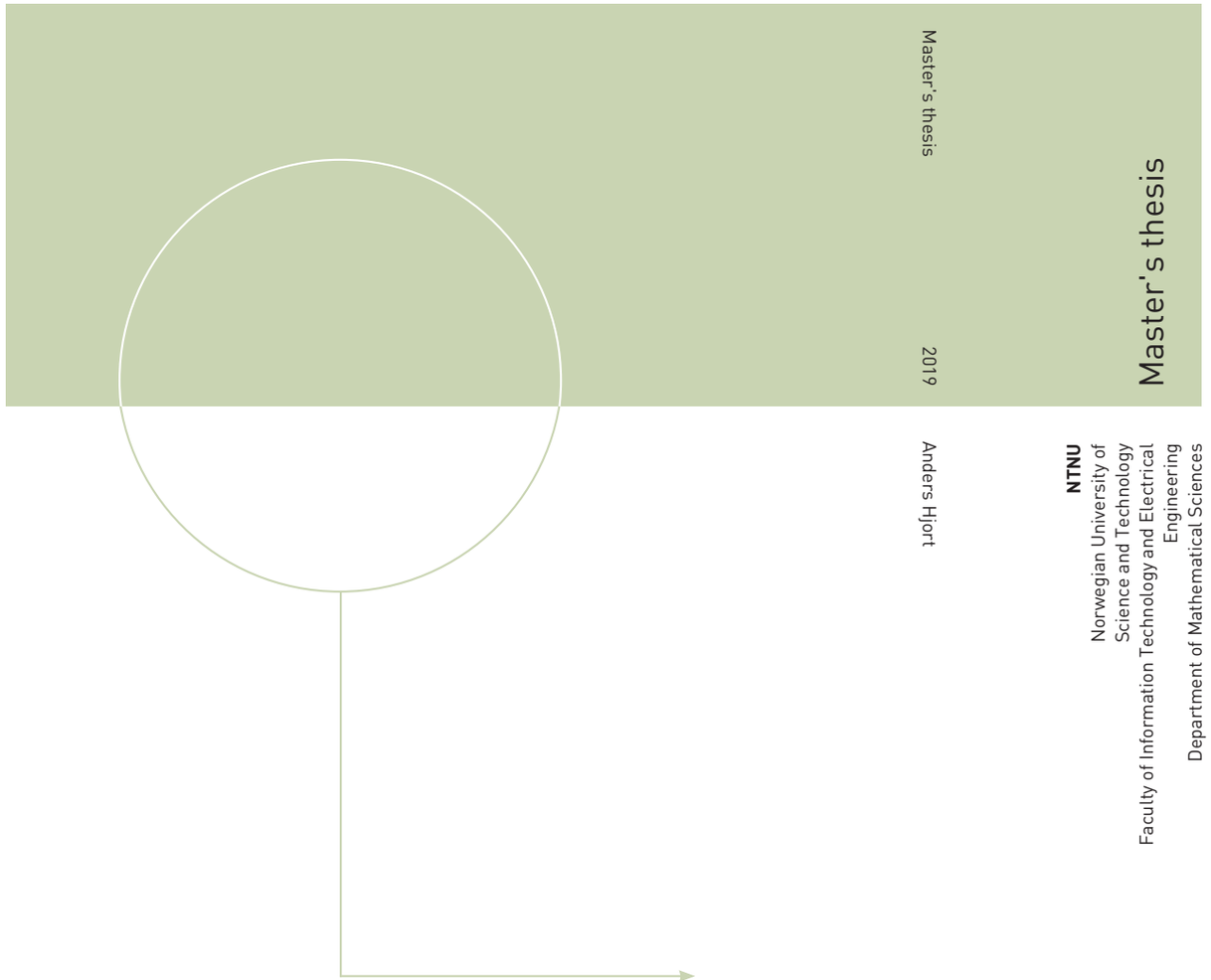

Master's thesis

2019

Anders Hjort

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology and Electrical
Engineering
Department of Mathematical Sciences

Anders Hjort

Strategies for Risk Based Inspection of Corrosion Damage in Pipeline Systems

June 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Strategies for Risk Based Inspection of Corrosion Damage in Pipeline Systems

Anders Hjort

Master of Science in Physics and Mathematics
Submission date: June 2019
Supervisor: Jo Eidsvik

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Pipes and other assets at oil refineries are subject to internal corrosion over time. The corrosion can, if not carefully monitored, cause leakages and other substantial and costly damages. Regular inspection of the pipes is therefore necessary to monitor the corrosion development. However, inspections are also costly and time-consuming, and for these reasons we aim to minimize the number of inspections through a Risk Based Inspection strategy.

A data set consisting of multiple inspections from 1019 pipes from an undisclosed oil refinery is used to train a Bayesian regression model with monthly corrosion rates and various pipe features as covariates. This data set is subsequently used to simulate synthetic time series. These time series are used to test various inspection strategies.

Two inspection strategies are suggested and compared: The Adaptive Monitoring Algorithm (AMA), which uses the probability of leakage as an inspection criterion, and the Informative Monitoring Algorithm (IMA), which uses expected gain in Value of Information as an inspection criterion. These strategies are compared with a non-adaptive method, which conducts inspections at fixed time frequencies.

The proposed strategies rely on several decision thresholds used to decide whether an inspection should take place, and whether a pipe needs to be repaired. Optimal decision thresholds are investigated and used to simulate and compare the different strategies. The simulations indicate that both the AMA and the IMA yield fewer inspections than a non-adaptive inspection strategy, while simultaneously reducing the number of leakages substantially. For certain values for the decision thresholds, AMA achieves a 0.6% probability of leakage and IMA achieves a 8.9% probability of leakage, compared with $> 30\%$ for the non-adaptive methods.

Sammendrag

Rør og andre konstruksjoner på oljeraffinerier er gjenstand for innvendig korrosjon over tid. Korrosjonen kan føre til lekkasjer og andre kostbare skader, og jevnlige inspeksjoner av rørenes tilstand er derfor nødvendig for å monitorere skadeomfanget. Å gjennomføre inspeksjoner er imidlertid både kostbart og tidkrevende, og det er derfor et mål å minimere antall inspeksjoner gjennom strategier for risikobasert vedlikehold.

Vi bruker et datasett bestående av inspeksjonsresultater fra 1019 oljerør til å trene en Bayesiansk regresjonsmodell hvor månedlig korrosjonsrate og egenskaper ved røret er brukt som kovariater. Dette datasettet brukes videre til å generere syntetiske tidsserier som vi tester ulike inspeksjonsstrategier på.

Vi foreslår to strategier for inspeksjon: En Adaptiv Monitoreringsalgoritme (AMA), som bruker sannsynlighet for lekkasje som inspeksjonskriterium, og en Informativ Monitoreringsalgoritme (IMA), som bruker forventet endring i verdien av informasjon som inspeksjonskriterium. Disse strategiene blir sammenlignet med en ikke-adaptiv modell, som foretar inspeksjoner med en fast tidsfrekvens.

De foreslåtte strategiene avhenger av flere beslutningsparametre for å avgjøre hvorvidt et rør skal inspiseres, og eventuelt repareres. Vi eksperimenterer med ulike verdier for parametrene, og bruker disse til å sammenligne de ulike strategiene. Resultatet av simuleringene indikerer at både AMA og IMA fører til en reduksjon i antall inspeksjoner, samtidig som antallet lekkasjer reduseres betraktelig. For enkelte verdier av beslutningsparametrene oppnår AMA 0.6% sannsynlighet for lekkasje og IMA 8.9% sannsynlighet, sammenlignet med $> 30\%$ for den ikke-adaptive metoden.

Preface

This master thesis is the final work of the Master of Science degree in Industrial Mathematics at Norwegian University of Science and Technology (NTNU). The thesis is written in cooperation with Oceaneering Asset Integrity in Trondheim. All the work has been conducted during the spring of 2019.

I want to express my gratitude towards Professor Jo Eidsvik for his supervision and help throughout the entire process. I am also thankful to Haaken Ahnfelt, Knut Nordanger and the rest of the team at Oceaneering Asset Integrity for providing me with a data set and for fruitful and interesting discussions.

Anders Hjort
Trondheim, June 2019

Table of Contents

Chapter 1

Introduction

1.1 The Corrosion Problem

Oil processing plants are facilities where oil is processed and refined. These facilities consist of an intricate system of pipes, pressure vessels and storage units that transfer hydrocarbons and other chemicals. This equipment is vulnerable to corrosion damage over time. The corrosion might occur on the *inside* of the pipes due to the chemical substances that move through the pipe, and it might occur on the *outside* of the pipe due to condensation, temperature fluctuations, leakages, etc. Because of the great costs associated with corrosion damages, petrochemical facilities work intensively on inspecting the pipes in order to gain information about the state of the pipe, so that mitigating measures can be conducted if needed. These inspections happen in two ways:

- **Manual inspection:** A team of human inspectors measure both qualitatively and quantitatively the current state of the oil pipe. This can yield a good data set about the current state of the pipe, but it is time-consuming and costly. A physical inspection also requires that production halts.
- **Automatic inspection:** An ultrasonic scanner, sometimes referred to as a *pig* or *scraper*, can be sent through the pipe to measure wall thickness at various points and detect damages; if the wall thickness is lower than the nominal wall thickness, this is most likely due to corrosion damages. This kind of automated inspection yields far better coverage of quantitative inspection results, but is also time-consuming and sometimes costly, as production in some cases must be paused in order to let the ultrasonic pig run through the pipes.

Automatic inspection is the most usual when inspecting an extensive network of pipelines covering large geographical areas. The data set stems from pipes at one specific oil plant, and the measurements are made manually by human inspectors, although with measurement tools that yield a precise measurement of wall thickness at any given place.

It might seem intuitive to approach the problem of corrosion damage from a strictly chemical perspective. However, even though the chemical properties of a corrosion pro-

cess is well known in theory, the practical case is often more complicated. Due to the chaotic nature of the problem, it can be challenging to measure the amount of water, oxygen and other vital chemicals that are present at any given time at any given point in the pipe. This might be due to subtle human errors in the construction of the pipe, fluctuating temperatures or unpredictable turbulence in the flow of the fluids in the pipe. It is difficult to evaluate the corrosion process form a chemical perspective. For these reasons, the problem is attacked from a statistical perspective in this work, where we try to build statistical models to predict and analyze corrosion progression. These models can be used to plan a predictive maintenance scheme.

1.2 Predictive Maintenance and Risk Based Inspection

As soon as we have a statistical model for the corrosion progression, we can start to investigate the following questions:

- What is the **Remaining Useful Life (RUL)** of a pipe?
- What is the probability that a pipe reaches its **Minimum Allowed Wall Thickness (MAWT)** within the next k time units (e.g. months)?
- Is it possible to introduce some methodology that classifies each pipe into for instance "High risk" (of reaching a critical wall thickness) and "Low risk"? Other more nuanced categories might also be used.

The answer to these questions would yield a good starting point for planning a **Risk Based Inspection (RBI)** scheme. DNV (2009) defines RBI as an "[...] approach designed to aid the development of optimized inspection, and recommendations for monitoring and testing plans for production systems". In other words, an RBI strategy aims to monitor and inspect the pipes only when the risk of leakage is high.

The opposing inspection strategy is to inspect with a fixed frequency, i.e. strategies where every pipe is inspected every n-th month, regardless of any statistical predictions. For instance, one can inspect every pipe every month. This routine would surely give us much information about all the pipes, and likely give a small probability of failure. However, some practicalities hinder us from doing this. Every inspection comes with a cost – practically, time-wise and financially – and we must therefore choose carefully when and where we want to inspect. The most practical solution to the problem of high inspection costs would be to never measure, but this would potentially yield huge damages, whose costs are far higher than that of inspections. Consequently, we seek to find an inspection strategy that minimizes the chance that the pipes "fail" (for instance, reach a critical threshold), while simultaneously punishes strategies with a huge amount of inspections.

To plan an RBI strategy, we must have some a tool to measure risk. In this paper we introduce a time-dependent **random effects model** to model corrosion damage – referred to as **Loss of Material** – as a function of time and various features of the pipe. We proceed to update the model in a Bayesian fashion as new data are gathered. With this model, we have a framework to quantify risk, and can test different inspection strategies to see how they perform.

Loss of Material for a pipe is measured repeatedly over a period of many years. The difference between consecutive measurements is denoted the **corrosion rate**. For each pipe, we know the MAWT, which depends on several factors related to the material of the pipe, the pressure it endures and other factors. As the wall thickness of a pipe goes towards MAWT, we should consider to repair it. If we fail to repair it before it reaches MAWT we say that an event occurs. An event typically means leakage of hydrocarbons.

When trying to make a model for corrosion damage, the following assumptions are imposed:

- **The corrosion rate is not constant.** Predicting when we reach MAWT would be simple if we knew for certain that the corrosion rate always was constant. However, we assume that this is not the case and that the rate might vary over time due to changes in the production and other unknown factors.
- **Different pipe types have different decay rates.** Because of the difference in material, insulation and the formation of the pipe, the decay rates will vary between different pipe types.
- **Measurement noise is present.** The tools that measure the wall thickness are not perfect, and neither are the human operators that use them.

Figure 1.1 illustrates a simple example. Here we know the baseline wall thickness, i.e. the wall thickness at the installation date (denoted "Time 0" in the figure). In addition to this, three inspections are performed at Time 1, Time 2 and Time 3. The third inspection is above MAWT. Time has since passed without any new inspections being conducted. We want to model the probability that the wall thickness is less than MAWT.

1.3 Structure of Thesis

In Chapter 2 we present a data set of inspection results from a non-disclosed production site. This data is provided by Oceaneering. [1] In Chapter 3 we present some important ideas from Bayesian statistics and use this to derive the Gaussian Time-Dependent Random Effect model that will be used to model Loss of Material. In Chapter 4 we use this model as a building block when we derive an algorithm for adaptive monitoring. We also introduce a new inspection criterion, that monitors the pipes with the highest expected Value of Information. Finally, in Chapter 5 we test different inspection strategies on the data set and compare various metrics. Chapter 6 contains key findings and some ideas for further work.

[1] Oceaneering AS, Trondheim, is owned by Oceaneering International, Inc., Houston USA – a global subsea company.

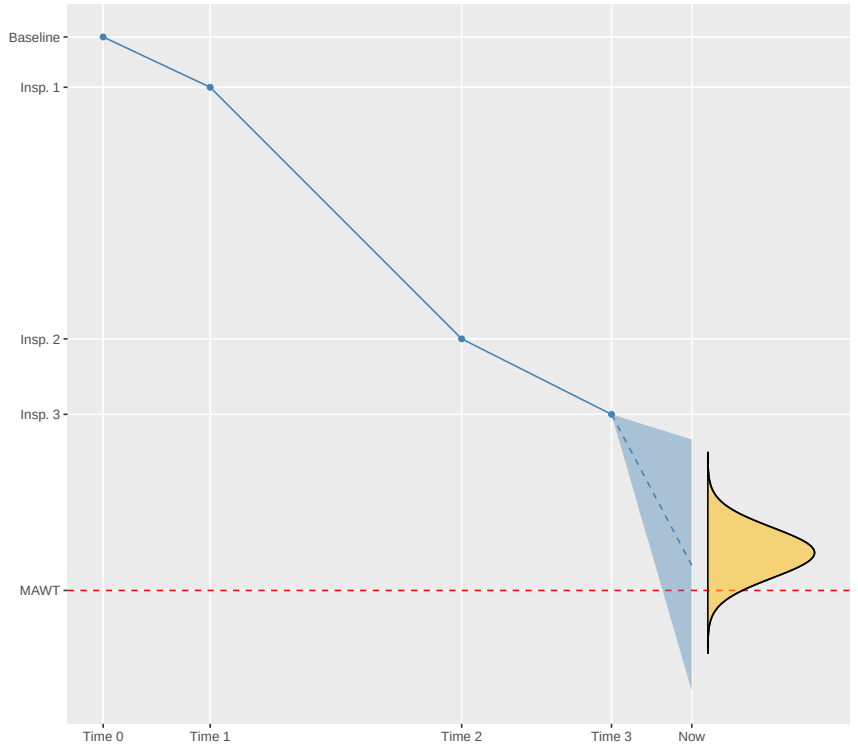

Figure 1.1: A conceptual illustration of the problem. We have done three inspections at Time 1, Time 2 and Time 3. In addition we know the baseline wall thickness at Time 0 and the MAWT, marked by a dashed red line. In none of the previous three inspections we have detected a wall thickness less than MAWT. However, some time has passed since last inspection at Time 3, and we use a statistical model to help us to decide if we should inspect again. The dashed blue line and the surrounding shaded area indicate the predicted wall thickness with a confidence band. Even though the prediction lies above MAWT, there is a probability that the actual wall thickness is below MAWT. Based on this probability – and our tolerance – we might choose to inspect. The prediction is normally distributed around the mean, as indicated by the yellow Gaussian distribution. The area of the Gaussian distribution that is under MAWT illustrates the probability that the unobserved wall thickness is below MAWT, according to the model.

Chapter 2

The Oceaneering Data Set

2.1 About the Data

Oceaneering provides a data set consisting of measurements from an oil refinery at an undisclosed geographic place, operated by an undisclosed petrochemical company. The data set consists of repeated measurements of wall thickness at multiple pipe locations. Human inspectors make the measurements with an ultrasonic measurement tool. There are $N = 1019$ distinct pipes in the data set, and a total of $k = 4$ measurements of wall thickness are conducted for each pipe at different times. At the request of Oceaneering, we have done some slight modifications to the data set to avoid disclosing confidential information. These modifications include adding Gaussian noise $\mathcal{N}(\mu = 0, \sigma^2 = 0.1^2)$ to the inspection data, as well as changing the names of variables, categories and materials. These techniques for working with private data are recommended by Mivule (2012).

2.2 An Overview

An overview of the data set can be seen in Table 2.1. Each data point consists of two types of data: Inspection data, i.e. measurements of wall thickness with corresponding time stamps, and pipe data, i.e. data about various features, insulation and material for the inspected pipe.

Figure 2.1 shows the corrosion damage (in mm) plotted against time since installation (months). As a convention we denote the corrosion damage as negative values, as this indicates that material is lost. Thus, corrosion damage of for instance -4 mm indicates that the wall thickness is 4 mm less than when the pipe was installed.

From Figure 2.1 we can observe the following:

- There seems to be a declining trend: The corrosion damage becomes larger (in absolute terms) over time. However, many observations still have close to zero corrosion damage even after many years. Generally, the variance seems to increase with time.

Covariate name	Description	Unit
`installation`	The exact date when the pipe was installed	dd/mm/yyyy
`time1`	Time until the first inspection	months
`time2`	Time until the second inspection	months
`time3`	Time until the third inspection	months
`baseline`	Wall thickness at installation date	mm
`insp1`	Wall thickness at inspection 1	mm
`insp2`	Wall thickness at inspection 2	mm
`insp3`	Wall thickness at inspection 3	mm
`feature`	Description of any specific feature at the pipe	-
`mawt`	Minimum allowed wall thickness	mm
`material`	Material of the pipe	-
`insulation`	Is the pipe insulated?	{Yes, No, Unknown}
`circuit`	ID of the corrosion circuit	-
`inch`	Diameter of the pipe	mm

Table 2.1: Covariates with description in the Oceaneering data set. The covariates above the dashed line are the ones used in our model.

- The oldest pipe in the data set is measured at 192 months – exactly 16 years – after installation. This seems to be one of a few outliers, however, as most of the observations are made up until approximately 150 months after installation. It might be difficult to make precise predictions about the wall thickness after month 150 because of the lack of data.

- After around 100 months there seems to be a slight decrease in variance, and many observations are concentrated around lower (absolute) corrosion damage. One might say that this is intuitively a bit surprising: How can material be gained? The corrosion process cannot be reversed. However, one possible explanation is that the pipes get clogged, i.e. that various materials (chemicals, sand, byproducts from the production, etc.) build up on the inside of the pipe. As we only have a measurement of the wall thickness, it is hard to know what exactly that causes this slightly surprising effect. Another possibility is that preventive actions are being taken from the manager of the facility, for instance treating the pipes with various chemicals or coatings in an attempt to hinder the damage.

In Figure 2.2 we plot the Loss of Material against time *since previous inspection*. This has a shorter time axis than Figure 2.1; the longest time between two inspections on one pipe is 111 months (9 years and 3 months). Apart from this and a couple of other outliers around 105 months, most inspections are done within 70 months after previous inspections. It also seems that many inspections are conducted in the same months – the data points seem to lie in a vertical row – which might be due to purely practical reasons. If we intended to inspect one pipe in one month and the neighboring pipe the month after, it is more practical to inspect both at the same time. Thus, we get these clusters with many inspections in the same month. Apart from this, the trend in Figure 2.2 seems to follow the same pattern as in Figure 2.1 with an increasing variance over time.

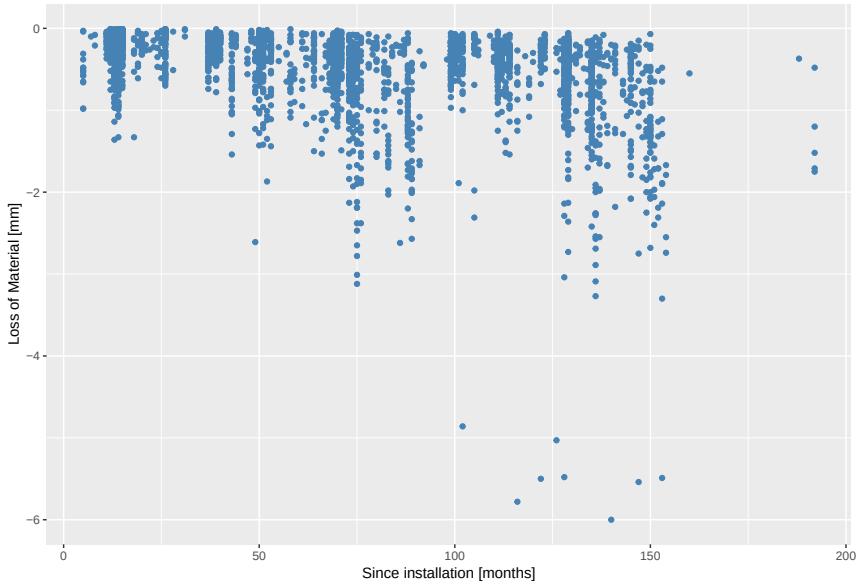

Figure 2.1: Accumulated corrosion (in mm) plotted against time since installation (in months). There is a total of 3057 data points; $N = 1019$ pipes that are each measured 3 times, as we exclude the initial measurement done at installation (we assume no corrosion damage at installation).

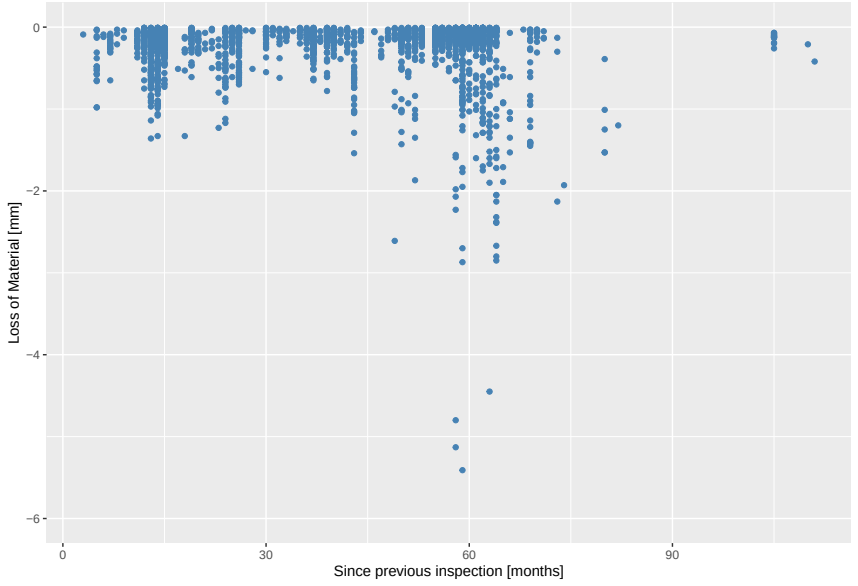

Figure 2.2: Accumulated corrosion (in mm) plotted against time since last inspection (in months). These are the same 3057 data points as in Figure 2.1, but with a different time axis.

The average monthly corrosion rate is -0.0091 mm, which translates to -0.1092 mm per year. The largest monthly rate is -0.196 mm per month, translating to a yearly Loss of Material of -2.352 mm.

2.3 Features, Insulation and Pipe Material

Every pipe is categorized with one of six distinct features, or pipe formations:

- **Drain**: A drain valve.
- **Elbow**: A 90 degree angle on the pipe.
- **Nozzle**: A specific shape of the pipe designed to increase the velocity of the fluids (e.g. hydrocarbons) that is flowing through it.
- **Piping**: An ordinary, straight piece of pipe.
- **Tee**: A "T" shape, i.e. one pipe that is orthogonal to another.
- **Other**: Other features or abnormal pipe shapes that do not fit in the other categories.

These features are of interest because we know empirically that the shape or formation of the pipe is important for the development of corrosion. This might be because different pipe shapes give different flow of the hydrocarbons through the pipe. In Figure 2.3 we plot corrosion damage against time for each of the 6 features. It is clear that most pipes are in the "Piping" and "Elbow" category (53% and 23%, respectively), while "Drain" and "Nozzle" both have around 3% of the observations each. Every feature seems to follow the same trend that we see in Figure 2.1, i.e. a generally decreasing trend, However, the "Drain" feature appears to stand out with all inspections yielding a very small Loss of Material; all observations are higher than -1 mm. This could probably be due to the small sample size.

Another important feature of the data set is the Minimum Allowed Wall Thickness (MAWT). If the wall thickness falls below MAWT we are experiencing an event (a leakage). MAWT is calculated based on various parameters about the fluids that will flow through it. Examples of these parameters are pressure, temperature and diameter of the pipe. These calculations are based on the American Society of Mechanical Engineers' standards.[1] The calculations are conducted by Oceaneering. In Figure 2.4 we have plotted MAWT against the baseline wall thickness, i.e. the wall thickness at installation. Both measurements are in millimeters. The plot gives an indication about the size of the pipes; with a very few exceptions, all pipes have a baseline wall thickness of less than 25 mm. The average wall thickness is just under 14 mm. At the same time, most MAWT values are well below 10 mm. The **Accepted Loss of Material** is the MAWT subtracted from the baseline wall thickness.

Additionally, the data set consists of some covariates that we will not use in the model. One of these covariates is "Material". The pipes are divided into three categories in terms

[1] The American Society of Mechanical Engineers' B31.3 2012 Process piping (304.1.2) standard.

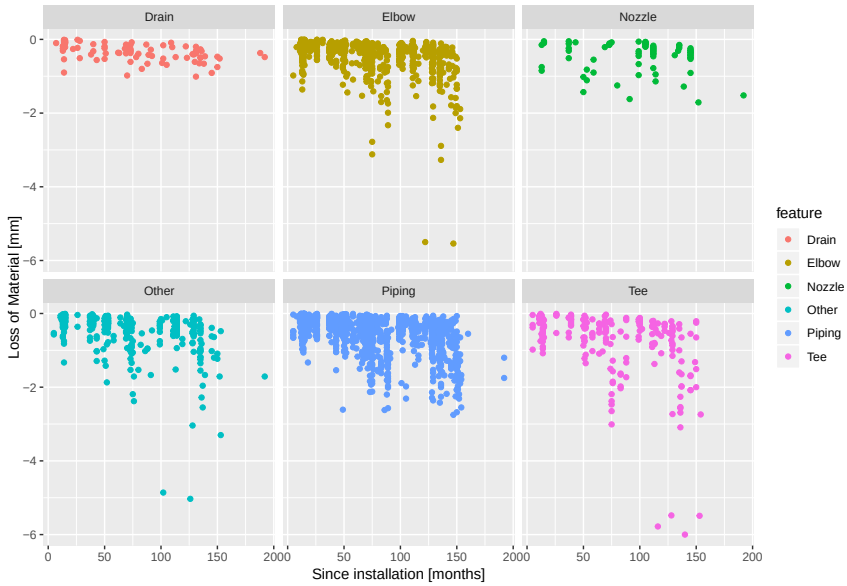

Figure 2.3: Accumulated corrosion damage plotted against time for various features

of material: Carbon Steel (”CS”), P11 (a composite steel material) and ”Other”. Additionally, every pipe is equipped with either ”Insulation”, ”No insulation” or ”Unknown”. Multiple pipes are also grouped together in a corrosion circuit, which are common paths in the pipe network where the same type of fluids often is transferred through.

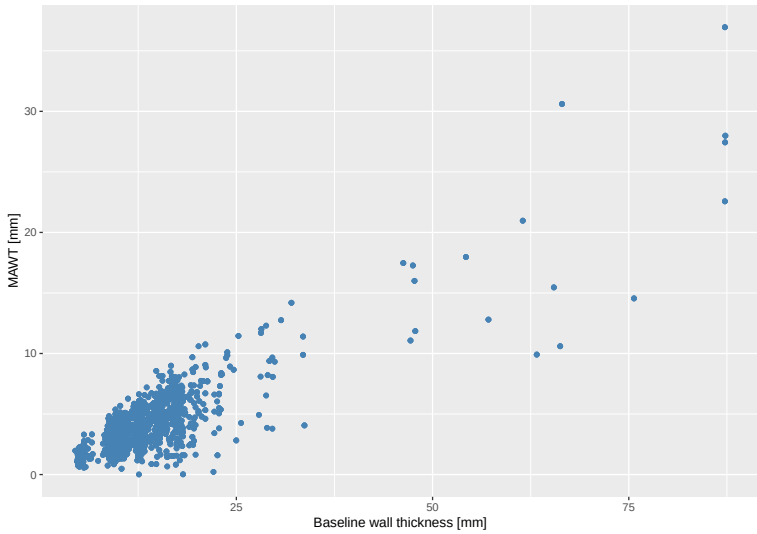

Figure 2.4: Minimum Allowed Wall Thickness (MAWT) plotted against baseline wall thickness.

Chapter 3

Statistical Framework

In this chapter we derive a statistical model to predict the Loss of Material in a given pipe at a given time. Denote this $y_i(t)$, where t is the time and i is the pipe number. The data set presented in Chapter 2 contains monthly data, and we will therefore use a monthly time grid in the model, so that time $t = 1$ indicates month one after installation, $t = 2$ indicates month two after installation etc. Note that the installation time might be different for the various pipes, but we nevertheless use the installation time as a starting point for every pipe.

We measure Loss of Material in millimeters (mm) with a negative sign. For instance, $y_i(t = 100) = -2$ means that pipe i has experienced an accumulated loss of 2 millimeters from installation up to and including month 100. We assume that the Loss of Material at installation is zero for every pipe, $y_i(0) = 0$ for every i.

3.1 Basic Model

If we assumed that the corrosion rate was constant in time, we could formulate a simple model of the type

$$y_i(t) = \beta t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, \tau^2), \tau \in \mathbb{R}, \tag{3.1}$$

where β is a constant damage rate (i.e. damage per time unit). Additionally we want to capture the effects that come from different features in the data set. Assuming p different features we have effects $\lambda_j, j = 1, ..., p$. We write

$$y_i(t) = (\beta + \boldsymbol{f}_i^\top \boldsymbol{\lambda})t + \epsilon_{it}, \tag{3.2}$$

where $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_p]^\top$ is the size p vector of different features and $\boldsymbol{f}_i^\top = [0, ..., 1, ..., 0]$ is a binary indicator vector that marks which of the p features that is present in pipe i.

As seen in Chapter 2, however, there are indications that the rate is not necessarily constant in time. For this reason we introduce a separate rate for each month. Let β_m be

the corrosion rate (in mm) in month m, and

$$\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_k]^\top \in \mathbb{R}^k \tag{3.3}$$

where k is the maximum number of months in the data set. Now, the damage of the pipe at time t is

$$\begin{aligned} y_i(t) &= \sum_{m=1}^{t} \left(\beta_m + \boldsymbol{f}_i^\top \boldsymbol{\lambda}\right) + \epsilon_{it} \\ &= \left(\sum_{m=1}^{t} \beta_m\right) + \boldsymbol{f}_i^\top \boldsymbol{\lambda} t + \epsilon_{it}, \end{aligned} \tag{3.4}$$

i.e. an accumulation of all the monthly damage rates up until month t, as well as the time effect for the specific category. The first part of the expression can be formulated via a design matrix of 0's and 1's:

$$\begin{aligned} \sum_{m=1}^{t} \beta_m &= [1, 1, ..., 1, 0, , ..., 0] \cdot [\beta_1, \beta_2, ..., \beta_t, \beta_{t+1}, ..., \beta_k]^\top \\ &= \boldsymbol{x}_i^\top \boldsymbol{\beta}. \end{aligned} \tag{3.5}$$

Thus, we arrive at the following model for $y_i(t)$:

$$y_i(t) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{f}_i^\top \boldsymbol{\lambda} t + \epsilon_{it}. \tag{3.6}$$

The above framework can be generalized to the case of N pipes. We can introduce the vectors $\boldsymbol{y} = [y_1, y_2, ..., y_N]^\top$ and $\boldsymbol{t} = [t_1, t_2, ..., t_N]^\top$ and formulate the joint model as

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{\lambda} + \boldsymbol{\epsilon}_{it}, \tag{3.7}$$

where $Z \in \mathbb{R}^{N \times p}$ is such that row i is

$$\boldsymbol{z}_i = \boldsymbol{f}_i t_i, \quad i = 1, ..., N, \tag{3.8}$$

and $X \in \mathbb{R}^{N \times k}$ is a design matrix where each row is as described in (3.5). The explicit matrix expressions for (3.7) is

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ & & \ddots & \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ & & \ddots & \\ z_{N1} & z_{N2} & \dots & z_{Np} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}. \tag{3.9}$$

In this model we have one effect of time, which is common for all feature types. This is incorporated in the $\boldsymbol{\beta}$ coefficients. Additionally, we have one effect of feature time, which is common for all the pipes that exhibit this feature. We can introduce the augmented notation

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{bmatrix}, \tag{3.10}$$

and

$$Q = \begin{bmatrix} X & Z \end{bmatrix} \in \mathbb{R}^{(k+p)\times N}, \tag{3.11}$$

such that

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{\lambda} + \boldsymbol{\epsilon} = Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \tag{3.12}$$

This will ensure a more compact notation.

3.2 The Bayesian Framework

We will now consider $\boldsymbol{\gamma}$ to be a random variable with a distribution. We cannot observe $\boldsymbol{\gamma}$ directly, but we can make observations of $\boldsymbol{y}$ and use that to assess the belief about $\boldsymbol{\gamma}$. To do this in a structured way, we apply a Bayesian framework. In this framework we aim to adjust the distribution of $\boldsymbol{\gamma}$ after observing new data $\boldsymbol{y}$. That is, we want to find a distribution $p(\boldsymbol{\gamma}|\boldsymbol{y})$ and do so by applying Bayes' rule:

$$p(\boldsymbol{\gamma}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\gamma}) \cdot p(\boldsymbol{\gamma})}{p(\boldsymbol{y})}. \tag{3.13}$$

The components of this equation can be explained as follows:

- $p(\boldsymbol{\gamma})$ is the **prior** distribution over the coefficients. This distribution incorporates any belief we might have – or the lack of any such belief – about the distribution before observing the data.
- $p(\boldsymbol{\gamma}|\boldsymbol{y})$ is the **posterior** distribution of the coefficients; the distribution *after* we have observed some data.
- $p(\boldsymbol{y}|\boldsymbol{\gamma})$ is the **likelihood**, as defined in equation (3.12); the probability of observing what we did (i.e. the observations $\boldsymbol{y}$) given a probability distribution for $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau^2 I_N)$.
- $p(\boldsymbol{y})$ is a **normalization** factor, also known as the **evidence**.

Omitting the normalization factor, we can re-state equation (3.13) as simply

$$p(\boldsymbol{\gamma}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\gamma}) \cdot p(\boldsymbol{\gamma}). \tag{3.14}$$

or, informally,

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}. \tag{3.15}$$

These equations often form the framework in an iterative scheme: We use a prior for $\boldsymbol{\gamma}$, calculate the posterior $\boldsymbol{\gamma}|\boldsymbol{y}_1$ after observing some data $\boldsymbol{y}_1$. Then we can in turn use $\boldsymbol{\gamma}|\boldsymbol{y}_1$ as a *prior* and update the belief once again with some new observations $\boldsymbol{y}_2$. This iterative scheme is depicted in Figure 3.1. We assume some known prior distributions

$$\begin{aligned} &\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_\beta) \\ &\boldsymbol{\lambda} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \Sigma_\lambda), \quad \Sigma_\lambda = \sigma_\lambda^2 I_p, \quad \sigma_\lambda^2 \in \mathbb{R} \end{aligned} \tag{3.16}$$

where I_p is the identity matrix of size p. We assume independence between $\boldsymbol{\beta}, \boldsymbol{\lambda}$ and $\boldsymbol{\epsilon}$. Since the model parameters now are equipped with a distribution, we have a **random effects** model

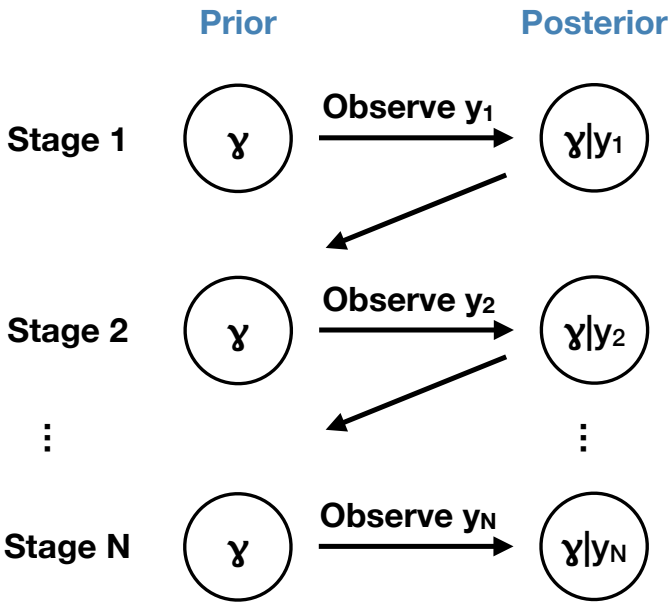

Figure 3.1: An illustration of the iterative process of prior and posterior distributions. After observing some data $\boldsymbol{y}_1$ we obtain a posterior distribution $\boldsymbol{\gamma}|\boldsymbol{y}_1$. This is then used as a prior before observing some new data $\boldsymbol{y}_2$. This procedure is repeated. We have written "Stage" rather than "Time" because we do not necessarily update our models after every time step.

3.3 Deriving Distributions

In this section we will derive the joint, conditional and marginal $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu_y}, \Sigma_y)$. To make notation simpler we again construct $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\mu_\gamma}, \Sigma_\gamma)$, with

$$\boldsymbol{\mu_\gamma} = \begin{bmatrix} \boldsymbol{\mu_\beta} \\ \boldsymbol{\mu_\lambda} \end{bmatrix}, \quad \Sigma_\gamma = \begin{bmatrix} \Sigma_\beta & \mathbf{0} \\ \mathbf{0} & \Sigma_\lambda \end{bmatrix}, \tag{3.17}$$

where we have assumed independence between $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. With $\boldsymbol{\mu_\gamma}$ and Σ_γ assumed to be known we can calculate

$$\begin{aligned} \boldsymbol{\mu_y} = \mathbb{E}(\boldsymbol{y}) &= \mathbb{E}(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}) \\ &= Q\mathbb{E}(\boldsymbol{\gamma}) + \mathbb{E}(\boldsymbol{\epsilon}) \\ &= Q\boldsymbol{\mu_\gamma}, \end{aligned} \tag{3.18}$$

and

$$\begin{aligned}\Sigma_y &= \text{Var}\left(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}\right) \\ &= \text{Var}(Q\boldsymbol{\gamma}) + \text{Var}(\boldsymbol{\epsilon}) \\ &= Q\Sigma_\gamma Q^\top + \tau^2 I_N. \end{aligned} \tag{3.19}$$

3.3.1 Joint Distributions

In summary, the marginal distribution for $\boldsymbol{y}$ is

$$\begin{aligned}\boldsymbol{y} &\sim \mathcal{N}\left(\boldsymbol{\mu_y}, \Sigma_y\right) \\ &= \mathcal{N}\left(Q\boldsymbol{\mu_\gamma}, Q\Sigma_\gamma Q^\top + \tau^2 I_N\right). \end{aligned} \tag{3.20}$$

With the independence between $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ it can be verified that this is the same as

$$\boldsymbol{y} \sim \mathcal{N}\left(X\boldsymbol{\mu_\beta} + Z\boldsymbol{\mu_\lambda}, X\Sigma_\beta X^\top + Z(\sigma_\lambda^2 I_p)Z^\top + \tau^2 I_N\right). \tag{3.21}$$

We might later be interested in predicting the value of $\boldsymbol{y}_o \in \mathbb{R}^{N_o}$ at unobserved locations corresponding to a different design matrix Q_o. Not surprisingly, the distribution of $\boldsymbol{y}_o$ is pretty similar to (3.20),

$$\begin{aligned}\boldsymbol{y}_o &\sim \mathcal{N}\left(\boldsymbol{\mu_{y_o}}, \Sigma_{y_o}\right) \\ &= \mathcal{N}\left(Q_o\boldsymbol{\mu_\gamma}, Q_o\Sigma_\gamma Q_o^\top + \tau^2 I_{N_o}\right). \end{aligned} \tag{3.22}$$

Assume we have a Gaussian prior distribution $p(\boldsymbol{\gamma}) \sim \mathcal{N}(\boldsymbol{\mu_\gamma}, \Sigma_\gamma)$ with known parameters. That is,

$$p(\boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\mu_\gamma})^\top \Sigma_\gamma^{-1}(\boldsymbol{\gamma} - \boldsymbol{\mu_\gamma})\right). \tag{3.23}$$

Using this, we can write the joint distribution as

$$p\begin{pmatrix}\boldsymbol{\gamma} \\ \boldsymbol{y} \\ \boldsymbol{y}_o\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\mu_\gamma} \\ \boldsymbol{\mu_y} \\ \boldsymbol{\mu_{y_o}}\end{pmatrix}, \begin{bmatrix}\Sigma_\gamma & \Sigma_\gamma Q^\top & \Sigma_\gamma Q^\top \\ Q\Sigma_\gamma & \Sigma_y & Q\Sigma_\gamma Q_o^\top \\ Q_0\Sigma_\gamma & Q_o\Sigma_\gamma Q^\top & \Sigma_{y_o}\end{bmatrix}\right), \tag{3.24}$$

where we have utilized that

$$\begin{aligned}\text{Cov}(\boldsymbol{y}, \boldsymbol{\gamma}) &= \text{Cov}(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \boldsymbol{\gamma}) \\ &= Q\text{Cov}(\boldsymbol{\gamma}, \boldsymbol{\gamma}) + \text{Cov}(Q\boldsymbol{\gamma}, \boldsymbol{\epsilon}) \\ &= Q\text{Var}(\boldsymbol{\gamma}) + 0 \\ &= Q\Sigma_\gamma, \end{aligned}$$

and

$$\begin{aligned}\text{Cov}(\boldsymbol{y}, \boldsymbol{y}_o) &= \text{Cov}(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}, Q_o\boldsymbol{\gamma} + \boldsymbol{\epsilon}_o) \\ &= \text{Cov}(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}, Q_0\boldsymbol{\gamma}) + \text{Cov}(Q\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon}_o) \\ &= \text{Cov}(Q\boldsymbol{\gamma}, Q_o\boldsymbol{\gamma}) + \text{Cov}(\boldsymbol{\epsilon}, Q_o\boldsymbol{\gamma}) + \text{Cov}(Q\boldsymbol{\gamma}, \boldsymbol{\epsilon}_o) + \text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_o) \\ &= Q\Sigma_\gamma Q_o. \end{aligned} \tag{3.25}$$

This will come in as a handy notation multiple times, especially given the following proposition:

Proposition 1. *Let* $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2]^\top$ *be multivariate normally distributed, i.e.* $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *with*

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Now, the conditional distribution $\mathbf{w}_1 | \mathbf{w}_2 \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\Sigma})$ *with*

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{w}_2 - \boldsymbol{\mu}_2) \\ \bar{\Sigma} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

(Proposition ended.)

The derivation of Proposition 1 can be found in Anderson (2003) and Johnson and Wichern (2007).

3.3.2 The Posterior Distribution

Assume now that we have a prior distribution for $\boldsymbol{\gamma}$ and make an observation $\boldsymbol{y}_{obs}$. The observation $\boldsymbol{y}_{obs}$ is thus not a random variable, but rather one realization of the random variable $\boldsymbol{y}$. Utilizing the formulas from Proposition 1 and the joint distribution (3.24) the posterior distribution is straight forward to compute $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{y}}, \Sigma_{\boldsymbol{\gamma}|\boldsymbol{y}})$:

$$\begin{aligned} \boldsymbol{\gamma} | \boldsymbol{y}_{obs} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{y}} &= \boldsymbol{\mu}_{\boldsymbol{\gamma}} + \Sigma_\gamma Q^\top \Sigma_y^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{\boldsymbol{y}}), \\ \Sigma_{\boldsymbol{\gamma}|\boldsymbol{y}} &= \Sigma_\gamma - \Sigma_\gamma Q^\top \Sigma_y^{-1} Q \Sigma_\gamma), \end{aligned} \tag{3.26}$$

where $\boldsymbol{\mu}_{\boldsymbol{y}}$ and Σ_y are as defined in (3.20). One peculiar observation is that the posterior covariance matrix $\Sigma_{\boldsymbol{\gamma}|\boldsymbol{y}}$ does not depend on the observed values $\boldsymbol{y}_{obs}$, but only the prior covariance matrix Σ_γ, the design matrix Q and the variance of the Gaussian noise, τ^2. The expected value $\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{y}}$, on the other hand, is dependent on $\boldsymbol{y}_{obs}$. From (3.26) we see that if $\boldsymbol{y}_{obs} = \mathbb{E}(\boldsymbol{y}) = Q\boldsymbol{\mu}_{\boldsymbol{\gamma}}$, then the posterior mean is simply equal to the prior mean, $\boldsymbol{\mu}_{\boldsymbol{\gamma}|\boldsymbol{y}} = \boldsymbol{\mu}_{\boldsymbol{\gamma}}$.

3.3.3 Bayesian Prediction

In any statistical model that aims to predict anything it is of interest to investigate how precise the predictions are. This can typically be done by holding back a part of the data set – the *test* set or *holdout set* – and using the rest – the *training* set – to train your model. A systematic way of repeatedly partitioning the data set into testing and training set in order to avoid overfitting the model, is a procedure known as cross-validation. This is described in chapter 3.4 of Fahrmeir et al. (2013).

In Section 3.3 we described the linear prediction of an unknown data point $\boldsymbol{y}_o$. A slightly more interesting and relevant situation occurs when we want to use the training set

$\boldsymbol{y}_{obs} \in \mathbb{R}^N$ to predict the unknown values $\boldsymbol{y}_o \in \mathbb{R}^{N_o}$ in the test set. Applying the formulas from Proposition 1, the conditional predictive density becomes

$$\begin{aligned}\boldsymbol{y}_o|\boldsymbol{y}_{obs} \sim \mathcal{N}(&Q_o\boldsymbol{\mu_\gamma} + Q_o\Sigma_\gamma Q^\top(Q\Sigma_\gamma Q^\top + \tau^2 I_N)^{-1}(\boldsymbol{y}_{obs} - Q\boldsymbol{\mu_\gamma}),\\ &Q_o\Sigma_\gamma Q_o^\top + \tau^2 I_{N_o} - Q_o\Sigma_\gamma Q^\top(Q\Sigma_\gamma Q^\top + \tau^2 I_N)^{-1}Q\Sigma_\gamma Q_o^\top)\\ &:\sim (\boldsymbol{\mu_{y_o|y}}, \Sigma_{\boldsymbol{y}_o|\boldsymbol{y}}),\end{aligned} \tag{3.27}$$

where Q_o is the design matrix corresponding to $\boldsymbol{y}_o$.

3.4 Gaussian Process

In Section 3.2 we saw how we could update the probability distributions after observing new data, and we introduced the concept of priors and posteriors. We will now introduce a useful framework for defining a covariance matrix to the prior distribution of $\boldsymbol{\gamma}$.

A Gaussian Process (GP) is well suited for regression. Rasmussen (2004) defines a GP as a collection of random variables with a joint Gaussian distribution. Let $\boldsymbol{\beta}$ be defined on the discretized time grid $\boldsymbol{t} = [t_1, t_2, ..., t_k]$ with corresponding function values $[\beta_1, \beta_2, ..., \beta_k]$. Moreover, a function $\boldsymbol{\beta}$ that is distributed as a GP is denoted

$$\boldsymbol{\beta} \sim \mathcal{GP}(\boldsymbol{\mu_\beta}, \Sigma_\beta),$$

where $\boldsymbol{\mu_\beta}$ is its mean function and Σ_β is its covariance matrix. While sampling from a distribution $\mathcal{N}(\cdot, \cdot)$ in $\mathbb{R}$ yields a single *value*, sampling from $\mathcal{GP}(\boldsymbol{\mu_\beta}, \Sigma_\beta)$ yields a *function*. In other words, a GP is a distribution over functions. The discretized $\boldsymbol{\mu_\beta}$ can be written as

$$\boldsymbol{\mu_\beta} = [\mu_\beta(t_1), \mu_\beta(t_2), ..., \mu_\beta(t_k)]^\top \in \mathbb{R}^k$$

and similarly $\Sigma_\beta \in \mathbb{R}^{k\times k}$. Thus, sampling from the GP yields $\boldsymbol{\beta} \in \mathbb{R}^k$, i.e. a set of k function values corresponding to the $\boldsymbol{t}$ grid. The power of GP's lie in the design of the covariance matrix, where we can encode different covariance patterns. One common example is to design Σ_β such that element (i, j) in the covariance matrix is

$$(\Sigma_\beta)_{i,j} = \sigma^2 \exp\left(-\phi|t_i - t_j|\right), \quad \sigma, \phi \in \mathbb{R}^+. \tag{3.28}$$

This covariance structure ensures that information is borrowed with other covariates that are close in time. If we make an observation in month i, the effect of this observation will not only affect the update of β_i, but also the surrounding coefficients.

In Figure 3.2 we have sampled 3 realizations of three different Gaussian processes with $\boldsymbol{\mu_\beta} = -t$, but with different covariance matrix. The first example has a covariance matrix $\Sigma_\beta = \sigma^2 I$, i.e. no covariance structure. The resulting samples are independent in time. The second and third examples have the covariance structure from (3.28), although with different values of ϕ ($\phi = 3$ and $\phi = 0.1$, respectively). The result indicates that samples with higher ϕ value yields curves with less oscillations. This also makes sense mathematically: From (3.28) it is clear that large values of ϕ yields a large negative value in the exponent, i.e. a low covariance between two observed values $\boldsymbol{\beta}(t_i)$ and $\boldsymbol{\beta}(t_j)$ for

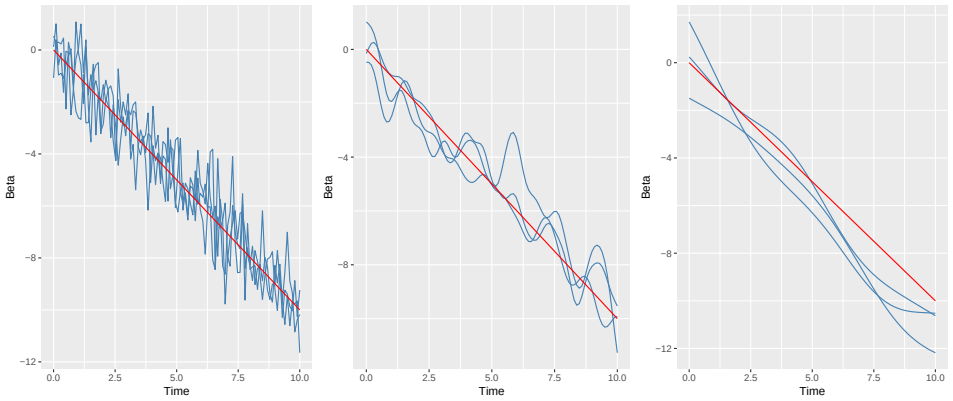

Figure 3.2: 3 realizations of a Gaussian process with same mean function $\boldsymbol{\mu_\beta}(t) = -t$, but different covariance matrix Σ_β. **Left:** $\Sigma_\beta = \sigma^2 I$. **Middle:** Σ_β as described in (3.28) with $\phi = 3$. **Right:** Σ_β as described in (3.28) with $\phi = 0.1$. In all three examples we have used $k = 100$ data points, $\sigma^2 = 1$.

$i \neq j$. In all of the three examples we have used $k = 100$ equidistant time points and $\sigma^2 = 1$.

We will utilize this idea and use a covariance structure as described in (3.28) in the prior distribution for $\boldsymbol{y}$.

3.5 Likelihood Function and Hyperparameters

In (3.20) we formulated a distribution for $\boldsymbol{y}$ with the hyperparameters σ_λ^2 and τ^2. In (3.28) we introduced the hyperparameters σ^2 and ϕ in the covariance matrix. For the sake of simplicity, let us denote the four hyperparameters as

$$\boldsymbol{\theta} = \begin{bmatrix} \sigma^2 \\ \phi \\ \sigma_\lambda^2 \\ \tau^2 \end{bmatrix}. \tag{3.29}$$

It is known that the likelihood of a multivariate normal distribution is

$$L(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_y|}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu_y})^\top \Sigma_y^{-1} (\boldsymbol{y} - \boldsymbol{\mu_y})\right) \tag{3.30}$$

and consequently the log-likelihood is

$$l(\boldsymbol{\theta}) = -\frac{k}{2} \log |\Sigma_y| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu_y})^\top \Sigma_y^{-1} (\boldsymbol{y} - \boldsymbol{\mu_y}), \tag{3.31}$$

where it is important to emphasize that $\Sigma_y = \Sigma_y(\boldsymbol{\theta})$, as derived in (3.20). We are seeking the Maximum Likelihood Estimators (MLE's) for $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}_{MLE}$. This is defined as

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg_{\boldsymbol{\theta}} \max l(\boldsymbol{\theta}). \tag{3.32}$$

3.5.1 Evaluating a Prediction

Assume now that we have made some predictions $\hat{\boldsymbol{y}} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_n]$ and want to compare them with the true values $\boldsymbol{y} = [y_1, y_2, ..., y_n]$ in the test set. One way of quantifying the prediction error is the **Root Mean Squared Error (RMSE)**,

$$\text{RMSE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}. \tag{3.33}$$

Another way is the **Continuous Ranked Probability Score (CRPS)**, which calculates a score for each prediction $\hat{y}_i$. The score is formulated as

$$\text{CRPS}_i(F_i, y_i) = -\int_{z=-\infty}^{\infty} \left(F_i(z) - \mathbb{I}\{z \geq y_i\}\right)^2 dz \tag{3.34}$$

where

- $F_i(z)$ is a cumulative probability distribution with mean $\mathbb{E}(\hat{y}_i)$ and a variance $\text{Var}(\hat{y}_i)$ from the predictive distribution;
- $\mathbb{I}\{z \geq y_i\}$ is an indicator function that has the value 1 if $z \geq y_i$ and 0 if $z < y_i$.

The integral is illustrated in Figure 3.3 for three different scenarios. The first scenario is a prediction with high accuracy and high precision. The second scenario has a high accuracy, but low precision. The third scenario has low accuracy and low precision. Note that (3.34) is only for a single prediction, and we can average the CRPS if we make more than one prediction. Gneiting and Raftery (2007) show that if the predictive distribution is Gaussian, i.e. if we let the predictive distribution be $\mathcal{N}\left(\mu, \sigma^2\right)$ and the true value is y, then (3.34) can be expressed in closed form as

$$\text{CRPS}\left(\mathcal{N}(\mu, \sigma^2), y\right) = \sigma\left[\frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{y-\mu}{\sigma}\right) - \frac{y-\mu}{\sigma}\left(2\Phi\left(\frac{y-\mu}{\sigma}\right) - 1\right)\right]. \tag{3.35}$$

3.6 Illustrative Example

We make $N = 1000$ observations, denoted $\boldsymbol{y} \in \mathbb{R}^{1000}$, of some damage phenomenon over a course of $k = 50$ time units (months). A large negative value indicates large damage. Each observation y_i is done at a time $t_i \in [1, 50]$. Half of the observations exhibit a specific feature, while the other half do not have this feature. For simplicity, we categorize the data as either "Feature" or "No feature". This could be one of the features introduced in Chapter 3, such as "Drain" or "Elbow". The data set can be seen in Figure 3.4. We see that most of the data from the "Feature" class lie somewhat lower than the data without the feature; it seems like there is a general, declining trend for all the data points, as well as a feature-specific effect for those who exhibit the feature.

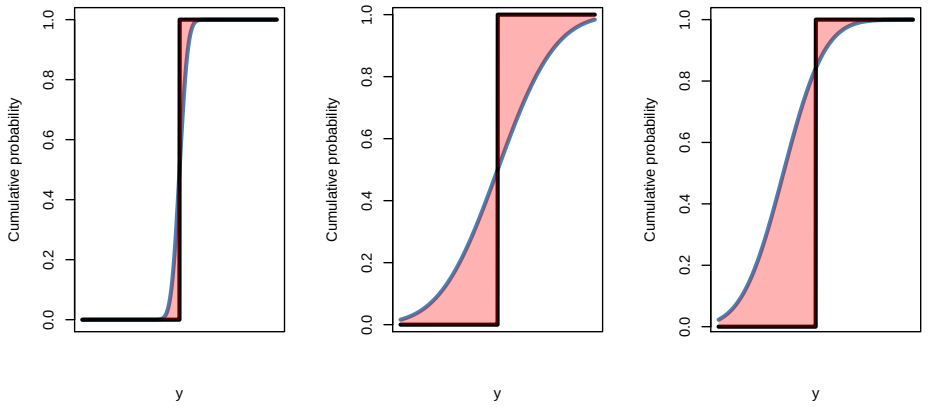

Figure 3.3: CRPS exemplified. The smooth, blue curve is a cumulative distribution. The black step function is the indicator function $\mathbb{I}\{...\}$. The colored area between the graphs is what we integrate over in (3.34) **Left:** The predicted mean lies close to the true value; hence the two curves cross close to where the cumulative probability is 0.5. Small variance indicates small area to integrate. **Middle:** The predicted mean is still close to the true value, but this time with a significantly larger variance. **Right:** The same variance, but less accurate, hence the higher intersection between the curves.

We want to model the relationship between damage and time on a discretized grid of $k = 50$ regression coefficients, denoted $\boldsymbol{\beta} = [\beta_1, ..., \beta_{50}]^\top$. More precisely, we model it as

$$\begin{aligned} y_i &= x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ik}\beta_k + z_i\lambda + \epsilon \\ &= \boldsymbol{x}_i\boldsymbol{\beta} + z_i \cdot \lambda + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \tau^2) \\ \implies \mathbf{y} &= X\boldsymbol{\beta} + Z\boldsymbol{\lambda} + \boldsymbol{\epsilon}. \end{aligned}$$

Here, $X \in \mathbb{R}^{1000\times 50}$ is a design matrix where each row consist of 1's up to position t_i and 0's afterward. $\boldsymbol{x}_i \in \mathbb{R}^k$ is row i in X. Similarly, Z is a design vector consisting of 0 for all the pipes from the "No feature" data, and t_i for all the pipes with the feature. Given that Z only has a single column, $p = 1$. We introduce flat priors of

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\beta}} &= [-0.05, ..., -0.05]^\top \\ \mu_\lambda &= -0.1 \end{aligned}$$

and the covariance structure introduced in equation (3.28) with $\phi = 0.5$ and $\sigma = 0.5$. Now, we can calculate the posterior distribution $\boldsymbol{\beta}|\boldsymbol{y}$ and $\lambda|\boldsymbol{y}$ using (3.26). The resulting posterior $\boldsymbol{\beta}$ can be seen in Figure 3.5, as well as the $\boldsymbol{\beta}^{true}$ – the coefficients used to generate the data – and the flat prior distribution. We can see that the posterior $\boldsymbol{\beta}$ is reasonably close to the real $\boldsymbol{\beta}^{true}$.

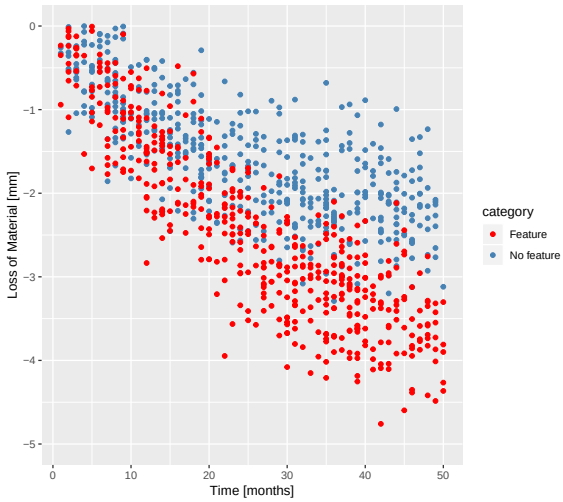

Figure 3.4: The data set $\boldsymbol{y}$, consisting of 500 data points from a class with some feature (red points) and 500 points from a class without the feature. The x axis show time in months and the y axis show Loss of Material in mm.

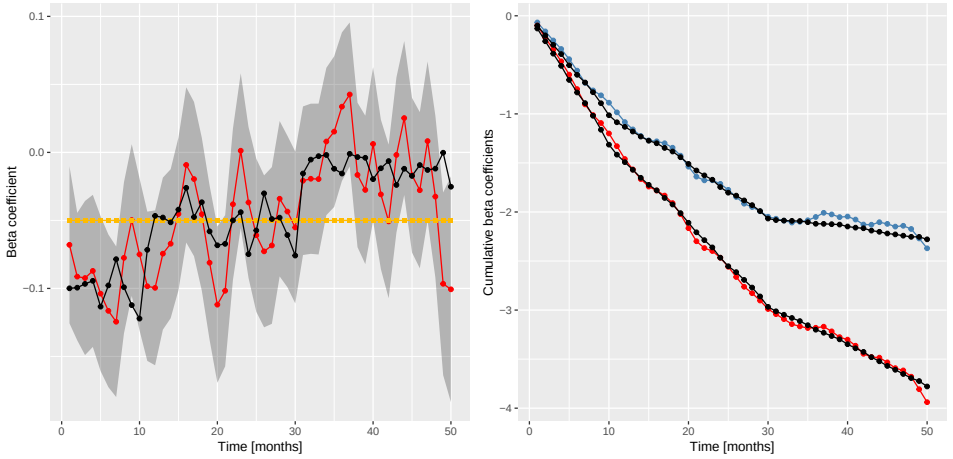

Figure 3.5: Left: The red line indicates the $\boldsymbol{\beta}$ posterior coefficients calculated from the data set. The grey area is a confidence band with $+/-$ one standard deviation, also using the posterior Σ_γ calculated from the data. The straight yellow line at $\beta = -0.05$ is our flat prior, and the black line is the true $\boldsymbol{\beta}^{true}$ used to generate the data. **Right:** Cumulative $\boldsymbol{\beta}$ coefficients for both the general trend without the random effect (blue), as well as with the random effect (red). The corresponding black lines show the true values used to generate the data.

Chapter 4

Where to Inspect?

In Chapter 3 we developed a statistical model to predict the wall thickness of a pipe at any given time. In this chapter we will use this model and all of its properties as building blocks when we derive various **inspection strategies**. First, we will talk about the general idea of an inspection strategy, before we move on to present the differences between an adaptive and a non-adaptive strategy. Finally, we present an inspection algorithm based on a **Value of Information (VOI)** approach.

An inspection strategy is a set of decision rules to plan and conduct inspections on a set of pipes. Which pipes should we inspect, and at what times? How shall we treat the information we gather from an inspection? Should we at any time stop inspecting? Are all inspections equally valuable, or can some kind of information turn out to be more *informative*, however we choose to define that, than others? At the same time there is a cost related to the collection of information: Inspections often come with a significant financial cost, but the cost of *not inspecting* might be worse.

4.1 What Constitutes an Inspection Strategy?

For a single pipe we have the following four scenarios at every time:

- **We do not inspect.** This yields no new information and has no cost.
- **We inspect.** This results in a new data point y_{obs}. An inspection has a cost.
- **We inspect and repair.** If the inspected wall thickness is dangerously close to MAWT, we might choose to preventively repair the pipe. This will set the wall thickness back to baseline wall thickness. A reparation has a cost that is larger than a simple inspection.
- **An event occurs.** An event is a pipe failure that we do not notice to notice before it is too late. In simple terms, this means that the pipe is leaking. This typically gets noticed without conducting a proper inspection, simply because parts of the pipe

network fails to function because of this event. An event has a cost that is higher than a repair.

We consider a repair and an event mutually exclusive: Either we detect the danger in time and repair the pipe and avoid any catastrophes, or we fail to notice the danger and the pipe experiences an event. Of course, from a practical point of view, the pipe that is leaking must be repaired somehow, but in this context we will talk about a repair as something that is done preventively, i.e. *before* we reach MAWT. Furthermore, we assume that a single pipe can *at most* experience one event or repair throughout its lifetime: If an event occurs, the pipe effectively gets taken out of use. If a repair occurs, the wall thickness gets reset to the baseline wall thickness, and we consider it highly unlikely that a pipe will go from baseline wall thickness to MAWT two times throughout its lifetime.

How do we value or give a score to an inspection strategy? The three parameters that we usually are interested in are

- the number of inspections;
- the number of repairs;
- the number of events.

Additionally, the number of **non-utilized months** is of interest. A conservative strategy for repair will be quick to repair a pipe, potentially missing out of many months that we *could* have utilized. In real life this is impossible to measure; if we repair a pipe, it is difficult to measure how many more months it *could* have been utilized. However, in simulations we might be able to measure this. How we choose to weigh these metrics typically depends on the costs related to each of the scenarios. These costs vary from facility to facility and from company to company. Additionally, the costs depend on what kind of chemical substance that is floating inside the pipes: Leakage of sulfuric acid, for instance, would be much more dramatic than leakage of hot water. Generally, we want to find the strategy that yields the overall lowest cost.

4.2 Non-adaptive Strategies

The simplest strategy for pipe inspection is to monitor the pipes at fixed time intervals, independently of any statistical model. A non-adaptive inspection strategy is easy to follow and both financially and practically very predictable. However, a non-adaptive strategy might in many cases be too simple, and lead to too many inspections in some situations and too few in other situations. Assume for instance that a pipe has an average lifetime of 480 months and we decide to follow a non-adaptive inspection strategy and inspect it every n-th month. What should n be? Should it be the same for every pipe? This would also lead to many unnecessary inspections in the first years of a pipe's lifetime, when we feel pretty confident that the pipe is well above MAWT. One could of course wait some years before starting to inspect the pipe, but this again yields the question: How long should we wait?

In the absence of precise statistical models, companies often tend to worst-case estimates to answer these questions (API (2009)). This might lead to too many inspections.

In Figure 4.1 we show an illustrative example of a non-adaptive inspection strategy where a pipe is inspected at a fixed time interval of 10 months. The figure shows the accumulated Loss of Material from installation and up to month 150. The accepted loss is at -0.85 mm. This is a synthetic example, where the true values and the inspection results are sampled from the distribution of $\boldsymbol{y}$. In month 142 the wall thickness falls below the accepted loss (i.e. the baseline wall thickness minus MAWT), but since we only inspect every 10th month we would not notice this before it was too late. This is a slightly exaggerated example: In a practical situation a Subject Matter Expert would probably make some considerations based on the inspection in month 140 and interfere. Figure 5.4 therefore serves as an illustrative example about *blindly* following a non-adaptive strategy.

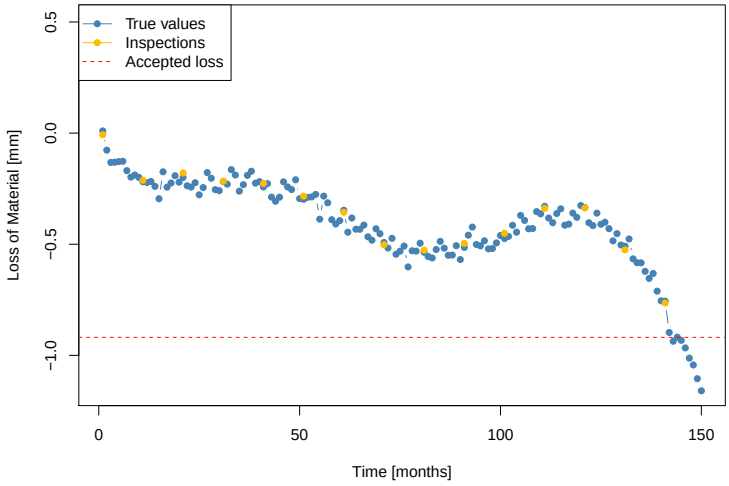

Figure 4.1: An example of a non-adaptive inspection strategy, where we choose to inspect the pipes every 10th month. The blue line indicate the true wall thickness every month. The yellow dots indicate inspections conducted (with noise). The red line marks the accepted loss (the baseline wall thickness minus MAWT).

4.3 Adaptive Strategies

In adaptive inspection strategies we do not necessarily conduct inspections at fixed time intervals, but rather when one or more of the risk criteria indicate that we should inspect. The most obvious such criterion is to calculate

$$p := p(y_o(t) \leq \text{MAWT}_o), \tag{4.1}$$

where $y_o(t)$ is an unobserved pipe at time t and MAWT_o is the corresponding (known) MAWT for this pipe. The probability p can be calculated with the distribution $\mathcal{N}(\mu_{\boldsymbol{y}_o}, \Sigma_{\boldsymbol{y}_o})$

as defined in Chapter 3. Based on some threshold L we choose whether or not we should conduct an inspection. That is, our **decision rule** is to inspect if

$$p > L.$$

An important feature of adaptive strategies is – as the name suggests – that we adapt the strategy when new information is acquired, i.e. update the underlying probability distributions that are the foundation of the inspection strategy. The Bayesian framework presented in Chapter 3 will be helpful in this setting.

The power of this idea lies partially in the fact that we use the predictive distributions to calculate a separate probability for every pipe. We do not necessarily inspect all pipes, but rather the subset of pipes that our model predicts has a high probability p.

Adaptive inspection strategies are structured as follows:

- Use an underlying statistical model to predict the probability that an unobserved pipe has a *critically low* wall thickness, i.e. lower than MAWT
- If this probability is above a user defined threshold L, we inspect the pipe and obtain a data point y_{obs}
- Use the observed values to update our model, i.e. calculate the posterior distribution $p(\boldsymbol{\gamma}|y_{obs})$ and the predictive distribution $p(\boldsymbol{y}|y_{obs})$
- Repeat this for all time steps t up until some maximum time step T

4.3.1 Repairing a Pipe

We have not yet talked about what to if we in fact believe that a pipe is dangerously close to, or even below, MAWT. In Figure 4.1, for instance, our predictions are *below* MAWT without any action being taken.

The obvious answer is to repair the pipe. We assume that if a pipe is repaired or changed, the Loss of Material goes back to 0 mm. But when should we change it? A simple decision rule is to change the pipe if the observed wall thickness, y_{obs}, is within some distance α form MAWT, i.e. if

$$y_{obs} < \text{MAWT} + \alpha,$$

for some $\alpha > 0$.

In Figure 4.2 we illustrate an adaptive strategy on the same pipe as in Figure 4.1. The clear distinction that can be seen is the lack of inspections early in the pipe's lifetime in Figure 4.2 – in fact, the first inspection is conducted in month 77, after more than 6 years – compared to the non adaptive strategy from Figure 4.1. However, the adaptive strategy conducts lots of inspections from month 130 and on wards, and actually ends up doing a total of 17 inspections, two more than the non-adaptive strategy. It also conducts a repair in month 142, which avoids that the wall thickness goes below the accepted loss, as happened in Figure 4.1. It is also worth noting how the predictive distribution is updated as new data is gathered: Up until the first inspection in month 130 the mean lies consequently a bit too low, but after we start inspecting the model is adjusting its mean. Additionally, the

variance seems to decrease as more data is gathered, especially when the inspection results lie within the current confidence band. This shows the strength of an adaptive algorithm that updates after new inspections.

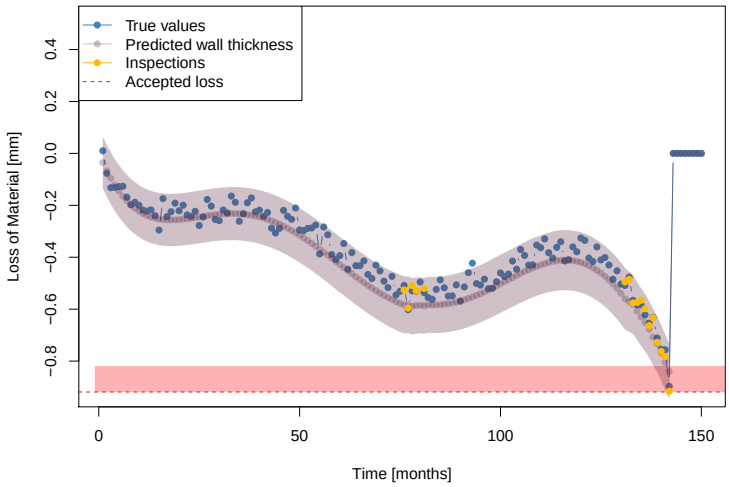

Figure 4.2: The same pipe as in Figure 4.1, but with an adaptive strategy. The blue dotted line indicates the true values, while the grey line is our prediction, with a confidence band of +/- one standard deviation. The red line indicates accepted loss (baseline wall thickness minus MAWT), and additionally a red shaded area which is the ”repair zone”; if we make an observation in this zone, we immediately repair the pipe. This is what happened in this example, and the wall thickness is reset to zero as the pipe is repaired. In this example we have used an $\alpha = 0.1$, meaning that if we make an observation within 0.1 mm of accepted loss, we repair. We have also used $L = 0.001$, meaning that we choose to inspect the pipe at time t if $p(y_o(t) \leq \text{MAWT}_o) > 0.001$. The result is 17 inspections and a successful repair.

A key motivation behind choosing an adaptive algorithm rather than a non-adaptive algorithm is the possibility to decrease the number of inspections by avoiding unnecessary inspections when the wall thickness is thought to be far from MAWT. However, the drawback of following this strategy blindly is that we end up doing many inspections when we are close to MAWT. After all, the criterion used to decide if we should inspect is $P(y_o(t) \leq \text{MAWT}_o)$, and this is likely to increase over time. So while an adaptive monitoring algorithm without a doubt does a good job of reducing inspections early in a pipe's lifetime, it will instruct us to inspect increasingly frequent near the end of the lifetime of the pipe. For instance, the algorithm tells us to inspect the pipes 12 months in a row in Figure 4.2! This seems a bit too often, and here is a potential point of improvement in the adaptive strategy.

One idea to avoid this is to stop inspecting after j consecutive inspections. We can then use the predictive distribution (updated after the last inspection) to estimate when we expect the pipe to reach MAWT, and come back and repair the pipe in this month, or some

months before the estimated time of MAWT.

Algorithm 1 Adaptive Monitoring Algorithm (AMA)

Input : A batch of pipes to be inspected;
MAWT for every pipe;
a threshold $L \in [0, 1]$;
a max number of iterations T;
a repair threshold α (in mm)
Output: A simulated inspection plan
while *time step* $t < T$ **do**
  $\mathcal{D} = \emptyset$
  for *pipe i* **do**
    $p_i = P(y_i(t) \leq \text{MAWT}_i)$
    if $p_i \geq L$ **then**
      Inspect pipe, obtain observed value y_{obs}
      $\mathcal{D} \leftarrow \mathcal{D} \cup y_{obs}$
      if $y_{obs} < MAWT_i + \alpha$ **then**
        Repair;
      end
    end
  end
  Update distribution: $P(\boldsymbol{y}) \leftarrow P(\boldsymbol{y}|\mathcal{D})$
end

How to choose the probability threshold L? In many cases this can be derived from a cost perspective: If the cost of an event is 100 times the cost of an inspection, we should inspect if $L > 1/100$, since the expected cost of an inspection then is lower than the expected cost of not inspecting. It can also be set by the operator of the facility based on empirical knowledge.

4.3.2 The Adaptive Monitoring Algorithm

In Algorithm 1, we have formulated the Adaptive Monitoring Algorithm (AMA) to formalize the adaptive monitoring procedure. For every time step, the algorithm calculates the probability that $y_i(t)$, the wall thickness of pipe i at time t, is close to MAWT_i. If this probability is reasonably large, defined by the threshold L, we inspect the pipe. After each time step we update our distribution for $\boldsymbol{y}$. Notice that we potentially have multiple pipes, and update the distribution after every time step. The variable $\mathcal{D}$ is used to store the observations in every time step, in order to update the distribution. If the pipe is "dangerously close" to MAWT we repair the pipe, as described in Section 4.3.1.

In the Adaptive Monitoring Algorithm we use $p = P(y_o(t) \leq \text{MAWT}_o)$ as the mechanism to decide if we should inspect or not. However, as discussed previously, this method has its drawbacks, for instance the huge amount of inspections towards the end of the lifetime of the pipe. Some possible solutions exist, such as the previously discussed idea of a simple stopping rule after j consecutive inspections, but these solutions tend to be less rigorous and more pragmatic. Because of this it might be tempting to ask if there are other

ways to define *informativeness* of an inspection, as we might doubt that the 12-th consecutive inspection of a pipe gives much new information. To investigate this, we will now present an alternative criterion to decide if a pipe should be inspected or not.

4.4 A Value of Information Approach

In Algorithm 1 we used the probability p as a decision rule; if p is large enough, we choose to inspect. We will not present another decision rule for inspections which uses p, but approaches the problem from a **Value of Information (VOI)** perspective.

What is the price and what is the value of information gathering? Eidsvik et al. (2015) describe three factors that make information valuable (the *pyramid of conditions*):

1. **Relevant.** A fundamental requirement – the information we gather must be relevant to the decision we want to make.
2. **Material.** Can the information, if we choose to pursue it, change the decision we want to make?
3. **Economic.** Once we know that the information is relevant and possibly will change our view, we must ask ourselves: Is it worth it? Does the value of the information exceed the cost?

4.4.1 The Intuition Behind VOI

Assume that the cost of repairing a pipe in time, i.e. before an event occurs, is C. Also, assume that the cost of an event is D. The exact relationship between C and D depends on the fluids in the pipe and various parameters in the financial situation of the operator of the facility, but an estimated figure is $C/D \approx 1/100$; it is around 100 times more expensive to wait until after a catastrophe has occurred, than it is to perform a preventive repair. In other words, we can formulate this as a decision rule based on the cost of the actions we want to decide between:

Alternative 1: Repair now, before any event occurs.
Alternative 2: Wait until event occurs.

This begs the question: What is the expected cost of our action for an unobserved instance $y_o(t)$? If we choose alternative 1 the cost is C. If we choose alternative 2, however, the cost is dependent on the probability of $y_o(t)$ being less than MAWT. In other words, the expected cost of alternative 2 is $D \cdot P(y_o(t) \leq \text{MAWT}_o)$. Let the *Prior Value* (PV) of the pipe cost be

$$\text{PV} = \min\{C, D \cdot P(y_o(t) \leq \text{MAWT})\}. \tag{4.2}$$

Now, assume we have conducted an inspection and obtained some new data $\boldsymbol{y}_{obs}$. Then the *Posterior Value* (PoV) is

$$\text{PoV}(\boldsymbol{y}_{obs}) = \min\{C, D \cdot P(y_o(t) \leq \text{MAWT}|\boldsymbol{y}_{obs})\}. \tag{4.3}$$

This way, we can define the Value of Information as

$$\text{VOI}(\boldsymbol{y}_{obs}) = \text{PV} - \text{PoV}(\boldsymbol{y}_{obs}). \tag{4.4}$$

Intuitively, $\text{VOI}(\boldsymbol{y}_{obs})$ measures how valuable the observation of $\boldsymbol{y}_{obs}$ is, based on how much it changed our idea of y_o. So in this case, we are not measuring the informativeness of y_o, but of the observation $\boldsymbol{y}_{obs}$, while y_o serves as a way of measuring the impact. A large $\text{VOI}(\boldsymbol{y}_{obs})$ indicates that the observation of $\boldsymbol{y}_{obs}$ made us change our belief about y_o significantly. However, if we have not made any observations $\boldsymbol{y}_{obs}$ we can still calculate the expected PoV by taking the expected value of $\boldsymbol{y}$:

$$\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] = \int_{\boldsymbol{y}} \min\{C, D \cdot P(y_o(t) \leq \text{MAWT}|\boldsymbol{y})\} p(\boldsymbol{y}) d\boldsymbol{y}. \tag{4.5}$$

Since the $\min\{\cdot,\cdot\}$ function is concave, Jensen's inequality yields that

$$\begin{aligned}
&\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] \leq PoV(\mathbb{E}_y[\boldsymbol{y}]) \\
\implies &\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] = \int \min\{C, D \cdot p_{y_o}(\mathbf{y})\} p(\mathbf{y}) d\mathbf{y} \\
&\leq \min\{\int C \cdot p(\mathbf{y}) d\mathbf{y}, \int D \cdot p_{y_o}(\mathbf{y}) p(\mathbf{y}) d\mathbf{y}\} \\
&= \min\{C, D \int P(y_o(t) \leq \text{MAWT}|\boldsymbol{y}) \cdot p(\boldsymbol{y}) d\boldsymbol{y}\} \\
&= \min\{C, D \cdot P(y_o(t) \leq \text{MAWT})\} \\
&= \text{PV} = \mathbb{E}_y\left[\text{PV}\right].
\end{aligned}$$

Consequently we can conclude that

$$\begin{aligned}
&\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] \leq \mathbb{E}_y\left[\text{PV}\right] \\
\implies &\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right] \geq 0
\end{aligned}$$

namely that the value of the information $\boldsymbol{y}$ is expected to be positive. Now, we can let this, the expected VOI, serve as an alternative criterion to decide weather or not to inspect; instead of choosing to inspect solely based on the probability of $y_o(t)$ being less than MAWT, we can prioritize to inspect in the places where the expected VOI is the highest.

4.4.2 Deriving a Distribution

The expression for $\mathbb{E}\left[\text{PoV}(\boldsymbol{y})\right]$ in (4.5) is not straight forward to compute. We will now derive a closed form solution given by a sum of cumulative multivariate normal distributions. For an unobserved instance y_o we can write

$$\begin{aligned}
p = P(y_o \leq \text{MAWT}) &= \int_{-\infty}^{\text{MAWT}} p(y_o) dy_o \\
&= \int_{-\infty}^{\text{MAWT}} \mathcal{N}(Q_o^\top \mu_\gamma, \sigma_o^2) dy_o \\
&= \Phi\left(\frac{\text{MAWT} - Q_o^\top \mu_\gamma}{\sqrt{\sigma_o^2}}\right)
\end{aligned}$$

where σ_o^2 is as described in (3.22). Similarly, denote

$$p_{y_o}(\boldsymbol{y}_{obs}) := P(y_o \leq \text{MAWT}|\boldsymbol{y}_{obs}) = \Phi\left(\frac{\text{MAWT} - \mu_{y_o|y}}{\sigma_{y_o|y}}\right), \tag{4.6}$$

where $\mu_{y_o|y}$ and $\sigma_{y_o|y}$ are from the predictive distribution described in (3.27). $p_{y_o}(\boldsymbol{y}_{obs})$ The expected PoV can then be expressed as

$$\begin{aligned}
\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] &= \int_{\boldsymbol{y}} \min\{C, D \cdot p_{y_o}(\boldsymbol{y})\} p(\boldsymbol{y}) d\boldsymbol{y} \\
&= \int_{\boldsymbol{y}} \min\{C, D \cdot p_{y_o}(\boldsymbol{y})\} \mathcal{N}\left(Q\boldsymbol{\mu_\gamma}, Q\Sigma_\gamma Q^\top + \tau^2 I\right) d\boldsymbol{y} \\
&= \int_{\mu_{y_o|y}} \min\left\{C, D \cdot \Phi\left(\frac{\text{MAWT} - \mu_{y_o|y}}{\sigma_{y_o|y}}\right)\right\} p(\mu_{y_o|y}) d\mu_{y_o|y}
\end{aligned} \tag{4.7}$$

as $\mu_{y_o|y}$ is the important linear combination of the data, since it is the only term that depends on $\boldsymbol{y}$. The mean and variance of the predictive mean $\mu_{y_o|y}$ from (3.27) can be calculated by utilizing $\mathbb{E}(\boldsymbol{y})$ and $\text{Var}(\boldsymbol{y})$ from (3.24), and the result is

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\mu_{y_o|y}}] &= Q_o \boldsymbol{\mu_\gamma} \\
\text{Var}[\boldsymbol{\mu_{y_o|y}}] &= Q_o \Sigma_\gamma Q^\top (Q\Sigma_\gamma Q^\top + \tau^2 I_n)^{-1} Q\Sigma_\gamma Q_o^\top
\end{aligned}$$

in the multidimensional case. If we only consider a single instance $y_o \in \mathbb{R}$, then the matrix Q_o reduces to simply a row vector, but the same equations would of course hold. For notational simplicity we will denote

$$z_1 = \frac{\text{MAWT} - \mu_{y_o|y}}{\sigma_{y_o|y}}.$$

Furthermore, we can split the integral of the $\min\{\cdot,\cdot\}$ function into two separate integrals; the first from $-\infty$ to $\bar{\mu}$ and the second from $\bar{\mu}$ to ∞, where $\bar{\mu}$ is defined by the equation

$$C = D \cdot \Phi\left(\frac{\text{MAWT} - \bar{\mu}}{\sigma_{y_o|y}}\right) \quad \Longrightarrow \quad \bar{\mu} = \text{MAWT} - \Phi^{-1}\left(\frac{C}{D}\right)\sigma_{y_o|y}.$$

Notice that $\bar{\mu}$ is simply a constant, as $\sigma_{y_o|y}$ is independent of $\boldsymbol{y}$. We can proceed to write (4.7) as

$$\begin{aligned}
\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] &= \int_{-\infty}^{\bar{\mu}} C \cdot p(\mu_{y_o|y}) d\mu_{y_o|y} \\
&\quad + \int_{\bar{\mu}}^{\infty} D \cdot \Phi\left(\frac{\text{MAWT} - \mu_{y_o|y}}{\sigma_{y_o|y}}\right) p(\mu_{y_o|y}) d\mu_{y_o|y} \\
&= C \int_{\mu_{y_o|y}=-\infty}^{\bar{\mu}} p(\mu_{y_o|y}) d\mu_{y_o|y} \\
&\quad + \int_{\mu_{y_o|y}=\bar{\mu}}^{\infty} \int_{z=-\infty}^{z_1} p(z) p(\mu_{y_o|y}) dz d\mu_{y_o|y} \\
&= C \cdot P(\mu_{y_o|y} < \bar{\mu}) + D \cdot P(z < z_1, \mu_{y_o|y} > \bar{\mu}),
\end{aligned} \tag{4.8}$$

where $z \sim \mathcal{N}(0,1)$. Let us introduce

$$\begin{aligned} V &= \mu_{y_o|y} - \bar{\mu} \\ W_1 &= \bar{\mu} - \mu_{y_o|y} \qquad W_2 = \sigma_{y_o|y} \cdot z - \text{MAWT} + \mu_{y_o|y}, \end{aligned}$$

such that we can express (4.8) as

$$\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] = C \cdot P(V < 0) + D \cdot P(W_1 < 0, W_2 < 0).$$

Now, we have reduced the expected value of the $\min\{\cdot,\cdot\}$ function to a sum of two multivariate normal calculations. If we know the expected value and covariance matrix of the multivariate distribution $P(V)$ and the joint multivariate distribution $P(W_1, W_2)$, this would be easy to calculate. Because of linear combinations we obtain that

$$\begin{aligned} \mathbb{E}(V) &= \mathbb{E}_y[\mu_{y_o|y}] - \mathbb{E}_y\left[\bar{\mu}\right] \\ &= Q_o \boldsymbol{\mu_\gamma} - \bar{\mu} \\ \text{Var}(V) &= \text{Var}(\mu_{y_o|y}) \\ &= Q_o \Sigma_\gamma Q^\top (Q \Sigma_\gamma Q^\top + \tau^2 I_n)^{-1} Q \Sigma_\gamma Q_o^\top, \end{aligned}$$

since $\bar{\mu}$ is just a constant. Equivalently for W_1, W_2:

$$\begin{aligned} \mathbb{E}(W_1) &= \bar{\mu} - Q_o \boldsymbol{\mu_\gamma} \\ \mathbb{E}(W_2) &= \mathbb{E}(\sigma_{y_o|y} \cdot z - \text{MAWT} + \mu_{y_o|y}) \\ &= \sigma_{y_o|y} \cdot z - \text{MAWT} + \mathbb{E}(\mu_{y_o|y}) \\ &= \sigma_{y_o|y} \cdot z - \text{MAWT} + Q_o \boldsymbol{\mu_\gamma} \\ \text{Var}(W_1) &= Q_o \Sigma_\gamma Q^\top (Q \Sigma_\gamma Q^\top + \tau^2 I_n)^{-1} Q \Sigma_\gamma Q_o^\top \\ \text{Var}(W_2) &= \sigma^2_{y_o|y} + Q_o \Sigma_\gamma Q^\top (Q \Sigma_\gamma Q^\top + \tau^2 I_n)^{-1} Q \Sigma_\gamma Q_o^\top \\ \text{Cov}(W_1, W_2) &= \text{Cov}(\bar{\mu} - \mu_{y_o|y}, \sigma_{y_o|y} \cdot z - \text{MAWT} + \mu_{y_o|y}) \\ &= \text{Cov}(-\mu_{y_o|y}, \sigma_{y_o|y} \cdot z + \mu_{y_o|y}) \\ &= \text{Cov}(-\mu_{y_o|y}, \sigma_{y_o|y} \cdot z) + \text{Cov}(-\mu_{y_o|y}, \mu_{y_o|y}) \\ &= 0 - \text{Var}(\mu_{y_o|y}) \\ &= -\text{Var}(\mu_{y_o|y}) \\ &= -Q_o \Sigma_\gamma Q^\top (Q \Sigma_\gamma Q^\top + \tau^2 I_n)^{-1} Q \Sigma_\gamma Q_o^\top \end{aligned}$$

Thus, we have

$$P(V < 0) = \Phi\Big(\mathbb{E}(V), \text{Var}(V)\Big) \tag{4.9}$$

and

$$P(W_1 < 0, W_2 < 0) = \Phi\left(\begin{bmatrix} \mathbb{E}(W_1) \\ \mathbb{E}(W_2) \end{bmatrix}, \begin{bmatrix} \text{Var}(W_1) & \text{Cov}(W_1, W_2) \\ \text{Cov}(W_2, W_1) & \text{Var}(W_2) \end{bmatrix}\right). \tag{4.10}$$

We finally arrive at the goal of all the calculations: We started out with $\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right]$ as an integral over a $\min\{\cdot,\cdot\}$ function, which now has been reduced to

$$\mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] = C \cdot P(V < 0) + D \cdot P(W_1 < 0, W_2 < 0), \tag{4.11}$$

i.e. a sum of two cumulative multivariate normal distribution, where we know the parameters in both distributions. Figure 4.3 shows a visual interpretation of the cumulative

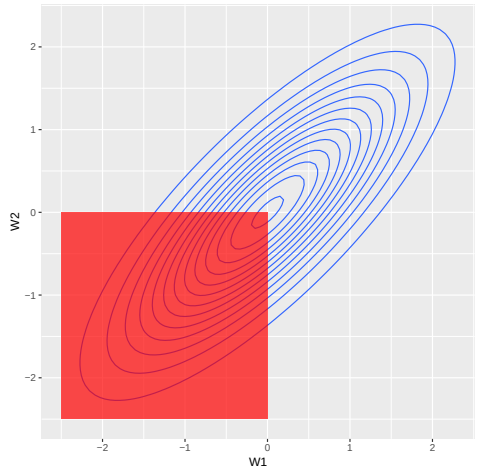

Figure 4.3: A visual interpretation of $P(W_1 < 0, W_2 < 0)$ (equation (4.11)). The contour lines show a multivariate normal distribution centered in $[0, 0]$ with $\text{Var}(W_1) = \text{Var}(W_2) = 1$ and $\text{Cov}(W_1, W_2) = 0.8$. The red area indicates the cumulative distribution, i.e. the volume under the multivariate distribution in the region $\{(W_1, W_2) : W_1 < 0, W_2 < 0\}$.

distribution. There are generally no easy-to-handle closed form for cumulative multivariate normal distributions, but numerical approximations are commonly known. See for instance Genz (2018). We can now calculate the expected $\text{VOI}(\boldsymbol{y})$ as

$$\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right] = \text{PV} - \mathbb{E}_y\left[\text{PoV}(\boldsymbol{y})\right] \tag{4.12}$$

and use that as a criterion to decide if we should inspect or not. An algorithmic presentation of the VOI procedure is presented in Algorithm 2. In Figure 4.4 we see a realized VOI strategy on the same pipe as in Figure 4.1 and 4.2. In the figure we have used the costs $C = 1, D = 100$, i.e. we assume that an event costs hundred times more than a repair in time. Furthermore, we have chosen to inspect if $\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right] > 0.5$. The $\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right]$ can be seen in the right plot. Notice that, in this particular example, we end up **not** repairing the pipe in time. While it of course is desirable to repair the pipe in time, this example is an illustrative example of how the VOI criterion differs from the adaptive strategy. In the VOI example in Figure 4.4 we stop inspecting close to the accepted loss because the VOI algorithm calculates that any more inspections not necessarily would be informative enough. This also shows the dangers of blindly following a VOI strategy; we might be encouraged not to inspect even though the current wall thickness is very close to MAWT.

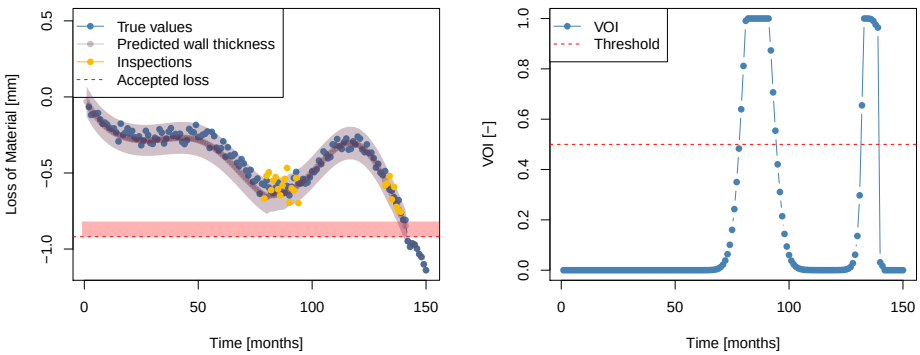

Figure 4.4: The same pipe as in Figure 4.1 and 4.2, but with the VOI strategy. We have used $C = 1, D = 100$ as costs for repair and event, respectively. Additionally, we have used the hyper parameter $L_{\text{VOI}} = 0.5$, i.e. we inspect if the predicted VOI is higher than 0.5. The result is 24 inspections, but one event: We fail to repair the pipe before it goes below the accepted loss. **Left:** The true wall thickness (blue) compared with our predictions (grey) and inspections (yellow). **Right:** The expected VOI for every month, used as a decision rule.

4.4.3 Choices Related to the VOI Strategy

We must choose some L_{VOI} as a threshold, such that we inspect if $\mathbb{E}_y\left[\text{VOI}(\boldsymbol{y})\right] > L_{\text{VOI}}$ and not inspect otherwise. This serves as the equivalent threshold to L in the AMA. While the L in an AMA is a probability threshold, this L_{VOI} is an *economic* threshold. In a non-adaptive method it might have been intuitive to set the threshold based a capacity or budget: We could for instance have a budget that allowed us to perform M number of inspections, and calculate beforehand which M inspections that would be the most informative. However, this approach does not work in an adaptive setting, as we continuously update our model. Something that would be deemed highly informative at one point in time can be deemed not so interesting in the next time point, if we have updated our model with new data.

The VOI metric measures the difference between $P(y_o(t) \leq \text{MAWT})$ and $P(y_o(t) < \text{MAWT}|\boldsymbol{y}_{obs})$. In other words, we try to measure how the additional information $\boldsymbol{y}_{obs}$ changes our belief about y_o. But what should y_o be, i.e. on *what* should we measure effect of $\boldsymbol{y}_{obs}$ on? If we inspect a batch of pipes together, then y_o could be the other pipes in the batch. Assume, for simplicity, that we operate 10 pipes. Then $\text{VOI}(y_1(t))$ could be the effect of inspecting pipe 1 on the other 9 pipes. This would ensure a *transfer effect*, where all the pipes in the batch utilize the effect of inspection at pipe 1. However, this is not always desirable: Pipes might be grouped together in a batch based on practicalities (they lie close to each other), and we might not always want this transfer effect. Therefore, another opportunity is to measure the expected *self-effect* of inspection, i.e. measure how informative an inspection of pipe 1 would be for our belief about pipe 1. In this case $Q = Q_0$ in (4.9) and (4.10). This is what we have done in Figure 4.4.

Algorithm 2 Informative Monitoring Algorithm (IMA)

Input : A batch of pipes to be inspected;
MAWT for every pipe;
a threshold L_{VOI};
a max number of iterations T;
a repair threshold α (in mm);
costs C and D for repair and events, respectively

Output: A simulated inspection plan

while *time step* $t < T$ **do**
  $\mathcal{D} = \emptyset$
  for *pipe i* **do**
    $\mathbb{E}[\text{PoV}(y_i)] = C \cdot P(V < 0) + D \cdot P(W_1 < 0, W_2 < 0)$
    $\mathbb{E}[\text{VOI}(y_i)] = PV - \mathbb{E}[\text{PoV}(y_i)]$
    if $\mathbb{E}[VOI(y_i)] \geq L_{VOI}$ **then**
      Inspect pipe, obtain observed value y_{obs}
      $\mathcal{D} \leftarrow \mathcal{D} \cup y_{obs}$
      if $y_{obs} < MAWT_i + \alpha$ **then**
        Repair;
      end
    end
  end
  Update distribution: $P(\boldsymbol{y}) \leftarrow P(\boldsymbol{y}|\mathcal{D})$
end

Chapter 5

Experiments and Results

In Chapter 2 we presented the Oceaneering data set. In Chapter 3 and 4 we presented a statistical framework for simulating various inspection strategies. In this chapter we apply these techniques to the Oceaneering data set. We first show the results of the Bayesian regression and estimate the coefficients in our Bayesian regression model, as presented in Chapter 3. Next, we proceed to test various configurations of the Adaptive Monitoring Algorithm presented in Chapter 4 and compare the results.

5.1 Model Specification from Real Data

5.1.1 Maximum Likelihood Estimators

We use the Oceaneering data set to calculate the $\boldsymbol{\theta}_{MLE}$, as described in (3.32). We can tune these hyperparameters with help from various optimizing tools in `R`, such as the `optim` function (Nash (2016)), the `PSO` package (Bendtsen (2012)) or the `MLR` (Bischl et al. (2016)). However, the different methods yield different numerical problems, and – most importantly – different results. It seems like optimizing the likelihood (3.29) might have multiple local optima that make the optimization problem vulnerable to small perturbations in starting values. As such, after testing various methods with various starting parameters for the optimization algorithm, we arrive at the MLE estimators

$$\boldsymbol{\theta}_{MLE} = \begin{bmatrix} \hat{\sigma}^2 \\ \hat{\phi} \\ \hat{\sigma}_\lambda^2 \\ \hat{\tau}^2 \end{bmatrix} = \begin{bmatrix} 0.1^2 \\ 0.0003 \\ 0.1^2 \\ 0.1^2 \end{bmatrix} \tag{5.1}$$

It is especially interesting to note how low $\hat{\phi}$ is. As seen in Chapter 3 a low ϕ indicate a stronger covariance between $\boldsymbol{\beta}$ coefficients that are far apart in time. The maximum likelihood estimate for ϕ indicate that our data set exhibit strong correlations in time. We will continue to use these hyperparameters throughout the rest of the experiments.

5.1.2 Results of Random Effects Model

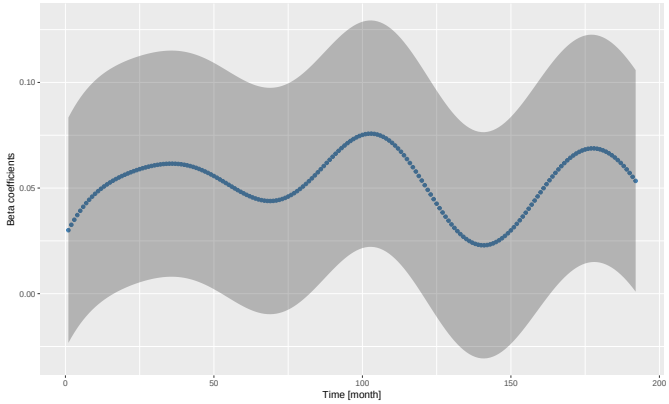

Figure 5.1: The $\boldsymbol{\beta}$ coefficients for all the 192 months, with a confidence band of one standard deviation.

We want to use the Oceaneering data set to find the distribution for the random variables $\boldsymbol{\gamma}$, i.e. we want to apply the posterior calculations from (3.26) to calculate $\boldsymbol{\gamma}|\boldsymbol{y}$. Here, $\boldsymbol{\gamma}$ is a collection of $\boldsymbol{\beta} \in \mathbb{R}^{192}$ and $\boldsymbol{\lambda} \in \mathbb{R}^6$, one for each of the six possible pipe features. To initialize this we apply flat prior means

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = [\mu, ..., \mu]^\top$$

where $\mu = -91 \cdot 10^{-3}$, which is the average rate in the whole data set from Oceaneering, regardless of feature type. Furthermore, we take the flat prior

$$\boldsymbol{\mu}_{\boldsymbol{\lambda}} = [-0.01, ..., -0.01]^\top \tag{5.2}$$

for $\boldsymbol{\lambda}$ and apply the Maximum Likelihood Estimators to calculate prior covariance matrices for $\boldsymbol{\gamma}$. More precisely,

$$\Sigma_\lambda = \hat{\sigma}_\lambda^2 I$$
$$\Sigma_\beta = \Sigma_\beta(\hat{\sigma}^2, \hat{\phi})$$

where $\Sigma_\beta(\hat{\sigma}^2, \hat{\phi})$ is the covariance structure introduced in Chapter 3 with MLE for $\hat{\sigma}^2$ and $\hat{\phi}$.

In Figure 5.1 we have plotted all the $\boldsymbol{\beta}$ coefficients. Interestingly enough, we observe that all the coefficients are positive. From this we might be lead to believe that the general trend is positive, i.e. that material is gained over time. This is not right, however, as every pipe also has a feature effect. These feature coefficients, $\boldsymbol{\lambda}$, are presented in Table 5.1. The largest effect comes from the "Tee" feature and the "Drain" feature, while we can see that "Elbow" and "Other" are almost the same. Similarly, the "Nozzle" and the "Piping" coefficients are also almost identical.

Coefficient	Estimate	Std. error
Feature:Nozzle	$-5.8145 \cdot 10^{-2}$	$53.5683 \cdot 10^{-3}$
Feature:Other	$-6.0724 \cdot 10^{-2}$	$53.5692 \cdot 10^{-3}$
Feature:Piping	$-5.9947 \cdot 10^{-2}$	$53.5694 \cdot 10^{-3}$
Feature:Drain	$-5.7684 \cdot 10^{-2}$	$53.5704 \cdot 10^{-3}$
Feature:Elbow	$-5.9903 \cdot 10^{-2}$	$53.5692 \cdot 10^{-3}$
Feature:Tee	$-6.4207 \cdot 10^{-2}$	$53.5692 \cdot 10^{-3}$

Table 5.1: The $\boldsymbol{\lambda}$ coefficients values and their standard error

We plot the cumulative $\boldsymbol{\beta} + \boldsymbol{\gamma} \cdot t$ in Figure 5.2. We can see clearly the idea behind including a feature specific effect in the random effects model: Every feature follows the same general trend, but is shifted up and down because of the feature specific coefficients in $\boldsymbol{\lambda}$. All the features experience a slight increase around month 120. This is not surprising given the data set: As seen in Chapter 1 the data has this trend, even though it intuitively might seem strange that the wall thickness increases.

We also see a slight tendency to this increase towards the end of the time period. This might be due to loss of data; as discussed in Chapter 1, more than 99.9% of the data set is from month 1 to month 150, with only some outliers around month 190. This makes it difficult to make precise predictions in the period of time between month 150 and 192. For these reasons, we will run our simulations up to month 150.

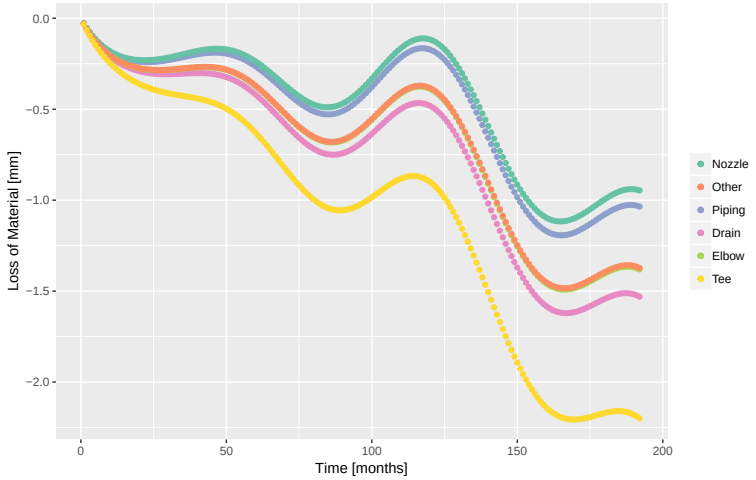

Figure 5.2: Left: Cumulative plot for $\boldsymbol{\beta} + Z\boldsymbol{\lambda}$. Note that we have **not** plotted only $\boldsymbol{\beta}$ alone, as every pipe also has one of the random effects. A cumulative plot of only $\boldsymbol{\beta}$ would be larger than zero almost all of the times; it relies on the random effect from $\boldsymbol{\lambda}$ to make sense.

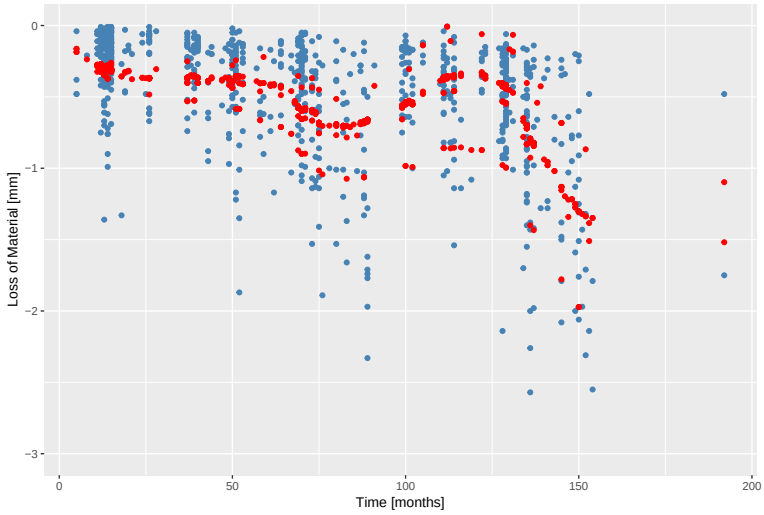

Figure 5.3: We train a model on 75% of the data set, and use this model to predict values from the test set. In this plot we see the true values from the test set (blue) and the predicted values (red).

5.2 RMSE and CRPS

We can test how precise our model is, measured by Root Mean Squared Error (RMSE) and Continuous Ranked Probability Score (CRPS). We have divided the original Oceaneering data set into a training set (75% of the data set) and used this to train a model. The final 25% percent of the data set is the hold out set used to test the predictions. One example of a prediction on a test set can be seen in Figure 5.3. The red points are the predictions (mean of the predictive distribution) and the blue points are the true values of the test set. In the figure we can see that some predictions lie slightly above or below others, instead of every prediction forming one line. This is due to the fact that different features have different $\boldsymbol{\lambda}$ coefficient, as shown in Figure 5.2.

In Table 5.2 we have calculated the average RMSE and the average CRPS after making 100 different splits into a training set and a testing set, each time with 75% in the test set and 25% in the test set. We compare with to other models:

1. The same kind of random effects model that we use, but *without* the features. That is, we use $\boldsymbol{y} \sim \mathcal{N}(X\boldsymbol{\beta}, X\Sigma_\beta X^\top + \tau^2 I_N)$, i.e. without the $\boldsymbol{\lambda}$ and Z. Apart from this we use the same methods as described in Chapter 3, including the same covariance structure in Σ_β.

2. A simple linear regression, where the Loss of Material at time t is modeled as

$$y(t) = \beta_0 + \beta_1 \cdot t + \epsilon, \tag{5.3}$$

   where β_0 and β_1 are intercept and slope, respectively, and t is time (in months). This model has neither a feature specific effect nor different effects for each month. This

model simply assumes that all pipes have the same monthly corrosion rate every month.

The results show that the random effects model we have used is indeed better than both the models we compare with. However, it also indicates that it is difficult to make precise predictions, as the average prediction error for our model is 0.2999 mm (measured by CRPS) and 0.4651 mm (measured by RMSE). This prediction error is not surprising, given the large variance in the data set. Over time, an increasing proportion of the pipes will experience corrosion damage, but there will also be many pipes that maintain 0 mm Loss of Material. This is the case for all features. Thus, however we choose to model Loss of Material, it is undeniable that – with the current covariates – we will get a notable prediction error.

Method	**RMSE [mm]**	**CRPS [mm]**
Random Effects Model With Features	0.4651	0.2999
Random Effects Without Features	0.5499	0.3044
Simple Linear Regression	0.5139	0.3219

Table 5.2: The average prediction error for the model, measured by RMSE and CRPS. We compare with a random effects model without the features and with a pure linear model.

5.3 Simulation Study

Now that we have estimated the best-fit estimators for the $\boldsymbol{\gamma}$ coefficients, we use these as the basis for a simulation study to compare the three different simulation strategies: The non-adaptive strategy, the adaptive strategy and the VOI strategy.

5.3.1 Experimental Setup

In order to evaluate an inspection strategy we need to know the ground truth, i.e. the true wall thickness of the pipe at any given point. As mentioned in Chapter 2, the Oceaneering data set consists of 4 inspections per pipe, but we need a ground truth for all of the k months we choose to simulate over for every pipe. We choose to simulate a ground truth from our data set. This can be done by sampling from the distribution of $\boldsymbol{y}$, as calculated in equation (3.12) in Chapter 3. Furthermore, the inspections are sampled from this ground truth with a normally distributed noise term with variance $\hat{\tau}$ (the maximum likelihood estimator).

When setting up the simulation study, we are also faced with another challenge: Empirically, the expected lifetime of the pipes is around 40 years, but since we only study less than half of this time period (192 of 480 months) the vast majority of the pipes end up being far from MAWT in our simulations. Therefore, to make the situation more interesting we must set an artificially high MAWT. This is done by setting MAWT to be a multiple of the baseline wall thickness. We choose this multiple to be 0.9, so that MAWT is 0.9 times baseline wall thickness. Using the same multiple for every pipe is not a problem: Given

that the baseline wall thickness varies between the pipes, the artificial MAWT also varies. For the simulation study we use a batch size of 1, meaning that we study the effect of one pipe at the time.

5.3.2 Some Example Results

In this section we test the three strategies – non-adaptive, adaptive and VOI – to the same set of pipes. The pipes are chosen at random from the Oceaneering data set. First, the non-adaptive strategy is depicted in Figure 5.4. Here we inspect every 10th month, thus totaling 45 inspections over the 150 months for the three pipes. We have also used a repair threshold of $\alpha = 0.1$. In two of the examples we fail to repair the pipe before it reaches the accepted loss.

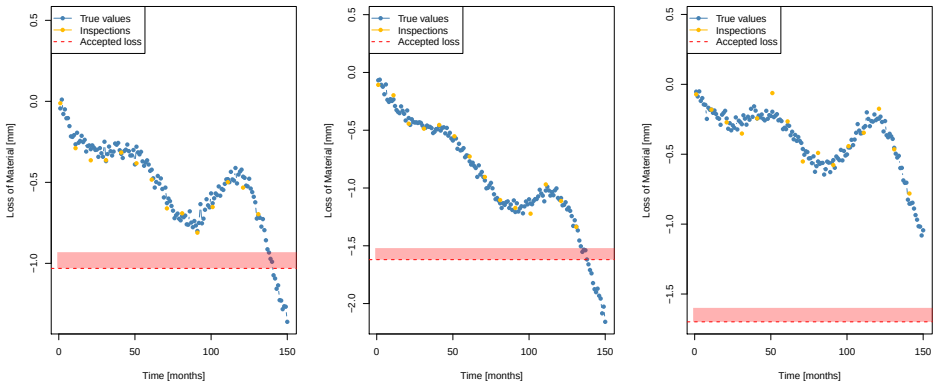

Figure 5.4: A non-adaptive strategy applied to three different pipes for $k = 150$ months. In the first two plots the pipe reaches the accepted loss without being noticed in time. In the final example the pipe stays well above the accepted loss throughout its lifetime. We have used an inspection interval of 10 months, thus we end up with a total of 45 inspections, zero repairs and two events for the three example pipes.

Next, the adaptive strategy is depicted in Figure 5.5 on the same three pipes. Here we have used $L = 0.001$ as the probability threshold and $\alpha = 0.1$ as the repair threshold. From these examples it seems clear that the adaptive algorithm performs better than the non-adaptive algorithm; in total it ends up at just 8 inspections for the three pipes, also conducting two repairs (while the non-adaptive fails to notice both of these dangers in time). From the first example (the left plot) in Figure 5.5 we see how the predictions initially diverge slightly from the true values, but after three consecutive inspections the mean is adjusted and the variance is decreased so that the prediction fits follows the true curve closely. This shows the power of an adaptive model. However, in the second example (middle plot) we almost experience the opposite: The predictions lie consistently *above* the true values, leading us to believe that no inspection is necessary. In this case we inspected just in time, but this is the disadvantage of the adaptive model. If we believe that we are far from MAWT we do not feel the need to inspect, and thus we are unable to find

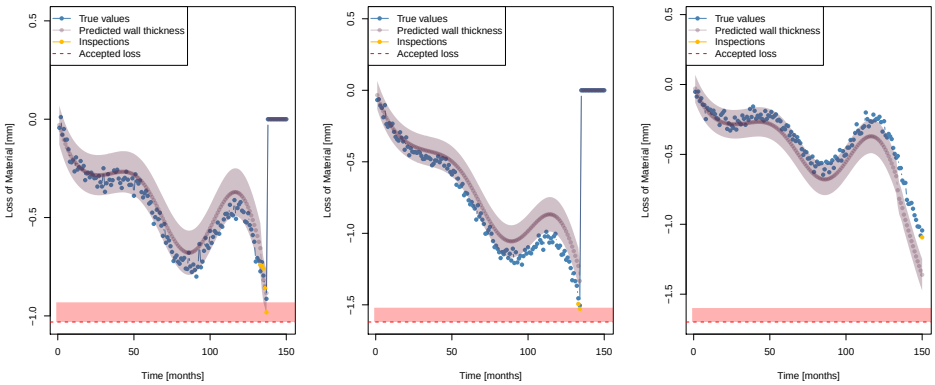

Figure 5.5: An adaptive strategy applied to the same three pipes as in Figure 5.4, with the same settings. We have used $\alpha = 0.1$ as the decision rule for reparation and $L = 0.001$ as the probability threshold for inspection.

out if our belief is correct. In the third example (right plot) we see the same thing, although we are further from MAWT. In this example we end up conducting a single inspection, in the final month. Finally, the VOI strategy applied to the same pipes can be seen in Figure 5.6. Here we use $\mathbb{E}[\text{VOI}(\boldsymbol{y})]$ as a criterion and $\alpha = 0.1$ as the repair threshold again. The corresponding expected VOI curves can be seen in Figure 5.7, with a threshold $L_{\text{VOI}} = 0.1$.

5.4 Comparing Strategies

In the Adaptive Monitoring Algorithm and the Informative Monitoring Algorithm presented in Chapter 4 (Algorithm 1 and 2, respectively) we have three *decision parameters* that must be set:

- L, the inspection threshold in the adaptive strategy;
- α, the repair threshold;
- L_{VOI}, the inspection threshold in the VOI strategy.

In some cases these are to be decided from Subject Matter Experts or based on the financial situation; as discussed in Chapter 4, the threshold L can be expressed in terms of C, the cost of reparation, and D, the cost of an event. However, this is not always the case. In this section we investigate how different decision hyper parameters perform when it comes to number of inspections, number of events, number of repairs and number of non-utilized months.

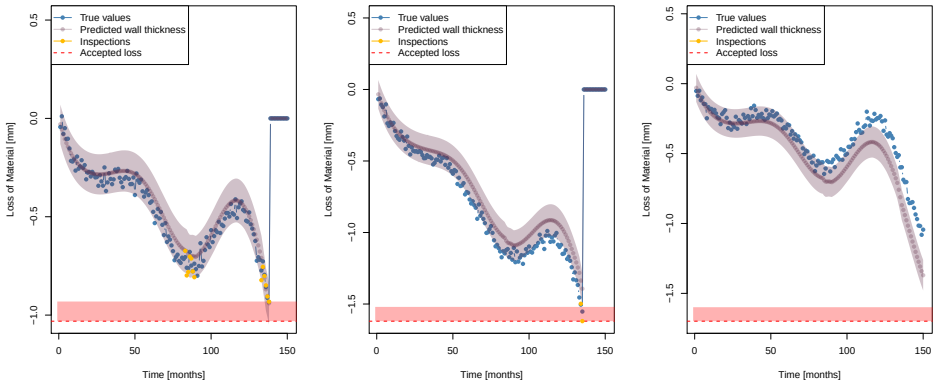

Figure 5.6: A VOI strategy applied to the same three pipes as in Figure 5.4 and 5.5, with the same settings. We have used $\alpha = 0.1$ as the decision rule for reparation and $L_{\text{VOI}} = 0.1$ as the decision rule for informativeness.

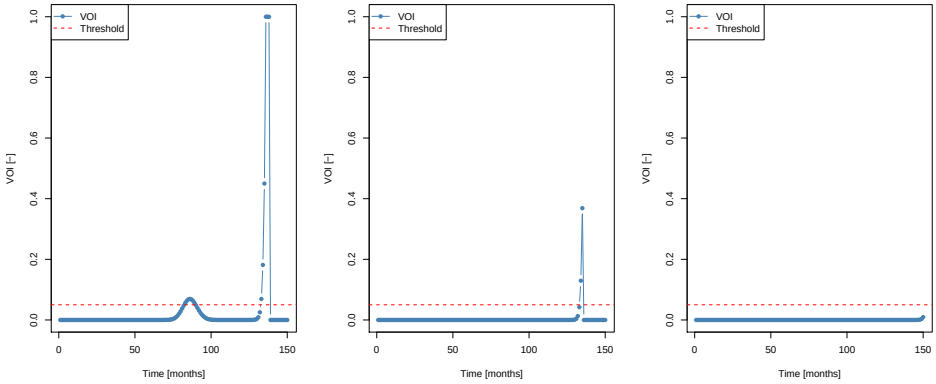

Figure 5.7: Predicted VOI for every time step which is used to decide if we should inspect or not, as seen in Figure 5.6. We use $L_{\text{VOI}} = 0.1$ as threshold.

Figure 5.8: We run the adaptive method for $\alpha \in [0, 0.01, 0.02, ..., 1]$ to see how different α values affect the outcome of inspection simulations with the three different strategies: non-adaptive (red), adaptive (blue) and VOI (yellow). For every α we run 100 simulations with batch size 1. Clockwise from top left corner: Number of inspections against α, number of events against α, number of non-utilized months against α and number of repairs against α.

5.4.1 The Repair Threshold α

We choose to repair a pipe in the Adaptive Monitoring Algorithm if the observed value is within a distance of α (in millimeter) from MAWT. In Figure 5.8 we have experimented with 100 different α values on an equidistant grid $[0.01, 0.02, ..., 1]$. For every α we ran 100 simulations of a pipe (with features chosen at random) for 150 months and observed the number of inspections, number of events, number of repairs and number of non-utilized months. This was done for the non-adaptive strategy, the adaptive strategy and the informative strategy. The results can be seen in Figure 5.8.

As α increase we are essentially quicker to repair the pipe; a large α would indicate that we conduct a repair even when the measured Loss of Material is far from MAWT. As a result, we actually make fewer inspections as α grows, because we repair the pipe in one of the first inspections. We can see that both the adaptive and the VOI strategy converges towards a single inspection for any $\alpha > 0.5$. As a result, the number of events quickly

also converges to zero, as we repair the pipe before any event is allowed to occur. Not surprisingly, we can also see that for all three strategies the number of inspections goes towards 1 as α grows, i.e. we repair almost all the pipes.

The number of non-utilized months stabilizes between month 15 and 25 for the adaptive and VOI strategy, as seen in the bottom right plot in Figure 5.8. That is, we end up repairing the pipe somewhere between month 125 and 135 (of 150 possible months). Seeing as the average number of inspections also goes towards 1, this essentially indicates that on average – when α is higher than 0.5 – we end up doing the first inspection between month 125 and 135, and repairing immediately.

We want to choose an α that yields few inspections and few events, but at the same time do not repair too early. This becomes a trade-off. From the top right figure (Events vs. α) we see that with $\alpha \approx 0.13$ we have already reduced the number of events as much as possible, while the number of non-utilized months is still pretty low. Choosing a higher α would yield more non-utilized months, but with little effect in number of events.

5.4.2 The Adaptive Monitoring Threshold L

In Figure 5.9 we run a similar experiment with varying the L parameter, i.e. the inspection threshold in AMA. We only run this simulation for the adaptive algorithm, as this is the only method that uses the probability threshold L. We make 100 simulations for every L value on an equidistant grid from $L = 0.00001$ to $L = 0.1$, each simulation running for 150 months. We have used $\alpha = 0.13$ as the repair threshold.

When L increases the number of inspections conducted goes down. This makes sense, as L is the probability threshold that we use to decide if we should inspect. As we increase the threshold, fewer pipes get accepted for an inspection. The number of events (top right corner) is slightly increasing, which comes as no surprise: As we conduct fewer inspections we have a smaller chance of discovering pipes that are very close to MAWT. For low L we experience approximately 0.01 events per pipe on average.

As L increase we also conduct fewer repairs and consequently we have less non-utilized months. This also comes as a direct consequence of fewer inspections.

5.4.3 The Informative Monitoring Threshold L_{VOI}

We do the same simulations for L_{VOI}, the inspection threshold in IMA. Just as with L in the adaptive algorithm, we experience fewer inspections as the inspection threshold increase. The results can be seen in Figure 5.10. Again the number of events increases slightly as L_{VOI} increases, but this is also more stable than for L in the adaptive case. The number of non-utilized months follows roughly the same trend as when we vary L, i.e. the number of non-utilized months decrease as the threshold L_{VOI} is increasing. The number of repairs stays relatively constant, although with a slightly decreasing trend. We would possibly have expected it to decrease even more, seeing that the number of inspections falls from 15 per pipe on average to less than 5 per pipe. This might indicate that the reduction in inspections primarily is happening early in the pipe's lifetime: The few inspections we do with a high L_{VOI} is concentrated towards the end of the pipe's lifetime, and thus we continue to make approximately the same number of repairs.

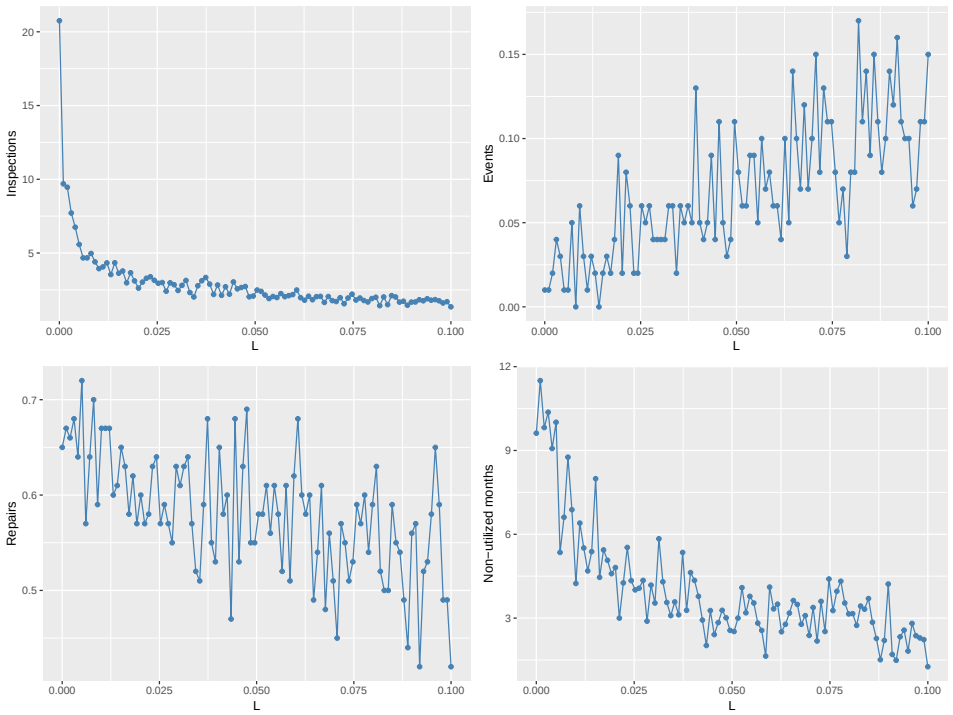

Figure 5.9: We run the adaptive method for L between 0.00001 and 0.1 to see how different L values affects the outcome of an adaptive monitoring strategy. For every L we run 100 simulations with a batch size of 1. Each simulation runs from month 1 to month 150. We have used a repair threshold of $\alpha = 0.13$. Clockwise from top left corner: Number of inspections against L, number of events against L, number of non-utilized months against L and number of repairs against L.

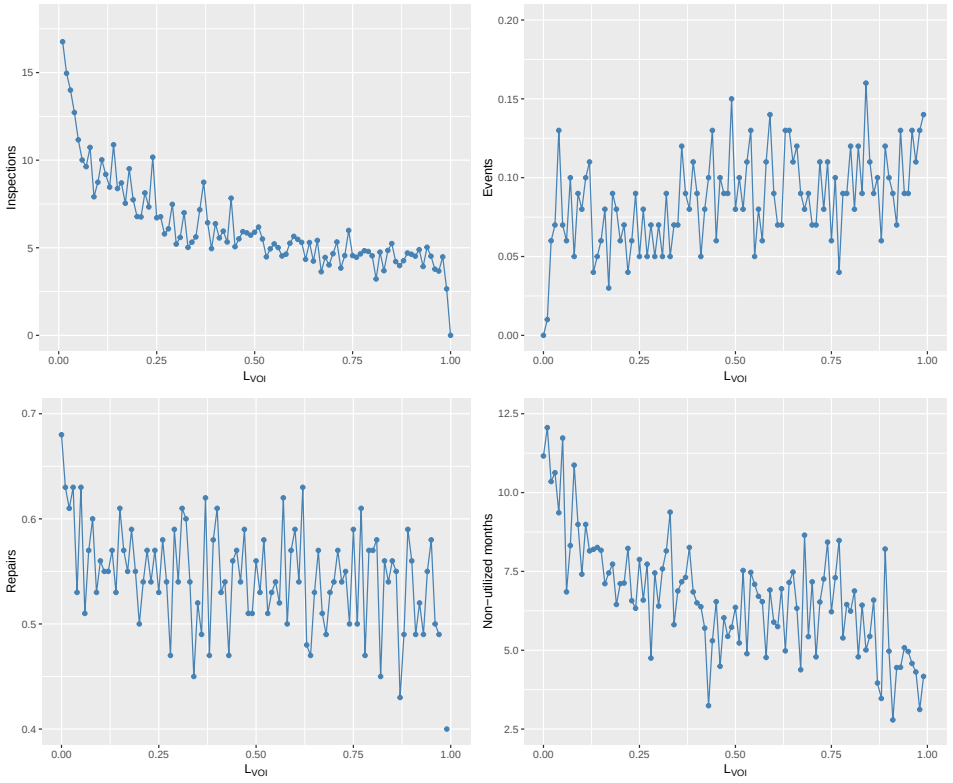

Figure 5.10: Simulations with 100 different L_{VOI} values for the informative strategy, from $L_{\text{VOI}} = 0.01$ to $L_{\text{VOI}} = 1$. We run 100 simulation of 150 months for every L_{VOI}. Clockwise from top left corner: Number of inspections against L_{VOI}, number of events against L_{VOI}, number of non-utilized months against L_{VOI} and number of repairs against L_{VOI}.

5.4.4 Comparing Value of Strategy

Let us now see how the three strategies – adaptive, non-adaptive and VOI strategy – compare in a larger simulation when we use fixed decision parameters. We use $L = 0.001, \alpha = 0.13$ and $L_{\text{VOI}} = 0.5$ and run 2000 simulations on the same pipes for each of the three strategies. We run two simulations for the non-adaptive strategy; one with inspections every 10 months and one with inspections every 20 months. We use a batch size of 1, and have set the costs in the VOI criterion to be $C = 1, D = 1000$. The results can be seen in Table 5.3.

	Inspections	**Events**	**Repairs**	**Non-utilized months**
Non-adaptive (freq. 10)	14.1545	0.3155	0.2215	3.5245
Non-adaptive (freq. 20)	7.5425	0.3605	0.1765	2.2910
Adaptive	9.2080	0.006	0.657	8.985
VOI	5.5525	0.0895	0.5400	7.0370

Table 5.3: Results of simulation study after running 2000 simulations. Each simulation is ran with a batch size of 1 and $k = 150$ months. We use repair threshold $\alpha = 0.13$ for all strategies. In the adaptive strategy we have used inspection threshold $L = 0.001$ and in the informative strategy we have used inspection threshold $L_{\text{VOI}} = 0.5$.

In the left column we can see the number of inspections. For the two non-adaptive simulations, the number of inspections is not so surprising, given that we have set a fixed inspection frequency. The adaptive method yields on average 9.2 inspections per pipe, which is approximately 35% less inspections than the non-adaptive method with frequency 10, but 22% more than the non-adaptive with frequency 20. The VOI method yields 5.5 inspections per pipe on average, which is more than 70% less than the non-adaptive method with frequency 10.

When looking at the number of events, it is clear that the adaptive strategy outperforms all the other strategies. On average every pipe experiences only 0.006 events (or: 0.6% of the pipes experience an event), whereas the non-adaptive strategies lie above 30%. The informative strategy is also significantly better than the non-adaptive ones, with 8.95% of the pipes experiencing an event. These results indicate strongly that the adaptive strategy is the best strategy to avoid events.

The price to pay for few events is multiple repairs and more non-utilized months. This pattern can clearly be seen in the final two columns of Table 5.3. While the non-adaptive methods have 2.29 and 3.52 non-utilized months on average, both the adaptive and the VOI method have more than 7 non-utilized months on average. This indicates that the repairs are being conducted a bit too early. This goes back to the choice of α, the repair threshold. A lower α would yield fewer non-utilized months, but also an increase in events.

5.5 Discussions About Methodology

While the results presented in Table 5.3 indicate that both the AMA and IMA yield significantly better results than following a non-adaptive strategy, one might argue that this

comparison is not entirely fair. In practical applications it is likely that the facility operator would interfere with the non-adaptive inspection scheme at certain times if some inspections are surprising.

It is also worth reminding that these simulations are run on an artificially high MAWT, as discussed in Section 5.3. This is because we only have data from 192 months – or even just 150 months, effectively. This makes some of the numbers artificially high. For instance, experiencing events in 30% of the pipes, as in the two non-adaptive strategies, is slightly unrealistic. However, the relationships between the different strategies are still valid and interesting.

Furthermore, the original data set consists of some covariates that we did not utilize, such as material type and insulation type. It would be highly interesting to include these covariates to make an even more precise predictive model.

We have discussed the opportunity of including a mechanism to stop inspecting after multiple consecutive inspections. There is reason to believe that this would decrease the number of inspections in the adaptive strategy and the VOI strategy, as we have experienced that these strategies often conduct inspections in as much as $14 - 15$ consecutive months. A stopping mechanism – either automated or with help from a human Subject Matter Expert – would most likely benefit these strategies.

Finally, the model relies heavily on choices of the decision parameters such as L, L_{VOI} and α. We have studied the effect of varying these parameters in Section 5.4 and made some choices based on these experiments, but this is also something that a facility operator would decide based on the financial situation of the company, practicalities or even laws that regulate accepted risk.

Chapter 6

Closing Remarks

We summarize the findings in the master thesis and propose some ideas for further work.

6.1 Key Results

In this thesis we have developed a random effect model to predict the Loss of Material – the loss in wall thickness due to corrosion – over time in oil pipes with different features. Based on this model we have derived two risk-based inspection strategies: an adaptive strategy and a VOI strategy. These strategies are compared with a non-adaptive model, where the pipes are inspected at fixed points in time, regardless of any new information gained in inspections.

We have simulated 150 months of lifetime for more than 2000 pipes. The key findings can be summarized as follows:

- For certain choices of decision parameters, the Adaptive Monitoring Algorithm (AMA) reduces the number of inspections with around 35% compared to a non-adaptive method with inspections every 10 months, but achieved 22% more inspections than a non-adaptive method with inspections every 20th month. However, the adaptive method reduces the number of events with more than 98% compared to the non-adaptive methods.
- The Informative Monitoring Algorithm (IMA) achieves the fewest inspections on average (5.55 per 150 months), while at the same time reducing the number of events with more than 70% compared with the non-adaptive methods.
- The number of non-utilized months is highest for the adaptive and VOI methods, indicating that these methods are possibly repairing the pipes too early.
- While the simulation results admittedly are sensitive to choices of decision parameters, we have demonstrated that both the AMA and the IMA consistently outperform the non-adaptive strategies in most realistic scenarios, when the risk of experiencing an event is used as a metric.

6.2 Further Work

The current data set consists of 4 inspections per pipe over a time frame of over 10 years. An even bigger data set with more observations would be of great interest to study further, as ironic as it might sound, as the purpose of this thesis is trying to limit the number of inspections.

Something that possibly is more available – and something that most likely would radically improve the predictive model – is production data. Examples are data about what kind of chemicals that flow through the pipes at various times, the temperature of the chemicals or the pressure inside the pipes. Sensors that monitor production in the oil and gas industry are increasingly used: Weinelt (2017) reported in a World Economic Forum report that modern offshore drilling facilities have about 80 000 sensors, producing 15 million gigabytes of data over the lifetime of an asset. Although offshore drilling facilities are somewhat different from a land-based refinery, it serves as an example of better utilization of production in the oil and gas industry. Furthermore, Mohammed reza Akhondi et al. (2010) reported that 25% of all wireless sensors deployed in 2009 were for the oil and industries, and predicted that refineries would benefit from this. Additionally, we know that production data is being monitored at the production facility that the Oceaneering data set comes from, although it was unfortunately not available for our use. This would definitely have been of interest to investigate further.

Furthermore, the calculated $\boldsymbol{\beta}$ coefficients from the Oceaneering data set indicate periods where wall thickness is gained. While this seems reasonable given the data set, it might be a bit surprising. Methods that enforce a monotonically decreasing wall thickness could be considered.

The optimization of the hyperparameters proved to be challenging, as different methods gave slightly different results. An in-depth study of this optimization problem might improve the predictions.

It would also have been interesting to study an even longer time frame; the oldest asset in the data set is 192 months old, but in many cases the expected lifetime of a pipe might be as long as 40 years, or 480 months. In the absence of these data, one could try various extrapolation techniques to create a synthetic data set to with longer history, in order to be able to do simulations over more than just the 192 months.

Bibliography

Anderson, T., 2003. An Introduction to Multivariate Statistical Analysis, 3rd Edition. Wiley Interscience.

API, 11 2009. Risk-based inspcetion: Api recommended practice 580. Tech. rep., http://www.irantpm.ir/wp-content/uploads/2013/08/API-RP-580-Risk-Based-Inspection-2009.pdf, (visited 20.05.19).

Bendtsen, C., 2012. Package 'pso': Particle Swarm Optimization. R Foundation for Statistical Computing, https://cran.r-project.org/web/packages/pso/index.html, (visited: 23.05.2019).

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z. M., 2016. mlr: Machine learning in r. Journal of Machine Learning Research 17 (170), 1–5, http://jmlr.org/papers/v17/15-066.html, (visited: 23.05.2019).

DNV, 2009. Dnv-rp-g101 recommended practice: Risk based inspection of offshore topsides static mechanical equipment. https://rules.dnvgl.com/docs/pdf/DNV/codes/docs/2009-04/RP-G101.pdf, (visited 20.05.2019).

Eidsvik, J., Mukerji, T., Bhattacharjya, D., 2015. Value of Information in the Earth Sciences, 1st Edition. Cambridge University Press.

Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2013. Regression: Models, Methods and Applications, 1st Edition. Springer-Verlag.

Genz, A. e. a., 2018. Package 'mvtnorm': Multivariate Normal and t Distributions. R Foundation for Statistical Computing, https://cran.r-project.org/web/packages/mvtnorm/mvtnorm.pdf, (visited: 01.12.2018).

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis, 6th Edition. Pearson.

Mivule, K., July 2012. Utilizing noise addition for data privacy, an overview. The 2012 International Conference on Information and Knowledge Engineering.

Mohammed reza Akhondi, A., Carlsen, S., Peterson, S., 2010. Applications of wireless sensor networks in the oil, gas and resources industries. 2010 24th IEEE International Conference on Advanced Information Networking and Applications.

Nash, V. e. a., 2016. Package 'optim': General-Purpose Optimization. R Foundation for Statistical Computing, https://cran.r-project.org/web/packages/optimr/optimr.pdf, (visited: 23.04.2019).

Rasmussen, C., 2004. Advanced Lectures on Machine Learning. Springer, Ch. Gaussian Processes in Machine Learning, pp. 63–71.

Weinelt, B., 01 2017. Digital transformation initiative: Oil and gas industry. Tech. rep., http://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/dti-oil-and-gas-industry-white-paper.pdf, (visited 20.05.19).