Extreme value analysis of ultrasonic thickness measurements

A thesis submitted for the degree of **Doctor of Philosophy**

Daniel Joseph Benstock

Department of Mechanical Engineering
Imperial College London
London, SW7 2AZ

April 2016

Abstract

Modern infrastructure and industrial plants have a finite design life. Their effective and profitable operation depends on well organized maintenance and condition assessments. Non-destructive evaluation and inspection is a key tool for condition assessment. However, despite best efforts, full inspection coverage of a plant is not always possible because of access problems, time constraints and limited budgets.

Many inspection companies are beginning to use partial coverage inspection (PCI) techniques to solve this problem. PCI describes the use of inspection data collected from a small area of the component to extrapolate to the condition an the rest of the component. Extreme value analysis (EVA) is a technique of particular interest for this application as it allows an inspector to construct a statistical model of the smallest thickness measurements across the component. This model can then be used to extrapolate to the most likely minimum thickness.

In this thesis, an analysis of the uncertainties that can arise when using EVA for extrapolation is performed. A clear outline of the uncertainties expected to result from EVA extrapolation is presented and it is made usable for inspectors. In addition, a simple test algorithm to analyse when EVA is suitable for a set of inspection data is described. It is hoped that the work described in this thesis will enhance confidence in the practical use of the technique in the field.

Furthermore, the effect of surface roughness on ultrasonic thickness measurements is investigated with joint experimental and computational studies. It is shown that the thickness measurement distribution can differ significantly from the actual thickness distribution, particularly for the smallest values of thickness and with rougher surface conditions. Consequently, extrapolations from extreme value models using ultrasonic thickness data are shown to be conservative compared to the true condition of the component.

Copyright Declaration

Declaration of Originality

The content of this thesis is the result of independent work carried out by myself under the supervision of Dr. Frederic Cegla. Appropriate references have been provided wherever use has been made of the work of others.

Daniel Joseph Benstock

Thursday 21st April, 2016

Acknowledgements

I would like to express my most sincere gratitude to my supervisor Dr. Fred Cegla. Your advice and mentoring over the past three years has help me to be productive and to develop the skills required to perform research. When I compare myself to when I joined the group, I can only conclude that you have made a huge contribution to both my personal and professional development.

I am indebted to Peter Cawley for offering me a place to study in the Non-destructive testing (NDT) group at Imperial College. Your feedback on both my written work and my conference presentations has helped me to become a more diligent and confident student.

I am most grateful to all of my colleagues in the Non-destructive testing group at Imperial College. It has been wonderful working with you all. In particular: Tristan, for organising frequent visits to Chelsea football club, Jack and Fan, for proving to be formidable squash partners, Nick Brierley, for lively debate at lunch, and Andrew Jarvis, for helping me settle in when I first started.

I'd like to thank Mark Stone and Gordon Davidson at Sonomatic Ltd. for their advice and assistance when I was developing my simulations of rough surface measurements. It was a great experience discussing the project with you and being able to see the technology you've developed in Warrington. I'd also like to thank Simon Burbidge, who runs the Imperial College High Performance Computing (HPC), for his help when I first started to use the HPC.

To have achieved this much, you have to have a strong foundation. At Leeds Grammar School I certainly developed that and I want to extend my thanks to everyone that taught me there. In particular: Peter Jolly and Mark Bailey, for being such great role models, and Derek Fry, for inspiring my lifelong love of science.

I extend my most heartfelt thanks to my friends, particularly Kristina, Hillary, Brad and Dan, whose support over the last few months will never been forgotten. Finally, I want to thank my parents, sisters and grandparents, to whom this thesis is dedicated. Your love and support and made all of this possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Assessments of the condition of in-service engineering components are an important part of ensuring operational safety and integrity during a facility's lifetime. Failure to assess damage can have catastrophic consequences [4] and lead to potential criminal prosecution [5]. In the United Kingdom, regulators require evidence that a plant is safe to run throughout its lifetime. Plant operators must provide evidence of plant integrity and safety. The operators identify the key components to survey and ask independent contractors to inspect these components at regular time intervals. The condition of the component is reported back to the operator who repairs or replaces the component if required and presents the results of this inspection to the industry regulator. As efficient and safe operation becomes increasingly important to nations and industries the value of inspection services will increase. The world wide NDE market is estimated to be worth \$5.6bn [1] [6].

Inspections are carried out using techniques from the field of non-destructive evaluation (NDE). NDE encompasses a range of methods such as ultrasonics [7, 8], radiography [9, 10], eddy current testing [11] and thermography [12, 13]. One of the most commonly used techniques in NDE is the ultrasonic thickness (UT) measurement (figure 1.1A). Measurements are made using an ultrasonic transducer which is coupled to the surface of the component using water or a water-based couplant.

[1] in 2012

An ultrasonic pulse is transmitted into the component, reflects from the rear wall of the component and the reflected pulse is recorded (figure 1.1B). A measurement of thickness can be extracted from this signal using one of a number of algorithms to extract a time of flight from the signal, each of which have varying accuracy [14]. The thickness of the component is given as:

$$t = \frac{\tau \cdot c_L}{2} \tag{1.1}$$

where τ is the time of flight, c_L is the speed of sound in the material and t is the thickness of the component.

UT measurements are usually performed as a C-scan over a large area. The transducer is moved across the surface of the component in a pre-described pattern with measurements taken at regular spatial intervals (figure 1.1). The thickness measurements are used to form a thickness map, which represents each thickness measurement using a coloured pixel (figure 1.1C).

It is common practice to schedule shut down periods to allow for these inspections as, more often than not, they require access to hazardous areas of a facility that cannot be accessed during operation. This incurs a loss of revenue to the facility owner due to the shut down periods in addition to the cost of the inspection. Even with the best efforts, access to the entire component is often restricted due to access problems (i.e. parts of the component concealed by pipework or insulation), time constraints or a limited inspection budget. Consequently, full inspection is not always possible.

For these reasons there is pressure to increase the efficiency of inspections, such that they can be performed in shorter periods of time or require access to smaller areas of the component. This has led to the development of new inspection strategies and technology, a prime example of which is guided wave testing (GWT). Guided wave based systems allow inspection of large areas of a component from a single inspection point [8]. Testing is performed by affixing a ring of transducers around the outside of the component. This ring of transducers excites an ultrasonic guided wave in the material.

The inspector can select a specific mode to excite in the component. Typically, a non-dispersive mode is selected as this allows the wave to propagate for long distances, and usually the shear

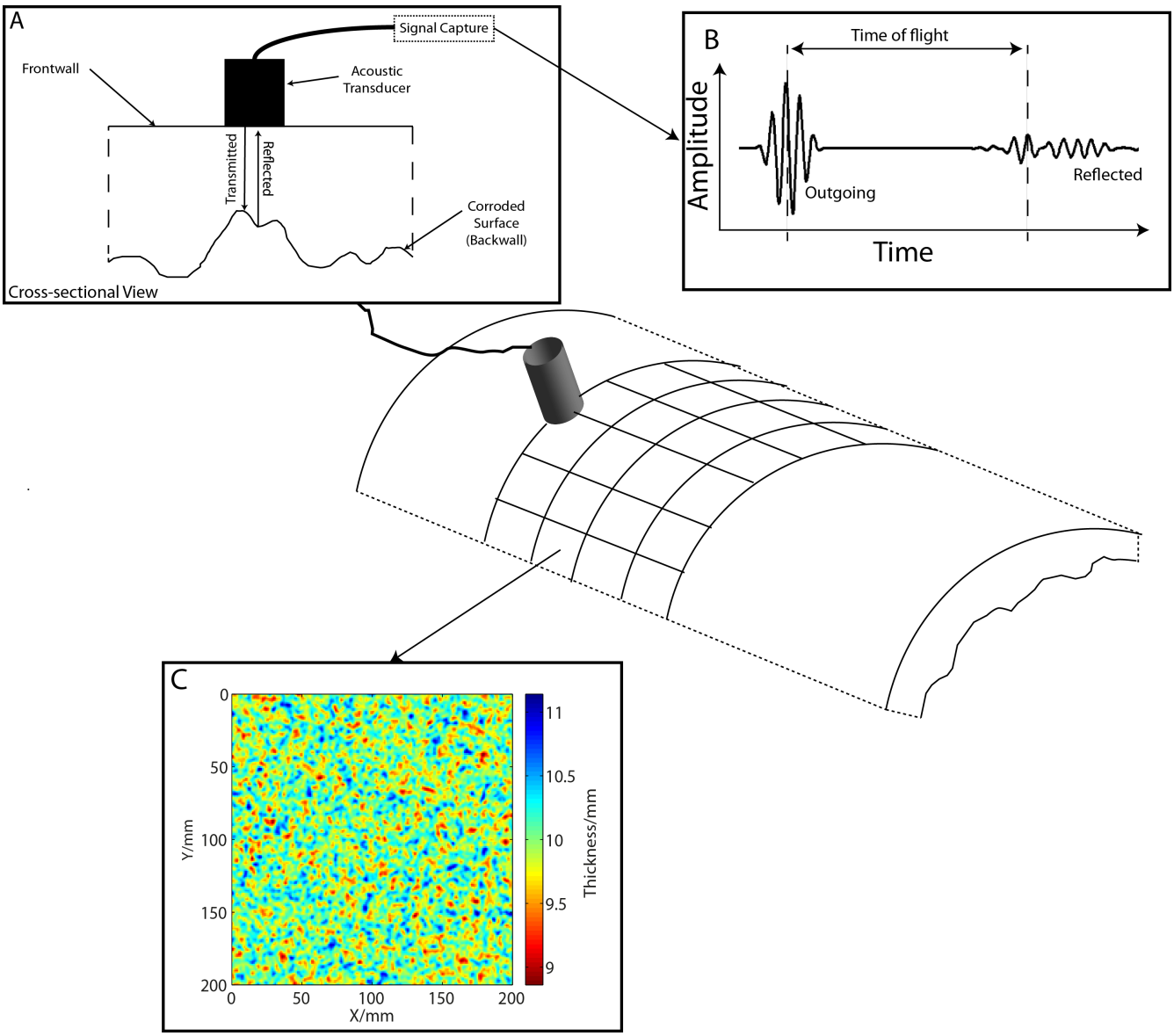


Figure 1.1: An illustration of a C-scan of a section of corroded pipe. The ultrasonic transducer is moved across the surface and at regular spatial intervals, indicated by the grid, an ultrasonic pulse is transmitted into the component. The pulse is reflected from the corroded back wall of the component (A). The reflected signal is recorded (B), the time of flight is extracted from it and used to calculate a measurement of thickness. Thickness measurements from across the inspection area are presented as a thickness map (C).

wave is deployed because it is relatively insensitive to fluid contents within the pipe. Defects can be detected by recording the reflected signal from the defects and examining both its frequency content and amplitude. This approach has been shown to effectively detect defects that have removed a few percent of the cross-sectional area of the pipe over tens of metres [15, 16]. GWT can significantly reduce the time required for an inspection. However, GWT is a screening method that doesn't directly determine the integrity of the pipe: instead it is used to identify locations of concern, and then these are followed up by local inspections using conventional methods of NDE.

Some approaches have removed the need for inspection enforced shut-down periods by permanently installing sensors at positions of interest across the component. Cegla *et al.* describe a permanently fixed sensor which separates the transducer from high temperature environments using a wave-guide [18]. These sensors have seen applications at many facilities around the world [19]. In contrast to guided wave based techniques these sensors are capable of measuring very small changes in wall thickness [20] at a limited number of locations. As the measurements from these sensors are associated with a small area directly under the sensor they may not adequately sample the component to detect localised degradation over a large area.

Both of these approaches increase the efficiency of condition assessment by reducing the amount of time or area required for inspection. However, neither of these approaches are suitable for all applications. Guided wave based methods can inspect large areas of a component, but are insensitive to small changes in wall thickness. Fixed sensors on the other hand are fixed in place, providing high resolution over very limited areas. There is a need for an intermediate between these two techniques; a method that combines sensitivity to wall thickness changes with large enough area coverage [21].

Statistical modelling is a promising approach to achieve this goal. If an inspector only has access to inspection data from a small area it can be used to build a statistical model. This statistical model can be used, assuming that the inspection area is representative of the condition of the component, to extrapolate to the condition of much larger areas of a component, as illustrated in figure 1.2. This approach, known as partial coverage inspection (PCI), has seen limited use in industry (for examples, see Stone [2], Terpstra [22] and Hawn [23]). However, progress towards

widespread use has been limited as there are no reliable guidelines for the statistical extrapolation of inspection data [24]. This thesis describes the development of guidelines for performing partial coverage inspection. In the next two sections, background literature around partial coverage inspection (PCI) is reviewed (section 1.2) and the aims and structure of the thesis are outlined (section 1.3).

1.2 Partial Coverage Inspection (PCI)

Inspection strategies based on limited data are becoming more commonplace in asset management. Certain areas are more safety critical or degradation mechanisms (such as corrosion) are known to be more aggressive in particular parts of the plant. These areas are considered to be at a higher risk of degradation than others. Therefore, to be most economical, asset owners prioritise inspections in these sample areas. This approach is know as risk based inspection (RBI) [25].

Sometimes inspectors can only access a fraction of these areas, due to time or budget constraints. In this situation partial coverage inspection [PCI] (also often called: non-intrusive inspection) can be used to infer the condition of inaccessible areas of the plant [22] (illustrated in figure 1.2). An inspector measures the thickness of a component across an accessible area, which is assumed to be under the same conditions as the inaccessible region. These measurements can be used to form a cumulative distribution function (CDF), which can be used to draw conclusions about the inaccessible area. The measured distribution can differ quite significantly depending on the mode of damage occurring. Typically, general corrosion shows Gaussian or skewed Gaussian wall thickness distributions, while pitting tends to exhibit a double exponential type distribution [2, 26, 27]. PCI is attractive as it has the potential to estimate the condition of very large areas of a component using small samples of data.

There are no existing guidelines as to when PCI is a suitable tool for condition assessment. This is problematic as the suitability of PCI is highly dependent on the nature of the damage occurring. The operator must be confident that the inspection data is representative of the rest of the component. Different parts of a component will be undergoing different modes of degradation, depending on, for example, local pressure variations, pH levels or geometry. For example, if very

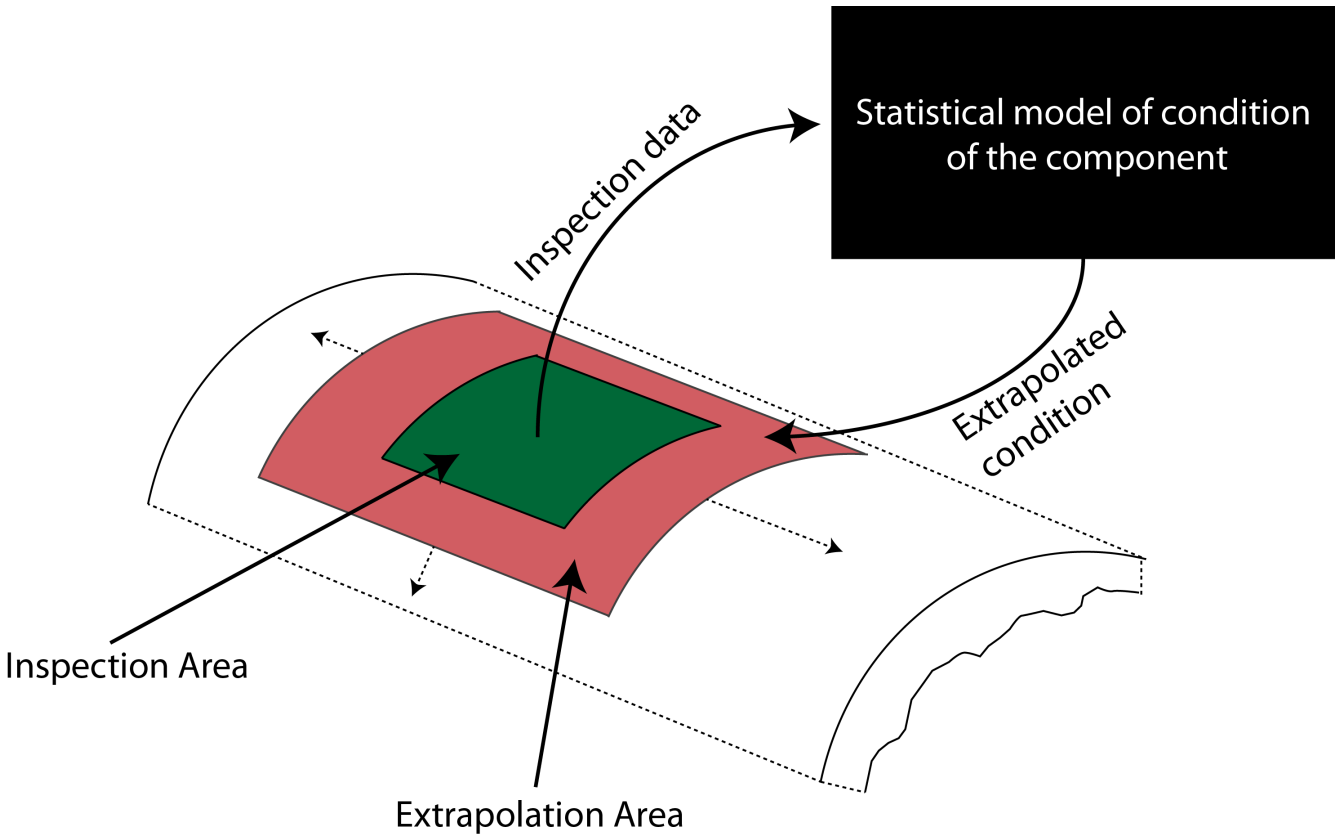


Figure 1.2: An example partial coverage inspection. An inspector collects data from the green area and uses it to construct a statistical model. The statistical model (represented by a black box) is used to extrapolate to the condition of the larger red area. Hypothetically this area could be as large as the entire component.

isolated pitting is present, an inspector could measure a thickness distribution unrepresentative of the rest of the condition of the component.

Furthermore the accuracy and reliability of PCI is dependent on the inspection data collected. Different spatial sampling or inspections of different areas will all produce different sets of data and, subsequently, different conclusions about the condition of the remainder of the component. Round robin tests between groups of different operators on the same flaw showed that differences in inspection teams can lead to up to 25% differences in flaw sizing [28]. In terms of condition assessment, this is less than ideal. It could lead to inspectors drawing different conclusions about the same component.

The central problem with PCI is how to reliably and accurately extrapolate from partial coverage to the full condition of the component. Great strides would be made towards this goal with the development of general guidelines for the use of PCI. Det Norske Veritas (DNV) has published a recommended practice document for the non-intrusive inspection of pressure vessels [3]. In this report, three different damage environments are outlined, these are repeated in table 1.1. Although these have been described for pressure vessels, they can be applied to other components and provide a useful reference for discussing PCI.

Type	**Definition**
A	Degradation mechanism not expected to occur. Inspection is required to confirm there is no onset of the degradation mechanism.
B	Degradation mechanism expected, with low to medium progression. Location of degradation can be predicted. However, it is not anticipated to impact on vessel integrity in the medium term (typically at least two outage periods). Inspection required to confirm corrosion risk assessment predictions.
C	Degradation expected with medium / high progression. Location of degradation can not be predicted. It may impact on vessel integrity in the medium term (typically at least two outage periods). Inspection required to confirm absence of flaws of critical size.

Table 1.1: Definitions of damage environments, as outlined in [3]

One of the keystones of PCI is that the inspection data must be representative of the damage

occurring across the extrapolation area (figure 1.2). This requirement will almost certainly be met for type A degradation. Type A is expected to have no or very slow degradation, only a small inspection area (1-10%) is required to sufficiently sample the damage. This is in contrast to type C corrosion. Degradation is occurring unpredictably, with different modes of damage occurring in different locations. A full inspection is required to ensure their detection [21].

At the moment, PCI is restricted to type A components as many contractors or facility operators are unsure as to when PCI will provide results that they can have confidence in. Applications of PCI could be extended to assess type B components as the degradation in most of these components is relatively predictable with more inspection data (10-50% of the component). The development of a structured approach towards performing PCI, which outlines how to verify whether PCI is suitable, will increase confidence in PCI as a tool for condition assessment [24].

1.3 Outline of the thesis

This thesis aims to outline guidelines for performing PCI with ultrasonic inspection data. The primary focus of this thesis is a method for PCI known as extreme value analysis (EVA). In chapter 2, the motivation for focussing on EVA will be described alongside a review of existing literature. A section in this chapter will describe the theory behind EVA and highlight key assumptions the model makes.

In chapter 3 a method for constructing extreme value models using inspection data is described. The method checks that the assumptions made by EVA are met by the inspection data. This approach is validated using a large amount of numerically generated inspection data.

The main goal of PCI is the extrapolation from an inspection of a sample area of a component to the condition of a larger area. Extrapolation itself is impossible to validate as validation would require data from outside of the inspection area which is typically unavailable to an inspector. However, it is possible to quantify the uncertainties associated with EVA through the calculation of confidence intervals. In chapter 4 a method for calculating confidence intervals around the smallest expected thickness in the extrapolated area is described. This method is validated by comparison to large amounts of numerically generated data.

Extrapolation is sensitive to errors from many different sources. The method for constructing extreme value models ensures that the inspection data will meet the assumptions made by EVA and the calculation of confidence intervals allows an inspector to quantify the uncertainty associated with an extrapolation made from the model. Most inspection data takes the form of an ultrasonically measured thickness map (figure 1.1) measured across a rough surface resulting from a degradation process. It has been shown that ultrasonic thickness measurements across rough surfaces can be inaccurate [14].

Any bias or uncertainty in the inspection data will be magnified by extrapolation. The effect of bias in the inspection data on extrapolation must be well understood to effectively use EVA. Subsequently, a joint computational and experimental study of ultrasonic thickness measurements across rough surfaces was performed. This study is described in chapter 5 and 6.

Chapter 5 describes the distributed point source method (DPSM), which was used to simulate the large number of ultrasonic signals collected as part of a C-scan. The development of a modification to the DPSM, to improve its accuracy when compared to an analytical benchmark, is described. Appropriate parameters for the simulation of a typical C-scan were found with a full convergence study.

The DPSM was used to simulate C-scans of three different rough surfaces. The results of these simulations are presented in chapter 6 alongside the results from experimental C-scans of the same surfaces machined onto steel plates. The simulations and experiments were compared to validate the results of the simulations. Extreme value models were constructed from the data to investigate the effect the use of ultrasonic thickness data has on an extrapolation.

Chapter 2

Extreme Value Analysis

2.1 Statistical modelling of corrosion data

Partial coverage inspection (PCI) is based on the construction of a statistical model from inspection data collected across a sample area (figure 1.2). Statistical modelling is an important tool in many areas of science and engineering. It has been used in applications as wide ranged as quantifying the uncertainty around the output from a measurement tool [29] or predicting the lifetime of engineering components [30, 31]. In NDT specifically, statistical modelling has been used to study the probability of missing critically sized defects in an inspection [32], to study the reliability of inspections [33, 34] and the probability of component failure [35]. A statistical model will reproduce the behaviour of a random variable, providing a probability of a given value of the random variable occurring. For PCI this random variable is the measured thickness of the component (or the wall loss) across the inspection area.

Inspection data usually comes in the form of a thickness map (figure 2.1a). The thickness at each position in the map is represented by a coloured pixel, providing a qualitative overview of the degradation in the inspection area. The thickness map can be converted into a more quantitative presentation of the data by calculating an estimate of the cumulative probability distribution of the thickness measurements. This is calculated by sorting the thickness measurements in ascending order, assigning each thickness measurement a rank and using this rank to calculate the empirical

cumulative distribution (ECDF) function:

$$F(x) = \frac{i}{N+1} \tag{2.1}$$

where x is a measurement of thickness, i is its rank and N is the total number of thickness measurements. $F(x)$ is the probability of measuring a thickness of less than x. An example of a cumulative distribution function generated from the thickness map in figure 2.1a is shown in figure 2.1b. The ECDF is an estimate of the probability of measuring a thickness less than a given value. A number of examples of ECDFs generated from inspection data collected from in service engineering components can be found in [2].

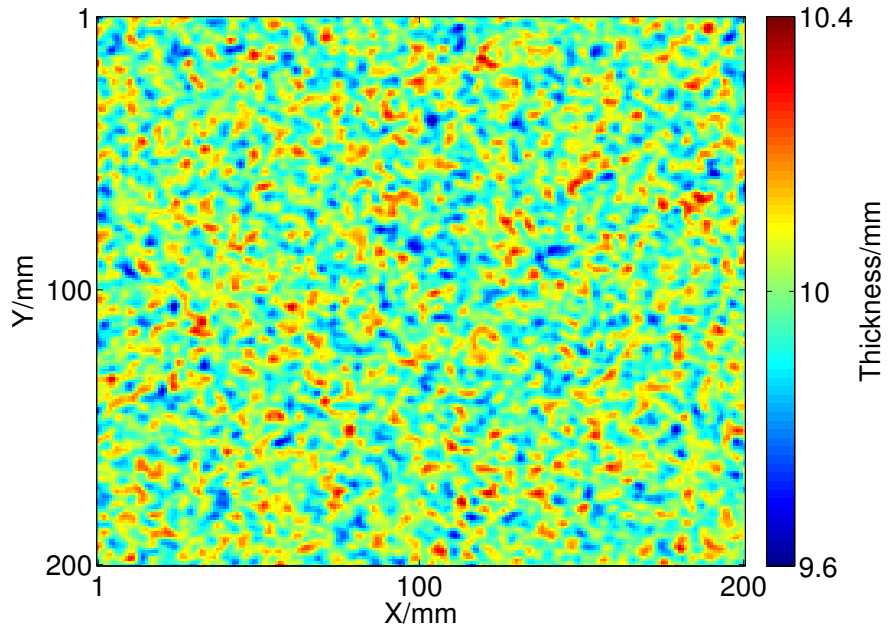
In addition to the empirical estimate of the cumulative probability distribution, it is common practice to fit a probability distribution to the thickness measurements. Thickness measurements can be distributed in many different ways. Their distribution is determined by the conditions the component is subjected to. Different temperatures, pH, and surface conditions can all produce different distributions. There are usually many different degradation mechanisms occurring across the inspection area, consequently, by the central limit theorem, the overall thickness distribution often tends to be Gaussian or Gaussian correlated [36]. However, localised corrosion mechanisms can often produce exponential [27] or more exotic distributions [2].

The Gaussian distribution is defined by two parameters: the mean (m) and the standard deviation (s). The probability of obtaining a measurement of less than x from a Gaussian distribution is given by:

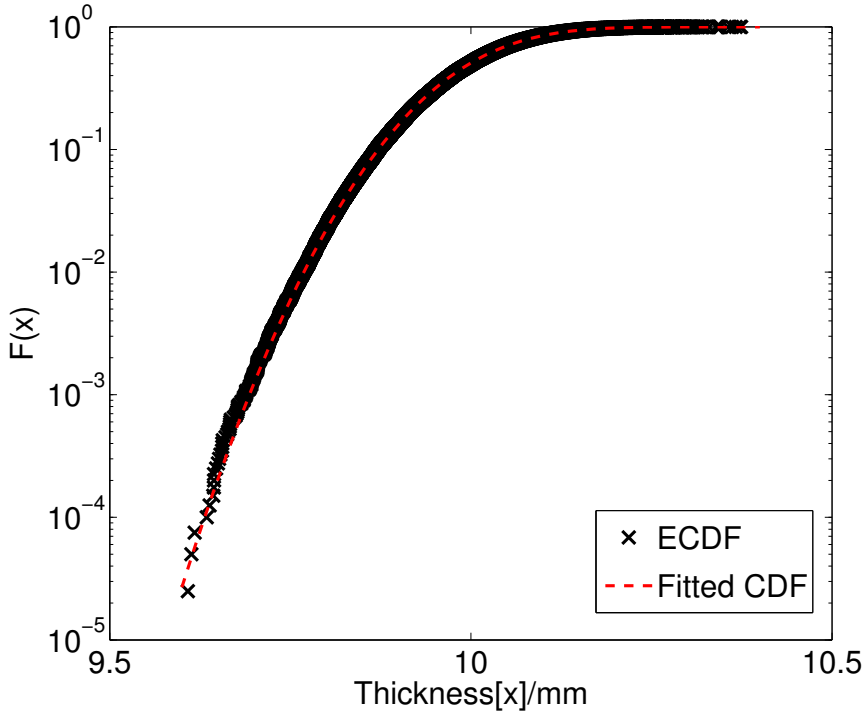
$$F(x) = \frac{1}{s\sqrt{2\pi}} \int_{\infty}^{x} e^{-\frac{(t-m)^2}{2s^2}} dt \tag{2.2}$$

where $F(x)$ is the probability of obtaining a measurement of less than x. In figure 2.1b this distribution has been fitted to thickness measurements using maximum likelihood estimation (MLE) [37, p. 824] and plotted alongside the empirical cumulative distribution function.

MLE is a method of estimating the parameters of a distribution from a given set of data. The method is based on the joint distribution of the set of measurements, conditioned on the parameters of the distribution. For example, for the normal distribution the joint distribution is



(a)



(b)

Figure 2.1: (a) A thickness map of a Gaussian distributed Gaussian correlated rough surface with RMS=0.1mm and correlation length 2.4mm. (b) The empirical cumulative distribution function extracted from the thickness map above. The black crosses are the estimated values of the cumulative distribution function while the red dashed line is a Gaussian distribution that has been fitted to the data.

given by:

$$f(x_1, x_2, x_3, ..., x_N|m, s) = f(x_1|m, s)f(x_2|m, s)f(x_3|m, s)...f(x_N|m, s) \tag{2.3}$$

where m and s are the mean and standard deviation of the distribution and x_i are a set of observations. Alternatively, this can be though of in terms of the likelihood function, which is defined as:

$$\mathcal{L}(m, s|x_1, x_2, x_3, ..., x_N) = f(x_1, x_2, x_3, ..., x_N|m, s) \tag{2.4}$$

where $\mathbb{L}$ is the likelihood of the parameters μ and σ given x_i. The larger the likelihood of a pair of parameters (μ,σ), the better the fit of model to the data.

Maximisation of the likelihood will provide a best estimate for the parameters given the observed data. In practice the likelihood is normally expressed as the log-likelihood as it is often easier to work with:

$$L(\mu, \sigma|x_1, x_2, x_3, ..., x_N) = \log(\mathcal{L}(\mu, \sigma|x_1, x_2, x_3, ..., x_N)) = \sum_{n=1}^{N} f(x_i|\mu, \sigma) \tag{2.5}$$

as the logarithm is a monotonic function, maximisation of equation 2.5 is equivalent to maximisation of equation 2.4. For the normal distribution the log-likelihood function is given by:

$$L(\mu, \sigma|x_1, x_2, x_3, ..., x_N) = \frac{-1}{\sigma\sqrt{2\pi}} \sum_{i=1}^{N} \frac{(x - \mu)^2}{2\sigma^2} \tag{2.6}$$

Maximisation of this function will provide estimates for the parameters of the normal distribution given the set of data. There are alternative methods to maximum likelihood estimation. such as the method of moments and least squares regression. However, in this thesis, MLE estimation is used exclusively as maximum likelihood methods are very flexible and they can be used to estimate distributional parameters for a large variety of problems.

The Gaussian distribution provides a good fit to the set of data, as shown by the red dashed line in figure 2.1b. Both the ECDF and the fitted distribution are models for the thickness measurements in the inspection area (blue area in figure 1.2). Both models provide information on the average behaviour of the thickness measurements. However, condition assessment is most

concerned with assessing the largest extent of the wall loss in a component. From a constructed model, the probability that x is the smallest thickness in an area the size of the inspection area is $F'(x) = 1 - F(x)$. Consequently, the probability that x is the smallest thickness in an area N times the initial inspection area is given by:

$$F'_n(x) = F'(x)^N = (1 - F(x))^N \tag{2.7}$$

where $F'_n(x)$ is an empirical distribution of the smallest thickness in an area N times the inspection area. Hypothetically, a fitted distribution such as equation 2.2 could be used with equation 2.7 to compute the exact analytical form of the distribution of the minimum thickness in a given area. However, any small inaccuracy in $F'(x)$ will be magnified in $F'(x)^N$.

An alternative approach is to accept that F' is unknown and to look for a limiting form of the distribution F'^N. Extreme value analysis describes such an approach. It shows that, under certain assumptions, the distribution of F'^N will be a generalised extreme value distribution (GEVD). In this chapter, the theory behind extreme value analysis (EVA) will be described with a review of existing literature concerning its application to inspection data. This will be followed by a simple example of the process of generating an extreme value model from a sample of simulated inspection data.

2.2 Extreme value analysis (EVA)

The study of extreme values is a well-developed topic, finding applications to topics as varied as finance [38, 39], structural design [40, 41, 42], environmental modelling [43, 44] and even the assessment of risk of terrorist attacks [45]. Its use for analysis of corrosion data is discussed in [31]. Kowaka provides a number of examples of the use of extreme value analysis to extrapolate from C-scans of reduced areas of a plant to larger areas of components. Further examples are provided in [46, 23, 26, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57].EVA gained limited popularity in Japan in the 1970s. However, due to the lack of available computational power, processing large amounts of C-scan data was infeasible, which led to it falling out of favour.

In recent years, the use of an extreme value approach has regained popularity. A report prepared for the Health and Safety Executive (in the UK) assessed current available methods and the barriers to their adoption. It concluded that there are readily available statistical methods for the analysis of corrosion data. However, these methods are not used due to poor dissemination to engineers and the lack of any readily available computational tools [24]. The report also mentions that extreme value theory can be used to extrapolate from sample inspection data to assess large areas of a plant. This extrapolation technique was first developed in the 1970's. Hawn made probabilistic assessments of the condition of an underground pipeline. He used EVA to show that the size of corrosive pits would not exceed a given limit in an area 5820 times the initial inspection area [23].

Previous examples of EVA take advantage of the Fisher-Tippet-Gnedenko theorem [58] which provides a limiting form for equation 2.7 (the probability of obtaining a minimum thickness x in an area N times the inspection area). The theorem shows that in the limit $N \to \infty$ equation 2.7 is a generalised extreme value distribution (GEVD):

$$\Phi(x|\mu,\sigma,k) = 1 - \exp\left\{-\left[1 + k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/k}\right\} \tag{2.8}$$

where μ is the location parameter, which determines the size of the minima; σ is the scale parameter, which determines the spread of the minima; and k is the shape parameter, which determines the shape of the distribution. The shape parameter can loosely be thought of as controlling the skewness of the distribution, that is the difference between the mean and the median of the distribution. $\Phi(x|\mu,\sigma,k)$ is the probability of measuring a thickness minimum of less than x. A full discussion of a suitable size for N for equation 2.8 to be a valid model can be found in chapter 3. Examples of GEVDs are shown in figure 2.2.

Equation 2.8 can be used as a model for the smallest thickness measurement in a prescribed area. However, for this to be possible, an inspector must calculate values of μ, σ and k. With a standard statistical model this can be accomplished by fitting the model to the collected data. From the authors experience thickness minima are usually modelled using distributions with $k \leq 0$, real examples of this can been found in Hawn [23, 27, 46]. The derivation of equation 2.8 is based on

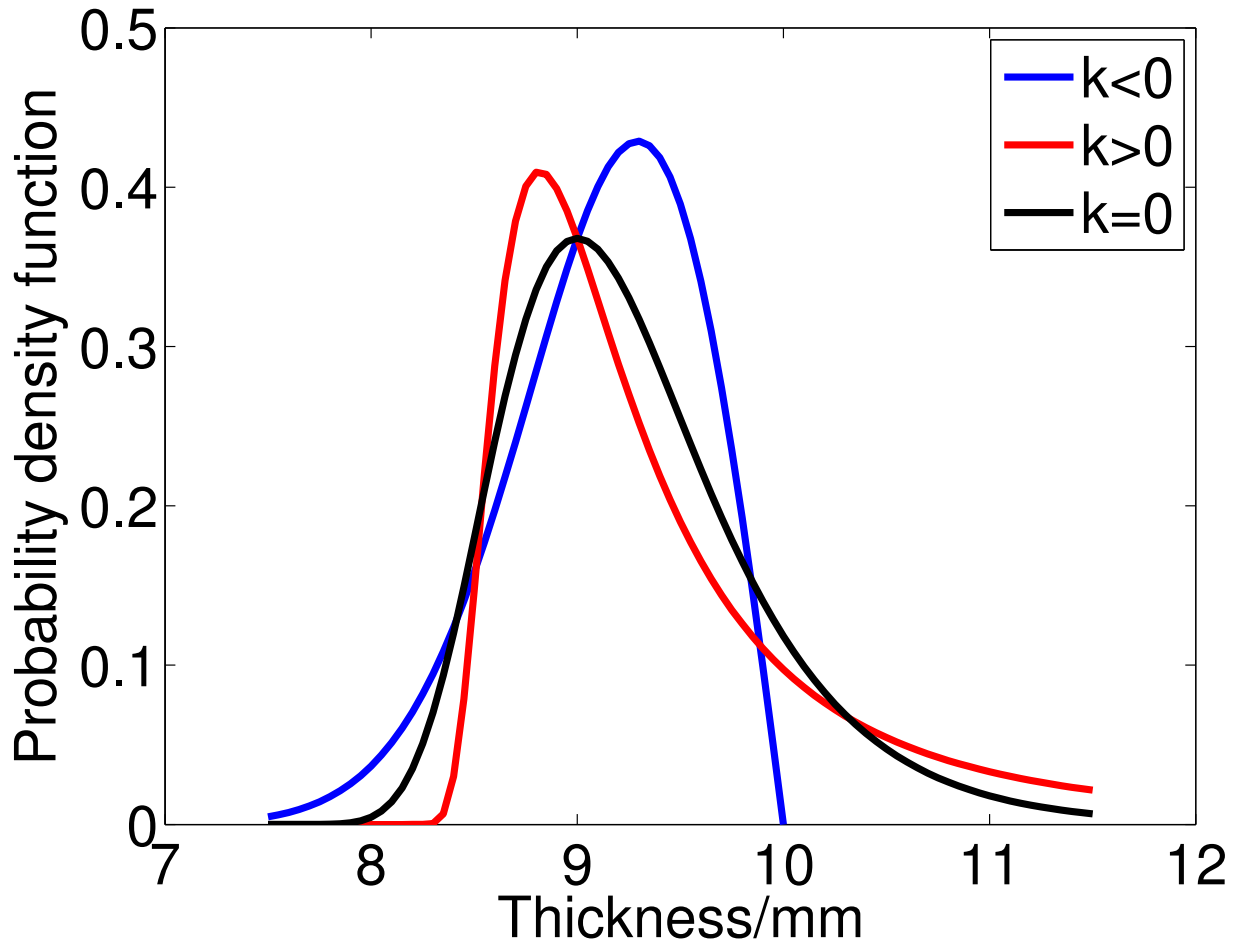


Figure 2.2: Three examples of the probability density function of the generalized extreme value distribution with different values of the shape parameter(k). The blue, red and black lines are the probability density functions for distributions with shape parameters of $k = -0.5, 0, 0.5$ respectively. Each distribution has the same location (μ) and scale (σ) parameters. The undamaged thickness of the component in these examples was 10mm. From the authors experience, thickness minima can usually be modelled by distributions with $k \leq 0$. An example of a distribution with $k > 0$ has been included here for completeness. This type of distribution could occur, for example, if a process in the component was causing a build up of deposits. In this thesis components only undergoing wall loss are considered so distributions with $k > 0$ are not discussed.

the convergence the sequence of renormalised thickness measurements. The shape of distribution is determined by the the way this sequence converges, which depends on the family of distributions the thickness measurements belong to. The thickness distributions studied in this thesis are restricted to be Gaussian and exponential, which both lead to extreme value distributions with $k \leq 0$ [41]. In practice, it would be unusual for a GEVD with $k > 0$ to occur in components undergoing only wall loss as this would imply wall thickness growth.

An extreme value (EV) model is a model for the minimum thickness in an area as opposed to the thickness. An inspector requires a sample of thickness minima to calculate values of μ, σ and k. A sample of minima can be extracted from a number of different thickness maps from different areas [23], or from a single thickness map [31]. Performing multiple inspections to obtain different thickness maps can often be infeasible for the reasons discussed in chapter 1. As a consequence, selecting a sample of minima from a single thickness map is often the best option for PCI applications. A sample of minima can be selected from a thickness map by dividing the map into a number of equally sized blocks (shown in figure 2.3). The minimum thickness is selected from each of these blocks. These measurements form a set of minima which can be used to calculate the parameters in equation 2.8. An example of a histogram of a set of 100 thickness minima extracted from a correlated Gaussian surface with RMS height $0.1mm$ and correlation length $2.4mm$ is shown in figure 2.4. Each thickness measurement is represented by a bar, the height of which is proportional to the frequency of the measurement occurring. A GEVD, which has been fitted to the set of minima, is shown by the red line. The GEVD provides a good description of the data.

2.3 Discussion

Currently, EVA is not widely used with inspection data for a number of different reasons. At the time of writing, there are no guidelines as to when EVA will be a suitable tool for analysing inspection data. EVA makes the assumption that the thicknesses in each block used to extract the minima are independent and identically distributed. DNV's classification scheme (summarised in table 1.1) provides a useful point of reference when discussing this [3]. It should be kept in mind that this classification system, although it is in use in the field, should not be taken as an absolute

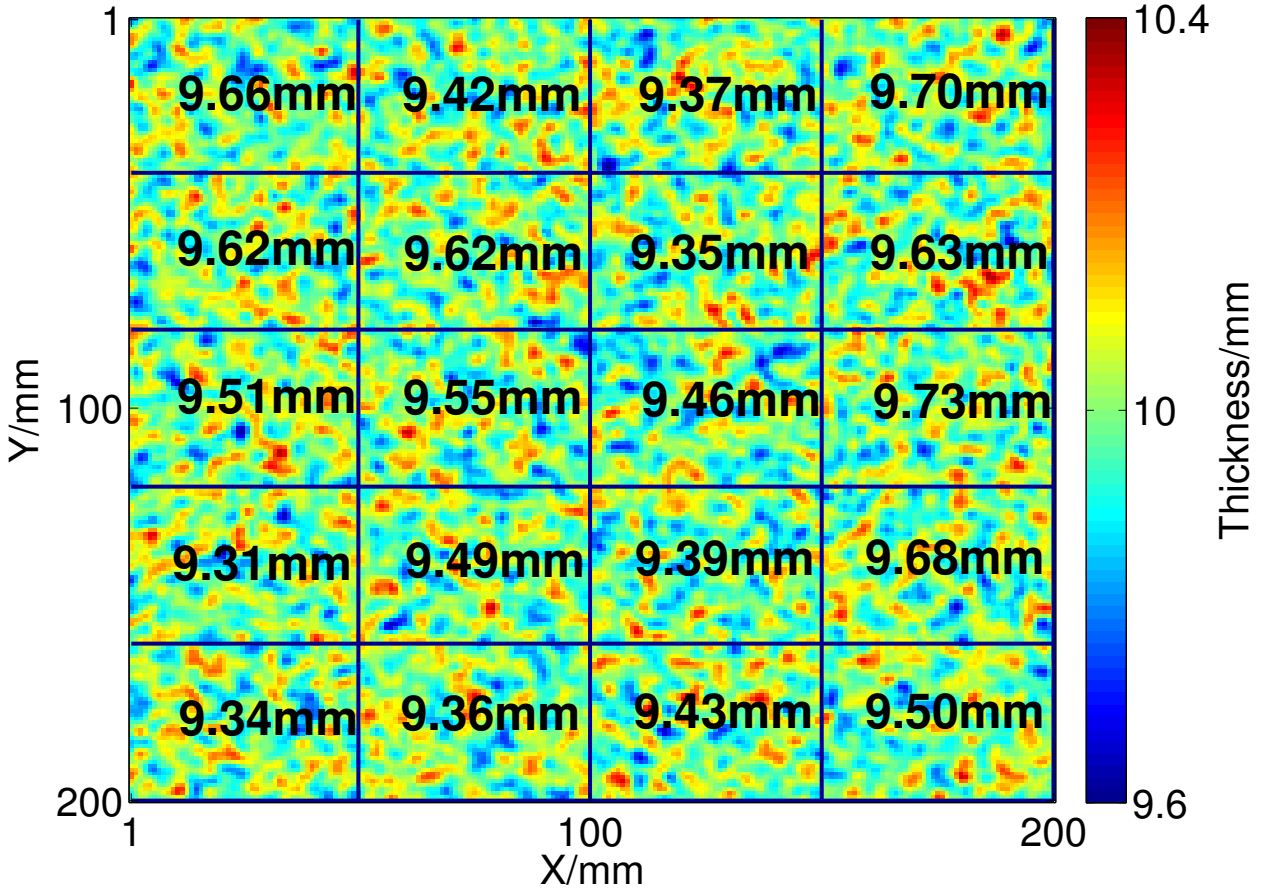


Figure 2.3: An example of a thickness map split into 25 blocks. From each block the minimum thickness (highlighted in red) is selected. This set of measurements form a sample of minima for the construction of the EV model.

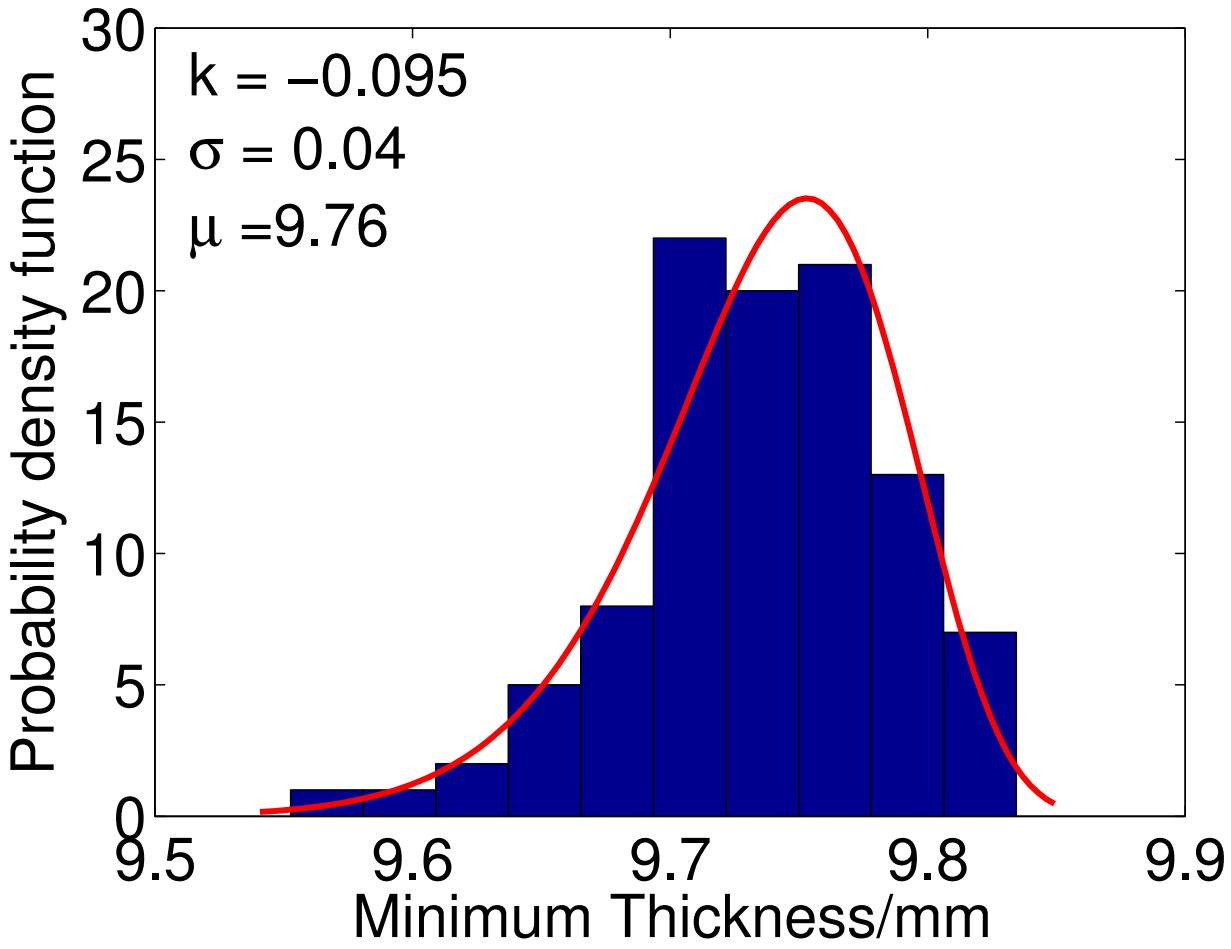


Figure 2.4: A histogram of a set of 100 thickness minima extracted from a correlated Gaussian surface with RMS height $0.1mm$ and correlation length $2.4mm$. Each thickness measurement is represented by a bar, the height of which is proportional to the frequency of the measurement occurring. The extreme value model which has been fitted to this set of minima is shown by the red line. The distributional parameters are shown in the top left corner.

authority on the types of damage that could occur. There is much more subtlety in the types of damage that could occur than can be described by categorical classification. However, for the purposes of describing the situations in which EVA it is a useful tool.

For some types of degradation the damage mechanisms will be slow and uniform across the entire component (type A in DNV's guidelines [3]), while some types of degradation can be aggressive and unpredictable (type C in DNV's guidelines [3]). Type B degradation encompasses most of the types of damage in-between. The degradation is usually progressing at a low to medium rate and any localised areas of degradation occur at predictable locations.

An inspector would be able to use EVA to build a model for type A degradation as EVA's assumptions will be met in the inspection area and the assumption is made that this area is representative of the remainder of the component. In contrast, it would be unsuitable for type C degradation as EVA's assumptions will not be satisfied by the inspection data nor will it be likely that the inspection area is representative of the remainder of the component. For type B, the decision is less clear, as different modes of damage can occur in different areas the inspector must have further information about the vessel in order to be sure EVA is a suitable tool.

Without suitable information about the types of damage occurring across the vessel an inspector can not be confident in the extreme value model regardless of the damage classification. The type of damage that is occurring can change over time or may not have been correctly identified in the first place. This can lead to the inappropriate classification of a component. This is a particular issue with type A damage as the assumption is made that the damage is occurring slowing and uniformly. Over time this could change. Small isolated defects could occur due to, for example, changes in operating conditions. Therefore any PCI inspection, particularly a scheme based on EVA, must be accompanied by screenings of the full vessel in order to confirm whether the assumptions made for PCI are valid. At the moment an inspector relies on their judgement to determine when this is the case, which is not ideal as it can be prone to human error. For an inspector to be truly confident in an EV model guidelines to determine when it is a suitable tool are required.

In addition to this there is no established method for selecting a block size. This is a key step in constructing an extreme value model. To generate a model one requires a sufficient sample size,

which requires a smaller block size. However, the smaller the block size the more likely that the thickness selected will not be an example of an extreme of the distribution. Current methods for selecting block sizes focus on examining the fit of the GEVD to the set of minima selected using that block size [59, 60]. However, these methods often do not ensure that the underlying thickness measurements satisfy the assumptions made by EVA.

There is a clear need for a structured approach to constructing EV models from inspection data. Such an approach will ensure that the assumptions made by EVA are met by the data and that an appropriate block size is selected. In the next chapter the development of such a framework is detailed, outlining an approach for selecting block sizes by ensuring that there is evidence the thickness measurements in each block are independently and identically distributed. Following this, chapter 4 investigates the uncertainties associated with an assessment of the minimum wall thickness of a component based using EVA.

Chapter 3

Block Size Selection

3.1 Introduction

A major problem with existing applications of EVA to corrosion data is that the analysis is dependent on the judgement of the analyst and does not necessarily check that the data is suitable for EVA (i.e. they do not check that there is evidence the assumptions made by EVA are fulfilled). For example existing methods for selecting a suitable block size have focussed on examining the fit of the GEVD to the set of minima selected using that block size. Glegola selected a block size by extracting sets of thickness minima using multiple block sizes [59]. For each set of minima the quality of the fit to the GEVD was examined and the block size which gave the best fit to the GEVD was used for the analysis.

Figure 3.1 shows the probability plot of two different extreme value models extracted from a 200 by 200mm correlated Gaussian surface using block sizes of 20mm (red crosses) and 40mm (blue crosses). Empirical cumulative distribution functions for each set of minima were calculated using equation 2.1 and plotted on a scale such that the points will lie on the black dashed line if they are a good fit to the extreme value model. Both of these models capture the majority of the thickness minima well; the majority of the thickness minima in each set lie on a straight line. For the 20mm block size R^2=0.95, while for a block size of 40mm, R^2=0.97. This suggests that the 40mm case is a better model for the thickness minima. However, as the difference is so slight, this metric cannot be used in isolation to make a decision on the most suitable block size. An inspector

must use their judgement to make a decision.

The EV model constructed using a $40mm$ block size models the smallest thickness measurements more accurately. For EVA of inspection data, these are the most important measurements. Therefore, in this case, the inspector should use a block size of $40mm$. As a general rule, the larger the block size, the better the model will capture the behaviour of the smallest thickness minima, assuming the underlying thickness distribution meets the assumptions made by EVA.

An EV model may show a good fit to the majority of the data, but if it does not accurately model the smallest measurements then it is not a good model for condition assessment purposes. Model fit does not provide a perfect method for selecting a block size to construct an EV model for condition assessment, but it can be used alongside the inspector's judgement to build an effective model. Schneider described an alternative approach. He selected a block size to ensure that the minima from each block were independent [52]. Schneider examined the two dimensional autocorrelation function of the thickness map and chose a block size, L, such that thickness measurements separated by L were weakly correlated. In contrast to Glegola's method this approach chooses a block size based on one of the assumptions of EVA. However, in addition to the independence of thickness measurements, EVA also assumes that the probability distribution of thickness measurements in each block is identical.

Ensuring that there is evidence that both of these assumptions are met is key to implementing an extreme value model for partial coverage inspection. Thickness maps from corroded components are often complex as they undergo damage from different modes (e.g. pitting corrosion as opposed to uniform corrosion). Each damage mode will produce a different thickness measurement distribution. If one naively builds an extreme value model, the complexity of the surface can violate the assumptions of an extreme value model. Due to the risks involved in drawing conclusions from an inappropriate model (e.g. component failure leading to loss of life), an analyst should gather a body of evidence which supports their choice of model. This will reduce the chance that an inappropriate model will be used for PCI.

In summary, the current state of the art offers methods for building extreme value models from inspection data which can lead to subjective models as they rely on the judgement of the

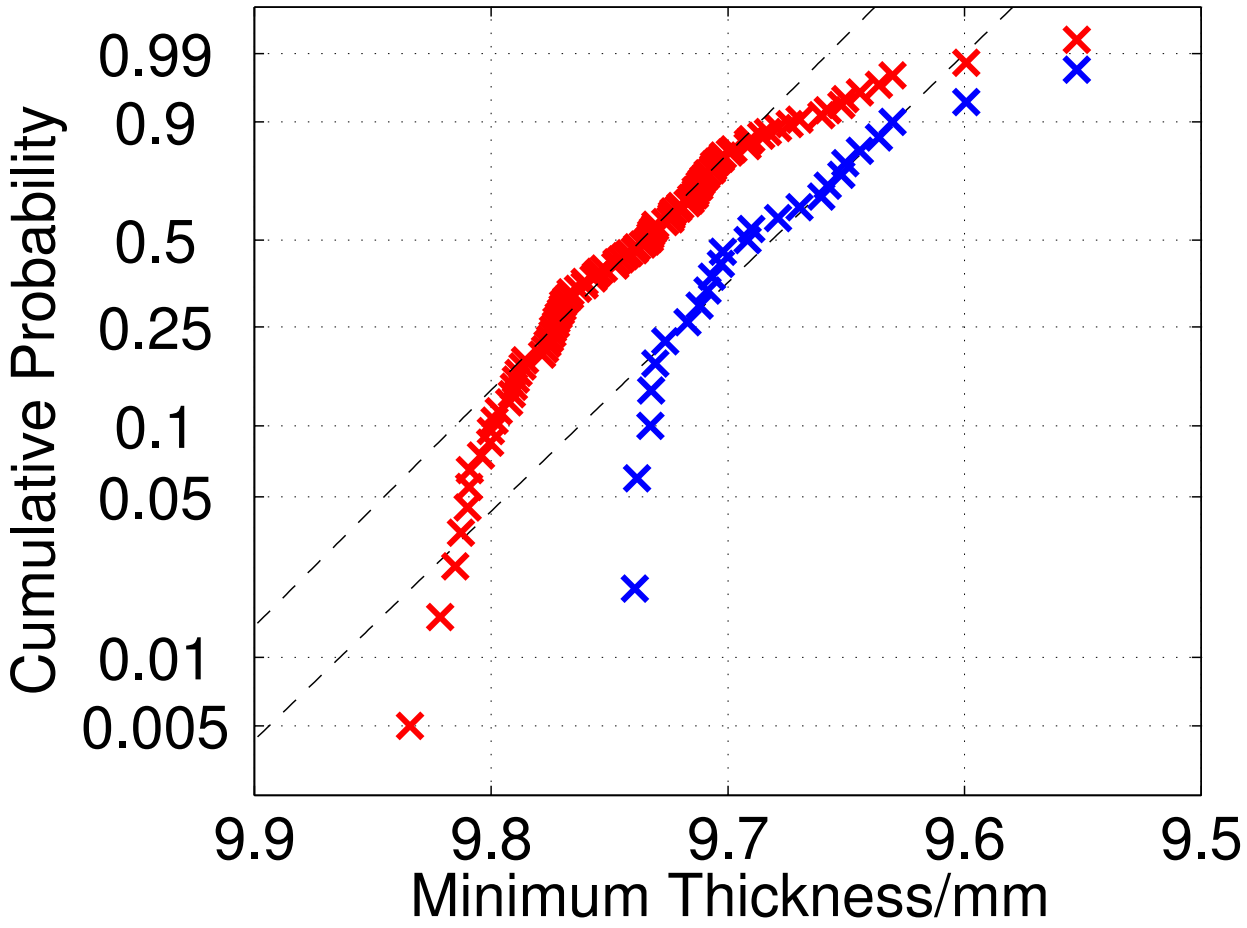


Figure 3.1: An extreme value probability plot of two sets of minima extracted from a 200 by 200mm correlated Gaussian surface with mean thickness 10mm, RMS=0.1mm and correlation length 2.4mm using block sizes of 20mm (red crosses) and 40mm (blue crosses). The scale of the axes has been designed such that, if the model provides a good description of the data, then the crosses will lie on a straight line (the black dashed line drawn through each set of points). For the 20mm block size the R^2=0.95, while for the 40mm block size was R^2=0.97.

analyst rather than compliance to an objective set of requirements. In this chapter a data analysis procedure is introduced that checks that all of EVA's assumptions are met and which an analyst can refer to when developing extreme value PCI models. This is a novel contribution to the field, there is currently no established framework for the construction of extreme value models from inspection data. This chapter begins with a description of the framework in which EVA can be applied to ultrasonic thickness maps of corroded engineering components (section 3.2). Section 3.4 presents evidence that the presented approach yields sensible results when applied to simulated surfaces and data acquired by ultrasonic measurements. This is followed by a discussion of these results and conclusions (section 3.6). The majority of this chapter has been submitted to *Corrosion Science* as part of a paper.

3.2 Blocking procedure to check EVA's assumptions are met (Blocking algorithm)

Schemes to partition thickness maps must tread the line between ensuring there are a sufficient number of sample minima and that the thickness measurements selected are extreme deviations from the median thickness. Too large a block size and there will not be enough minima to extract parameters for the GEVD; too small and the sample will not be representative of the extremes of the thickness distribution. An effective scheme will balance these requirements whilst ensuring that there is evidence that the assumptions made by EVA are met by the inspection data.

In this light, this chapter describes a framework for checking that all of EVA's assumptions are met prior to building a model. First, we check the independence of the underlying thickness measurements by calculating the autocorrelation function of the thickness map:

$$C(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(x, y) T(x' - x, y' - x) dx' dy' \tag{3.1}$$

where $C(x', y')$ is the correlation between a thickness measurement $T(x', y')$ and $T(x, y)$. $C(x', y')$ is a two-dimensional surface reflecting the fact that the thickness map spans two horizontal di-

mensions, described by the x and y coordinates.

The surfaces studied in this thesis have been restricted to isotropic autocorrelation functions. Consequently, the autocorrelation function for all of the test cases is symmetric in the x and y directions and all the information about the correlation structure can be obtained from $C(x', y' = 0)$. This is easily expandable to non-isotropic surfaces as the autocorrelation function can be calculated in any direction. Figure 3.2 shows an example of an autocorrelation function calculated from one of the test surfaces. The ordinate axis is the correlation between two measurements, where 1 indicates perfect correlation and 0 indicates no correlation, while the abscissa is the distance between the pair of measurements. Measurements separated by smaller distances are more correlated, which indicates that measurements closer together are likely to be interdependent.

It is common practice to define a correlation length λ_c for a surface, which is defined as $C(\lambda_c) = exp(-1)$. Using the correlation length, one can define a distance at which one can ensure two measurements are uncorrelated and likely to be independent. Figure 3.2 shows that the autocorrelation function of the surface has dropped to zero at a distance of $2\lambda_c$, therefore it is imposed that measurements must be at least $2\lambda_c$ apart in order to guarantee that data points are uncorrelated. Once a correlation length has been calculated, the surface is partitioned into a number of equally sized blocks. For the $200mm$ square surfaces studied in this chapter, we chose block sizes ranging from 10 to $60mm$. Starting with the smallest block size, a random sample of thickness measurements, including the minimum thickness measurement, is selected from every block. The sample is chosen such that every thickness measurement is separated by $2\lambda_c$, which ensures that the thickness measurements in the sample are independent of each other. The algorithm then checks the random samples from every pair of blocks are from the same underlying thickness measurement distribution using a two sample Kolmogorov-Smirnov (KS) test [61].

The two sample Kolmogorov-Smirnov test is a non-parametric hypothesis test for whether two random samples come from the same distribution. For a pair of blocks, the algorithm calculates the ECDFs for a random sample of thickness measurements from each block, an example of which is shown in figure 3.3. The largest vertical distance, denoted by D in figure 3.4, is extracted from between pair of ECDFs. This is the test statistic for the two-sample Kolmogorov-Smirnov test.

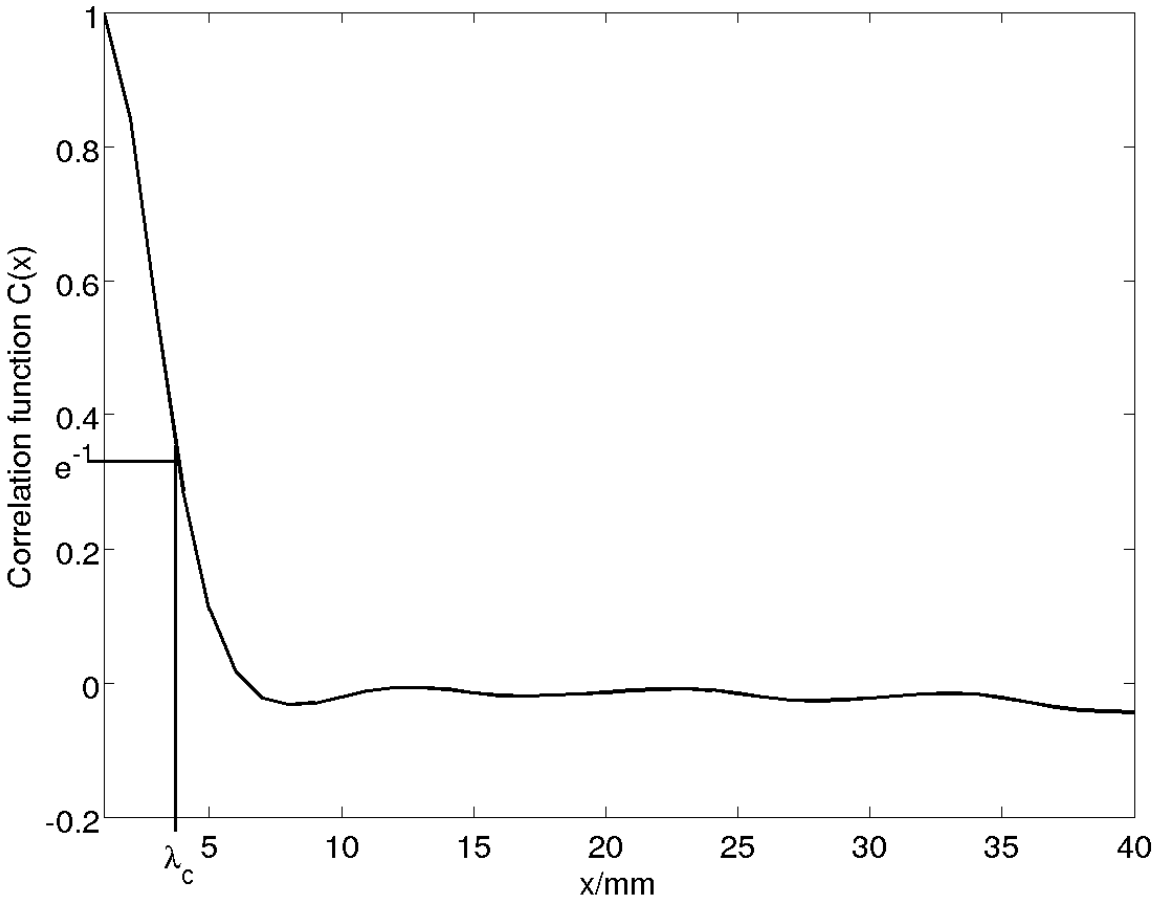


Figure 3.2: Autocorrelation function of a Gaussian surface generated with RMS=0.2mm and correlation length 2.4mm, showing the estimated correlation length λ_c and that the function drops to nearly zero at a distance of $2\lambda_c$.

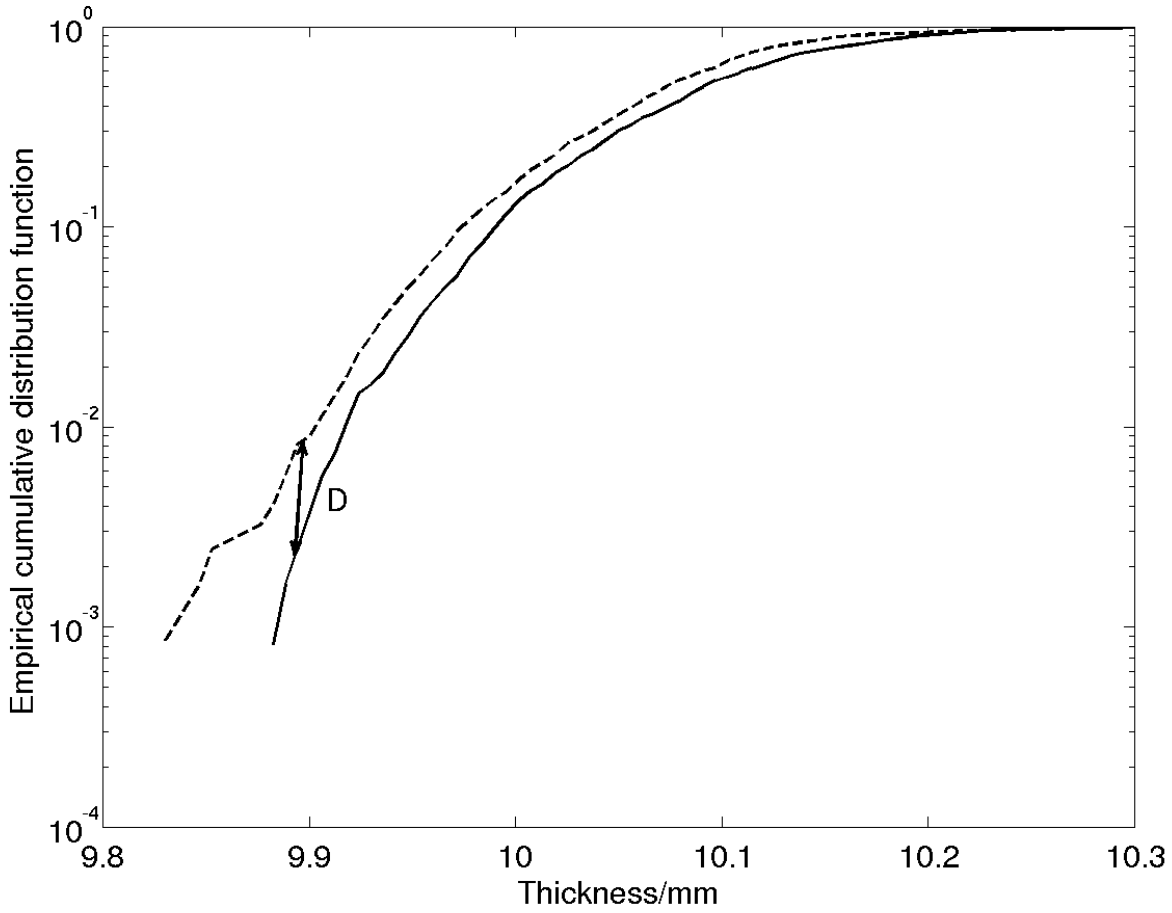


Figure 3.3: A pair of ECDFs generated from adjacent blocks. The Kolmogorov-Smirnov test extracts the largest vertical distance between the ECDFs and compares it to a critical value.

If one assumes that both samples are from the same probability distribution, the probability distribution of the test statistic can be calculated [61]. This is particularly useful as it provides information on how likely the difference between a pair of ECDFs given that they represent the same distribution. For example, using this distribution, an inspector can specify a level at which the distance between the between the two ECDFs becomes large enough that they deem it to be significant.

Figure 3.4 shows the probability density function of D. Integrations of this curve give the probability of measuring different values of the test statistic. If the probability of measuring D is high, the thickness measurements are from the same distribution, if it is low it's unlikely the samples are from the same distribution. This is formalised by defining a null and alternative hypothesis:

$$H_0 : \text{The thickness distributions are the same.}$$

$$H_a : \text{The thickness distributions are not the same.}$$

along with a user specified significance level, which is the probability at which the user deems it unlikely that the differences due to sampling variability (p in figure 3.4). This user specified significance level is shown by the red region of figure 3.4. If the test statistic lies in this region then the test would reject H_0 and the measurements would be deemed to be from different distributions.

Selection of a significance level is an important part of this process. Too large a significance level and the user risks rejecting samples which are from the same distribution, while with too small a significance level there is a risk samples from different distributions will be selected. Significance level selection is not a trivial matter as suitable levels of significance can vary from application to application. In this thesis, significance levels of 0.01 and 0.05 are used, as these are commonly used in statistical literature. These correspond to values of D that would occur only 1% or 5% of the time if the samples arose from the same distribution.

The algorithm performs a two-sample KS test on the random samples from every pair of blocks. If a single pair of blocks fails the two sample KS test, then the algorithm increases the block size and repeats the blocking process. Otherwise, if every pair of blocks does not fail the two sample

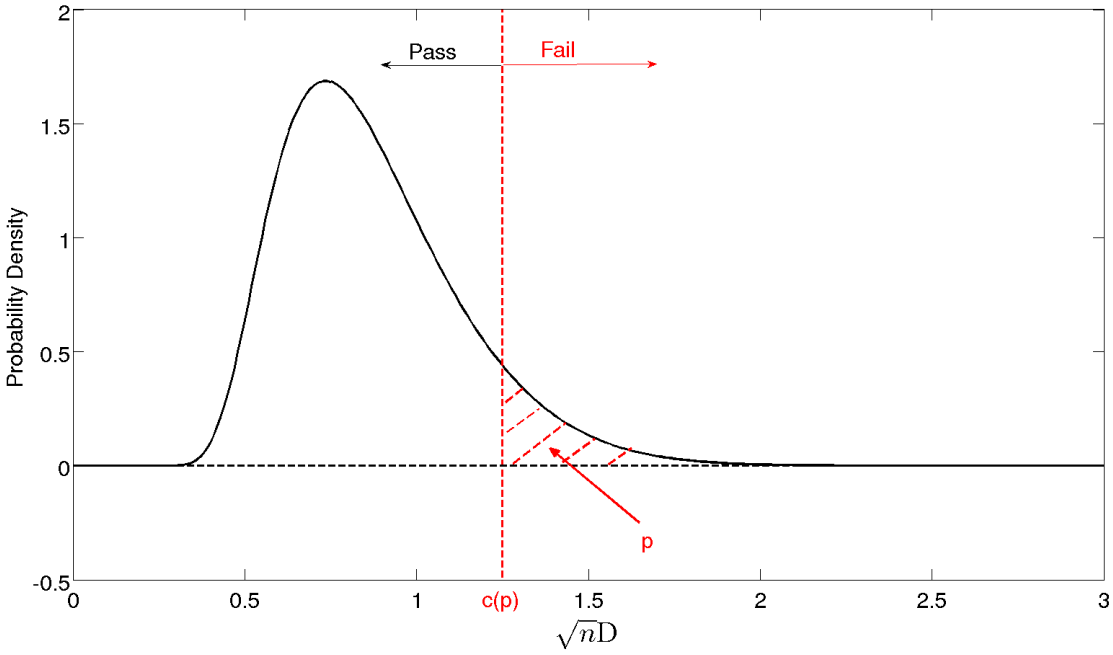


Figure 3.4: Kolmogorov probability density function and illustration of a a hypothesis test to check if two sample are from the same distribution.

KS test, the algorithm has found a block size for which there is evidence the distribution in each block is identical. The sample of thickness minima extracted using this block size can then be used to build an extreme value model for the thickness map. The parameters for the GEVD are then extracted from the sample of minima selected by the algorithm using maximum likelihood estimation (MLE) [62].

In summary, figure 3.5 shows a graphical form of the proposed algorithm. For the smallest block size, the thickness map is split into equally sized blocks and a two-sample Kolmogorov-Smirnov test is performed on the distributions from every pair of blocks. This tests that the thickness measurements in each block are from the same distribution. If the tests show that the thickness distribution is the same in every block, the algorithm terminates and this is the correct block size. Otherwise, the algorithm repeats this process for the next largest block size until there are no remaining block sizes and we conclude that the inspection data is not suitable for EVA. In this way the algorithm is selecting a block size by looking for evidence that the partitioned data meets the assumptions made by EVT.

3.3 Simulation set up

Surfaces produced by corrosion are generated by a large number of independent, random events. Although the height probability distributions of these events may be very different, the overall distribution will tend to a Gaussian [36]. There is strong evidence in the literature to support this claim, for surfaces generated by general uniform corrosion [63, 36, 2, 64]. For more localised corrosion, the height distribution can follow exponential distributions [27, 26, 2]. The type of corrosion that can be expected depends on the damage mechanism that is most likely to occur in the vessel, for the internal operating conditions (temperature, pressure) and chemical conditions (pH, species present). For the purposes of this study, we restrict ourselves to surfaces generated by a general uniform corrosion mechanism.

To demonstrate that a Gaussian profile is representative of the type of damage which should be expected in the case of general corrosion, surface profile measurements were taken from a pipe sample. The sample was retired from a unit exposed to high temperature sulfidation corrrosion

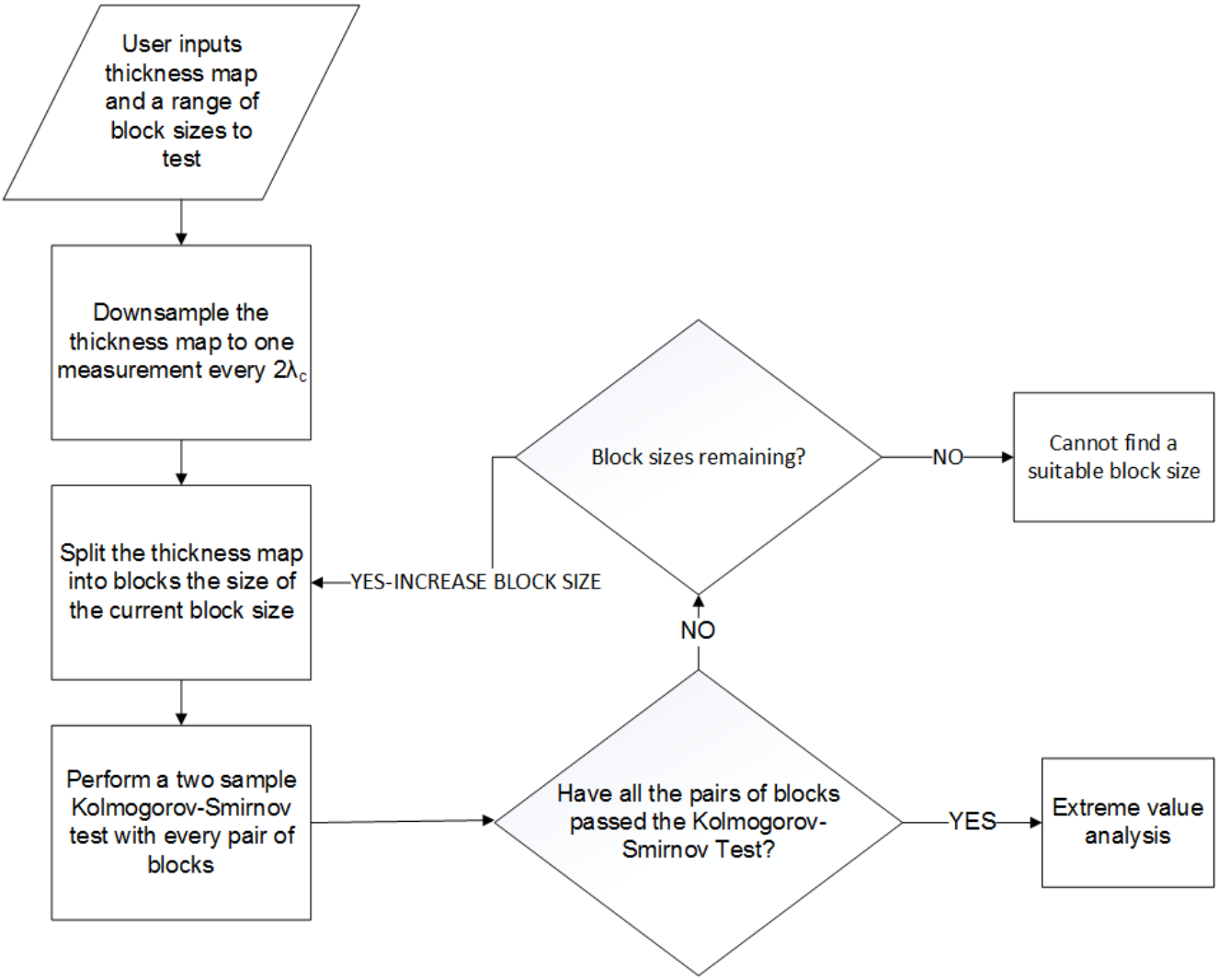


Figure 3.5: A flow chart of the proposed blocking method. A user will input an ultrasonic thickness map and the algorithm will split it into a number of equally sized blocks. The blocks will then be tested to ensure that there is evidence they meet the assumptions made by extreme value analysis.

(a uniform corrosion mechanism). A Talysurf^{TM}surface profile measurement instrument [65] was used to measure the profile. Talysurf^{TM}draws a stylus across the surface of the material, measuring the deflection of the stylus with an interferometer, extracting the surface profile. Several height profiles from around the inner radius of the pipe were measured. An example of a histogram of the measured heights is given in figure 3.6.

Rough surfaces were generated using the algorithm described by Oglivy [63]. Other approaches exist [66], however Oglivy's was chosen due to it's computational simplicity. Oglivy's method performs a moving average on a set of uncorrelated random numbers, which produces a set of Gaussian correlated random numbers. The probability distribution function of such a surface is given by:

$$p(z_i) = \frac{1}{\sqrt{2\pi}} exp(\frac{z_i^2}{2\sigma^2}) \tag{3.2}$$

where z_i is the height of the point at position i above the plane of the mean level of the surface and σ is the r.m.s. surface variation, which controls the vertical extent of the roughness. The algorithm commences by drawing a sequence of $2N$ uncorrelated Gaussian distributed random numbers, v_i, such that:

$$\sigma_v^2 = \langle v_i^2 \rangle = \frac{1}{2N+1} \sum_{i=-N}^{N} v_i^2 \tag{3.3}$$

The sequence of random numbers is correlated using a series of weights, such that:

$$z_i = \sum_{j=-M}^{j=M} w_i v_{i+j} \tag{3.4}$$

where the weights, w_j, are normalised:

$$\sum_{j=-M}^{j=M} w_i = 1 \tag{3.5}$$

The aim is to pick these weights such that the resulting sequence in equation 3.4 are correlated, in the horizontal direction, with a Gaussian weighted correlation function:

$$C(x) = \frac{< h(x_0 >< h(x_0 + x) >}{h^2} = exp(\frac{x^2}{\lambda_c^2}) \tag{3.6}$$

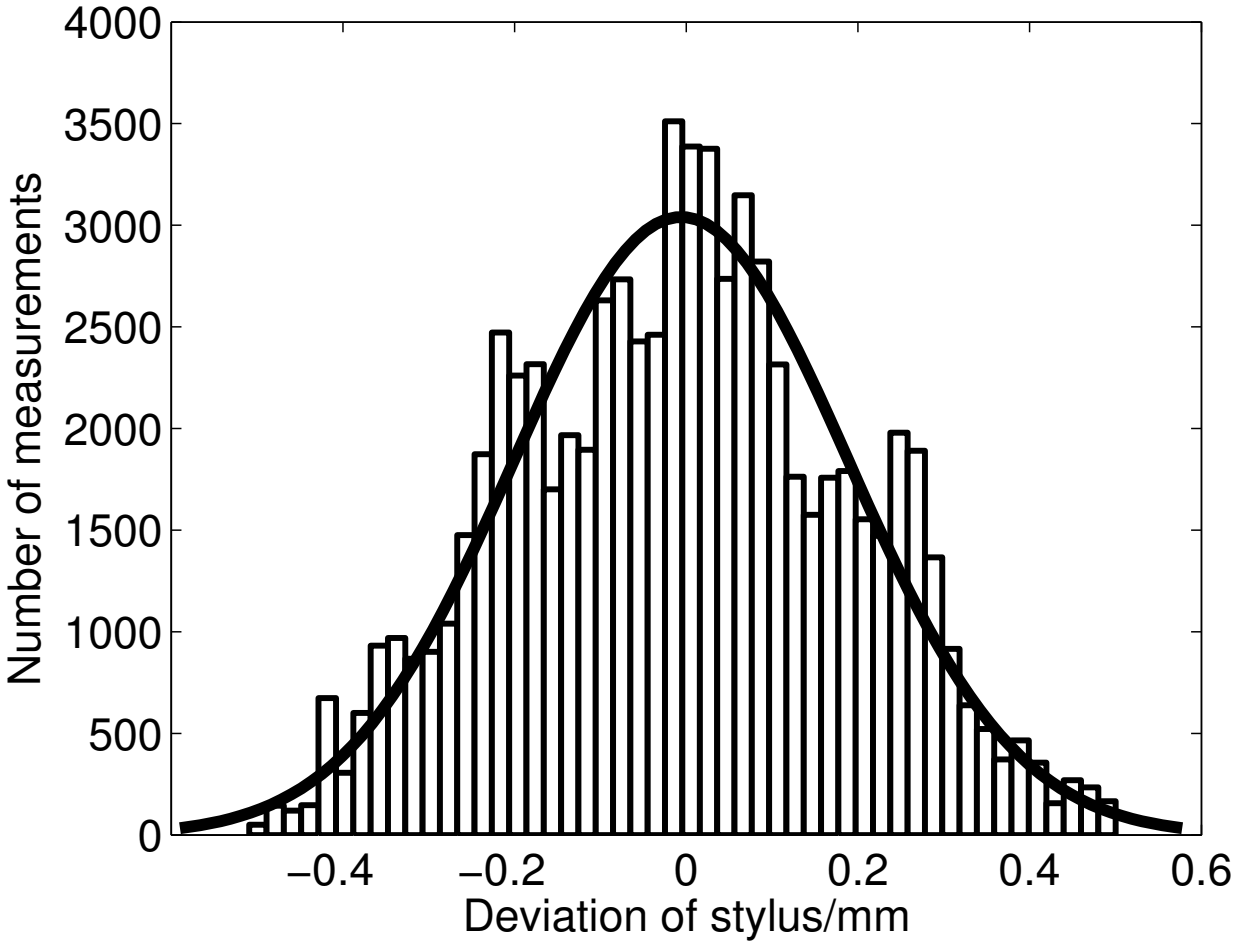


Figure 3.6: A histogram of the Talysurf measurements extracted from several lines (100mm in length) of the inner surface of a pipe (mean thickness 7.7mm) which has undergone sulphidation corrosion. The black line is a Gaussian, whose parameters have been estimated from the measurements which have been expressed as a deviation from the mean thickness of the pipe. This figure is taken from a published article by the author [1].

with a correlation length λ_c, which controls the distance at which the heights of two points becomes statistically independent. The weights can be calculated by comparing equation 3.6 to the correlation function as calculated from the height of each point above the surface. It is shown by Oglivy that these weights are given by:

$$w_i = \exp(-\frac{2(i\delta x)^2}{(2\lambda_0)^2}) \tag{3.7}$$

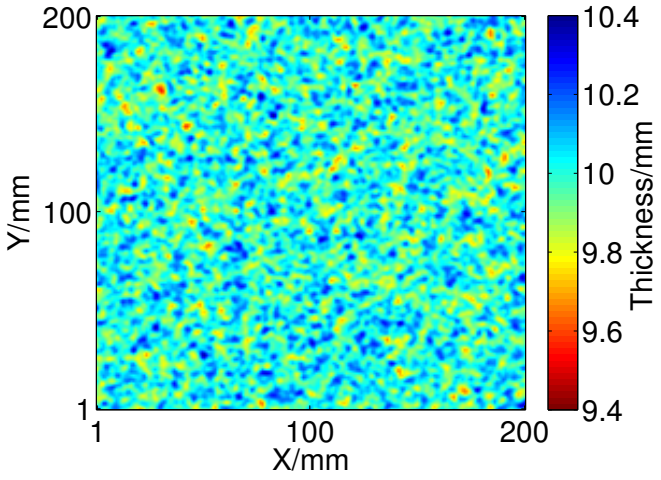
The script used to generate the rough surfaces is the same as in [14] and implemented in MATLAB[67]. Full details of the roughness generation algorithm can be found in Ogilvy[63]. Surfaces which have different correlation lengths in different directions do occur in practice. However, throughout this thesis an isotropic correlation length is used. Methods to generate non-isotropic correlation profiles can be found in Hu[66]. Examples of the thickness maps and CDFs produced by Oglivy's algorithm can be seen in figure 3.7

The blocking algorithm was tested using a large number of both Gaussian and exponential surfaces. Gaussian surfaces were generated using sequences of uncorrelated random numbers drawn from a normal distribution, while the exponential surfaces were generated with sequences drawn from an exponential distribution. Correlated surfaces were generated from these sequences using a weighted moving average with weights chosen such that points on the surface had a root mean squared (RMS) height of 0.1, 0.2 and 0.3mm (examples of Gaussian and exponential surfaces with RMS=0.3mm can be seen in figure 3.7b), a mean position of 10mm and a Gaussian autocorrelation function.

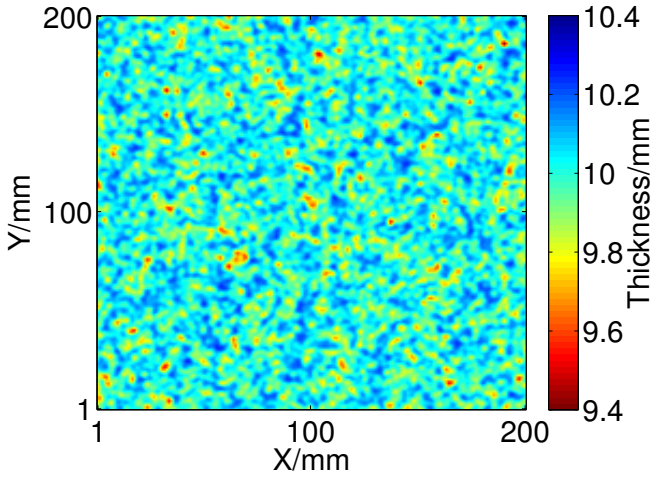
For both the Gaussian and exponential height distributions 1000 200 by 200mm surfaces were generated (with a distance of $1mm$ between each measurement in both the x and y directions) and each surface was processed using the blocking algorithm, with candidate block sizes ranging from 20 to 60mm (in steps of 5mm). If the algorithm successfully chose a block size for the surface an extreme value model was generated using that block size. This model was validated using the scan return period:

$$SRP = \frac{R(t_{min})}{N_{blocks}} = \frac{1}{\Phi(t_{min}|\mu,\sigma,\xi)N_{blocks}} \tag{3.8}$$

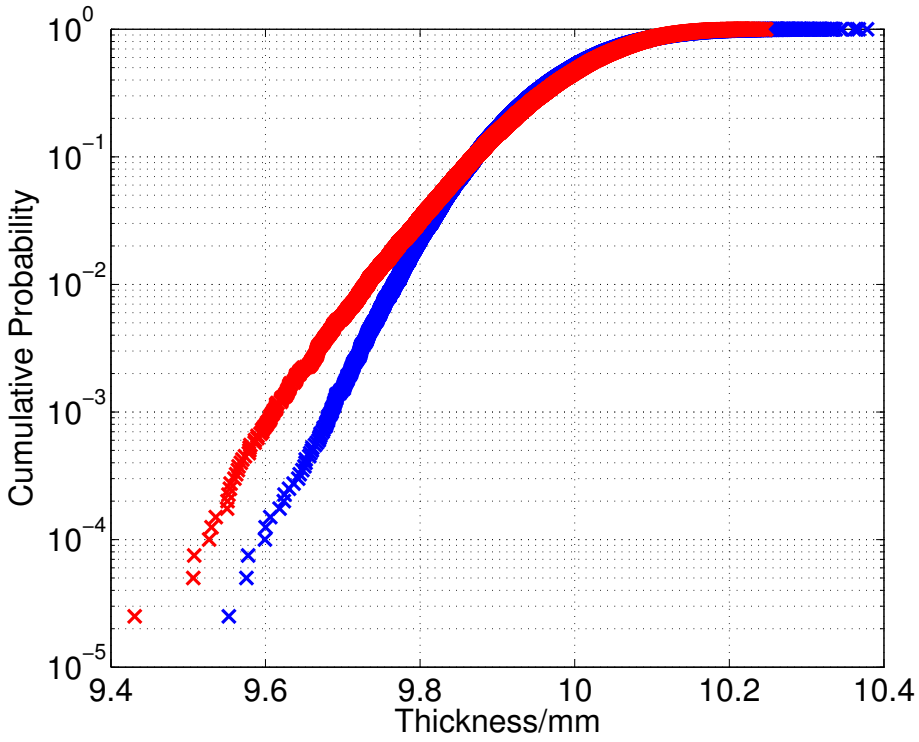
where N_{blocks} is the number of blocks the thickness map has been partitioned into, t_{min} is the smallest thickness across the surface and $R(t_{min})$ is the return level for the smallest thickness



(a)



(b)



(c)

Figure 3.7: (a,b) Example thickness maps from correlated Gaussian surface (a) and correlated exponential surface (b) with RMS height 0.1mm and correlation length 2.4mm, showing the position of each measurement with each point colour coded proportional to its magnitude. (c) Empirical cumulative distribution functions calculated from the Gaussian (blue) and exponential (red) thickness maps. The ordinate axis shows the probability of measuring a thickness of less than the corresponding value on the abscissa.

measurement (equation 4.6). If the model adequately describes the surface then $SRP \approx 1$ as there is at least one thickness measurement of t_{min}. Therefore, this metric provides evidence of the quality of the extreme value model for that surface.

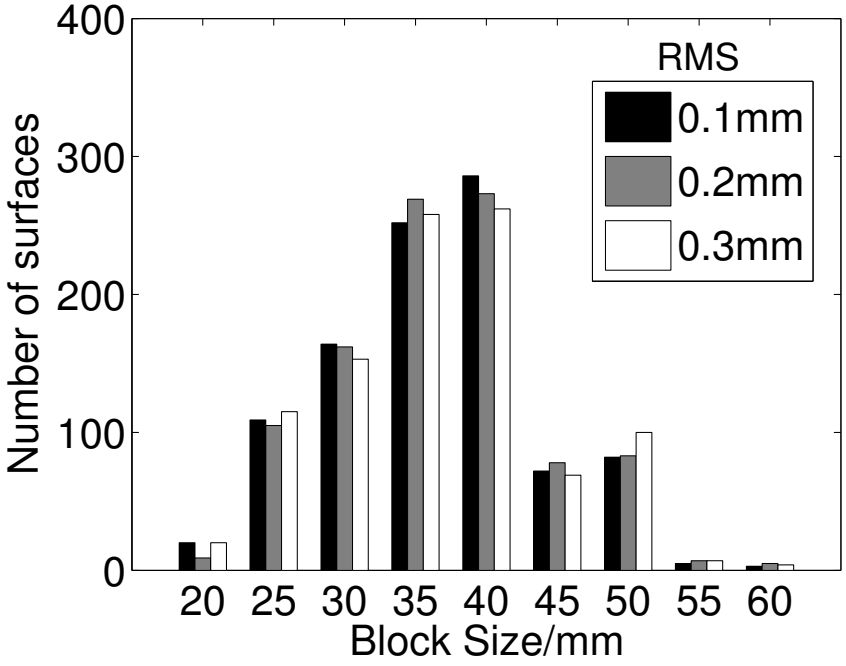
In the remaining sections of this chapter, the performance of the algorithm described in section 3.2 above will be tested using a large number of surfaces generated using the algorithm described in section 3.3. The SRP defined in equation 3.8 will be used to examine the behaviour of the algorithm with different significance levels and conclusions about the suitability of the algorithm will be drawn.

3.4 Performance of the algorithm

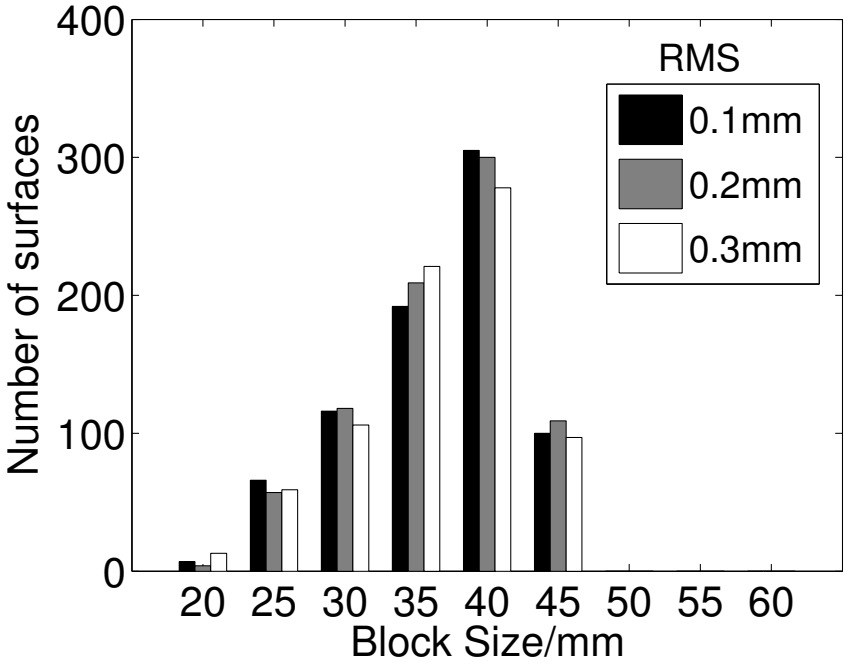
The statistics of the block sizes and scan return periods were calculated to examine the performance of the algorithm. Figures 3.8a, 3.8b, 3.9a and 3.9b show histograms of the block sizes selected for each surface. The histograms are a visualisation of the distribution of the block sizes. For each block size, the number of surfaces for which that block size was selected is visualised using a bar. The height of the bar is proportional to the number of surfaces for which that block size was selected.

Figure 3.8a shows the distribution of the blocks selected for Gaussian surfaces using a significance level of 0.01. The black, grey and white bars show the results from surfaces with RMS=0.1, 0.2 and 0.3mm. The mode block size selected by the algorithm for the Gaussian surfaces is 40mm. For comparison, the mode block size for the exponential surfaces (figure 3.9a) was 35mm for both RMS=0.2 and 0.3mm, and 30mm for RMS=0.1mm. This difference originates in the height distributions of the surfaces. For an exponential distribution, the average deviations from the median of the distribution are larger than for a Gaussian distribution. Consequently, one requires a smaller number of thickness measurements and therefore a smaller block size to measure an extreme than with a Gaussian distribution.

The mode block size of 40mm corresponds to thickness minima sample sizes of 25 for the Gaussian surfaces and the mode block size of 35mm corresponds to 32 sample minima for the exponential surfaces. In general this was a sufficient number of minima to be confident about the

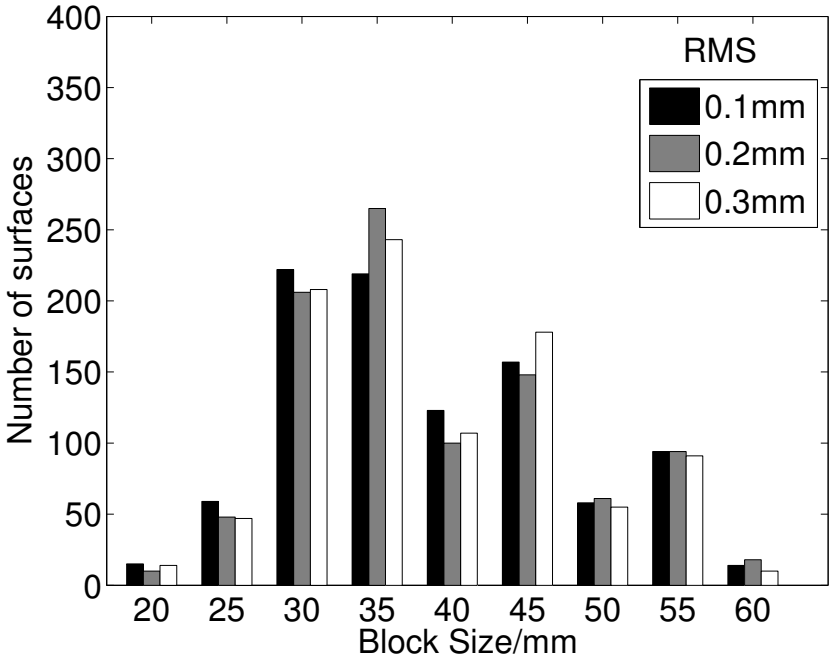


(a) Significance level 1%

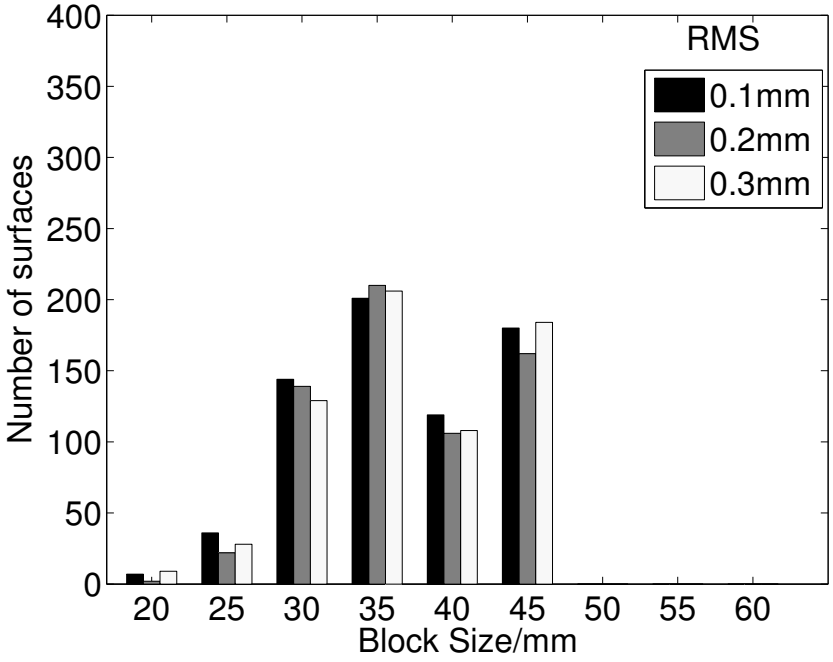


(b) Significance level 5%

Figure 3.8: Histograms of the number of Gaussian surfaces against block size at different significance levels, showing the number of surfaces for which the algorithm has selected a given block size. With a significance level of 1%, the algorithm could not find a suitable block size for 1% of the surfaces, this increased to 20% with a signficance level of 5%.



(a) Significance level 1%



(b) Significance level 5%

Figure 3.9: Histograms of the number of exponential surfaces against block size at different significance levels, showing the number of surfaces for which the algorithm has selected a given block size. With a significance level of 1%, the algorithm could not find a suitable block size for 3% of the surfaces, this increased to 30% with a signficance level of 5%.

quality of the generated extreme value model. However, there are a fraction of the surfaces for which a block size of greater than 50mm has been selected, which corresponds to smaller samples of minima (16 for 50mm and 9 for 60mm). Consequently the models generated using these block sizes will produce poor descriptions of the surface as there is less information from which to estimate the model parameters. This is evident in figures 3.10a and 3.11a which show the distribution of the SRPs for the models generated by the algorithm. The mode scan return period is around 1 for both types of surface, which is expected from our definition of SRP. However, some of the models have a very large SRPs. These models were generated using the larger block sizes (and the associated smaller sample sizes). In these cases the algorithm has required a much larger block size in order to find a sufficient level of evidence that the thickness measurements come from identical distributions.

This level of evidence is determined by the choice of the significance level of the KS test. A lower significance level means that the algorithm requires less evidence that the distributions are identical, increasing it raises the amount of evidence required. When the blocking algorithm fails to find a suitable block size we conclude that there is insufficient evidence that the assumptions made by EVA are met by that surface. As with any method, there are circumstances in which EVA is suitable and those in which it is not. Although the assumptions made to generate the surfaces are congruent with those of EVA, each surface is a random process. Consequently, it will not necessarily show evidence that the assumptions of EVA are met.

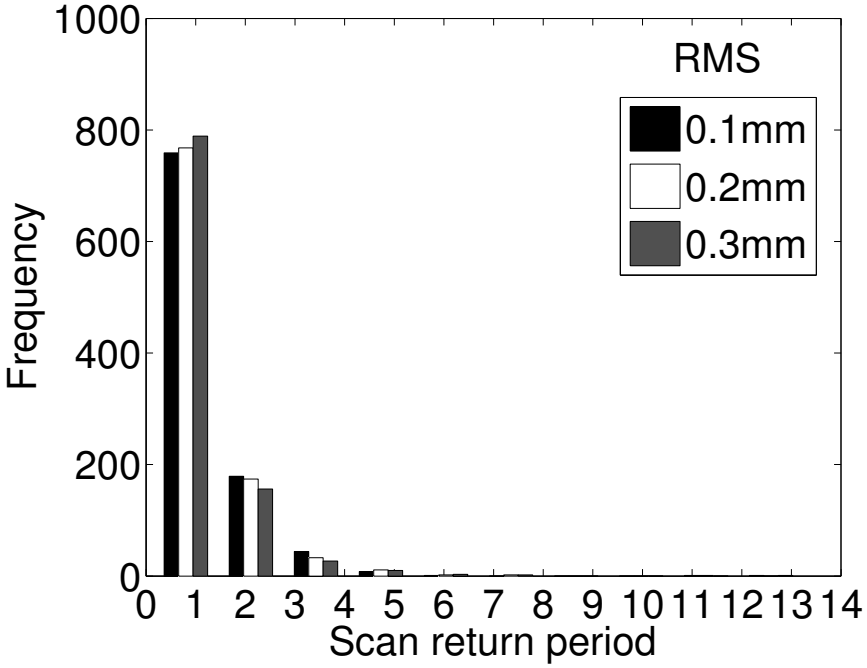
Figures 3.8b and 3.9b show the distributions of block sizes using a significance level of 0.05 for the Gaussian and the exponential surfaces respectively. The mode block sizes remain the same, however, there are no longer any surfaces for which a block size of greater than 50mm has been selected. In fact, the algorithm has failed to find a suitable block size for around 20% of the Gaussian surfaces and 30% of the exponential surfaces, compared to 1% and 3% at a significance level of 0.01. These surfaces mostly correspond to the larger block sizes in figures 3.8a and 3.9a. As a result the distributions of SRPs at a significance level of 0.05 (figures 3.10b and 3.11b) do not show SRPs greater than 5. For the surfaces in this chapter, this indicates that a significance level of 0.05 leads to models which more accurately describe the surface. As a general rule, an inspector

should choose the largest possible significance level as this ensures that the blocking algorithm does not produce models with very large SRPs.

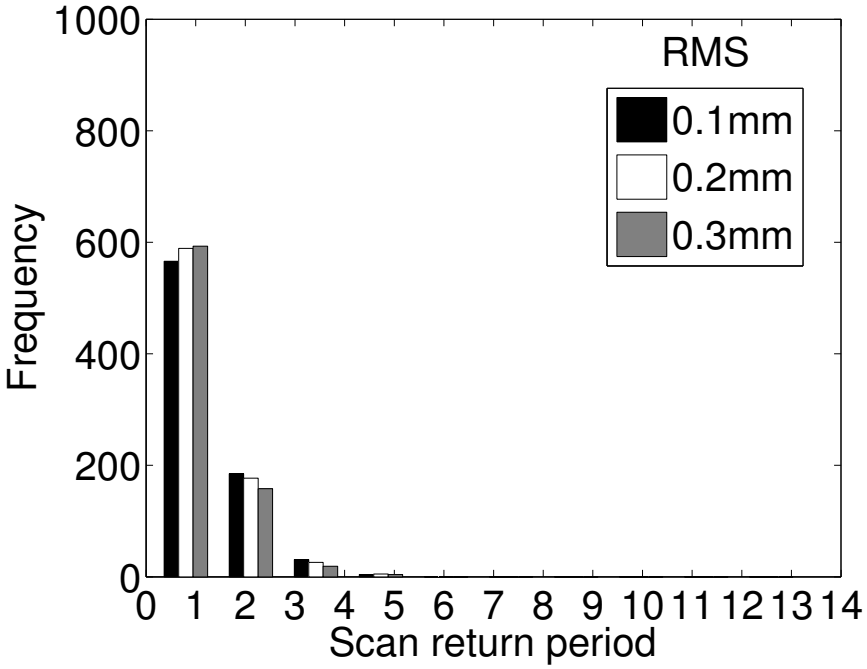
Figures 3.12 and 3.13 show box plots of the SRP for each block size for the Gaussian and the exponential surfaces. Box plots provide a visualisation of the distribution of the SRP calculated from each model. The interquartile range (IQR), represented by the length of each box, is the bounds within which half of the values of SRP lie. The median of the distribution is shown by the line in the middle of each box and the whiskers show the range (scan return periods within the 1% and 99% quantiles) which does not contain any outliers. Any values outside of this range are plotted individually as crosses.

Figures 3.12(a),(b) and (c) show the box plots for Gaussian surfaces with RMS=0.1,0.2 and 0.3mm at a significance level of 0.01. The black dashed line on the figures indicates a scan return period of 1. For block sizes of 55 and 60mm, the median scan return period deviates significantly from 1, as the models using these block sizes rely on a small number of minima. There are also a number of large outliers for some of the smaller block sizes. With a significance level of 0.05, there is a large reduction in the number of outliers shown in figures 3.12(d),(e) and (f) and there are no longer any models generated using block sizes greater than 50mm. The average deviation of the median from the black dashed line is also reduced. This is a consequence of the more stringent requirements for surfaces deemed suitable for EVA.

This pattern is continued for the exponential surfaces. Figures 3.13(a),(b) and (c) show the box plots for exponential surfaces with RMS=0.1,0.2 and 0.3mm at a significance level of 0.01. The median of the SRP for the models generated for each block size was close to 1 with IQRs of around 2 scans. In a similar manner to the Gaussian surfaces, the median SRP deviates significantly from 1 for block sizes greater than 40mm, which indicates that these models are poor descriptions of the data. At a larger significance level of 0.05, there is a reduction in both the IQR for each set of data and the number of outliers shown in figures 3.12(d),(e) and (f). This suggests that making the test requirements more stringent increases the quality of the models produced by the algorithm.

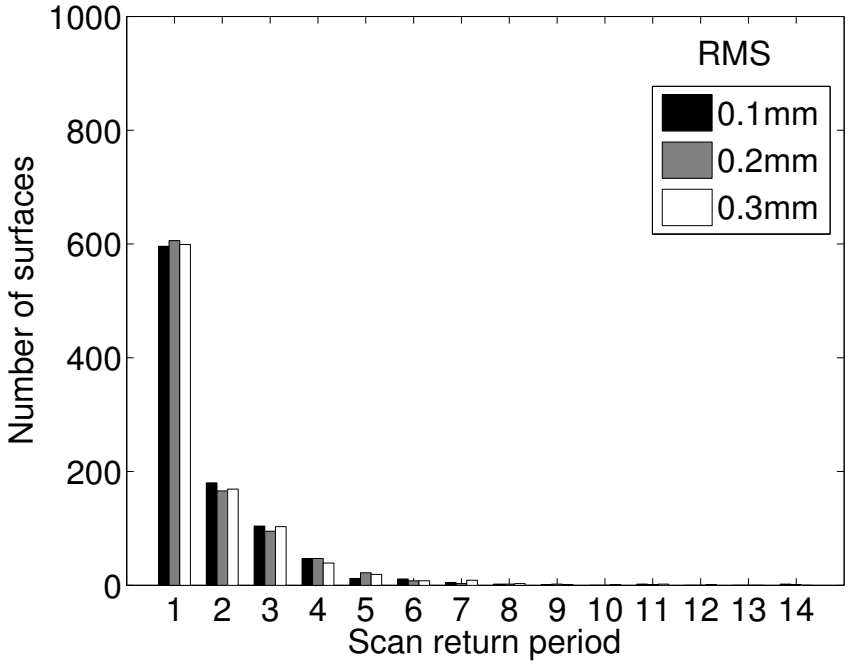


(a) Significance level 1%

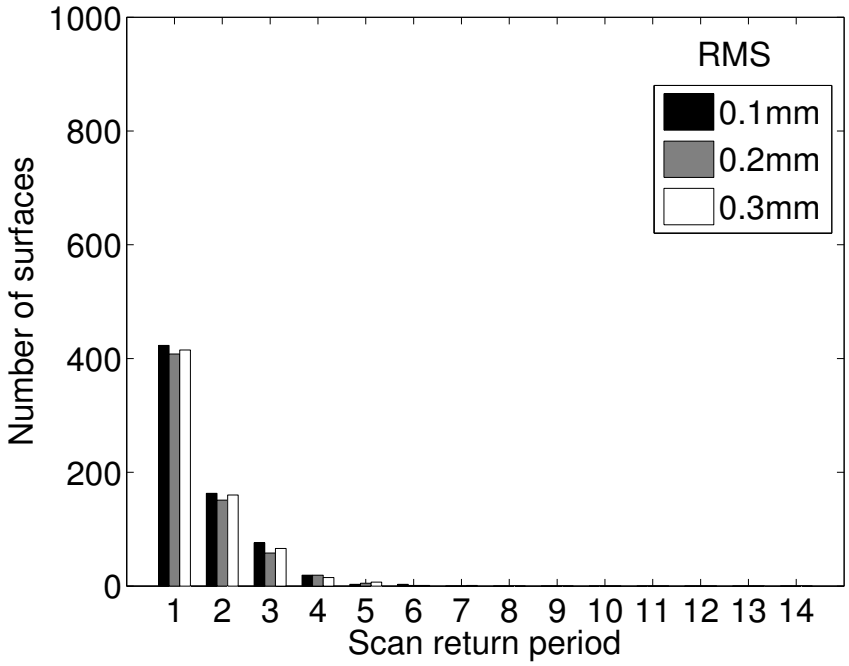


(b) Significance level 5%

Figure 3.10: Histograms of the number of Gaussian surfaces against scan return period at different significance levels. With a significance level of 1%, scan return periods ranged as far as 14 scans, which corresponded to block sizes greater than 40mm. On increasing the significance level to 5%, the algorithm did not select a block size for these surfaces and the range of scan return period decreased.



(a) Significance level 1%



(b) Significance level 5%

Figure 3.11: Histograms of the number of exponential surfaces against scan return period at different significance levels. With a significance level of 1%, scan return periods ranged as far as 14 scans, which corresponded to block sizes greater than 40mm. On increasing the significance level to 5%, the algorithm did not select a block size for these surfaces and the range of scan return period decreased.

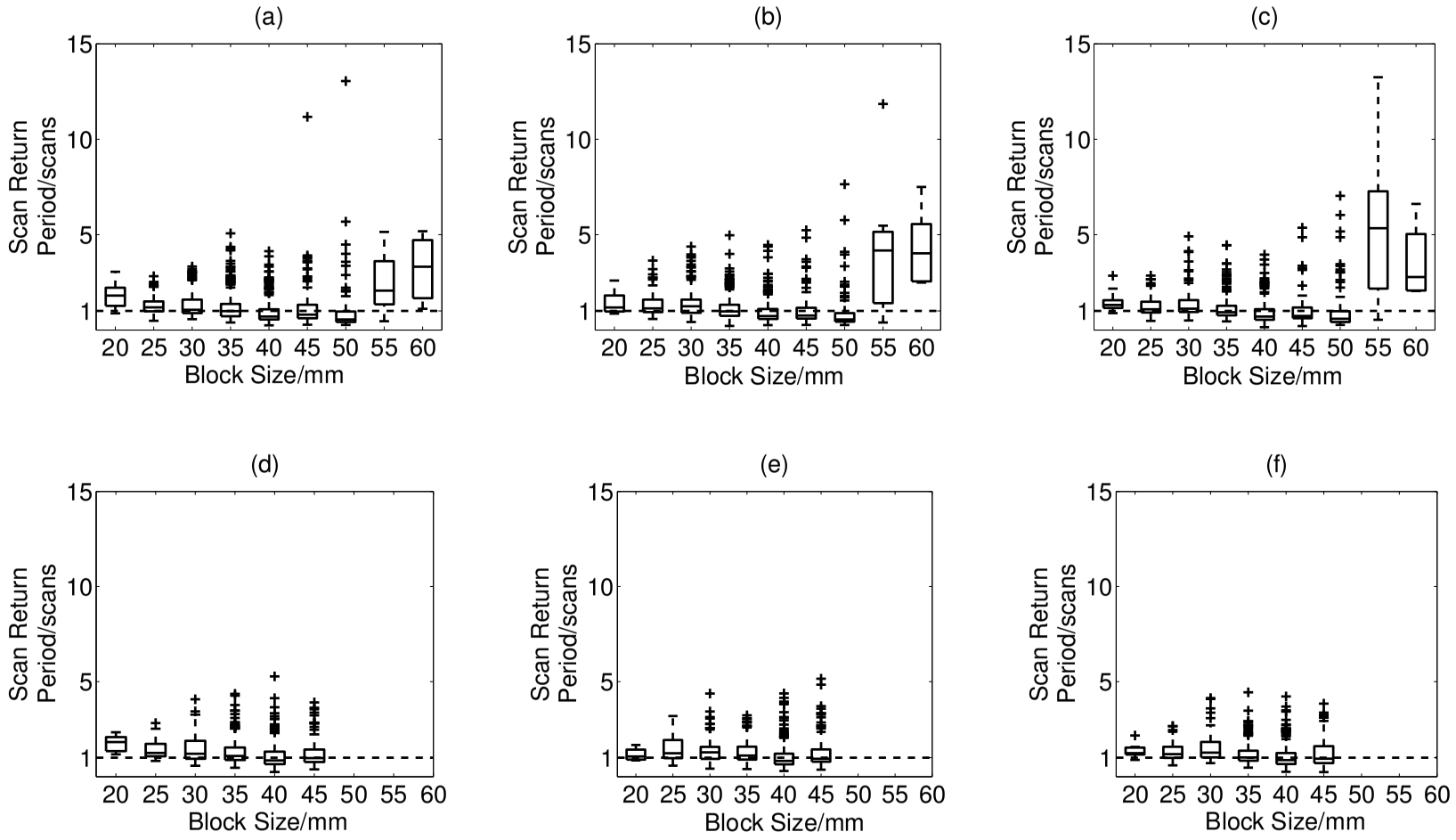


Figure 3.12: Box plots showing the spread in the return period for the block size selected by the algorithm for the Gaussian surfaces with (a) RMS=0.1mm at the 1% significance level, (b) RMS=0.2mm at the 1% significance level, (c) RMS=0.3mm at the 1% significance level, (d) RMS=0.1mm at the 5% significance level, (e) RMS=0.2mm at the 5% significance level and (f) RMS=0.3mm at the 5% significance level.

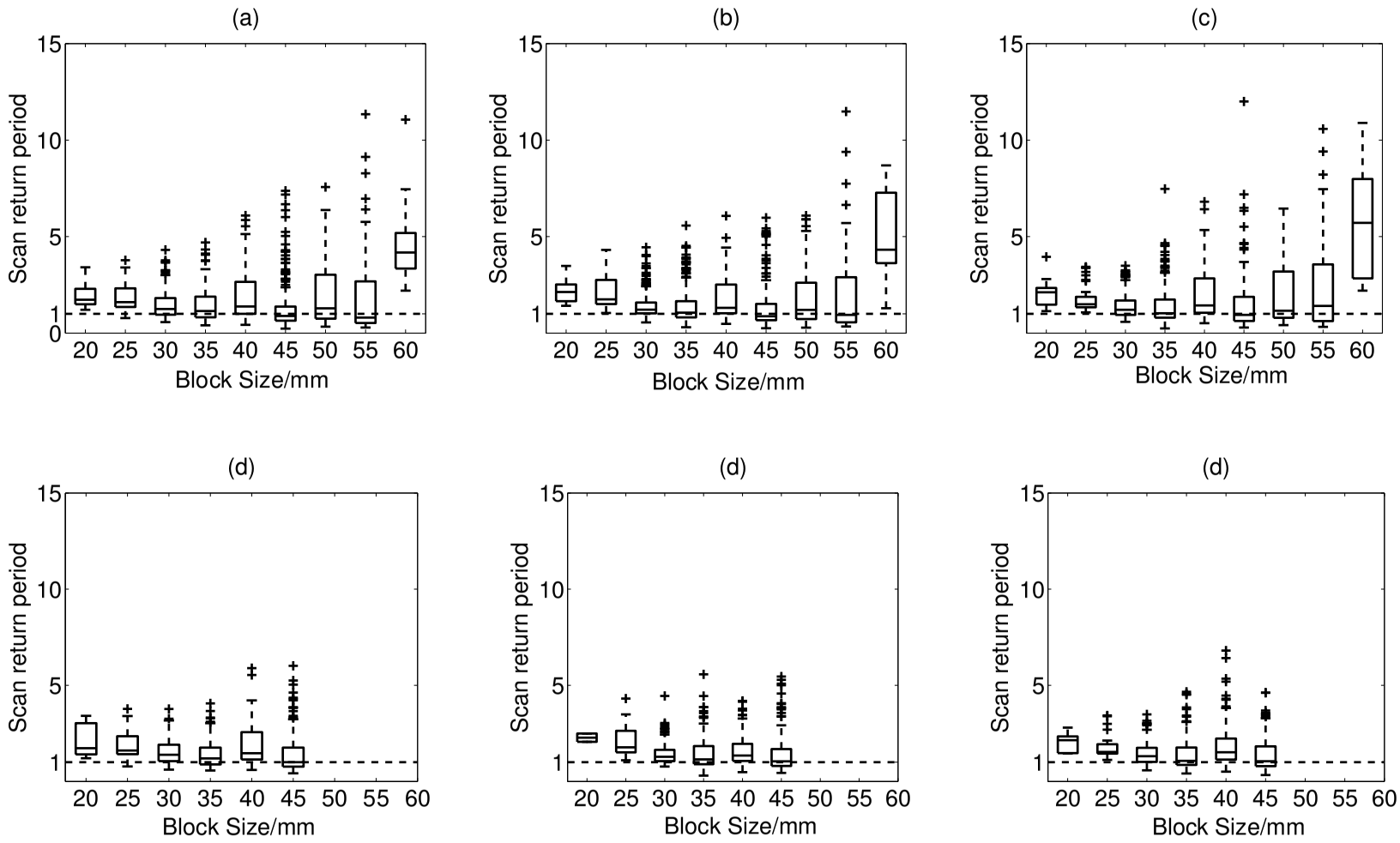


Figure 3.13: Box plots showing the spread in the return period for the block size selected by the algorithm for the exponential surfaces with (a) RMS=0.1mm at the 1% significance level, (b) RMS=0.2mm at the 1% significance level, (c) RMS=0.3mm at the 1% significance level, (d) RMS=0.1mm at the 5% significance level, (e) RMS=0.2mm at the 5% significance level and (f) RMS=0.3mm at the 5% significance level.

3.5 More realistic corrosion

3.5.1 Introduction

The research into EVA presented in this chapter has exclusively dealt with inspection consisting of a single thickness distribution. This is the type of thickness measurement distribution expected to be produced by a general corrosion mechanism, a type A degradation mechanism (table 1.1). In real in-service components degradation can occur by multiple mechanisms simultaneously. An example of a thickness map measured as part of an inspection of an operational subsea oil pipeline is shown in figure 3.14 (an example of a type B damage environment).

Figure 3.14 is clearly different from other examples of thickness maps in this thesis (e.g. figure 2.1a). There are areas over which a localised degradation mechanism (pitting) is occurring (the green and red areas in the image). In these areas the thickness measurement distribution is different than across the other (blue areas) of the thickness map. This poses a problem for an inspector who wants to use EVA. If they partition figure 3.14 into a number of equally sized blocks, the thickness measurements in each block and, therefore, the minima, will originate from different distributions. This will violate the assumptions made by extreme value theory. A model constructed from this set of minima will be incorrect.

The ability to handle the data collected from real components is key to increasing the usage of EVA in industry. The framework outlined in this thesis offers a route to achieving this goal. The blocking algorithm in chapter 3 would fail to find a suitable block size for figure 3.14 as the thickness distributions in each block would be different. If the inspector possessed a method to determine which blocks contain localised corrosion and which do not, then they would be able to classify each minimum belonging to either the localised distribution or the general distribution. This would result in two (or more) sets of minima, each corresponding to a different damage mechanism. An extreme value model could then be constructed for each of them.

Classification can be achieved by hand. An inspector will be able to see which blocks correspond to the localised corrosion and which do not. However, in some cases it can be hard to distinguish between the localised areas and the general distribution. This becomes particularly difficult when

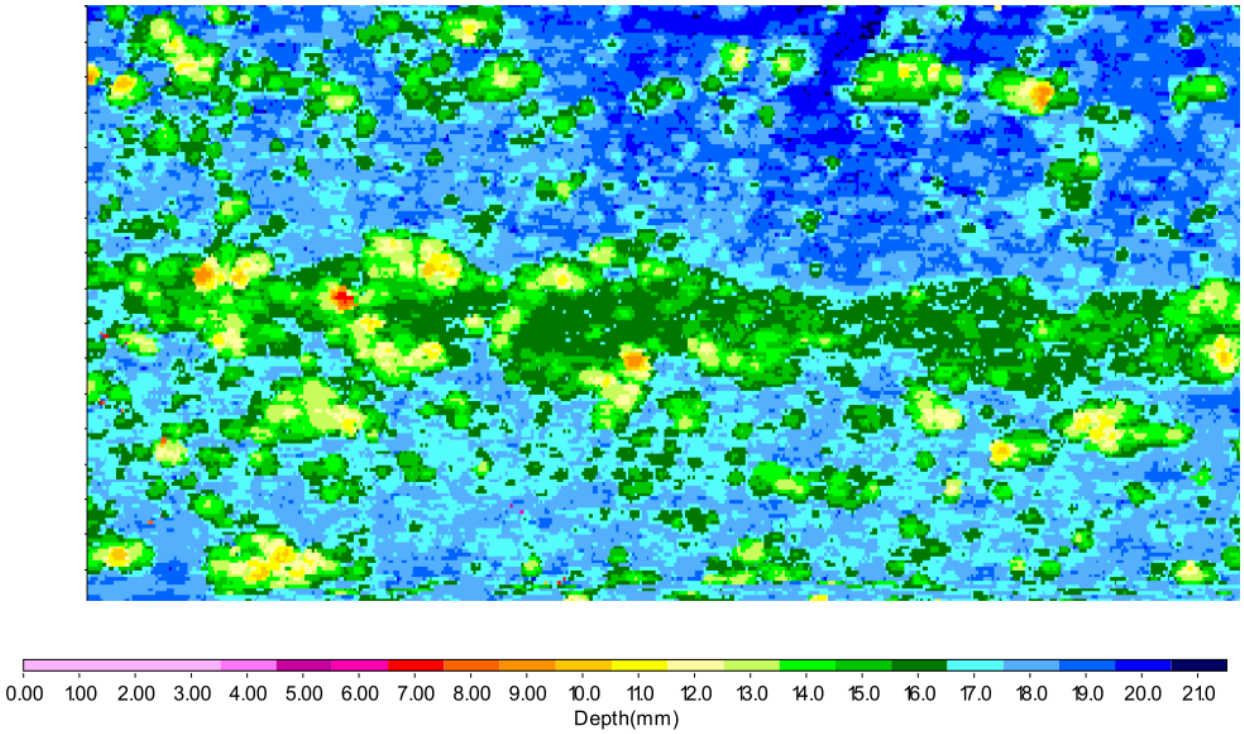


Figure 3.14: An example of a thickness map taken from a subsea pipeline. Reproduced from Stone [2, fig. 1].

there are more than two thickness measurement distributions present (the result of many different degradation mechanisms). A method of automating damage classification, to separate out the thickness distributions, is required.

An initial investigation into a method of achieving this goal is described in this section. This investigation used an approach which tried to determine a threshold separating the thickness distributions associated with the pitting and the general corrosion. A basic example of the use of this approach is presented, showing that it can achieve some separation between the general and pitting corrosion distributions with simulated inspection data. However, there are a number of limitations to this type of approach. The section concludes with a discussion of these shortcomings with highlights for further work and lines of investigation.

3.5.2 Pit selection algorithms

Localised corrosion often results in clusters of thickness measurements which are significantly smaller than the majority of the thickness measurements. This is usually exhibited in an ECDF of the thickness measurements as an exponential tail, an example of which is shown in figure 3.15. Figure 3.15 is a ECDF of thickness measurements collected as part of an inspection of an in-service separator undergoing corrosion [2].

The points enclosed in the red ellipse are the pitting tail, while those in the blue box are a general degradation mechanism. A review of ECDFs by Stone showed that localised corrosion is usually exhibited by these exponential-like tails [2]. There is a clear transition in figure 3.15 between the measurements associated with general corrosion (highlighted by a black dashed line). The aim of the algorithm in this section is to calculate the thickness at which this transition occurs.

Figure 3.16 shows an example of simulated inspection data from across an area in which localised pitting is occurring. The depth of the pits was simulated using the gamma-process model described by [68]. The resulting surface was overlaid with a correlated Gaussian roughness generated using the algorithm described in section 3.3. A colour map of the resulting inpsection data is shown in figure 3.16a and the corresponding ECDF in figure 3.16b. For the purposes of this discussion this simulated inspection data is used to demonstrate how the method works.

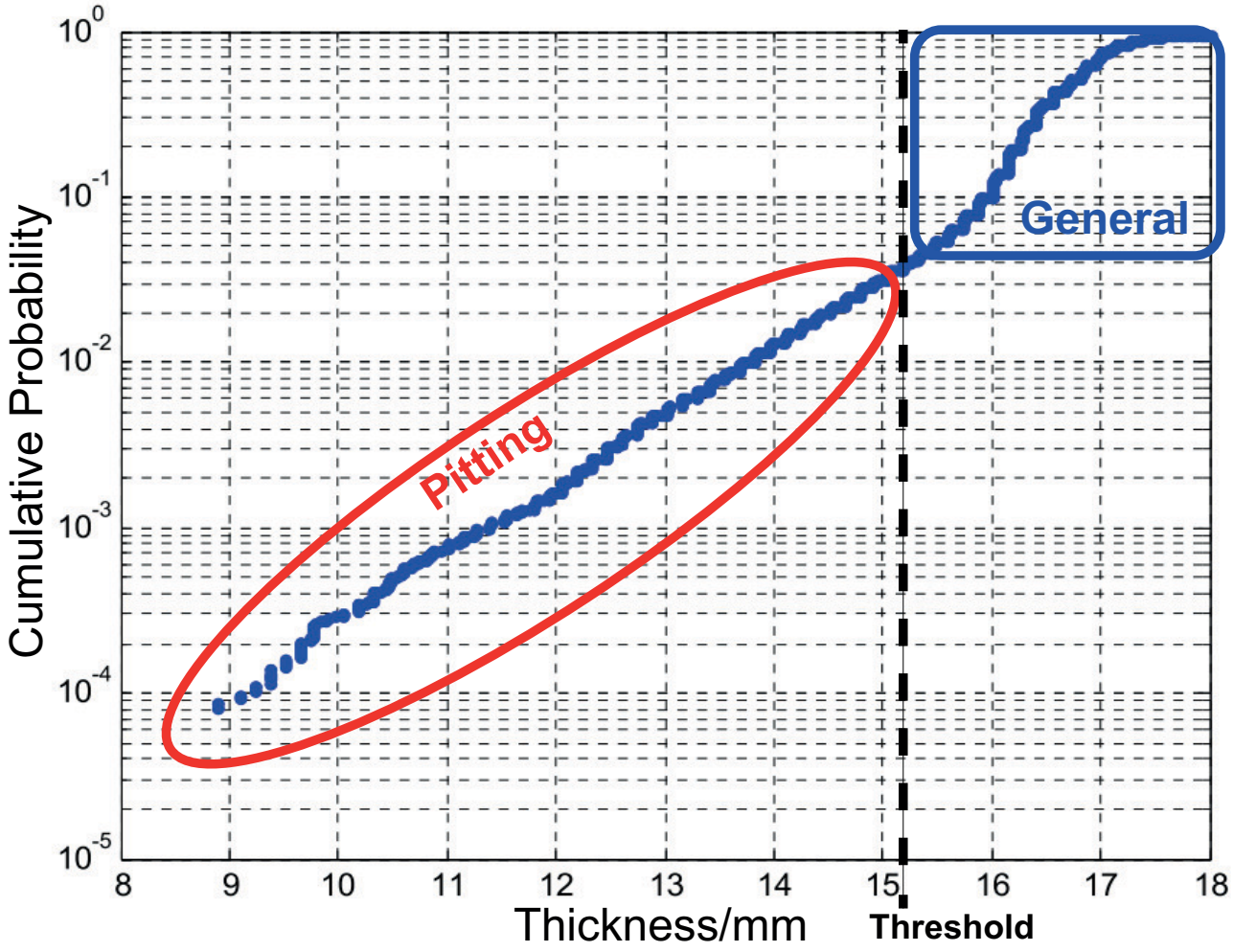
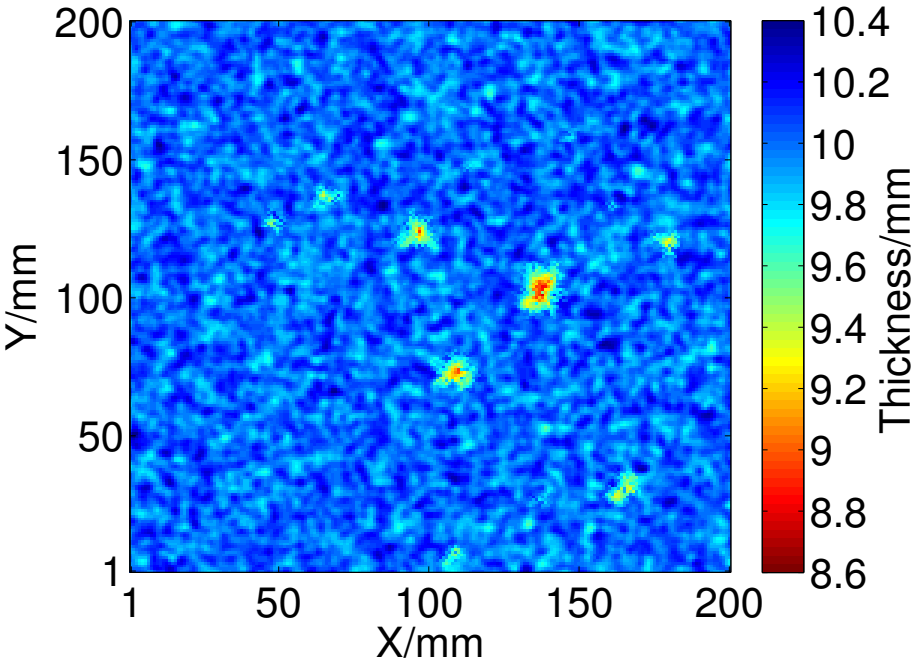
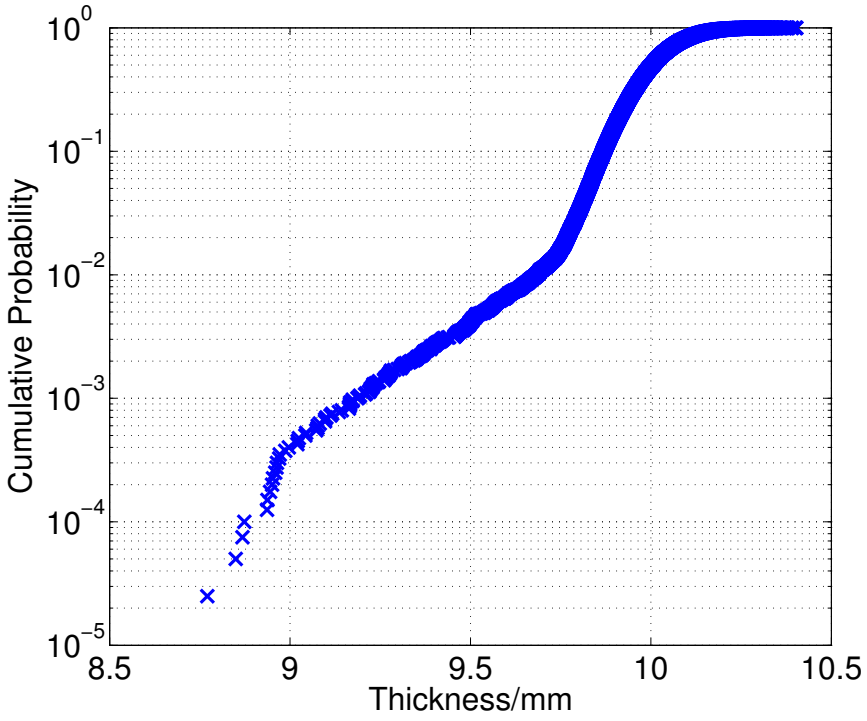


Figure 3.15: An ECDF extracted from an inspection of an in-service separator undergoing corrosion.Reproduced from Stone [2].



(a)



(b)

Figure 3.16: (a) A colour map of simulated inspection data from across a numerically generated pitted surface. (b) The empirical cumulative distribution function calculated from the colour map shown in (a)

As an initial step the inspector chooses a threshold. The thickness measurements above this threshold are used to generate an ECDF. This ECDF is compared to the ECDF of all of the thickness measurements using a two sample Kolmogorov-Smirnov test (see section 3.2). If the test determines that the ECDFs are the same distribution, then the inspector increases the threshold and repeats this process. If the test determines that the ECDFs are different, then the inspector has found a threshold separating the two distributions. This algorithm is summarised in figure 3.17.

This algorithm was used to process the thickness measurement distribution shown in figure 3.16. The initial threshold was chosen to be the smallest thickness measurement across the thickness map ($8.7mm$). At each step of the algorithm this threshold was increased by $0.05mm$, until the ECDFs were significantly different (as determined by the KS test). For this data, the algorithm selected an threshold of $9.58mm$.

Figure 3.18 shows a comparison of the ECDF of all of the thickness measurements (blue) with the ECDF of all the thickness measurements above the threshold (red). The threshold selected by the algorithm is shown as a black dashed line. These distributions are clearly quite different, the exponential tail, indicative of a pitting distribution, is much less pronounced. The algorithm has successfully selected a threshold which can be used to separate the thickness measurement distributions.

A further check of the quality of the threshold can be obtained using a classification map. A classification map for the inspection data is formed by taking the colour map in figure 3.19a and changing the colour of every pixel below the threshold to black. All pixels above the threshold are coloured white. Consequently, if the threshold is appropriate, the location of the pits will be indicated by the black pixels. A classification map for the inspection data is shown in figure 3.19.

By comparing the classification map in figure 3.19b to the thickness map in figure 3.19a, it is clear that the algorithm has selected a reasonable threshold. The majority of the thickness measurements belonging to the pits are shown in black. However, the threshold has also selected a few thickness measurements which do not belong to the pitting distribution, as demonstrated by the isolated black pixel and it has failed to select a number of thickness measurements around

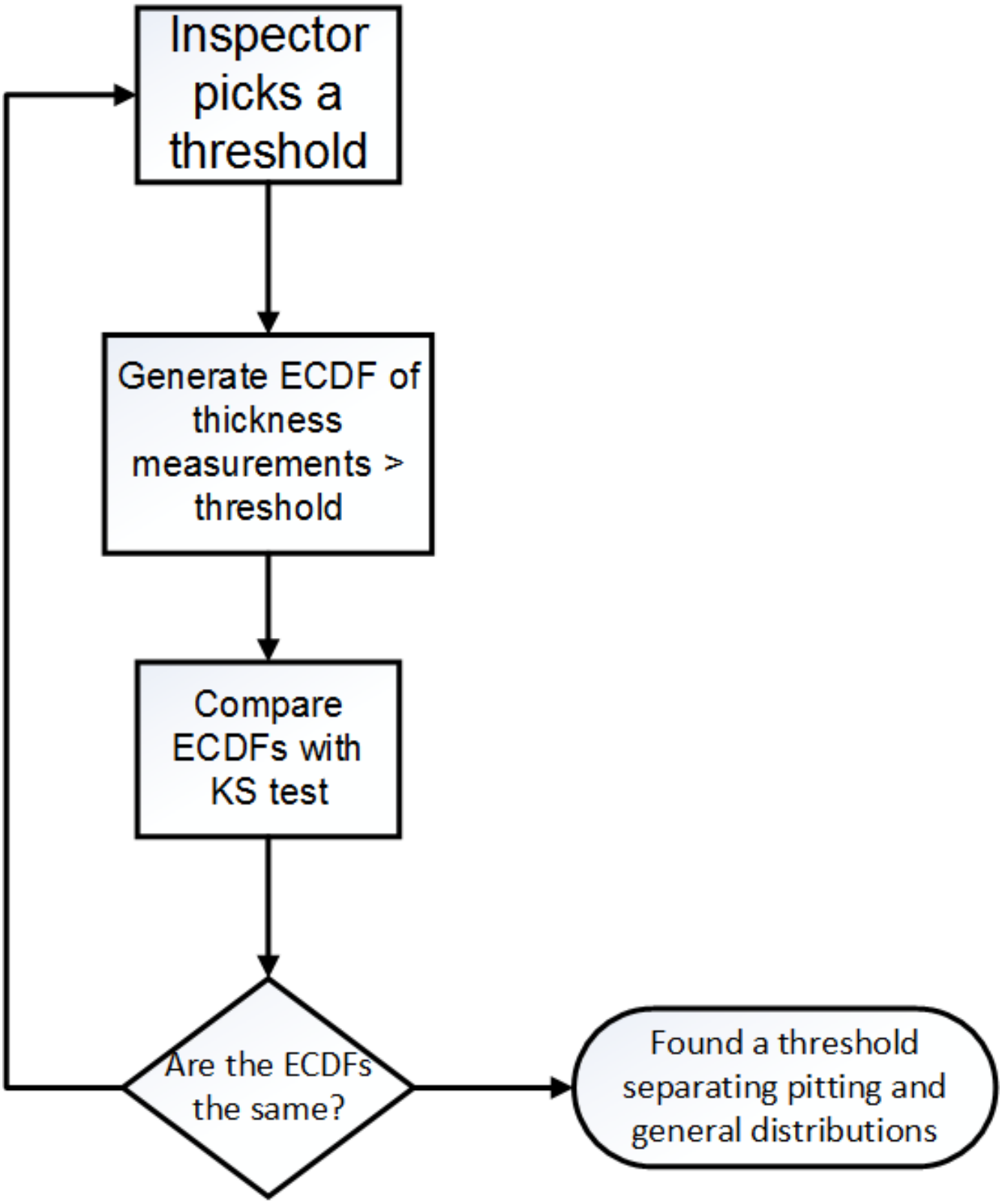


Figure 3.17: A flow chart summarising the threshold selection method.

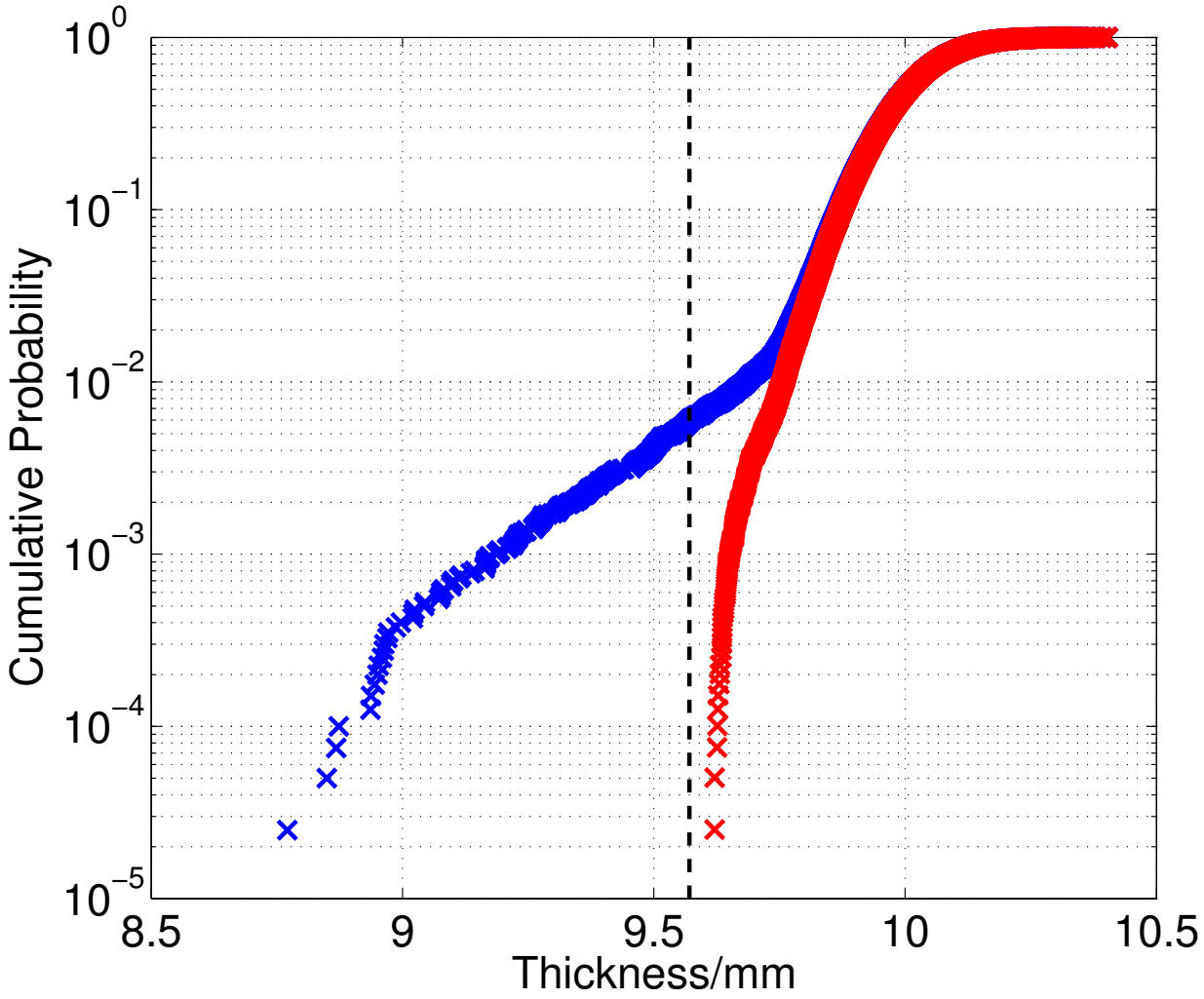
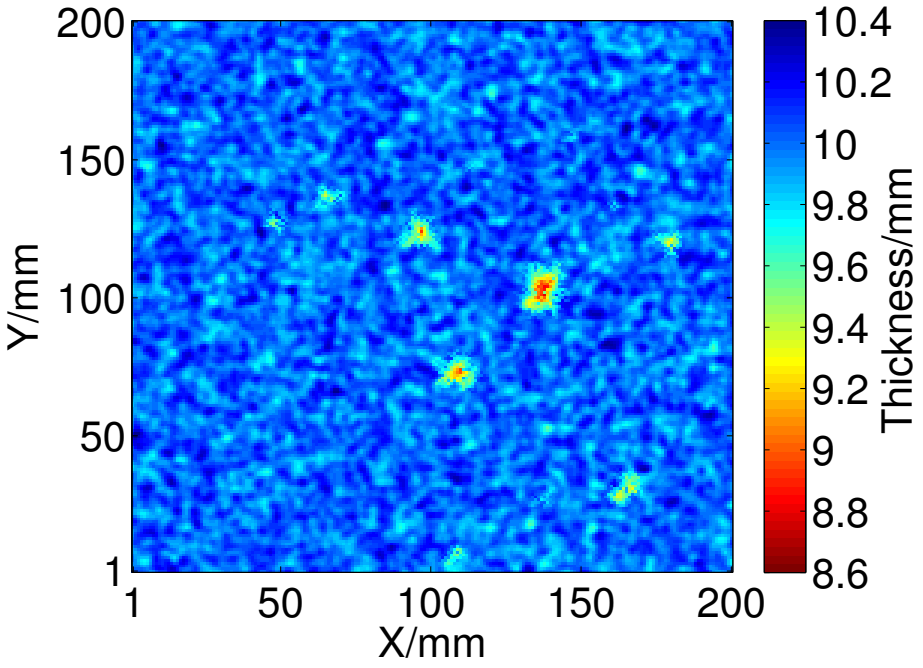
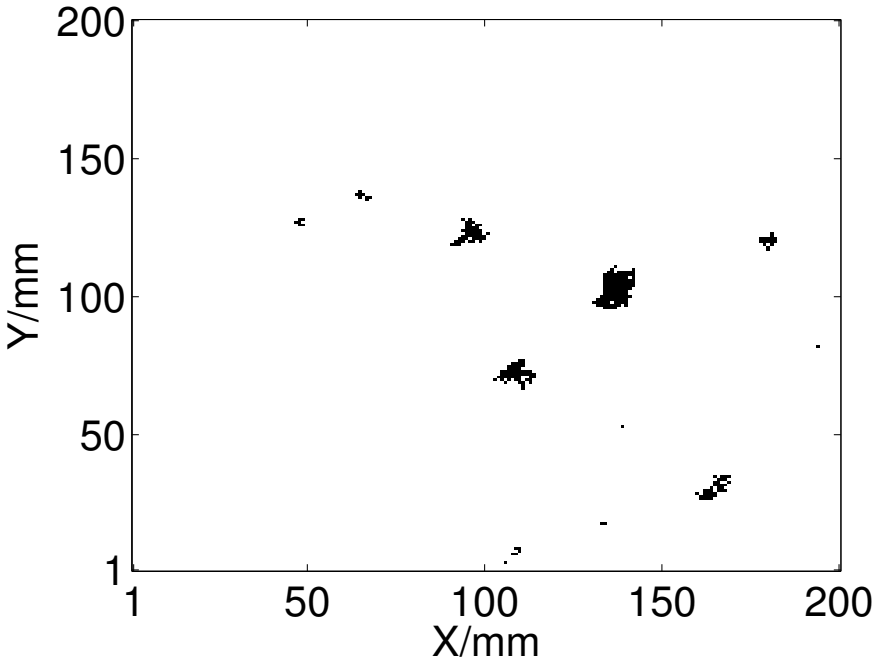


Figure 3.18: A comparison of the ECDF of all of the thickness measurements (blue) with the thickness measurements greater than the threshold (red). The algorithm has selected a threshold which has removed the exponential tail from the distribution.



(a)



(b)

Figure 3.19: (a) A colour map of simulated inspection data from across a numerically generated pitted surface. (b) A classification map corresponding to the colour map in (a). All of the thickness measurements above the threshold (general corrosion distribution) are coloured white, those below it are coloured black (the pitting corrosion distribution).

the outside of the pits.

This is a consequence of the fact that the algorithm does not take into account the spatial location of the thickness measurements. Pits are clusters of thickness measurements that have a thickness significantly different than the rest of the distribution. The algorithm does not consider this. It only looks at the thickness measurement relative to the rest of the distribution, searching for a threshold which separates out the pits from the general corrosion. Subsequently, when thickness measurements across pits are relatively close to the values of the thickness in the general corrosion the algorithm will misclassify them.

This can be worked around in the context of the blocking algorithm described in this chapter. The inspector needs to know which blocks contain pits so that the minima selected from these blocks can be processed separately. The algorithm described here allows one to achieve this: it gives an approximate location of the pits across the surface. An inspector can then use these locations and an adapted blocking algorithm to generate extreme value models for both the general corrosion distribution and the pitting distribution.

There are two different approaches an algorithm could take it could either: 1) separate out the blocks which contain pits, or 2) remove the thickness measurements corresponding to the pits and process them separately. Both of these approaches require further investigation. For example, with the first approach, it is possible it will lead to an insufficient number of blocks and therefore minima to perform EVA with either the pitting measurements or the measurements from the general corrosion. Therefore, this method is likely to only be useful in cases where there are enough blocks containing both pits and general corrosion (25 each according to the studies performed in chapter 4).

The second approach also requires further investigation. Simply separating the pitting thickness measurements from the inspection data does not allow one to construct an extreme value model from the data. For the general corrosion distribution, it will lead to some blocks containing different numbers of thickness measurements than the others. A minimum selected from these blocks will follow a different extreme value distribution than the others. In addition, there is no obvious area with which to associate the separated thickness measurements.

3.6 Conclusions

Extreme value analysis can be used to model the thinnest areas of a component and to extrapolate to the condition of much larger areas that are exposed to the same degradation mechanism. There is currently no standard methodology to sample the minimum thickness from an ultrasonic inspection thickness map. This chapter has described an approach to sample the thickness minima by looking for evidence that the assumptions made by EVA are reasonable for the inspection data. The algorithm was applied to a large number of surfaces with both Gaussian and exponential height distributions. It successfully selected a block size for the majority of surfaces and generated extreme value models which provided good descriptions of the data. Block size selection is a trade-off: too small a block size and the GEVD will not provide a good model for the minimum thicknesses; too large a block size and accurate estimates of the model parameters are unattainable.

Smaller block sizes correspond to larger samples of thickness minima ($\tilde{1}00$ minima). The variation in the quality of models generated using the smaller block sizes is larger than that for the larger block sizes, which is the result of some of the minima not being extremes of the distribution. In contrast, larger block sizes ensure that all the thickness measurements in a sample are extremes at the expense of sample size ($\tilde{1}6$ minima). Smaller samples lead to increased uncertainty in the parameter estimates for the extreme value models. The data presented in this chapter indicates that, for the set of surfaces described here, block sizes of 35 to 45mm are most suitable. This block size selection provides a balance between ensuring the minima are extremes of the distribution, while limiting the uncertainty associated with a small sample size. Uncertainty associated with EVA models is a key part to the development of a framework for using EVA as a tool for PCI. In the next chapter, a study of the uncertainties associated with EVA of thickness minima is presented. This study provides further important information which can be used by an inspector when constructing an EVA model.

Chapter 4

Extrapolation using extreme value analysis

4.1 Introduction to extrapolation using extreme value models

To perform partial coverage inspection (PCI), an inspector will perform an ultrasonic thickness C-scan of a fraction of the component. The resulting thickness map can be used to construct an extreme value model for the smallest thickness measurements by partitioning the thickness map into N_{blocks} equally sized blocks. The blocks are used to construct a sample of thickness minima by selecting the smallest thickness measurement in each block.

If the thickness minima are selected from sets of thickness measurements that are independently and identically distributed, the generalized extreme value distribution (GEVD) is the limiting form of the thickness minima distribution. The GEVD is defined using three parameters: the location parameter, $\mu \in \mathbb{R}$, which determines the size of the minima; the scale parameter, $\sigma > 0$, which determines the spread of the minima; and the shape parameter, $k \in \mathbb{R}$, which determines the shape of the distribution. The probability of measuring a thickness minimum less than x is given by:

$$\Phi(x|\mu,\sigma,k) = 1 - \exp\left\{-\left[1 + k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/k}\right\} \tag{4.1}$$

for $1 + k(x - \mu)/\sigma > 0$, where $\Phi(x|\mu, \sigma, k)$ is the probability of measuring a minimum thickness (in a block) of less than x. Examples of probability density functions of GEVDs can be found in figure 2.2.

Estimates of μ, σ and k can be extracted using maximum likelihood estimation (MLE), a widely used technique of estimating distributional parameters [69]. The likelihood of a set of thickness minima is a measure of how probable a set of distributional parameters are given the available data. For an extreme value model of a set of thickness measurements $X = \{x_1, ..., x_N\}$, the likelihood function is, derived using Coles [70, eq. 2.7, p. 30]:

$$\begin{aligned} \mathcal{L}(\mu, \sigma, k|X) &= \prod_{i=1}^{N} P(x_i) \\ &= \prod_{i=1}^{N} \frac{1}{\sigma} \left[1 + k\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\left(\frac{1}{k}\right)-1} \exp\left(-\left[1 + k\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\frac{1}{k}}\right) \end{aligned} \tag{4.2}$$

where $\mathcal{L}(\mu, \sigma, k|X)$ is the likelihood of a set of parameters (μ, σ, k) given the set of thickness minima X. A more detailed description of the maximum likelihood method can be found in section 2.1.

The likelihood function can be thought of as the probability that (μ, σ, k) are the distributional parameters associated with the set of thickness minima. Maximisation of $\mathcal{L}(\mu, \sigma, k|\mathbf{x})$ with respect to (μ, σ, k) provides the maximum likelihood estimates, $(\hat{\mu}, \hat{\sigma}, \hat{k})$, for the true values of (μ, σ, k). For a real application of MLE, an inspector would use the log-likelihood function as it is easier to work with, from Coles [70, p. 55]:

$$\begin{aligned} L(\mu, \sigma, k|\mathbf{x}) &= \ln\left(\mathcal{L}(\mu, \sigma, k|\mathbf{x})\right) \\ &= N \ln(\sigma) - \left(1 + \frac{1}{k}\right) \sum_{i=1}^{N} \ln\left(\left[1 + k\left(\frac{x_i - \mu}{k}\right)\right]\right) \\ &\quad - \sum_{i=1}^{N} \left[1 + k\left(\frac{x_i - \mu}{k}\right)\right] \end{aligned} \tag{4.3}$$

where $L(\mu, \sigma, k|\mathbf{x})$ denotes the log-likelihood of the model. Maximisation of the log-likelihood with respect to (μ, σ, k) is equivalent to maximisation of the likelihood function [70, p. 31]. For the remainder of this thesis, references to the likelihood will refer to the log-likelihood.

Once an extreme value model has been constructed from the thickness minima, an inspector can use this model to perform PCI. This will require extrapolations to areas larger than the initial inspection using the extreme value model. There are two methods avaliable for extrapolation of an extreme value model: the return level method and the distributional method. The return level method is used to draw conclusions about areas larger than the initial inspection by mapping them to quantiles of the EV model, whilst the distributional method attempts to directly construct a model for the minimum in the extrapolated area.

In the literature, the return level method is by far the most widely used for extrapolation. Examples of return level calculations can be found in fields as varied as hydrology[40] and finance[38]. There are a limited number of examples of the calculation of return levels using corrosion data in the public domain. For example, Hawn used the return level to show that the pit depth along a buried pipeline is not expected to exceed a prescribed limit in 5280 times the initial inspection area[23].

The M^{th} return level, r_M is defined as the thickness threshold which will be surpassed only once (on average) in M blocks of an inspection. The expected number of thickness measurements greater than the return level in a sample of M thickness minima is given by:

$$E(N(x > r_M)) = MP(x > r_M) = M(1 - \Phi(x|\mu, \sigma, k)) \tag{4.4}$$

where $P(x > r_M)$ is the probability of measuring a thickness measurement greater than the return level. From the definition of return level ($E(N(x > r_M)) = 1$),

$$\begin{aligned} MP(x > r_M) &= 1 \\ P(x > r_M) &= \frac{1}{M} \\ \Phi(x|\mu, \sigma, k) &= 1 - \frac{1}{M} \end{aligned} \tag{4.5}$$

Equation 4.5 is visualised in figure 4.1 as the area bound to the right of the red dashed line and the probability density function[1] It is equivalent to the M^{th} quantile of the GEVD. The position

[1] An explanation of the probability density function can be found in chapter 2

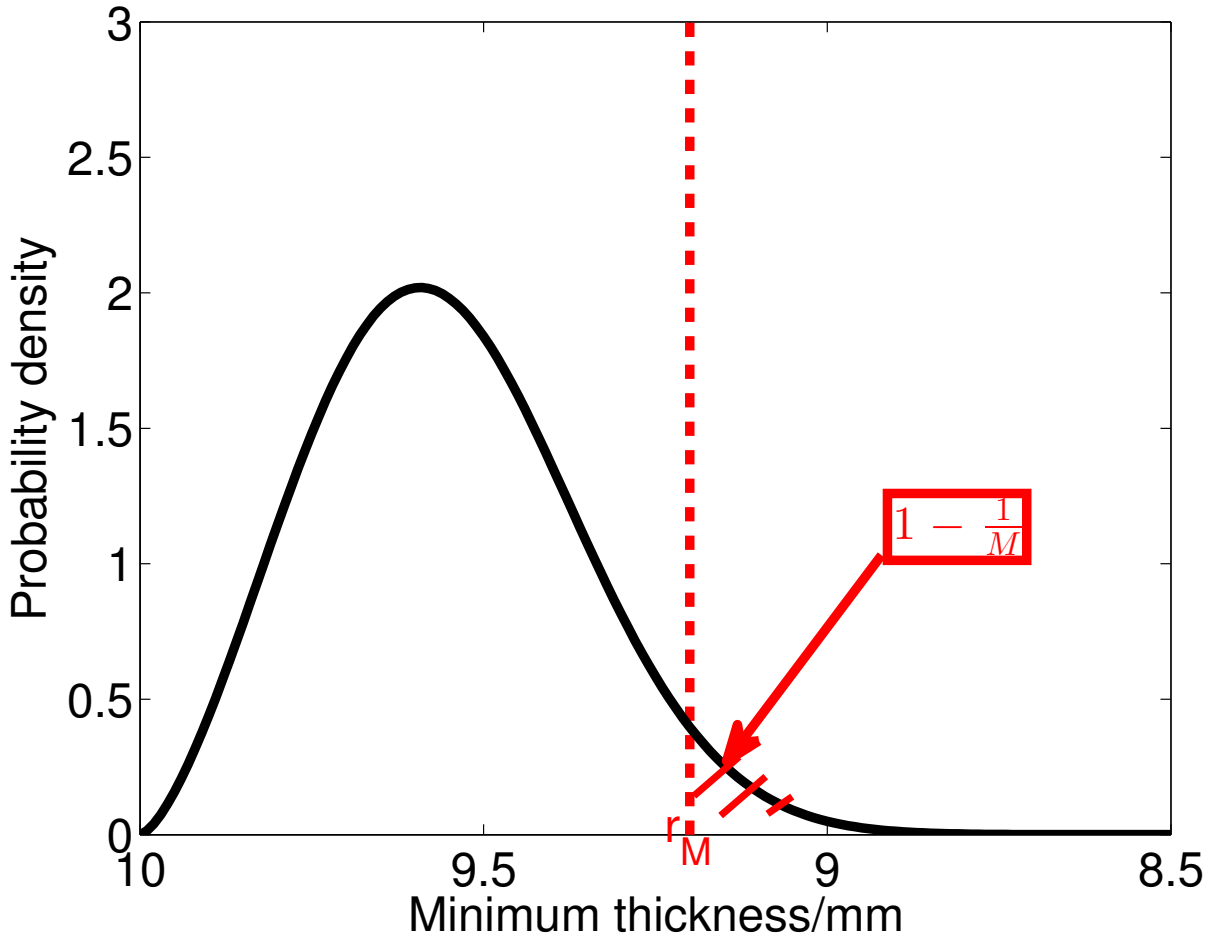


Figure 4.1: An example of a probability density function of a generalized extreme value distribution. The red dashed line is the M^{th} return level and the red shaded area is defined in equation 4.5.

of the red dashed line can be calculated be rearranging the GEV distribution (a derivation can be found in Coles [70]):

$$r_M = \mu - \frac{\sigma}{k}\left[1 - \left\{-\log\left(1 - \frac{1}{M}\right)\right\}^{-k}\right] \tag{4.6}$$

This is the value of thickness that the model predicts will be exceeded at least once in M blocks. It can be interpreted as an estimation of the smallest thickness that is expected to be found in an area the size of M blocks.

Extrapolations to areas larger than the inspection region can be performed by calculating the return level corresponding to a number of blocks greater than the initial sample of minima. For example, an inspector could estimate the minimum thickness in an area twice the size of the inspection area by calculating the return level corresponding to $M = 2N_{blocks}$. This return level is the threshold that the model expects would not be exceeded in an area two times the initial inspection area.

Alternatively, an inspector could extrapolate with the distributional method developed by Glegola. Glegola showed that the distribution of thickness measurements in an area larger than one block will be a GEVD with scaled distributional parameters [59]. Prospectively, this method has the capability to provide more information than the return level method. An extrapolated distribution would be able to provide a probabilistic description of the possible condition of the uninspected area, in contrast to a limit which the inspector does not expect to be surpassed.

However, for the purposes of PCI, much of this information is superfluous. An inspector is mainly interested in ensuring that the thickness of a component remains inside safe limits. This can be achieved much more easily using the return level method than the distributional method as an inspector will only have to calculate the return level to achieve this rather than the parameters of the extrapolated distribution. Subsequently, the focus of this chapter will be on the return level method, as it promises to be a more suitable tool for the extrapolation of inspection data.

It is impractical to directly validate an extrapolation, it would require data from outside of the inspection area, which is unavailable to an inspector. However, one can show that the model constructed is reasonable given the available data and the assumptions made by EVA (which can be achieved using the blocking method in chapter 3). Once the assumptions made by the model have

been shown to be reasonable, the uncertainty associated with the extrapolations can be quantified by calculating a 95% confidence interval around the return level.

In this chapter the uncertainty associated with extrapolation form extreme value models will be quantified for the first time. To the author's knowledge no other attempts have been made to quantify the possible errors that could occur when constructing an extreme value model. This is an important step in developing a framework in which extreme value models of inspection data can be used with confidence in industry.

4.1.1 Confidence intervals for the return levels

The uncertainty around the return level that arises from statistical variations can be quantified with a confidence interval. A 95% confidence interval for the return level is the bounds which contains 95% of the possible estimates of the return level. For example, if 100 models were generated from different samples from inspections of the component (with the same thickness distribution), 95 of the estimates will lie within these bounds. Confidence intervals are a reflection of the inspector's belief in the return level estimate.

Confidence intervals around the return level can be calculated using the profile likelihood method. The likelihood function in equation 4.2 can be written as a one dimensional function of any of the three distributional parameters μ, σ and k by maximising it with respect to the other two. For example, equation 4.2 can be written in terms of only μ by maximising with respect to σ and k:

$$L(\mu) = \max_{\sigma,k} \left[L(\mu, \sigma, k|\mathbf{x}) \right] \tag{4.7}$$

Equation 4.7 is known as the profile likelihood for μ. This equation can be parametrised in terms of the return level. This is achieved by writing the location parameter, μ, in terms of the return level in equation 4.6 and substituting into equation 4.7:

$$\mu = r_M + \frac{\sigma}{k} \left[1 - \left\{ -\log\left(1 - \frac{1}{M}\right) \right\}^{-k} \right] \tag{4.8}$$

Thus, for the return level, the profile likelihood is Coles [70, p. 57]:

$$
\begin{aligned}
L\left(r_m\right) &= \max_{\sigma,k}\left[L\left(r_m,\sigma,k|\mathbf{x}\right)\right] \\
&= N\ln\left(\sigma\right) \\
&- \left(1+\frac{1}{k}\right)\sum_{i=1}^{N}\ln\left(\left[1+k\left(\frac{x_i-\left(r_M-\frac{\sigma}{\xi}\left[1-\left\{-\log\left(\frac{1}{M}\right)\right\}^{-\xi}\right]\right)}{k}\right)\right]\right) \\
&- \sum_{i=1}^{N}\left[1+k\left(\frac{x_i-\left(r_M-\frac{\sigma}{\xi}\left[1-\left\{-\log\left(\frac{1}{M}\right)\right\}^{-\xi}\right]\right)}{k}\right)\right]
\end{aligned}
\tag{4.9}
$$

Equation 4.9 can be used to calculate a confidence interval for the return level with the deviance function. The deviance function is defined as:

$$
D(r_M) = 2(L(\hat{r_M}) - L(r_M)) \tag{4.10}
$$

where $L(\hat{r_M})$ and $L(r_M)$ are the profile likelihoods for an estimate of r_M, $\hat{r_M}$, and the true value of r_M. It can be shown that the statistic $D(r_M) \sim \chi^2_d$, where χ^2_d is the chi-squared function [71] with d degrees of freedom. d describes the number of factors affecting the likelihood function.

As the deviance statistic follows a χ^2 distribution, we can calculate the bounds in which 95% of its estimates lie. This is defined as the set $\{\theta : D(\theta) \geq c_\alpha\}$ where c_α is a $(1-\alpha)$ quantile of the χ^2_d distribution. This set is described graphically by figure 4.2.

In all EVA examples in the literature, extrapolated return levels are reported as a single value. This is a misleading representation of the data. No information about the uncertainty associated with the extrapolation is revealed. Confidence intervals allow an inspector to quantify some of the uncertainty associated with the model, which is key understanding the conclusions drawn from extreme value models.

4.1.2 Chapter Overview

The research in this chapter is split into three sections. The first is a study of the return level as a method of extrapolation which was performed to investigate whether the return level behaves

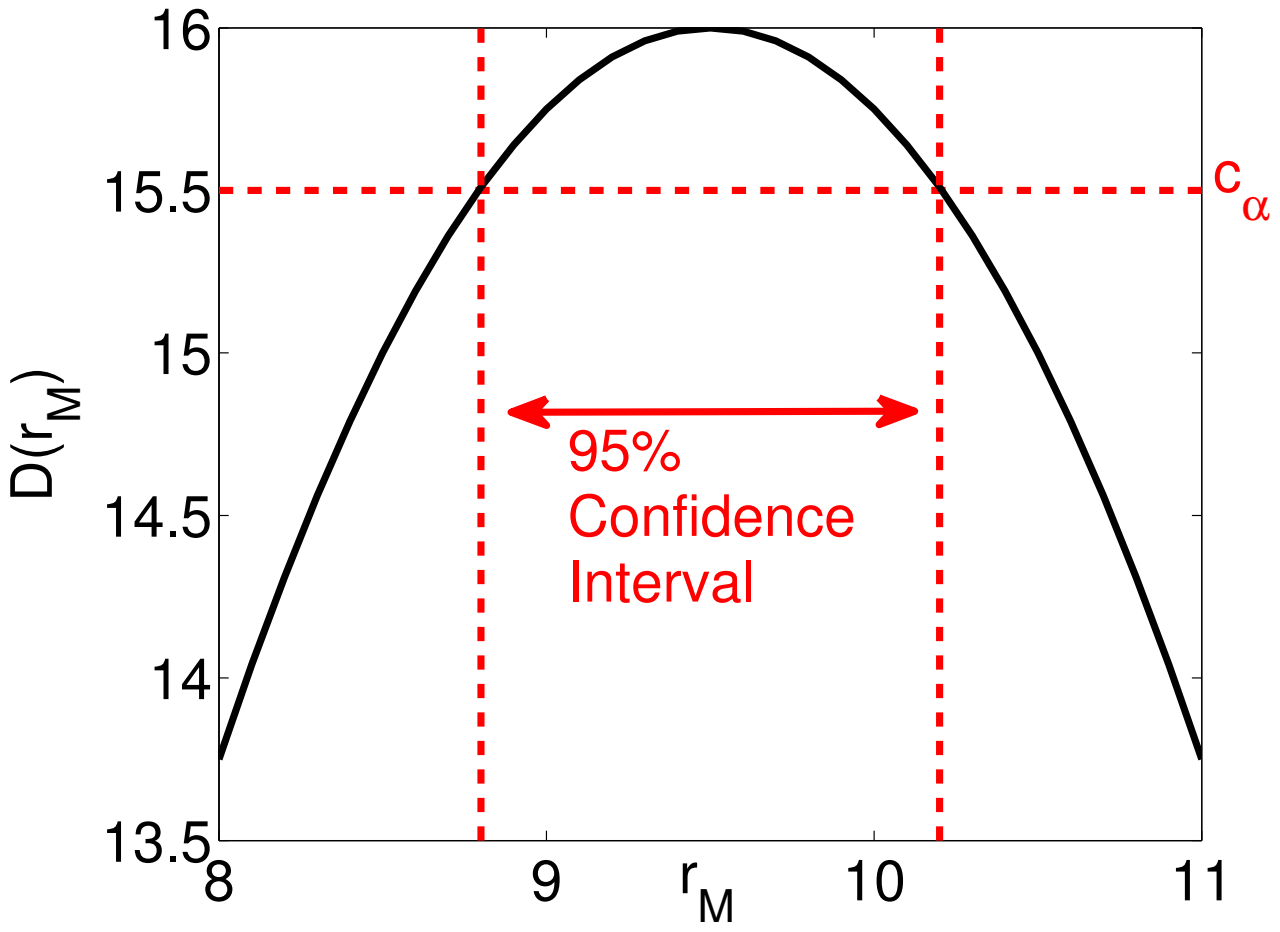


Figure 4.2: An example of a deviance function for a return level, r_M, of a generalized extreme value distribution. A (1-α)% confidence interval is the set of r_M for which $D(r_M)$ exceeds c_α.

as expected for simulated inspection data. Published research which uses return levels for extrapolation from inspection data has used the return level without verifying whether it behaves as expected for corrosion data [23, 26].

To verify the use of the return level with inspection data, an extreme value model is generated from simulated inspection data. Return levels corresponding to an increasing number of blocks are calculated using this model and they are compared to data from a simulated inspection of an area much larger than the initial inspection. This data is used to verify whether the return level performs as an effective metric for extrapolation.

In the second study, the calculation of the confidence intervals using the profile likelihood is described. A large number of extreme value models are generated from different sets of simulated inspection data. From each model return level confidence intervals are calculated for different sample and extrapolation sizes. This data is used to examine the mean width of confidence intervals with increasing extrapolation area.

The third and final section details a direct analysis of the accuracy of the extrapolation. Extreme value models were generated from a simulated inspection of a very large surface. The extreme value model was used to calculate the return level and its confidence intervals corresponding to an extrapolation to the total size of the surface. The return level and confidence intervals were compared to the smallest thickness across the surface. This analysis is used to determine how well extreme value models perform when they are used to extrapolate to larger areas of a component and the likelihood of an incorrect assessment using this technique.

4.2 Validating the return level for inspection data

The aim of these simulations is to show that the return level and the calculated confidence intervals are useful metrics for condition assessment purposes. With this in mind samples of 1000 48 by 48mm Gaussian height distributed Gaussian correlated surfaces and 1000 48 by 48mm exponential height distributed Gaussian correlated surfaces with mean thickness 10mm, RMS=0.1mm and $\lambda_c = 2.4mm$ were generated using the rough surface algorithm described in chapter 3. These samples are equivalent to inspections of $2,304,000mm^2$ ($2.3m^2$) of a component. Extreme value

models were constructed from a 240 by 240mm Gaussian and a 240 by 240mm exponential surface with RMS heights 0.1mm and $\lambda_c = 2.4mm$. The sample minima were extracted from the surfaces using a block size of 48mm. The model provides a description of the minimum thickness in an area the size of a single block and was compared to the sample of 1000 48 by 48mm Gaussian correlated surfaces.

Extrapolations using these models can, conveniently, be considered as extrapolations to multiples of this block size. The model was used to calculate return levels and their corresponding confidence intervals to areas ranging from 2 to 1000 blocks. These areas are rescaled in terms of the initial inspection area, which consisted of 25 blocks, to define the extrapolation ratio:

$$ER = \frac{EA}{IA} \tag{4.11}$$

where ER is the extrapolation ratio, EA is the extrapolated area and IA is the inspected area. In terms of the extrapolation ratio, return levels corresponding to extrapolation ratios ranging 0.08 to 40 were calculated. The sample of 1000 surfaces corresponds to $ER = 40$.

Figure 4.3 and figure 4.4 show box plots of the thickness minima selected from 1000 48 by 48mm correlated Gaussian surfaces and 1000 48 by 48mm correlated exponential surfaces respectively. The length of the box is the inter-quartile range of the sample of thickness minima, which is a measure of spread of the measurements. The solid line in the middle of the box is the median thickness minimum, whilst the whiskers contain 99% of the thickness minima. Return levels corresponding to the extrapolation ratios on the x-axis were calculated from the model generated from both the Gaussian and the exponential surfaces using equation 4.6. These are shown as black crosses. 95% confidence intervals around these return levels are shown as blue crosses. With an increasing extrapolation ratio, the return levels decrease, indicating that in a larger area a smaller minimum thickness is expected.

Broadly speaking figures 4.3 and 4.4 show the same patterns of behaviour. This is encouraging as it indicates that return level provides a measure that is independent of the distribution. However, the widths of the confidence intervals are different. For example, at an extrapolation ratio of 40, the Gaussian data has a confidence interval width of 0.2mm compared to 0.5mm for the exponential

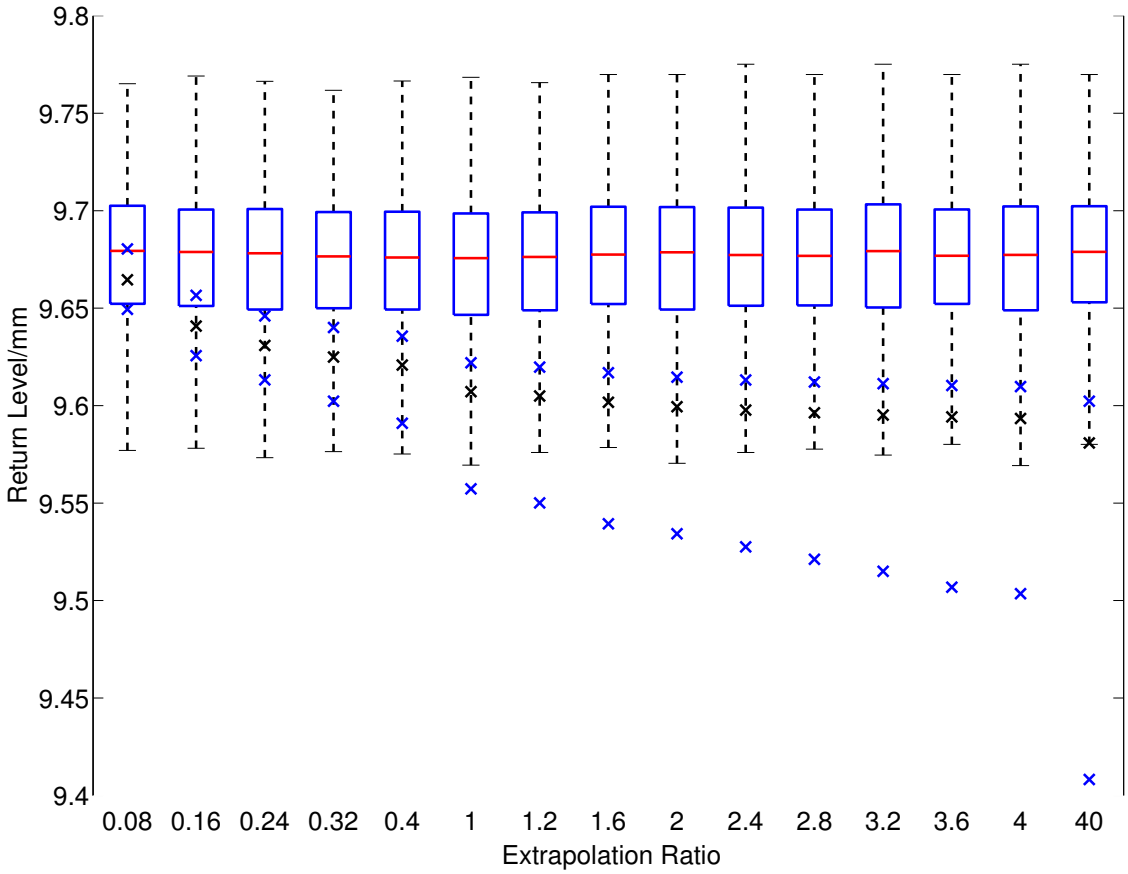


Figure 4.3: A histogram of the minimum thickness measurements from 1000 48 by 48mm correlated Gaussian surfaces with mean thickness 10mm, RMS=0.1mm and correlation length 2.4mm. The return levels (black crosses) and confidence intervals (blue crosses) at each extrapolation ratio calculated from an extreme value model constructed from a 240 by 240mm Gaussian surface of the same statistics have been overlaid.

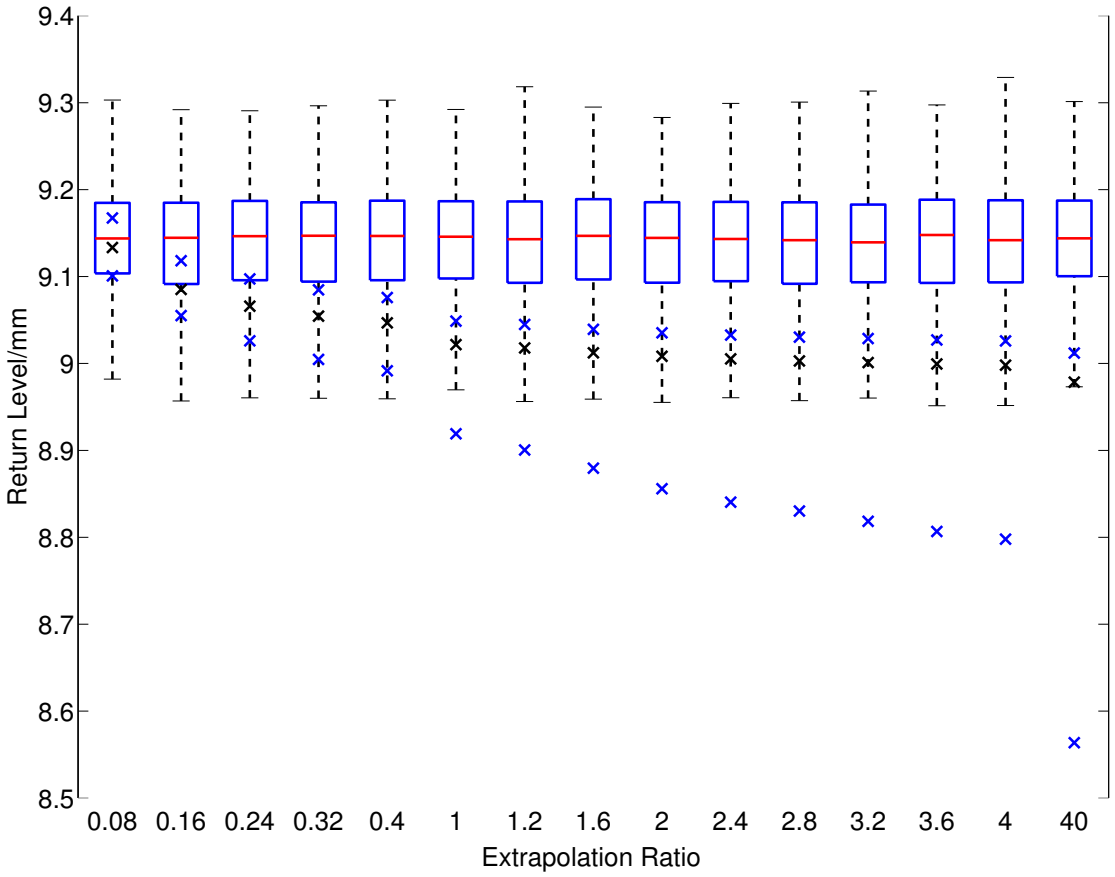


Figure 4.4: A histogram of the minimum thickness measurements from 1000 48 by 48mm correlated exponential surfaces with mean thickness 10mm, RMS=0.1mm and correlation length 2.4mm. The return levels (black crosses) and confidence intervals (blue crosses) at each extrapolation ratio calculated from an extreme value model constructed from a 240 by 240mm exponential surface of the same statistics have been overlaid.

data. This is due to differences in the shape of the generalised extreme value likelihood functions arising from the Gaussian and exponential distributions. It indicates that the width of the return levels could have some dependence of the underlying thickness distribution alongside the quality of the data used to construct the model.

For extrapolation ratios less than 1, the return level is modelling the minimum thickness in an area less than the inspection area (an area from which the inspection data is taken). Subsequently, the confidence intervals for this region are narrower and the return level provides a better description of the data than the return levels for extrapolation ratios greater than 1. For example, the return level for the extrapolation ratio of 0.08 is very close to the median of the sample of thickness minima. This is expected as the return level for this extrapolation ratio will be exceeded [2] once every two blocks, so around 50% of the thickness measurements should be less than this value. This trend is continued with the return levels for the extrapolation ratios ranging from 0.16 to 1, with the return levels matching up with the appropriate quantiles in the sample.

Once the extrapolation ratio increases to greater than 1, the exact value of the return level predicted by the model does not necessarily match up with the correct quantile of the thickness sample. This is expected as extrapolations, by their very nature, will not provide a perfect description of the data. In these situations, the confidence interval around the return level is key to interpreting the results of the model. For example, as the extrapolation ratio is increased to the point where it corresponds to the size of the sample (ER=40), the return level gets closer and closer to providing bounds for the smallest thickness measurement in the sample. For the sample of data used for this study, there are no thickness measurements less than the return level with an extrapolation ratio of 100. However, for a different sample, it is feasible that there could be thickness measurements less than this value due to the extrapolation leading to inaccurate prediction for return level.

However, the confidence interval around the model's estimate of return level will contain the true value of the return level 95% of the time. The confidence interval can be interpreted as the bounds inside which the smallest thickness measurement in an extrapolated area will lie. Rather

[2]To clarify, by *exceed* the return level, I mean that there will be at least one thickness measurement less than the return level.

than reporting just a single value for the minimum thickness, the return level, an inspector can state a range of values and a measure of his confidence that the minimum thickness lies in this bounds. This will allow a plant operator to make a decision about the condition of a component with knowledge of the uncertainty around the inspector's estimate.

The width of the confidence intervals increases with the extrapolation ratio. Ideally an inspector will try to minimise the extrapolation ratio for the extrapolation they are performing as it will minimise the width of the confidence bounds around the return level. Knowledge of how large a confidence interval will be on average, as a function of extrapolation ratio, will allow an inspector to make a decision about the amount of inspection area required to obtain bounds on the minimum thickness of a given width. This is addressed in the next section.

4.3 Errors associated with extrapolation

Direct validation of an extrapolation requires data from outside the inspection area. This is impossible for an inspector to obtain in many applications of PCI due to, for example, constrained access to the component under inspection. However, the inspector can check that the inspection data meets the assumptions made by EVA (using the blocking algorithm in chapter 3). If the data meets the assumptions made by EVA, then the model generated from the inspection data will be able to provide reasonable description of the condition of the component outside the inspection area.

In the previous section (section 4.2) it was shown that 95% confidence intervals around an extrapolated return level are likely to contain the smallest thickness in the extrapolated area. The width of the confidence intervals are a reflection of the uncertainty around the value of the return level. Larger confidence intervals imply greater uncertainty around the return level predicted by the model. There are a number of factors, determined by the way the EV model was constructed, which determine the width of the confidence intervals. For example, if the inspector has a large sample of minima, the confidence intervals will be narrower and they will be more confident in the predictions made by the model. The sample size is determined by the ratio of the block size to the inspection area. If the minima have been selected using a block size which is a small fraction of

Block Size/mm	Number of minima
24	100
30	64
40	36
48	25
60	16

Table 4.1: The different block sizes and the corresponding number of minina used to generate the extreme value models.

the inspection area, the measurements in the sample will not be representative of extremes of the distribution. The model will not provide a good description of the condition of the component.

In terms of an inspection the interpretation of the return level confidence interval is very important. An inspector may have a large sample of minima, which leads to narrow confidence intervals, but these minima may not be representative of the extremes of the distribution. Any predictions made by the model will underestimate the severity of the minimum thickness. In addition, the narrow confidence intervals could lead the inspector to believe they have a "good" estimate of the smallest thickness across the component. It is vital that the model is constructed using minima which are representative of extremes of the distribution and, as we saw in chapter 3, this is often best achieved with small samples of minima collected using a large block size.

Knowledge of how the confidence intervals behave with different inspection designs will allow for improvements in the design of partial coverage inspections. With this goal in mind many confidence intervals were generated using simulated inspections partitioned using different block sizes. The inspection data was simulated by generating 100 correlated Gaussian surfaces of size 240 by 240mm and 100 correlated exponential surfaces of the same size (using the algorithm described in 3). The surfaces all had an RMS height 0.1mm and a correlation length of 2.4mm.

From each surface an extreme value model was generated using block sizes corresponding to different numbers of minima, summarised in table 4.1. The return levels corresponding to extrapolation ratios ranging from 0.08 to 400 were calculated, along with the corresponding confidence intervals. For each set of simulations the average width of the confidence intervals is calculated and expressed as a percentage of the return levels. This percentage is plotted as a function of the extrapolation ratio.

Confidence intervals are an expression of the inspector's belief in the estimate of the return level. In the case of the confidence intervals calculated here, they are the bounds in which the true value of the return level will lie 95% of the time. As the extrapolation ratio is increased past the size of the inspection area, the size of the confidence intervals relative to the estimate of the return level will increase. This reflects that, as the extrapolation ratio increases past 1, there is much less certainty of any predictions made by the model.

Figures 4.5 and 4.6 show the average size of the confidence intervals over all the models, expressed as a percentage of the return level, as a function of extrapolation ratio. Although these figures were produced using data collected from surfaces with different distributions, they show very similar behaviour. The width of the confidence intervals is nearly constant up to an extrapolation ratio of 1. For extrapolation ratios less than one, the return levels correspond to areas less than the inspection area. Therefore, the width of the confidence interval is a reflection of the uncertainty resulting from the sampling variation.

Once the extrapolation ratio increases past 1, the return levels correspond to areas greater than the inspection area. Consequently the average size of the confidence intervals begins to increase as there is more uncertainty due to the extrapolation to an area larger than the inspection area. The rate at which the width of the confidence interval increases with increasing extrapolation ratio, as the level of uncertainty increases very rapidly with the extrapolation ratio.

Past a certain point, the confidence intervals indicate that estimates of the return level are no longer a useful tool. The inspector's role to ensure that the component under inspection is safe for operation. This includes ensuring that the smallest wall thickness measurement found is within acceptable bounds[3]. If the return level confidence interval is too wide then an inspector would not, with any certainty, be able to draw conclusions about the condition of the component. For example, in both figure 4.5 and 4.6, at an extrapolation ratio of 400 the average width of the confidence intervals is at least 30% of the size of the return level, e.g. for a return level of 5mm and confidence interval of 3.5 to 6.5mm. This is a large range of values in which the return level could lie. In this case the inspector should either collect more data to improve this extreme value

[3]"Acceptable bounds" can vary from application to application. The author here assumes that the inspector has been given some figure for an unacceptable level of damage by the company ordering the inspection.

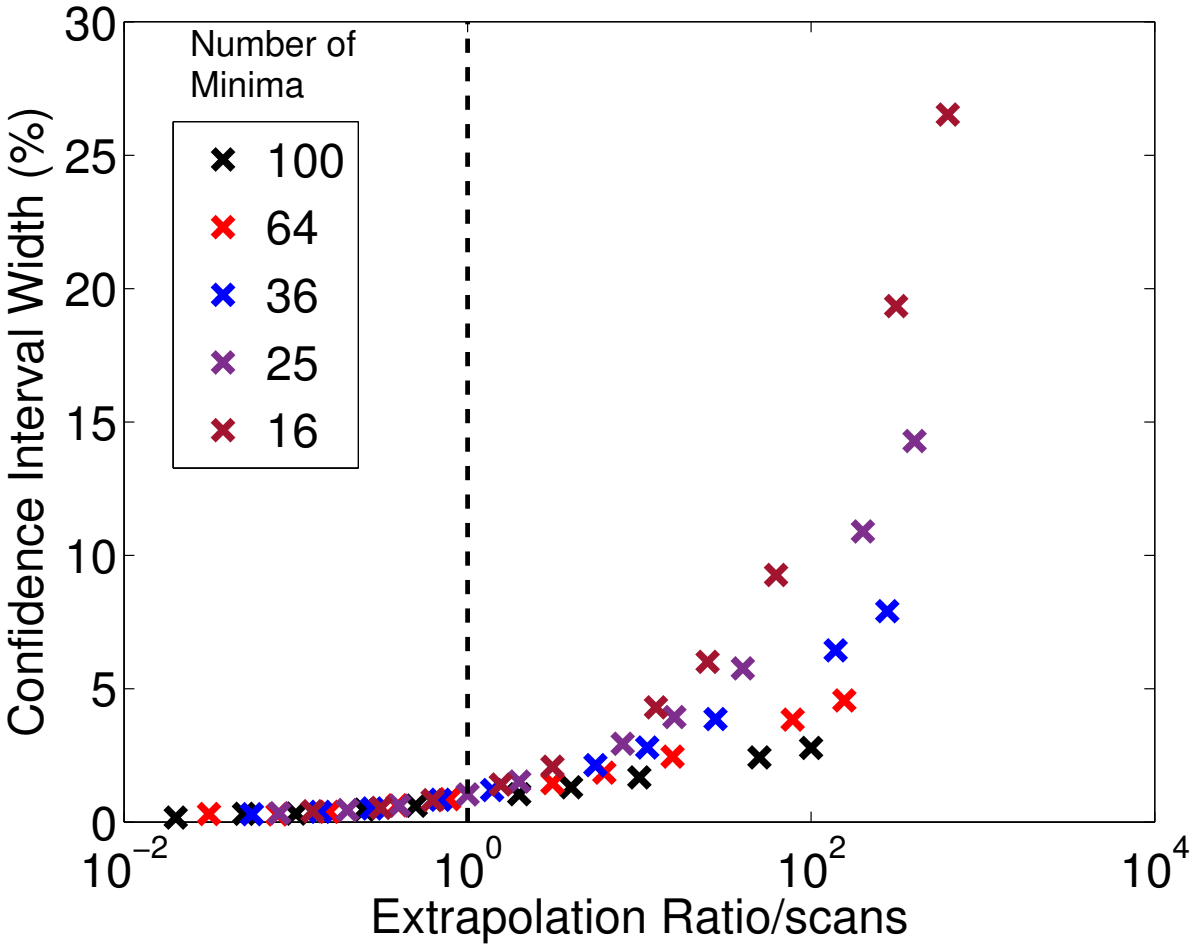


Figure 4.5: The average width of the return level confidence intervals, expressed as a percentage of the return level, calculated from 100 extreme value models constructed from 240mm by 240mm Gaussian surfaces, as a function of extrapolation ratio. Up to an extrapolation ratio of 1 (the size of the inspection area), the width of the confidence intervals is constant. Past this point the model is being used to extrapolate and the confidence intervals grow. The rate of growth is determined by the number of minima used to construct the model.

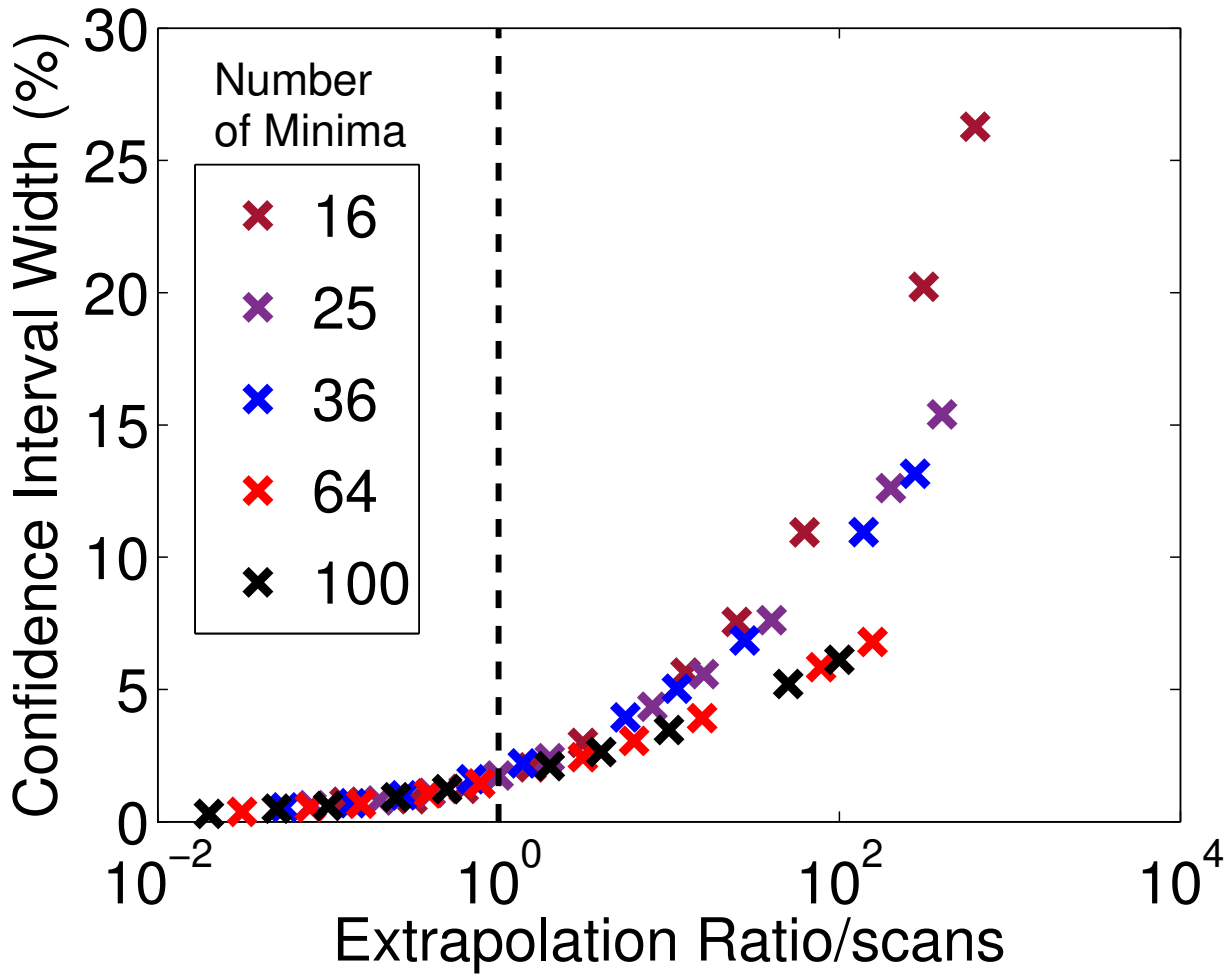


Figure 4.6: The average width of the return level confidence intervals, expressed as a percentage of the return level, calculated from 50 extreme value models constructed from 240mm by 240mm exponential surfaces, as a function of extrapolation ratio. Up to an extrapolation ratio of 1 (the size of the inspection area), the width of the confidence intervals is constant. Past this point the model is being used to extrapolate and the confidence intervals grow. The rate of growth is determined by the number of minima used to construct the model.

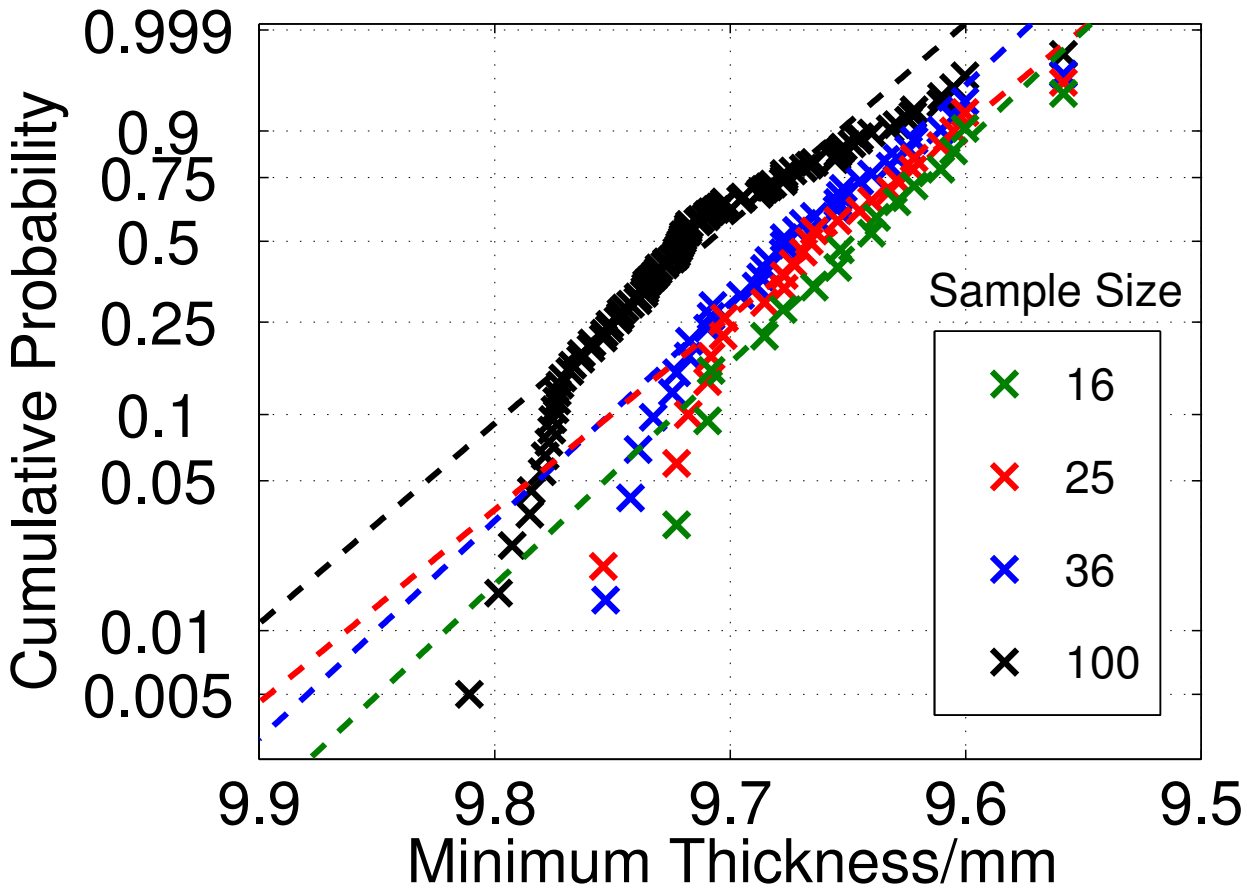


Figure 4.7: Extreme value probability plots for sets of minima extracted from a 240mm square correlated Gaussian surface with RMS=0.1mm and $\lambda_c = 2.4mm$. Thickness minima were extracted using block sizes of 24mm (black), 40mm (blue), 48mm (red) and 60mm (green). As the block size is increased, the extreme value models (black dashed lines) become better descriptions of the smallest minima in each sample.

model, or perform a full inspection of the component.

From this data it seems as if using more minima is the correct approach, the widths of the confidence intervals are consistently smaller using more minima. However, this should be considered carefully. To obtain a larger sample, the minima are selected using a smaller block size. The smaller the block size, the lower the number of thickness measurements in each block. Each block is a smaller sample of the underlying thickness measurement distribution. As the size of the sample of thickness measurement in the block decreases, the expected number of extremes in the sample will decrease. Consequently, it is likely that the minimum thickness measurement in this block may not be representative of the extremes of the distribution. A model constructed from a set of minima from these blocks may be inaccurate.

Figure 4.7 shows probability plots of an example model generated from a square Gaussian surface of size 240mm by 240mm with block sizes from to 24-60mm (100 down to 9 minima). A probability plot is a method of graphically comparing the model to the set of minima used to generate it. The y-scale is designed such that the EV model fitted to each set of data is represented by the straight dashed lines. Estimates of the probability of each thickness measurement are calculated empirically using equation 2.1 and plotted as crosses. If the model is a good fit the crosses will lie close to the dashed line.

The models generated using smaller samples of minima (25 and 16) provide better descriptions of the smallest minima in the sample, with R^2=0.99 and 0.98 respectively. This is in contrast to the model generated using 100 minima, which has R^2=0.91. The 100 minima model provides a good description of the majority of the minima. However, the smallest minima in the sample do not lie on the straight line. It fails to provide an adequate description of the smallest thicknesses in the inspection area. This is a key prerequisite for making engineering decisions based on extreme value models. The model must provide a good description of the smallest thicknessess that could occur. It is important that the thickness minima used to construct the extreme value model are taken from the areas of the component most likely to undergo the worst degradation.

Any inaccuracy in the model will propagate into the extrapolation. Models generated using a small block size, such as the 100 minima model in figure 4.7, will lead to inaccurate extrapolations,

as the minima selected from the blocks will not provide adequate examples of the thinnest areas of the component. In this next section, this problem is investigated using a series of simulated inspections taken from across very large surfaces.

4.4 Testing an extrapolation

In order to determine the quality of extrapolations from models generated using different block sizes, Gaussian and exponential surfaces (with RMS height of 0.1mm and correlation length of 2.4mm) of size 2400 by 2400mm, 12000 by 12000mm and 24000 by 24000mm were generated. From each surface 50 different inspection areas of 240 by 240mm (such that the total size of the surfaces corresponded to extrapolation ratios of 10, 50 and 100) were chosen and EV models were generated from each inspection area using block sizes ranging from 24 to 60mm (corresponding to minima sample sizes ranging from 100 down to 16). The EV models generated from these inspection areas were used to calculate return levels and confidence intervals for extrapolation ratios corresponding to the size of the surfaces. The return levels and confidence intervals were averaged over the 50 inspection areas and compared to the smallest thickness measurement across the surface.

The average return levels and average confidence intervals were plotted as a function of the number of minima in figures 4.8, 4.10 and 4.12 for the Gaussian surfaces corresponding to extrapolation ratios of 10, 50 and 100 respectively. The data collected from the exponential surfaces is shown in figures 4.9, 4.11 and 4.13. The red crosses are the average confidence bounds on the return level, the blue cross is the average return level and the dashed line is the smallest thickness across the entire surface. The average return level in each figure overestimates[4] (but is close to) the minimum thickness across the surface. This is expected as the return level is expected to overestimate the minimum thickness as, by definition, it is a threshold which will be exceeded at least once in the extrapolated area. The average return level decreases slightly as the number of minima in the sample is decreased. This is reflective of the fact that a smaller number of minima is collected using a larger block size, the average minimum thickness in the sample will be smaller,

[4]The return level is larger than the minimum thickness.

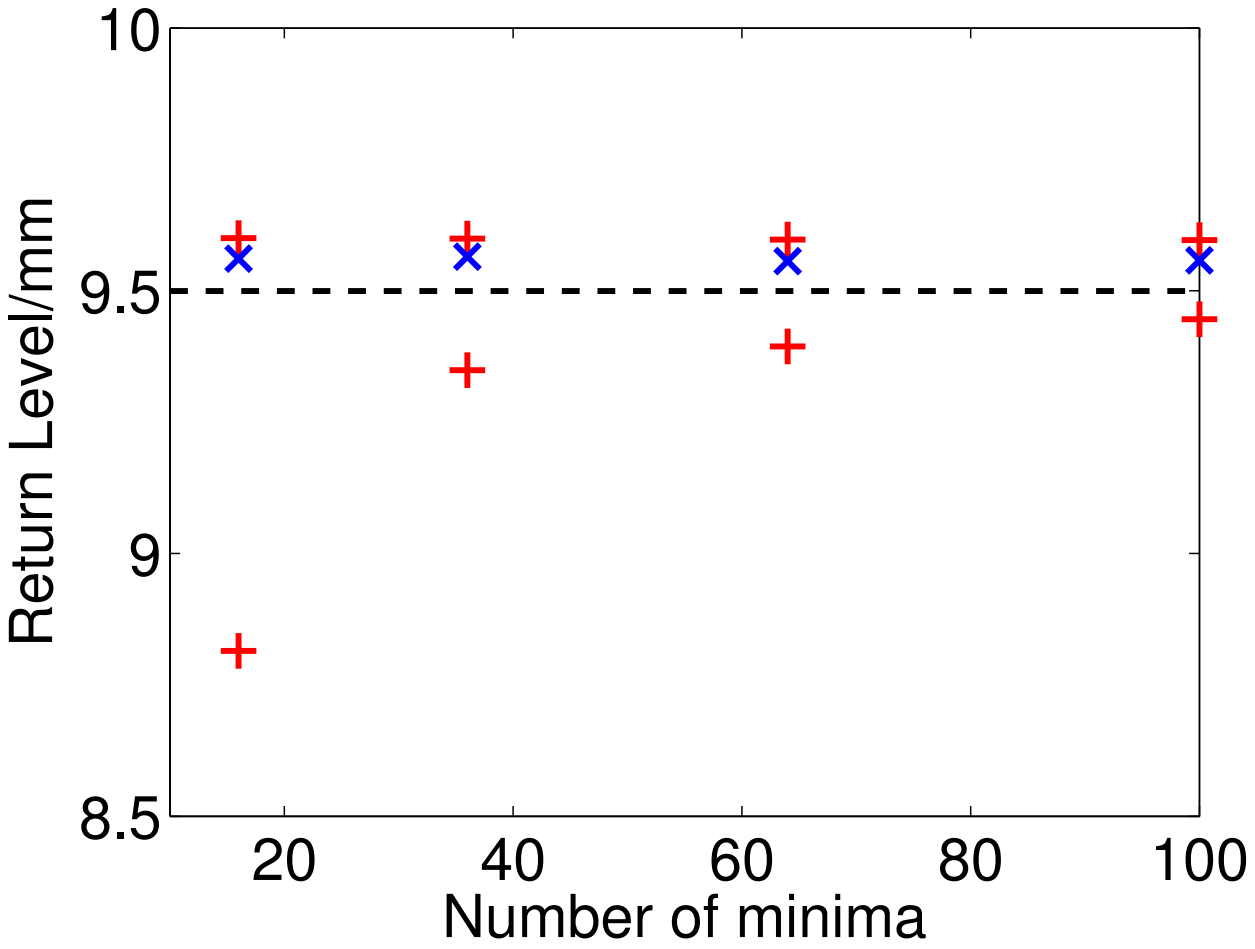


Figure 4.8: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 50 different extreme value models constructed using 240mm by 240mm inspections of a Gaussian surface corresponding to an extrapolation ratio of 10. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

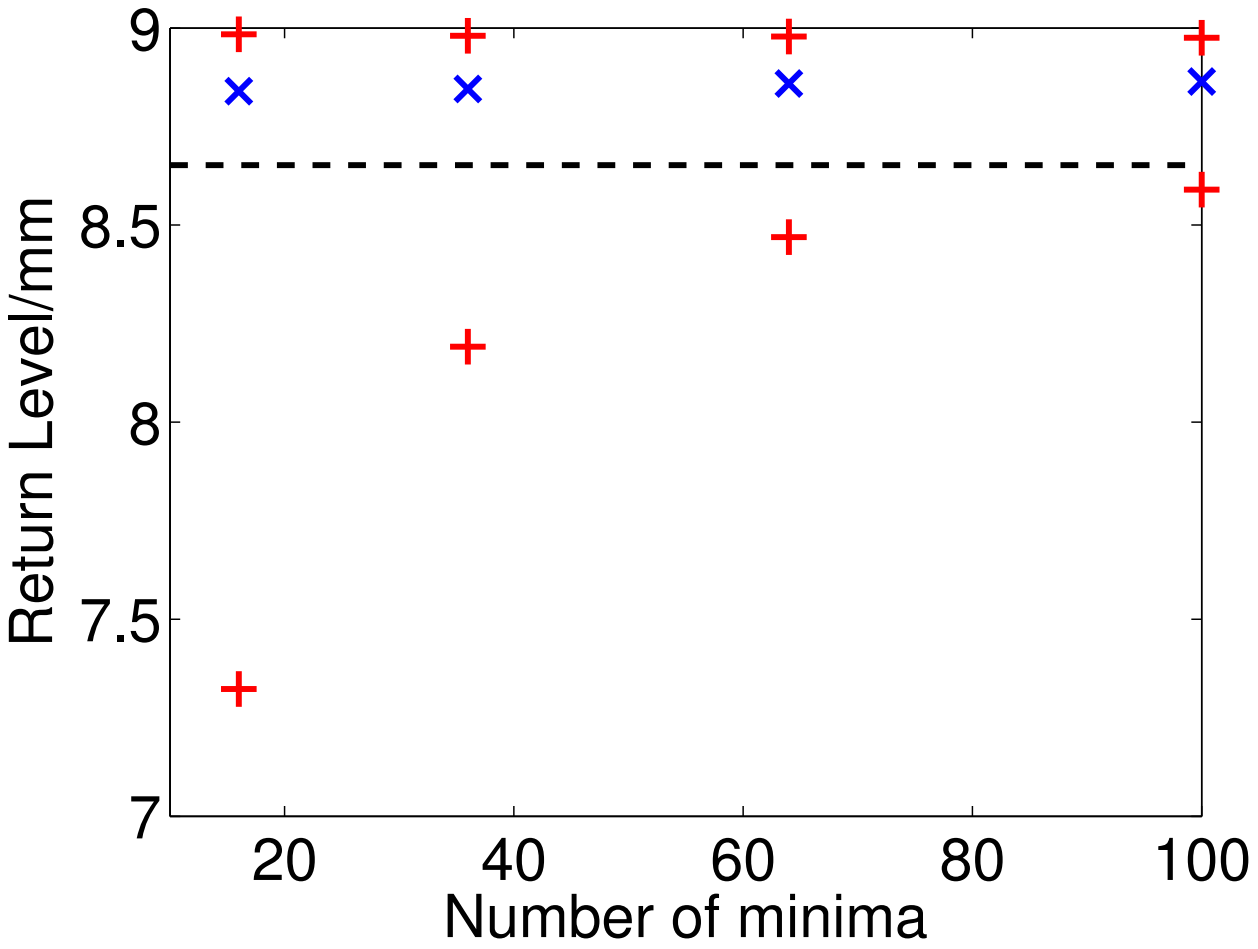


Figure 4.9: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 50 different extreme value models constructed using 240mm by 240mm inspections of an exponential surface corresponding to an extrapolation ratio of 10. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

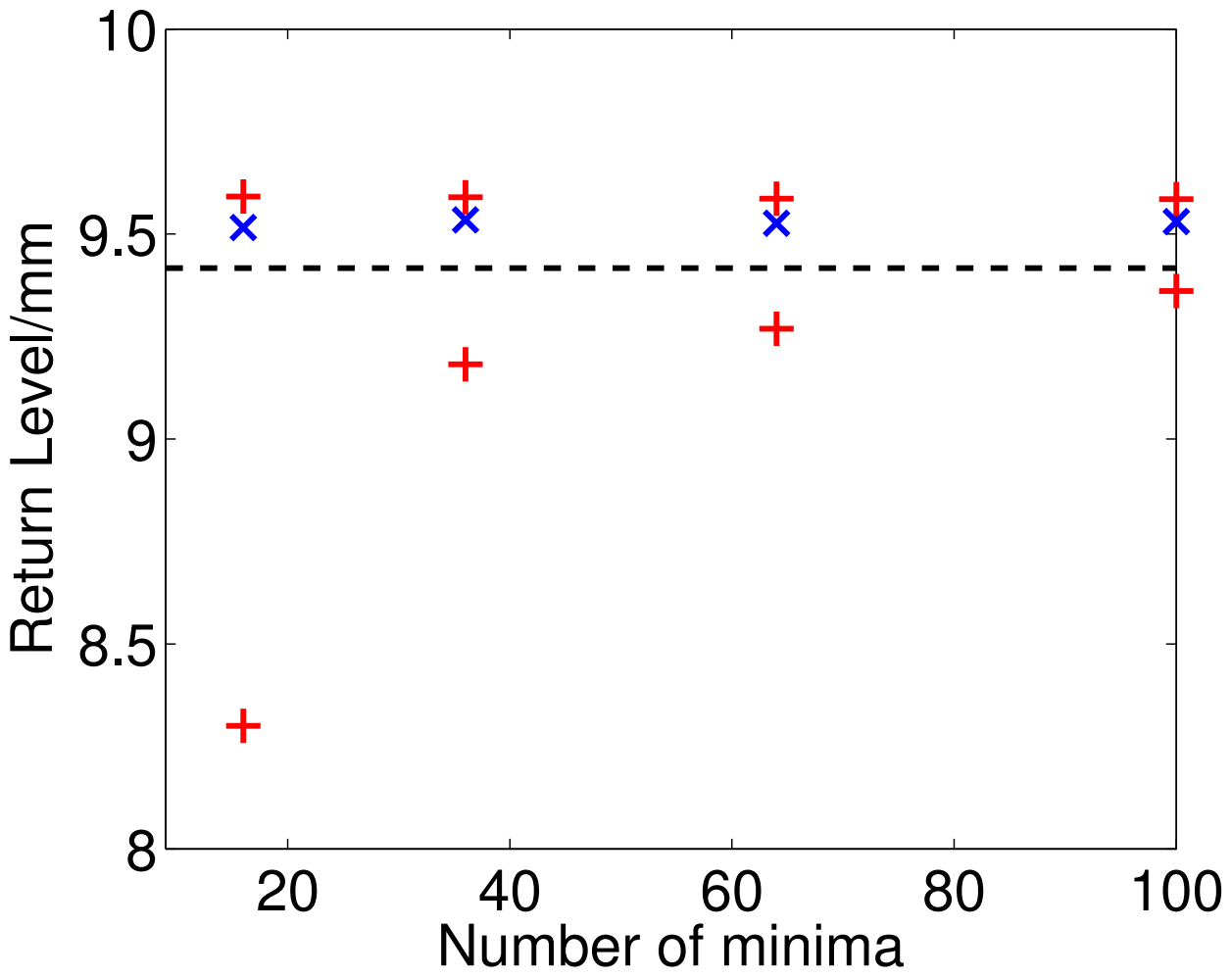


Figure 4.10: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 50 different extreme value models constructed using 240mm by 240mm inspections of a Gaussian surface corresponding to an extrapolation ratio of 50. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

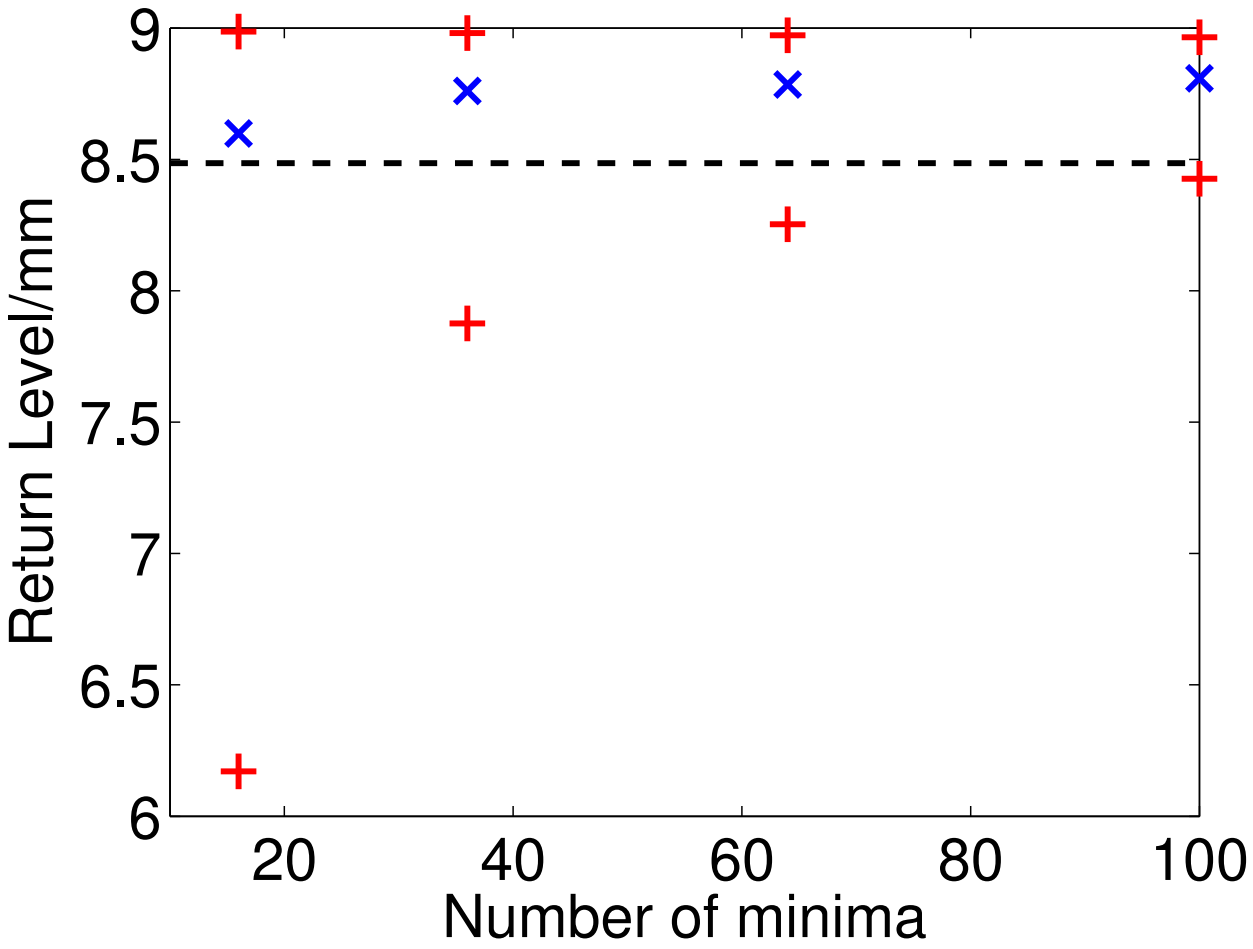


Figure 4.11: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 50 different extreme value models constructed using 240mm by 240mm inspections of an exponential surface corresponding to an extrapolation ratio of 50. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

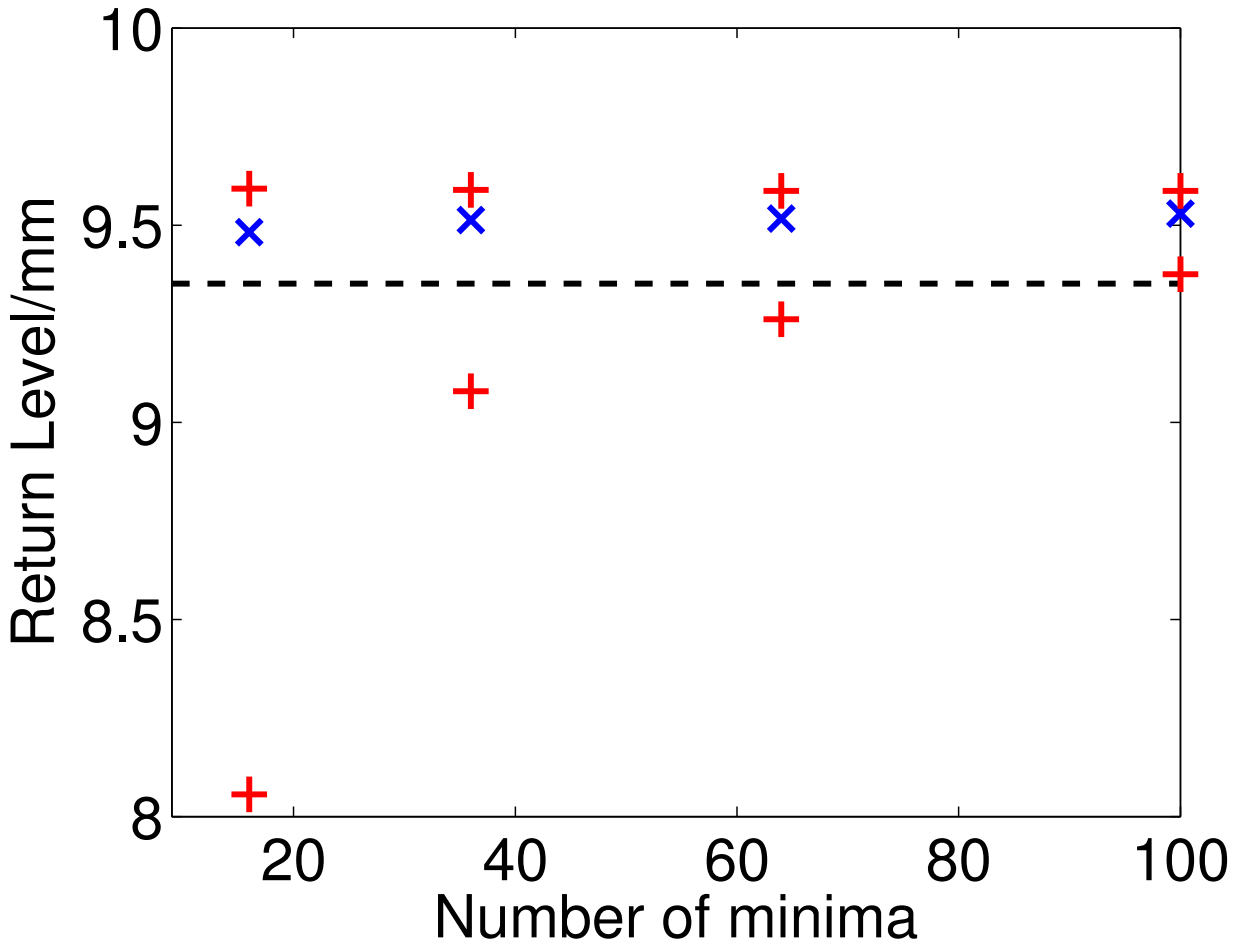


Figure 4.12: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 50 different extreme value models constructed using 240mm by 240mm inspections of a Gaussian surface corresponding to an extrapolation ratio of 100. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

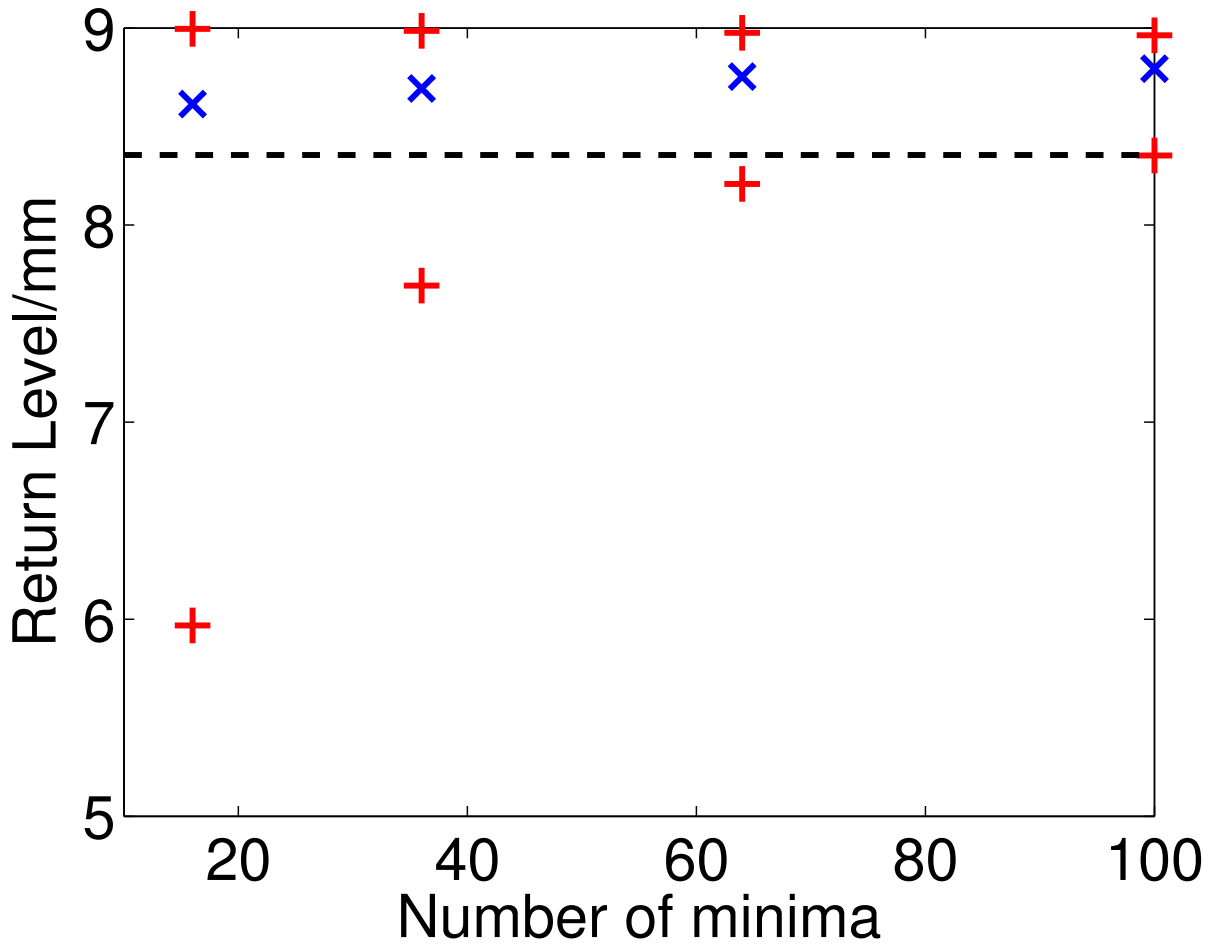


Figure 4.13: The average return level (blue crosses) and confidence intervals (red crosses) calculated from 100 different extreme value models constructed using 240mm by 240mm inspections of an exponential surface corresponding to an extrapolation ratio of 10. Each model was constructed using a different number of minima (x-axis) and the average confidence intervals were compared to the smallest thickness across the surface (black dashed line).

resulting in smaller return levels.

The size of the confidence intervals increases as the number of minima is decreased. This is in agreement with figures 4.5 and 4.6 in the previous section. Encouragingly, and as speculated, for extrapolation ratios of 10 and 50, the minimum thickness measurement lies within the average confidence interval. That is, on average, the confidence interval provides bounds in which the minimum thickness lies. However, this is the average behaviour over 50 different examples. Models from individual inspections can sometimes predict bounds which do not contain the minimum thickness. An example of this is shown in figure 4.14, which is an example of the return levels and confidence intervals for an extrapolation ratio of 50 predicted by a model from a single inspection area from the 12000 by 12000mm Gaussian surface. The minimum thickness is contained by both the confidence bounds calculated from the EV models constructed using 16 and 36 minima, but it lies outside the confidence bounds calculated using 64 and 100 minima.

As the size of the surface increases, the minimum thickness decreases. Simultaneously, the size of the average return level confidence intervals increases. If the extreme value models are providing an adequate description of the surface (on average), should remain well within the confidence intervals. However, the minimum thickness decreases at a faster rate than the growth of the confidence intervals. Consequently, there is a point where the minimum thickness is no longer contained by the confidence bounds. At the point the model is not providing an adequate description of the damage in the extrapolation area. The number of minima (block size) used to generate the model is key to ensuring that this is unlikely to occur.

In general, the confidence intervals from models generated using larger block sizes (fewer minima) more consistently provide a bounds for the minimum thickness. For the Gaussian surface, the return level confidence intervals calculated from the models with 16 and 36 minima contained the minimum thickness in 82% and 76% of cases respectively for an extrapolation ratio of 50, compared to 68% and 60% for sample size of 64 and 100. For an extrapolation ratio of 100, the number of models which predict bounds which contain the minimum thickness reduces further. The confidence intervals only contain the minimum thickness in 44%, 58%, 72% and 68% of case for models generated with 100, 64, 36 and 16 minima respectively. Arguably the models generated

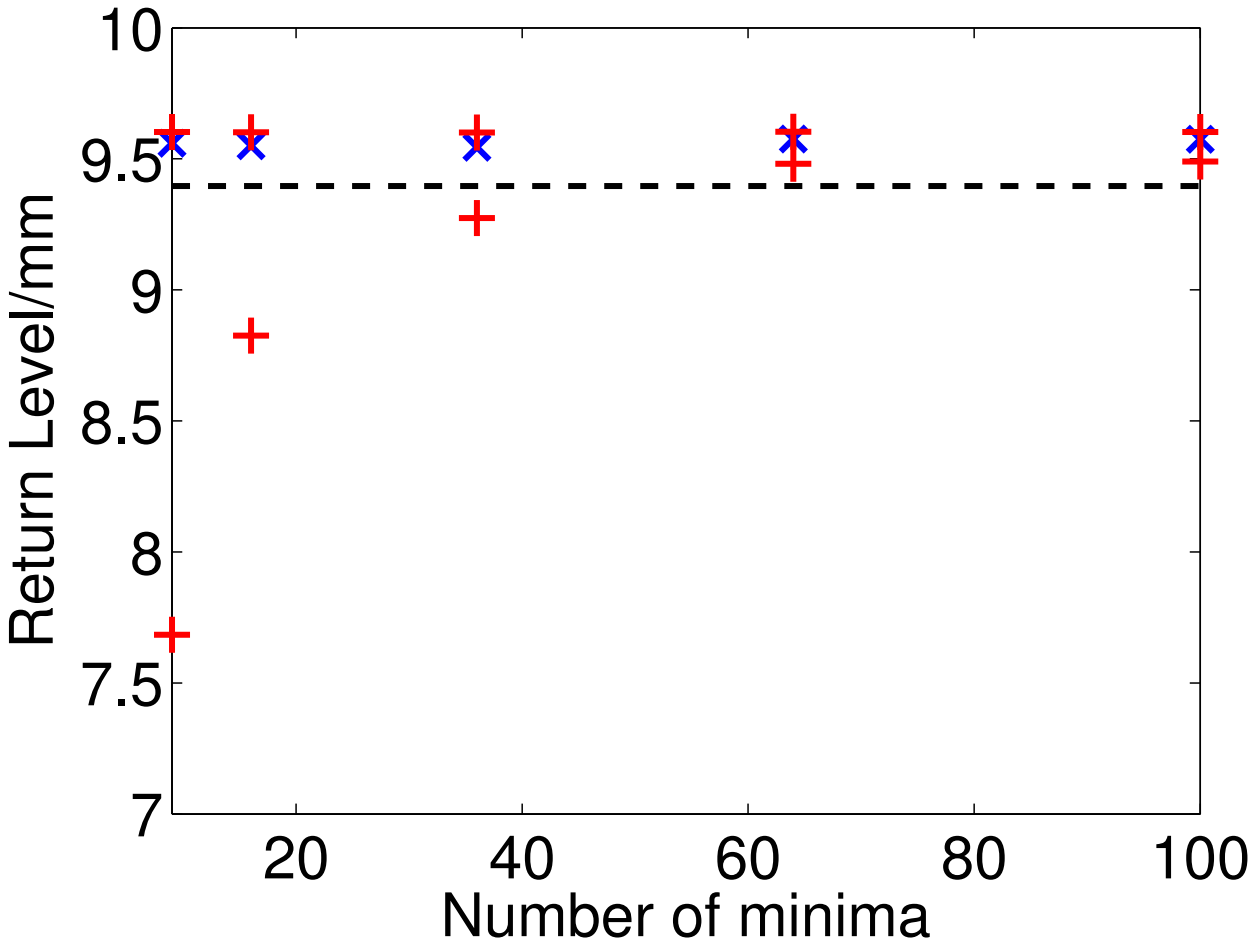


Figure 4.14: The return level (blue cross) and confidence intervals (red cross) corresponding to an extrapolation ratio of 50 calculated using an extreme value model constructed from a single, randomly selected, 240mm by 240mm inspection of a Gaussian surface corresponding to an extrapolation ratio of 50. The smallest thickness across the surface is shown as the black dashed line.

using the larger block sizes could still be used for extrapolations of this size, the minimum thickness will lie outside the bounds 30% of the time, but for most applications this error rate will be too high.

Both of these cases can be compared to the results for an extrapolation ratio of 10. For this case, the return level confidence intervals contained the minimum thickness for 76%, 82%, 86% and 90% of the models constructed using 100, 64, 36 and 16 minima respectively. The confidence intervals contain the minimum the majority of the time. However, it is clear that, even at a small extrapolation ratio a model generated with fewer minima is able to more accurately model the thickness measurements in the extrapolation area.

The confidence bounds calculated from models constructed using the exponential behaviour show a similar trend. For an extrapolation ratio of 10, the confidence interval contained the minimum thickness across the surface 92%, 86%, 66% and 54% of the time for sample sizes of 16, 36, 64 and 100 minima. With an extrapolation ratio of 50, the minimum thickness lay in the bounds 92%, 74%, 80% and 70% of the time of the time for sample sizes of 16, 36, 64 and 100. Finally, with an extrapolation of 100, the bounds contained the minimum thickness in 92%, 80%, 74% and 74% of cases.

The increased error rate that occurs with the combination of a model generated using a large sample of minima (small block size) and a large extrapolation ratio originates in the bias introduced into the EV model by the sample of minima. Sample minima collected using smaller block sizes are on average larger than those collected with larger block sizes. Therefore, the predictions made by an extreme value model will overestimate the size of the minimum thickness (predict it to be thicker than it actually is) in the extrapolated area. This leads to an overestimation of the return level. Consequently, the model will be biased, it will underestimate the probabilities of the smallest measurements of thickness occurring. This is demonstrated by figure 4.7 and discussed in the previous section (section 4.3). In addition, as the sample is larger, there is more evidence the model is 'correct' so the width of the confidence intervals increases at a slower rate than for smaller sets of minima (as shown in figure 4.5), which leads to the minimum thickness lying outside the confidence bounds of the return level.

4.5 Conclusions

Understanding uncertainty associated with extrapolation is a key part of using extreme value analysis for partial coverage inspection. In the existing literature, there have not been any studies of the errors associated with extrapolation from extreme value models, which has been a barrier to the use of extreme value models in the inspection industry. In this chapter a series of studies of the uncertainty associated with extrapolation using the return level method are described. Both Gaussian and exponential surfaces were studied at every stage of this study and it was found that models constructed from them performed similarly. This is encouraging as it provides evidence that confirms that EVA is independent of the thickness distribution across the surface as would be expected from the discussion in section 2.1.

To date, there has been no verification of the return level as a tool for extrapolation. This is addressed with the first set of simulations which compared the return levels calculated using an extreme value model constructed using simulated inspection data to simulated inspection data from a much larger area. The return levels at extrapolation ratios less than 1 match up with the equivalent quantiles in the larger inspection data. Once the extrapolation ratio increases past 1 the return level diverges from the equivalent quantile, due to the inherent inaccuracy of extrapolation. However, it's shown that the confidence intervals around the return level do contain the true value of the return level.

Through further study of this data it was suggested that the return level confidence interval could be used as a range for the minimum thickness in an extrapolated area. The second and third studies were designed to investigate this possibility. Initially, a large number of extreme value models were generated from simulated inspection data. These models were used to calculate the return levels and confidence intervals for various extrapolation ratios. The average width of the confidence intervals, expressed as a percentage of the return level, was plotted as a function of the extrapolation ratio.

The width of the confidence interval increased with the extrapolation ratio which reflected that extrapolations to larger areas are associated with more uncertainty. There is a higher chance that predictions will be incorrect. It also appears from this data that a larger sample of minima leads

to narrower confidence intervals. If the return level confidence intervals do contain the smallest thickness across the extrapolated area, this would suggest that smaller block sizes (more minima) would offer narrower bands for the minimum thickness in the extrapolated area.

This was shown to be misleading. The larger samples, extracted from inspection data using a smaller block size, have a larger mean than samples extracted with the larger block sizes. Consequently, models constructed using this data will tend to overestimate the minimum thickness, this bias will propagate through the model during an extrapolation. This can be seen both in figure 4.7, where the fit of the data to the model constructed using a smaller block size (the 100 minima model) was much worse than those constructed with larger block sizes (the 16 and 25 minima models). This effect can be seen again in the third study, which compared the average confidence interval generated using models from inspections of very large Gaussian and exponential surfaces to the smallest thickness across those surfaces.

In section 4.4, it was shown that the minimum thickness will lie in the range of return level confidence intervals on average. The effectiveness of this method for extrapolation was shown to strongly depend on the construction of the model and the area to which the inspector is extrapolating. In general, the use of a larger block size (smaller sample of minima) will provide an extreme model which performs more effectively when extrapolating to larger areas. From the data presented in this section, it can be suggested that a practical rule of thumb for field use would be to select a block size which produces a sample of at most 25 minima. This model could be used to extrapolate to an area with an extrapolation ratio of (at most) 50. The expected width of the confidence intervals is around 10% of the return level i.e. an inspector could be confident that the minimum thickness would lie within 10% of the predicted return level.

An inspector can use the data provided in this chapter to look up the error range that can be expected on the minimum thickness value for the extrapolation ratio and block size they want to use. This provides an inspector with insights into the amount of area that requires inspection to achieve an expected level of accuracy on their prediction of minimum thickness. This is a formidable insight when planning or applying EVA to real inspection cases.

In this chapter, the use of extreme value models for extrapolation has been investigated. The

data used in the chapter has been generated using a numerical algorithm. Although this algorithm produces realistic results, it is missing some key features that are found in thickness data collected using ultrasonic measurement techniques. For example, the data in this chapter does not take into account any errors or uncertainty that could arise from the scattering of the ultrasonic pulse by a rough surface. In chapter 5, a method for simulating a large number of ultrasonic thickness measurements is described. This method will be used in chapter 6 to conduct a joint computational and experimental study of the effect of surface roughness on the accuracy of ultrasonic thickness measurements.

Chapter 5

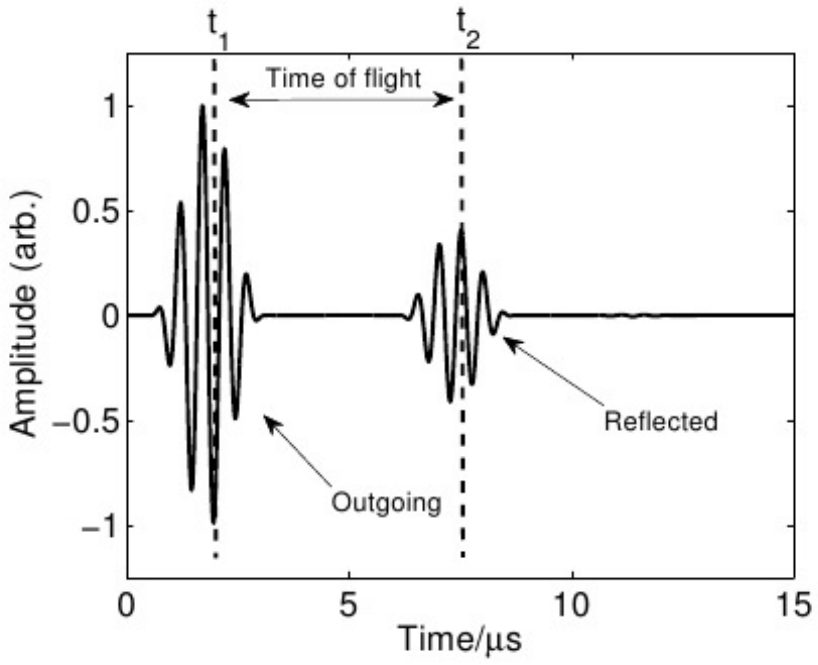
Distributed Point Source Method

5.1 Introduction

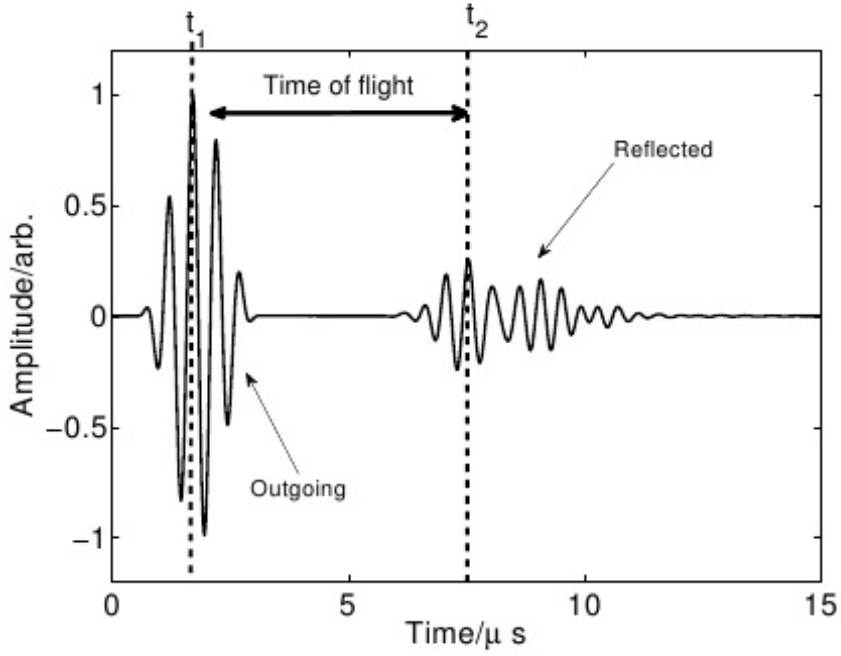
The majority of data that is used for partial coverage inspection (PCI) is collected using well-established non-destructive testing techniques. A brief overview of these techniques can be found in Blitz [8], with ultrasonic thickness measurement being the most commonly used. Ultrasonic tests are often performed as a C-scan. C-scans are used to construct area colour maps of the measured thickness. At every point across the inspection area, a thickness measurement is taken (an example of a scanning pattern is shown in figure 1.1). An image of the measured thickness distribution is formed by representing each measurement with a coloured patch, the colour or gray scale chosen to be representative of the measured thickness (figure 1.1C).

Ultrasonically measured thickness measurements are a combination of the interaction of the ultrasonic pulse with the corroded surface and any noise introduced by the signal processing algorithm and the signal acquisition system. Figure 5.1 shows examples of ultrasonic signals collected from a flat surface (figure 5.1(a)) and a rough surface (figure 5.1(b)). Surface roughness introduces distortion to the signal in figure 5.1(b). This distortion makes it difficult to extract a measurement of thickness from the signal.

Jarvis and Cegla investigated the stability of three commonly used timing algorithms used to extract wall thickness measurements from signals collected using a permanently installed shear wave monitoring system [14]. It was found that the wall thickness estimate changes significantly



(a)



(b)

Figure 5.1: Examples of signals scattered from a flat surface ((a)) and a rough surface ((b)), showing the sent signal (labelled "outgoing" in figures), the first backwall reflection (labelled "reflected" in figures) and the time of flight (calculated using a peak to peak algorithm, described later in this chapter) that is used to evaluate thickness. Reproduced from Benstock *et al.* [1].

for different instances of a rough surface (with the same statistical description), if a different algorithm is used. Different timing algorithms will lead to different thickness measurements. This uncertainty is problematic for PCI as it leads to uncertainty in the extrapolation.

To investigate this effect in relation to PCI, equivalent simulated and experimental ultrasonic thickness C-scans were performed. This study was published in the Journal of the Acoustical Society of America [1]. Much of the material presented in the next two chapters (chapters 5 and 6) has been reproduced from this paper. In this chapter, the model used to simulate the ultrasonics is described.

Although the DPSM is a promising method for simulating ultrasonics, it has been shown that there are discrepancies between analytical solutions and the pressure fields calculated by the DPSM. This has been well-documented and is attributed to poor matching to the uniform pressure boundary condition, due to the limited number of point sources [72, 73, 74, 75]. Alternative formulations of the DPSM have been developed with the goal of improving the matching to the uniform pressure boundary condition [76, 77]. These have shown some success, however, in this chapter it will be shown that with a slight alteration to the DPSM to better match the boundary conditions these discrepancies can be removed.

The chapter begins by describing the distributed point source method (DPSM), alongside a brief comparison to alternative techniques and a review of the literature. It continues with a demonstration of the problems with the current formulation of the DPSM via a comparison to an analytical solution. Limitations to the DPSM are discussed and addressed with the development of an extension to the DPSM which ensures better matching to the boundary conditions of the model (section 5.4.3).

In this chapter a new extension to the DPSM method is formulated. This extension removes the limitations of earlier incarnations of the DPSM and allows for fast, accurate simulations of ultrasonic thickness measurements. The extension of the DPSM described in this chapter is used in chapter 6 to simulate large numbers of ultrasonic thickness measurements

5.2 Modelling reflected ultrasonic signals

Finite element methods (FEM) are the most widely used method for simulating ultrasonics. They have been used in applications as wide ranging as modelling elastic wave propagation in waveguides [78] to simulating the propagation of ultrasonic waves in dental implants [79]. FEMs have also seen some application in modelling rough defects. For example, Pettit modelled a pulse-echo inspection of planar rough defects, using a bespoke elastic finite element method [80].

Inspections of engineering components often consist of large numbers of ultrasonic measurements. While FEMs are a sufficient tool for small numbers of simulations of measurements from rough surfaces, they are too computationally expensive for modelling the number of signals[1] required to simulate a full inspection. Jarvis and Cegla performed a comparative study of the computational load required by various methods to simulate the reflection of an ultrasonic pulse from a rough surface [14]. They found that the time taken for the simulation using the FEM was 236 seconds compared to 13 seconds for the distributed point source method (DPSM).

A large contribution to the difference in computational load arises from inclusion of the shear stress components in FEM models. C-scans are usually performed using a zero-degree longitudinal wave transducer. It can be assumed that mode conversion has little effect on the received signal i.e. the acoustic component of the incident and reflected fields can be used to model the physics of the system. As the systems modelled in this chapter deal exclusively with zero-degree longitudinal transducers it is prudent to model only the acoustic components.

This could be achieved by changing the FEM to use only the acoustic components. However, the FEM requires full discretization of the volume of interest. This is expensive from both the perspective of the memory required to store the mesh and the time required to set it up. Alternative formulations of the FEM, such as the boundary element method (BEM) can reduce the computational cost by only discretizing the boundaries of the volume of interest. However, for problems with large surface to volume ratios, such as those that will be considered in this chapter, this efficiency can be lost due to the need for absorbing boundaries.

As an alternative to the family of finite element methods, the distributed point source method

[1] For a scan of $200mm^2$ of a component with a separation between measurements of $1mm$, the number of thickness measurements taken is $40,000$.

(DPSM), which models acoustic fields, can be used to reduce the computational cost of ultrasonic modelling. A number of different authors have shown that DPSM outperforms finite element methods in terms of the time required for a simulation [75, 14]. Additionally, DPSM has been shown to be a very promising tool for modelling the reflection of acoustic waves from rough defects by Jarvis and Cegla [14].

The DPSM is a mesh-free semi-analytical technique, first used to study the magnetic fields of complex shaped magnets [81]. The method has been extended to acoustic field calculations [82] and is being applied to an increasing number of problems. Banerjee *et al.* have used DPSM to study elastic wave propagation in sinusoidally corrugated waveguides [83] . It has also been used to study a variety of scattering problems [82, 72, 84] and to predict the beam from phased array transducers [85]. More recently it has been used to simulate the scattering from a rough surface [14] in 2D. A brief outline of the method is given here. Readers interested in more detail on the DPSM are referred to Placko *et al.* [86].

5.3 The distributed point source method (DPSM)

5.3.1 Frequency domain calculations

The distributed point source method consists of the placement of point sources across the transducer face and at the rough surface. Amplitudes of the point sources are calculated directly from the boundary conditions of the system. Point sources (*S2*) placed where the field strength is known (*S1*) are termed active point sources, with passive point sources (*I2*) placed where reflections or scattering occurs (*I1*).

The amplitudes of the active point sources, labelled $i = 1..N$, for N sources are labelled A_{Si}. The pressure from the i^{th} point source is given by:

$$p(r) = \frac{A_{Si}e^{jkr}}{r} \tag{5.1}$$

where r is the distance of a target point from the source, k is the wave number and j is the

imaginary unit. The point sources (S2 and I2) are offset from the boundary (S1 and I1) by a small distance (r_s), to avoid singularities.

A suitable offset from the transducer face can be calculated analytically. Equating the hemispherical area of a sphere of radius r_s to the area associated with each source point a^2 as in Placko *et al.* [86], gives:

$$a^2 = 2\pi r_s^2 \rightarrow r_s = \frac{a}{\sqrt{2\pi}} \tag{5.2}$$

where a is the point source separation. Equation 5.2 provides a guide of what value to choose for r_s, providing consistently good results [86].

Equation 5.1 can be written in matrix form for a set of source points S and a set of target points T:

$$\mathbf{P_T} = \mathbf{Q_{TS}}\mathbf{A_S} \tag{5.3}$$

where the propagation matrix, $\mathbf{Q_{TS}}$:

$$Q_{TS} = \begin{pmatrix} \frac{e^{ikr_{11}}}{r_{11}} & \frac{e^{ikr_{12}}}{r_{12}} & \frac{e^{ikr_{13}}}{r_{13}} & \cdots & \frac{e^{ikr_{1M}}}{r_{1M}} \\ \frac{e^{ikr_{21}}}{r_{21}} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{e^{ikr_{N1}}}{r_{N1}} \cdots & \cdots & \cdots & \cdots & \frac{e^{ikr_{NM}}}{r_{NM}} \end{pmatrix} \tag{5.4}$$

and $\mathbf{P_T}$ is the vector of pressures at the set of target points T and $\mathbf{A_S}$ is the vector of source amplitudes.

If the set of target points is across the transducer face, the pressure at each point is known. The propagation matrix can be calculated from the positions of the point sources and the target points. This leaves one vector unknown, the source amplitudes, which can be found by inverting equation 5.3:

$$\mathbf{A_{S2}} = \mathbf{Q^{-1}_{S2S1}}\mathbf{P_{S1}} \tag{5.5}$$

Then, the pressure at any set of target points can be calculated by applying the relevant propagation matrix to $\mathbf{A_{S2}}$.

If the field from the transducer is calculated at points along the rough surface, the amplitudes of the passive point sources can be found:

$$\mathbf{A_{I2}} = -\mathbf{Q^{-1}_{I2I1}}\mathbf{Q_{I1S2}}\mathbf{A_{S2}} \tag{5.6}$$

The reflected field at any point can then be calculated by propagating $\mathbf{A_{I2}}$ to the points required.

5.3.2 Time Domain

Common practice in wall-thickness measurements is to use time-domain pulses. The calculations outlined so far are performed in the frequency domain and require extension for time domain pulse propagation. The method used here follows Jarvis [14]. Alternatively, the time-domain Green's function could be used for the calculation, although this approach is unsuited to long pulses [84].

The frequency-domain representation of the outgoing pulse, f(t), is calculated at *S1*:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) exp(j\omega t) d\omega \tag{5.7}$$

where F(ω) is the frequency domain of the incident pulse, ω is the angular frequency and t is time. A monochromatic DPSM calculation is then performed for each frequency component:

$$\mathbf{P_{S1}} = \mathbf{Q_{S1I2}}\mathbf{A_{I1}} \tag{5.8}$$

A single reflection is modelled and the reflected frequency domain at each point in S1 is constructed. The arithmetic mean of all of these frequency domain representations, gives the frequency domain representation of the received signal. The time-domain received signal is calculated by performing an inverse Fourier transform:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) exp(-j\omega t) dt \tag{5.9}$$

The outgoing pulse used was a 5MHz Hanning windowed toneburst, with 5 cycles. To reduce the computational time taken for these simulations, any frequency component with amplitude less

than 1% of the largest component was set to zero. These components were found to have little effect on the overall signal.

5.4 Validation of the DPSM code by comparison to an analytical solution

There is no commercially available software which can perform DPSM calculations. Consequently, a program was written in C++, to perform the calculations efficiently on Imperial College's high performance computing cluster [87]. The program uses the Eigen libraries [88], to handle the matrix inversion and multiplication required for DPSM. The calculations are performed in parallel, using OpenMP [89] for the field calculations and MPI [90] for the signal calculations, which handle communication between the processors. Fourier transforms are performed using FFTW [91]. There is additional functionality to perform calculations in two dimensions. These require Bessel functions, which are included using the Boost libraries [92]. To ensure the validity of the program that was coded, its solutions were compared to an analytical benchmark and a convergence study was performed.

5.4.1 Analytical Solution for the on-axis field of the resilient disc

The transducer was modelled by assuming a constant pressure across its active element (a disc). There is an analytical solution for the on-axis pressure [93] of this type of transducer. The pressure, p, is given as a function of distance, y, as:

$$p(y) = e^{-jky} - \frac{y}{\sqrt{y^2 + a^2}} e^{-jk\sqrt{y^2+a^2}} \tag{5.10}$$

where k is the wave-number and a is the radius of the disc. The parameters used for these simulations are for a 6mm diameter transducer, operating at 5MHz in steel ($v_L = 5960m/s$), unless otherwise stated.

To derive equation 5.10, two assumptions are made:

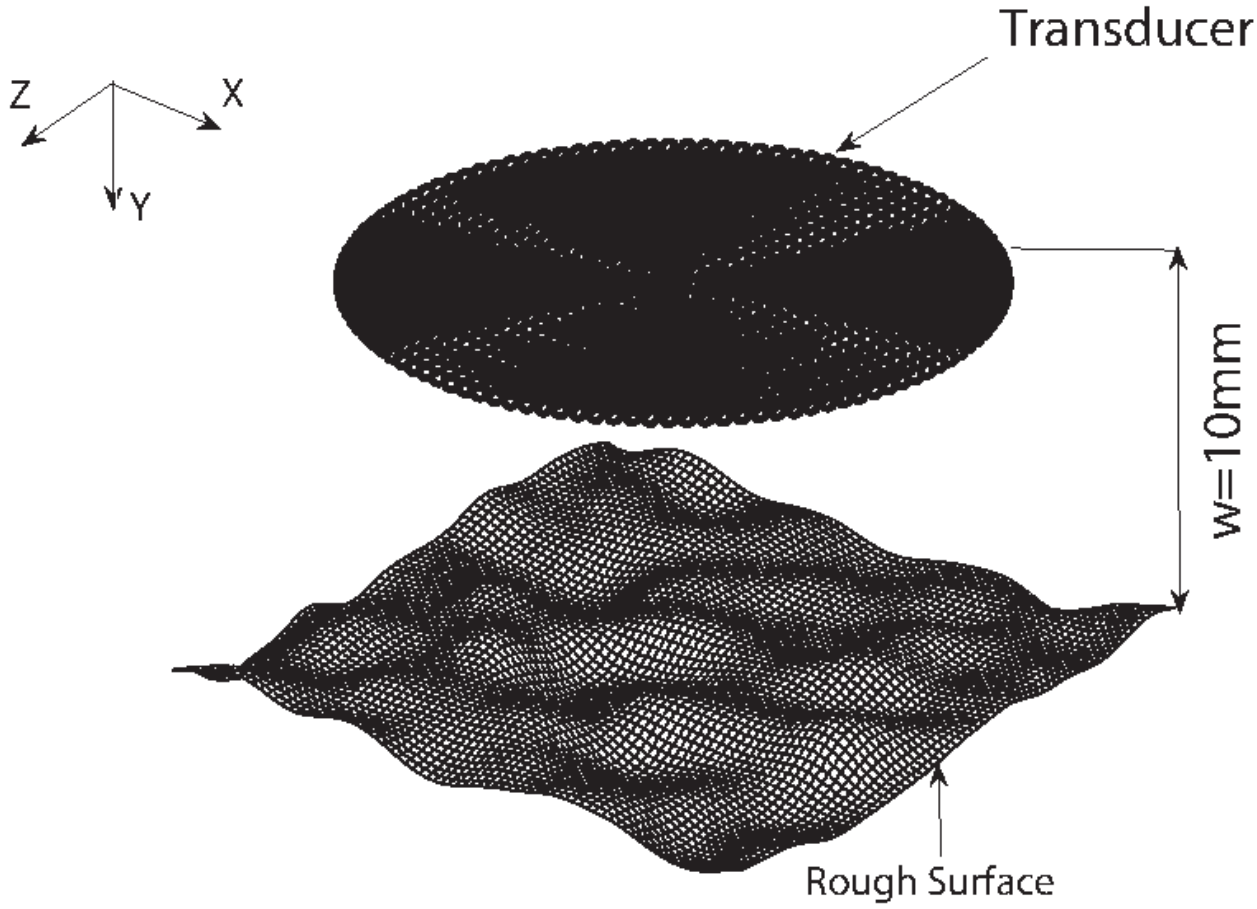


Figure 5.2: Schematic showing the simulation cell for a single thickness measurement. The 6mm 5MHz longitudinal wave transducer and the rough surface are represented by point sources separated by 0.1mm, $w = 10mm$ is the mean thickness of the material. The size of the rough surface patch required for the simulation is determined by the mean thickness of the component. For a $10mm$ mean thickness the size of the patch required was $9mm$ by $9mm$. Reproduced from Benstock *et al.* [1].

1. *Uniform pressure boundary condition*: the pressure on the disc is
   $p(x, y, z) = 1$.

2. *Zero pressure boundary condition*: everywhere on the plane *(x,0,z)*, $p(x, 0, z) = 0$, where $x^2 + z^2 > a^2$.

The resilient disc is an idealised source. In real life, transducers would emit different beams as the vibration across the surface (and hence the pressure) might not be uniform. However, a uniform pressure distribution is a commonly used assumption employed to model 0-degree longitudinal probes. Additionally, the analytical solution provides a readily available standard to verify the results of the DPSM and to check for convergence.

Equation 5.10 can be used to calculate a reflected analytical signal in a similar manner to the time domain signal calculation outlined for DPSM (section 5.3). The Fourier transform of the outgoing signal is taken (equation 5.7) and equation 5.10 is used to calculate the the pressure for each frequency component (above 1% of the maximum amplitude). These pressures are then used to form the frequency spectrum of the reflected signal, which is then subjected to an inverse Fourier transform (equation 5.9) to give the time-domain representation of the reflected signal.

5.4.2 Direct comparison to the analytical solution

The geometrical set-up for these calculations is the same as in figure 5.2[2]. The active source points (the transducer) are placed on the plane *Y=0*. The DPSM is used to calculate the pressure radiating into an acoustic medium with the same density and compressional wave velocity as steel (i.e. $v_l = 5960m/s \Rightarrow k = 5.27 \times 10^3 radsm^{-1}$), from the active point sources.

Figure 5.3 shows the on-axis pressure as calculated by DPSM compared to the analytical solution. This figure clearly demonstrates the discrepancies between the analytical solution and the field calculated by the DPSM. This difference has been attributed to poor matching of the uniform pressure boundary condition to the pressure distribution provided by the point sources [75]. If the uniformity of the boundary conditions was the cause, then simply increasing the number

[2]with a flat surface of the same dimensions in place of the rough surface

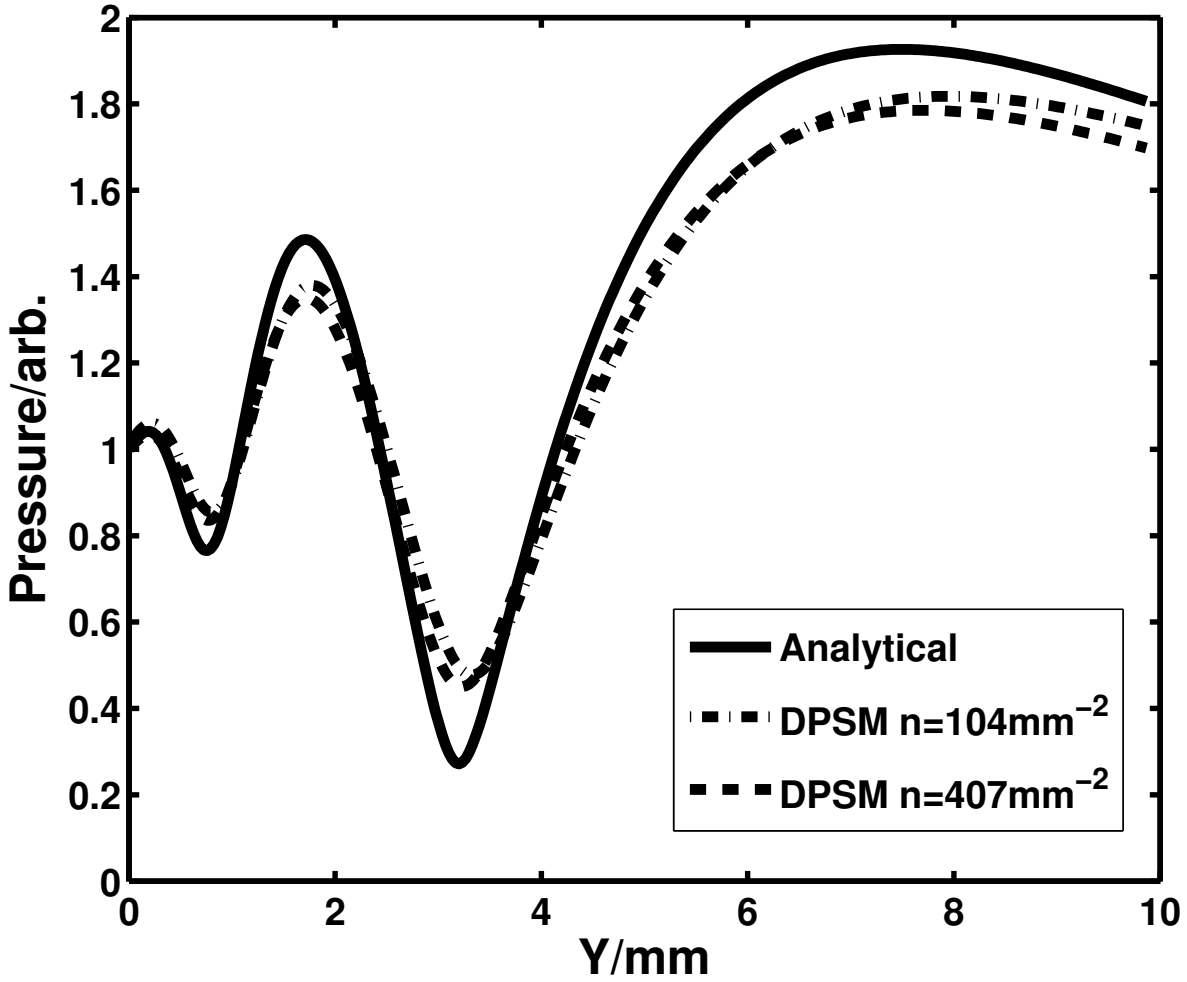


Figure 5.3: Comparison between the analytical solution and DPSM solution with different point source densities for the on-axis field (f=5MHz), for a disc of radius 3mm into a medium with c=5960m/s. Reproduced from Benstock *et al.* [1].

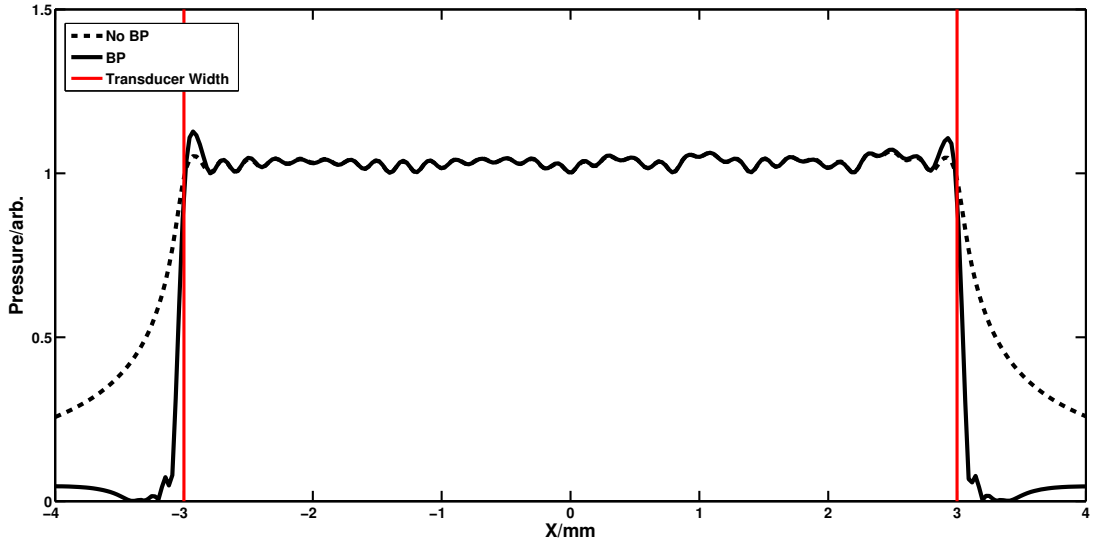


Figure 5.4: A comparison of the pressure distributions across the surface of the transducer, calculated on the line (Z=0, Y=0,X). The red line gives the diameter of the transducer. The pressure distribution with no boundary source points (black dashed line) shows poor matching to the zero pressure boundary condition. The addition of boundary points (BP), leads to much better matching to the boundary conditions used in the analytical solution, which rises to 1 at x=-3, is then constant up to x=+3 where it returns to zero. Reproduced from Benstock *et al.* [1].

of point sources should improve the results. However, this is not the case. Figure 5.3 shows that with increasing point source density, the discrepancy remains.

The DPSM does not explicitly satisfy the zero-pressure boundary condition used by the analytical solution. For the analytical solution a cross section of the pressure distribution across its surface, would show uniform pressure (equal to unity) across the transducer. At its edges, the pressure drops immediately to zero. In comparison, the cross-section of the pressure distribution calculated by DPSM, tapers off slowly (figure 5.4), leading to a 'halo' around the outside of the transducer (where the zero-pressure boundary condition is not fulfilled).

5.4.3 Boundary points

The zero-pressure boundary condition can be matched by the addition of a number of *boundary point sources* around the outside of the transducer face (figure 5.4). Figure 5.5 is an example of a discretised 6mm diameter transducer. The boundary points are shown by the red points, which are positioned around the outside of the transducer. At these points the pressure is set to zero. The pressure at the blue points is set to one. The addition of the boundary points requires an extension to the DPSM.

To include boundary point sources, one alters the boundary condition vector and the propagation matrix in equation 5.5. For M active source points and K boundary points:

$$Q_{TS'} = \begin{pmatrix} \frac{e^{ikr_{11}}}{r_{11}} & \frac{e^{ikr_{12}}}{r_{12}} & \frac{e^{ikr_{13}}}{r_{13}} & \cdots & \frac{e^{ikr_{1,M+K}}}{r_{1,M+K}} \\ \frac{e^{ikr_{21}}}{r_{21}} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{e^{ikr_{N1}}}{r_{N1}} \cdots & \cdots & \cdots & \cdots & \frac{e^{ikr_{N,M+K}}}{r_{N,M+K}} \end{pmatrix} \tag{5.11}$$

where S' is the set of source points and boundary points. The boundary condition vector is given by:

$$P = \begin{pmatrix} \mathbb{I}_{M\times 1} \\ 0_{K\times 1} \end{pmatrix} \tag{5.12}$$

where $0_{K\times 1}$ is a K dimension vector of zeros. The source amplitudes can then be calculated using

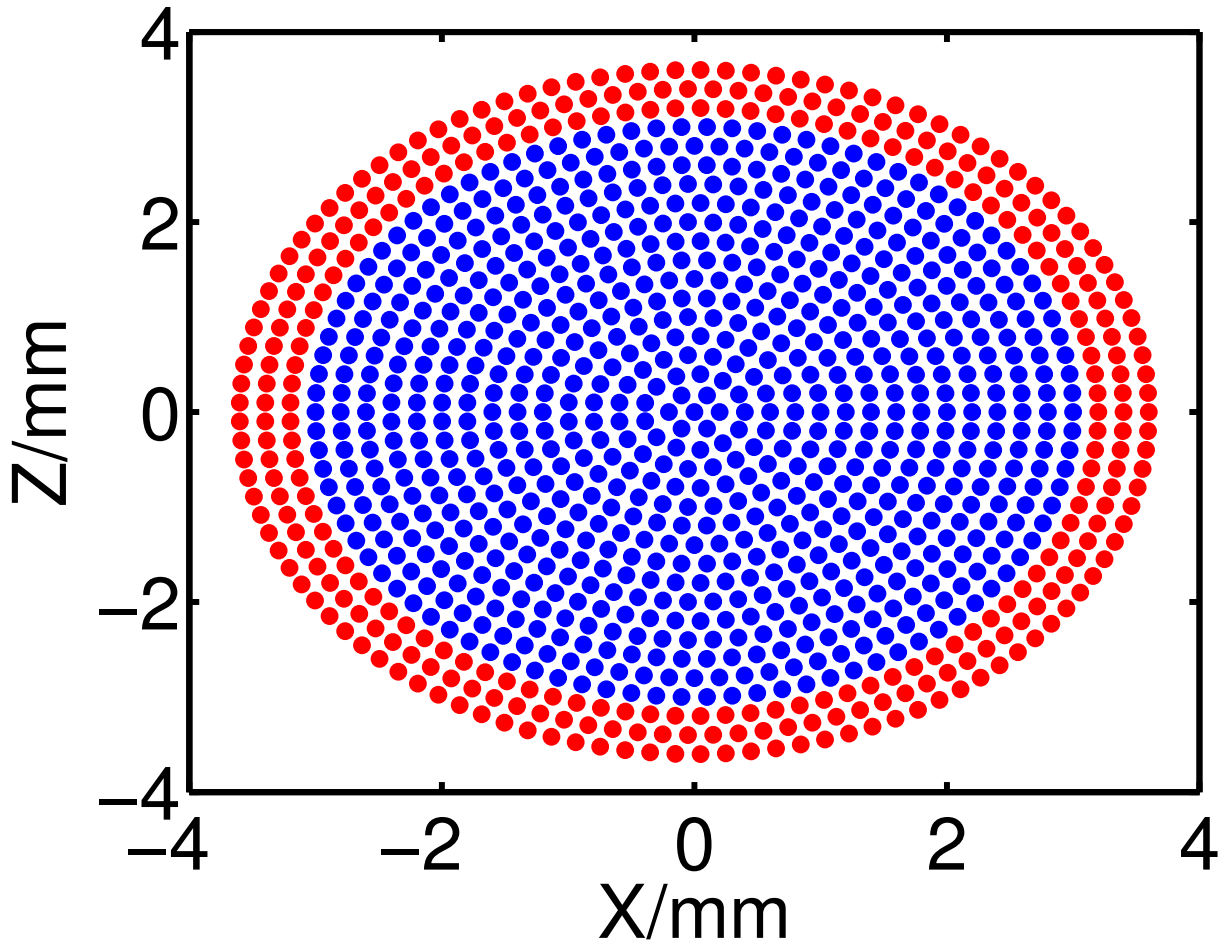


Figure 5.5: An example of a discretised 6mm diameters circular transducer. The pressure at the blue points was set to one; the pressure at the red boundary points was set to zero. Reproduced from Benstock *et al.* [1].

equation 5.5, with the altered matrix and vector. Boundary points were found to reduce the pressure around the edge of the transducer to close to zero (figure 5.4, solid black line), with the addition of four layers of boundary points (with a point separation of $\frac{\lambda}{12}$).

While the pressure conditions are much more closely matched, there are still some differences in the pressure profile produced by the DPSM and the profile used for the analytical solution. This is a consequence of trying to model a continuous boundary condition by a number of point sources. The pressure profile radius is slightly larger than the prescribed transducer radius (figure 5.4), which would lead to phase differences in the near field and amplitude discrepancies in the far-field.

An effective radius for this pressure distribution can be calculated. The difference between the DPSM and analytical solutions across the axis is minimized with respect to the radius. This leads to close agreement between the analytical and the DPSM solutions (figure 5.6).

Equation 5.10 was also used to calculate the pulse propagated to a single point at $(0, \frac{50}{3}\lambda, 0)$ and compared to the equivalent DPSM signal. The frequency domain representation of the outgoing signal was calculated using equation 5.7. At each frequency the amplitude at $(0, \frac{50}{3}\lambda, 0)$ was calculated and used to form the frequency domain representation of the propagated pulse. This representation was then transformed into the time domain using equation 5.9, which gives the received signal at $(0, \frac{50}{3}\lambda, 0)$. Close matching between the DPSM (using an active point source separation of $\frac{\lambda}{12}$) and analytical signals was achieved (figure 5.7), with a 2% maximum difference of the Hilbert envelopes of the signals. This analytical signals forms the benchmark for the convergence studies.

5.4.4 Passive point source density

The next part of the validation which needs to be addressed is the passive point source density across the backwall. An analytical solution for a reflected signal from a flat backwall can be calculated by reversing the phase of the analytical signal (section 5.4.3). This models the total distance travelled by the pulse and its reflection from a flat backwall $\frac{25}{3}\lambda$ away from the transducer.

The DPSM was used to calculate the reflected signals from four backwalls with different point

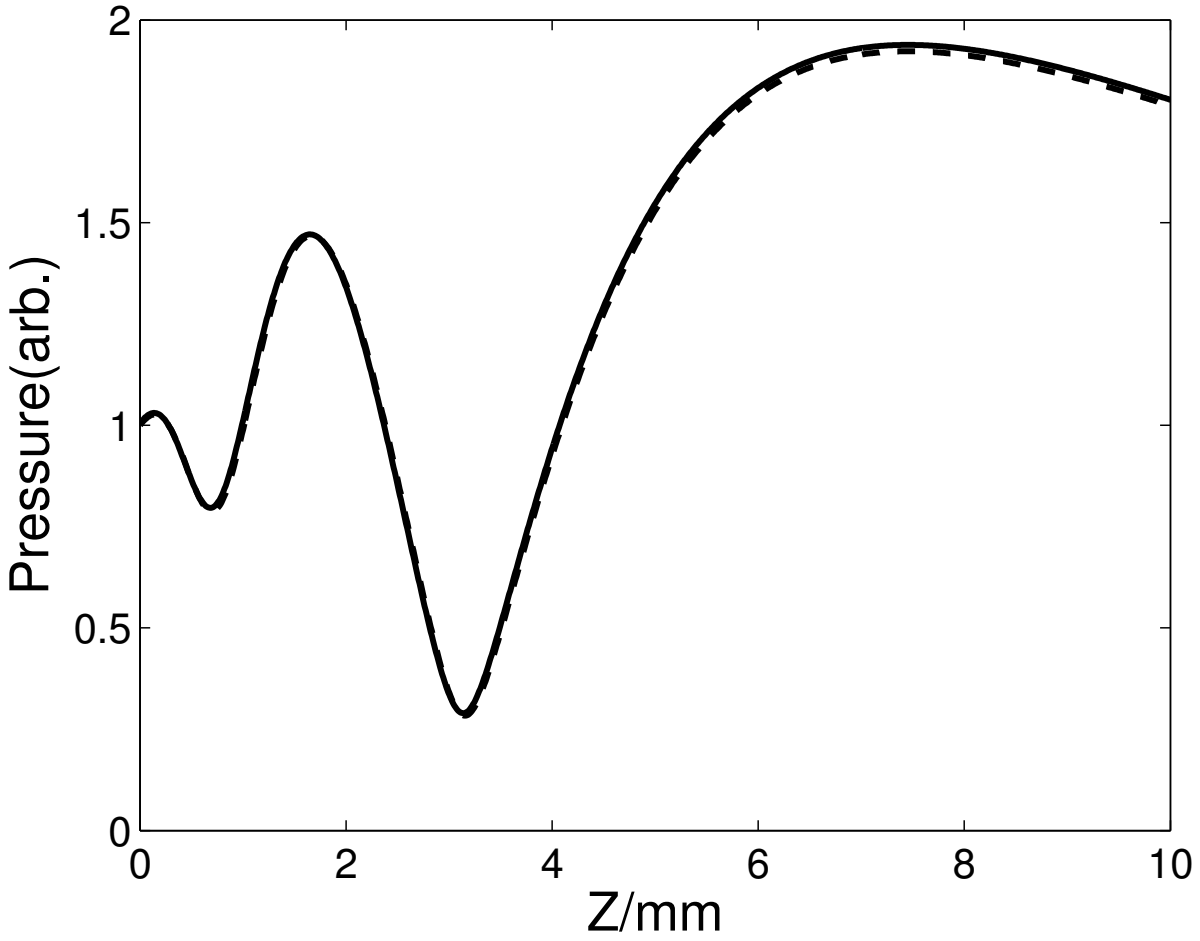


Figure 5.6: Comparison between analytical solution (solid black line) with radius 3.05mm and DPSM (dashed black line) with boundary points (f=5MHz). The DPSM calculation was performed with an active point source separation of $\frac{\lambda}{12}$ $(0.1mm)$. The frequency was chosen as it will be used as the centre frequency of the Hann window pulse used in the simulations in chapter 6. The effective bandwidth of a 5MHz Hann window pulse is 3MHz. Reproduced from Benstock *et al.* [1].

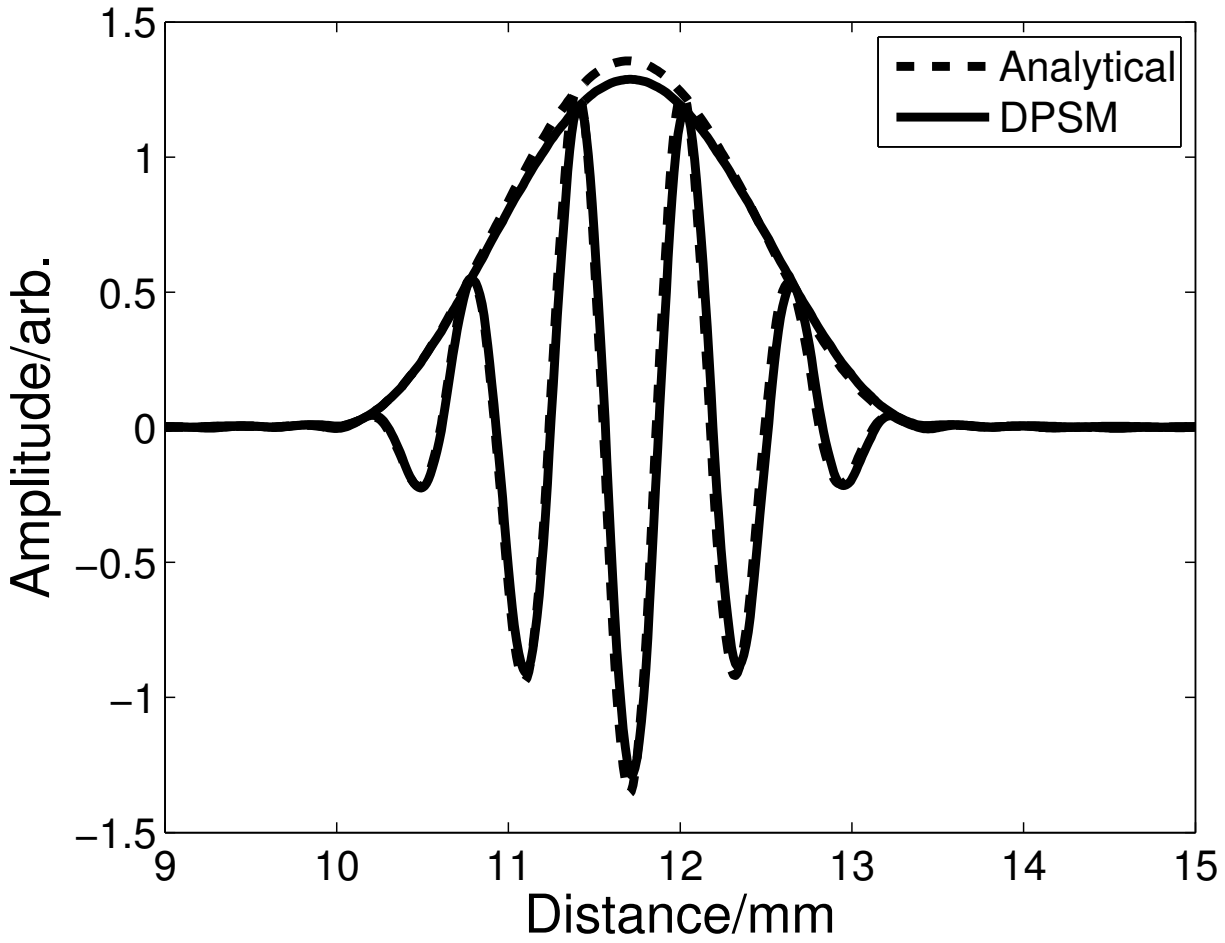


Figure 5.7: Comparison between the analytical outgoing signal at $\frac{50\lambda}{3}$ and the DPSM calculated signal at $\frac{50\lambda}{3}$ (and their Hilbert envelopes). It is believed that the major source of error is the slightly different effective radius of the disc as a result of the discretisation. Reproduced from Benstock *et al.* [1].

source densities. The maximum error on each signal was found by calculating the percentage difference between the maximum of the DPSM solution and the analytical solution. The passive point source density when the error is reduced to 1% or less, was used for the calculations. This was found to be $147\lambda^{-2}$, corresponding to a passive point source separation of approximately $\frac{\lambda}{12}$ (figure 5.8).

5.4.5 Backwall size

The active and passive point source densities found in sections 5.4.3 and 5.4.4 were then used to determine the minimum patch size required. The reflected signal from square patches ranging from $5 - 11mm$ ($\frac{25\lambda}{6}$ to $\frac{45\lambda}{6}$) in dimension were calculated using the DPSM and compared to the analytical signal. The maximum error was calculated in the same way as in section 5.4.4 and reduced to less than 2% with a patch size of $9mm$ (figure 5.9), so this was chosen as the patch size for the C-scan model.

5.4.6 Roughness discretization

The previous sections have found the parameters for which the DPSM simulation of a signal reflected from a flat backwall has converged. However, this study is concerned with rough surfaces. Therefore, the parameters need to be checked for a surface of varying height.

A surface with a sinusoidal height variation was used as a model surface. Its amplitude and wavelength were chosen to be of a similar extent to RMS and correlation lengths which were to be modelled. The reflected signals from this surface with various point source separations were calculated using DPSM.

For this case there is no available analytical solution for the reflected signals. Therefore, the maximum error was found by calculating the percentage difference from the signal calculated with the largest passive point source density. The calculation was taken to be converged when the reduction in error was less than 1% upon doubling the number of point sources per unit area. This was achieved for a point source density of $147\lambda^{-2}$ $[100mm^{-2}]$ (circles in figure 5.8).

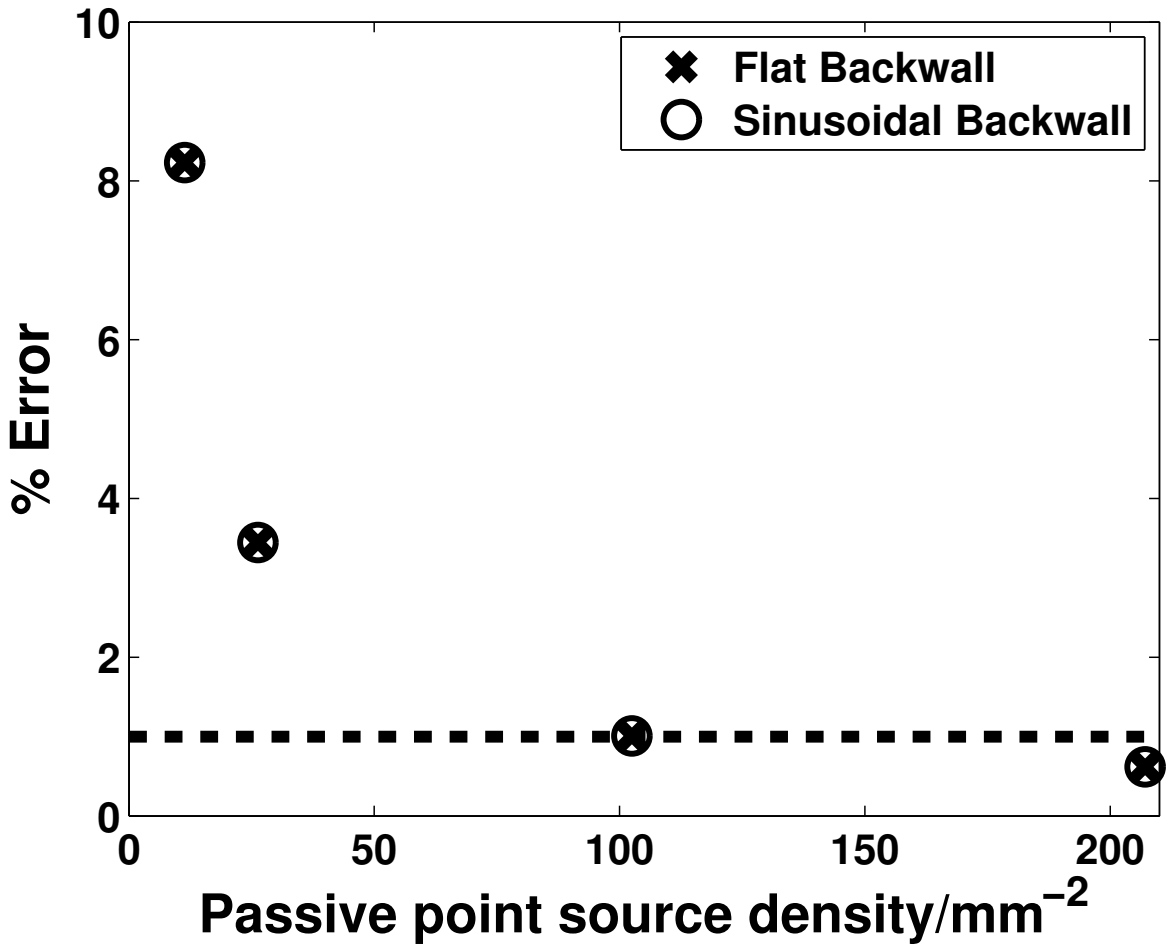


Figure 5.8: The variation of the difference between the maximum of the Hilbert envelope of the analytical signal and the maximum of the Hilbert envelope of the DPSM calculated signal as a function of increasing point source density. A point source density of $147\lambda^{-2}(103mm^{-2})$ was chosen as this was the first point where the error reduced to 1% of the maximum of the Hilbert envelope of the analytical signal. The crosses show data from a flat backwall reflection and the circles from sinusoidal surfaces. Reproduced from Benstock *et al.* [1].

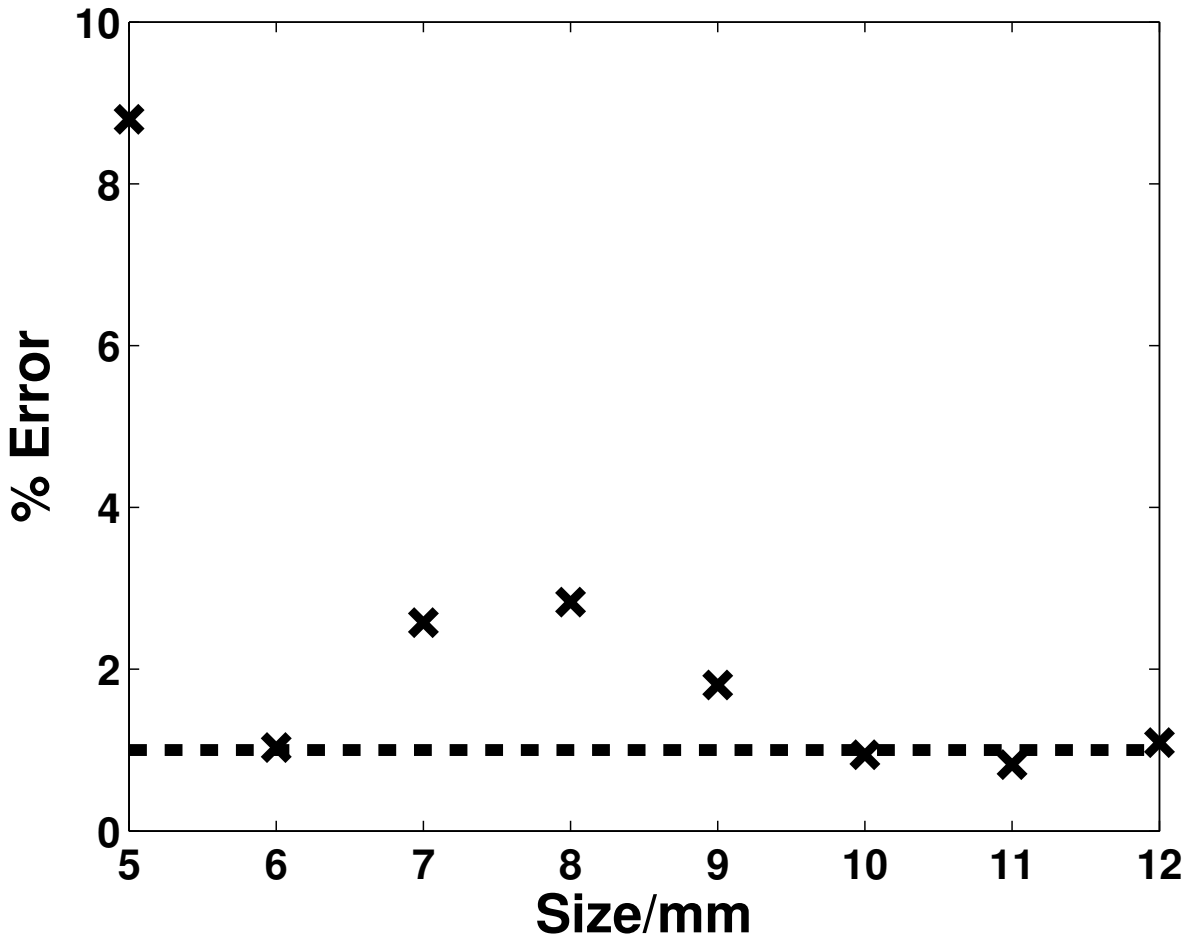


Figure 5.9: The variation of the error, calculated as the percentage difference between the maximum of the Hilbert envelope of the DPSM and analytical signals, with square patch size. A patch size of $9mm^2$ was chosen. Reproduced from Benstock *et al.* [1].

5.5 Summary

The distributed point source method is a fast alternative to more common finite element methods for simulating ultrasonic thickness measurements. It has been shown that, for the simulation of acoustic waves, the DPSM is faster than the finite element method by a factor of 10 [14]. In this thesis it has been used to simulate a large number ultrasonic thickness measurements across surfaces representative of damage that could occur in a corroded component. Prior to these simulations, the DPSM was investigated as part of the development of a C++ program. It was found that, although the DPSM offers a promising new approach to simulating ultrasonic thickness measurements, results produced by the method differed from equivalent analytical results.

Subsequently, in this chapter, the limitations of the DPSM were investigated and it was found that the discrepancies between the DPSM and analytical benchmarks arose from differences in their assumptions. The DPSM does not automatically fulfil the zero pressure assumption made by the benchmark. This was addressed by the addition of a number of boundary point sources around the outside of the transducer. At these points, the pressure is set to zero. This led to a significant improvement in the matching of the results of the DPSM to an analytical benchmark.

A convergence study of the parameters of the DPSM model were performed. The point source separation, size of the rough surface patch and discretization of the surface roughness were investigated in order to find the best parameters for simulating an ultrasonic C-scan. It was found that convergence of the DPSM results was achieved using a $9mm$ by $9mm$ square patch of rough surface with active and passive point source separations of $0.1mm$. The simulation parameters selected in this chapter, along with the new extensions to the DPSM, were used to simulate a large number of ultrasonic thickness measurements across rough surfaces. The results of these simulations are presented in the next chapter (chapter 6).

Chapter 6

Accuracy of ultrasonic thickness C-scans

6.1 Modelling a C-scan

Ultrasonically measured thickness measurements are a combination of the interaction of the ultrasonic pulse with the corroded surface and any noise introduced by the signal processing algorithm and the signal acquisition system [14, 94, 95]. The interaction between the signal processing algorithm and the distorted signal produced by the rough surface leads to inaccurate thickness measurements. This is one of many factors which can contribute towards thickness measurement uncertainty. Most of these can be controlled through careful inspection set-up and procedures. However, the uncertainty introduced by surface roughness cannot be controlled as it is a feature of the measured system. This effect need to be quantified in order to understand its effect on condition assessment as any inaccuracy in the inspection data will propagate through an extrapolation model and could lead to incorrect conclusions.

Jarvis and Cegla investigated the effect of surface roughness and the choice of timing algorithm on the accuracy of the thickness measurements [14]. They modelled the behaviour of reflections of ultrasonic pulses, transmitted from a rectangular transducer, from rough surfaces. A two dimensional distributed point source method was used for the simulations. It was found that increases in the surface roughness increased the median measured thickness i.e. led to an overestimation of the actual thickness. The amount of the overestimation varied with the choice of the timing algorithm.

In this chapter, a series of simulations and equivalent experimental scans are performed using the modified three dimensional distributed point source method described in the previous chapter (chapter 5). In contrast to the work by Jarvis and Cegla [14], the simulated ultrasonic thickness measurements are taken as a C-scan, as an inspector would collect in the field with a circular transducer. An analysis of the results of these simulations and experiments more insight into the effect of surface roughness and choice of timing algorithm on a set of thickness measurements. The experimental data is used to produce an extreme value model, which is used to provide an example of these effects on an extrapolation. Prior to this thesis, there has been no study of the effect of surface roughness on ultrasonic thickness measurements taken as part of a C-scan, nor has any study of the effect of these errors on an extreme value model been performed.

6.2 Simulation Set-up

The corrosion mapping set-up used in this chapter consists of a 6mm diameter 5MHz longitudinal wave transducer coupled to the surface using water (or another suitable couplant). There are two alternatives to couple the transducer to the part: directly placing the transducer on the surface of the part with a small amount of couplant (a contact scan) or placing the part in a water bath. In industry most scans are performed as contact scans as it is infeasible to use immersion scans for in service pipework. The simulations in this chapter model a contact scan as it is less computationally expensive to model than an immersion scan. In contrast, the experiments use an immersion scanning set-up to ensure consistent coupling between the transducer and the plate across the surface of the part.

There are differences between contact and immersion scans which require discussion here. The transducer is much further away from the rough surface for an immersion scan and the transmitted ultrasonic pulse has to travel through a much larger region of couplant (the water). Subsequently, both the ultrasonic footprint on the rough surface will be larger than the contact measurements, increasing the amount of scattering from the rough surface, and the received signal amplitude will be smaller than the contact scan, due to the increased distance both the transmitted and reflected pulse must travel. To reduce the impact of these differences the transducer in the immersion scans

was positioned as close to the surface of the steel plate as possible. Subsequently the effect of these differences should be minimized.

For the simulations, the front surface is assumed to be in good condition, so that the transducer makes a flat contact (figure 5.2) with the component. To make a thickness measurement, a 5MHz Hanning windowed pulse is then radiated into the steel wall, which is reflected from the internal, corroded, rough surface and recorded (figure 5.1). The thickness is calculated from the time of flight of this pulse using one of the algorithms described in 6.2.2. In real situations, multiple reflected pulses will be received. However, the signal processing is restricted to the first reflection for the purposes of our study, as it is usually used for thickness measurements.

To obtain a map of wall thickness measurements, the transducer is moved a small distance in either the X or the Z directions, collecting an ultrasonic signal at each point across the inspection area. The beam profile at the backwall is collimated, so each measurement will only probe a small area directly under the transducer. The reflected ultrasonic signal from this patch will consist of scattered energy from the surface roughness (figures 5.1). Consequently, the signal shape can change substantially between closely spaced measurements. These two effects lead to a variation in the thickness measurements across the surface, determined by the characteristics of the transducer and the surface roughness.

6.2.1 Simulation parameters

The previous chapter described an analysis of the DPSM model which showed which parameters were required to ensure DPSM simulations with an error of less than 2% in the maximum amplitude of the Hilbert envelope of the signal. These parameters were used throughout this chapter to simulate signals collected from a wall thickness C-scan. The beam from the transducer is very collimated, only probing a small footprint directly under the transducer (a $9mm$ by $9mm$ patch). Therefore, each simulated measurement only needs a small region of the surface to accurately model the reflected signal. This allows for the C-scan to be split into a large number of independent measurements, which makes the problem computationally tractable.

Each measurement is represented by a single simulation cell. A cell consists of the transducer

placed on the (X,0,Z) plane (as in figure 5.2), with a corresponding patch of the rough surface placed in the (X,$\frac{25}{3}\lambda$,Z) plane. The active point sources which model the transducer are surrounded by four layers of boundary point sources at which the pressure is set to zero (as in section 5.4.3). The patches of rough surfaces were taken from three different $200mm$ square surfaces.

The surfaces were generated with a Gaussian distributed height profile using the algorithm described in section 3.3, with RMS surface variations of $\frac{\lambda}{12}$, $\frac{\lambda}{6}$ and $\frac{\lambda}{4}$, and correlation length $\lambda_c = 2\lambda$. These surfaces were split up into 9 by $9mm$ (7λ by 7λ) patches, with a lateral separation of $1mm$. These patches each correspond to a thickness measurement taken $1mm$ apart in a square grid. Ultrasonic simulations for each cell were performed producing a reflected signal. Thicknesses were calculated from each cell using a timing algorithm described in section 6.2.2 and they were used to calculate cumulative distribution functions (section 6.4.2) for the surfaces.

6.2.2 Timing algorithms

The transducer operates as a pulse-echo sensor with the only scattering occurring at the pipe boundaries. The thickness of the component is calculated as:

$$w = \frac{1}{2}(t_2 - t_1)c \tag{6.1}$$

where t_2 is the time of arrival of the pulse reflected from the rough surface, t_1 is the time of arrival of the reflection from the front of the component and c is the speed of sound. The time of flight (TOF) is defined as $t_2 - t_1$.

There are many ways of measuring the time of flight. The performance of three common timing algorithms, presented below, is compared as part of this study. In all of the algorithms, the signals were interpolated, to increase the accuracy of the TOF measurement.

Envelope peak detection

The Hilbert transform was used to calculate the envelope of the pulse. The time at which the maximum occurs in this envelope is t_2:

$$t_2 = \max\left(\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(x)}{t-x}dx\right) \tag{6.2}$$

where $f(x)$ is the Hilbert transform of the reflected pulse, x is an integration variable and t is time. Similarly, t_1 is the time at which the maximum of the envelope of the transmitted pulse occurs.

Cross-correlation

The outgoing pulse f(t) and the received pulse g(t) are cross-correlated. For J samples of the signal, the cross-correlation at sample k is given by:

$$s(k) = \Sigma_{j=0}^{J} - g(k+j)f(j) \tag{6.3}$$

The minus sign in front of $g(t)$ accounts for phase reversal on reflection. The maximum point in $s(k)$ is the time of flight:

$$t_2 - t_1 = t(\max\left(s(k)\right)) \tag{6.4}$$

where $t(\max\left(s(k)\right))$ is the time at which the maximum value in $s(k)$ occurs.

Threshold first arrival

The Hilbert transform was used to calculate the envelope of the pulse. The amplitude of the envelope is normalised relative to its maximum amplitude.

A threshold amplitude was selected and the time when the part of envelope corresponding to the reflected pulse, crosses the threshold is t_2. t_1 is the point at which the the envelope of outgoing pulse crosses the threshold.

6.3 Experimental Set-up

To further validate the model, $200mm$ square patches of roughness (RMS=$\frac{\lambda}{12}$, $\frac{\lambda}{6}$ and $\frac{\lambda}{4}$, and $\lambda_c = 2\lambda$) were machined onto three $300mm$ square mild steel plates. To manufacture the plates, roughness profiles were generated using the method in section 3.3 and used to build a Solidworks^{TM}(Dassault, Vélizy-Villacoublay) model of the plate. The height profile was machined onto the plate using a Bridgeport^{TM}CNC machine (Series 2 Interact 4, Bridgeport, New York).

The minimum radius of curvature achievable by the CNC machine was $2mm$. Any surface features with a radius of curvature less than $2mm$ were filtered from the point cloud data, by performing a spatial Fourier transform of the surface and removing the relevant frequencies. Furthermore the rough surfaces taper down gradually to a flat surface, for the last 2mm of the rough patch, to allow cutter access. An example of a plate is shown in figure 6.1.

To validate the surface variation of the plates, the Talysurf^{TM}(Taylor Hobson Ltd, UK) surface profilometer was used to measure the roughness profile of several lines across the surface of each of the plates. From these profiles, the RMS and correlation length of the surface was calculated. Figure 6.2 shows the measured RMS surface variation against the target RMS surface variation. There is good agreement between the RMS of the real surfaces and the RMS of the computer generated surface. Furthermore, figure 6.2 shows that the measured correlation lengths are all within $0.2mm$ of the target correlation length.

The experiments were performed using the set up in figure 6.3(a). An Ultrasonic Sciences Limited USL Scanner, with an Olympus NDT 5MHz 6mm diameter plane longitudinal transducer (near field distance $10.8mm$[1]), was used to scan the plates containing the rough surface. The scanning frame arm draws the transducer across the surface of the plate, sending an ultrasonic pulse at pre-set positions (a $1mm$ by $1mm$ resolution raster scan).

The pulse travels through the water, entering the plate at a distance of $32mm$ from the transducer face. Then, the pulse travels through the metal, until it reflects off the rough surface. The scattered signal is then recorded with the same transducer in the form of a time-trace (A-scan). The thickness corresponding to each time-trace was extracted in the post-processing using the

[1]It should be noted that the near field distance for both the simulations and the experiments is approximately equal i.e. the rough surface is in the far field for both simulations and experiments.

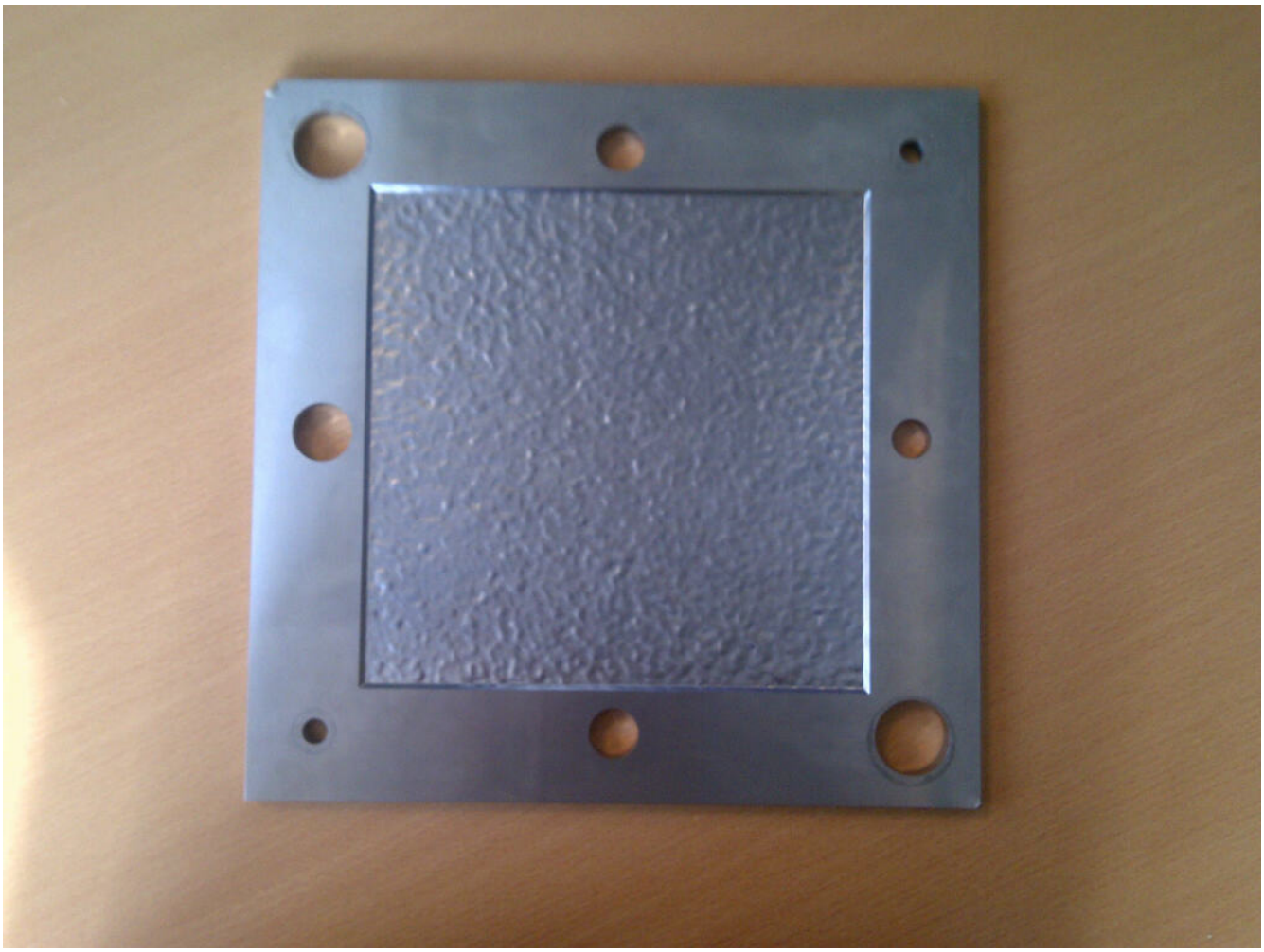


Figure 6.1: An example of a mild steel plate (300mm by 300mm external dimensions) with a Gaussian height distributed roughness profile. Reproduced from Benstock *et al.* [1].

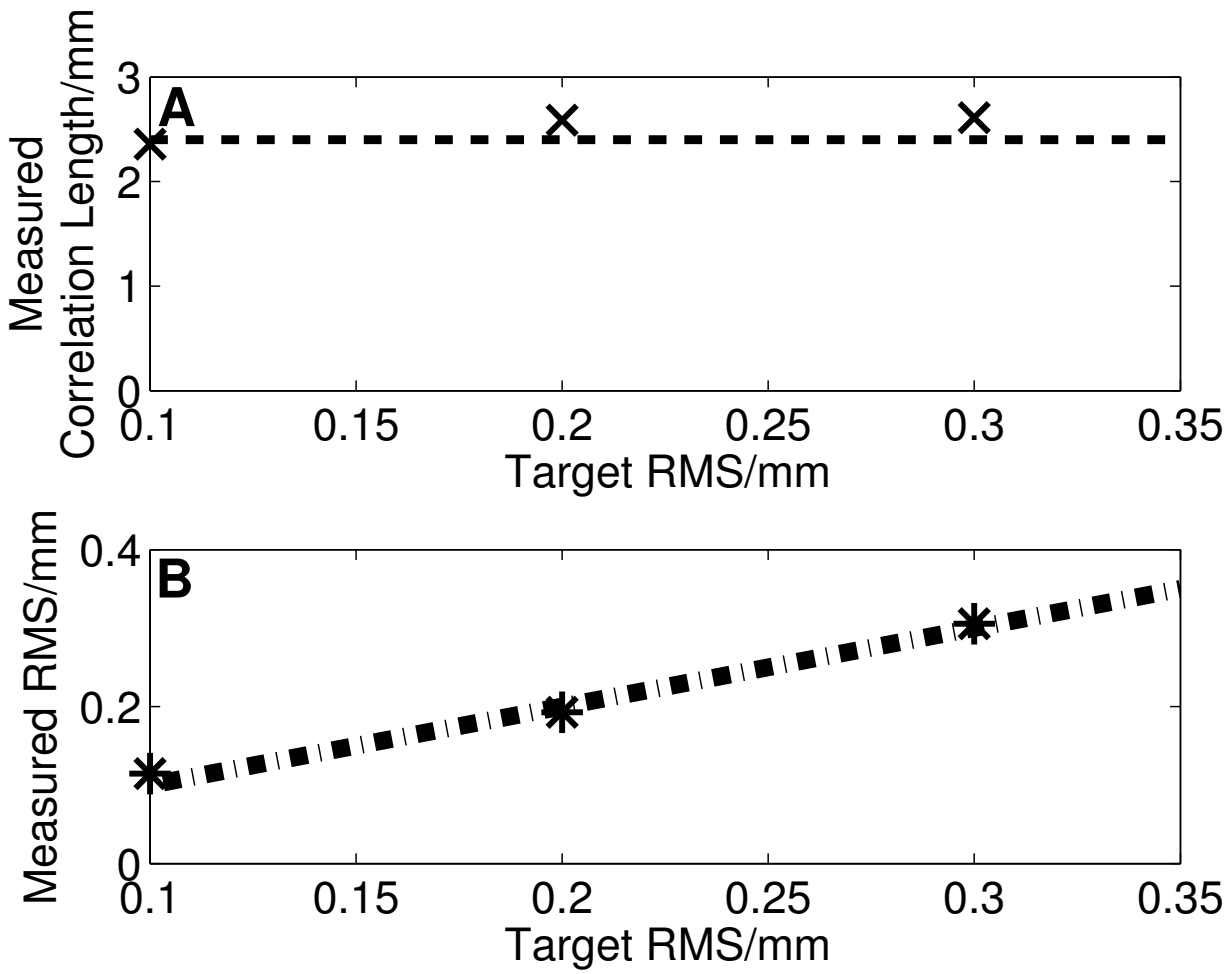


Figure 6.2: A: a comparison between the measured RMS surface variation using Talysurf plotted as a function of the RMS surface variation of the surface and the target RMS (black dashed line). B: the correlation length of the plates extracted using Talysurf plotted as a function of the RMS surface variation of the surface and the target RMS (black dashed line). Reproduced from Benstock *et al.* [1].

previously described timing algorithms.

The rough surface is in the far-field of the transducer. As a consequence the field at the backwall will be of a very similar shape to that in the simulations, albeit with a reduced amplitude [96].

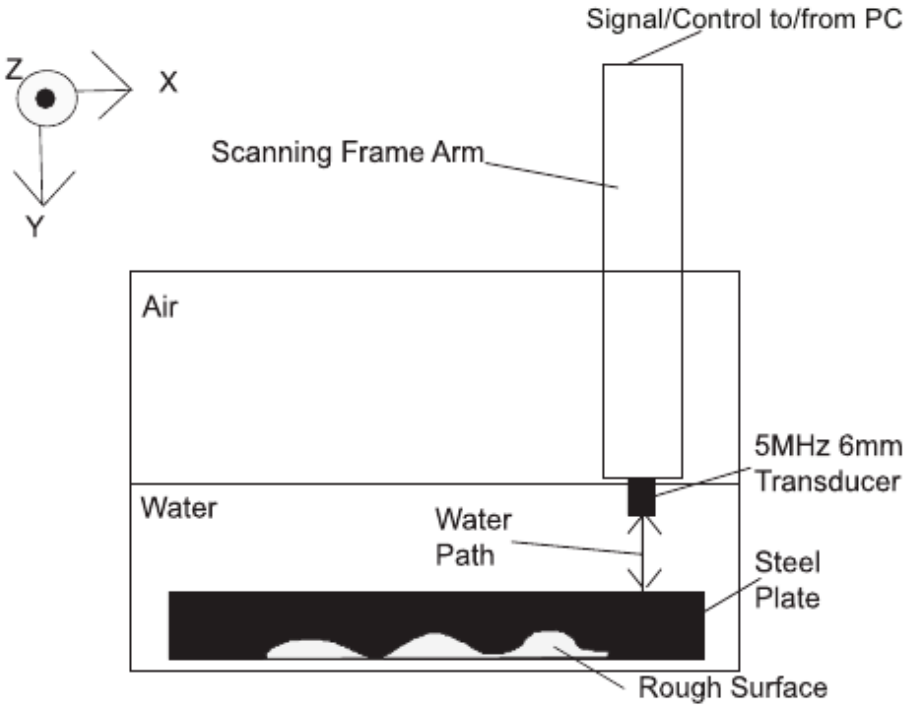
6.4 Results

6.4.1 Signals

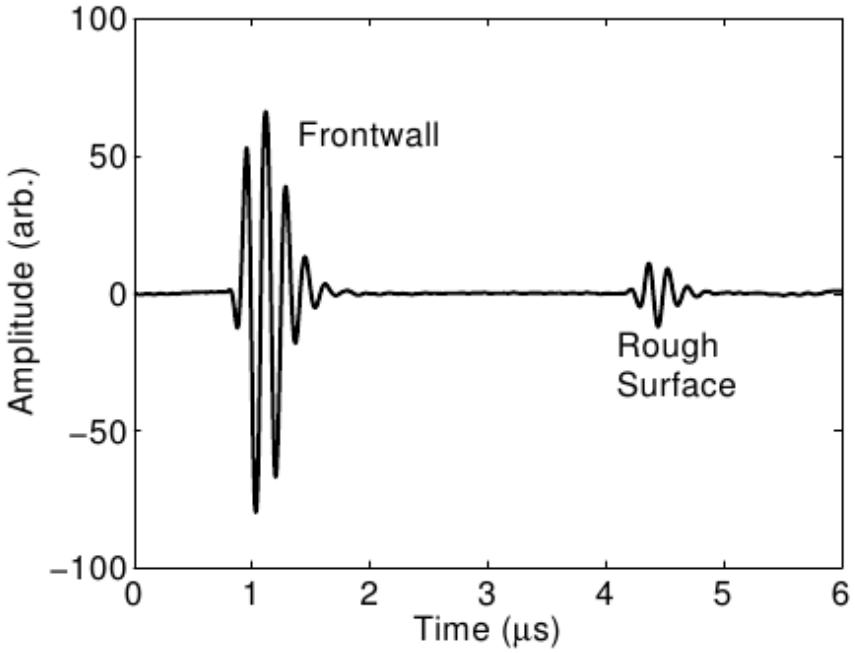
The time domain signals collected from across the surface give insight into the measurement process. The transmitted pulse travels through the steel plate and is reflected from the rough surface. Figure 6.3(b) shows a raw A-scan signal acquired by the equipment. It shows the front reflection from the flat surface of the steel plate and the reflection from the rough surface. We were only interested in the reflection from the backwall and therefore the time travelled through the plate was computed by setting the time at which the pulse enters the plate as time zero. The distance travelled in the steel plate was calculated by multiplying the arrival time of the reflection by the longitudinal wave velocity in steel (5960 m/s). Subsequently traces containing only the backwall reflection will be shown (see Figure 14)

The signals in figure 6.4.A were collected from the RMS=0.3mm plate (using the immersion scanner), from points on the surface a small distance apart (1mm). There is a clear difference between the two signals. For the solid signal, there is constructive interference between components of the signal reflected from different parts of the rough surface, leading to a large signal amplitude. In contrast, for the dashed signal, there is destructive interference between components of the signal that reflected from different parts of the rough surface, leading to reduced amplitude and changes in signal shape.

Variations in signal amplitude and shape lead to different thickness measurements. Figure 6.4.B shows the Hilbert envelope of the signals in figure 6.4.A plotted on a dB scale. The 0dB point on these lines corresponding to the maximum of the Hilbert envelope, which envelope peak detection (EPD) uses to calculate $c \cdot t_2$. For the solid signal, this point is at $23.5mm$, compared to $24.5mm$ for the dashed signal. Although the actual mean thickness at these two points is about



(a)



(b)

Figure 6.3: (a) Schematic showing the experimental set-up. The water path was set to as small as possible ($32mm$) to minimize the differences between the contact set up used in the simulations and the immersion set up in the experiments. This is in the far-field of the transducer (near field distance $10.8mm$). (b) An example of a full A-scan collected by the experimental set up. Reproduced from Benstock *et al.* [1].

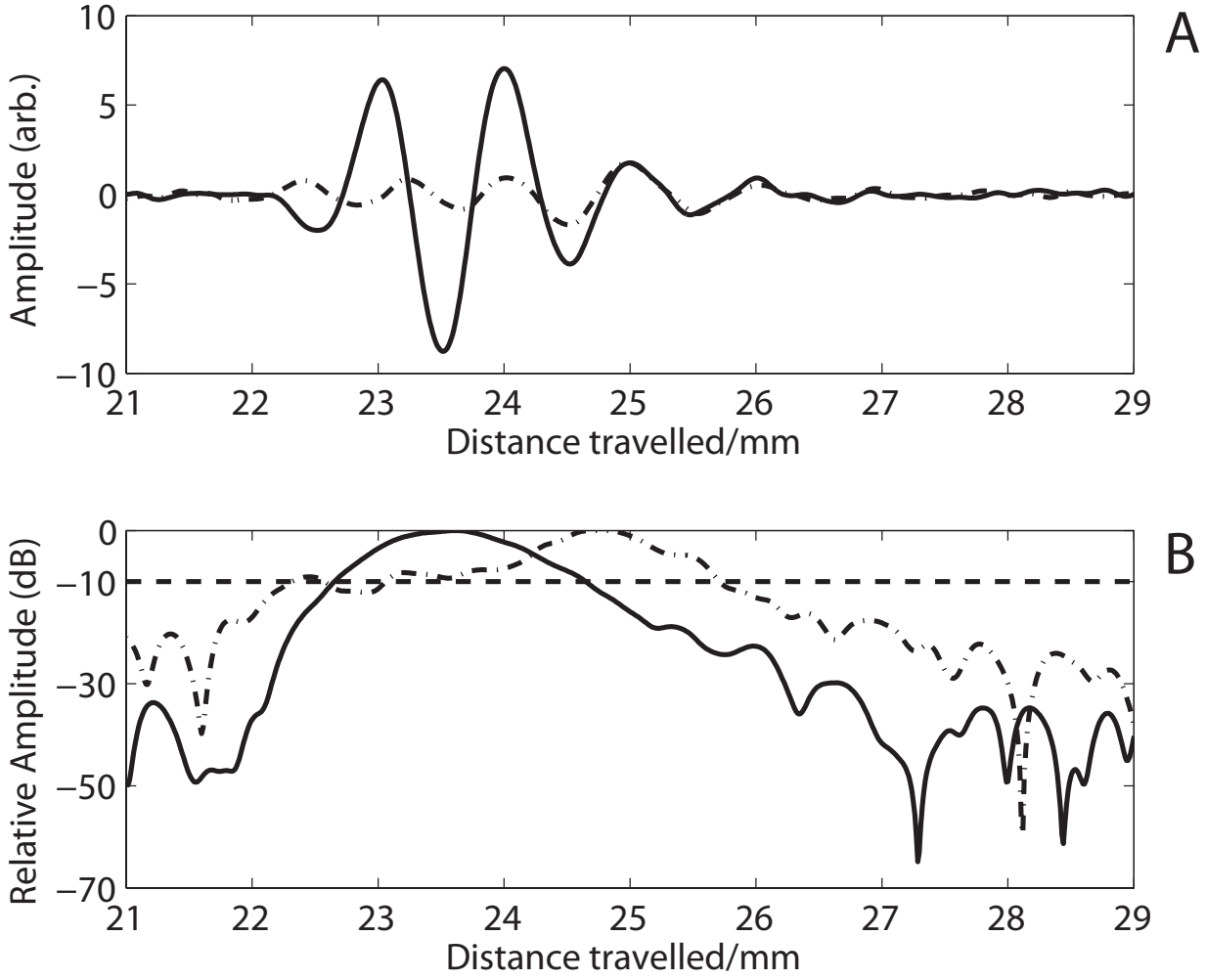


Figure 6.4: Two signals from the same rough surface (RMS=0.3mm), illustrating the amount of distortion that the rough surface can introduce to the pulse shape. The top figure (A) is the raw signal shape and the bottom (B) is the Hilbert envelope plotted on a logarithmic scale. Reproduced from Benstock *et al.* [1].

the same, the underlying surface morphology leads to alterations in the reflected signal shape and, therefore, a difference in the measured thickness.

Thickness measurement variation due to changes in pulse shape is determined by the timing algorithm. The black dashed line in figure 6.4 (bottom) is the -10dB line. The first point where the signal crosses the line, is used by threshold first arrival (TFA) to calculate the thickness. In contrast to EPD, there is only a small difference between the thickness measured from the solid and dashed signals using TFA.

The collected signal can be thought of as superposition of an average coherent pulse, corresponding to the mean thickness of the plate and a random diffuse component, introduced by the backscatter from the surface morphology. Averaging signals collected across the plate reveals the shape of the average coherent pulse. As the reflected energy is shared between the diffuse and the coherent components and the fraction of energy of the diffuse component increases with surface roughness, for rougher surfaces one would expect the average pulse to drop in amplitude. This is shown in figure 6.5. For the experimental and simulated data, it can be seen that the average signal amplitude drops rapidly with increasing RMS surface variation. The collected signal is being dominated by the random component introduced by the surface roughness. Therefore, for surfaces with larger RMS surface variations, there will be much larger changes in signal shape and amplitude, leading to larger variations in the thickness measurement.

One should note that, in figure 6.5, there is some variation in the position of the average signal. This is because the manufacturing process of the rough surfaces is difficult and complex, and offsets between the mean planes of the surfaces could not be avoided. Furthermore, the average pulse shape between the experimental and simulated mean signals (figure 6.5) is different as the transducer used in the experiments was driven by a pulser, as opposed to the Hanning windowed pulse used in the simulations. The bandwidth of both pulses is approximately the same.

6.4.2 Empirical cumulative distribution functions (ECDF)

Corrosion maps provide an overview of the condition of a component. However, it is difficult to draw quantitative conclusions from them. The cumulative distribution function (CDF), offers a

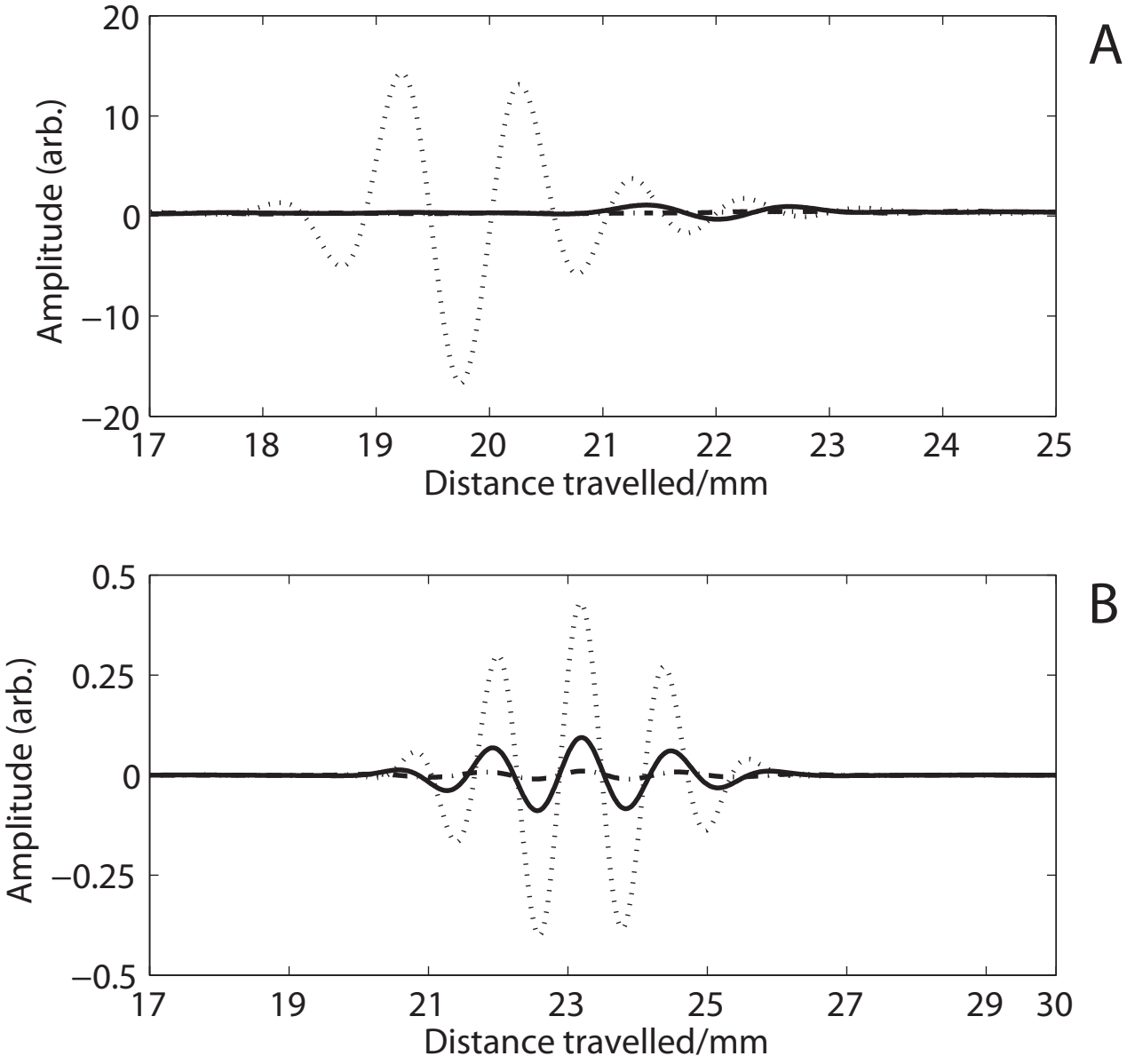


Figure 6.5: The mean signals from the RMS=0.1mm (dotted), 0.2mm (solid) and 0.3mm (dot-dashed) surface. In the top figure (A) are the mean signals from the experimental data and in the bottom (B) are the simulated results. There is a shift in time for the experimental results, as the plates have slightly different mean thicknesses. Reproduced from Benstock *et al.* [1].

compact presentation of the thickness measurements. Independent statistical distributions can be clearly distinguished and form the basis for drawing conclusions about the corrosion processes occurring in a vessel [2]. The CDF can be calculated by sorting the thickness measurements into ascending order and assigning each a rank from 1 to N (where N is the number of thickness measurements). The CDF is then:

$$F(X) = P(x < X) = \frac{i}{N+1} \tag{6.5}$$

where X is the thickness measurement, i is its rank and N is the total number of measurements. $F(X) = P(x < X)$ is the probability of measuring a thickness less than X.

Figure 6.6 shows the ECDFs of the thickness measurements from the three different plates, calculated using EPD. The crosses denote the experimental results and the circles denote the simulated results. There is close agreement between the simulations and the experiments, both showing an increasing spread in the measurements with the increase in RMS surface variation. This is expected, as for increasing RMS surface variation, the reflected signals will become more incoherent.

The majority of differences between the experimental and simulated ECDFs arises in the tail ($F(x) < 10^{-3}$), which corresponds to a small number of measurements (0.1%). The measurements in the tail underestimate the thickness of the component. These measurements correspond to signals which have undergone quite large pulse shape changes due to the surface roughness. They have a large random component resulting from the scattering from the rough surface. This is a limitation of performing ultrasonic thickness measurements across rough surfaces. In the field, an inspector will have limited knowledge of the underlying surface morphology and it is, therefore, very difficult to improve the accuracy of these thickness measurements. However, as it will be seen later in this chapter, the choice of timing algorithm can help control the extent of the differences.

Figures 6.7 (A,B), 6.8 (A,B) and 6.9 (A,B) show ECDFs for the thickness measurements extracted from the RMS=0.1, 0.2, 0.3mm plates with different timing algorithms. The graphs labelled A show the experimental results (with crosses) and the graphs labelled B show the simulated results (with circles). Each timing algorithm produces a different tail.

The tail of the distribution can be used for extrapolation purposes [2]. For example, one

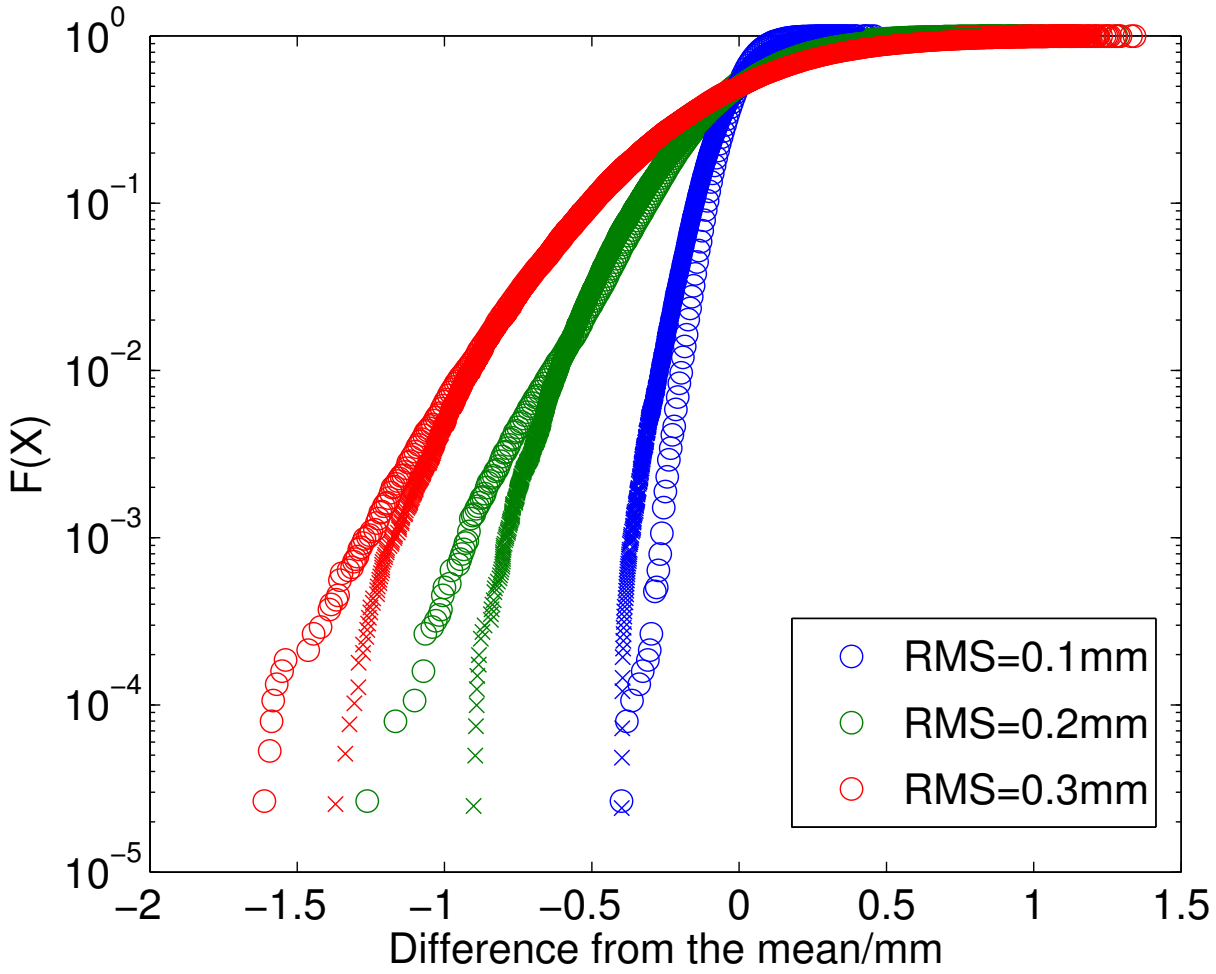


Figure 6.6: The empirical cumulative distribution function of thickness measurements extracted using the envelop peak algorithm from the RMS=0.1mm (blue), 0.2mm (green) and 0.3mm (red) surfaces. The crosses indicate the experimental results and the circles indicate the simulated results. Reproduced from Benstock *et al.* [1].

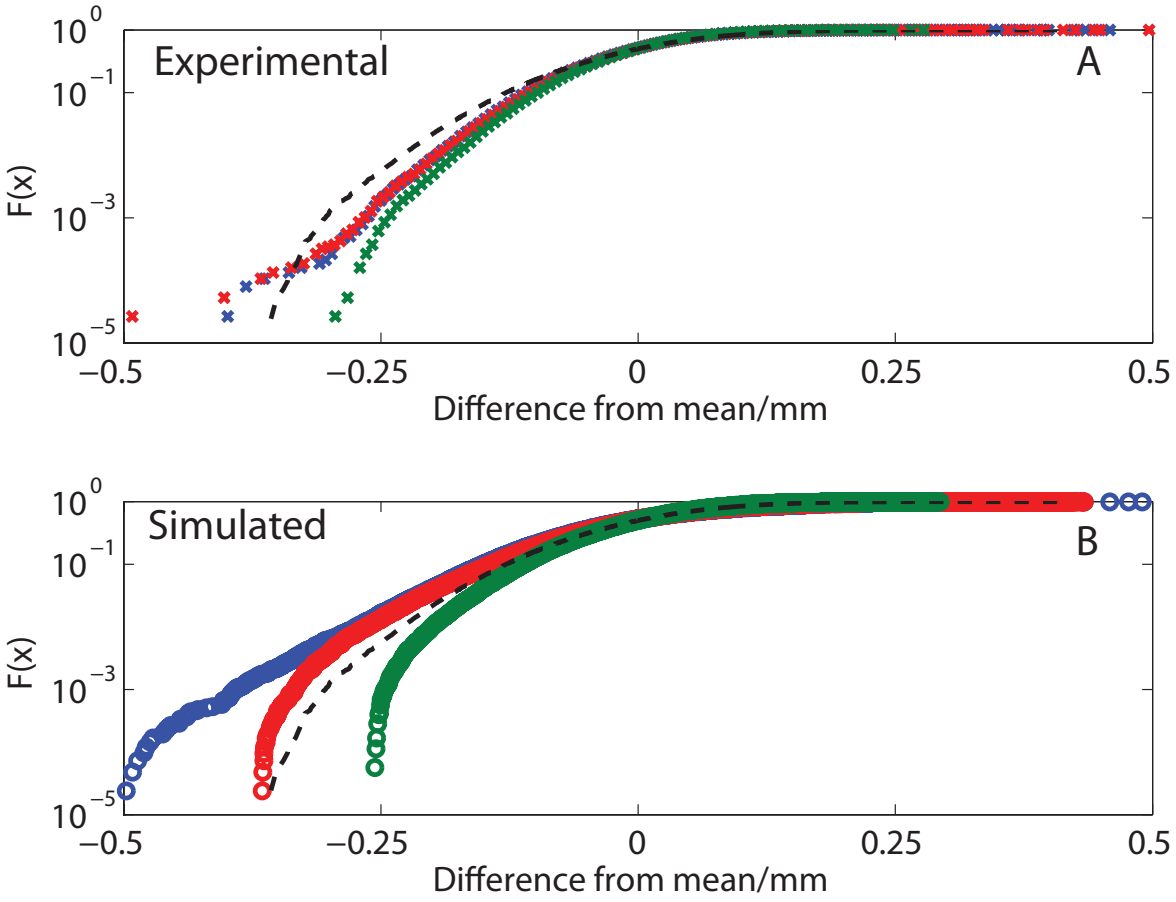


Figure 6.7: The empirical cumulative distribution functions for the RMS=0.1mm surface with different timing algorithms. The blue markers are from EPD, the red from TFA and the green from XC. The crosses are the experimental data points and the circles are the simulations. The black dashed line is the ECDF calculated from the point cloud of thickness values used to generate the surface. Reproduced from Benstock *et al.* [1].

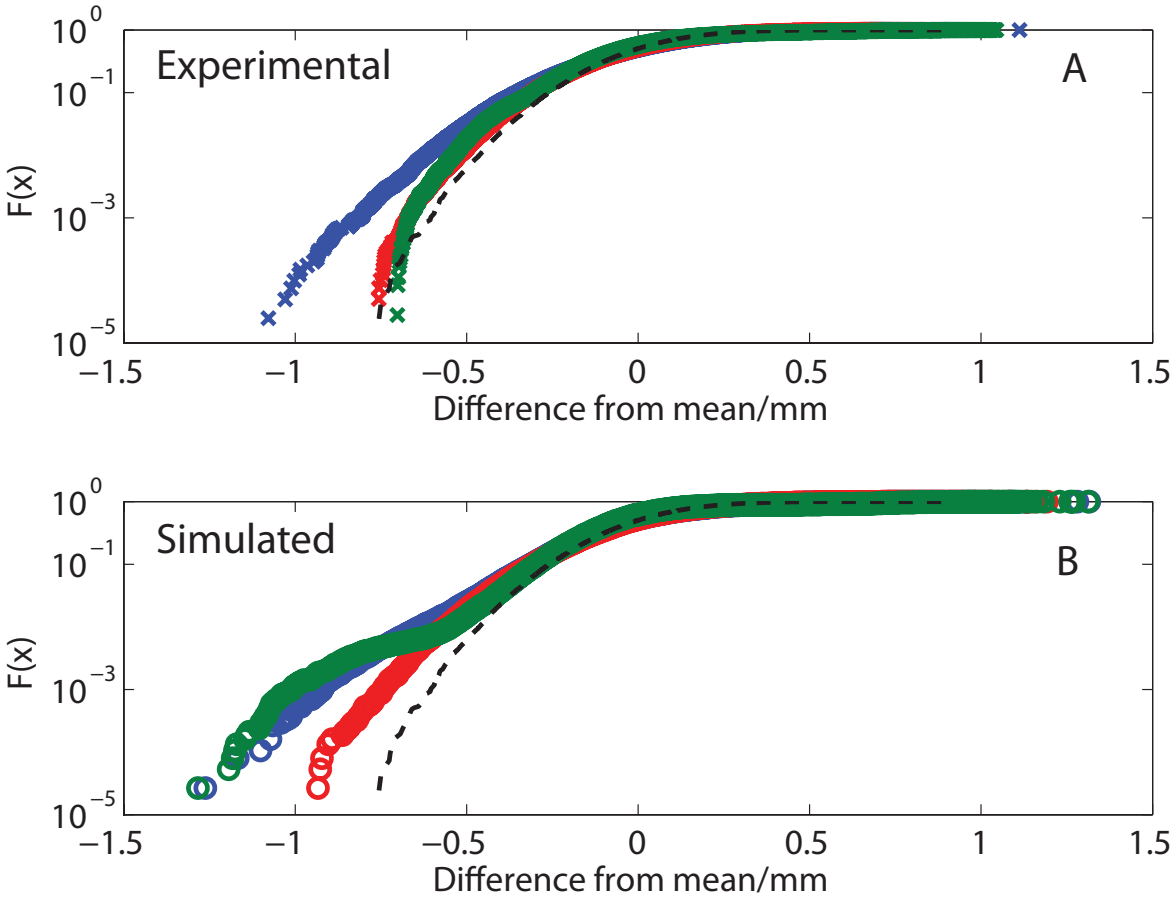


Figure 6.8: The empirical cumulative distribution functions for the RMS=0.2mm surface with different timing algorithms. The blue markers are from EPD, the red from TFA and the green from XC. The crosses are the experimental data points and the circles are the simulations.The black dashed line is the ECDF calculated from the point cloud from the point cloud of thickness values used to generate the surface. Reproduced from Benstock *et al.* [1].

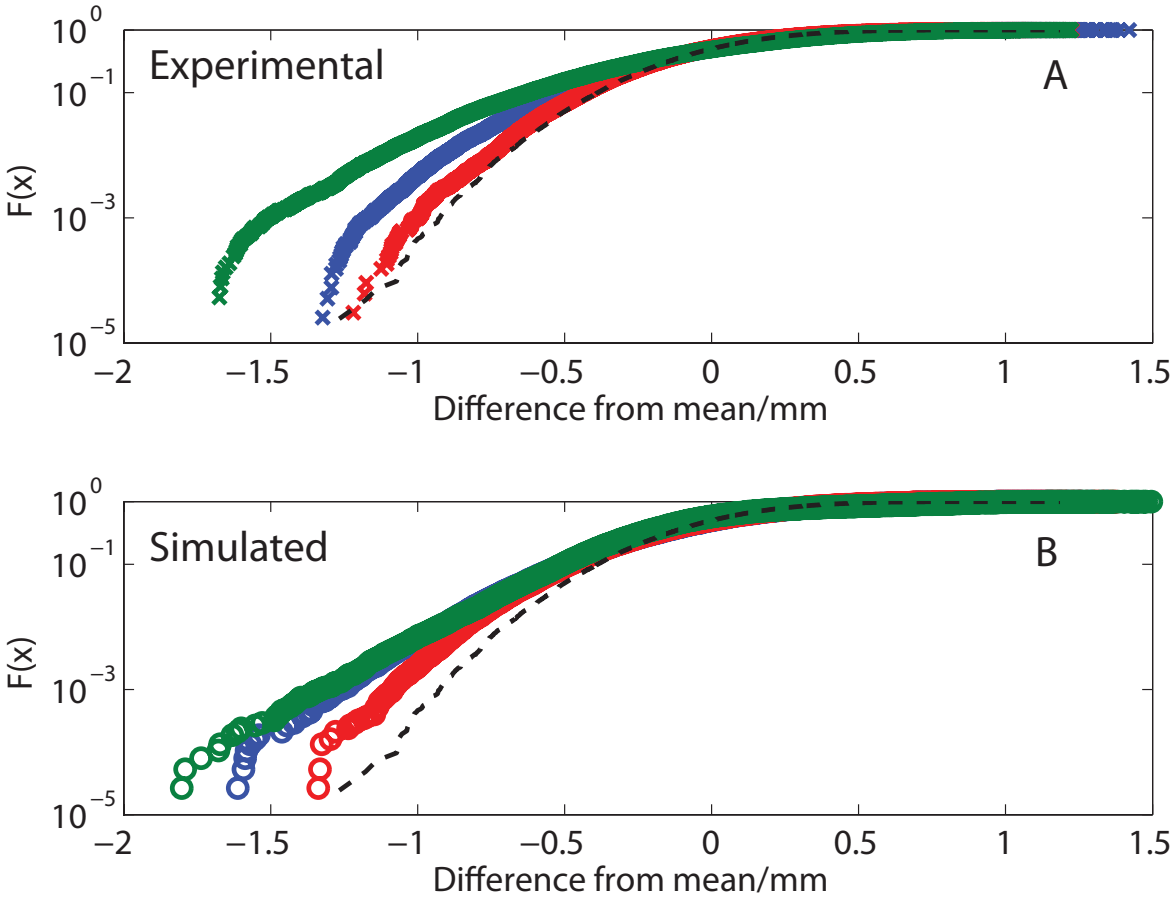


Figure 6.9: The empirical cumulative distribution functions for the RMS=0.3mm surface with different timing algorithms. The blue markers are from EPD, the red from TFA and the green from XC. The crosses are the experimental data points and the circles are the simulations.The black dashed line is the ECDF calculated from the point cloud from the point cloud of thickness values used to generate the surface. Reproduced from Benstock *et al.* [1].

could assume that uninspected areas of a component have the same thickness distribution as that measured over a small sample area. The ECDF can be interpreted as the fraction of measurements with less than a given thickness [2]. From figure 6.9A, the probability of measuring a thickness less than $\bar{x} - 1mm$ (at any random measurement point), where $\bar{x}$ is the mean thickness, is $F(x = -1mm) \approx 0.002$ for EPD and $F(x = -1mm) = 0.006$ for XC. Interpreting these values as a percentage of the area, an inspector would conclude that 0.2% of the structure would have a thickness of less than $\bar{x} - 1mm$, using EPD, or 0.6% with XC. The actual percentage of the component with less than $\bar{x} - 1mm$ thickness is 0.05% (from the point cloud).

Clearly, this has consequences for any extrapolation scheme using ultrasonic thickness measurements. Firstly, the minimum thickness in the uninspected area will be underestimated. Secondly, ultrasonic thickness measurements of the worst case defect for these surfaces will lead to overestimations of the probability of measuring less than a given thickness; the size of the overestimation is determined by the timing algorithm used.

The overestimation will get worse with increasing RMS surface variation. In figures 6.7, 6.8 and 6.9 the difference of measured thickness distribution to the point cloud of the actual thickness values, grows with increasing RMS surface variation; the rougher the surface, the larger the overestimation of the size of the worst case defect.

6.4.3 Standard deviations of the thickness measurements

ECDFs plotted on a semi-log axis are very good for showing differences in the distribution tails; however, they suppress differences in the bulk of the distribution. For example, it is hard to see quantitatively from figures 6.7, 6.8 and 6.9 how the overall spread in the measurements varies with the choice of timing algorithm or surface roughness. The measured thickness distributions in this chapter are all Gaussian (due to the nature of the the surfaces). Therefore, calculating the standard deviation of the thickness measurements will give a measure of how the spread in the measurements changes. By examining the standard deviation as a function of surface roughness, one can draw conclusions about how much measurement error could be introduced by each timing algorithm. To determine this, one needs a measure of how much of the standard deviation can be

attributed to the surface roughness alone.

The expected standard deviation from the surface roughness can be calculated by considering the patch of surface insonifed by the transducer for each measurement (section 5.4.5). It is assumed that the thickness which should be measured by the transducer is the mean thickness of this patch. The expected standard deviation of the thickness measurements is the standard deviation of the means of the patches.

In figure 6.10A the standard deviation of the thickness measurements is shown as a function of the RMS surface variation of each surface, for both the simulated and experimental results. The black dashed line in each figure is the standard deviation expected from the surface roughness. The top graph shows the standard deviation calculated using EPD, the middle using TFA and the bottom using XC.

For the RMS=0.2 and 0.3mm plates, the standard deviation of the simulations agrees closely with the experiments, for all of the timing algorithms. However, for EPD and TFA, the standard deviation of the RMS=0.1mm experimental thickness measurements is slightly higher than the simulated results. This is due to other noise sources in the experimental data which become more significant compared to the signal changes introduced by the roughness when the overall noise is low.

The effect of incoherent noise is less pronounced for XC, as it relies on pulse shape and it is very good at rejecting random noise, whilst TFA and EPD rely on the signal exceeding the noise floor [14]. This is shown by the excellent agreement between the simulated and experimental standard deviations for XC in figure 6.10A. However, it should be noted that thickness measurements extracted using XC have a larger standard deviation than both TFA and EPD. The increased spread introduced by the use of XC is caused by its reliance on pulse shape. For increasing RMS surface variation, the surface roughness has an increasingly large effect on the reflected pulse; this leads to a larger standard deviation.

EPD and TFA perform better than XC with increasing surface roughness, as they are not as susceptible to changes in pulse shape as XC. TFA performs the best out of the three algorithms as the starting point of a signal, is a more stable estimate of the thickness measurement than the peak

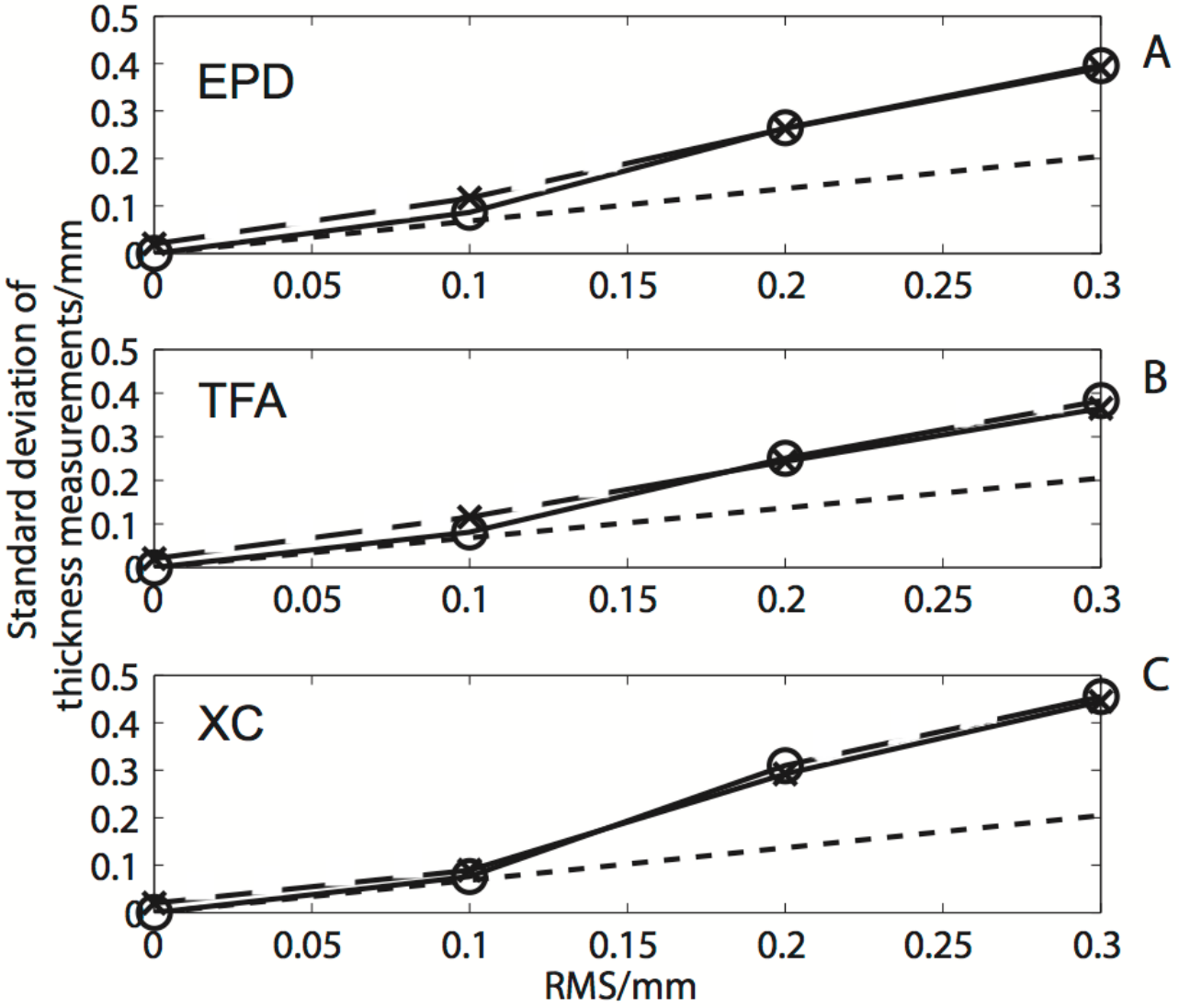


Figure 6.10: The standard deviation of the thickness measurements plotted as a function of RMS surface variation, for the EPD (A), TFA (B) and XC (C). The crosses (long dashed line) indicate the simulated results and the circles (solid line) indicate the experimental results. The black dashed line is the standard deviation which would be expected, given the point cloud. The wavelength of the centre frequency of the pulse is $\lambda = 1.2mm$. Reproduced from Benstock *et al.* [1].

of a pulse. The peak of the pulse moves around with changes in pulse shape (see section 6.4.1). Up to RMS=0.1mm ($\frac{\lambda}{12}$), the standard deviations of the measurements for all the timing algorithms match up well with the expected standard deviations. However, past this point, the standard deviations increase at a much larger rate than expected; this rate is determined by the choice of timing algorithm. Therefore, one can conclude that up to 0.1mm ($\frac{\lambda}{12}$) RMS surface variation, the spread in the measurements is dominated by the surface roughness, while for RMS surface variations greater than 0.1mm, it is dominated by errors introduced by the timing algorithm.

6.4.4 Frequency Dependence

A study of the frequency dependence of the standard deviation of the thickness measurements was also performed, with the same experimental set-up. The plates were scanned using a 3.5MHz 6mm diameter longitudinal transducer and the standard deviations of the thickness measurements were calculated. The standard deviations were plotted as a function of the RMS surface variation of the plates (figure 6.11). It should be noted that RMS surface variation is given here as a fraction of the incident wavelength. The expected standard deviation (black dashed line) was calculated using the field at the backwall for a 5MHz transducer (section 6.4.3); the line for 3.5MHz has not been included as it does not differ significantly from this line.

It is clear from figure 6.11 that the standard deviation increases as a function of $\frac{RMS}{\lambda}$. Up to $\frac{RMS}{\lambda} = 0.1$ the standard deviation is close to the expected values. However, past this point, it increases faster than anticipated. This suggests that to obtain a thickness distribution that is more representative of the actual surface, one should increase the wavelength of the incident pulse by using a lower frequency transducer. This is somewhat counter-intuitive to the usual thinking in ultrasonics where an increased frequency (shorter wavelength) is associated with improved resolution and accuracy. However, this may conflict with the need for a higher frequency to resolve the back face echoes in cases where wall-thickness is low.

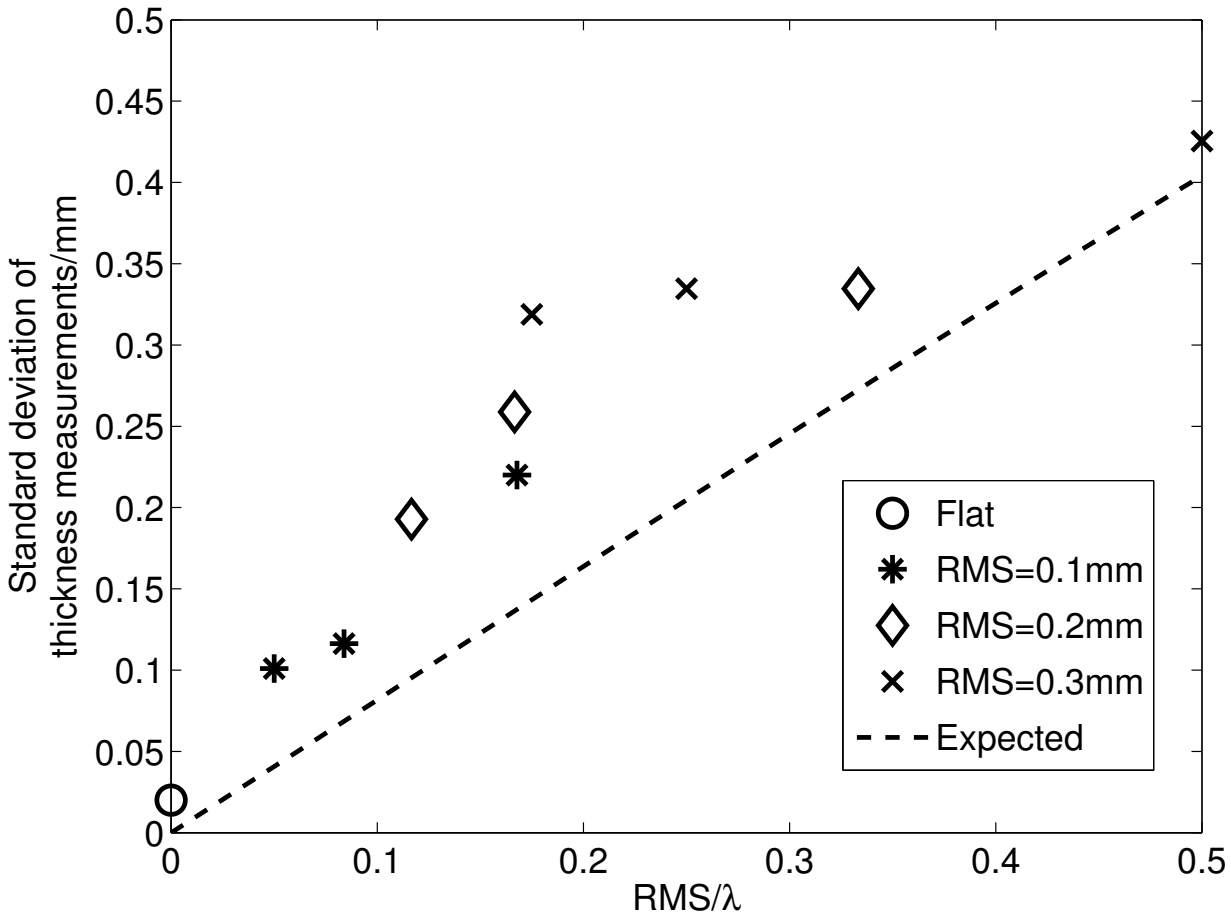


Figure 6.11: The standard deviation of the thickness measurements (extracted using EPD) plotted as a function of the RMS surface variation, given as a fraction of the wavelength of the transmitted pulse. The black dashed line is the expected standard deviation calculated by averaging over the footprint of each measurement. Reproduced from Benstock *et al.* [1].

6.4.5 Example of an extreme value model generated from inspection data

The data presented in this section has been submitted as part of a paper to the journal *Corrosion Science*.

As an additional step in the analysis of the thickness measurements, an example of an ultrasonic thickness C-scan of a steel plate machined with a Gaussian surface with RMS=0.3mm and λ_c=2.4mm was processed using the blocking algorithm (chapter 3. As the smallest thickness measurements from a UT scan are underestimates of the true thickness of a component, one would expect an extreme value model generated from UT thickness data to overestimate the severity of damage in an area. Any estimates of the minimum thickness made using an extreme value model will be smaller than the true minimum thickness.

To investigate this effect the thickness map was processed using the blocking algorithm with a significance level of 0.05. A block size of 25mm was selected. A set of minima was extracted from each using the selected block size and a GEVD was fitted to the sample minima. Figure 6.12 shows the extreme value models generated using the sample minima from both the ultrasonic data (circles) and the point cloud data (crosses). The extreme value models for each set of data are shown by the dashed lines. Graphically, the models provide a good fit to each set of minima. The SRP for the ultrasonic data was calculated to be 1.16 for the ultrasonic data and 0.93 for the model generated from the point cloud data, indicating that the models offer a good representation of the data.

The model generated from the ultrasonic data overestimates the differences from the mean thickness compared to the actual surface condition. As a consequence the extreme value model generated from the ultrasonic data overestimates the severity of the damage across a component. For example, the smallest thickness in the point cloud data is 8.7mm, which corresponds to a deviation from the mean thickness of -1.3mm. The return period of this thickness calculated from the ultrasonic extreme value model is 0.04 scans. The extreme value model generated from the ultrasonic data is very conservative compared to the true surface condition. This arises from the differences between the distribution of the minimum thickness for the point cloud data and the

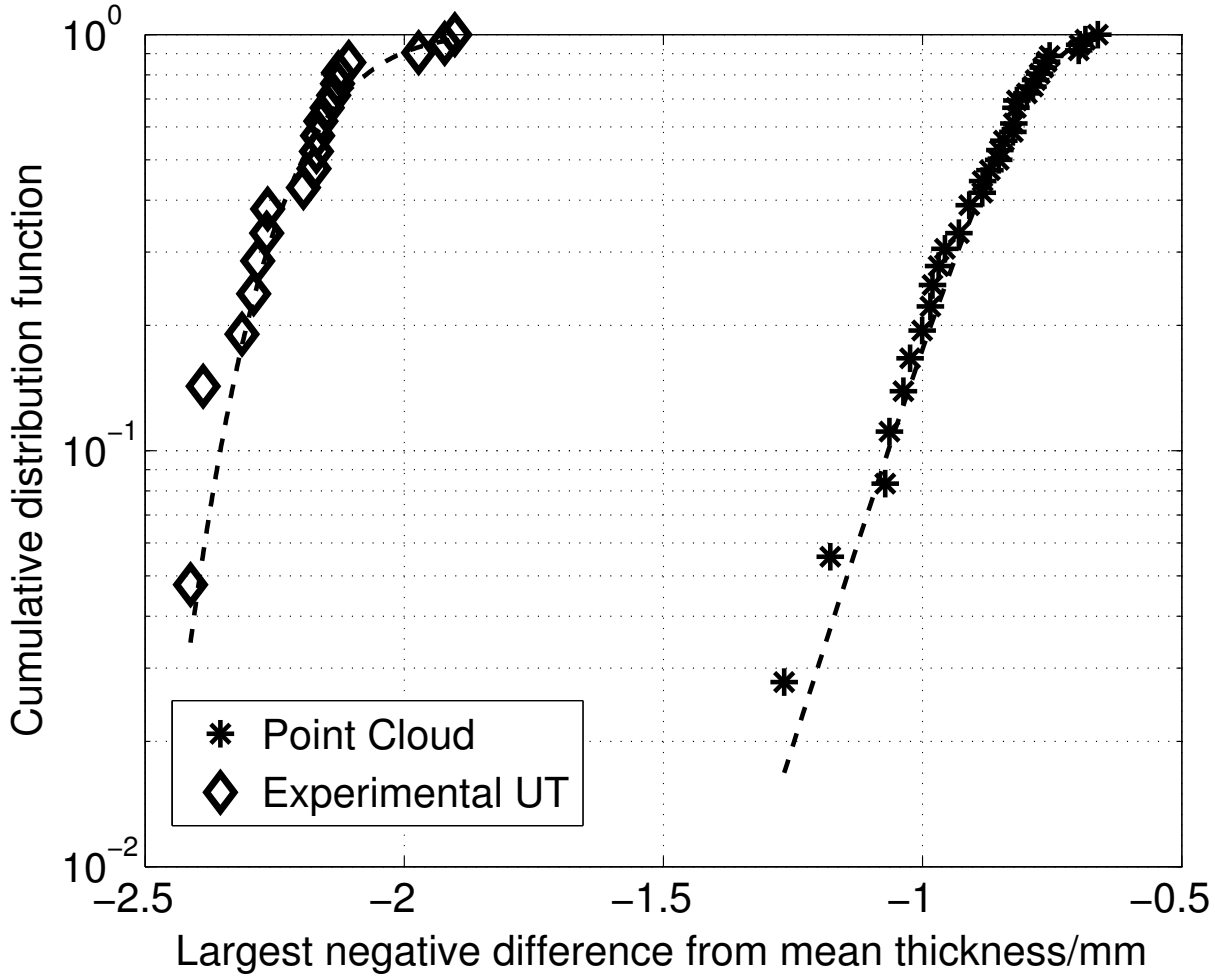


Figure 6.12: The cumulative distributions functions for the minimum thicknesses selected by the blocking algorithm from the experimental ultrasonic C-scan of the RMS=0.3mm plate. The minima selected from the point cloud used to machine the surface are denoted by the stars while the minima selected from the experimentally measured thickness map are show by the triangles. The dashed lines are maximum likelihood fits of the generalised extreme value distribution to each set of minima.

ultrasonic thickness data, discussed sections 6.4.2 and 6.4.4.

6.5 Conclusions

The modified distributed point source method (DPSM) described in chapter 5 was used to simulate C-scans of surfaces representative of a corroded engineering component. Ultrasonic reflections from around 100,000 rough surfaces were modelled and compared to experimental ultrasonic scans of the same surfaces. The statistics of the model and the experiments agreed well, showing that the DPSM simulations are an effective tool to simulate populations of ultrasonic signals from rough surfaces.

The simulated and experimentally acquired ultrasonic signals from surfaces with three different RMS surface variations were analyzed using three different timing algorithms in order to extract thickness measurement data. It was found that the thickness measurement distribution can differ significantly from the actual surface distribution, especially in the tail of the distributions and at larger RMS values. Furthermore, the shape of these distributions changed with the choice of the timing algorithm, which implies that the assessment of a component will be dependent on the timing algorithm used to extract the thickness measurements from the A-scans.

The effect of the differences between the ultrasonic thickness measurements and the actual condition of the surface is of particular importance to EVA. Due to these differences, which are largest in the tails of the distribution, extreme value models generated from the ultrasonic thickness data will be overly conservative compared to the true condition of the component. Extrapolations from extreme value models using ultrasonic thickness data will predict thinner thickness values than actually present.

The standard deviation of the thickness measurements was also investigated. It was found that up to 0.1mm ($\frac{\lambda}{12}$) RMS surface variation the standard deviation increased in proportion to the change in actual surface roughness. However, for larger RMS surface variation the standard deviation increase of measured thicknesses was larger than that of the underlying surface and dependent on the choice of timing algorithm. A study of the frequency dependence of the standard deviation was also performed. It was found that reducing the frequency of the interrogating wave,

reduced the error introduced by the timing algorithm and the overall standard deviation of the thickness measurements.

This is counter-intuitive to the general ultrasonic thinking, where increased frequency is associated with increased resolution and accuracy. The results in this chapter show that thickness distributions of rough surfaces might be more precisely assessed when interrogated at lower frequency so that $\frac{RMS}{\lambda} < 0.1$.

Additionally, it is shown in this chapter that, the blocking algorithm described in chapter 3, is capable of processing real ultrasonic thickness measurements. It successfully selected a block size for an ultrasonic thickness scan of a correlated Gaussian surfaces with RMS height of 0.3mm and a correlation length of 2.4mm. The model provided a good description of the inspection data. Furthermore, congruent to the other finding presented in this chapter, extreme value models constructed using ultrasonic inspection data can overestimate the severity of the damage across a component because of the physics of the inspection technique.

Chapter 7

Conclusions

7.1 Thesis Review

Condition assessment is an important part of the maintenance cycle for many engineering components. Inspections of a component are carried out at regular intervals, usually during plant shut down periods. These inspections are time consuming and expensive, due to both the cost of the inspection and the loss of revenue to the plant operator. Furthermore, full access to a component may be restricted due to, for example, other components or insulation. As a result, there is increasing interest in techniques to improve the efficiency of an inspection.

Partial coverage inspection (PCI) is an example of an approach to solve this problem. PCI describes the assessment of the condition of an engineering component using a representative sample of inspection data which is used to extrapolate to and estimate the condition of a larger area of the component. This thesis has focussed on statistical methods for performing PCI, namely, extreme value analysis.

Extreme value analysis (EVA) can be used to construct a statistical model of the thinnest areas of a component from an inspection of a sample area of a component (a full description of EVA can be found in chapter 2). This model can be used to extrapolate to the minimum thickness in an area much larger than the initial inspection area. The technique can be applied to data from conventional inspection techniques, such that all existing sensing technologies can be used.

EVA has not been used more extensively in industry as there is no clear understanding of the

uncertainties in the analysis, when it is a suitable approach and when it's not. These are the key issues addressed by this thesis with the development of a basic framework for performing EVA with ultrasonic inspection data. This framework includes an approach for selecting the block size (described in chapter 3) used to partition the thickness map. The approach selects a block size by looking for evidence that the thickness measurements in each block are independent and identically distributed. This method was validated using a large amount of simulated inspection data from surfaces with both Gaussian and exponential roughness profiles.

Once the blocking algorithm has checked that the inspection data meets the assumptions made by extreme value analysis an extreme value model can be constructed from the inspection data. An extreme value model gives estimates of the probability of a thickness minimum occurring in an area the size of one block. By assuming that the damage in the inspection area is representative of the damage occurring across the uninspected area of the component, this model can be used for extrapolation.

Extreme value models can be used to calculate return levels. Return levels, associated with a number of blocks M, are a threshold which will be exceeded at least once in M blocks. Extrapolations are performed by calculating the return level associated with an area larger than the inspection area.

Return levels are a standard tool for estimating the most extreme members of a distribution in fields such as hydrology [40] and insurance [38]. However, they have only seen limited use with inspection data and there were no existing studies on the uncertainties associated with extrapolations using return levels. Consequently, the use of return levels for extrapolation was validated using simulated inspection data (see chapter 4).

Understanding the uncertainty associated with extrapolation is key to the use of EVA. The profile likelihood method allows the calculation of 95% confidence intervals around return levels. Confidence intervals provide a range in which 95% of the estimates of return level will lie. As the return level is the limit which will be surpassed only once on average, the confidence interval will provide a bounds in which the smallest thickness measurement in the extrapolated area will lie.

The width of the confidence interval rapidly increases with the size of the extrapolation. Larger

extrapolations lead to larger confidence intervals. There is more uncertainty around the value of the return level and as a result the bounds in which the minimum thickness could lie are much larger. The calculation of confidence intervals for a large number of extreme value models was used to determine the width of the confidence intervals as a function of the extrapolation ratio[1] (see chapter 4). This information can be used to design an inspection such that the confidence intervals around the return levels for extrapolations from the inspection data will be of a prescribed size (on average).

Extrapolations are sensitive to the quality of the inspection data. Errors can be introduced into inspection data in a number of different ways. A large amount of literature on the source of errors in inspection data exists. However, it has been shown that the uncertainty introduced to a thickness measurement by surface roughness is uncontrollable and one of the key parameters which influences uncertainty [14]. Understanding of the effects of surface roughness on thickness measurements is important to PCI as any uncertainty in the data will lead to uncertainty in the extrapolation.

Consequently, simulations of ultrasonic thickness measurements from three Gaussian height distributed surfaces were performed. These simulations were performed using a bespoke distributed point source method (DPSM). During the development of the DPSM, a review of the the theory behind DPSM was performed. It was found that the performance of the DPSM can be improved by the inclusion of a number of boundary source points, which improve the adherence of the model to its boundary conditions (chapter 5).

The bespoke DPSM was used to develop a program, written in C++, to perform large numbers of simulations using Imperial College London's High Performance Computing systems [87]. It was used to simulate ultrasonic C-scans of three different rough surfaces, alongside equivalent experimental C-scans of steel plates onto which the rough surfaces were machined (chapter 6). The thickness measurement distribution differs from the actual surface distribution, particularly in the tail of the distributions, an effect which is magnified with increasing surface roughness. In addition, distribution shape varied with the choice of the timing algorithm, indicating that different timing algorithms would lead to different extrapolations. The simulated and experimental ultrasonic data

[1]The ratio of the inspection area to the extrapolated area.

was used to generate extreme value models. The distortions in the tail of the distributions lead to overestimations of the extent of the damage in a component. Extrapolations using extreme value models from ultrasonic thickness data will be conservative compared to the true condition of the component.

7.2 Main findings of the thesis

7.2.1 Extreme value analysis for partial coverage inspection

Extreme value analysis has been suggested as a possible tool for partial coverage inspection. However, examples of the use of EVA with real inspection data are quite limited and, prior to this thesis, limited understanding of the uncertainties and extreme value methodology reduced confidence in its practical applicability. For example, in existing literature, there is no consensus for a method for selecting the block size required for partitioning the inspection data, nor has there been any research into the uncertainties associated with extrapolation from an extreme value model.

In this thesis, research into these problems has been described. A framework for using EVA to build statistical models of the condition of large areas of an engineering component has been described. The framework begins with a method for selecting the block size, which selects a suitable block size for a large amount of simulated inspection data from both Gaussian and exponential surfaces. A method for estimating the uncertainty associated with an extrapolation has been described and it has been used to estimate the size of inspection area required to extrapolate with a given level of uncertainty for the first time. The combined knowledge of when EVA is appropriate and quantification of the uncertainty associated with the model will hopefully have an impact on EVA's use in the industry.

Method for selecting block size

Prior to this thesis there was no accepted method for selecting the block size used to construct an extreme value model which ensured that the assumptions made by EVA were met by the inspection data. An approach to solve this problem has been described in this thesis. A set of candidate block

sizes is chosen by the inspector and, for the smallest block size, the thickness map is partitioned into blocks of that size.

From each block, the empirical cumulative distribution function (ECDF) is calculated and the ECDF from every pair of blocks is compared using a two-sample Kolmogorov-Smirnov test. This test checks that the thickness measurements in each block are from the same distribution. If the tests show that the thickness distribution is the same in every block, the algorithm terminates and this is the correct block size. Otherwise, the algorithm repeats this process for the next largest block size until there are no remaining block sizes and we conclude that the inspection data is not suitable for EVA. The algorithm selects a block size by looking for evidence that the partitioned data meets the assumptions made by EVA.

This approach was tested with a large amount of simulated inspection data from simulated inspections of both exponential and Gaussian rough surfaces. It was found to select a suitable block size for the majority of inspections and generated extreme value models which provided good descriptions of the data. Block size selection is a trade-off: too small a block size and the GEVD will not provide a good model for the minimum thicknesses; too large a block size and accurate estimates of the model parameters are unattainable. The algorithm selected block sizes which provided a trade-off between minima that are "extreme enough"[2], whilst ensuring there are an adequate number of minima to fit to a generalised extreme value distribution. As a rule of thumb, for the surfaces in this study, it was shown that selection of a block size which provided 16 or 25 minima allowed for the construction of a useful extreme value model.

Uncertainty in extreme value extrapolations

Validation of an extrapolation is impossible. It requires data from outside of the inspection area, which is unavailable to an inspector. However, one can show that the model constructed is reasonable given the available data and the assumptions made by EVA, which can be achieved using the blocking method in chapter 3. With evidence that the assumptions behind the model are reasonable, confidence intervals can be used to quantify the uncertainty around the return level.

Extreme value models of simulated inspection data were generated and used to validate the

[2]That is, they occur in the tails of the underlying thickness distribution.

return level for extrapolation of inspection data. It was found that the return levels matched up with the relevant quantiles in a set of data corresponding to an extrapolated area, however, for larger extrapolations they began to diverge from the correct values. The profile likelihood method was used to calculate the confidence intervals around the return level and it was shown that the smallest thickness in the extrapolated area would lie within these bounds. The confidence intervals around the return level allow an inspector to determine the range in which the smallest thickness measurement lies.

The confidence intervals for a large number of models were calculated at a number of extrapolation ratios. The average width of the confidence intervals was plotted as a function of the extrapolation ratio. This provides insight into the size of the inspection area required to obtain an estimate of the minimum thickness in the extrapolation area to within a given percentage error. Models generated using smaller block sizes, which lead to an increased number of minima, were shown to give narrow confidence bounds on the return level. However, the minima selected for these models are not from the tails of the underlying thickness distribution. Consequently, these models will underestimate the severity of the smallest thickness minima. In congruence with the conclusions in chapter 3, models generated using 16 or 25 minima tended to provide reasonable confidence bounds on the return level, whilst ensuring that the model provides accurate estimates of the most severe minimum thickness.

7.2.2 Accuracy of ultrasonic thickness measurements

Surface roughness introduces distortion to ultrasonic signals collected as part of an inspection. It is difficult to consistently extract a thickness measurement from a distorted signal. Different timing algorithms lead to different thickness measurements. This results in uncertainty as to what the true value of thickness is. Any uncertainty in the inspection data will lead to uncertainty in the extrapolation.

Prior to this research project, the effects of surface roughness and the timing algorithm were poorly documented. Consequently, joint computational and experimental studies were performed to investigate the effect of the surface roughness on ultrasonic C-scan data. This research was

presented in two parts: the development of a fast and accurate distributed point source method, and the results of the simulated and joint experimental scans of three Gaussian height distributed surfaces.

To simulate a large number of ultrasonic thickness measurements a fast, accurate simulation method is required. The method chosen was the distributed point source method as it has been shown to be faster than more common finite element methods [14]. An extension to the DPSM was described. It was shown that, with the addition of boundary source points, improved the matching of the DPSM solution to an analytical benchmark. This extension to the DPSM was implemented in a C++ code to perform calculations efficiently on Imperial College's high performance computing cluster [87]. A full suite of convergence studies was performed in order to ensure the accuracy of the program.

The program was used to simulate C-scans of surfaces representative of a corroded engineering component using a 6mm diameter 5MHz longitudinal transducer. Concurrently, experimental C-scans of the same surfaces machined onto steel plates were performed. The simulated and experimentally acquired ultrasonic signals from the surfaces, were processed with three different timing algorithms to extract thickness measurements. The tail of the thickness measurement distribution differs significantly from the actual surface distribution, particularly for surfaces with larger RMS heights. The shape of these distributions changed with the timing algorithm, which implies that the assessment of a component will be dependent on the timing algorithm used to extract the thickness measurements from the A-scans.

It was found that up to 0.1mm ($\frac{\lambda}{12}$) RMS surface variation (with a correlation length of 2.4mm the standard deviation of the thickness measurements increased in proportion to the change in actual surface roughness. For RMS heights greater than this the standard deviation increase of measured thicknesses was larger than that of the underlying surface and dependent on the choice of timing algorithm. A further study of the frequency dependence of the standard deviation was also performed, it was found that decreasing the frequency of the interrogating wave reduced the error introduced by the timing algorithm and the overall standard deviation of the thickness measurements. This is counter-intuitive to the general ultrasonic thinking, where increased frequency

is associated with increased resolution and accuracy. As a final step, to investigate the effect of this bias on extrapolation models, the simulated ultrasonic thickness maps were processed using the blocking algorithm. It was found that, relative to the true component condition, the extreme value models overestimated the extent of the damage occurring across the component.

In terms of recommendations that could be made from this data, it is the author's opinion that the use of a threshold first arrival type timing algorithm will reduce the effect of scattering from a rough surface on the smallest thickness measurements. Furthermore, the frequency of the interrogating ultrasonic pulse should be reduced in order the further reduce the underestimation of the smallest thickness measurements.

7.3 The next steps

In this thesis, a framework for performing EVA with inspection data has been described. The possible errors that can occur during extrapolation have been investigated and discussed. However, this work has been performed using inspection data generated using simulations and experiments in a laboratory. Real inspection data is much more complex than these examples. Further work is required to develop the framework in this thesis into a robust tool which can handle the complicated inspection data which can arise during a real inspection. Inspectors must be confident that the methodology they are following is suitable for the majority of data they are collecting. Any framework developed for use in industry needs be able to handle more complicated thickness distributions than those described in this thesis.

Real inspection data can consist of multiple thickness measurement distributions each resulting from a different damage mechanism. In order to use EVA to build effective and accurate models using this data, a method for separating out the distributions from different mechanisms is required. An initial investigation into methods of separating the general corrosion thickness measurements from the pitting thickness measurements was described in chapter 3 and the complexity of implementing such a method inside an EVA framework is discussed. This investigation shows that it is indeed possible to automate the classification of thickness measurements. However, it requires further work in order to handle an arbitrary number of damage mechanisms and to determine exactly

how it could be implemented within an extreme value framework. The author recommends that this method would be a good starting point for further investigation of the application of EVA to real inspection data.

Furthermore, real inspection data can also contain errors or missing data. A procedure to handle this erroneous data needs to be developed and incorporated into the framework. The development of an approach to detect or to remove this data will allow for the full automation of EVA of real inspection data. If future work is able to address the issues described in this chapter, EVA will develop into a powerful tool for condition assessment of engineering components.

Bibliography

[1] Daniel Benstock, Frederic Cegla, and Mark Stone. The influence of surface roughness on ultrasonic thickness measurements. *The Journal of the Acoustical Society of America*, 136(6):3028, December 2014.

[2] Mark Stone. Wall Thickness Distributions for Steels in Corrosive Environments and Determination of Suitable Statistical Analysis Methods. In *4th European-American Workshop on Reliability of NDE*, 2009.

[3] DNV. DNV-RP-G103 Non-intrusive inspection. Technical Report January, Desk Norske Veritas, 2011.

[4] Health and Safety Executive. Public Report of The Fire and Explosion at the ConocoPhillips Humber Refinery, 2001.

[5] Great Britain. Prevention of Oil Pollution Act, 1917.

[6] Robin Young, Carlos Huggins, Keith Newton, and Tony Dunhill. A landscape for the future of NDT in the UK economy. Technical report, British Institute of Non-destructive Testing, Manchester, 2014.

[7] Lester Schmerr. *Fundamentals of Ultrasonic Non-destructive Evaluation*. Plenum Press, New York, USA, 1998.

[8] J. Blitz and G. Simpson. *Ultrasonic Methods of Non-destructive Testing*. Springer, November 1995.

[9] Baldev Raj, T. Jayakumar, and M. Thavasimuthu. *Practical Non-destructive Testing*. Woodhead Publishing, 2002.

[10] Eastman Kodak Company. *Radiography in Modern Industry*. Eastman Kodak Company, Rochester, NY, USA, 4th editio edition, 1980.

[11] Louis Cartz. *Nondestructive Testing: Radiography, Ultrasonics, Liquid Penetrant, Magnetic Particle, Eddy Current*. ASM International, Novelty, OH, USA, 1995.

[12] Oral Büyüköztürk. Imaging of concrete structures. *NDT & E International*, 31(4):233–243, August 1998.

[13] V. Vavilov, X. Maldague, B. Dufort, F. Robitaille, and J. Picard. Thermal nondestructive testing of carbon epoxy composites: detailed analysis and data processing. *NDT & E International*, 26(2):85–95, April 1993.

[14] A J C Jarvis and F B Cegla. Application of the distributed point source method to rough surface scattering and ultrasonic wall thickness measurement. *The Journal of the Acoustical Society of America*, 132(3):1325–1335, 2012.

[15] D.N. Alleyne and P Cawley. Long Range Propagation of Lamb Waves in Chemical Plant Pipework. *Materials Evaluation*, 55(4):504–508, 1997.

[16] D. N. Alleyne. Rapid, long range inspection of chemical plant pipework using guided waves. In *AIP Conference Proceedings*, volume 557, pages 180–187. AIP, April 2001.

[17] Pierre Belanger. *Feasibility of thickness mapping using ultrasonic guided waves*. PhD thesis, Imperial College London, 2009.

[18] F B Cegla. Monitoring Using Dry-Coupled Ultrasonic Waveguide Transducers. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 58(1):156–167, 2011.

[19] Permasense. `http://www.permasense.com/solutions.php`, Last Accessed: 2013-10-26.

[20] Attila Gajdacsi. *High accuracy ultrasonic degradation monitoring*. PhD thesis, Imperial College London, 2015.

[21] Peter Cawley, Frederic Cegla, and Mark Stone. Corrosion Monitoring Strategies—Choice Between Area and Point Measurements. *Journal of Nondestructive Evaluation*, 32(2):156–163, December 2012.

[22] Sieger Terpstra. Use of statistical techniques for sampling inspection in the oil and gas industry. *4th European-American Workshop on Reliability of NDE*, pages 1–19, 2009.

[23] D.E. Hawn. Extreme Value Prediction of Maximum Pits on Pipelines. *Materials Performance*, 16:29–, 1977.

[24] TWI. Guidelines for use of statistics for analysis of sample inspection of corrosion. Technical report, The Welding Institute, 2002.

[25] J B Wintle, B W Kenzie, G J Amphlett, and S Smalley. Best practice for risk based inspection as a part of plant integrity. Technical report, Health and Safety Executive, London (United Kingdom), 2001.

[26] T. Shibata. Application of extreme-value statistics to corrosion. *Journal of Research of the National Institute of Standards and Technology*, 99(4):327, July 1994.

[27] J.E. Strutt, J.R. Nicholls, and B. Barbier. The prediction of corrosion by statistical analysis of corrosion profiles. *Corrosion Science*, 25(5):305–315, 1985.

[28] WG Yi, MR Lee, JH Lee, and SH Lee. A study on the ultrasonic thickness measurement of wall thinned pipe in nuclear power plants. In *Asia-Pacific Conference on NDT*, pages 4–10, 2006.

[29] Ifan Hughes. *Measurements and their uncertainties: A practical guide to modern error analysis*. Oxford University Press, 2010.

[30] A Balkema and L de Haam. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.

[31] Masamichi Kowaka. *Introduction to Life Prediction of Industrial Plant Materials: Application of Extreme Value Statistical Method for Corrosion Analysis*. Allerton Press, Inc., New York, USA, 1994.

[32] Abdul-Hannan Ali, Daniel Balint, Andrew Temple, and Pat Leevers. The reliability of defect sentencing in manual ultrasonic inspection. *NDT & E International*, 51:101–110, October 2012.

[33] Nick Brierley, Trevor Tippetts, and Peter Cawley. Improving the reliability of automated non-destructive inspection. In *AIP Conference Proceedings*, volume 1430, pages 1824–1831. American Institute of Physics, May 2012.

[34] S F Burch, B A Stow, and M Wall. Computer modelling for the prediction of the probability of detection of ultrasonic corrosion mapping. *Insight - Non-Destructive Testing and Condition Monitoring*, 47(12):761–764, December 2005.

[35] Shiyao Liu. *Statistical methods for extreme values and degradation analysis*. PhD thesis, Iowa State University, 2013.

[36] T. R. Thomas. *Rough Surfaces*. Longman, 1982.

[37] John Harris and Horst Stocker. *Handbook of Mathematics and Computational Science*. Elsevier Science, Burlington, USA, 2014.

[38] Paul Embrechts, Claudia Kluppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

[39] James B. Wiggins. Estimating the volatility of S&P 500 futures prices using the extreme-value method. *Journal of Futures Markets*, 12(3):265–273, June 1992.

[40] E Gumbel. *Statistics of Extremes*. Dover Publications, Inc., New York, 2004.

[41] S Coles and J Tawn. Statistical methods for multivariate extremes - an application to structural design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):1–48, 1994.

[42] S. G. Coles and J. A. Tawn. Statistics of Coastal Flood Prevention. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 332(1627):457–476, September 1990.

[43] Jonathan a. Tawn. An extreme-value theory model for dependent observations. *Journal of Hydrology*, 101:227–250, 1988.

[44] Richard W. Katz. Statistics of extremes in climate change. *Climatic Change*, 100(1):71–76, May 2010.

[45] Hamid Mohtadi. Assessing the risk of terrorism using extreme value statistics. In *Proceedings of the Institute of Food Technologists' First Annial Food Protection and Defense Research Conference*, Atlanta, Georgia, 2005.

[46] P Aziz. Application of the Statistical Theory of Extreme Values To the Analysis of Maximum Pit Depth Data for Aluminum. *Corrosion*, 12(10):35–46, October 1956.

[47] Toshio Shibata. Evaluation of corrosion failure by extreme value statistics. *ISIJ International*, 31(2):115–121, May 1991.

[48] D. Rivas, F. Caleyo, A. Valor, and J.M. Hallen. Extreme value analysis applied to pitting corrosion experiments in low carbon steel: Comparison of block maxima and peak over threshold approaches. *Corrosion Science*, 50(11):3193–3204, November 2008.

[49] P Scarf and P Laycock. Applications of extreme-value theory in corrosion engineering. *Journal of Research of the National Institute of Standards and Technology*, 99(4):313, July 1994.

[50] J.J Vajo, R Wei, A.C Phelps, L Reiner, G.a Herrera, O Cervantes, D Gidanian, B Bavarian, and C.M Kappes. Application of extreme value analysis to crevice corrosion. *Corrosion Science*, 45(3):497–509, March 2003.

[51] F. Caleyo, J.C. Velázquez, A. Valor, and J.M. Hallen. Probability distribution of pitting corrosion depth and rate in underground pipelines: A Monte Carlo study. *Corrosion Science*, 51(9):1925–1934, September 2009.

[52] C Schneider. Application of extreme value analysis to corrosion mapping data. In *4th European-American Workshop on Reliability of NDE*, pages 1–8, 2009.

[53] Robert E. Melchers. Extreme value statistics and long-term marine pitting corrosion of steel. *Probabilistic Engineering Mechanics*, 23(4):482–488, October 2008.

[54] P.J. Laycock, R.A. Cottis, and P.A. Scarf. Extrapolation of Extreme Pit Depths in Space and Time. *Journal of Electrochemical Society*, 137(1):64–69, 1990.

[55] Jon Wallace, Ramani Reddy, Dylan Pugh, and Jorge Pacheco. Sour Service Pit Growth Predictions of Carbon Steel using Extreme Value Statistics, January 2007.

[56] D. Najjar, M. Bigerelle, C. Lefebvre, and A. Iost. A New Approach to Predict the Pit Depth Extreme Value of a Localized Corrosion Process. *ISIJ International*, 43(5):720–725, May 2003.

[57] Malte Isacsson, Mats Ström, Holger Rootzén, and Otto Lunder. Galvanically Induced Atmospheric Corrosion on Magnesium Alloys: A Designed Experiment Evaluated by Extreme Value Statistics and Conventional Techniques. Technical report, SAE International, February 1997.

[58] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample, 1928.

[59] Marcin Glegola. *Extreme value analysis of corrosion data*. PhD thesis, TU Delft, 2007.

[60] Yuzhi Cai and Dominic Hames. Minimum Sample Size Determination for Generalized Extreme Value Distribution. *Communications in Statistics - Simulation and Computation*, 40(1):87–98, December 2010.

[61] M Hollander and D A Wolfe. *Nonparameteric statistical methods*. Wiley-Interscience, 1973.

[62] J Harris. *Handbook of mathematics and computational science*. Springer London, 1998.

[63] J.a. Ogilvy. Theoretical comparison of ultrasonic signal amplitudes from smooth and rough defects. *NDT & E International*, 19(6):371–385, December 1986.

[64] Peter B. Nagy and Laszlo Adler. Surface roughness induced attenuation of reflected and transmitted ultrasonic waves. *The Journal of the Acoustical Society of America*, 82(1):193–197, 1987.

[65] Taylor Hobson Ltd. `http://www.taylor-hobson.com/`, Last Accessed: 2015-10-20.

[66] YZ Hu and K Tonder. Simulation of 3-D random rough surface by 2-D digital filter and Fourier analysis. *International Journal of Machine Tools and Manufacture*, 32(1):83–90, 1992.

[67] MATLAB. *version 8.0.0.783*. The MathWorks Inc., 2012.

[68] Agnieszka Ostrowska. *Simulating inspections on corroded surfaces*. PhD thesis, Delft University of Technology, 2006.

[69] George Box and George Tiao. *Bayesian Inference in Statistical Analysis*. Wiley-Interscience, 1992.

[70] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag London, London, UK, 2001.

[71] Mary Boas. *Mathematical Methods in the Physical Sciences*. Wiley-Interscience, New York, USA, 2006.

[72] Sourav Banerjee, Tribikram Kundu, and Nasser a Alnuaimi. DPSM technique for ultrasonic field modelling near fluid–solid interface. *Ultrasonics*, 46(3):235–250, June 2007.

[73] Tribikram Kundu, Sourav Banerjee, and Kumar V. Jata. An experimental investigation of guided wave propagation in corrugated plates showing stop bands and pass bands. *The Journal of the Acoustical Society of America*, 120(3):1217, 2006.

[74] Tribikram Kundu, editor. *Advanced Ultrasonic Methods for Material and Structure Inspection*. ISTE, London, UK, January 2007.

[75] Tribikram Kundu, Dominique Placko, Ehsan Kabiri Rahani, Tamaki Yanagita, and Cac Minh Dao. Ultrasonic Field Modeling : A Comparison of Numerical Techniques. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 57(12):2795–2807, 2010.

[76] Jiqi Cheng, Wei Lin, and Yi-Xian Qin. Extension of the distributed point source method for ultrasonic field modeling. *Ultrasonics*, 51(5):571–580, July 2011.

[77] EK Rahani and Tribikram Kundu. Gaussian-DPSM (G-DPSM) and Element Source Method (ESM) modifications to DPSM for ultrasonic field modeling. *Ultrasonics*, 51(5):625–631, July 2011.

[78] Friedrich Moser, Laurence J. Jacobs, and Jianmin Qu. Modeling elastic wave propagation in waveguides with the finite element method. *NDT & E International*, 32(4):225–234, June 1999.

[79] Romain Vayron, Vu-Hieu Nguyen, Romain Bosc, Salah Naili, and Guillaume Haïat. Finite element simulation of ultrasonic wave propagation in a dental implant for biomechanical stability assessment. *Biomechanics and Modeling in Mechanobiology*, January 2015.

[80] J. R. Pettit, A. E. Walker, and M. J. S. Lowe. Modelling NDE pulse-echo inspection of misorientated planar rough defects using an elastic finite element method. In *41st Annual Review of Quantitative Non-destructive Evaluation*, volume 1650, pages 1730–1737. AIP Publishing, March 2015.

[81] Dominique Placko and Tribikram Kundu. Theoretical study of magnetic and ultrasonic sensors: dependence of magnetic potential and acoustic pressure on the sensor geometry. In Tribikram Kundu, editor, *Prococeedings of SPIE 4335*, pages 52–62, July 2001.

[82] Dominique Placko, Tribikram Kundu, and Rais Ahmad. Ultrasonic field computation in the presence of a scatterer of finite dimension. In Tribikram Kundu, editor, *Proceedings of SPIE 5047*, pages 169–179, July 2003.

[83] Sourav Banerjee and Tribikram Kundu. Elastic wave propagation in sinusoidally corrugated waveguides. *The Journal of the Acoustical Society of America*, 119(4):2006, 2006.

[84] EK Rahani and Tribikram Kundu. Modeling of transient ultrasonic wave propagation using the distributed point. *Ultrasonics, Ferroelectrics and Frequency Control*, 58(10):2213–2221, 2011.

[85] Rais Ahmad, Tribikram Kundu, and Dominique Placko. Modeling of phased array transducers. *The Journal of the Acoustical Society of America*, 117(4):1762, 2005.

[86] Dominique Placko and Tribikram Kundu. *DPSM for modeling engineering problems*. Wiley-Interscience, 2007.

[87] Imperial College London. Imperial College High Performance Computing Facilities. `http://www.imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing`, Last Accessed: 2013-04-15.

[88] Gael Guennebaud, Benoit Jacob, and Others. Eigen v3. `http://eigen.tuxfamily.org`, Last Accessed: 2015-10-20.

[89] OpenMP Architecture Review Board. OpenMP: The OpenMP API Specification for Parallel Programing. `http://openmp.org`, Last Accessed: 2013-02-26.

[90] MPI Forum. MPI: A Message-Passing Interface Standard. Version 2.2. `http://www.mpi-forum.org`, Last Accessed: 2013-02-26.

[91] M Frigo and S G Johnson. The Design and Implementation of FFTW3. In *Proceedings of the IEEE*, volume 93, pages 216–231, February 2005.

[92] Boost C++ Libraries. `http://www.boost.org`, Last accessed: 28/08/13.

[93] Tim Mellow. On the sound field of a resilient disk in free space. *The Journal of the Acoustical Society of America*, 123(4):1880–91, April 2008.

[94] M. V. Berry. The Statistical Properties of Echoes Diffracted from Rough Surfaces. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 273(1237):611–654, February 1973.

[95] MV Berry. On deducing the form of surfaces from their diffracted echoes. *Journal of Physics A: General Physics*, 272, 2001.

[96] Gordon Kino. *Acoustic Waves*. Prentice-Hall, 1987.

Appendix A

List of publications

Benstock, D. & Cegla, F. (2013). *Modelling the scattering of ultrasonic signals from 3D rough surfaces using the distributed point source method (DPSM)*, In Proceedings of the British Institute of Non-destructive Testing Conference 2013.

Benstock, D., Cegla, F. & Stone, M. (2014). *The influence of surface roughness on ultrasonic thickness measurements.*, The Journal of the Acoustical Society of America, 136(6), 3028.

Benstock, D. & Cegla, F. (2015). *The effect of surface roughness on extrapolation from thickness C-scan data using extreme value theory*, A paper presented at 41st Review of Progress in Non-destructive evaluation, AIP Conference Proceedings 1650, 1677.

Benstock, D. & Cegla, F. (2015). *Partial coverage inspection of corroded engineering components using extreme value analysis*, A paper presented at 42nd Review of Progress in Non-destructive evaluation (under review).

Benstock, D. & Cegla, F. (2015). *Sample selection for extreme value analysis of inspection data collected from corroded surfaces*, Corrosion Science (under review).

Benstock, D. & Cegla, F. (2015). *Uncertainty and extrapolation from extreme value models of*

inspection data, Corrosion Science (submitted).

Appendix B

Copyright Permissions

The documents provided in this appendix provide evidence that permission has been sought to reuse material which has previously been published. The first document give the author permission to use the text and figures from [1] which are used extensively in chapters 5 and 6. The second document gives the author permission to reproduce the figures in chapter 3.

Daniel Benstock <dbenstock@gmail.com>

Thank you for your RightsLink / AIP Publishing LLC transaction

1 message

Copyright Clearance Center <rightslink@marketing.copyright.com> 23 September 2015 at 10:30
Reply-To: Copyright Clearance Center <reply-fe4d1079716102757c11-14153369_HTML-1244781878-114453-91026@info.copyright.com>
To: dbenstock@gmail.com

To view this email as a web page, go here.

Do Not Reply Directly to This Email

To ensure that you continue to receive our emails,
please add rightslink@marketing.copyright.com to your address book.

RightsLink | Copyright Clearance Center

Thank You For Your Order!

Dear Mr. Daniel Benstock,

Thank you for placing your order through Copyright Clearance Center's RightsLink service. AIP Publishing LLC has partnered with RightsLink to license its content. This notice is a confirmation that your order was successful.

Your order details and publisher terms and conditions are available by clicking the link below:
http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=758005bb-3283-4071-b1f9-3c77c38013bd

Order Details
Licensee: Daniel Benstock
License Date: Sep 23, 2015
License Number: 3714700286071
Publication: The Journal of the Acoustical Society of America
Title: The influence of surface roughness on ultrasonic thickness measurements
Type Of Use: Thesis/Dissertation
Total: 0.00 USD

To access your account, please visit https://myaccount.copyright.com.

Please note: Online payments are charged immediately after order confirmation; invoices are issued daily and are payable immediately upon receipt.

Subject: RE: Figures in "Wall thickness distributions for steels in corrosive environments and determination of suitable statistical analysis methods"
From: Mark Stone <Mark.Stone@sonomatic.com>
Date: 19/10/15 17:25
To: Daniel Benstock <daniel.benstock08@imperial.ac.uk>

Hi Daniel
Good to hear from you. Please feel free to use the images, no restriction from Sonomatic.
Good luck with writing up, I look forward to seeing the thesis when its published!
Regards
Mark

-----Original Message-----
From: Daniel Benstock [mailto:daniel.benstock08@imperial.ac.uk]
Sent: 15 October 2015 17:54
To: Mark Stone <Mark.Stone@sonomatic.com>
Subject: Figures in "Wall thickness distributions for steels in corrosive environments and determination of suitable statistical analysis methods"

Dear Mark,

I'm currently writing up my PhD. I'd like to reproduce figures 1 and 3 from your paper, "Wall thickness distributions for steels in corrosive environments and determination of suitable statistical analysis methods", in my thesis. The source of the figures will be fully acknowledged. As my thesis will be made publicly available, I need to make sure that I have permission from the copyright owner to do so. Is it acceptable for me reproduce these figures? Is there any process I need to go through with Sonomatic?

Best wishes,

Daniel