School of Electrical Engineering, Computing and Mathematical Sciences

Centre for Transforming Maintenance Through Data Science

Practical Bayesian model building for reliability analysis in industry settings.

Ryan K. Leadbetter

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

November 2023

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

Ryan K. Leadbetter

"*The Quote*"

— AUTHOR

Acknowledgements

Write your acknowledgements here

Abstract

Write your abstract here

Contents

List of Figures

Chapter 1

Introduction

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

It was apparent from my placement time that there is a disconnect between the reliability modelling literature and the methods used by reliability practitioners in the mining and mineral processing industry, often referred to as a theory-practice gap. There are many well established models for reliability data, such as the Weibull distribution for lifetime data, or stochastic process models for degradation data. There is also a desire by mine, processing plant, and refinery operators to use these models, since once an asset is put into service completely new information becomes available which can be utilised so that operators can make better maintenance decisions based on the specific reliability characteristics of their assets, rather than estimates based on experimental data (Jardine & Tsang, 2013). But the reality of collecting reliability data in the field (where there are many sources of noise) and within large companies (where it is difficult to regulate data collection practice and where there is a constant trade off between cost, safety, and time verses the quality and amount of data collected). This results in unique observational processes that need their own sophisticated modelling approaches before reliability practitioners in the mining and mineral processing industry can take advantage of the well established reliability models in the literature. Two examples of issues confronting reliability engineers are 1) obtaining reasonable estimates of lifetime distributions when lifetime data are heavily censored due the pre-emptive replacement of assets because of the risk their failures pose to safety or production, and 2) forecasting complex degradation processes with noisy and sparsely observed condition monitoring data. In this thesis I show novel expansions of some of these reliability methods through the Bayesian model building approach (Gelman et al., 2020) and demonstrate how they can be applied to observational industry data sets provided by the centres industry partners.

The Bayesian paradigm became a strong backbone of this thesis. This is 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 that is obtained through Bayesian analysis allows us to easily produce estimates and uncertainty intervals for complicated functions of the model parameters (Meeker et al., 2022), which is extremely useful for propagating uncertainty through a decision making process. While there is a well developed subfield of Bayesian analysis in the reliability literature (Hamada, Wilson, Reese, & Martz, 2008; Meeker et al., 2022), the Bayesian framework is underutilised in industry.

This is most likely because, for most cases, inference must be obtained through Monte Carlo simulation, and in the past this has meant constructing Marcov Chain Monte Carlo (MCMC) algorithms by hand. However, the recent increase in the popularity of Bayesian methods has lead 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 alleviates the analyst form the need to construct bespoke MCMC algorithms. The result is a newfound ability to fit and explore complex models relatively quickly.

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

The rest of the chapter provides a general background for the rest of the thesis. First, in *section* 1.1, I provide some context around maintenance decision making in the mining and mineral processing industry. Then, in *section* 1.2, I give a high level overview of reliability modelling and how it informs maintenance decisions. *Section* 1.3 outlines Bayesian methods and the key components of the Bayesian model building workflow which will be a strong thematic thread through out the remainder of the thesis. Finally, in *section* 1.4, I lay out the structure of the thesis.

1.1 Maintenance decision making

The maintenance of an asset can be considered as *"all activities aimed at keeping an {asset} in, or restoring it to, the physical state considered necessary for the fulfilment of its production function"* (Geraerds, 1985). In other words, the main objective of maintenance actions is to fix/replace an asset's components to ensure that the asset is able to perform its desired duty at an acceptable level of performance. In this context, the only consideration when deciding when to maintain the asset is weather 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 one specific asset, and there are some assets more critical to production than others. This "big picture" management of an assets maintenance is what we call asset health management. It is in this bigger context that reliability engineers must make their decisions about how and when to maintain an asset. The process of asset health management requires foresight, planning, and–most importantly–risk management.

Maintenance strategies help to roughly allocate resources and plan maintenance schedules ahead of time. There are three general strategies; reactive, preventative, and predictive maintenance (). We provide a more detailed overview of these strategies a little later. An asset can have different strategies for its different components and typically the choice of strategy is dictated by how critical the component is, how expensive, and what type of data we are able to collect. But even with a maintenance strategy, once an asset is put into service, we gain new

Once an asset is put into operation, we begin to added information

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). This strategy would only be used

Example of conveyor belt?

Reactive is not used unless component is non-critical. Preventative is used in the case where CM data is not available AND the number of similar assets is large, they are not highly critical and the replacement is reasonably low.

1.2 Reliability modelling

General introduction to reliability data is SMRD2 Meeker, Escobar, and Pascual

Soft and hard failure

Lifetime data

Degradation data A good bridge between lifetime data and degradation modelling is Bayesian Reliability Hamada, Wilson, Reese, and Martz.

1.3 Bayesian reliability modelling

Bayesian inference An overview of Bayesian methods.

Bayesian workflow Probably also some discourse on the Bayesian workflow.

1.3.1 Simulation for model checking

An ongoing thread through the thesis is the concept of using simulated date from a known underlying truth to assess model performance…

1.4 Structure of this thesis

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

The first part of the thesis focuses on lifetime analysis. Particularly, on how we can obtain reliable inference from Weibull analysis when data are heavily censored by thoughtfully constructing an informative joint prior. The part is composed of three chapters: chapters two, three, and four. *Chapter Two* starts with the introduction of Weibull lifetime analysis and censoring of lifetime data and then proceeds to demonstrate how, when lifetime data are heavily censored, fitting the model with maximum likelihood or Bayesian methods with commonly used priors results in biased parameter estimates. The chapter concludes by demonstrating a method for constructing an informative joint prior, which encodes information about how the parameters covary with one another and shows how encoding information into the model in this way reduces the effects of the bias caused by heavy censoring, allowing us to obtain usable estimates of the lifetime parameters. *Chapter Three* then presents a simulation study demonstrating the systematic reduction in bias provided by the informative joint prior and explores the method's limitations. From the findings in the simulations study, we consolidate our recommendations to practitioners when analysing heavily censored lifetime data. In *Chapter Four*, we apply the methods and recommendations from chapters two and three to an industry dataset. The case study analyses the censored lifetime data of idlers on an overland iron ore conveyor to inform bulk replacement strategy.

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

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

Part I

Part one: lifetime analysis

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

Chapter 2

Heavily censored lifetime data

Computerised maintenance management systems (CMMS) such as SAP (SAP SE, 2023) are now embedded in companies maintenance procedures, meaning that these companies now posses large scale datasets of component installation and replacement times. A natural use of these personalised failure time data sets is for tailoring replacement strategies for the companies specific operating environments (Meeker et al., 2022, p. 13), rather than solely relying on the manufacturers recommendations. One problem however, is that these large observational datasets collected through CMMS are much messier that the experimental ones used by manufacturers in traditional reliability/warranty analysis. This messiness comes about because of reporting issues and the fact that most components are pre-emptively replaced before they fail because of the risk to production and employees safety. The result is that many of the valuable data sets stored in CMMS systems are heavily censored.

This heavy censoring results in biassed reliability estimates. Therefore, we propose a method of mitigating this bias so that we can still use the data to inform decisions.

In this chapter we...

2.1 Background

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

2.1.1 Lifetime distribution

We model the lifetimes of the units as a random variable defined in terms of t, the exposure time, on $[0, \infty)$. t is some continuous or discrete exposure time from a clearly defined origin, the installation of the component, to a well defined event, the failure of the component.

These distributions are defined on a support. In reliability, the exposure is typically absolute time or the operating time of the unit. Say we choose a specific parametric lifetime distribution for the population of units, $p(t|\theta)$ expressed as the probability density function (CDF). Once we have estimated the parameters of the lifetime distribution we can draw useful interpretation of the

- CDF (The probability that a unit will have failed by time T, i.e. $P(t <=$

T).)

- Survival function (The compliment of the CDF.)
- Hazard (The instantaneous failure rate at a given exposure.)

2.1.2 The Weibull distribution

Details of the Weibull.

2.1.3 Censoring

What is Censoring?

2.1.4 An example from industry

2.2 Bias in heavily censored lifetime data

Introduce the problem that, when data are heavily censored, the estimates of lifetime parameters become bias. We will show via simulated data so that the underlying truth is known.

2.2.1 Simulation method

How do we simulate censored lifetime data?

2.2.2 Bias in results

How does the estimated CDF differ from the truth?

2.3 Informative Bayesian analysis

How can informative priors help us in this case?

Independent Construction of independent priors.

Joint Construction of the joint prior.

2.3.1 Effect of informative priors

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

2.4 Discussion

...

Chapter 3

Simulation study

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

3.1 Simulation structure

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

3.2 Results

Visualise/summarise the results of the simulation experiment.

3.3 Discussion

Typical results section.

3.4 Recommendations

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

Chapter 4

Case study

A case study.

4.1 Idler replacement data

4.2 Results

4.3 Discussion

Part II

Part two: Degradation modelling

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

Chapter 5

Noisy gamma process for modelling degradation measurements with uncertainty

5.1 Background

5.1.1 BHM

5.1.2 Gamma process

5.1.3 Gamma process with noise

Chapter 6

Conveyor belt wear forecasting

6.1 Background

6.1.1 FDA

6.2 Process models

6.2.1 Linear

6.2.2 Noisy GP

Chapter 7

Conveyor belt wear forecast with spatial random effect

Chapter 8

Discussion

8.1 Tie together discussion

8.2 Strengths and limitations

8.3 Future directions

8.4 Industry practitioner implications

Appendices

Appendix A

Appendix Title

Appendix B

Copyright Information

References

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.

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

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

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

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

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

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

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.

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

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