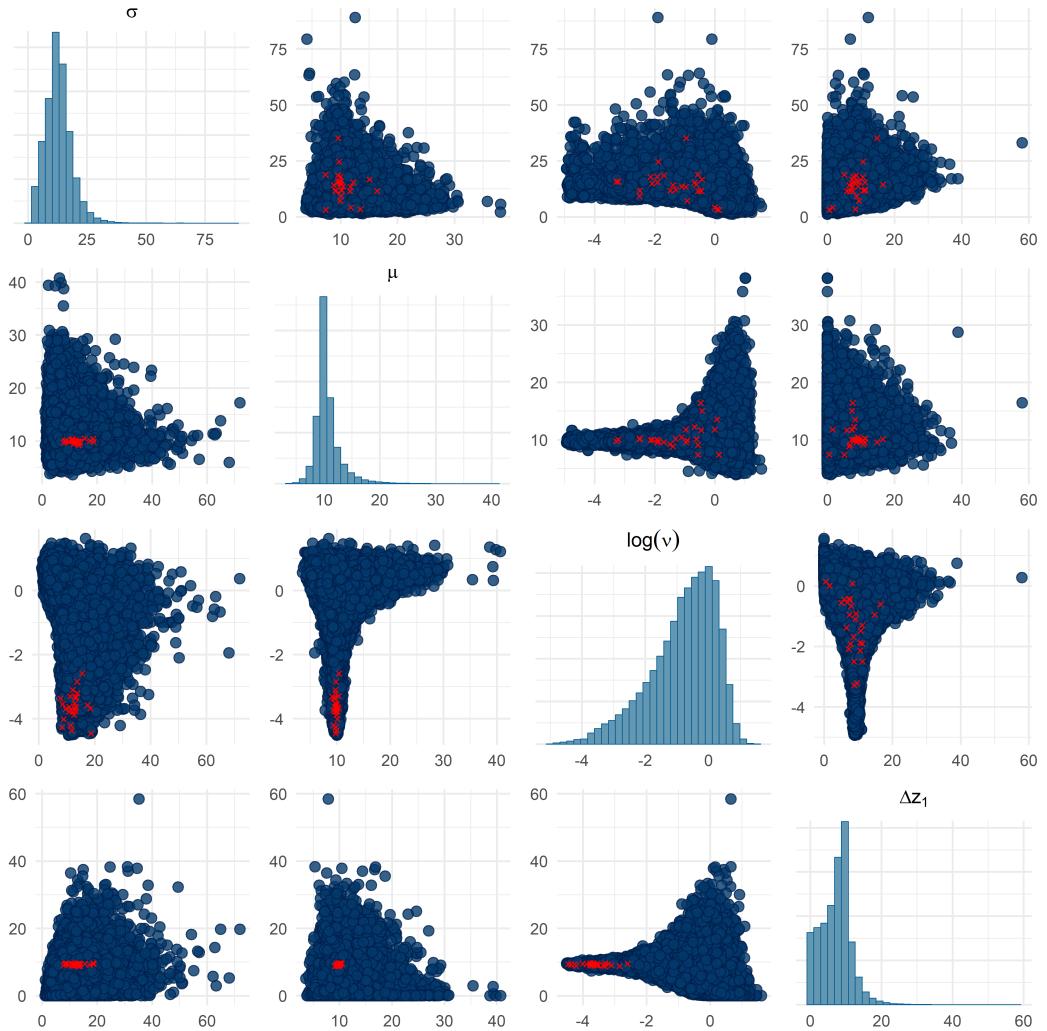


Figure 4.7: Pairs plot showing the MCMC draws from the posterior distribution of the parameters σ , μ , $\log \nu$, and the filtered value Δz_1 when the BHM of the noisy gamma process is fitted to the small dataset. The red points indicate divergences, which congregate at the end of the funnel in the pairwise plots of $\log \nu/\Delta z_1$ and $\log \nu/\mu$.

```
./figures/ch-4/parcoord.png
```

Figure 4.8: The parallel coordinate plot of the draws from the posterior conditioned on the small dataset with the divergent traces plotted in red.

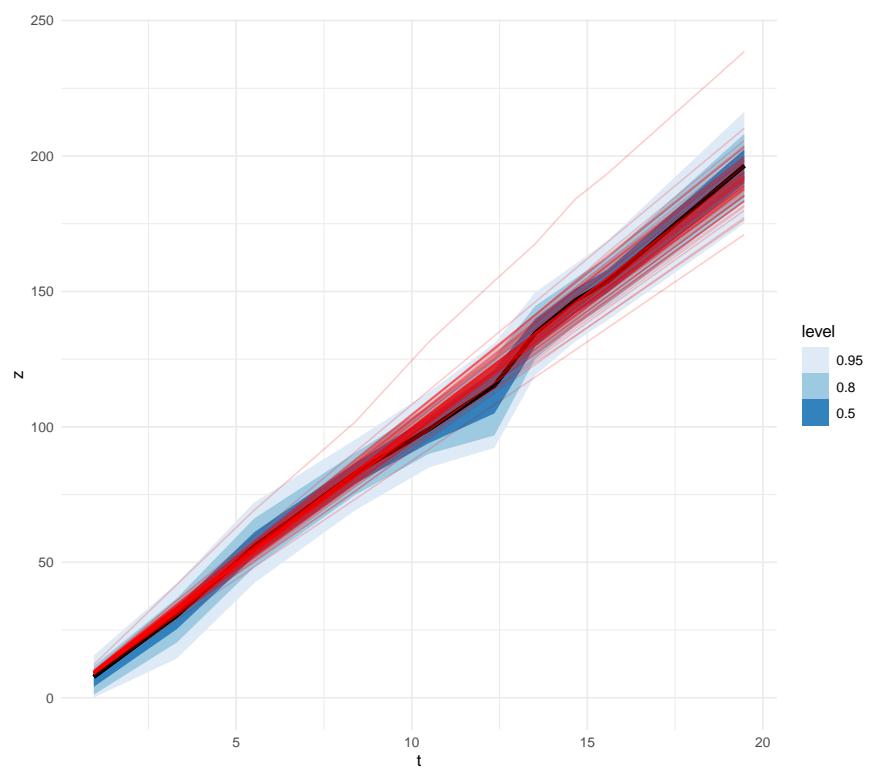


Figure 4.9: The divergent traces overlaid (in red) on the posterior predictive distribution of the degradation path for the small data set.

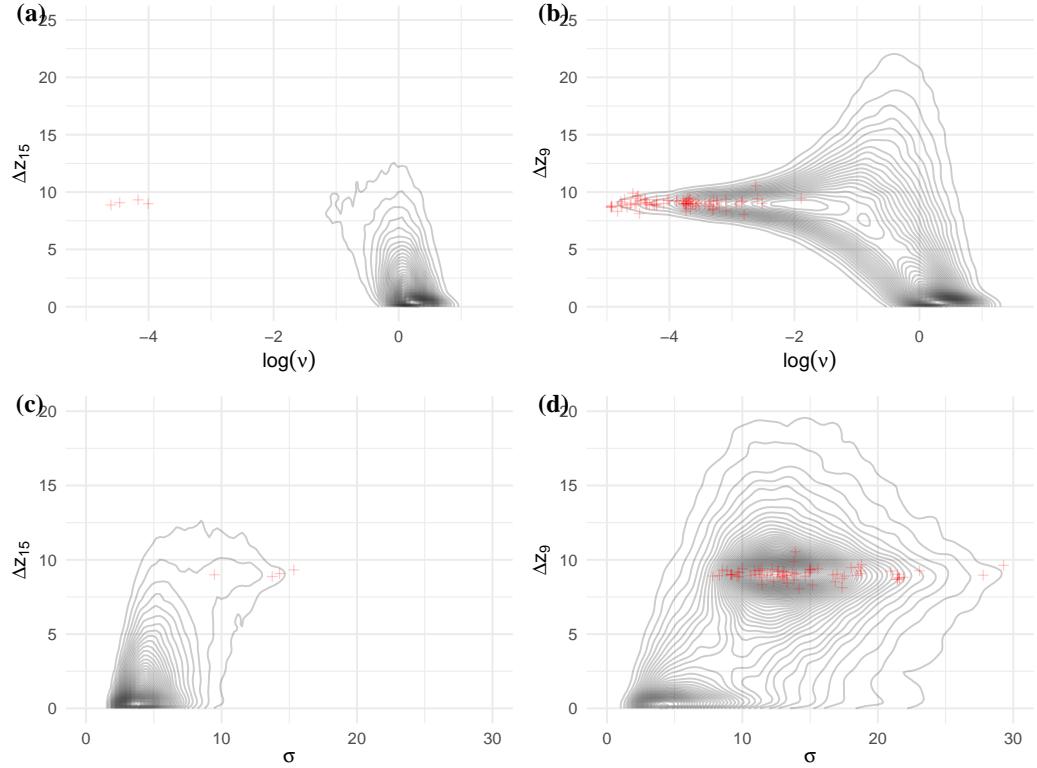


Figure 4.10: The contours of the posterior joint distribution of Δz_{15} and $\log(\nu)$ (a) and Δz_{15} and σ (c) from the posterior when fit to all twenty observations and Δz_9 and $\log(\nu)$ (a) and Δz_9 and σ (c) from the posterior when fit to only ten observations. The divergent transitions are overlaid in red and show the areas in the posterior that are difficult to sample from.

and Δz_9 ((b) and (d) in Fig. 4.10) 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 $\log(\nu)$ (Fig. 4.10 (b)), 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 (Fig. 4.10 (d)). However, in the joint distribution of Δz_{15} and $\log(\nu)$ (Fig. 4.10 (a)) there is very little mass around $\Delta z_{15} = 9$ and no second mode in the joint distribution of Δz_{15} and σ (Fig. 4.10 (c)). This behaviour suggests the nonidentifiability only exists when there are few observations.

4.5.4 Solutions to computational issues

The identifiability issues of the small data set can also be solved by injecting more information into the analysis that helps disentangle σ and ν . This information can come in the form of either supplementary data or prior information that informs one of the nonidentifiable parameters. Obtaining extra information about the measurement error would typically be much easier than the coefficient of variation of the gamma process. Here, I show that adding a small amount of supplementary information using either supplementary data or a stronger prior helps to identify σ and, therefore, ν , resulting in much smoother geometries in the posterior and, therefore, much more efficient sampling. The results regarding inference are arguably better than when I fit the model to all twenty degradation observations without supplementary information about σ .

Prior information In Sec. 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

$$\sigma \sim N(4, 1).$$

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 the noisy gamma process model with a stronger prior, conditioning on the small dataset, is much quicker, and no divergent transitions occur. Figure 4.11 shows the chain energies of the sampler when a stronger prior is used on sigma. The marginal energy distribution and the first differenced distribution now match closely, showing that the chains have efficiently explored the posterior.

A pairs plot of the MCMC samples is shown in Fig. 4.12. In the plots, the geometry looks much smoother than in Fig. 4.7, and there is little remanence of the deep funnel-shaped degeneracies between $\log \nu$ and μ and Δz_1 . In Fig. 4.13,

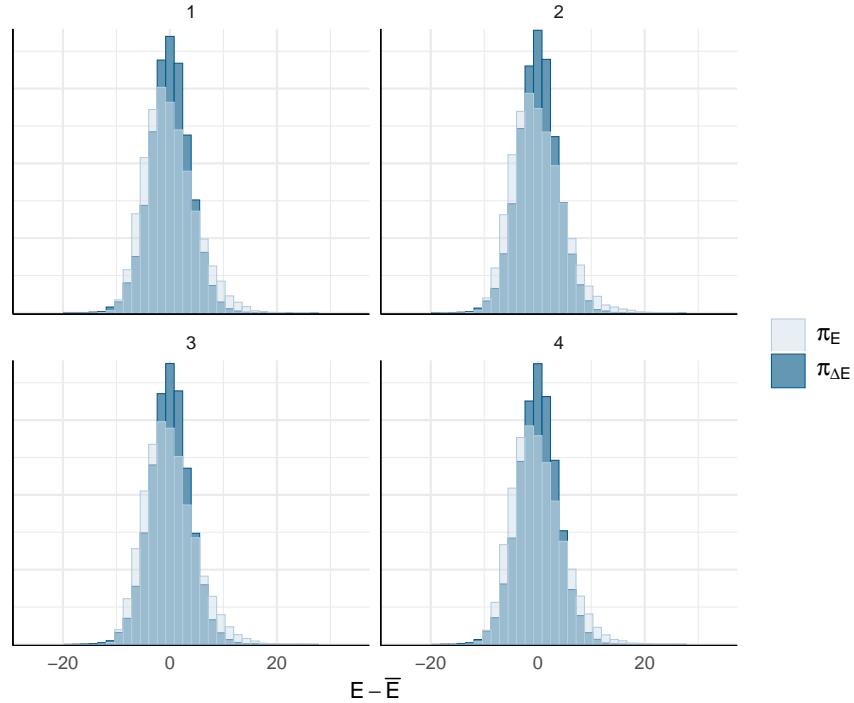


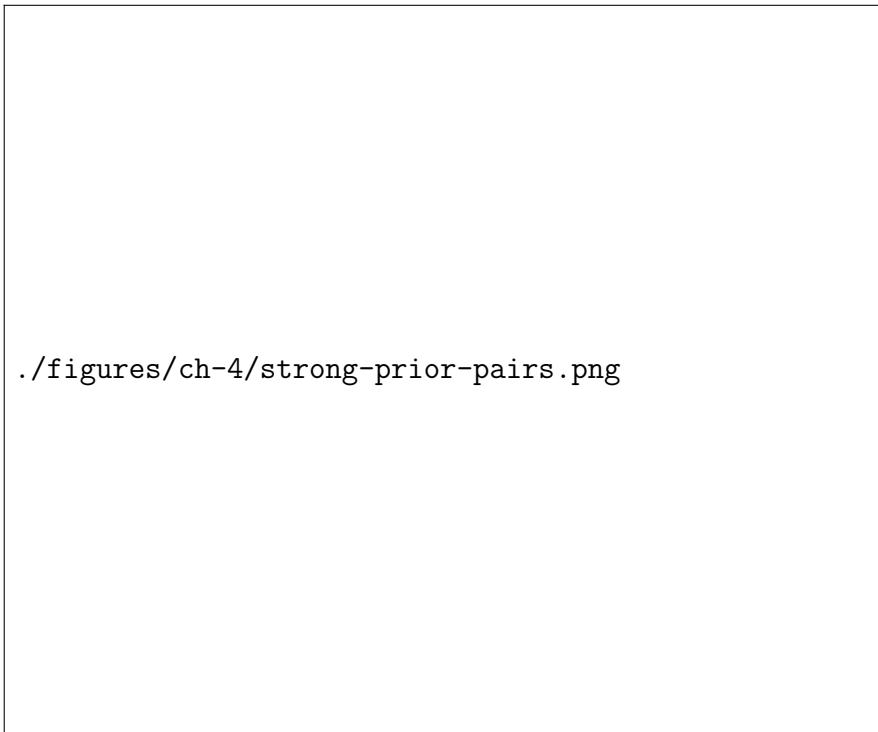
Figure 4.11: The chain energy diagnostics when the model is fit with a stronger prior on σ .

I compare the marginal distribution for σ , μ , and ν from this posterior with the true values and the model fits of Sec. 4.5.3. The marginal distributions of the model with the stronger prior are much smoother, and there is now no mass around zero in the posterior of ν .

Supplementary data Alternative to using a stronger prior, extra data that informs σ can be used to improve the behaviour of sampling. I sample five supplementary observations of the measurement error from the distribution

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

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 by performing a small experiment. The supplementary



./figures/ch-4/strong-prior-pairs.png

Figure 4.12: The pairs plot of the MCMC draws when using a stronger prior on σ .

observations can be easily incorporated into the Hierarchical model by extending the data model to

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

Similar to the more informative prior, the sampler is much more efficient, no divergencies occur during sampling, and the geometry of the posterior distribution looks much smoother (not shown). The resulting marginal posterior distributions of σ , μ , and ν for the model fit with both the small dataset and the supplementary data are also compared in Fig. 4.13. The marginal distributions of the parameters are very much the same as when a stronger prior is used, and the model successfully reclaims the true parameter values.

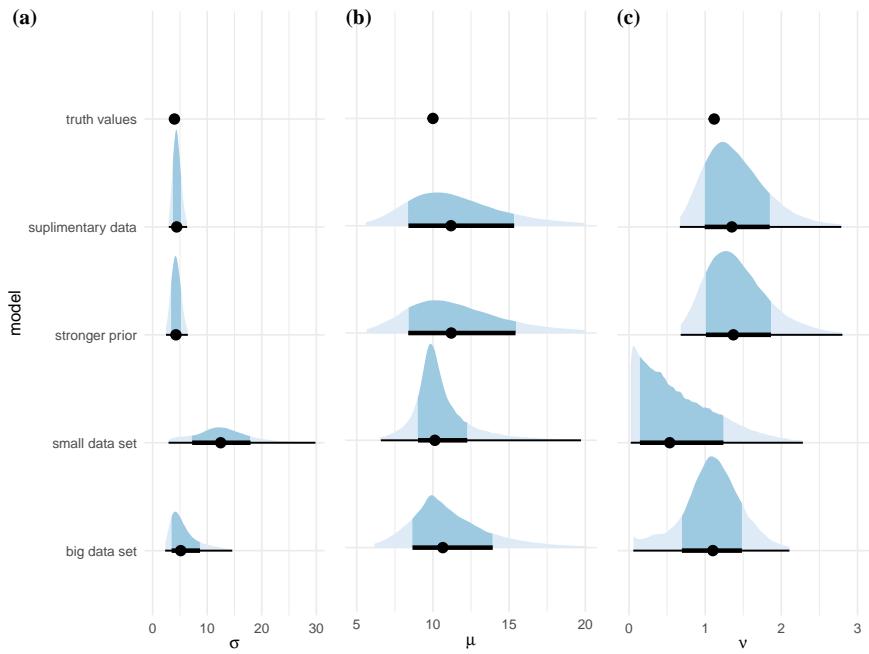


Figure 4.13: The marginal posterior distributions of the parameters σ , μ , and ν when information is added into the analysis either through a stronger prior or supplementary data compared to the original two posteriors.

4.6 Discussion

The main focus of this chapter was 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. Below, I summarise the main element of this chapter and highlight the important findings and contributions, as well as areas for future work.

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, such as unit-to-unit variability (which I show in the next chapter) or covariates, can be incorporated into the model. Finally, the parameters μ and ν are orthogonal, which has desirable computational benefits.

Under this new parameterisation, I re-assessed good choice of priors and shown a principled way of assessing these priors. I draw away from the conventional gamma priors used for the gamma process and think deeply about the justification of priors to construct a weakly informative set of prior distributions that conform with my understanding of the underlying data-generating process. I then evaluate if the priors are, in fact, weakly informative through prior predictive checking. The use of a weakly informative prior is particularly important in the case of a noisy gamma process, since the noisy observation of the degradation trace means that the data do not strongly inform the underlying degradation model and in such cases using non-informative prior distributions can put large amounts of mass in unrealistic parts of parameter space (Tian et al., 2024).

In fitting the proposed noisy gamma process to simulated data, I identify issues with sampling from the posterior when there are only a few observations. Investigating the poor sampling uncovers an identifiability issue between the volatility of the gamma process (expressed by ν) and the measurement error (σ), which is only present when the sample size is small. The observed degenerate behaviour of inference from the small-data posterior results from what Betancourt (2020) refers to as ‘pre-asymptotic non-identifiability’.

Because variation in the noisy degradation signal can be a result of both the randomness of jumps of the GP and the randomness of the measurement

error, it can be difficult to separate these two sources when there is only a small number of observations. In the small dataset, the data do not strongly inform the parameters σ and ν , and it is therefore difficult to distinguish between competing models—the noisy gamma process and one where ν approaches zero—as reflected in their multi-modal posterior distributions. Using the terminology of Betancourt (2020), we can say that these parameters are pre-asymptotically non-identifiable.

I further confirm this in Sec. 4.5.4 by showing that the computational issues and pathological behaviour in the posterior are completely resolved by adding additional information that specifically informs one of the pre-asymptotically non-identifiable parameters. Although fitting the model to all twenty noisy observations results in a much better behaved posterior than fitting it to only ten, there are still remanence of the degenerate areas seen in the posterior conditioned on only ten observations; in addition, there are still a few divergencies and the chain energy plots (Fig. 4.4) show that sampling is still slow. In contrast, when I use a stronger prior for σ or add a small amount of supplementary data that informs σ , there is no sign of degenerate behaviour in the posteriors and sampling becomes very efficient. With enough noisy observations, the model eventually becomes identifiable from the noisy degradation data alone. However, in a typical reliability application, the analysis will have small sample sizes, in which case adding additional information to the analysis can help to identify the model.

The issue of pre-asymptotically non-identifiability is not unique to the noisy gamma process. In an early paper on noisy Wiener processes, Whitmore (1995) also remarked on the difficulty in estimating the measurement error variance of a noisy Wiener process. In a Bayesian reanalysis of the same data, Hamada, Wilson, Reese, and Martz (2008b) imposed strong prior distributions on the measurement error variance and the variance of the Wiener process in order to ensure identifiability, although they do not explicitly justify their reasons for doing so. More work should be done to understand the interplay between the scale of the measurement error, the volatility of the underlying stochastic degradation process,

and these small sample identifiability issues.

In the context of real noisy degradation data, there is no way of checking if there is enough information in the data to properly identify the model. Therefore, practitioners applying the noisy GP model should use as much information as they have available to them. This includes encoding their domain-expert knowledge into the prior distributions of the parameters, rather than choosing a default non-informative prior; incorporating supplementary data that informs one of the pre-asymptotically non-identifiable parameters; and, if available, modelling the degradation of groups of similar units jointly as to ‘borrow’ information. With respect to the latter, in the next chapter, I extend the noisy gamma process for a single degradation trace to model noisy degradation paths from multiple units while assuming that the measurement error is the same for all units. Doing so pools information from between multiple units to inform the parameters, drastically improving the problems with identifiability and MCMC sampling that I have demonstrated in this chapter.

Chapter 5

Noisy gamma process with unit-to-unit variability

In Chapter 4, I discussed how to model a single degradation path using a noisy gamma process postulated through the Bayesian hierarchical modelling framework. I concluded that there is an identifiability issue between the noise and the volatility of the underlying gamma process when there are only a few degradation measurements. A resolution to this preasymptotic nonidentifiability is to add extra information into the analysis of the degradation trace, which can be done by modelling the degradation of a population of m nominally identical units simultaneously. This raises the question of how the degradation traces of each unit are related to one another. For example, we may assume that all of the units are realisations from the same underlying gamma degradation process.

However, this assumption may be too restrictive in practice since there may be additional variability in their degradation resulting from slight variations in operating conditions or their manufacture. The most common approach to modelling this extra layer of heterogeneity between units beyond what can be explained by the volatility of the gamma degradation process and any covariates is to use a ‘mixed effects’ model, in which some of the parameters of the model—so-called ‘random effects’—vary between units or individuals, whereas others, the ‘fixed

effects', do not¹. Early examples in the degradation literature include C. Lu and Meeker (1993) and Lawless and Crowder (2004), who incorporated random effects into a general path model and gamma process, respectively. A more recent example is Rodríguez-Picón et al. (2018), who modelled the GaAs laser dataset that was also analysed by Meeker and Escobar (1998). To model the heterogeneity in the degradation paths, Rodríguez-Picón et al. (2018) incorporate random effects into a noise-free gamma process by specifying the effect in either the mean or variance of the gamma process. By contrast, Peng and Ai (2018) follow the methodology of Lawless and Crowder (2004) and specify random effects in the scale parameter of a gamma process.

In this chapter, I show how the hierarchical model for the noisy gamma process can be neatly extended to incorporate unit-to-unit variability through the same BHM formalism and show the advantages of using the mean/coefficient of variation parameterisation in this context. Before going any further, however, it is worth clarifying the terminology that I use. As pointed out above, the terms random and fixed effects are used when mixed effects models are used to describe unit-to-unit variability. However, as Gelman (2005) and Gelman and Hill (2006) point out, all parameters in a Bayesian analysis are random variables; furthermore, because there is a multiplicity of definitions of fixed and random, such terms can engender considerable confusion [Section 6](Gelman, 2005). Consequently, Gelman (2005) and Gelman and Hill (2006) make a plea for abandoning these long-used terms in place of more descriptive ones: *varying*, for parameters that differ between groups or units, and *constant*, for parameters that are identical for all groups or units. In this chapter, I simply identify which parameters are common across units, those that are unique to each unit, and, most importantly, the specification of the hierarchical prior distribution(s) for parameters that vary from unit-to-unit.

To demonstrate the models for multiple units, I use a data set from an exper-

¹This definition is just one of five that Gelman (2005) lists.

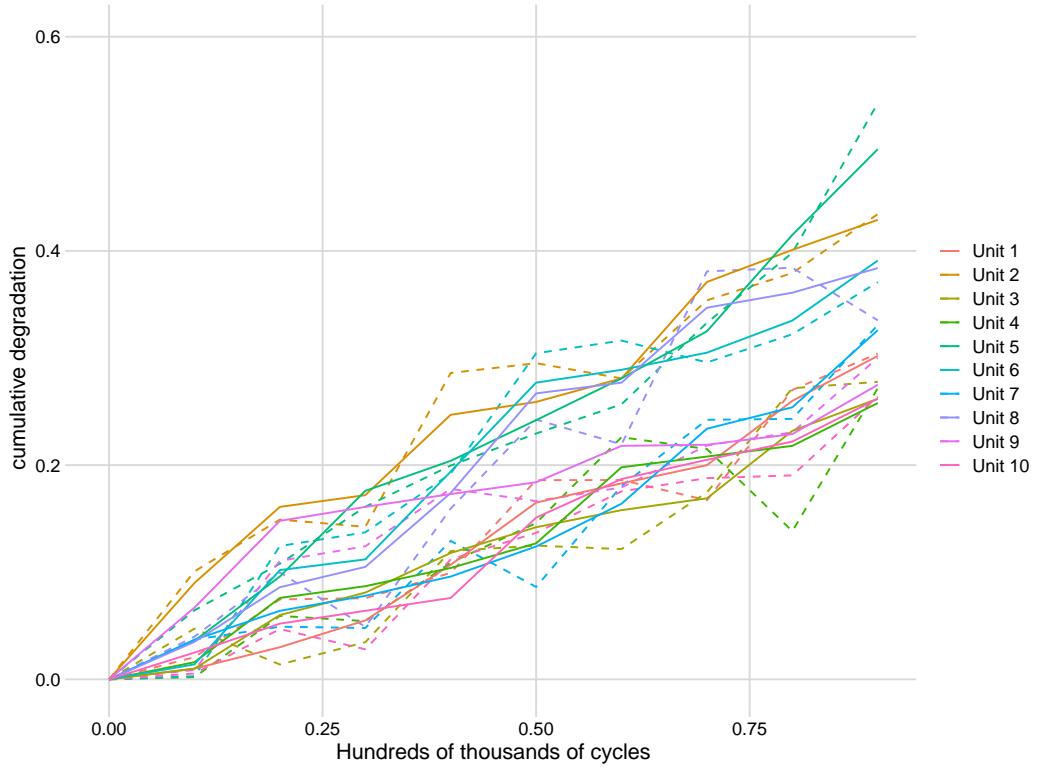


Figure 5.1: Crack-growth propagation data of Rodríguez-Picón et al. (2018): original data (solid lines); noisy data (dashed lines).

iment to measure and then model crack-propagation in the terminal of nominally identical electronic devices originally published by Rodríguez-Picón et al. (2018). The non-noisy data are shown in Fig. 5.1 as solid lines. To emulate measurement error, I add a small amount of $N(0, 0.025)$ noise. These new noisy degradation traces are shown as dashed lines in Fig. 5.1. The soft failure of the terminals is considered to be when the crack length reaches 0.4 mm, and we can see from the figure that by the end of the experiment, several units have yet to fail. There are two reasons why such data may be collected (Robinson & Crowder, 2000): to estimate the remaining useful life or failure time distributions of units that have yet to fail during operation, or the corresponding quantities for new units. In the analysis that follows, I model the noisy data and show how to estimate such failure time distributions along with uncertainty intervals.

In the rest of this chapter, I first propose several noisy gamma process models

for the multiple noisy degradation traces in Fig. 5.1 in Sec. 5.1; a complete pooling model (i.e., where all parameters are *constant*) and three where either one or both parameters of the gamma process vary across units. In Sec. 5.2, I then go on to sample from and evaluate the posterior distributions of these models conditioned on the noisy crack-growth data set. The models are also compared using elppd and cross-validation methods. I then show how to construct failure time distributions for a new unit and a unit that is currently under test but yet to fail in Sec. 5.3 using both a complete and a partial pooling model. Finally, the findings and conclusions of the analysis are discussed in Sec. 5.4.

5.1 Models for multiple units

There are three ways in which we might consider allowing a model for the degradation data in Fig. 5.1 to vary, each of which leads to a different form of *pooling*, or, alternatively, of how information is shared or not among the units (Johnson, Ott, & Dogucu, 2022). The added advantage of the mean/coefficient of variation parameterisation is that it makes it explicit which characteristics of the model we are sharing between units. For example, first, we might make the assumption that for a particular characteristic of the model, such as the mean degradation rate (described by the parameter μ), the units do not contain information that may be relevant to each other and therefore estimate completely separate values of the parameter for each unit; this corresponds to *no pooling*. Secondly, we might assume that all of the units have the same mean wear rate, and the variation that we observe in Fig. 5.1 is only due to the volatility of the gamma process and the fact that we have only observed them over such a short period, and hence estimate the parameter μ by averaging the data from all units—this is *complete pooling*. Finally, although the units are different from each other, they have the same specifications, so we might expect their average degradation rates to share similar characteristics: this supposition can be modelled by allowing the parameter μ to vary from unit-to-unit yet arise from a common distribution. Doing so

results in *partial pooling* of information, which is especially useful when sample sizes are small, but we do not want to make the assumption that all units are identical (McElreath, 2020, Section 13.1).

Different forms of pooling can be applied to each parameter. For example, Lawless and Crowder (2004) allow the rate parameter of the gamma process to vary from unit-to-unit and assume these unit-specific rate parameters arise from a common distribution whose parameters are estimated from the data (partial pooling) while also assuming that the shape parameter is the same for all units; i.e. completely pooled. The multitude of possible pooling combinations is one reason why it is useful to specify a hierarchical model so that the parameters have separate, clear effects on the outcome—like μ , ν , and σ do. By doing so, we can use our understanding of the data-generating process to select a sensible cohort of models to explore.

In the analysis that follows, I confine my exploratory modelling of the crack growth data to models where σ is completely pooled, and μ and ν are either completely or partially pooled across the units. Since the crack growth data is from a formal experiment, it is reasonable to assume that measurement error in the degradation measurements is constant across the units and observation times and hence to completely pool σ . Moreover, because the degradation traces are from nominally identical units, it makes sense to assume that their mean wear rate and volatility are in some way related; i.e. either partially or completely pool μ and ν . Furthermore, if there is no pooling of either μ or ν , there is no way to make statements about new units without using heuristics and one of the motivations for modelling the crack growth data is to produce reliability estimates for new units. In the next section, I define the complete and partial pooling models that I explore in the rest of the chapter. If it was suspected that the measurement error varied between the units, then the same methods that I describe for varying μ and ν in Sec. 5.1.2 bellow could be used for σ .

5.1.1 The complete pooling model

I denote by y_{ij} , $j = 1, 2, \dots, J$, the measured degradation of J identical units, and without loss of generality, assume that they are measured at the same times t_i , $i = 1, 2, \dots, I$. In a complete pooling model for the crack growth data, both parameters μ and ν are completely pooled between the ten units. In other words, each unit is a realisation from the same underlying gamma degradation process.

I specify the complete pooling model

$$\begin{aligned}
y_{ij}|z_{ij}, \sigma &\sim N(z_{ij}, \sigma) && \text{data model} \\
\Delta z_{ij}|\mu, \nu &\sim Ga\left(\frac{\Delta t_i}{\nu^2}, \frac{1}{\mu\nu^2}\right) && \text{process model} \\
\sigma &\sim \text{Unif}(0, 10) && \text{parameter model} \\
\mu &\sim N^+(0.5, 0.2) \\
\nu &\sim t_3^+(0, 0.5)
\end{aligned}$$

for the crack growth data. Note that this model is essentially the same model as I explored in Chap. 4 except there are now multiple realisations from the gamma process for each Δt_i corresponding to the jump in degradation from each unit. I also use new values of the hyperparameters that are adjusted to the scale of the crack growth phenomenon. These new priors were selected using prior predictive checking.

Figure 5.2 shows four prior predictive simulations from the noisy crack growth model with complete pooling. In the figure, each simulated dataset contains the same number of units and observations as the true data set in Fig. 5.1. Clearly, the simulations are noticeably different from the true data; in the first simulation, the units wear much faster; in the second and third, the degradation traces are much more volatile; and in the fourth, there is almost no variability between the pathways. However, as I stressed in Sec. 1.4, the point of performing prior predictive checks is not to tune the prior until it matches the observed data

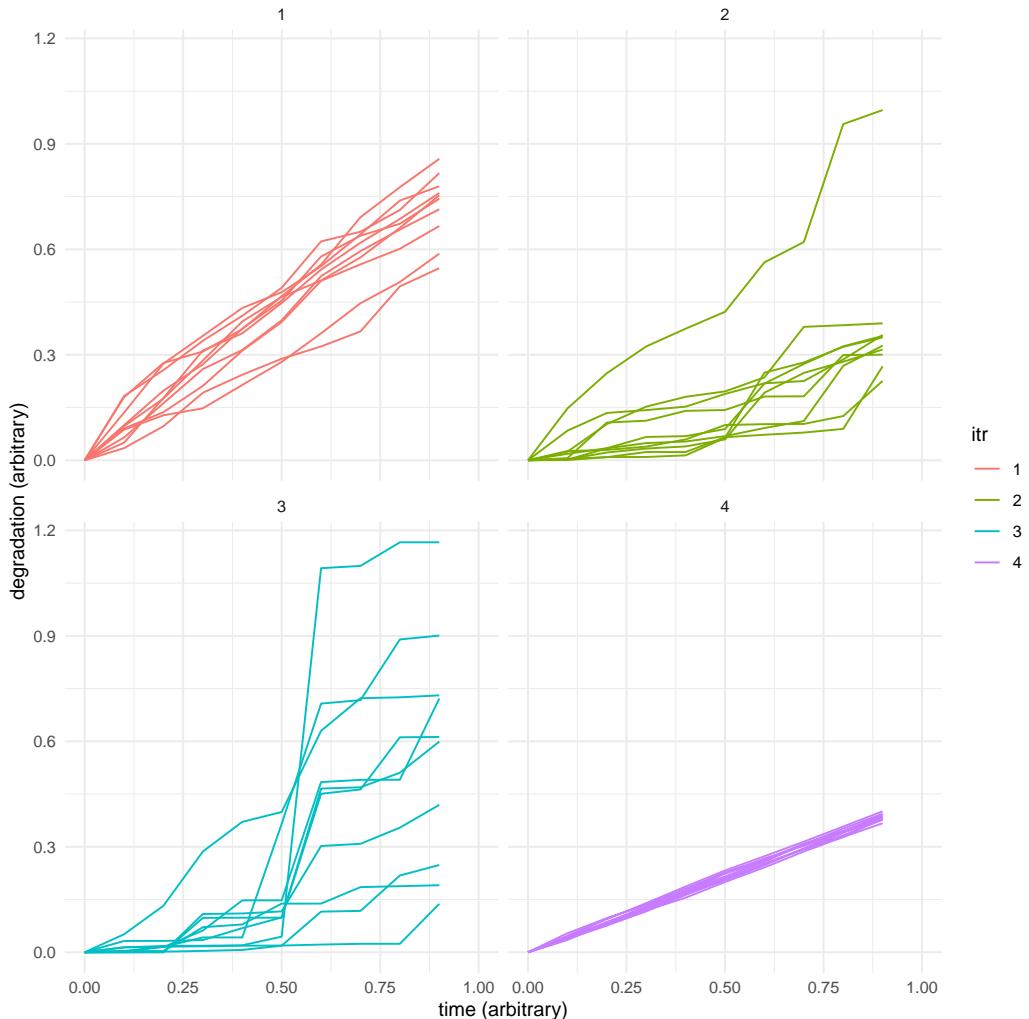


Figure 5.2: Four non-noisy prior predictive simulations from the complete pooling model. Each simulation contains the same number of units and observations as the crack growth data in Fig. 5.1.

but rather to ensure that the model produces plausible realisations of the data, which appears to be the case here. Interestingly, through the prior predictive simulations, we can see that the volatility of the gamma process with no varying parameters already allows for a reasonable amount of variation in the degradation traces of the units. Next, I explain how I extended the complete pooling model to allow the parameters to vary between units.

5.1.2 Partial pooling models

To incorporate unit-to-unit variability into the complete pooling model above, I assign a hierarchical prior on μ , ν , or both μ and ν and then estimate the hyperparameters from the data. I use the general structure

$$\begin{aligned}\theta_j | \mu_\theta, \sigma_\theta &\sim N^+(\mu_\theta, \sigma_\theta) \\ \mu_\theta &\sim {}_\theta\pi \\ \sigma_\theta &\sim \text{Cauchy}^+(0, 1)\end{aligned}$$

for the hierarchical prior, where θ represents the parameter that we are allowing to vary between units, θ_j are the unit-specific parameters, and ${}_\theta\pi$ the corresponding prior for that parameter in the complete pooling model. In this hierarchical prior distribution, I assume that the θ_j arise from a Gaussian distribution whose hyperparameters μ_θ and σ_θ are to be estimated. In this way, information is shared across units since a unit-specific estimate of the parameter influences the hyperparameters and, hence, the other unit-specific parameters. I use the same prior for μ_θ as the prior used for the completely pooled case of the parameter θ in the complete pooling model since μ_θ now expresses the expected value of the θ_j . Finally, I use a vague truncated Cauchy hyperprior for the standard deviation of the hierarchical prior following the recommendations of Gelman, Carlin, et al. (2020, chap. 17). These choices are just a general starting point; the more mass close to zero in the hyper prior for σ_θ , the more information is pooled between the units (McElreath, 2020), and if I were to use a distribution with heavier tails than a Gaussian, such as the Student's t , then inference about the hyperparameters would be more robust to outlying units (Gelman, Carlin, et al., 2020, chap. 17). Using this general structure of a hierarchical prior, I explore a varying μ model, varying ν model and a model where both μ and ν vary from unit-to-unit.

varying μ model In the varying μ model, I am assuming that each of the degradation traces in Fig.5.1 arise from different gamma processes where these processes have the same volatility (described by ν) and similar, but not the same, average degradation rates. The varying μ model is specified as

$$\begin{aligned}
y_{ij}|z_{ij}, \sigma &\sim N(z_{ij}, \sigma) && \text{data model} \\
\Delta z_{ij}|\mu_j, \nu &\sim Ga\left(\frac{\Delta t_i}{\nu^2}, \frac{1}{\mu_j \nu^2}\right) && \text{process model} \\
\sigma &\sim \text{Unif}(0, 10) && \text{parameter model} \\
\mu_j &\sim N^+(\mu_\mu, \sigma_\mu) \\
\nu &\sim t_3^+(0, 0.5) \\
\mu_\mu &\sim N^+(1, 0.2) \\
\sigma_\mu &\sim \text{Cauchy}^+(0, 1).
\end{aligned}$$

varying ν model In the varying ν model, I once again assume that each degradation trace is a realisation from a different gamma process. However, this time, I assume that all of these processes share the same average degradation rate μ but have varying degrees of volatility, ν . I specify the varying ν model as

$$\begin{aligned}
y_{ij}|z_{ij}, \sigma &\sim N(z_{ij}, \sigma) && \text{data model} \\
\Delta z_{ij}|\mu, \nu_j &\sim Ga\left(\frac{\Delta t_i}{\nu_j^2}, \frac{1}{\mu \nu_j^2}\right) && \text{process model} \\
\sigma &\sim \text{Unif}(0, 10) && \text{parameter model} \\
\mu &\sim N^+(1, 0.2) \\
\nu_j &\sim N^+(\mu_\nu, \sigma_\nu) \\
\mu_\nu &\sim t_3^+(0, 0.5) \\
\sigma_\nu &\sim \text{Cauchy}^+(0, 1).
\end{aligned}$$

varying μ and ν model In the final and most flexible model, I assume that the gamma processes that each degradation trace arises from have unique values

of μ and ν . This model where both μ and ν are partially pooled is

$$\begin{aligned}
y_{ij}|z_{ij}, \sigma &\sim N(z_{ij}, \sigma) && \text{data model} \\
\Delta z_{ij}|\mu_j, \nu_j &\sim Ga\left(\frac{\Delta t_i}{\nu_j^2}, \frac{1}{\mu_j \nu_j^2}\right) && \text{process model} \\
\sigma &\sim \text{Unif}(0, 10) && \text{parameter model} \\
\mu_j &\sim N^+(\mu_\mu, \sigma_\mu) \\
\nu_j &\sim N^+(\mu_\nu, \sigma_\nu) \\
\mu_\mu &\sim t_3^+(0, 0.5) \\
\sigma_\mu &\sim \text{Cauchy}^+(0, 1) \\
\mu_\nu &\sim t_3^+(0, 0.5) \\
\sigma_\nu &\sim \text{Cauchy}^+(0, 1).
\end{aligned}$$

These models can be seen as a set of nested models where the varying μ , varying ν , and complete pooling models are special cases of the model where both μ and ν vary. In models where either μ , ν , or both are constant across the different units, the complete pooling is equivalent to a model in which the hyperparameters σ_μ or $\sigma_\nu \rightarrow 0$ and hence the unit specific parameters are forced to be equal to the mean hyperparameters μ_μ or μ_ν .

5.2 Computation, posteriors, and predictive distributions

I fit all four models using the probabilistic programming language Stan (Stan Development Team, 2022). The code is available on a GitHub repository. The HMC algorithm is remarkably efficient, particularly for the complete pooling model, and exploring the posterior distributions requires only six chains of length 1000 after a burn-in period of 1000 iterations. The n_{eff} and \hat{R} statistics for the parameters of interest for all four models indicate that, in all cases, the chains

Table 5.1: The number of divergent transitions that occur during sampling.

| model | number of divergent transitions |
|---------------------------|---------------------------------|
| complete pooling | 0 |
| partial pooling mu | 23 |
| partial pooling nu | 64 |
| partial pooling mu and nu | 47 |

have mixed well (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2021), these statistics are shown in Tabs. 5.2–5.5 for the different models. During sampling from the posteriors of the four hierarchical models, some divergent transitions occur—particularly in the models where ν varies from unit-to-unit. Table 5.1 lists the number of divergent transitions that occur while sampling from the posterior of each model.

In the first part of this section, I summarise the inference from each model and show that all models have been able to reclaim the scale of the measurement error and the true underlying degradation paths from the noisy data. I also investigate the cause of the divergent transitions for the hierarchical models, identifying the cause to be the tight curvature in the posterior distributions where the partial pooling collapses towards the complete pooling case. Posterior predictive checking is used to understand the practical differences between the fitted models. Section 5.2.1 then compares the four models using the $\text{elpd}_{\text{LOO-cv}}$ scoring method described in Chap. 1 Sec. 1.4.

Complete pooling The posterior samples from the complete pooling model for parameters σ , μ , and ν are summarised in Tab. 5.2. The model has done a reasonable job at reclaiming the standard deviation of the measurement error (which is 0.025) since the expected value of σ is 0.030 and the majority of the posterior mass sits between $\sigma = 0.020$ and $\sigma = 0.040$. The BHM also provides posterior distributions of the underlying degradation. These are shown in Fig. 5.3 as 95% credible intervals, along with the noisy data and the true underlying degradation traces from which they were generated. As we can see, the credible

Table 5.2: Output from fitting a model with complete pooling to the noisy data of Fig. 5.1. We assume that the data from all units is a manifestation of a single underlying gamma process, and hence the mean and coefficient of variation of the process do not vary from unit-to-unit.

| Parameter | Mean | 2.5% | 50% | 97.5% | n_{eff} | \hat{R} |
|-----------|------|------|------|-------|------------------|-----------|
| σ | 0.03 | 0.02 | 0.03 | 0.04 | 2003 | 1.00 |
| μ | 0.38 | 0.33 | 0.38 | 0.44 | 7889 | 1.00 |
| ν | 0.21 | 0.15 | 0.21 | 0.29 | 652 | 1.01 |

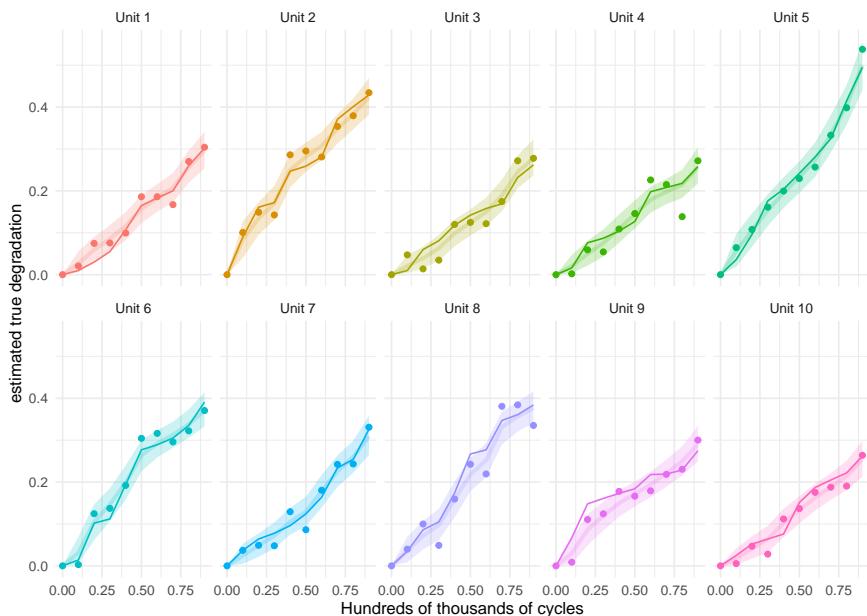


Figure 5.3: Marginal posterior distributions of the underlying gamma process from a BHM where all parameters are completely pooled.

intervals contain the underlying true degradation over the entire time span for each of the ten units, with few exceptions.

Varying μ Table 5.3 shows some summary statistics of the marginal posterior distributions from the varying μ model for select model parameters, and Fig. 5.4 shows the marginal posterior distributions of the parameters $\sigma, \nu, \mu_1, \dots, \mu_{10}, \mu_\mu$ and σ_μ . As Tab. 5.3 and Fig. 5.4 show, a posteriori, the mean degradation rates of the units arise from the distribution $N^+(0.38, 0.07)$. The small expected standard deviation of 0.07 indicates that the unit-specific mean degradation rates vary in

Table 5.3: Partial output from fitting a BHM to the noisy data of Fig. 5.1 where mean degradation μ_j varies between units. Only statistics for Units 1–4 are shown.

| Parameter | Mean | 2.5% | 50% | 97.5% | n_{eff} | \hat{R} |
|--------------|------|------|------|-------|------------------|-----------|
| σ | 0.03 | 0.02 | 0.03 | 0.04 | 747 | 1.01 |
| μ_1 | 0.36 | 0.26 | 0.36 | 0.47 | 2150 | 1.00 |
| μ_2 | 0.42 | 0.33 | 0.42 | 0.56 | 954 | 1.00 |
| μ_3 | 0.35 | 0.25 | 0.35 | 0.46 | 1001 | 1.00 |
| μ_4 | 0.34 | 0.23 | 0.34 | 0.45 | 746 | 1.01 |
| ν | 0.18 | 0.09 | 0.18 | 0.28 | 248 | 1.03 |
| μ_μ | 0.38 | 0.32 | 0.38 | 0.46 | 3079 | 1.00 |
| σ_μ | 0.07 | 0.01 | 0.06 | 0.17 | 343 | 1.01 |

a relatively narrow range, as the posterior distributions in Fig. 5.4 indicate. In addition, the lower tail of the marginal posterior of σ_μ has a considerable mass near zero, and hence, there is reasonable evidence that the average degradation rate is constant across units.

Interestingly, the marginal posterior of μ_μ is wider relative to μ in the CP model; however, both have the same mean. The uncertainty intervals of the unit-specific μ_j are even wider still. A possible reason for this loss of precision is that because I am not making the simplifying assumption that all of the μ_j are equal, the data do not inform the parameters as strongly since there are now more parameters to estimate in the model, and hence the uncertainty is larger. Additionally, the estimate of ν has shrunk slightly (particularly in the lower tail). From the shrinkage of ν , it appears that because more of the variability between the traces is being attributed to the variation of the μ_j , the resulting traces are less volatile. Despite these slight changes in inference regarding the mean wear rates and coefficient of variation, the varying μ model reclaims the true value of σ to effectively the same degree as the complete pooling model.

During the sampling from the posterior of the varying μ model, a small number (twenty-three) of divergent transitions occur. Figure 5.5 show the pairs plots of the samples from the posterior for the parameters μ_1, μ_2, μ_μ , and σ_μ . The twenty-

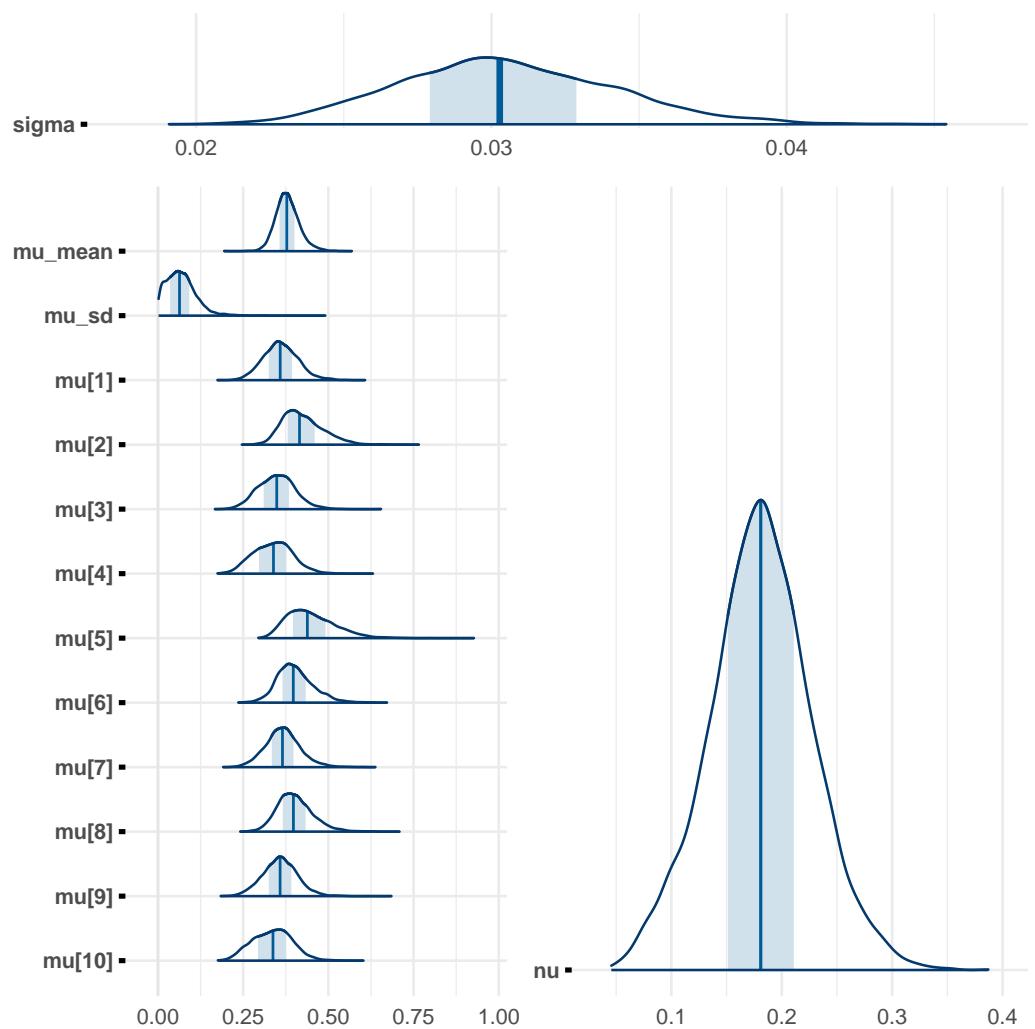


Figure 5.4: Marginal posterior distributions of parameters from the model where mean degradation rate μ_j varies from unit-to-unit.

three divergent transitions are plotted in red in the bivariate scatter plots. In the plots (particularly those in the lower off-diagonal), the divergent trajectories tend to occur for very small values of σ_μ . This pattern suggests that in the area of the posterior where the partial pooling model collapses towards the simpler complete pooling case—when $\mu_\mu = \mu_1 = \dots = \mu_{10}$ and $\sigma_\mu = 0$ —there is tight curvature, and hence the sampler starts to misbehave. To confirm this hypothesis, Fig. 5.6 show the parallel coordinate plot for σ , μ_μ , μ_1, \dots, μ_{10} , and σ_μ with the non-divergent traces plotted in blue and the divergent traces plotted in red. Tracking the divergent traces through the parameter space, it is clear that the divergencies occur when all the unit specific μ_j are very close to the mean hyperparameter μ_μ and the standard deviation hyperparameter is very close to zero.

Despite the issues with sampling, in the varying μ model’s posterior predictive distributions for the reclaimed ‘non-noisy’ degradation traces match the true degradation traces of the units very closely. Figure 5.7 shows the posterior predictive distributions of each unit’s degradation from the varying μ model. Like with the complete pooling case, the varying μ model’s the 95% posterior predictive intervals for the underlying degradation contain the true non-noisy degradation traces most of the time.

Varying ν When ν is allowed to vary between units instead of μ , all of the ν_j are effectively equal to the mean μ_ν , resulting in most of the mass of σ_ν being very close to zero; Tab. 5.4. In other words, there is little evidence that ν varies unit-to-unit. Like in the posterior of the varying μ model, the expected value of the hierarchical prior, in this case, μ_ν , is effectively the same as the expected value of the parameter in the complete pooling posterior, ν , and the uncertainty intervals are slightly wider in the partial pooling case. As for the completely pooled parameters, the marginal posteriors of μ in Tab. 5.4 and 5.2 are almost identical, as are the marginal posteriors of σ . The posterior predictive distributions of the degradation traces (not shown) also look similar to the complete pooling and varying μ cases.

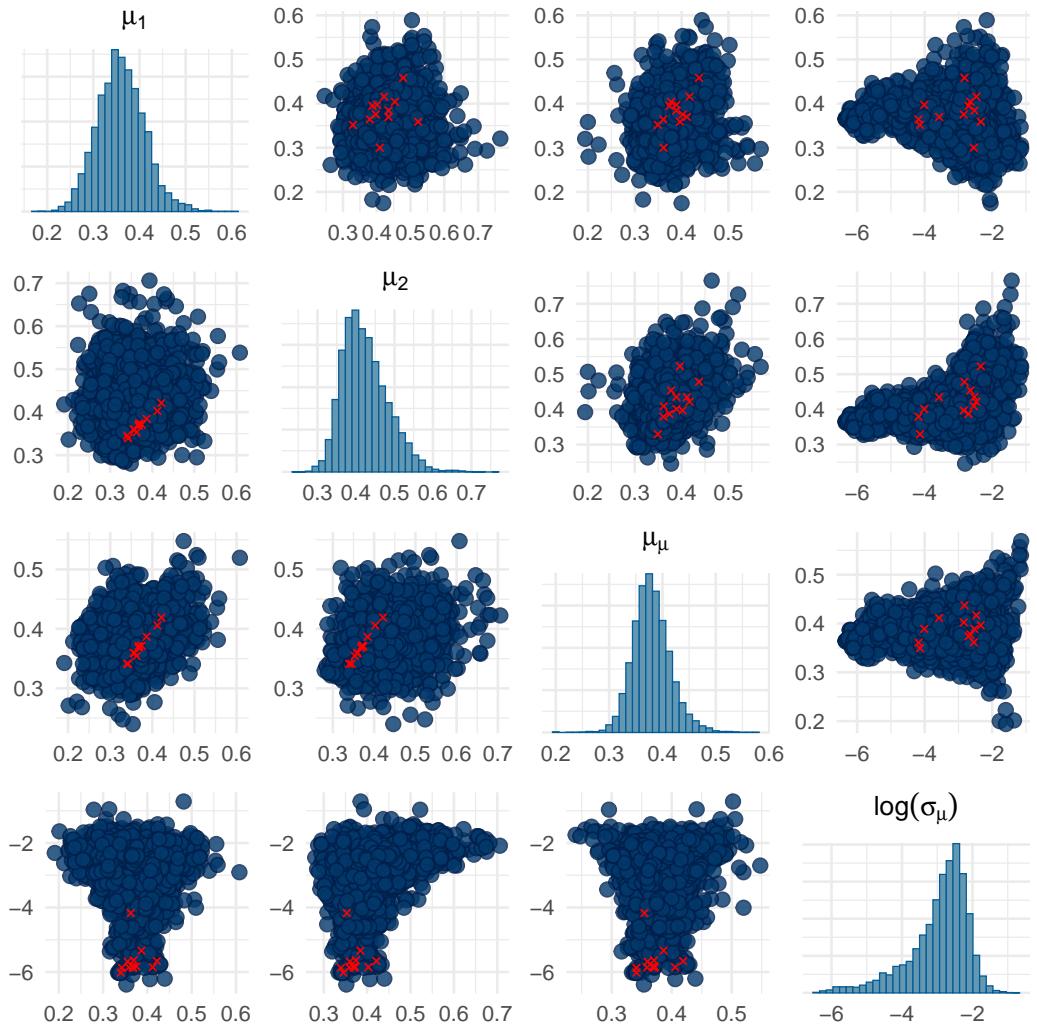


Figure 5.5: A pairs plot of the posterior samples of μ_1 , μ_2 , μ_μ , and $\log(\sigma_\mu)$ from the varying μ model. In each bivariate plot, divergencies are plotted in red.

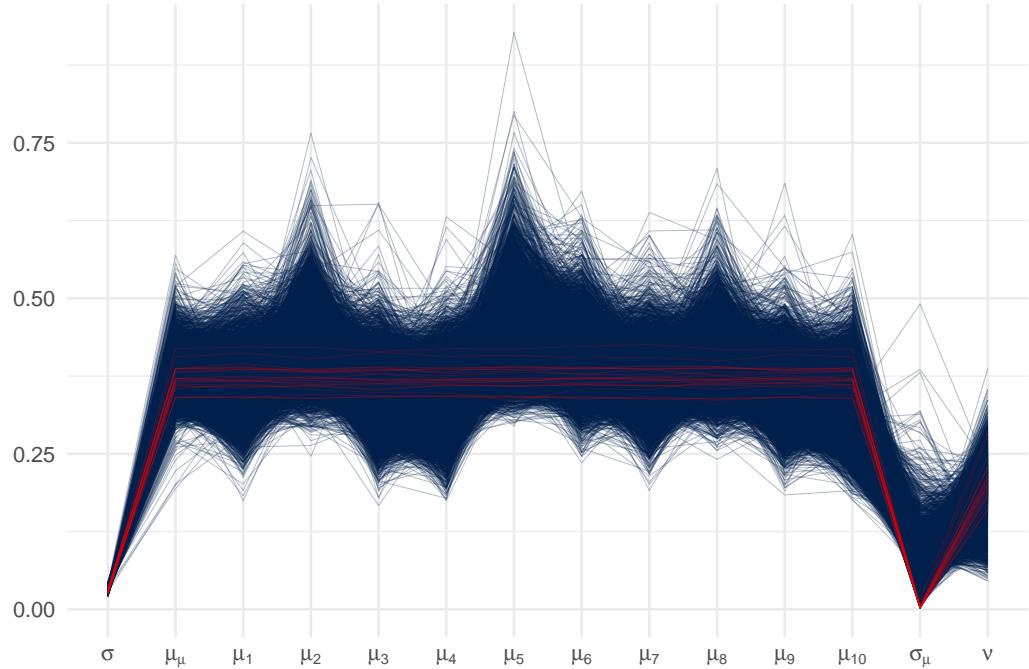


Figure 5.6: Parallel coordinate plot for the parameters and hyper parameters of the varying μ model. The divergent traces are plotted in red.

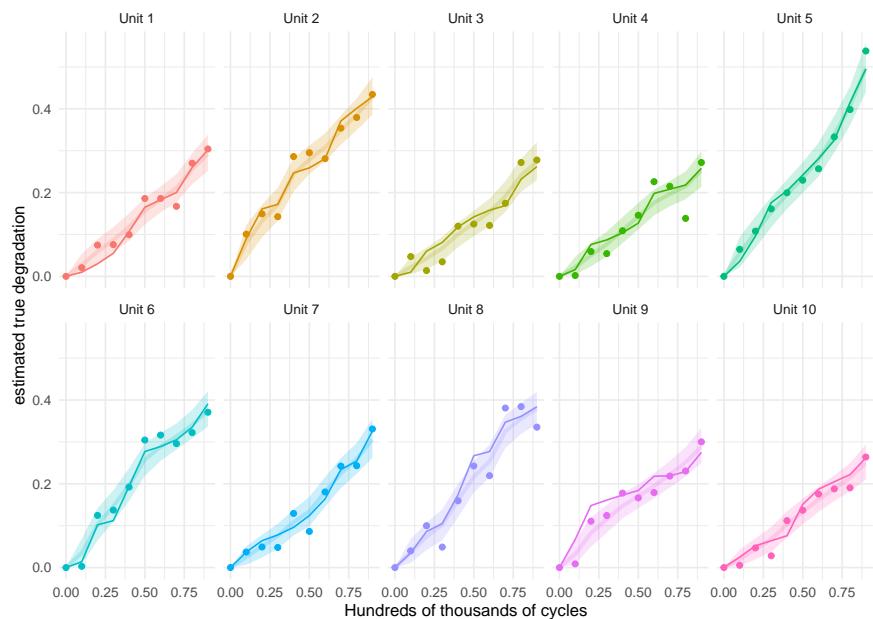


Figure 5.7: Marginal posterior distributions of the underlying gamma process from a BHM where mean degradation rate μ_j varies from unit-to-unit.

Table 5.4: Partial output from fitting a BHM to the noisy data of Fig. 5.1 where the coefficient of variation ν_j varies between units. Only statistics for Units 1–4 are shown.

| Parameter | Mean | 2.5% | 50% | 97.5% | n_{eff} | \hat{R} |
|--------------|------|------|------|-------|------------------|-----------|
| σ | 0.03 | 0.02 | 0.03 | 0.04 | 1856 | 1.00 |
| ν_1 | 0.21 | 0.10 | 0.21 | 0.32 | 735 | 1.00 |
| ν_2 | 0.22 | 0.14 | 0.22 | 0.33 | 910 | 1.01 |
| ν_3 | 0.22 | 0.13 | 0.22 | 0.33 | 755 | 1.00 |
| ν_4 | 0.22 | 0.13 | 0.22 | 0.33 | 811 | 1.00 |
| μ | 0.38 | 0.32 | 0.37 | 0.44 | 2598 | 1.00 |
| μ_ν | 0.22 | 0.15 | 0.22 | 0.31 | 598 | 1.00 |
| σ_ν | 0.03 | 0.00 | 0.03 | 0.10 | 284 | 1.02 |

Like the varying μ model, divergent transitions occur while sampling from the posterior of the varying ν model. However, there are almost five times more divergencies when fitting the varying ν model: Tab. 5.1. Figure 5.8 shows the parallel coordinate plot of the MCMC draws for the parameters σ , μ_ν , ν_1 , \dots , ν_{10} , and σ_ν . Like with varying μ model, the divergent transitions diagnose a degenerate area in the posterior around $\mu_\nu = \nu_1 = \dots = \nu_{10}$ and $\sigma_\nu = 0$. In the case of the varying ν model, there is even more posterior mass around this area since there is little evidence that ν should vary from unit to unit, and so a higher number of divergent transitions occur.

varying μ and ν The posterior of the varying μ and ν model is summarised in Tab. 5.5. When both parameters are allowed to vary unit-to-unit, the results are almost a synthesis of the two models where only one of the parameters varies from unit-to-unit. In Tab. 5.5, the summaries of the unit specific μ_j match the results of the varying μ model almost exactly (Tab. 5.3), and like for the varying ν model, there is very little variation between ν_1 , ν_2 , and μ_ν and σ_ν has considerable mass near zero. In Tab. 5.5, the marginal posterior distributions of the unit specific ν_j and their mean μ_ν match the completely pooled estimate of ν from the varying μ model. While generating samples from the posterior of the varying μ and ν

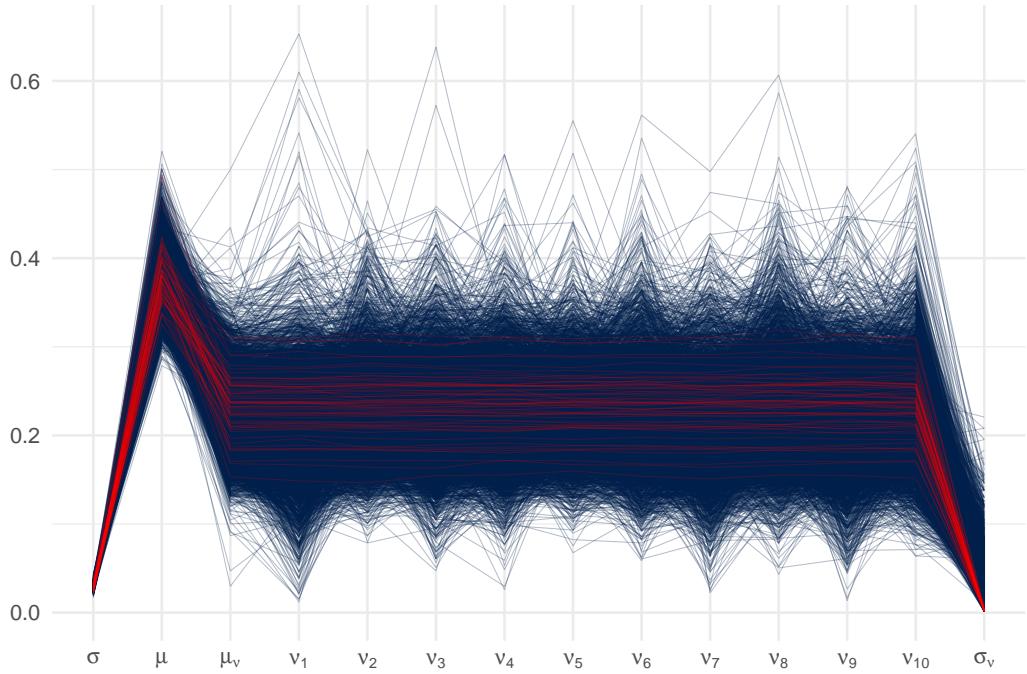


Figure 5.8: Parallel coordinate plot for the parameters and hyper parameters of the varying ν model. The divergent traces are plotted in red.

model, the poor behaviour of the sampler is even more severe than the models where only μ or ν vary: Tab. 5.1 shows that roughly 3% of the transitions are divergent. Still, the varying μ and ν model is able to reclaim the true scale of the measurement error and underlying degradation traces of the units.

Posterior predictive checks Besides the sampling issues, it is hard to differentiate one model from the others since in all of the posteriors σ and the underlying degradation paths of the units are reclaimed to roughly the same level. The easiest way to understand the practical differences between the four models is to look at posterior simulations of new datasets. Figure 5.9 shows three posterior predictive simulations generated from the posteriors of the different models. In each subplot, each of the different colours/line types indicates a posterior predictive simulation with the same number of units and observations as the original data. These simulations from the posterior now look a lot more like the observed data than the prior predictive simulations in Fig. 5.2 did. In Fig. 5.9, there is

Table 5.5: Partial output from fitting a BHM to the noisy data of Fig. 5.1 where both the coefficient of variation ν_j and the mean wear rate μ_j varies between units. Only statistics for Units 1–4 are shown.

| Parameter | Mean | 2.5% | 50% | 97.5% | n_{eff} | \hat{R} |
|--------------|------|------|------|-------|------------------|-----------|
| σ | 0.03 | 0.02 | 0.03 | 0.04 | 1005 | 1.01 |
| ν_1 | 0.19 | 0.07 | 0.19 | 0.31 | 467 | 1.01 |
| ν_2 | 0.20 | 0.10 | 0.19 | 0.32 | 508 | 1.02 |
| ν_3 | 0.20 | 0.08 | 0.20 | 0.33 | 529 | 1.02 |
| ν_4 | 0.20 | 0.08 | 0.19 | 0.33 | 493 | 1.02 |
| μ_1 | 0.36 | 0.26 | 0.36 | 0.48 | 2926 | 1.00 |
| μ_2 | 0.42 | 0.32 | 0.41 | 0.57 | 1049 | 1.01 |
| μ_3 | 0.35 | 0.25 | 0.35 | 0.47 | 2000 | 1.00 |
| μ_4 | 0.34 | 0.23 | 0.34 | 0.47 | 1532 | 1.01 |
| μ_μ | 0.38 | 0.31 | 0.38 | 0.46 | 5668 | 1.00 |
| σ_μ | 0.07 | 0.01 | 0.06 | 0.17 | 321 | 1.02 |
| μ_ν | 0.19 | 0.10 | 0.19 | 0.29 | 370 | 1.03 |
| σ_ν | 0.04 | 0.00 | 0.03 | 0.11 | 329 | 1.02 |

little difference between plots (a) and (c) as well as between plots (b) and (d) since there is so little variation between the unit-specific ν_j when we allow ν to vary unit to unit. The main difference between the plots is when μ is allowed to vary between units. In Fig. 5.9 (b) and (d), where μ varies from unit-to-unit, the spread of the degradation traces is wider, and the paths are slightly straighter than in Fig. 5.9 (a) and (c), where μ is constant. The posterior simulations from the models where μ is completely pooled look more like the true data. However, it is not possible to say that the data in Fig. 5.1 did not arise from any one of the models; some of the simulated datasets in the two models where μ varies from unit-to-unit also look very similar to the observed data. With no obvious choice by visually evaluating the posteriors, I quantitatively compare the models using cross-validation in the next section.

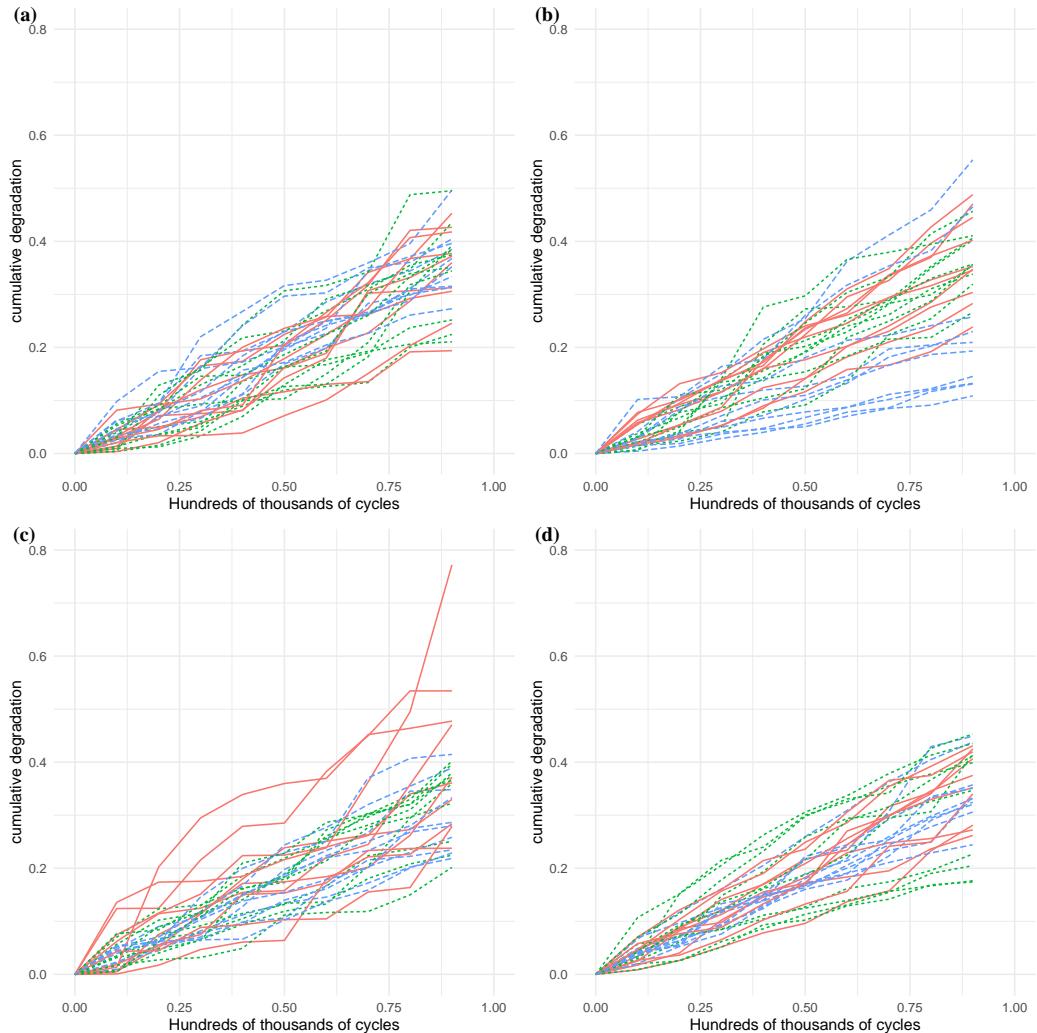


Figure 5.9: Posterior predictive simulations from each of the four models. Each simulated dataset, is indicated by a different colour and line type. Plot (a) shows three simulated datasets from the fitted complete pooling model, (b) from the model with varying μ , (c) from the varying ν model, and (d) from the model where both μ and ν vary from unit-to-unit.

5.2.1 Model comparison

Having chosen four suitable models, fitted them to the data, and checked that the sampling and inference appear reasonable, the next step is to compare the models. To do so, I use the elppd and cross-validation methods introduced in Sec. 1.4. In the case of the crack growth data, as mentioned previously, the motivation for analysis is to predict the failure time of new units and the units currently under test but yet to fail. Doing so requires that the models predict the future degradation of the units under test and any new, unobserved units. Hence, it makes sense to compare the four models with respect to these two things.

The ability to predict the degradation of new units can be estimated from eq. (1.12) by treating the I observations from each unit j collectively as a single new observation, i.e., $y_j = \{y_{1j}, \dots, y_{Ij}\}$. I refer to this method as leave-one-unit-out cross-validation (LOUO-CV) and the elppd score calculated in this way the $\text{elppd}_{\text{LOUO-CV}}$. By contrast, the elppd for new observations from the units under test can be approximated using eq. (1.12) by sequentially withholding the final observation, y_{Ij} , from each of the J units; I refer to this method as step-ahead cross-validation (SA-CV) and the corresponding score as $\text{elppd}_{\text{SA-CV}}$. In both cases, I construct the likelihood of the withheld observations in the same way since in the data model it is assumed that the noisy observations y_{ij} from the different units are independent and normally distributed conditional on the true underlying degradation, $\tilde{z}_{i,j}$, and the measurement error, σ . Although, the definition of the posterior predictive distribution of \tilde{z} that should be used depends on both the cross-validation method (LOUO-CV or SA-CV) and the model structure (i.e., complete pooling or partial pooling). The details of constructing the posterior predictive distribution of \tilde{z} for LOUO-CV or SA-CV and the results are outlined below, and the code may be found on the GitHub repository.

Leave-one-unit-out cross-validation To calculate $\text{elppd}_{\text{LOUO-CV}}$, I iteratively withhold the data from each unit j , condition on the data from the remaining $J-1$

Table 5.6: Leave-one-out cross-validation statistics for the models fitted in Section 5.2.

| | elppd _{LOOO-CV} | elppd _{SA-CV} |
|-------------------------|--------------------------|------------------------|
| complete pooling | 154.3189 | 15.17704 |
| varying μ | 153.5409 | 14.00906 |
| varying ν | 153.2858 | 15.12410 |
| varying μ and ν | 153.4567 | 15.07771 |

units and then calculate the log-likelihood of the withheld unit's observations, $y_j = \{y_{1j}, \dots, y_{Ij}\}$. This calculation is based on posterior predictive draws of a new unit's (the withheld unit) filtered degradation path under the fitted model, $\tilde{z}_j = \{\tilde{z}_{1j}, \dots, \tilde{z}_{Ij}\}$, and the posterior draws of σ . Thus, elppd_{LOOO-CV} is expressed as

$$\text{elppd}_{\text{LOOO-CV}} = \sum_{j=1}^J \sum_{i=1}^I \log \frac{1}{S} \sum_{s=1}^S p(y_{ij} | [\tilde{z}_{ij}, \sigma]_{-j}^s). \quad (5.1)$$

To generate posterior predictive draws of the non-noisy degradation path of a new unit, I sample $I - 1$ jumps $\Delta \tilde{z}_{ij}^s$ in degradation from $\text{Ga}([\tilde{\mu}_j, \tilde{\nu}_j]_{-i}^s)$ and then calculate their cumulative sum to generate the degradation path. If μ is completely pooled, the $\tilde{\mu}_j^s$ are taken from posterior draws μ^s ; similarly, if ν is completely pooled, the $\tilde{\nu}_j^s$ are posteriors draws ν^s . If, however, the mean degradation varies across units, $\tilde{\mu}_j^s$ is sampled from the (hierarchical prior) $N^+(\mu_\mu^s, \sigma_\mu^s)$; in the same way, $\tilde{\nu}_j^s$ would be also sampled from $N^+(\mu_\nu^s, \sigma_\nu^s)$ if the coefficient of variation varied across units. For the models discussed in Sec. 5.1.1 and 5.1.2, the first column of Tab. 5.6 shows the elppd_{LOOO-CV} scores calculated in this way.

Step-ahead cross-validation Step-ahead cross-validation is carried out by iteratively withholding the most recent observation from each of the units under test, and the SA-CV estimate of elppd is calculated as

$$\text{elppd}_{\text{SA-CV}} = \sum_{j=1}^J \log \frac{1}{S} \sum_{s=1}^S p(y_{Ij} | [\tilde{z}_{Ij}, \sigma]_{-[Ij]}^s). \quad (5.2)$$

To generate the posterior predictive draws in this case, I sample the jump in degradation for unit j from $\text{Ga}(\mu_j, \nu_j)$ and then add this jump to the posterior draws of $\tilde{z}_{I-1,j}$. Where either μ or ν are completely pooled, μ_j and ν_j are posterior draws μ^s and ν^s , respectively; otherwise, the draws from the posterior distributions of the unit-specific parameters of the gamma process are used. The $\text{elppd}_{\text{SA-CV}}$ scores for each of the different models are shown in the right-hand column of Table 5.6.

As the elppd results in Table 5.6 show, for the LOOU case, the models where μ is allowed to vary between units perform slightly better, whereas, for the SA scores, the units where μ is constant across units perform slightly better. However, the difference in both cases is marginal. The fact that the complete pooling model has the highest $\text{elppd}_{\text{SA-CV}}$ score—and that the inference from the complete pooling model and the partial pooling models is so similar—suggests that the completely pooled gamma process with measurement error is sufficient to model the variability in the degradation traces. However, the added variability of the unit-specific μ_j helps slightly when generalising to new units, although it is worth noting that the closeness of the models means that the elppd results are sensitive to the priors. In Leadbetter, Caceres, and Phatak (2024) we used a slightly miss-specified prior for μ and got a different result: the complete pooling model had the highest elppd scores for both LOOU-CV and SA-CV.

In this section, I have shown how to fit different noisy gamma process models for multiple units and demonstrated a principled way of evaluating and comparing them to identify the most appropriate one for the data set being analysed. In the case of the crack growth data, there is a marginal difference between all of the models I explore; this may not be the case for other datasets. The elppd scores suggest that both the completely pooled and a model where μ varies between units are useful for predicting the degradation of the current units under test and new units, respectively. From both posterior predictive checking and elppd scores, there is little motivation to allow ν to vary between units compared to the

simpler constant ν alternative. Holding ν constant also results in better-behaved sampling. Therefore, in the next section, I construct failure time distributions for the units under test as well as for new units using only the complete pooling and varying μ model.

5.3 Failure time distributions

As noted in the introduction, the crack growth measurements are collected to estimate the failure time distribution of individual units that are in-service but have not failed and/or of new units with the same nominal specifications as the experimental units. The soft failure of the terminals is considered to be when the crack length exceeds $z_f = 0.4\text{mm}$. For degradation models and a soft definition of failure, the failure time T can be defined as the first passage time when the true degradation path crosses the failure threshold z_f (Balakrishnan, Tsai, & C.-T., 2017), that is,

$$T = \inf [t | Z_t \geq z_f].$$

Note that it depends on the true degradation path, not on the observed one, and hence, it does not involve the measurement error (Hamada et al., 2008b). In the Bayesian context, we write the failure time distribution as $F_{T|\Theta}(t)$, where Θ denotes the (hyper) parameters in the model. It can, therefore, be written as

$$F_{T|\Theta}(t) = p(T < t | \Theta) = p(Z_t > z_f | \Theta).$$

One of the advantages of using a fully Bayesian treatment is that we can use the posterior predictive distribution of the underlying degradation Z_t to calculate the failure time distribution, thereby incorporating uncertainty in the parameters, which will be reflected in credible intervals for $F_{T|\Theta}(t)$. Alternatives to a fully Bayesian treatment include, for example, bootstrapping (Peng & Ai, 2018).

Although there is no explicit expression for $F_{T|\Theta}(t)$ it is straightforward to

Should this instead be the posterior? Or the params, int quantities, and hyper-params?

| Complete pooling | Varying μ |
|--|---|
| <ol style="list-style-type: none"> 1. Draw a sample from the posterior distribution of (μ, ν); 2. Given that $Z_t \mu, \nu \sim \text{Ga}(t/\nu^2, 1/\mu\nu^2)$, calculate $p(Z_t > z_f)$ numerically for a range of values of t to generate one draw of $F_T(t)$; 3. Repeat Steps 1. and 2. n_{sim} times. | <ol style="list-style-type: none"> 1. Draw a sample from the posterior distribution of $(\mu_\mu, \sigma_\mu, \nu)$; 2. Generate μ_j from $N^+(\mu_\mu, \sigma_\mu)$; 3. Using the μ_j from Step 2. and the corresponding value of ν in Step 1., generate a draw of $F_T(t)$ for a range of values of t as in Step 2. for complete pooling; 4. Repeat Steps 1.–3. n_{sim} times. |

Table 5.7: Algorithms for calculating the posterior distribution of the failure time distribution $F_T(t)$ for the complete pooling and varying μ models.

obtain the posterior distribution of $F_{T|\Theta}(t)$ by simulation and numerical evaluation of the distribution function of a gamma distribution (e.g., by using the R function `pgamma`), using a modified version of the procedure outlined by Hamada et al. (2008b, Sec. 8.2.1). For the complete pooling and varying μ models, the algorithms are shown in Tab. 5.7.

Figure 5.10 shows the posterior of the failure time distributions for new units, calculated from the complete pooling and varying μ models. There is considerably greater uncertainty in $F_T(t)$ from the partial pooling model, but this is not surprising: in addition to the inherent variability of the gamma process, the partial pooling model also includes the variability in the μ_j .

Using slight modifications of the algorithms shown in Tab. 5.7, we can also calculate the failure time distribution for a unit that is currently under test and that has yet to fail, for example, Unit 3. This distribution, also known as the predictive failure time distribution (Lawless & Crowder, 2004), is conditional on the unit not having failed by t_I and having attained a degradation level z_I . Since z_I is a latent variable in the model, the posterior distribution contains samples of z_I from which we can calculate the jump in degradation that corresponds to a soft failure. Figure 5.11 shows the posterior predictive failure time distributions

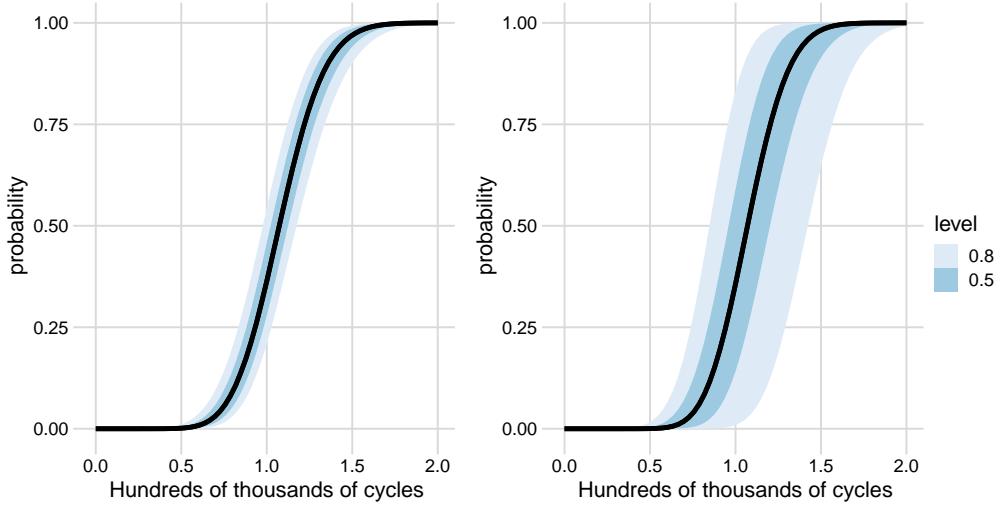


Figure 5.10: Posterior distributions of the failure time distributions from the complete pooling model (left) and the varying μ model (right).

for Unit 3, which has not failed; again, we see that because of the additional layer of uncertainty, the credible intervals for the varying μ model are wider than those from the complete pooling model, although the difference is not as extreme compared to the failure time distributions for new units. Comparing Fig. 5.10 with Fig. 5.11, the unit-specific failure time distributions have tighter uncertainty intervals than their new unit counterparts. This difference is more significant for the varying μ model since the unit-specific estimate does not average over the variability in the μ_j and so is a much more precise estimate.

5.4 Discussions

In this chapter, I showed how the noisy gamma process model for a single degradation path from Chap. 4 could be extended—using the same hierarchical modelling framework—to incorporate unit-to-unit variability when modelling the noisy degradation traces of multiple nominally-identical units, and how allowing some parameters to vary determines how information is shared between observational units. I then demonstrated the fitting, evaluation, and comparison of these models on an experimental crack growth dataset with added measure-

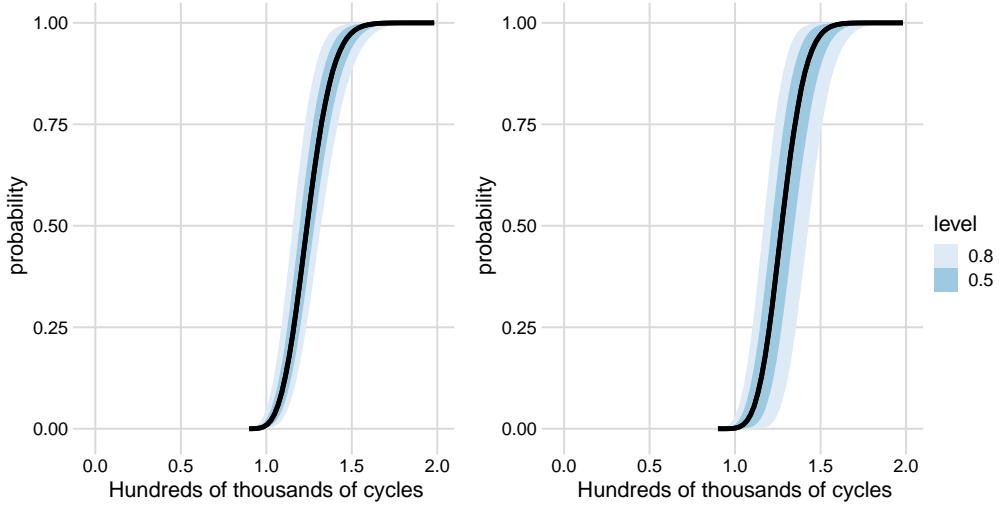


Figure 5.11: Posterior distributions of the predictive failure time distributions from the complete pooling model (left) and the varying μ model (right) for Unit 3.

ment error. Lastly, I showed how failure time distributions (with uncertainty bands) can be constructed for both new units and units under test using the posterior distributions of either a complete pooling or partial pooling gamma process model. This final section reviews the chapter’s main points and discusses the key findings, contributions, and areas for future work.

To model the ten units’ noisy degradation traces in Fig. 5.1 simultaneously, I explored several models: one where all of the degradation traces arise from the same underlying gamma process (complete pooling) and others where either μ , ν , or both are allowed to vary between units (partial pooling). Parametrising the gamma process in terms of μ and ν clarifies how unit-to-unit variability can be incorporated into the model and forces the analyst to be explicit in how he/she expect the units to vary: should the mean wear rate vary among units, or the volatility? In the same way, the new parameterisation also clarifies how we might model the effect of covariates on the degradation: for example, should environmental variables such as varying temperature or humidity affect the mean wear or the volatility? There are no covariates for the crack growth data, but extending the BHM to include covariate information would be interesting future work if data were available to do so. The fact that parameters μ and ν have clear

and separate effects also helped diagnose the sampling issues and interpret the posterior of the parameters and hyperparameters directly.

Based on the elppd criterion, the models where μ varies between units perform best for predicting out-of-sample. In contrast, models where μ is constant across observational units best predict the future observations of the units under test. Although, the differences in the elppd scores are relatively small. All of the models explored fit the data: the predictive distributions of each unit's underlying degradation path contains the true degradation path, and the marginal posterior of σ include the true value $\sigma = 0.025$. In the posterior distributions of the partial pooling models, there is evidence that μ should be allowed to vary between units since there is some variability in the modes of the marginal posteriors of the unit-specific μ_j . However, the spread of these distributions are wide enough to encompass the mean μ_μ —in addition, the posterior of σ_μ has mass near zero—therefore it could well be that all units share the same μ_i value, even under a model where we allow them to vary. In the models where ν varies, the marginal posterior distributions of the unit specific ν_j are almost identical, and the marginal posterior of σ_ν has considerable mass near zero, showing that there is little evidence that the coefficient of variation varies among the units.

Given the weak evidence in the hierarchical models' posteriors that μ varies from unit-to-unit, and even weaker evidence that ν varies, it is understandable that the complete pooling model performs at the same level as the partial pooling models for the crack growth data. The complete pooling model performs best in the case of $\text{elppd}_{\text{SA-CV}}$, possibly because assuming the simpler model structure results in the data more strongly informing the three parameters and hence more precisely estimating the volatility of the gamma process (ν): which is important for accurately forecasting future degradation. Because of the marginal difference between the four models, it is not surprising that in Leadbetter et al. (2024), where we used a slightly different prior for μ , the complete pooling model performs best with respect to both $\text{elppd}_{\text{LOOU-CV}}$ and $\text{elppd}_{\text{SA-CV}}$; showing that

the ordering is sensitive to the model specification and prior. Rodríguez-Picón et al. (2018) analyses the same data using gamma processes that incorporate unit-to-unit variability but without measurement error. They find that, when the data do not include measurement error, one of the partial-pooling models they explore outperforms the complete pooling model according to information criteria methods. However, there is also very little difference among the models they explore.

The struggle to clearly identify the best model could be a result of the nested structure of the models. When the observed variability of the degradation traces can easily be explained by the complete pooling model (which appears to be the case for the crack growth data), all of the models I have explored here contain this ‘true’ model. Future work exploring how well the elppd methods identify the true model from complete and partial pooling models when the data are generated from one or the other would be interesting. The presence of measurement error adds an additional *degree of freedom* to the models, making it even more difficult to clearly identify the best *underlying* model candidate. To this end, it would also be helpful if future work exploring elppd through simulation looked at the effect of sample size and noise level on how clearly the true model is identified. In an early work on unit-to-unit variability, Lawless and Crowder (2004) devise a statistical test to determine whether or not a random effect should be included in a degradation model when working in a non-Bayesian framework. Similar guidance for Bayesian models could be helpful.

The crack growth data that I have analysed do not show strong signs that there is variability among the units outside of the usual ‘jumpiness’ of a gamma process. For other data, identifying a suitable partial pooling model may be much more obvious. Nevertheless, I have shown how analysts can propose, fit, check, and then, finally, choose between suitable Bayesian model candidates using a fully Bayesian framework.

Chapter 6

Conveyor belt wear forecasting

In the past two chapters, I have looked at expanding the gamma stochastic process model to account for common situations encountered in practice, namely, the need to account for measurement error and the ability to borrow information across similar processes. In this chapter, I look at the practical example of modelling the degradation of a conveyor belts wearing surface using gamma processes. In doing so, I use both the noisy gamma process construction from chapter 4 and the partial pooling structures in chapter 5.

Conveyors are critical to the productivity of iron ore mines and other mining operations. As such, their unplanned failure can cause a significant loss of production and, subsequently, a substantial loss of profits. On the conveyor, one of the main components that can fail is the belt, and a major failure mode of the belt is wear (Bortnowski, Kawalec, Król, & Ozdoba, 2022). To manage the risk of failure due to wear, reliability engineers monitor the thickness of the belt's protective topcoat using ultrasonic thickness (UT) measurements. An example of this data is shown in *Figure 6.1*. Engineers then use this condition monitoring data to estimate the failure time of the belt and plan when to replace it. However, at each observation time, the UT data only provide a detailed description of the wear profile across the belt's width at one random location along its length. Furthermore, the observation times are sparse. These factors result in

lots of uncertainty around the underlying degradation of the belt. Therefore, estimating the failure time of the belt from the UT measurement data to inform maintenance decisions requires robust statistical modelling and the quantification of uncertainty. In this chapter, I show how the Bayesian hierarchical approach can be used to extend functional data analysis (FDA) of the wear profiles to different degradation models in order to forecast the belt’s wear and predict the remaining useful life. Particularly I compare a gamma stochastic process and a linear general path model for modelling the underlying degradation process of the belt.

Although there are papers that address the condition monitoring of conveyor belts, for example, identification of damage from puncture or modelling cord damage (Bortnowski et al., 2022), very few academic works focus on wear from abrasion. This is surprising considering that wear is a major failure mode of the belt (Bortnowski et al., 2022), especially for shorter, highly-utilised belts like stackers and reclaimers (which are also highly critical and difficult to maintain). Webb et al. (2020) demonstrate one typical method that an engineer would use to estimate when the failure of the belt will occur due to wear. For each measurement location along the belt’s width, the engineer fits a linear relationship to the UT measurement at that location using cumulative tonnes as the predictor variable. Next, at the location with the most aggressive wear rate—the steepest gradient—they extrapolate the line up to some predetermined soft failure threshold, indicated by a dashed line at 25mm of wear in *Figure 6.1*. The time at which the extrapolated line intersects the soft failure threshold is the predicted failure time. An alternative but similar approach is to take the maximum wear measurement from each profile and fit a linear relationship to these maximum wear measurements. This second approach is used by some conveyor condition monitoring software. *Figure 6.2* demonstrates these two methods using the data in *Figure 6.1*. Unfortunately, these method neglects the many sources of uncertainty in the data-generating and observation processes. Firstly, trend fitting of

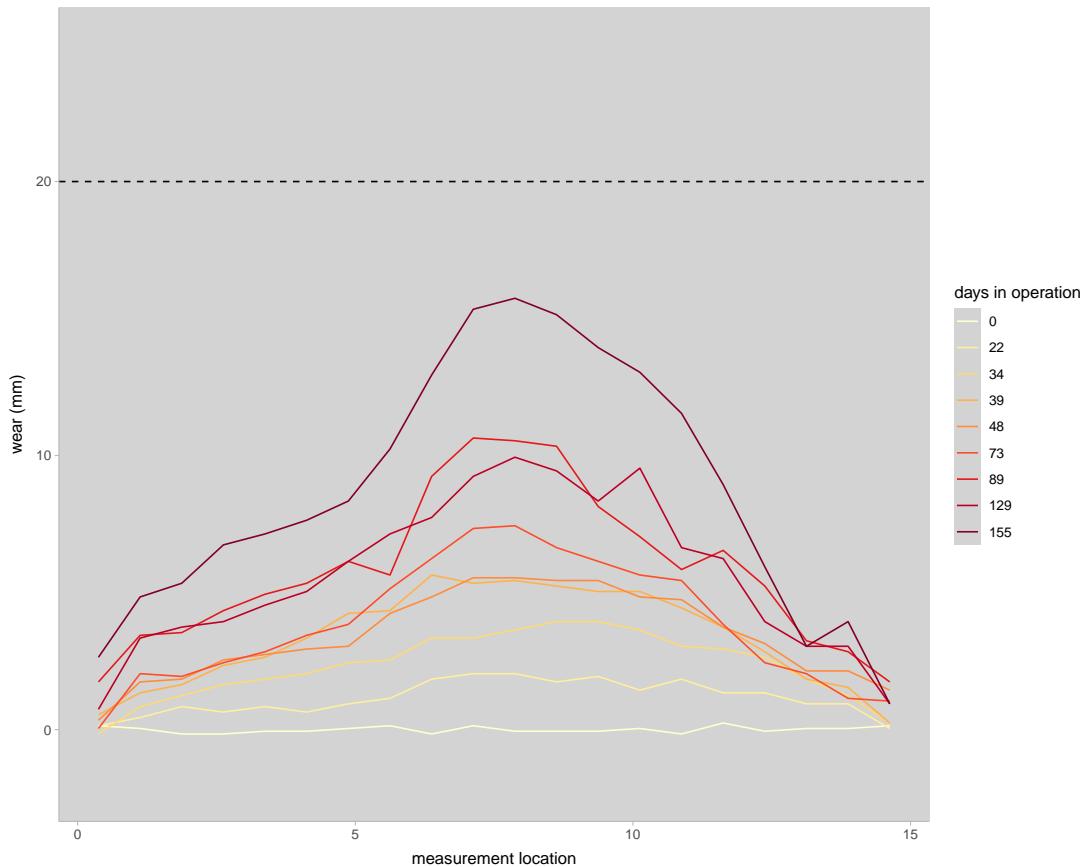


Figure 6.1: The growing wear profile of a conveyor's belt over time. Ultrasonic thickness measurements are taken at $N = 20$ measurement locations across the width of the belt at repeated times. The vertical axis is wear in mm , calculated by subtracting the measured thickness from the original thickness of the belt. The horizontal dashed line at $25mm$ of wear indicates the maximum allowable wear before the belt needs to be replaced, i.e. the soft failure threshold. The colour gradient indicates the time the belt has been in operation. Measurements taken at the beginning of the belt's life are shown in light yellow, and the most recent set of measurements are plotted in dark red

the raw UT measurements is sensitive to noise in the data, especially early on in the belt's life when there are few observations. Secondly, there is no formal structure for managing the different sources of uncertainty, for example, uncertainty in the UT measurements due to measurement error, uncertainty in the wear profile because of spatial variation along the length of the belt, uncertainty in the wear rate due to variations in operating conditions, and uncertainty in the parameters of the degradation process. Finally, the prediction is based solely on the forecast from only a few measurements on that particular belt with no uncertainty quantification. Therefore, the engineer cannot quickly assess the future wear of the entire belt surface, nor can they interpret risk, which limits their ability to justify and defend their maintenance decisions.

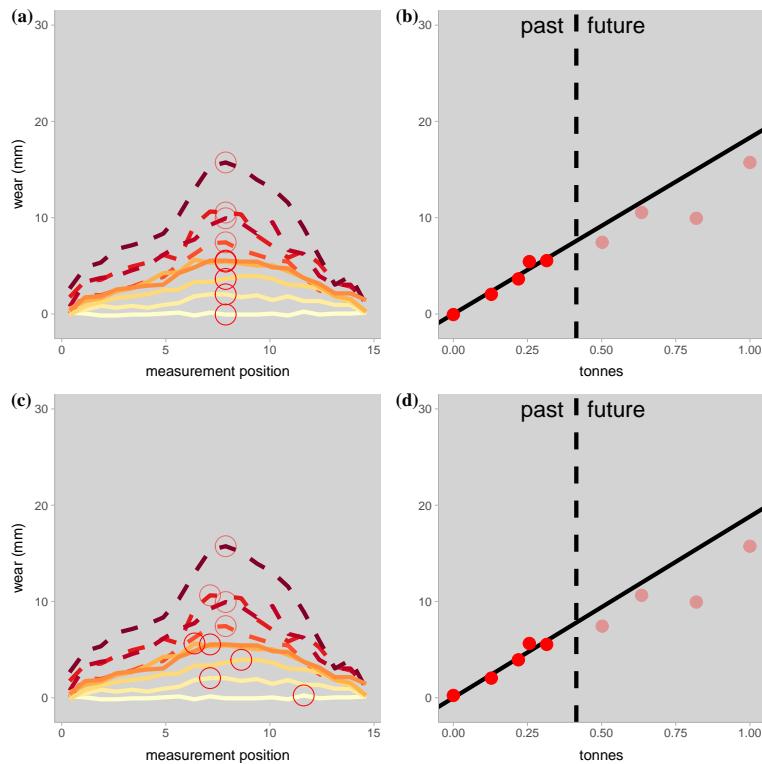


Figure 6.2: A demonstration of some typical approaches used to trend belt wear. (a) and (b) demonstrate the method of (Webb et al., 2020), where the measurements at the fastest wearing location which are circled in (a) are trended into the future in figure (b). (c) and (d) demonstrate an alternative but similar approach used by some reliability software, where the maximum wear measurement in each profile is trended instead.

Rather than forecasting the wear at a single measurement location or only for the maximum wear measurement, I propose a Bayesian hierarchical method for forecasting the evolution of the entire wear profile of the belt through time. The method I use expands degradation models, such as the previous work in chapters 4, and 5, to functional time series. In the first level of the hierarchical model, I apply a functional data analysis (FDA) approach to modelling the condition monitoring data in Figure 6.1. The FDA interpretation smooths the observations, which helps account for measurement error in the UT testing process; it also reduces the dimension of the data, making it easier to model the degradation at all the measurement locations simultaneously without too much of a computational burden. This functional data model, when paired with a suitable process model for the underlying degradation—either general path or stochastic process—and a suitable parameter model, formally manages the different sources of uncertainty. The proposed method produces forecasts of the belt’s wear that:

1. properly quantifies the uncertainties in the prediction, and
2. produces an intuitive forecast of the entire wear profile.

The method is not limited to conveyor belt wear; it can be used for any degrading surface monitored over a grid of locations. Here, I show an application for a one-dimensional profile, but the method could be expanded to two.

I begin in Section 6.1 by providing a brief overview of functional data analysis and showing how it can form the data model in a Bayesian hierarchical model for degradation. I then define two process models for the degradation of the belt in Section 6.2: a noisy gamma process and a linear general path. In Section 6.3, I define the prior distributions for the two models, show how an informative prior can be constructed from historic belt wear datasets, and check the plausibility of the process model through prior predictive simulation. Section 6.4 describes sampling and investigates the posterior draws from each model in terms of the marginal posterior distributions of the parameters, the posterior distribution of

the intermediate quantities in each model that describe the underlying wear process, and the prior predictive distributions for replications of the data in Fig. 6.1. In Section 6.5, I describe how to generate forecasts for the belt’s wear profile using the two different models and demonstrate with the ninth wear profile observation that I withhold when fitting the two models. I then evaluate and compare the two models based on their ability to predict the degradation of the belt at future time points in Section 6.6. Lastly, in Section 6.7, I demonstrate how to construct failure time distributions for the two models conditioned on the current condition of the belt. I finish in Section 6.8 by revisiting the main results and pointing out areas of useful future work.

6.1 Functional data analysis (the data model)

In functional data analysis, the data in each observation are considered as coming from a smooth underlying random function rather than being a scalar or vector-valued random variable (Gelman, Carlin, et al., 2020, p. 512). That is, we interpret the set of UT measurements at each observation time as noisy observations at discrete locations of some smooth underlying random function

$$z_{i,n}|f_i(n), \sigma \sim N(f_i(n), \sigma). \quad (6.1)$$

In this way, the standard deviation σ describes the noise in the UT measurement process—which we assume to be normally distributed—and the function $f_i(n)$ describes a smooth wear profile across the width of the belt at time t_i . The aim of a functional data analysis is, therefore, to model the collection of functions $\{f_i(\cdot)\}_{i=1}^I$. To do so requires the analyst to choose a functional form for $f_i(\cdot)$, a decision on which the model is implicitly conditioned. In this analysis, I use a B-spline to model the wear profiles of the conveyor belt.

B-splines B-splines are piecewise continuous functions (Ramsay, Hooker, & Graves, 2009, p. 33-38) that are constructed as the weighted sum of a set of M locally defined polynomial B-spline basis functions,

$$f_i(n) = \sum_{m=1}^M y_{i,m} b_m(n). \quad (6.2)$$

Here, $y_{i,m}$ is the weight of the m^{th} basis function, $b_m(\cdot)$, for the i^{th} observation. The number of basis functions and their shapes and locations are defined by a set of knots and the order of the basis functions. To describe the wear profiles, I use eight evenly spaced knots and third-order basis functions. However, I also drop the outer two sets of basis functions to constrain how flexible the spline can be towards the edges of the belt and to ensure that the wear profile is fixed at zero at the boundaries. These choices were made by measuring the goodness of fit for many different conveyor wear profiles from the industry partner's condition monitoring system and weighing up simplicity and flexibility. Figure 6.3 (c) shows the set of un-weighted basis functions on which we condition our model. The spline is fit to the UT data by estimating the weights each basis function. As an example, Fig. 6.3 (a) shows the fitted spline for the fifth observation and Fig. 6.3 (b) shows the weighted set of basis functions that make up the wear profile. Fitting a B-spline to each set of UT measurements using the set of basis functions in Fig. 6.3 (c) yields the set of spline coefficients $\{y_{i,m}\}_{m=1}^M$ that fully describes the wear profile at time t_i . The next level of the BHM, the process model, models how these spline coefficients evolve through time.

Functional time series Before moving on, I note that modelling the evolution of spline coefficients through time is not new. Functional time series analysis is an area of statistics that models how time-ordered functional observations evolve (Hormann & Kokoszka, 2012). It has also been done in a Bayesian context (Kowal, 2017). Usually, these methods use functional PCA (Ramsay et al., 2009, p. 16) and, for the most part, autoregressive processes to model the evolution

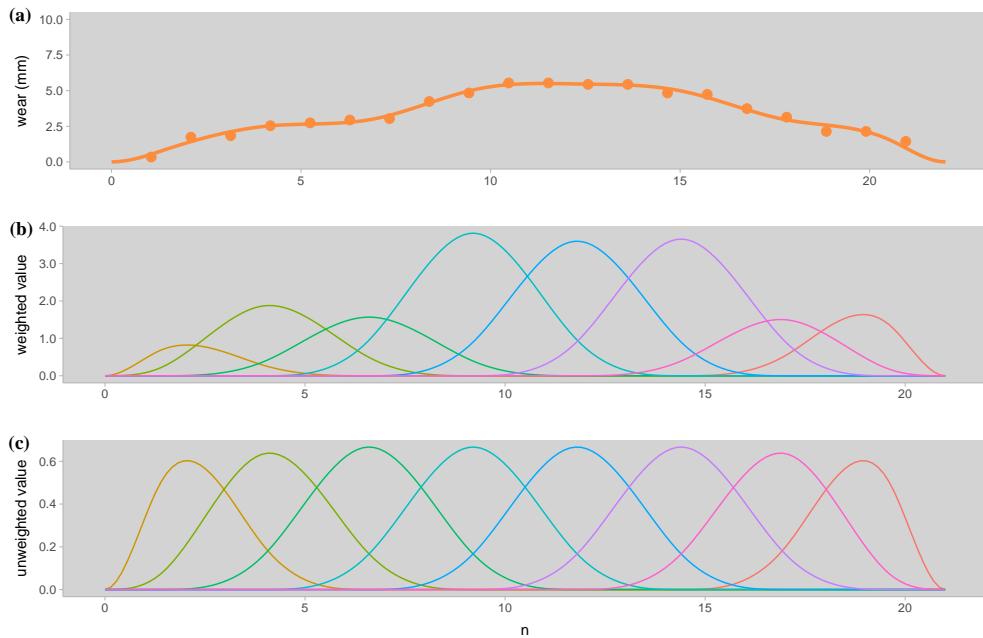


Figure 6.3: (a) shows a fitted wear profile for the fifth observation; (b) shows the weighted set of basis functions that make up the profile in (a); and (c) shows the unweighted set of basis functions

of the eigen weights (the coefficients of the eigenfunctions). Using functional PCA first reduces the number of coefficients to model. Furthermore, because the eigenfunctions are orthonormal, the eigen weights can be modelled independently. This dimension reduction is helpful for cases where many basis functions are needed to fit the data because, in these cases, modelling the basis coefficients directly would require a very large covariance matrix. However, in our case, the number of functional observations of the belt is too small to perform PCA reliably, and the degradation process is not stationary. Rather, we use a degradation model to model the spline coefficients directly and, for the moment, assume that the coefficients are independent of one another. In this way, our approach may be more down the path of the spatio-temporal models of Wikle et al. (2019, p. 218-224).

6.2 Process models

As potential process models for the underlying degradation process of the belt, I explore both a noisy gamma process and a linear general path model. In the noisy gamma process model, I model the spline coefficients with effectively the same model that I presented in chapters 4, and 5, except with some minor adjustments to make the model more robust to outlying observations. In the linear path model, I use a much simpler linear model for the spline coefficients. The major difference between the two process models is that the noisy stochastic process model explicitly breaks down the variation in the degradation signal into variation attributed to changes in degradation rate through time—which could arise from changes in operation or environmental conditions—and the uncertainty of the measurement process—which in this case arises because the cross-section is not always measured in the same location. In contrast, the general path model assumes that the degradation path is deterministic conditional on the model’s parameters and the time t and lumps the variability from both the variation in signal through time and measurement error into a single error term. When using a general path model, additional structure in the noisy degradation measurements can be accounted for by adjusting the error distribution, for example, adding autoregressive error terms. In some cases, the two approaches are directly comparable; for example, Whitmore (1995) formulates a Weiner degradation process with measurement error as a multiple linear regression with added covariance structure. Some discussions of stochastic process and general path models can be found in Ye and Xie (2015). The benefit of using the noisy gamma process, in this case, is that by modelling these two sources of variation separately, we can assign them different distributions (Gaussian-distributed measurement error and gamma-distributed jumps), which explicitly splits up the two sources of uncertainty. However, there may be a limit of measurement error and/or volatility of the degradation process where it is simpler and better to use the general path model even though it is less physically motivated. Below, I outline the two process

models.

6.2.1 Noisy gamma process

In the gamma process version of the process model, I model each spline coefficient as coming from a t distribution with 10 degrees of freedom where the location depends on the estimated ‘average’ value of the spline coefficient along the length of the belt at that time and the scale depends on the estimated average value and scale parameter ϕ . Assuming a t distribution for the spline coefficients makes the model more robust to outlying observations (Gelman, Carlin, et al., 2020, Chapter 17). I then model the progression of the ‘average’ degradation of the spline coefficient through the gamma process parametrised in terms of the mean and coefficient of variation

$$\begin{aligned} y_{m,i}|y_{m,i}^*, \phi &\sim t_{10}(y_{m,i}^*, \phi y_{m,i}^*) \\ \Delta y_{m,i}^* &= y_{m,i}^* - y_{m,i-1}^* \\ \Delta y_{m,i}^* | \nu_m, \mu_m &\sim \text{Ga}\left(\frac{\Delta t_i}{\nu_m^2}, \frac{1}{\mu_m \nu_m^2}\right). \end{aligned}$$

In this process model, each of the m spline coefficients has specific μ_m and ν_m .

6.2.2 Linear model

In the linear general path version of the process model, I use the same t_{10} structure to model the noisy spline coefficients conditional on their estimated average value and ϕ , but now model the progression of the average value of each spline coefficient as a deterministic function of t that is specified in terms of the average wear rate μ_m

$$\begin{aligned} y_{m,i}|y_{m,i}^*, \phi &\sim t_{10}(y_{m,i}^*, \phi y_{m,i}^*) \\ y_{m,i}^* &= \mu_m t_i. \end{aligned}$$

6.3 Parameter model

Because I have used the mean/coefficient of variation parameterisation of the gamma process and a conditional structure for both the models, they share many of the same parameters— σ , ϕ , and the μ_m . Hence, I use the same priors in both models. For the variance parameters σ and ϕ I use the priors

$$\begin{aligned}\sigma &\sim U(0, 100) \\ \phi &\sim \text{Cauchy}^+(0, 25)\end{aligned}$$

based on the same justifications presented in Section 4.4 (Gelman, Carlin, et al., 2020, chap. 17). For the mean wear rate of the different spline coefficients, I now use the prior

$$\mu_m \sim N(\hat{a}, \hat{b})$$

where \hat{a} and \hat{b} are estimated from historic data. I go into more detail on how this is done below, but first, define the remaining priors for the gamma process.

The gamma process model has slightly more parameters than the simpler linear path model. For the coefficient of variation parameters ν_m , I use the hierarchical prior

$$\begin{aligned}\nu_m &\sim N(\mu_\nu, \sigma_\nu) \\ \mu_\nu &\sim t_3(0, 0.5) \\ \sigma_\nu &\sim \text{Cauchy}^+(0, 0.25)\end{aligned}$$

defined in the same way as the unit-to-unit variability model in Section 5.1.2. The hierarchical prior partially pools information between the ν_m of the gamma processes for each spline coefficient (because they all belong to the same belt we expect them to be similar) while still allowing them to vary.

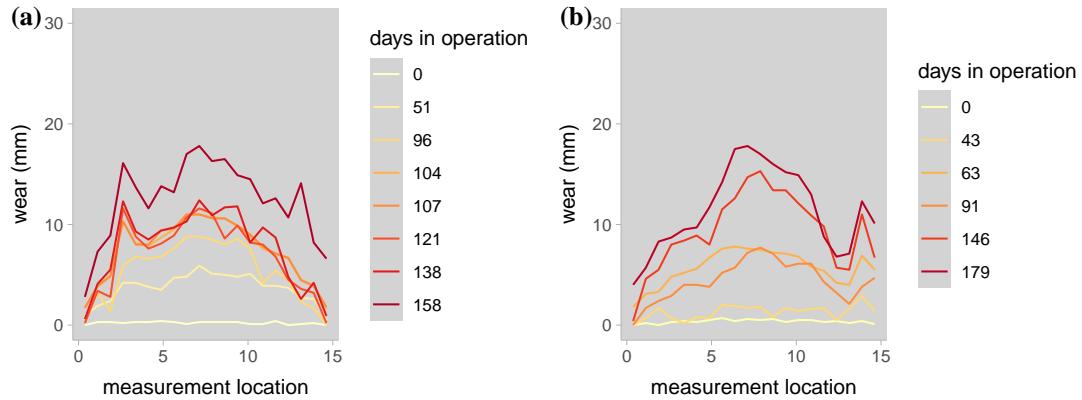


Figure 6.4: Two historic belt wear data sets from the same conveyor

Choosing informative hyperparameters for μ There are also historic belt wear datasets for the particular conveyor that the data in Fig. 6.1 come from. Two of these historical datasets are shown in Figure 6.4. According to the industry partner from whom we got the data, the future wear behaviour of the belt is expected to be different from the historic wear datasets because of variations in ore composition, belt manufacturer, and operational strategies. However, the historic wear behaviour of the belt should indicate future behaviour. Although the historical data cannot be directly used in the analysis, in the Bayesian framework, the historical data can be used to formulate an informative prior and, therefore, still supplement the analysis.

I encode the prior information in the model through the parameter μ . To do so, I fit B-splines to each of the historic wear profiles using the same set of basis functions in Fig. 6.3 (c) to calculate the spline coefficients and then estimate the average wear of each coefficient. Figure 6.5 shows the linear regressions for each spline coefficient. Based on the estimates, I set \hat{a}_m in the prior for μ_m as the estimated slope of the coefficient based on the historical data and \hat{b}_m as five times the standard error of the estimate.

Prior predictive checking To check the plausibility of the proposed priors in the context of the two process models, I perform prior predictive checking.

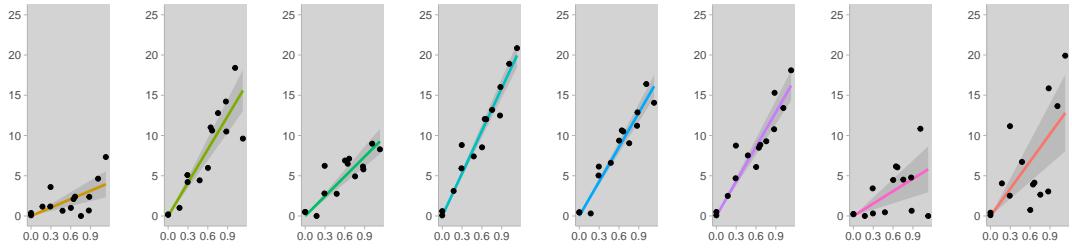


Figure 6.5: The estimated average wear rate for each spline coefficients calculated from the datasets in Fig. 6.4. The colour of each fit corresponds to the basis functions in Fig. 6.3 (c). The horizontal scale of each plots is in tonnes and the vertical scale is the value of the spline coefficient.

Figure 6.6 shows sixteen prior predictive simulations from the gamma process (solid lines) and linear model (dashed lines). Because I have used very vague priors for the parameters σ and ϕ , simulating the noisy wear profiles and UT measurements will not make sense until after conditioning on the observed data. Therefore, the simulations shown in Fig. 6.6 are for the belt's non-noisy (average) wear profile.

To generate each fictitious data set, I sample a value of each μ_m , μ_ν , and σ_ν from their prior. I then sample the values of ν_m from the hierarchical prior using the realisation of μ_ν and σ_ν . Next, I generate values of the smoothed spline coefficients for each observation time t_i from each of the two process models and apply the coefficients to the set of basis functions to calculate the smooth wear profiles. For the linear general path model, this means simply multiplying each of the μ_m by each t_i to get the values of the spline coefficients. Whereas, for the gamma process, I simulate the sets of jumps in degradation using the distributions $\text{Ga}(\Delta t_i / \nu_m^2, 1 / (\mu_m \nu_m^2))$ and take the cumulative sum to calculate the values of the filtered spline coefficients.

The fictitious data resulting from each model, for the most part, look somewhat realistic in scale and shape. Some of the gamma process simulations look unrealistic—there are some where the wear jumps 60mm between observation times and others where the belt does not wear at all—but many look sensible. The shapes of most of the linear simulations look reasonable, but their growth

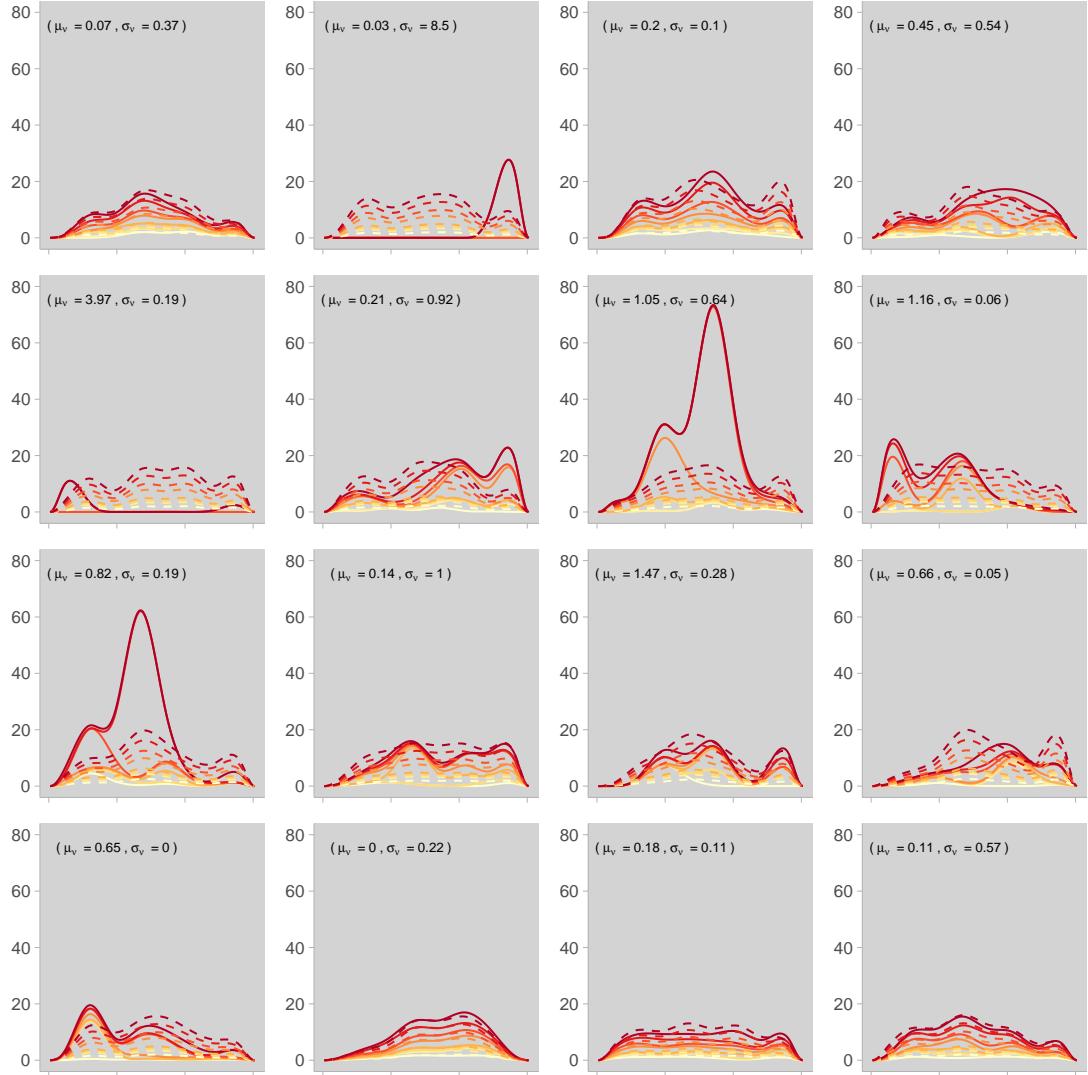


Figure 6.6: Sixteen prior predictive simulations of the average wear profile generated from the linear path (dashed line) and gamma process (solid line) process models with the priors defined in Sec. 6.3. The values of the hyper parameters for ν in the gamma process that were used to generate each simulation are displayed on each plot.

looks synthetic and not natural enough—this should change when I add in the noise layer. In each plot, I have used the same set of realisations of the μ_m to simulate from both the gamma process and linear general path model. The realised values of the hyperparameters of the hierarchical prior for the ν_m used in the gamma process models are displayed in each subplot. Notice that when μ_ν and $\sigma_\nu \rightarrow 0$ the profiles from the gamma process match those from the linear model very closely and when the values of the hyperparameters allow for large values of the ν_m —i.e. when either μ_ν , σ_ν , or both are large—the simulated data from the gamma process model look unrealistic. However, in these simulations, the gamma process model looks to be able to generate the most natural-looking datasets. One thing of note is that many of the simulations, particularly for the gamma process, appear unrealistically wiggly. This ‘wiggliness’ is due to a lack of large-scale spatial structure in the model and is a feature of the postulated model, not the prior. Incorporating large-scale spatial structure could be addressed in future works, but I do not address it here since this behaviour is ‘smoothed out’ when all the possible curves are interpreted as a distribution. All in all, the simulations from both models are in the realms of plausibility, which is what is desired of a weakly informative prior (Gabry et al., 2019b).

6.4 Posterior sampling and inference

Conditioning on the first eight observations, I generate samples from the posterior distributions of each model using the No-U-Turn sampler implemented in *Stan* (Stan Development Team, 2022). In total, I generated 12000 samples from each posterior, using four chains that are 4000 iterations in length with no thinning and a burn-in of 1000 iterations. For the case of the gamma process, sampling results in a small number of divergent transitions—roughly 0.9%—which appear to arise in the hierarchical prior for the same reasons as previously discussed in Section 5.2; however, because of the very limited number of divergent transitions, I do not try and rectify the issue. The linear general path model, on the other

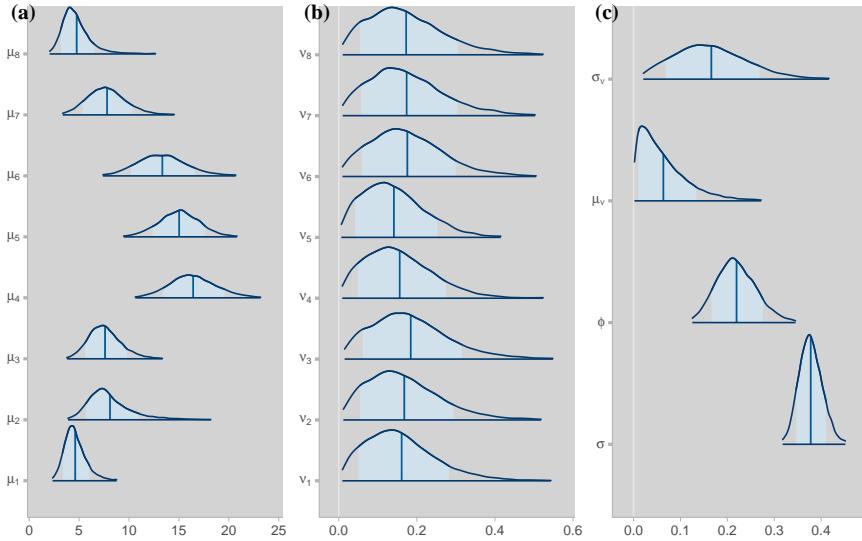


Figure 6.7: Marginal posterior distributions of the gamma process model parameters.

hand, fits very efficiently without any divergent transitions, and both models show good \hat{R} statistics and effective sample sizes for the parameters. In this section, I visually analyse and compare the posterior samples from each model through the marginal posterior distributions of the parameters, the posterior distributions of the intermediate quantities in the models, and the posterior predictive distributions of the UT measurements and underlying wear profiles.

Marginal posteriors of parameters The marginal posteriors of the parameters show our updated belief after conditioning on the data. Essentially, compared with the prior, they show what we have learnt from the data. Additionally, the posterior draws of the parameters are important because we rely on them when forecasting the underlying process into the future in Section 6.5. Figure 6.7 and 6.8 show the marginal posterior distributions of the parameters in the gamma process and linear general path model, respectively.

In the marginal distributions of the parameters of the Gamma process model, shown in Fig. 6.7, we can see three main things. Firstly, by looking at the distribution of the different μ_m , Fig. 6.7 (a), we can see that the model has

successfully captured the belt’s general ‘dishing out’ behaviour since the mean wear rates of the spline coefficients closer to the centre of the belt are higher. Secondly, the general wear of the belt appears to be lop-sided since there is a lack of symmetry in the μ_m —the posterior median of μ_6 is much greater than that of μ_3 . These first two points may be obvious to the reader when looking at the raw data in Fig. 6.1, but it is important that the model has identified the general behaviour as chronic and not just noise, especially when our main goal is to forecast the wear profile through time. The third main observation is how little variability there is amongst the ν_m in Fig. 6.7 (b). Their expected values are very similar, and the marginal posterior of σ_ν , shown in Fig. 6.7 (c), has a lot of mass near zero. However, it is hard to say whether or not the variability is negligible.

In addition to these three main points, in Fig. 6.7 (c), it looks as though the marginal posteriors of the two variance parameters σ and ϕ and the two hyperparameters of ν encode similar levels of uncertainty into the model. However, this is not the case, even though the marginal posteriors have similar scales. σ can be interpreted directly in mm , but the effect of ϕ is scaled by the values of the filtered spline coefficient and should rather be interpreted as a proportion of y^* . The influence of the posterior distributions of σ and ϕ is much more obvious in the posterior predictive distributions. On the other hand, the uncertainty in the degradation path that results from the underlying gamma process and the posterior distribution of its parameters and hyperparameters— μ_ν , σ_ν , ν_m , and μ_m —is almost impossible to interpret from the marginal distributions in Fig. 6.7. We gain a better intuition of the uncertainty in the underlying gamma degradation process by looking at the marginal posterior of the filtered spline coefficients $y_{i,m}^*$. I look at the posterior distributions of these intermediate quantities and the posterior predictive distributions next, but first, I compare the marginal posteriors from the fitted general path model to the marginal distributions from the gamma process that I have looked at here.

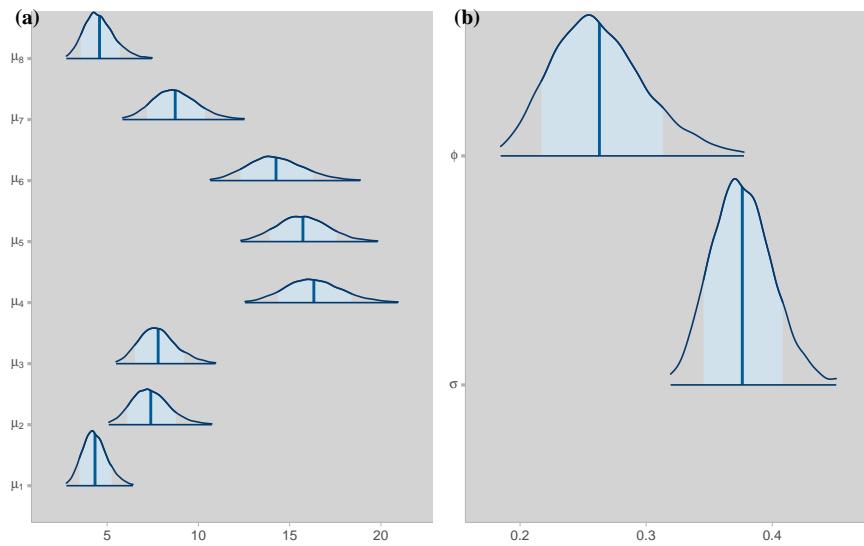


Figure 6.8: Marginal posterior distributions of the linear general path model parameters.

Figure 6.8 shows the marginal posteriors of parameters in the general path model. The marginal posterior distributions of the μ_m from the general path model in Fig. 6.8 (a) show the same general dishing out behaviour of the belt and asymmetric wear pattern. In fact, comparing Fig. 6.8 (a) with Fig. 6.7 (a), it appears that the posterior distribution of the μ_m are almost identical in the two models. The only observable difference is that the marginal distributions have heavier upper tails in the posterior of the gamma process model. In the marginal posteriors of the parameters σ and ϕ in Fig. 6.8 (b), the estimated value of sigma and corresponding uncertainty is also very similar to the gamma process model; however, the estimated value of ϕ is clearly higher. As expected, the scale of the t_{10} distribution in the parameter model has inflated in the case of the general path model to account for the variability in the degradation rate that would otherwise be accounted for by the jumpy gamma process. The distinction between the two process models is clearest in the posteriors of the intermediate quantities in the two models.

Intermediate quantities The $y_{i,m}^*$ are the filtered degradation paths of each spline coefficient, which essentially describe the ‘average’ wear along the length of the belt at each time. The intermediate quantities in the model are treated similarly to parameters in the Bayesian framework; as such, we also obtain posterior draws of the $y_{i,m}^*$ during MCMC sampling. Figure 6.9 shows the mean posterior value of the filtered degradation of each coefficient—plotted as coloured lines. Fig. 6.9 (a) shows the posterior of the gamma process model, and Fig. 6.9 (b) shows the posterior of the general path model. The colours of the mean path correspond to the basis functions in Fig. 6.3 (c). One hundred individual draws from each of the joint posteriors of the y^* are also plotted in Fig. 6.9 as dark grey lines.

The draws of the spline coefficient from the gamma process mode in Fig. 6.9 (a) are noticeably ‘jumpy’ while the mean paths of the filtered spline coefficients look reasonably linear. For the general path model in Fig. 6.9 (b), on the other hand, all of the posterior draws are perfectly straight lines. Despite this difference, the spread of the draws of the pathways in each equivalent subplot in Fig. 6.9 (a) and (b) are very similar. The fact that most of the draws from the gamma process appear as jumpy processes—even though we specified a prior which favours straight processes—seems to suggest that the wear rate varies through time, and as such, we should allow the model to do so too. The general path model has accounted for this added variability in the data by inflating the value of ϕ , as we previously noted. This difference is why some argue that the gamma process is a more realistic and physically motivated model (Ye & Xie, 2015). However, the simplification of reality that linear models apply is notoriously successful. To understand if the simplified linear model is flexible enough to sufficiently describe the data and if the data contain enough information and a clear enough signal to identify the gamma process with its extra parameters, I look at the posterior predictive distributions for the two different models and compare them with the observed data.

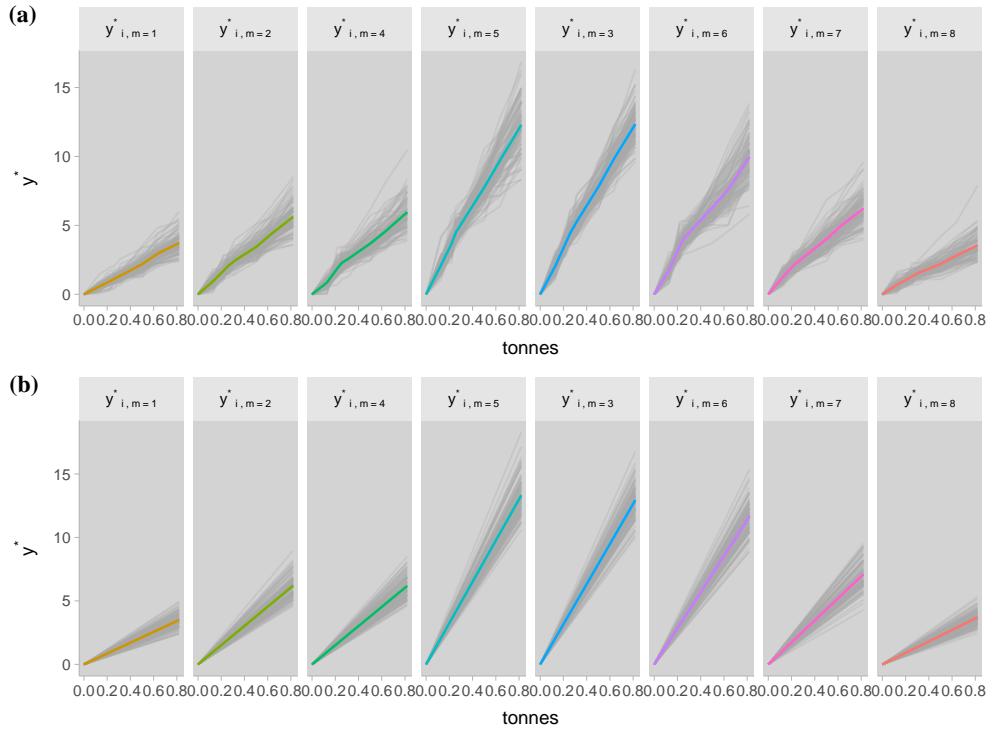


Figure 6.9: The posterior draws of the filtered spline coefficients for (a) the gamma process model and (b) the linear general path model.

Posterior predictive distributions As discussed in section 1.4, a method for checking the fit of Bayesian models is to check if the data look plausible under posterior predictive distribution. In hierarchical models, this comparison can be performed at different levels of the model to check the fit of the data model and process models. Here, I generate a compare the posterior predictive distributions of replications of the UT measurements at the same times and locations along the belt as the data in Fig. 6.1 as well as the predictive distribution of replications of the wear profile at different locations along the belt's length at the same observation times.

To generate posterior predictive distributions for replications of the UT measurements, \tilde{z} , I sample from the data model conditional on the draws of the spline coefficients and σ

$$\tilde{z}_{i,n} | \underline{y}_i^s, \sigma^s \sim N(f_i^s(n), \sigma^s). \quad (6.3)$$

The posterior predictive distribution is generated in the same way for both models. The joint posterior predictive distribution of \tilde{z} from the gamma process model is shown in Fig. 6.10 (a), and from the general path model in Fig. 6.10 (c). The observed data are also plotted in both sub-plots for comparison. At the noisy UT observation levels, the prior predictive distributions of \tilde{z} for both models are very similar and fit the observed data well. In both cases, all of the observed UT data sit very close to the median wear profiles (showing that the models are flexible enough to fit the data) and sit neatly within the 95% uncertainty intervals.

To generate the posterior predictive distribution for replications of the measured wear profiles at each observation time, I sample new values of the noisy spline coefficients from the first level of the process model conditioned on the posterior draws of the filtered (mean) values of the spline coefficients and ϕ ,

$$\tilde{y}_{i,m} | y_{i,m}^{*s}, \phi^s \sim t_{10}(y_{i,m}^{*s}, \phi^s y_{i,m}^{*s}), \quad (6.4)$$

and then calculate the values of the spline functions $f_i(n) = \sum_{m=1}^M b_m(n) \tilde{y}_{i,m}$. This process is also the same for both models. The posterior predictive distribution of the wear profiles at each observation time generated from the posterior of the gamma process is shown in Fig. 6.10 (b); Fig. 6.10 (d) shows their posterior predictive distribution generated from the general path model. In Fig. 6.10 (b) and (d), I plot each observation time separately so that the 95% uncertainty intervals are clear. In all sub-plots, the observed data are included for comparison.

In the case of the posterior predictive distribution of \tilde{y} , the average wear profiles of the belt at each time are now clearly monotonic increasing and the uncertainty around the average wear profile also grows with time. In both figures 6.10 (b) and (d), the observed noisy UT wear profiles sit inside the uncertainty intervals, showing that the observed data are also plausible under both Bayesian hierarchical models at the process model level, but comparing Fig. 6.10 (b) and (d), the linear model appears to predict a greater wear rate and larger uncertainty than the gamma process model.

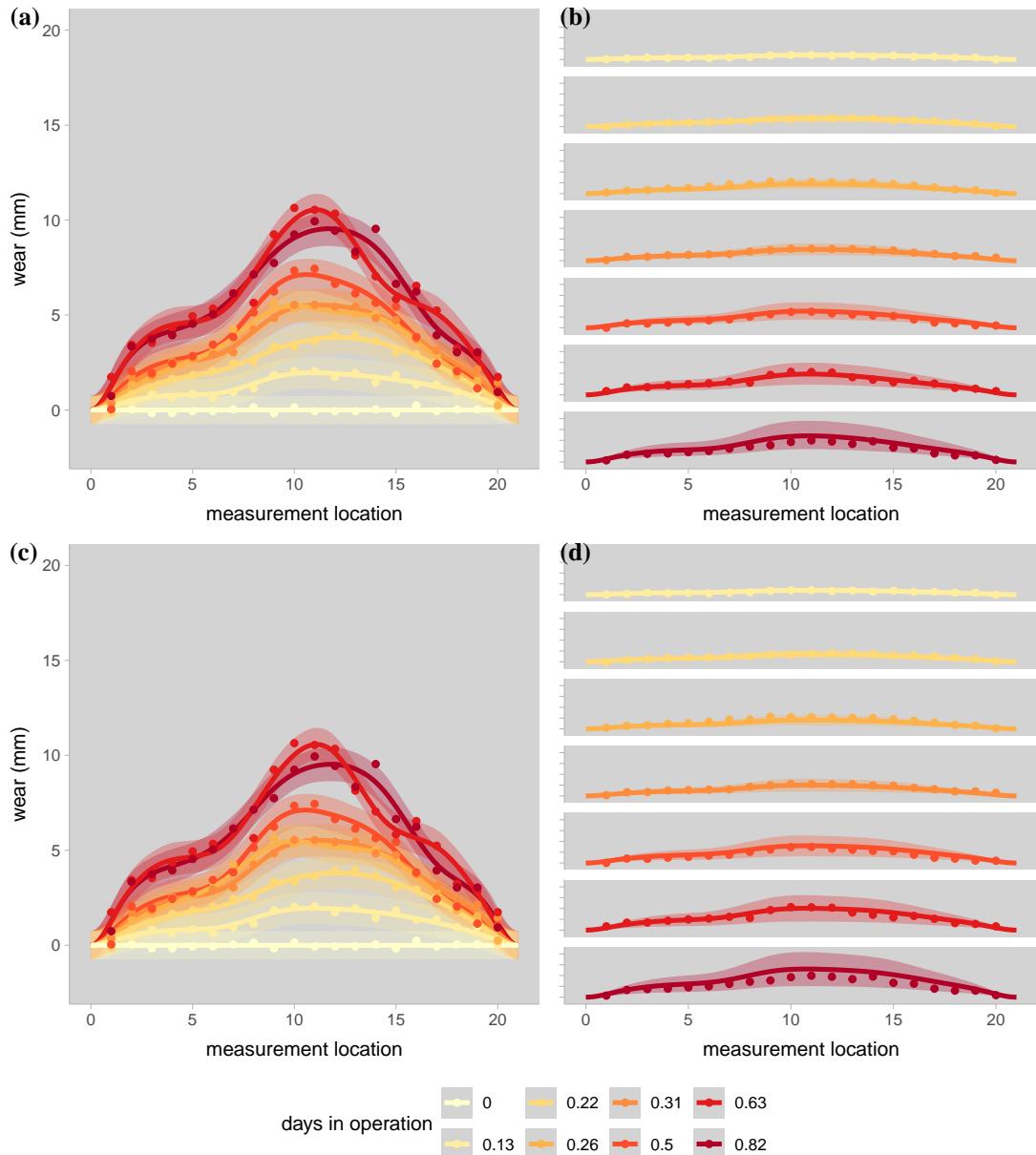


Figure 6.10: Four posterior predictive distributions, two form each model. (a) and (c) shows the predicted smooth functional observation of the wear profiles underlying the sets of UT measurements from the gamma process and general path model, respectively. The UT measurements are also plotted as points as a reference. (b) and (d) shows the posterior predictive distribution of a new observation of the wear profile along the length of the belt at each observation time for the gamma process and general path model, respectively. In (b) and (d), the observed wear profiles are also plotted for comparison.

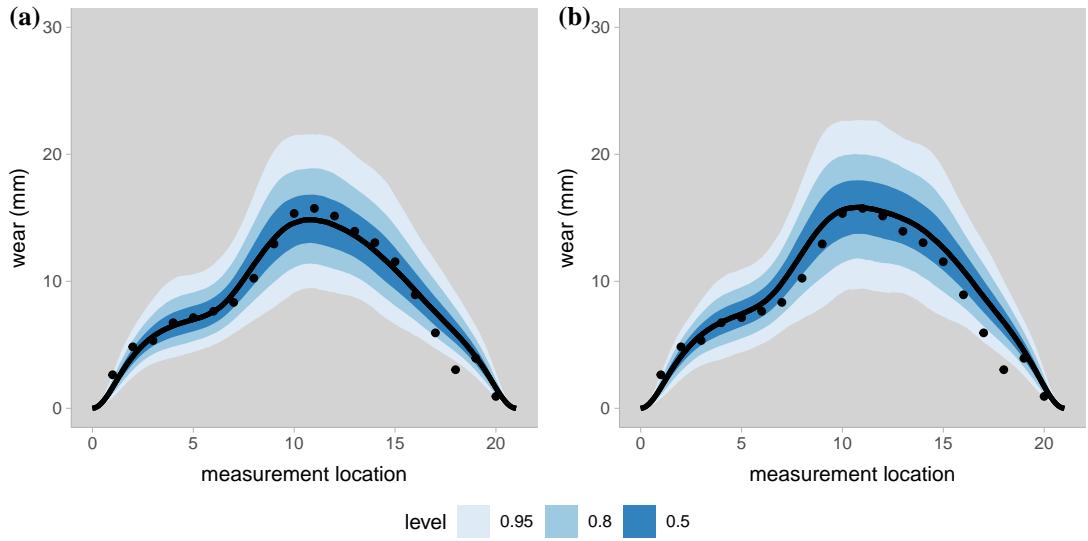


Figure 6.11: Forecasts of the ninth, withheld, observation from (a) the gamma process model and (b) the linear general path model. The true, observed, measurements are plotted for reference.

6.5 Forecasting degradation curves

Using the posterior draws from the two posteriors, we can forecast the degradation process of the spline coefficients using the process model and predict the belt’s wear profile at any time, t_{I+1} , in the future, conditioned on the belts current state of degradation. When producing the forecast, we should do so along the entire belt’s length since we want to predict when any part of the belt will exceed the soft failure threshold. In other words, we want to predict the distribution of the noisy spline coefficients at the forecast time t_{I+1} . The forecasts to the time $t_9 = 1$ of the ninth (withheld) belt wear observation are shown in Fig. 6.11 for the gamma process model (a) and general path model (b) (approximately 0.18 tonnes or 3.7 weeks into the future). In each sub-plot, the median forecast is shown as a solid black curve, and the 0.50, 0.80, and 0.95 uncertainty intervals are shown in various shades of blue. The observed UT measurements at t_9 —withheld from the model—are also plotted in each figure for comparison.

The process of generating the two forecasts in Fig. 6.11 are quite different. For the gamma process model, I first sample jumps of the spline coefficients from

the gamma distribution

$$\Delta y_{I+1,m}^* | \nu_m, \nu_m, z \sim Ga \left(\frac{t_{I+1} - t_I}{\nu_m^2}, \frac{1}{\mu_m, \nu_m^2} \right) \quad (6.5)$$

using the posterior draws of the gamma process parameters. I then add these jumps to the filtered values of the spline coefficients at time t_I , $\{y_{I,m}^*\}_{m=1}^M$, to predict the values of the filtered spline coefficients at time t_{I+1} . Lastly, I average over the spatial variability in the wear profiles along the length of the belt by sampling values from

$$y_{I+1,m} | y_{I+1,m}^*, \phi, z \sim t_{10}(y_{I+1,m}^*, \phi y_{I+1,m}^*). \quad (6.6)$$

For the general path model, I simply calculate the values of the filtered splice coefficients using the deterministic wear function $y_{I+1,m}^* = \mu_m \times t_{I+1}$ and then average over the spatial variability in the wear profiles along the length of the belt in the same way as I did for the last step of the gamma process forecast.

Despite the difference in how the two forecasts are produced, they both appear to reasonably predict the observed data at $t_9 = 1$. For both distributions, the observed data sit close to the median and comfortably within the uncertainty intervals. However, like for the posterior predictive distributions, the linear general path results in a slightly higher prediction of the wear and slightly wider uncertainty intervals. In the next section, I compare both models based on their predictive performance in terms of the whole wear profile and in terms of the maximum wear observation.

6.6 Comparison of methods

To distinguish between the two methods, I compare them based on both their ability to predict the whole wear profile at N steps ahead using the expected log scores and also visually on their ability to predict the maximum wear measure-

ment using a resampling technique similar to bootstrapping and cross validation. In the latter, I also compare the predictions of the method described in Webb et al. (2020). The first of these comparisons is mechanistically the same as elppd_{cv} , except that in this case it is unclear what predictive density I am trying to approximate and I reuse the withheld wear profiles as test sets more than once. Therefore, I refer to the method as the expected log score rather than elppd_{cv} , but for the purpose of model comparisons, the method is the same as that described in Sec 1.4.

Expected log score First I compare the two models based on the expected log probability of the spline coefficients of a withheld observation under the forecasted distribution. I fit the models to a portion of the data (i.e. observations 1:5, 1:6, 1:7, and 1:8) and then evaluate the expected log score of the withheld future wear profiles. To calculate the log score, I estimate the underlying spline coefficients of the withheld observation using maximum likelihood, assuming that $\sigma = 0.38$ (the median estimate from both models). I then calculate the expected log score of the set of spline coefficients under the distribution $t_{10}(\tilde{y}_{m,I}, \phi \tilde{y}_{m,I})$, using the eq. 1.12, as I would to calculate $\text{elppd}_{\text{LOO-cv}}$. The expected log scores of the N-step-ahead predictions are presented in table 6.1. The summations of the expected log score (similar to an elppd_{cv} measure) are presented in the final row of table 6.1.

In the expected log scores, the linear general path model outperforms the gamma process model in every scenario, even for the case of the forecasts shown in the previous section (Fig. 6.11). So, it appears that the linear general path model is a better model for predicting the overall wear profile at future times. However, since the main purpose for modelling the belt wear data is to predict the soft failure of the belt, reliability practitioners may be willing to forgo some accuracy in the prediction around the edges of the belt—where the wear rate is slow—for a more accurate prediction of the maximum wear.

Table 6.1: The expected log score (ELS) for each model when fitting to a portion of the data and predicting n-steps ahead. I is the maximum observation that the model was fit to and $I+1$ is the withheld observation that the forecast is generated for. The summation of the elppd scores are displayed at the bottom of the table.

| I | $I + 1$ | $ELS_{gamma\ process}$ | $ELS_{linear\ path}$ |
|-----------------|---------|------------------------|----------------------|
| 5 | 6 | -15.304 | -14.543 |
| 5 | 7 | -14.275 | -13.268 |
| 5 | 8 | -16.124 | -15.156 |
| 5 | 9 | -14.369 | -13.724 |
| 6 | 7 | -12.919 | -12.207 |
| 6 | 8 | -13.172 | -13.142 |
| 6 | 9 | -12.177 | -11.928 |
| 7 | 8 | -13.412 | -13.287 |
| 7 | 9 | -11.767 | -11.480 |
| 8 | 9 | -11.084 | -10.582 |
| -134.603 | | -129.317 | |

Test quantity An additional way of scrutinising the predictive performance of the two models is to choose a test quantity $T(z)$ (Gelman, Carlin, et al., 2020, p. 145) to evaluate the predictions solely on an aspect of the forecast that is important for the decision we are trying to inform; which in this case is the maximum wear measurement. To do so, I re-fit the models to all possible combinations of five, six, seven, and eight observations and predict from the most recent observations to the withheld future observations. For example, in one combination, I fit the models to the observations at $(t_1, t_3, t_4, t_5, t_6, t_7)$ —leaving out t_2 , t_8 and t_9 —and then predict the wear profile at t_8 and t_9 . I drop out observations in this way to remove the effect of individual observations; a similar concept to resampling techniques such as bootstrapping. I then calculate the predictive distributions of the maximum wear measurement from the two forecasts and compare them with the observed maximum wear measurement. I also compare the point estimate method of Webb et al. (2020) demonstrated in Fig. 6.2.

To generate the predictive distributions for the maximum wear measurement,

I first average over the UT measurement error by sampling from

$$\tilde{z}_{n,I+1} \sim N \left(\sum_{m=1}^M b_m(n) y_{m,I+1}, \sigma \right) \quad (6.7)$$

using the forecasted joint distribution of the $\{y_{m,I+1}\}_{m=1}^M$ and the posterior draws of σ . I then calculate $\text{Max}(\{z_{n,I+1}\}_{n=1}^N)$ for each set of draws. I average over the UT measurement error now because, in this case, we are directly comparing the prediction with the noisy observations. The predictive distributions are compared with the observed maximum UT measurements and the method of Webb et al. (2020) in Figure 6.12. Each subplot shows the the predictive distribution of the gamma process in blue and the general path model in red as well as the observed max wear measurement as a black vertical line and the point estimate of the max wear generated according to Webb et al. (2020) as a vertical red line. The title of each plot shows a vector indicating which functional observations the models were fit to and a number indicating the forecast time. For example, '[1, 0, 1, 1, 1, 1, 1, 0, 0] → 8' indicates that the models were fit to the observations at $(t_1, t_3, t_4, t_5, t_6, t_7)$ and that the forecasts are generated for t_8 .

The comparisons of Figure 6.12 show that the point estimates of Webb et al. (2020) and the maximum a posteriori (MAP) estimates of the Bayesian models are similar, and both reasonably predict the maximum wear measurement. However, the Bayesian MAP estimates are more robust when fewer observations are used to produce the forecasts. Interestingly, the point estimate of Webb et al. (2020) is typically closer to the MAP of the gamma process model than the linear general path model. Comparing the two predictive densities, the gamma process degradation model is much more optimistic than the linear general path. It also looks like it is generally as good or better than the linear general path model at predicting the maximum wear. In general, the observed maximum wear measurement is always contained in both the predictive distributions; with the exception being observation 8. A closer look at the eighth observation shows

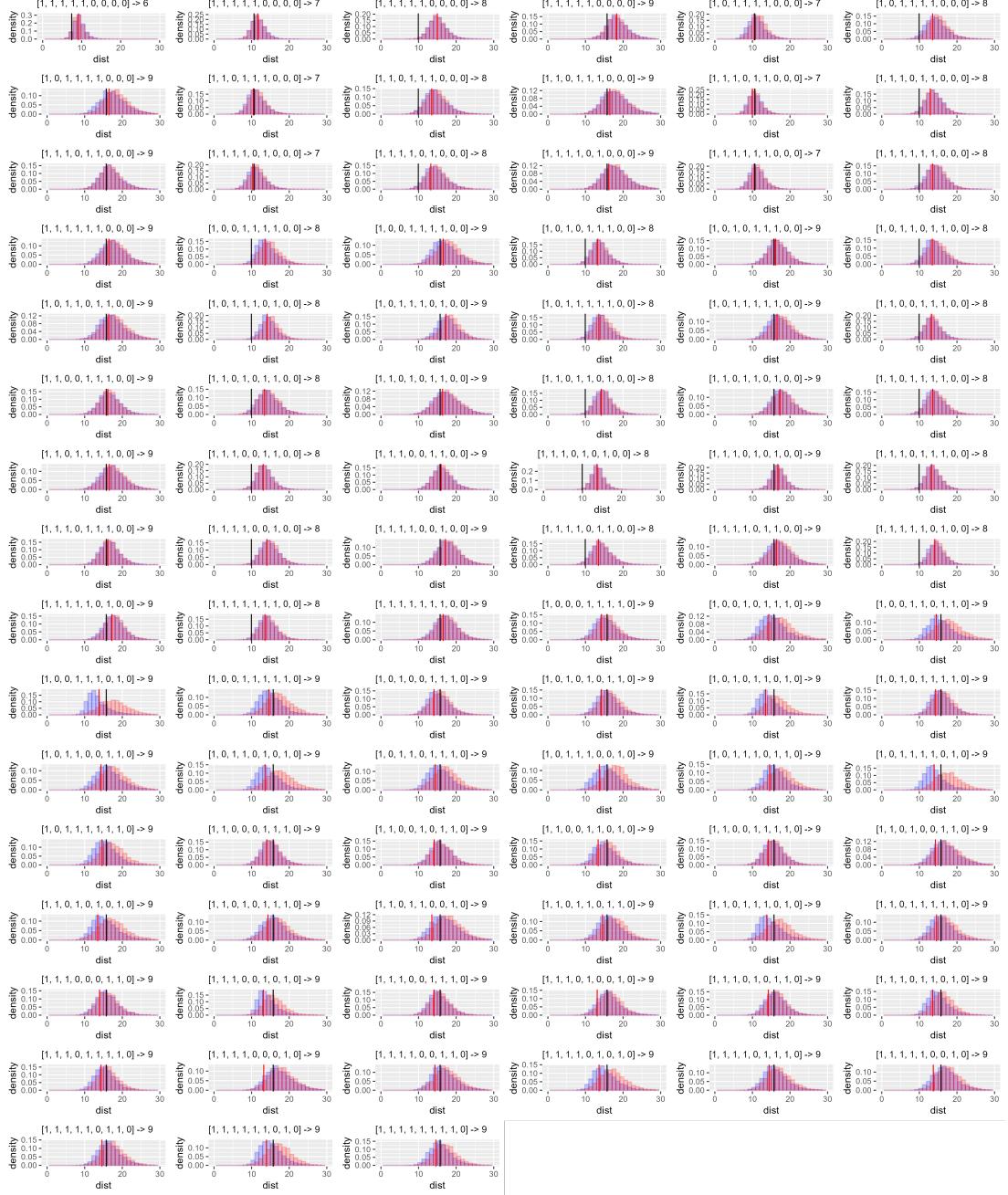


Figure 6.12: Comparisons of the forecasted maximum wear according to the two Bayesian models (the blue histogram shows the posterior predictive distribution from the gamma process and the red from the linear general path) and the method of Webb et al. (2020) (red line) with the observed maximum wear (black line). In each plot, the different methods are fit to a subset of the data and forecast a withheld future observation. For example, the plot title ' $[1, 0, 1, 1, 1, 1, 1, 0, 0] \rightarrow 9$ ' indicates that the model was fit to the observations at $(t_1, t_3, t_4, t_5, t_6, t_7)$ —leaving out t_2 , t_8 , and t_9 —and that the forecast is generated for t_8 .

that it sits below the seventh observation in many places along the profile, even though it is far from the seventh observation in tonnage. Hence, the eighth observation may well be an extreme outlier. As a reliability practitioner, if you were being conservative, you would use the predictions of the linear general path model. However, if reducing maintenance cost was important, then effort should be spent to properly validate the two methods since if the gamma process model is in fact a better representation of the true data generating mechanism then making decisions based on the linear general path model's predictions would result in the belt being replaced while it still has remaining useful life and, hence, overspending on maintenance.

6.7 Failure time distributions

The original motivation for forecasting belt wear was to inform practitioners decisions about when to replace the belt. To facilitate this decision, I use the fitted models to generate failure time distributions for the belt conditioned on its current state of degradation. I demonstrate calculating the failure time distributions from the seventh and eighth observation since at the time of the ninth observation there is a non-negligible probability that some point along the belt has already exceeded the soft failure threshold of $25mm$ —this is reflected in the max wear predictions for the ninth observation time in Figure 6.12—and so the failure time distribution is not very helpful. The failure time distribution describes our uncertainty about when the belt will reach the soft failure threshold (the first passage time). Note that unlike the failure time distributions in Chapter 5, in this case, the noise in the degradation model for the spline coefficients is important. For some situations, the failure time distribution $F(t)$ has an analytical solution, however, there may be no analytical solution for more complicated processes, and so the failure time distribution must be simulated using Monte Carlo evaluation (Meeker et al., 2022, p. 504-506). In the case at hand, multiple processes are driving the degradation of each spline coefficient and we must average over the spatial variability along

the length of the belt; hence, there is no straightforward analytical solution, and I must simulate the failure time distribution.

Algorithm 3 Numerical procedure for calculating the failure time distribution conditional on the fitted gamma process model and current state of degradation.

```

for each posterior draw do
2:   for  $j = 1$  to 1000 do
      Starting from most recent filtered spline coefficients.
5:      $\Delta t \leftarrow 0.001$ ,  $t \leftarrow t_I$ ,  $\underline{y}^* \leftarrow \{y_{n,I}^*\}_{m=1}^M$ 
      Average over the variability along the length of belt.
6:      $\underline{y} \sim t_{10}(\underline{y}^*, \underline{y}^* \phi)$ 
      Calculate the value of B-spline by multiplying the design matrix by
      column vector of coefficients.
8:      $\underline{z} \leftarrow B \cdot \underline{y}$ 
     while  $\text{Max}(\underline{z}) < 25$  do
10:     $\Delta y_m^* \sim Ga\left(\frac{\Delta t}{\nu_m^2}, \frac{1}{\nu_m^2 \mu_m}\right)$ 
         $t \leftarrow t + \Delta t$ ,  $\underline{y}^* \leftarrow \underline{y}^* + \{\Delta y_m^*\}_{m=1}^M$ 
12:     $\underline{y} \sim t_{10}(\underline{y}^*, \underline{y}^* \phi)$ 
         $\underline{z} \leftarrow B \cdot \underline{y}$ 
14:   end while
      $FT[j] \leftarrow t$ 
16: end for
      Calculate the empirical cdf from  $FT$ .
18: end for

```

To calculate the failure time numerically from the gamma process model I use the procedure in *Algorithm 3*. For each of the 12000 posterior draws, I simulate 1000 pathways from the most recent observation until soft failure. Each pathway is simulated by incrementing the GP forward by time steps of 0.001 and averaging over the variability along the length of the belt until the degradation at some point exceeds the soft failure threshold. Using the times at which each of 1000 simulated pathways cross the soft failure threshold, I construct an empirical failure time CDF. The result of *Algorithm 3* is 12000 empirical CDFs—one for each of the posterior draws. I construct the failure time distribution from the general path model in the same way, except that when I calculate the filtered values of the spline coefficients, I do so using the deterministic degradation function $\underline{y}^* = \underline{\mu}t$. The distributions of the CDFs are shown in Figure 6.13 for the gamma

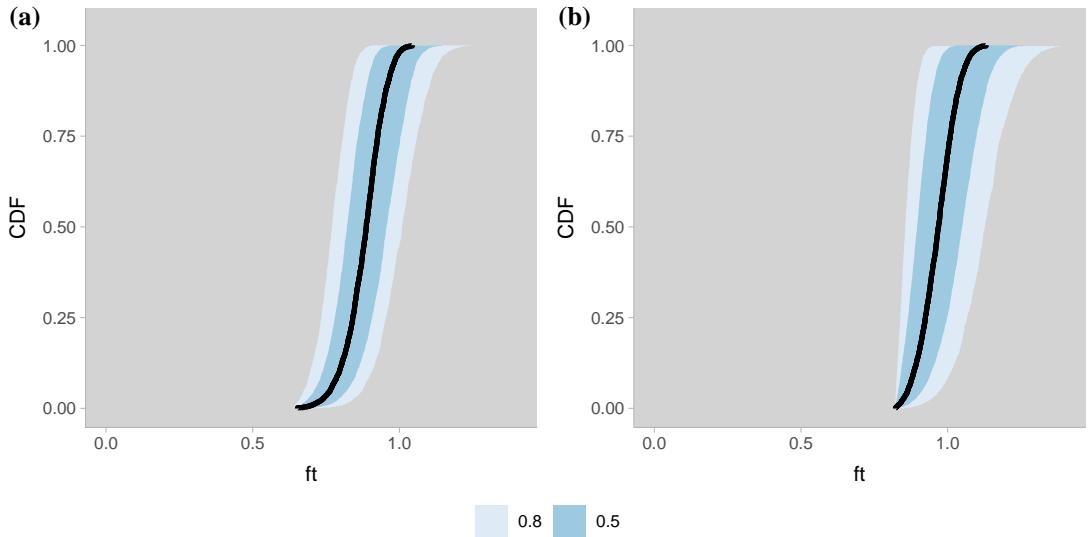


Figure 6.13: The failure time CDFs generated from the posterior of the gamma process model conditioned on the first (a) seven and (b) eight observations. The median CDF is indicated by the black line and the uncertainty intervals are shown by the different shades of blue ribbons.

process and Figure 6.14 for the general path model fitted to the first seven observations (Fig. (a) and 6.14 (a) respectively) and for the first eight observations (Fig. 6.13 (b) and 6.14 (b) respectively). The average CDF is shown as a black line and the 0.5, 0.8, and 0.95 uncertainty intervals in different shades of blue. From the distributions of the failure time CDFs, we can quickly interpret the risk of the soft failure of the belt as we delay the replacement time.

Remembering that maintenance decisions are made in the context of the whole business, not just the specific asset; a reliability engineer may plan to replace belts when there is a 25% chance that they have exceeded the soft failure threshold—remembering that soft failure means that the belt is still technically operational. At the time of the seventh observation, for this particular belt that I have analysed, the predicted time from the gamma process (linear general path) model that there is a 25% chance of soft failure is at 0.82(0.81) tonnes with a lower and upper bound of 0.75 and 0.95(0.74 and 0.90) respectively. Say there are two planned maintenance shutdowns for the particular conveyor coming up; a reliability practitioner can use this distribution to inform which shutdown they

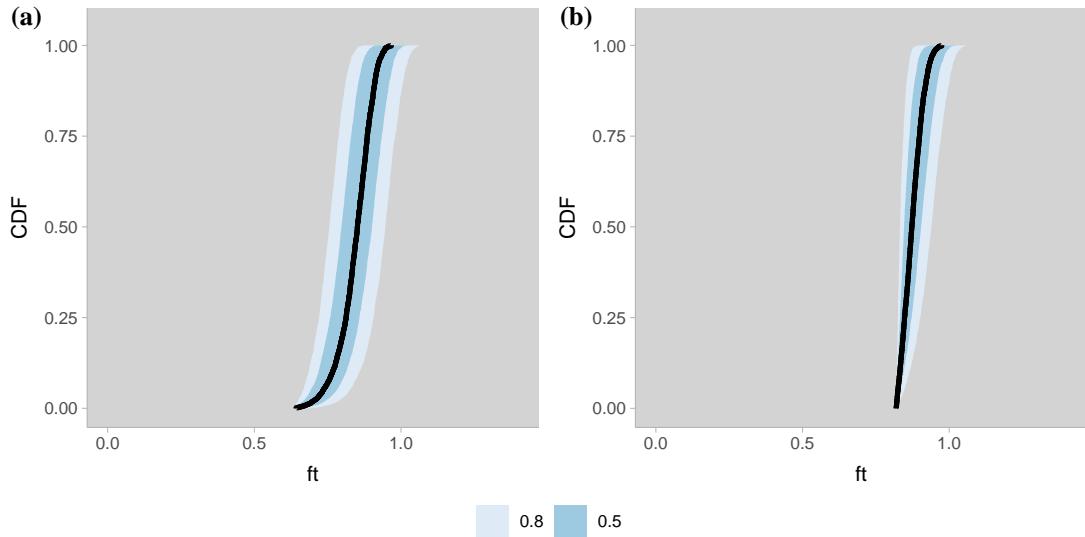


Figure 6.14: The failure time CDFs generated from the posterior of the linear general path model conditioned on the first (a) seven and (b) eight observations. The median CDF is indicated by the black line and the uncertainty intervals are shown by the different shades of blue ribbons.

should replace the belt in. The uncertainty intervals can also help to select one belt over another. Furthermore, since the failure time prediction is with respect to tonnes, the reliability practitioner could look at reducing (or even increasing) the planned operation of the belt to manage the probability of soft failure at the calendar time of the maintenance shutdown. Once we observed the eighth set of measurements, our belief is then updated, and the estimate for the time at which there is a 25% chance of soft failure becomes 0.95 (0.85) tonnes, and the lower and upper uncertainty bounds become 0.83 and 1.09 (0.83 and 0.91), respectively. Using this updated belief, the reliability practitioner could re-evaluate their decision and make last-minute changes to a shutdown plan while formally managing risk.

6.8 Discussions

In this chapter, I've constructed, fit, evaluated, and compared two BHM's for conveyor belt wear, and, in doing so, demonstrated an end-to-end example of the

Bayesian workflow for an applied problem in the mining industry. In the data models for the two BHMs, I extend FDA to degradation modelling in order to model a degrading surface. In the two process models, I compared the noisy gamma process model from chapters 4 and 5 with a linear general path model. Lastly, in the parameter models, I show how historical belt wear data can be used to inform the analysis of the current belt through an informative prior. I compare the two models with one another through the expected log score of N step ahead predictions and also compare them alongside the method of Webb et al. (2020) in their ability to predict the maximum wear measurement. Lastly, I've shown how to construct failure time distributions based on the current degradation of the belt for the two BHMs. In this last section, I distil the main points of the chapter, discuss the advantages and limitations of the BHM models I have explored and point to areas of future work.

The comparison of the two Bayesian models shows that while both models appear to have reasonably fit the data, the linear general path model is a better predictive model for the overall belt's wear. However, when the two models are compared based on their ability to predict the maximum wear observation, which defines the soft failure of the belt, both models' predictions appear reasonable when compared with the observed data and the gamma process model's forecasts are far more optimistic. Consequently, this is also reflected in the failure time distributions generated from the two models. If these models were to be implemented in practice, it would be worth investing in collecting more detailed data for a short period of time to properly validate the models—i.e. collecting wear profiles more frequently and measuring more than one location along the belt's length at each time.

An added advantage of the BHM structure applied to belt wear is that the structure can easily incorporate additional observations into the model without the model becoming over-parameterised and can take full advantage of such extra information to reduce the uncertainty of parameter estimates and predictions.

This is particularly true for the noisy gamma process model. For example, if we were to have a much finer grid of measurements across the width of the belt’s surface at each observation time, these measurements would still be summarised by the same number of spline coefficients. So, the result would be better uncertainty quantification of σ with no additional parameters in the lower levels of the hierarchical model. Alternatively, if more than one functional observation was recorded at each observation time, this could be incorporated by drawing more than one realisation of the noisy spline coefficients. To elaborate, instead of a single set of noisy spline coefficients at each time, we could use $y_{j,i,m} \sim N(y_{i,m}^*, y_{i,m}^* \phi)$, where the new subscript j identifies the different functional observation at time t_i . The result would be better identification of ϕ —the spatial variation in wear profiles along the length of the belt—and subsequently a more precise filtered estimate of the $y_{i,m}^*$. Lastly, if observations were collected more frequently, then the finer temporal resolution would better identify the ‘jumpiness’ of the gamma process—i.e. ν would be estimated more precisely—which refines the uncertainty quantification of the forecasts from the gamma process.

An alternative extension of the data would be if the set of UT measurements at each time were two-dimensional. Here, we have shown an application of this model to a one-dimensional surface; however, the method could easily be extended to a two-dimensional wearing surface by using two-dimensional spatial basis functions such as in Wikle et al. (2019, p. 84). For the case of belt wear, we could extend the model in this way if there were multiple functional observations at each time, and we knew the location along the length of the belt of each observation. By expanding the model to two dimensions, the method could be applied to the degradation of many other assets—for example, the wear liners in transfer station shoots or haul truck beds.

As I touched on briefly when discussing the prior predictive checks in Section 6.3, I have made the simplifying assumption that the spline coefficients are independent of one another. In doing so, we neglect any large-scale spatial struc-

ture in the wear profiles. For both prior and posterior predictive checks, if we simulate a single belt profile, it will look unrealistically ‘wiggly’. Although the B-spline accounts for small-scale spatial structure in the UT measurements, the assumed independence of the spline coefficients means that there is no way to capture any large-scale spacial structure. In reality, we would expect neighbouring coefficients to behave somewhat similarly. A possible way of accounting for this large-scale spatial correlation would be to include a spatial random effect, which would most likely result in better uncertainty quantification in the parameter estimates and forecasts since we are adding information about the underlying process through the model’s structure. One hurdle to implementing a spatial random effect in the gamma process model is there is no straightforward way of coercing correlation in the jumps of multiple gamma processes. This would be an interesting area for future work. Because of this hurdle, it may be simpler to implement large-scale spatial structure in the linear model. This could be done using a conditional autoregressive structure. Nevertheless, the simplifying assumption of independence appears acceptable since the ‘wiggliness’ of the individual realisations is ‘washed out’ when I average over all of the posterior draws, making any predictions look smooth, such as in Fig. 6.11.

When confronted with analysing complicated, small, and messy datasets—something very common in reliability and condition monitoring—a very natural approach is to simplify the data and apply methods we are familiar with, such as regression. However, here we have demonstrated that if we instead take the time to think deeply about how the data arise and what extra knowledge we possess about the data-generating process, we can construct statistical models that take full advantage of all the information available in both the data and our understanding of the problem. In doing so, we get more detailed predictions and defensible uncertainty qualification for the reliability predictions and accompanying maintenance decisions.

Chapter 7

Thesis discussion

In this thesis, I have developed and built upon reliability methods using the Bayesian framework to make these models applicable to maintenance problems in the mining and mineral processing domain. The works in the two main body parts of the thesis were motivated by industry problems—found through my time on industry placements—that were not solvable by using methods from the reliability literature due to incompleteness and noisiness of the industry data. Through the two parts, I address two main *theory-practice gaps* that will support reliability engineers in domains such as mining and mineral processing industry. The first is handling partially observed lifetime data that can be both left-truncated with unknown exposure history and right-censored; a case that naturally arises when units have been repeatedly replaced for many years and only the failure records after a given date are available. The second part looked at the gamma process and how through the Bayesian hierarchical modelling formalism necessary, non-trivial extensions can be included, such as the need to account for measurement error or unit-to-unit variation and how to accommodate wearing surfaces.

In these final pages of the thesis, I start in section 7.1 by summarising the work from the body of the Thesis and emphasis the key contributions. I then revisit the most significant areas for future work in section 7.2. In section 7.3, I discuss the impacts of the work for reliability practitioners in industry and conclude the

thesis.

7.1 Overview and main discussions

Part I of the thesis—consisting of chapters 2 and 3—focused on lifetime analysis, specifically lifetime data that can be both left-truncated with unknown exposure history and right-censored. In chapter 2, I developed a Bayesian model for such lifetime data that imputes the true values of the partially observed lifetimes and, along with them, the missing truncation times in order to evaluate the likelihood. I also built upon the method, originally proposed by Kaminskiy and Krivtsov (2005), for eliciting a joint prior for the Weibull parameters by embedding it fully within a Bayesian lifetime model so that the prior is properly filtered through the likelihood of the observed data. Using simulated data, I demonstrated the imputation method with both weakly and strongly informative priors, comparing it with the case where exposure times for the left-truncated lifetimes are known and where the left-truncated lifetimes are discarded. Chapter 2 ended with a small simulation study to evaluate when it is suitable to use the imputation treatment of the left-truncated lifetimes with unknown exposure history.

Left truncation with unknown exposure time is difficult to identify in a dataset and is easy to confuse with simple censoring. It is important to either account for the truncation in the lifetimes that begin before the beginning of the observation period or discard the lifetimes completely; otherwise, the results of the analysis will be biased. However, when formulating the likelihood of the data given the parameters in the usual way—that is, integrating out the censored observation—, there is no obvious way of accounting for the missing truncation times. I have shown that if we instead take the approach of imputing the partially observed censored lifetimes, any lifetimes that begin before the observation period are a case of censoring, and by imputing their values, we can then calculate a corresponding truncation time. I’ve also shown that this is easily implemented in probabilistic programming languages like Stan.

The second main body chapter of Part I, chapter 3, applies the methods developed in Chap. 2 to the industry dataset of idler-frame lifetimes. In the chapter, I analysed the idler-frame lifetimes with an informative prior, imputing the partially observed samples and truncation times to perform inference and retain as much of the information in the dataset as possible. I then showed how, by imputing the partially observed lifetimes in the model, the posterior naturally contains predictive draws for the frames still under test conditioned on their age. Using these draws, I showed how to generate predictive distributions for the RUL of frames still in operation at the end of the observation period and how to generate predictive distributions for the cumulative number of failures going forward. Lastly, in the chapter, I showed how the joint draws of the Weibull parameters can be passed through a cost function to include uncertainty in the analysis when choosing a fixed time replacement interval for the idlers on the conveyor.

Part II of the thesis focused on degradation modelling, mainly using gamma stochastic processes. The first two chapters, chap. 4 and 5, focused on theoretical developments of the gamma process model to include noisy observations and unit-to-unit variability. The third and final chapter, chap. 6, applied the gamma process to the industry dataset for conveyor belt wear and made comparisons with a linear general path model.

Chapter 4 organises some of the literature on the noisy gamma process by showing how 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 degradation model into data, process, and parameter models removes the need for complex deconvolution that requires the evaluation of, or approximations to, multidimensional integrals. The chapter also presents a reparametrisation of the gamma process in terms of the mean μ and coefficient of variation ν , which simplifies prior specification since μ is clearly interpretable as the average wear rate and ν is the volatility of the process. In addition, the

parameters μ and ν are orthogonal, which has desirable computational benefits.

I demonstrated the hierarchical noisy gamma process with parameters μ and ν on a simulated dataset and, in doing so, highlighted a presymptomatic non-identifiability between the measurement error and the volatility of the gamma process when the model is fit to a small number of noisy observations. This observation is particularly important in reliability applications since degradation datasets are often small. The weak identifiability between the standard deviation of the measurement error and the coefficient of variation of the gamma process highlights the importance of incorporating domain knowledge or supplementary data into the model or the partial pooling of information between similar units.

Chapter 5 then showed how the same BHM for the noisy gamma process can be naturally extended to include unit-to-unit variability from multiple noisily-observed degradation signals and how allowing some parameters to vary determines how information is shared between observational units. The proposed parameterisation in terms of μ and ν makes expansions of the model, such as unit-to-unit variability, more obvious, i.e. how similar do we expect the units to be in terms of their wear rate, their volatility, or both? In Chap. 5, I fitted several variations of the noisy gamma process with unit-to-unit variability to experimental crack growth data with added noise. I demonstrated how to check and compare the models in a fully Bayesian framework using posterior predictive distributions and *elppd* and cross-validation. In the last part of the chapter, I show how to construct failure time distributions, with uncertainty, for units that are under test but have not failed and for new units.

In Chapter 6, I analysed the wear profile dataset from an iron ore conveyor's belt. To model the profiles, I proposed a functional data analysis approach, where the smooth B-spline functions that describe the wear profile at each observation time evolve according to a degradation model. I compare, side-by-side, a noisy gamma process and a linear general path model for the evolution of the spline coefficients. To the best of my knowledge, there is no work of functional degrada-

tion models in the literature. Using the functional models, I show how to forecast the degradation profile of the belt into the future and how to construct failure time distributions that account for the uncertainty in the wear profile along the length of the belt.

In all of these chapters, I follow the precepts of good practice in the Bayesian statistical workflow and demonstrate: carefully constructing informative priors and checking these assumptions through prior predictive checking; assessing models using simulation; checking computation with HMC diagnostics; visualising the posterior through posterior predictive distributions or checking; and comparing models in a fully Bayesian framework. I place a particularly strong emphasis on how to encode prior information since either small sample sizes or weekly informative data (i.e. due to partially observed data or signals obscured by noise) are prominent problems in reliability datasets from the mining industry.

7.2 Future directions

pt.1 The method I propose for imputing the partially observed left-truncated lifetimes and their truncation times needs further investigation. For example, I did not fully separate the issues with the approach due to implementation from those due to breaches in the assumptions made. When implemented in Stan, there appears to be some updating in the imputed truncation times when the lifetime is both left-truncated by the beginning of the observation period and right-censored by its end, even though there should not be any information in the data to inform this parameter. The method could be implemented in Bugs (Lunn et al., 2013), where it is possible to include nuisance parameters that are not updated in the MCMC routine to overcome this issue, or the same could be achieved by constructing the MCMC algorithm from scratch. However, in the case of the idler-frames, only one lifetime was left truncated and right censored, so this limitation with the method should not affect results. There are also issues with the approach when the start of the observation period is too close to the

origin of the repeated replacement process ($t = 0$). However, the boundary for what is acceptable is not clearly defined. More rigorous simulation would be necessary to identify this boundary.

The approach I have proposed assumes that all units (frames) are identical. However, additional information about the idler frames could be included in the model, such as where along the conveyor they are located—in the impact zone, at the head or tail pulley transitions—or the manufacturer type—which is available for recently installed idlers. It would be both interesting and useful to allow for similar but different groups of idlers to share information, such as the Hierarchical models in Part II of the thesis, and to include covariates in the model. Additionally, In the analysis of the idler-frames, there are a number of very short lifetimes that I treat as right-censoring events. These early failures still threaten to cause downtime. The infant mortalities could be included in a mixture model such as in Mittman (2018); however, it would be important to understand how a mixture of lifetime distributions would affect the imputation of the partially observed left-truncated lifetimes and their truncation times.

pt.2 In fitting the noisy gamma process, there is a point when the noisy observation of the signal obscures the ‘jumpiness’ of the degradation trace, and it is much easier to fit a linear model rather than a stochastic process. More work needs to be done to understand the interplay between the volatility of a degradation signal and its noisy observation. In addition, noisy observations also add uncertainty to whether or not unit-to-unit variability is necessary in the model. For example, if we identify that unit-to-unit variability is present for a non-noisy set of degradation traces, is there a level of measurement error where the conclusion is not the same? A further question is how sensitive are metrics such as elppd_{CV} in these cases, and would they lead us to the correct model in these circumstances? These questions would need to be explored through simulation, simulating data from a known model and seeing if the model can be identified under different conditions.

In Chapter 6, I have shown how FDA can be used to model a one-dimensional degrading surface. There are now companies that perform point cloud scans of a surface. Extending the method to two dimensions would be useful for this type of data. I also highlight in section 6.8 that the model I propose does not account for large-scale spatial dependence between the spline coefficients. When analysing a two-dimensional dataset with observations on a densely observed grid, it would be interesting to incorporate large-scale spatial effects into the model. One possible way of doing this could be to use a conditional autoregressive (CAR) prior.

7.3 Implications for industry practitioners

Throughout the thesis, the Bayesian approaches I have presented address real problems in reliability data from the mining and mineral processing sector—retaining the information from lifetimes that are left truncated with unknown exposure history; modelling multiple noisy degradation signals with stochastic degradation models; and modelling the degradation of a wearing surface using functional data analysis. In addition to addressing these practical problems, the Bayesian approaches are implemented in Stan, an accessible and open-source probabilistic programming language, so that the models can be implemented and adjusted easily by a wide audience. Throughout the work, I have stressed the components of what is considered to be a good Bayesian statistical workflow and demonstrated their application in the reliability domain. These include thinking carefully about how a model can be encoded in order to simplify its implementation; evaluating and refining the model and priors through simulation before fitting the data; interpreting uncertainty in the posterior through posterior predictive distributions; propagating this uncertainty through useful utility functions such as cost functions or the failure time distribution; and comparing sets of Bayesian models using elppd and cross-validation methods. In the interests of reproducible research, I provide all of the code to implement the models and perform the analysis of this thesis on a Github repository [Leadbetter et al.,

2024] so that other reliability practitioners can follow along, reproduce and apply our work in practice, and adapt the models and workflow to their own industry problems.

Of these Bayesian workflow components, I have focused on defining a well-constructed prior and propagating posterior uncertainty. I've stressed how supplementary information can be included in the analysis through the prior (or directly in the likelihood in the case of section 4.5.4) and the benefits when analysing small datasets because this seems to be a common occurrence with reliability data in industry. New monitoring devices promise to supply detailed and large datasets in the future, but good decisions must be made now. Much of the historical datasets are small and messy, and the wealth of industry domain expert knowledge is incredibly helpful in these cases. I've also stressed the demonstration of simple examples of using the posterior of a Bayesian model to inform short and long-term maintenance decisions. Little in the literature guides practitioners on how to incorporate the uncertainty encoded in a posterior distribution into maintenance decisions. More works in Bayesian reliability, like this one, should present examples of this last step in the Bayesian *reliability* workflow.

Appendices

Appendix A

Authorship Agreement

References

- Abdel-Hameed, M. (1975). A gamma wear process. *IEEE transactions on Reliability*, 24(2), 152–153. (Publisher: IEEE)
- 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>
- Balakrishnan, N., Tsai, C.-C., & C.-T., L. (2017). Gamma degradation models: Inference and optimal design. In D.-G. Chen (Ed.), *Statistical modeling for degradation data* (pp. 171–191). Singapore: Springer Nature Singapore Pte. Ltd.
- 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. (2015). *Adiabatic monte carlo*.
- 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)
- Betancourt, M., & Girolami, M. (2015). Hamiltonian monte carlo for hierarchical

models. In S. K. Upadhyay, U. Singh, D. K. Dey, & A. Loganathan (Eds.), *Current trends in bayesian methodology with applications* (p. 24). Chapman and Hall/CRC. doi: 10.1201/b18502

Bortnowski, P., Kawalec, W., Król, R., & Ozdoba, M. (2022). Types and causes of damage to the conveyor belt - review, classification and mutual relations. *Engineering Failure Analysis*, 140, 106520. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1350630722004940> doi: <https://doi.org/10.1016/j.engfailanal.2022.106520>

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B*, 49(1), 1–39.

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.

Emura, T., & Michimae, H. (2022). Left-truncated and right-censored field failure data: Review of parametric analysis for reliability. *Quality and Reliability Engineering International*, 38(7), 3919-3934. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.3161> doi: <https://doi.org/10.1002/qre.3161>

Esposito, N., Mele, A., Castanier, B., & Giorgio, M. (2022, January). A new gamma degradation process with random effect and state-dependent measurement error. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1748006X2110672. Retrieved 2022-10-25, from <http://journals.sagepub.com/doi/10.1177/1748006X211067299> doi: 10.1177/1748006X211067299

Frenk, J., & Nicolai, R. (2007, May). *Approximating the randomized hitting time distribution of a non-stationary gamma process* (Tech. Rep. No. 2007-18).

Erasmus University Rotterdam, Rotterdam, The Netherlands: Econometric Institute.

- Gabry, J., & Mahr, T. (2024). *bayesplot: Plotting for bayesian models*. Retrieved from <https://mc-stan.org/bayesplot/> (R package version 1.11.1)
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019a). Discussion on the meeting on ‘Data Visualization’. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2), 419–441.
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019b). Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2), 389-402.
- Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1), 1 – 53.
- 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., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457 – 472. Retrieved from <https://doi.org/10.1214/ss/1177011136> doi: 10.1214/ss/1177011136
- 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.

- Giorgio, M., Mele, A., & Pulcini, G. (2019, March). A perturbed gamma degradation process with degradation dependent non-Gaussian measurement errors. *Applied Stochastic Models in Business and Industry*, 35(2), 198–210. Retrieved 2022-05-10, from <https://onlinelibrary.wiley.com/doi/10.1002/asmb.2377> doi: 10.1002/asmb.2377
- 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. (2008a). *Bayesian reliability*. Springer New York, NY.
- Hamada, M. S., Wilson, A. G., Reese, C. S., & Martz, H. F. (2008b). *Bayesian reliability*. New York, NY: Springer Science+Business Media, LLC.
- Hazra, I., Bhadra, R., & Pandey, M. D. (2022, December). Likelihood-free Hamiltonian Monte Carlo for modeling piping degradation and remaining useful life prediction using the mixed gamma process. *International Journal of Pressure Vessels and Piping*, 200, 104834. Retrieved 2022-11-23, from <https://www.sciencedirect.com/science/article/pii/S0308016122002198> doi: 10.1016/j.ijpvp.2022.104834
- Hazra, I., Pandey, M. D., & Manzana, N. (2020, June). Approximate Bayesian computation (ABC) method for estimating parameters of the gamma process using noisy data. *Reliability Engineering & System Safety*, 198, 106780. Retrieved 2023-02-16, from <https://linkinghub.elsevier.com/retrieve/pii/S0951832019309901> doi: 10.1016/j.ress.2019.106780
- Heckman, J. J., & Singer, B. (1986). Chapter 29: Econometric analysis of longitudinal data. In (Vol. 3, p. 1689-1763). Elsevier. Retrieved from <https://www.sciencedirect.com/science/article/pii/S157344128603009X> doi: [https://doi.org/10.1016/S1573-4412\(86\)03009-X](https://doi.org/10.1016/S1573-4412(86)03009-X)
- 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
- Hodges, J. S. (2014). *Richly Parameterized Linear Models. Additive, Time Series, and Spatial Models Using Random Effects*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Hoffman, M. D., & Gelman, A. (2014). The No U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1351–1381.
- 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
- Hormann, S., & Kokoszka, P. (2012, 12). Functional time series. *Handb. Stat.*, 30. doi: 10.1007/978-1-4614-3655-3_16
- Huzurbazar, V. S. (1956). Sufficient statistics and orthogonal parameters. *Sankhyā: The Indian Journal of Statistics*, 17(3).
- 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 equipment*. 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.
- Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). *Bayes rules! an introduction to applied bayesian modeling*. Boca Raton, FL: CRC Press.

- Kallen, M., & van Noortwijk, J. (2005, November). Optimal maintenance decisions under imperfect inspection. *Reliability Engineering & System Safety*, 90(2-3), 177–185. Retrieved 2022-03-02, from <https://linkinghub.elsevier.com/retrieve/pii/S0951832004002443> doi: 10.1016/j.ress.2004.10.004
- Kaminskiy, M., & Krivtsov, V. (2005, 01). A simple procedure for bayesian estimation of the weibull distribution. *IEEE Transactions on Reliability*, 54, 612-616.
- Kowal, D. R. (2017). *Bayesian methods for functional and time series data* (Doctoral dissertation, Cornell University). Retrieved from <https://higherlogicdownload.s3.amazonaws.com/AMSTAT/7328eb46-c57e-49af-8d4f-0460ca4a6a2c/UploadedImages/Kowal.pdf>
- Krivtsov, V., & Frankstein, M. (2017). A bayesian estimation procedure of reliability function for lifetime distributions. *International Journal of Performativity Engineering*, 13(2), 129.
- 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.
- Leadbetter, R., Caceres, G. G., & Phatak, A. (2024). *Bayesian hierarchical modelling of noisy gamma processes: Model formulation, identifiability, model fitting, and extensions to unit-to-unit variability*.
- 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. (Cited by: 215; All Open Access, Bronze Open Access, Green Open Access) doi: 10.1214/088342306000000330

- Lu, C., & Meeker, W. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*.
- 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.
- Lu, D., Pandey, M. D., & Xie, W.-C. (2013, August). An efficient method for the estimation of parameters of stochastic gamma process from noisy degradation measurements. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 227(4), 425–433. Retrieved 2022-06-08, from <http://journals.sagepub.com/doi/10.1177/1748006X13477008> doi: 10.1177/1748006X13477008
- 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.
- McElreath, R. (2020). *Statistical rethinking* (2nd. ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data* (1st. ed.). New York, NY: Wiley.
- Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2022). *Statistical methods for reliability data* (2nd. ed.). Hoboken, NJ: Wiley.
- 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.

- Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo*. Chapman and Hall/CRC. doi: 10.1201/b10905
- 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
- Peng, C.-Y., & Ai, H.-F. (2018). Degradation analysis with measurement errors. In D. C. et al. (Ed.), *Proceedings of the pacific rim statistical conference for production engineering* (pp. 95–122). Singapore: Springer Nature Singapore Pte. Ltd.
- 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.
- Pradhan, B., & Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81(9), 1187-1198.
- Pulcini, G. (2016, August). A perturbed gamma process with statistically dependent measurement errors. *Reliability Engineering & System Safety*, 152, 296–306. Retrieved 2023-02-02, from <https://linkinghub.elsevier.com/retrieve/pii/S095183201630014X> doi: 10.1016/j.ress.2016.03.024
- Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional data analysis with r and matlab*. New York, NY: Springer New York. Retrieved from <https://doi.org/10.1007/978-0-387-98185-7> doi: 10.1007/978-0-387-98185-7
- 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.

- Rodríguez-Picón, L. A., Méndez-González, L. C., Romero-López, R., Pérez-Olgún, I. J. C., Rodríguez-Borbón, M. I., & Valles-Rosales, D. J. (2021, January). Reliability Analysis Based on a Gamma-Gaussian Deconvolution Degradation Modeling with Measurement Error. *Applied Sciences*, 11(9), 4133. Retrieved 2022-11-23, from <https://www.mdpi.com/2076-3417/11/9/4133> doi: 10.3390/app11094133
- Rodríguez-Picón, L. A., Rodríguez-Picón, A. P., Méndez-González, L. C., Rodríguez-Borbón, M. I., & Alvarado-Iniesta, A. (2018, July). Degradation modeling based on gamma process models with random effects. *Communications in Statistics - Simulation and Computation*, 47(6), 1796–1810. Retrieved 2022-11-03, from <https://doi.org/10.1080/03610918.2017.1324981> doi: 10.1080/03610918.2017.1324981
- 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>
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1), 1 – 28.
- 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)
- Tian, Q., Lewis-Beck, C., Niemi, J. B., & Meeker, W. Q. (2024). Spec-

- ifying prior distributions in reliability applications. *Applied Stochastic Models in Business and Industry*, 40(1), 5-62. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2752> doi: <https://doi.org/10.1002/asmb.2752>
- van Noortwijk, J. (2009, January). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1), 2–21. Retrieved 2022-04-18, from <https://linkinghub.elsevier.com/retrieve/pii/S0951832007001111> doi: 10.1016/j.ress.2007.03.019
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5), 1413–1432.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667 – 718.
- Webb, C., Sikorska, J., Khan, R. N., & Hodkiewicz, M. (2020). Developing and evaluating predictive conveyor belt wear models. *Data-Centric Engineering*, 1, e3. doi: 10.1017/dce.2020.1
- Whitmore, G. (1995). Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Analysis*, 1, 307–319.
- 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>

Every reasonable effort has been made to acknowledge the owners of copyright

material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.