Durham University

Durham E-Theses

Bayesian spatio-temporal modelling for inspection and prediction of complex problems in the petrochemical industry

Little, John

How to cite:

Little, John (2003) *Bayesian spatio-temporal modelling for inspection and prediction of complex problems in the petrochemical industry*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/1252/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

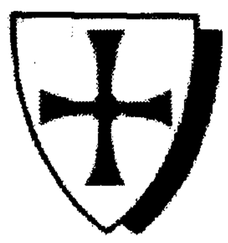
Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Bayesian spatio-temporal modelling for inspection and prediction of complex problems in the petrochemical industry

John Little

A thesis presented for the degree of
Doctor of Philosophy



Statistics Group

Department of Mathematical Sciences

University of Durham

Durham, DH1 3LE

England

2003

A copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

1 2 MAR 2004

Bayesian spatio-temporal modelling for prediction and control of complex multivariate problems in the petrochemical industry

John Little

Submitted for the degree of Doctor of Philosophy
2003

Abstract

Optimal inspection and maintenance of complex systems in modern industry is important for safety and economic reasons. With appropriate statistical modelling, the utilisation of inspection resources and quality of inferences can be greatly improved. Modelling and inspection of a full-scale industrial furnace subject to corrosion is considered. A suitable Bayesian spatio-temporal dynamic linear model for wall thickness is developed by eliciting the beliefs of experts and incorporating other relevant data for related systems. The use of the model to derive efficient inspection schedules for corrosion detection is then described and the considerable reduction in the inspection burden which the model allows demonstrated.

Concern is also with problems where the inspection method used collects transformed data, for example minimum regional remaining wall thicknesses.

The use of the model to derive efficient inspection schedules by identifying when, where and how much inspection should be made in the future is described.

Key words

Spatio-temporal, DLM, corrosion, inspection, correlation, simulation, Bayesian, minima.

Declaration

The work in this thesis is based on research carried out in the Statistics Group, at the Department of Mathematical Sciences, Durham University, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Copyright © 2003 by John Little.

"The copyright of this thesis rests with the author. No quotations from it should be published without the author's prior written consent and information derived from it should be acknowledged".

Acknowledgements

I gratefully acknowledge the help and support I have received from my family and friends, especially Anna for highlighting many areas of this thesis that benefited from additional clarification (and commas).

I would like to thank St. Aidan's College. Without their support completing my thesis would have been a much more difficult task.

I would also like to thank Shell UK for backing the CASE studentship, and everyone at Chester and Amsterdam who has helped me during my studies especially Phil Jonathan.

Special thanks go to the Durham Statistics Group but especially to Michael Goldstein, my supervisor, for his continual support.

I would like to thank the EPSRC for funding the research and making the whole project possible.

Contents

Abstract 2

Declaration 4

Acknowledgements 5

1 The inspection problem in the petrochemical industry 17

1.1 The problem 17

1.2 Overview of the approach 20

1.3 Literature review 21

1.3.1 Corrosion problems 21

1.3.2 The spatial and temporal literature 23

1.4 Outline of the thesis 28

2 Spatio-temporal models 30

2.1 The Dynamic Linear Model 31

2.1.1 Markov properties 34
2.2 The Bayes Linear approach 35
2.3 Updating the model 37
2.4 Spatial models 41
2.4.1 Discounting 44
2.5 Missing data 44
2.6 Modelling spatial correlation 46
2.6.1 Possible models 47
2.6.2 Markov covariance structures 48
2.6.3 More complex models 49
2.7 Simulation 50
2.7.1 Accuracy of simulations 52
2.8 The existing solutions 54

3 Furnace application **58**

3.1 Motivating example 59
3.2 Historical data 61
3.3 Prior beliefs 66
3.3.1 General considerations 66
3.3.2 Parameter choice for the example 67

3.4 Diagnostics 70
3.4.1 Introduction 71
3.4.2 Successive differencing 71
3.4.3 Univariate forecast error 78
3.4.4 Multivariate forecast error 78
3.4.5 Diagnostic variance 80
3.4.6 Alternative estimates for uncertainty 82
3.4.7 Further diagnostics 83
3.4.8 Estimating variances from the data 84
3.5 Validation of model 87
3.5.1 A simple check 87
3.5.2 Application of diagnostics 88
3.5.3 Sensitivity analysis 89
3.6 The effect of correlated noise 91
3.7 Concluding remarks 93

4 Future inspections **94**

4.1 Inspection criteria 95
4.2 Candidate inspection schemes 99
4.3 How many measurements 103

4.4 More advanced inspection regimes 107

4.5 Assessing the risk 109

4.5.1 Comparing inspection schemes 112

4.6 Determining the maximum inspection interval 123

5 Minimum data **127**

5.1 Motivating example 128

5.2 An illustrative model 136

5.2.1 Model parameters 137

5.3 Observing minima 138

5.3.1 More about minimum 141

5.3.2 Observational noise 146

5.4 Approximating a continuum

with a finite grid 147

6 Bayes linear forecasts based on minima observations **151**

6.1 Updating beliefs 153

6.2 Further simulation approaches 157

6.2.1 The partial simulation approach 157

6.2.2 Data with its spatial structure removed 160

6.3 Comparing full Bayesian modelling to the Bayes linear update 160

7 Finding good inspection schemes **168**

7.1 Region minima in this example 169

7.2 Criteria for comparison 174

7.3 Possible search strategies 177

7.4 Application to inspection planning 181

7.4.1 Correlation between minima 181

7.4.2 The effect of observation noise 184

7.5 Comparison between inspection schemes 186

7.6 How much improvement 190

7.6.1 Robustness 193

7.6.2 Intermediate regions 194

7.7 Prediction of global minimum 195

7.8 Discussion of global minima 202

Discussion **203**

Appendix 1: Data **214**

Furnace F4102 214

Tanks 3005 and 3006 218

Appendix 2 **227**

General closed form DLM updates 227

Comparison between a full Bayes update and a Bayes linear update 230

Comparing inspection schemes 235

List of Figures

3.1 Descriptive diagram of the furnace 60

3.2 Mean wall thickness ± 3 data s.d. 63

3.3 Wall thickness for pipe 1 64

3.4 Wall thickness for height 7' across coil 1 65

3.5 Data and forecasts at the measurement points of pipe 1, coil 1 87

3.6 Univariate forecast error density at the time of the fifth inspection 88

3.7 One step ahead forecast error density time 7 90

3.8 Data and forecasts at the measurement points of pipe 1, coil 1 92

4.1 Observation locations 102

4.2 Heuristic value using stepwise delete 105

4.3 Variance using stepwise delete 106

4.4 Number of points inspected against heuristic value

Upper line - mean heuristic, lower line - minimum heuristic . 108

4.5 Optimal inspection heights using criterion 110

4.6 Optimal inspection pipes using criterion 110

4.7 Optimal inspection coils using criterion 111

4.8 Correlations between scheme forecasts 116

4.9 Correlations between scheme failure proportions 116

4.10 Proportions in risk for full inspection against candidate schemes 118

4.11 Proportions in risk for full inspection against candidate schemes (matrix) 119

4.12 Probability of system failure through time 126

4.13 Long term probability of system failure 126

5.1 Wall thickness summary 133

5.2 Density histogram 139

5.3 Paths plot 143

5.4 Density plot (Single point) 145

5.5 Density plot (regional minima) 145

5.6 Intensities of observation 149

6.1 Forecasting through the underlying model 155

6.2 Forecasting directly 156

6.3 Simple example 161

6.4 Correlation at time 1 163

6.5 Matrix of simulations 165

7.1 64 regions of 25 points considered within the model 172

7.2 Covariance between the 64 regions at time 8 173

7.3 Variance of the 64 regions through time 173

7.4 Potential inspection locations 179

7.5 Matrix of local simulations 188

7.6 Values of $F(\Delta^*)$ using a spatial(.) and a non spatial scheme(*) 192

7.7 Gain from spatially informative sampling 193

7.8 Standard deviation of the global minimum through time following each adjustment 199

7.9 Standard deviation of the global minimum through time for different parameter values 201

List of Tables

3.1 Double data difference and estimate for W for different t, k, and n 86

3.2 Sensitivity to model parameter change 90

4.1 Some possible criteria 99

4.2 Some possible inspection schemes 101

4.3 Different criteria versus different schemes 103

4.4 Frequencies for the 4 quadrants about 0.05 for different schemes 121

4.5 Comparison between schemes 122

5.1 Data summary for tank 3005 134

5.2 Data summary for tank 3006 134

5.3 Brief description of schemes 148

5.4 Mean minimums for times 1 to 8 for schemes A to I 150

6.1 Changes in variance for different schemes 166

6.2 Sensitivity of changes in variance 167

7.1 Temporal and spatial correlation of minima (extreme values) . 183

7.2 Temporal and spatial correlation of minima 184

7.3 Temporal and spatial correlation, IID noise 185

7.4 Temporal and spatial correlation, correlated noise 186

7.5 Different inspection schemes 187

Chapter 1

The inspection problem in the petrochemical industry

1.1 The problem

Many industrial systems, including, for example, pipes, tanks, vessels and furnaces, corrode through time and with use. The integrity of these industrial systems has to be regularly monitored. A system with substantial corrosion will need to be repaired or replaced. Accurate prediction of the remaining system life via forecasts of the minimum remaining wall thickness is important for safety reasons as well as for planning replacement strategies. Forecast accuracy should, however, be balanced against the cost of making inspections and the additional safety risks associated with thinner system walls. Interest is in the remnant life of the system and the aim is to establish the system's integrity with least inspection cost. There are a huge number

of potential applications for these models, for example, a single oil refining plant may inspect thousands of these complex systems each week to try and identify corrosion regimes.

Full inspection across an entire system will usually be financially unviable, physically impossible and technically unnecessary. Often, when expensive non-intrusive inspections have to be made and critical decisions taken as a result, measurements are instead made at isolated points or local regions across the system and subsequent inferences used to check the integrity of the entire system.

There are many different non-intrusive methods used to collect data on remaining wall thickness including Magnetic flux, Ultrasound and Pigs, each of which provides data with distinct characteristics. A general introduction to non intrusive inspection is (Terpstra and Ritchie 2000) which discusses many different industrial systems and inspection devices. These and many other inspection methods widely applied in industry record only summary statistics for the area inspected rather than full spatial information. Additionally, these inspections often only cover a small proportion of a system. Common summaries are regional minima or the proportion of the area that has passed a given threshold (for instance, a protective coating). These functional outputs

may be due to the design of the data collection device, the time taken to record more extensive details or the requirements placed on the contractor. By carefully modelling the development of industrial processes, it is possible to evaluate the strengths and weaknesses of different inspection methods and schemes that could be used to monitor the system.

Within Shell, the time allowed between inspections is usually determined using an interval factor [0,1], where more critical items have a smaller interval factor. The inspection interval is then the product of the interval factor and the predicted remnant life. Some flexibility exists here since as a system comes towards the end of its life the number of inspections remains finite.

Many industrial systems can be divided into strata. Stratification is used to divide the system into parts where the corrosion regime can be considered to be influenced by the same parameters and general effects across space. For example, a simple stratification of a liquid holding vessel has a gas stratum, an interface stratum and a liquid stratum. Different corrosion characteristics will generally be expected within different strata. Many models model each stratum separately although there are often common input parameters (see for example (den Heijer and Festen 1996)).

1.2 Overview of the approach

Two distinct types of inspection are considered, firstly where measurements are made at isolated points and used to check the integrity of the entire system and secondly where measurements are made over a local area but only a data summary, for example, the local minimum wall thickness is recorded (as opposed to, for instance, the mean of the area inspected). A suitable model for the data observations and the intermediate points is essential and allows inferences to be made over the entire structure, as well as allowing us to borrow strength from neighbouring measurements.

The approach used in this thesis to handle these complex industrial problems is to establish an underlying model for the system which represents the true wall thickness across the entire system. Analysis of this underlying model allows calculations of general beliefs about complex derivations, for example, the correlation between observed minima and the rest of the system.

Using these derivations, the relationship between any observations and the underlying model can be established. Beliefs about the system quantities in the underlying model can then be updated using, for example, a Bayes linear framework. The remaining level of uncertainty within the underlying

model can then be assessed and a direct comparison made between a wide range of candidate inspection schemes.

Using the underlying model, questions about the confidence in the system's integrity in the future can be addressed. Similarly, difficult issues such as identifying when and where future inspections should be made and deciding which inspection techniques should be used may also be considered.

1.3 Literature review

The problems caused by corrosion in general and in industrial systems within Shell are discussed, as are the general spatio-temporal models available in the wider literature.

1.3.1 Corrosion problems

Historically the use of statistical modelling for dealing with wall thickness measurements in particular has been rather limited. The application of extreme value analysis for localised corrosion, however, is more common. The estimation of corrosion rate is treated very generally in industry guidelines, for example those issued by UK Health and Safety Executive (UK Health And Safety Executive 2002) and (ASTM Standard G16). No specific advice on modelling is given although the guidelines warn of the problems of chang-

ing parameters over time. Due to the lack of statistical modelling rigour. (Tallin and Conley 1994) note that the outcomes of inspections are often inconclusive. They advise the incorporation of expert opinion and historical evidence using Bayes theorem as well as for updating beliefs following inspection.

There is an extensive literature on corrosion studies both in the public domain and within Shell. Corrosion is often measured using sophisticated equipment which generally allows wall thickness measurements to be made without physically damaging the system - these are called non-intrusive inspections (NII). Detailed descriptions of some of the different inspection problems (including for example furnaces, vessels, pipes, nozzles,...) and the techniques used to measure wall thickness (including for example ultrasound, magnetic flux resonance,...) along with the inspection tools and general information about their accuracy are given in (Schipper and van Nisselroij 1998).

Inspection is usually controlled by a broader risk based inspection scheme. Risk assessments are discussed in general in (Dunn 1997). The inspection scheme will specify the requirements for the inspection tool, the maximum period allowed between inspections (the inspection interval) and the design

and implementation of the scheme - examples of these aspects of the inspection scheme may be found in (Terpstra, Ritchie, and den Heijer 2002) and (Terpstra 2002). (Gabrielli, Huet, Keddam, and Oltra 1990) write in great detail about the physics behind different inspection tools and their effect on corrosion modelling. For a detailed analysis of a particular system, with a particular inspection tool, specific and detailed literature should be consulted, and beliefs elicited directly from the experts. The operators of industrial systems, like those studied here, will usually have extensive reports about any given vessel, and the operating system and conditions, but due to the technical nature of many of these reports, they need to be interpreted by experts in the field.

1.3.2 The spatial and temporal literature

In this thesis, corrosion is modelled using a spatio-temporal dynamic linear model. The dynamic linear model (or DLM) is a well established multivariate temporal model related to the Kalman filter (Kalman 1960). There is much common ground between the Kalman filter and the DLM but while 'the Kalman filter provides an efficient computational (recursive) solution of the least-squares method' (Welch and Bishop 2003) the DLM usually has a probabilistic basis.

In addition to the standard Kalman filter, there are many nonlinear filtering systems which can sometimes be used when a linear system is inappropriate. The extended Kalman filter is the nonlinear equivalent of the standard Kalman filter and is widely used when observations and updates are a nonlinear combination of parameters. A Taylor approximation is used in place of the nonlinear components, but this in turn requires the nonlinear function to be at least once differentiable. Higher order Taylor approximations make the calculations more complicated. For a comprehensive introduction to the extended Kalman filter see (Anderson and Moore 1979). Another popular nonlinear filter is the unscented Kalman filter which was devised in (Julier and Uhlmann 1997). According to (Wan and van der Merwe 2000) "The unscented Kalman filter consistently achieves a better level of accuracy than the extended Kalman filter at a comparable level of complexity". The unscented filter approximates the model state and not the nonlinear function which describes the evolution directly. Because posterior means and variances are calculated with standard matrix calculations, the computational load is considerably lower. Calculating the unscented transformation also requires continuous derivatives. The ensemble Kalman filter was first proposed by (Evensen 1994) as an alternative to the extended Kalman filter and

subsequently extended in (Burgers, van Leeuwen, and Evensen 1997). Means and covariances are calculated using Monte Carlo integrations. This model however generally has a lower level of accuracy (see (Heemink and Segers 2002)).

Environmental problems which are commonly temporally rich in data have motivated an extensive spatio-temporal literature. A dynamic linear model (DLM) framework is a common approach used in these circumstances to update uncertainties about general model parameters as observations are made, for example, (Shaddick and Wakefield 2002) and (Stroud, Muller, and Sansó 2001). MCMC methods for DLMs are described in (West and Harrison 1996). Uncertainty in the hyperparameters is considered in a space time context by (Sansó and Guenni 2000) and (Wikle, Milliff, Nychka, and Berliner 2001), but both comment on the high computational load of their approaches. A full MCMC approach for uncertainty in the hyperparameters is described in (Muller 1998) but is only computationally tractable in a limited range of problems.

Kriging is a common interpolation technique used for making spatial inferences at locations where measurements have not been made. Kriging uses the assumption that the parameter of interest varies continuously between

measurement locations. Hence, points which are close together will have a higher spatial correlation than points that are far apart. Kriging incorporates a set of routines which, in some sense, minimise estimation variance from a predefined covariance model. Beliefs at unmeasured points are based on a weighted average of measured points, where weights are assigned to all points based on the rate of information decay when moving away from a measured point. There are many variations on Kriging - Ordinary Kriging uses only local samples for making estimates and is commonly used in environmental engineering. Universal Kriging is similar to ordinary Kriging but is used when a trend, or slow change in average values, in the samples exists. (Ripley 1981) provides a good introduction to the subject of Kriging. The use of Kriging in spatial temporal problems is demonstrated by (Mardia, Goodall, Redfern, and Alsonso 1998) and (Huang and Cressie 1995). In (Stroud, Muller, and Sansó 2001) a DLM is used directly in time but only a restricted number of spatial points are modelled. Inferences for intermediate points are then made using Kriging. A particular advantage of this approach is that data are not restricted to a lattice framework and the updates have a relatively low computational load. However, accurate modelling of localised spatial variation is likely to be more difficult.

Many authors have considered ways of reducing the computational load of the multivariate DLM. A common approach is to reduce the number of dimensions, for instance by constructing the update using summary variables. This sort of technique is most appropriate to systems which are relatively smooth across space, that is to say strongly spatially correlated. Of particular interest is (Wilke and Cressie 1999). The reduction of the number of inspections made within a spatio-temporal problem is considered in (Oehlert 1996) based on the model constructed in (Oehlert 1993), where inspection locations are fixed.

The extreme value literature is extensive especially focusing on temporal extremes given an extensive historical data set of regular observations, for example (Coles and Tawn 1994) consider a multivariate model using a limited Poisson process while (Coles and Tawn 1996) apply Bayesian methods using a limiting Poisson process to the extreme value problem. Spatio-temporal extremes are considered in (Barão and Tawn 1999) in the context of a bivariate model which uses multivariate observations to make a single prediction based on asymptotic results. Censored observations are considered in (Weissman 1978) using an asymptotic approach but not in a spatial environment.

An extreme value analysis is carried out on corrosion data based on asymp-

totic results in (Laycock, Cottis, and Scarf 1990) where measurements are of maximum pit depths. Extreme value analysis is sometimes used on corrosion problems, for example, (Shibata, Akashi, Ikematsu, and Tsuge 1998) and (Laycock, Cottis, and Scarf 1990) who both use asymptotic results to estimate maximum corrosion where sufficient data in a suitable form both temporally and spatially exists.

1.4 Outline of the thesis

In Chapter 2 a spatio-temporal dynamic linear model which can be used to model continuous systems where measurements are made over time on a finite grid is discussed, as well as the covariance structure used in the spatial model. In Chapter 3 the motivating example is introduced - an industrial furnace used in the oil refining process and the limitations of the existing model discussed. The data set and experts' prior beliefs about the means, variances and covariances of this data set and their assessment are also discussed. The discussion also covers the validation of the model using a range of diagnostics and a sensitivity analysis to check the ability of the model to distinguish changes in beliefs. In Chapter 4 future inspection plans, when and where inspections should be made, and different criteria for making these choices are

discussed. To do this a range of possible inspection strategies are discussed and their expected outcomes compared. Simulation is used to assess the overall risk of the system and to directly compare the best of the schemes from the previous section. The considerable improvements that can be made in the case of the example are also discussed. In Chapter 5 a further example and the corresponding data set of regional minima is introduced and the experts' prior beliefs and their assessment discussed. In Chapter 6 making inferences using minima data via simulation and Bayes linear adjustments is discussed. A comparison is also made for a particular example between a fully Bayesian approach and a Bayes linear approach. In Chapter 7 the potential solutions found are discussed as is the prediction of global minima using ranked observations. The thesis is then concluded with a discussion.

The material discussed in chapter 4 will be published in a special issue of Applied Stochastic Models in Business and Industry (ASMBI) which is focusing on "Innovative statistical models in the European businesses and industries" (Little, Goldstein, Jonathan, and den Heijer 2004).

Chapter 2

Spatio-temporal models

A Bayesian spatio-temporal model is a natural way to model many industrial systems with a continuous spatial element that develops in time, where there is extensive experience on the part of the operators but a relatively small amount of data. A suitable model of the underlying system is essential if good plans for the future are to be made. The approach discussed is developed using a multivariate spatio-temporal dynamic linear model (DLM). The DLM is developed extensively in (West and Harrison 1996) and their notation is followed as far as possible.

In the first section of this chapter the DLM is introduced and the sources of uncertainty discussed. In section two the Bayes linear approach to updating beliefs and circumstances when it may be especially useful are discussed.

In section three the revision of beliefs about the underlying DLM system

based on current data is discussed. In section four the DLM is extended to cover spatial models in a natural way, while in section five handling of missing spatial and temporal data is discussed. In section six a range of possible models for spatial correlation between points are discussed. Simulation of possible realisations of the system in the future is discussed in section seven. Simulating from the underlying system allows more complex distributions, for example the global minimum to be estimated. In the final section the existing non-spatial multiprocess DLM used by Shell is discussed.

2.1 The Dynamic Linear Model

The DLM uses an underlying system model to model what is happening within an unobservable system. Observations made on this system are then modelled separately, and may include observation error. Beliefs about the underlying model can be updated following one or more observations being made. It is often assumed that temporal realisations are made at times equivalent to some re-scaling of the positive integers.

Letting t index time and modelling a single point in space, θ_t represents the value of the underlying system quantity at time t, while y_t is an observation made of the value of θ_t also at time t. In the standard DLM structure,

observations may be decomposed into an underlying part and the remaining random noise. A DLM describes the structured relationship between observable time series components $y_1, y_2, \ldots$ in terms of the evolution of the unobservable state vectors $\theta_1, \theta_2, \ldots$ given the prior information about θ_0.

The DLM naturally extends to include trend terms, for example corrosion of the underlying system. Where a locally linear trend is appropriate, the system can be modelled with three coupled equations: an observation equation, a system equation and a system slope equation.

In this simple form the DLM may be summarised as follows:

$$\begin{array}{lll} \theta_t = & \theta_{(t-1)} + \omega_{\theta t} + \beta_t & \omega_{\theta t} \sim < 0, W_\theta > \\ \beta_t = & \beta_{(t-1)} + \omega_{\beta t} & \omega_{\beta t} \sim < 0, W_\beta > \\ y_t = & \theta_t + \nu_t & \nu_t \sim < 0, V_t > \end{array} \tag{2.1}$$

where $\omega_{\theta t} \sim < 0, W_\theta >$ is used to denote a random value $\omega_{\theta t}$ with expectation 0 and a variance W_θ and similarly for $\omega_{\beta t}$ and ν_t. In the context of this example W_θ, W_β and V_t correspond to expert judgements about the variance of $\omega_{\theta t}$, $\omega_{\beta t}$ and ν_t. In this example, y_t is a scalar, namely the observed wall thickness value, θ_t is the actual wall thickness and β_t is the corrosion rate, all at time t. Additional higher order terms may also be incorporated. A quadratic DLM may be used when modelling constant acceleration but higher

order terms are not widely applied in practice in this industrial context.

Uncertainty thus arises from three sources:

- Prior uncertainty about θ_0 and β_0
- Uncertainty about the data, captured in ν_t
- Uncertainty about evolution, captured in $\omega_{\theta t}$ and $\omega_{\beta t}$

For this univariate model, ν_t represents observation and calibration errors which will not have any effect on the actual wall thickness, $\omega_{\theta t}$ represents shocks in wall thickness, for example due to mishaps in operation and $\omega_{\beta t}$ represents changes in corrosion rate, for example due to long term changes in operating conditions. Throughout this thesis V, W_θ and W_β are constant in time. However, dependance on t can be introduced if it is considered appropriate. In particular, if the accuracy of the measurements changes over time, this should be reflected in changes in the value of ν_t.

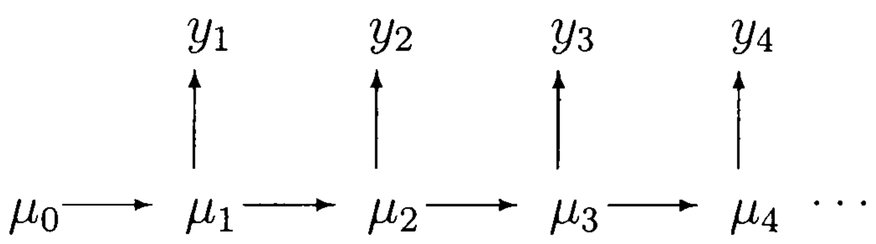
It is also assumed that the prior beliefs about the initial system level and slope are in a similar form $\theta_0 \sim < 0, C_{\theta 0} >$ and $\beta_0 \sim < 0, C_{\beta 0} >$. It is assumed that ν_s, ν_t, $\omega_{\theta s}$, $\omega_{\theta t}$, $\omega_{\beta s}$, $\omega_{\beta t}$, θ_0 and β_0 are independent for $t \neq s$ given V, W_θ, W_β, $C_{\theta 0}$ and $C_{\beta 0}$.

These uncertainties are discussed further in Section 2.4.

2.1.1 Markov properties

A key characteristic of the DLM is its conditional independence or Markov properties. That is within the underlying system, if the system values are known at time t no information from or before time t will change current or future beliefs about the system. In terms of the underlying system, θ_t is independent of all previous θ's given θ_{t-1}. A consequence of this is that an observation y_t is independent of previous observations and state vectors given θ_t. Hence, the DLM is only appropriate if, given sufficient information about the underlying system at time t, no further information from before time t would change future inferences.

This Markov property of the underlying level and slope beliefs μ_t may be summarised graphically as:



The underlying Markov structure has significant computational advantages allowing forecasts to be made in a simple and straightforward manner. Observations of a Markov system which include observational noise are non-Markov, since, as more realisations are observed beliefs about the underlying

system will become more accurate.

2.2 The Bayes Linear approach

There are many possible procedures that may be used to update beliefs within the DLM framework. These include potentially, a full analytical update, an MCMC simulation based update, or a Bayes linear update. Where every observation is a point value plus an observational error a full analytical update is possible (see Section 2.3). Where more complex observations are made, for example, the 'global minimum after observation noise', then alternative update methods should be considered.

For a given set of data a fully Bayesian approach to the problem will often be possible using MCMC techniques. MCMC simulations will give a good approximation for sufficiently large simulations. MCMC methods for DLMs are described in (West and Harrison 1996). Uncertainty in the hyperparameters is considered in a space time context by (Sansó and Guenni 2000) and (Wikle, Milliff, Nychka, and Berliner 2001), but both comment on the high computational load of their approaches. Where a range of design problems are to be considered, it is important to use an approach which can be run reasonably quickly and the faster Bayes linear approach is much bet-

ter suited to the problem. In addition to being much more computationally tractable, the Bayes linear approach bases beliefs on expectations, variances and covariances and avoids the necessity to make complex joint prior probability statements. A Bayes linear approach will give answers equivalent to full updating when observation and system noise both follow a Gaussian distribution and will give the best linear approximation for other distributions. The Bayes linear approach may also be considered appropriate when experts are only prepared to specify the mean and variance structure and not a full distribution.

The Bayes linear updates for expectation and variance are

$$E_D(\phi) = E(\phi) + cov(\phi, D)var(D)^{\dagger}[D - E(D)] \tag{2.2}$$

$$var_D(\phi) = var(\phi) - cov(\phi, D)var(D)^{\dagger}cov(D, \phi) \tag{2.3}$$

where $\dagger$ is a generalised inverse and ϕ is a vector whose values it is intended to learn about via D the observed data. $E_D(\phi)$ and $var_D(\phi)$ are termed the adjusted expectation and the adjusted variance for ϕ given D respectively. In particular $var_D(\phi)$ depends only on the observation locations and not on the observed values themselves and so can be used as the basis of a tractable design methodology. These adjustments are equivalent to a full

Bayesian update when the joint distribution between θ and D is joint multivariate normal. For more information about the Bayes linear approach, see (Goldstein 1999).

2.3 Updating the model

When observations are made, beliefs about the underlying system across time and space must be updated. Making this update when each observation is a point value plus an observational error is now discussed.

The underlying state components (the system level and slope) are combined into a vector and regression and evolution matrices are defined. Following the standard notation from (West and Harrison 1996):

D_t	All data up to time t
F	Regression matrix (known)
G	Evolution matrix (known)
a_t	Expectation of forecast state parameters at time t
R_t	Variance of forecast state parameters at time t
m_t	Expectation of state parameters at time t
C_t	Variance of state parameters at time t
f_t	Expected value of data at time t

Q_t Forecast data variance at time t

In the univariate case for a linear trend model define $\mu_t = \begin{pmatrix} \theta_t \\ \beta_t \end{pmatrix}$, $F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $W_t = \begin{pmatrix} W_{\theta t} & 0 \\ 0 & W_{\beta t} \end{pmatrix}$.

The DLM equations may now be written as:

$$\begin{aligned} \mu_t &= G\mu_{(t-1)} + \omega_t \quad & \omega_t \sim < 0, W_t > \\ y_t &= F'\mu_t + \nu_t & \nu_t \sim < 0, V_t > \end{aligned}$$

In addition to the Bayesian expectation and variance updates, the update equations for the DLM discussed below also use:

$$E(x) = E(E(x|y))$$

$$var(x|y) = var(E(x|z)|y) + E(var(x|z)|y)$$

The DLM can be sequentially updated in a closed form for example using a fully Gaussian formulation. The same updates result if a Bayes linear update is made.

$$
\begin{aligned}
a_t &= E(\mu_t|D_{t-1}) = E(E(\mu_t|\mu_{t-1})|D_{t-1}) = E(G\mu_{t-1}|D_{t-1}) = Gm_{t-1} \\
R_t &= var(\mu_t|D_{t-1}) = var(E(\mu_t|\mu_{t-1})|D_{t-1}) + E(var(\mu_t|\mu_{t-1})|D_{t-1}) \\
&= var(G\mu_{t-1}|D_{t-1}) + W_t = GC_{t-1}G' + W_t \\
f_t &= E(y_t|D_{t-1}) = E(E(y_t|\mu_t)|D_{t-1}) = E(F'\mu_t|D_{t-1}) = F'a_t \\
Q_t &= var(y_t|D_{t-1}) = var(E(y_t|\mu_t)|D_{t-1}) + E(var(y_t|\mu_t)|D_{t-1}) \\
&= V_t + var(F'\mu_t|D_{t-1}) = V_t + F'R_tF \\
m_t &= E(\mu_t|D_t) = E(\mu_t|y_t, D_{t-1}) \\
&= E(\mu_t|D_{t-1}) + cov(\mu_t, y_t|D_{t-1})var(y_t|D_{t-1})^{-1}[y_t - E(y_t|D_{t-1})] \\
&= E(\mu_t|D_{t-1}) + cov(\mu_t, E(y_t|\mu_t)|D_{t-1})Q_t^{-1}[y_t - f_t] \\
&= E(\mu_t|D_{t-1}) + var(\mu_t|D_{t-1})F'Q_t^{-1}[y_t - f_t] \\
&= a_t + R_tF'Q_t^{-1}[y_t - f_t] \\
C_t &= var(\mu_t|D_t) = var(\mu_t|y_t, D_{t-1}) \\
&= var(\mu_t|D_{t-1}) - cov(\mu_t, y_t|D_{t-1})var(y_t|D_{t-1})^{-1}cov(\mu_t, y_t|D_{t-1})' \\
&= R_t - R_tF'Q_t^{-1}FR_t'
\end{aligned}
$$

These equations can be used for updating a single data point model or more generally for a multivariate model with a vector of correlated data points (see Section 2.4).

It is assumed that G and F are time invariant although F (and V) may have to be adapted to suit the available data (see Section 2.5). W may either be explicitly modelled (potentially constant) in time or discounted (see (West and Harrison 1996)).

The variance/covariance structure between 2 underlying points μ_t and μ_{t+k} at times t and $t+k$ is:

$$\begin{pmatrix} G^t C_0 G'^t + \sum_{i=0}^{t-1} G^i W G'^i & G^t C_0 G'^{t+k} + \sum_{i=0}^{t-1} G^i W G'^{k+i} \\ G^{t+k} C_0 G'^t + \sum_{i=0}^{t-1} G^{k+i} W G'^i & G^{t+k} C_0 G'^{t+k} + \sum_{i=0}^{t+k-1} G^i W G'^i \end{pmatrix} \quad (2.4)$$

where it is assumed for simplicity that $W = W_t$ for all t.

The simplest polynomial growth model that can be applied is the constant model where the matrices F and G are identity matrices of order equal to the number of data points. A forecast expectation from this model is constant in time. In general for a model of order p, with n data points, F is np by n with 1's at points $(1,1)$, $(2,p+1)$, $(3,2p+1)$, etc., while G is np by np constructed block diagonally from n, p by p matrices with ones on and above the diagonal and zeros elsewhere. Different orders may be mixed within a single model although practically it may be simpler to set the higher order

terms to have expectation zero and a very small variance.

2.4 Spatial models

The DLM can incorporate a single observation (as described above) or a vector of observations at each time point. If a multivariate system is being modelled then the system and observation vectors ($\underline{\theta_t}$ and $\underline{y_t}$) may be indexed by space (s), that is, θ_{ts} and y_{ts}. Index s may be a scalar, a function of, for example, height, width and depth, or a vector.

The system and observation equations for any location s may now be written (with each component being a scalar) as:

$$\begin{array}{lll} \theta_{ts} = & \theta_{(t-1)s} + \omega_{\theta ts} + \beta_{ts} & \omega_{\theta ts} \sim < 0, W_{\theta s} > \\ \beta_{ts} = & \beta_{(t-1)s} + \omega_{\beta ts} & \omega_{\beta ts} \sim < 0, W_{\beta s} > \\ y_{ts} = & \theta_{ts} + \nu_{ts} & \nu_{ts} \sim < 0, V_s > \end{array} \tag{2.5}$$

Throughout this thesis $W_{\theta s}$, $W_{\beta s}$ and V_s, are constant in time. However, dependance on t can be introduced if it is considered appropriate. It is further assumed, given V_s, that ν_{ts} is correlated over s for a given t but is independently and identically distributed (IID) for different values of t. Similarly, given $W_{\theta s}$, $\omega_{\theta ts}$ is correlated over s, but IID over t and, given $W_{\beta s}$, $\omega_{\beta ts}$ is correlated over s, but IID over t. It is also assumed that the prior beliefs about the initial system level and slope are in a similar and spatially correlated form $\theta_{0s} \sim < 0, C_{\theta 0s} >$ and $\beta_{0s} \sim < 0, C_{\beta 0s} >$. Given V_s, $W_{\theta s}$,

$W_{\beta s}$, $C_{\theta 0}$ and $C_{\beta 0}$, ν_{ts}, $\omega_{\theta ts}$, $\omega_{\beta ts}$, θ_0 and β_0 are all mutually independent. ν_{ts} represents observation and calibration errors which will not have any effect on the actual wall thickness, $\omega_{\theta ts}$ represents shocks in wall thickness, for example due to mishaps in operation and $\omega_{\beta ts}$ represents changes in corrosion rate, for example due to long term changes in operating conditions.

ν_s, $\omega_{\theta s}$ and $\omega_{\beta s}$ can be modelled by any three joint distributions. In practice, to aid efficient computation and due to the difficulties of specifying more exact distributions meaningfully, the following assumptions are often made. V_s, $W_{\theta s}$ and $W_{\beta s}$ can either be considered known matrices, and then $\omega_{\theta ts}$, $\omega_{\beta ts}$ and ν_{ts} can be modelled using the Gaussian distribution, or the Bayes linear approach may be taken, considering V_s, $W_{\theta s}$ and $W_{\beta s}$ as the variances assigned by experts for the corresponding quantities. Where observations follow the form of y_{ts} in equation 2.5, that is the observation was made at a known location and comprises of a system value plus an observational error, beliefs may be updated in equivalent closed form by either approach (see section 2.3).

The fundamental advantage for our purposes is the ability to model spatially as well as temporally.

The update equations described in the previous section for a scalar prob-

lem naturally generalise to the multivariate setting. The regression and evolution matrices now take the form:

$$F = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & & 0 \\ 0 & 1 & & 0 \\ 0 & 0 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \\ 0 & 0 & & 0 \end{pmatrix} \text{ and } G = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & & 0 & 0 \\ 0 & 0 & 1 & 1 & & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & 0 & & 0 & 1 \end{pmatrix}$$

$$\text{while } \mu_t = \begin{pmatrix} \theta_{t1} \\ \beta_{t1} \\ \theta_{t2} \\ \beta_{t2} \\ \vdots \\ \theta_{tn} \\ \beta_{tn} \end{pmatrix}$$

Updates follow using exactly the same update equations as before.

Prior beliefs about the observed values at time t may be calculated using:

$$E(\underline{y}_t|D_0) = F'E(\underline{\theta}_t|D_0) = F'G^t\underline{m}_0 \tag{2.6}$$

$$var(\underline{y}_t|D_0) = V + F'var(\underline{\theta}_t|D_0)F = V + F'(Gvar(\underline{\theta}_{t-1}|D_0)G' + W_{t-1})F = \dots \tag{2.7}$$

which may be iterated until it is in terms of $var(\underline{\theta}_1|D_0) = R_1$.

The covariance between two vectors of observations at times i and j may

also be calculated. If $0 < i < j$

$$cov(\underline{y}_i, \underline{y}_j|D_0) = cov(F'\underline{\mu}_i + \underline{\nu}_i, F'\underline{\mu}_j + \underline{\nu}_j|D_0) = F'cov(\underline{\mu}_i, \underline{\mu}_j|D_0)F$$
$$= F'cov(\underline{\mu}_i, E(\underline{\mu}_j|\underline{\mu}_i)|D_0)F = F'cov(\underline{\mu}_i, G^{j-i}\underline{\mu}_i|D_0)F = F'var(\underline{\mu}_i|D_0)(G^{j-i})'F$$

2.4.1 Discounting

A simple example of discounting estimates the update system variance as a fixed proportion of the underlying system variance. This is done by weighting the current variance by the discount factor δ.

$$R_t = \frac{C_{t-1}}{\delta}$$

If the correlation between any two points is unchanged this generalises to the multivariate case.

Discounting requires no less careful prior specification. Discount factors have more effect in longer time series but many industrial inspection problems have very few measurements in time, and discounting is not applied to the model for this reason.

2.5 Missing data

Missing data is naturally handled in the DLM framework so there is no requirement for full data to be observed across all sites. Results follow through

the updates discussed in Section 2.3. To summarise, in the scalar case when data at a particular time is missing or in the multivariate case when all data at a particular time is missing no adjustment is made hence:

$$\begin{aligned} a_t &= Gm_{t-1} \\ R_t &= GC_{t-1}G' + W_t \\ m_t &= a_t \\ C_t &= R_t \end{aligned}$$

In the multivariate case where partial data is missing the update can be made as normal but the regression matrix F and the observation noise matrix V are adapted first. In F the columns and in V the rows and columns corresponding to the missing data are removed. This new form of V leaves only observation noise for the locations where observations are made. The closed form update proceeds following the updates in Section 2.3. The expectation of the observation $\underline{f}_t = E(F'\underline{\mu}_{t-1}|D_{t-1}) = F'\underline{a}_t$ where with the suitably adapted form of F, $\underline{f}_t$ includes only the expected values correspond-

ing to the values of $\underline{y}_t$ observed. The data variance follows similarly $Q_t = var(E(\underline{y}_t|\underline{\mu}_t)|D_{t-1}) + E(var(\underline{y}_t|\underline{\mu}_t)|D_{t-1}) = V_t + var(F\underline{\mu}_t|D_{t-1}) = V_t + F'R_tF$ and is again a reduced rank matrix. The adjusted system beliefs $\underline{m}_t$ and C_t follow similarly but are not of reduced rank. The underlying state is assumed not to be directly affected by a missing observation, that is, that the observation is missing is uninformative. Hence, missing observations do not effect system forecasts (a_{ts} and R_{ts}). As described above, where full multivariate data is missing at a particular time the update becomes equivalent to that described above for the scalar case.

2.6 Modelling spatial correlation

There are many different ways to introduce spatial relationships within the DLM structure. One possible approach is to adapt the evolution matrix (G) so updated beliefs become a weighted sum of neighbouring beliefs. Throughout this thesis the spatial covariance is instead modelled through the prior beliefs for the system level, the system slope ($C_{\theta 0s}$ and $C_{\beta 0s}$) and the system update matrices ($W_{\theta s}$ and $W_{\beta s}$). Additionally, for the sake of simplicity within each example the same correlation structure is assumed for $C_{\theta 0s}$, $C_{\beta 0s}$, $W_{\theta s}$ and $W_{\beta s}$ and only the variance term is changed. However, no assumption

is made, in the underlying methodology, that these variables have the same correlation structure and the method we develop allows different structures for these four components (if required) without further complication.

Once these spatial covariances are defined, using the model discussed earlier, beliefs for any location at the current time and in the future can be calculated.

2.6.1 Possible models

Many applications use data from several potentially dependent sources simultaneously. The correlation between two points of the system can be used to reduce uncertainty within the model. Often it will not be possible either to state explicitly every correlation or calculate every correlation from historic data so a model based on the various parameters is required. A possible model for correlation between two points is $ke^{-\alpha d}$ (Ripley 1981) where k and α do not functionally depend on d, a measure of distance. An alternative model for correlation which is used by (Muñoz, Carey, Schouten, Segal, and Rosner 1992) is $k\alpha^{d\phi^{\upsilon}}$ where υ is a damping parameter. An alternative geometric model αd^{υ} is proposed by (Ripley 1981). The geometric model loses the same proportion of information (α) for every unit of distance. Many more variations are possible - see for example (Cressie and Huang 1999).

Any non-negative definite matrix may be used to define a covariance structure between points in space, A non-negative definite matrix will only have non-negative eigenvalues. However, confirming that a large matrix is in general non negative definite can be difficult.

The choice of correlation model must be adapted to reflect the prior beliefs. As discussed below, models can be combined to produce a better overall system.

2.6.2 Markov covariance structures

Markov covariance structures can help give a simple inverse matrix. Working in one dimension, a covariance structure is Markov if two points are independent given an intermediate point. Using a Markov structure can decrease computational load and hence decrease run time.

Using $e^{-|d|}$ generates a Markov covariance structure where d is the distance between the two points being correlated. If the points lie on a line the inverted covariance matrix has form of a diagonal plus both neighbouring off diagonals. While moving into more dimensions destroys the simple covariance structure $e^{-|d|}$ continues to give a "simpler looking" inverse than for example e^{-d^2}. The covariance structure with the Markov property is called an "Ornstein-Uhlenbeck process". Chapter 5 of (Whittaker 1990) contains

information about the structure of the inverse of the variance matrix.

The decrease in time taken when calculating the eigenvectors and eigenvalues for the Markov covariance structure relative to more general covariance structures is insignificant compared to the computational load due to other aspects of the analysis in the examples considered in this thesis. Naturally a good model is most important.

2.6.3 More complex models

A further development of the model constructs the covariance matrix as a weighted sum of independent components. For example,

$$\omega_{\theta ts} = \sum_{n=1}^{N} \hat{\omega}_{\theta ts,n}$$

where

$$cov(\hat{\omega}_{\theta ts,m}, \hat{\omega}_{\theta ts',n}) = 0, m \neq n \tag{2.8}$$

In the furnace example (Chapter 3) the covariance structure between two points is built based on the difference in their height, the difference in their pipe number, the difference in their coil number and the difference in distance

through the coil (as the hydrocarbon flows) and

$$cov(\omega_{its}, \omega_{its'}) = W_{itss'} = \sum_{n=1}^{4} \tau_{in} e^{-c_n |d_n|^{\Phi_i}} \qquad i = \theta, \beta \qquad (2.9)$$

used where τ_{in} is a measure of the importance of the component, c_n is a relative scaling parameter (> 0), $\Phi_i \in [1, 2]$ is a general constant and d_n is a Euclidian measure of distance between s and s' for the n^{th} component of variation.

2.7 Simulation

Simulation is an ideal way of evaluating different methods for a particular problem, and thereby deciding which model forecasts should be used. The sensitivity of decisions to different model parameters can also be assessed.

Simulation may be used to construct realisations of a system consistent with given prior beliefs. Simulated data can be used to estimate expectations, variances and covariances where it would be difficult or impossible to calculate these by analytical means. For example, if the k step ahead covariance between two regional minima is required full data sets for both regions can be simulated repeatedly and hence the covariance between the minima of the two regions estimated.

Gaussian data with a given covariance structure are simulated by first

calculating the eigenvectors and eigenvalues of the covariance matrix. The matrix of eigenvectors is then multiplied with the diagonal matrix of the square-rooted eigenvalues and their product (A) by a vector of IID $N(0,1)$ draws. This is then added to the expected value of the data to obtain simulated data with the required characteristics.

That is, if the eigenvectors are $V_1, \ldots, V_n$, the eigenvalues $\lambda_1, \ldots, \lambda_n$ and $\underline{X} \sim N(\underline{0}, I_n)$ then since every eigenvector is independent with variance λ

$$(V_1, \ldots, V_n)' \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix} = \sqrt{\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

where $\underline{S}$ is the simulated draw from the distribution.

Then since $(V_1, \ldots, V_n)\,(V_1, \ldots, V_n)' = I_n$

$$\begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix} = (V_1, \ldots, V_n) \sqrt{\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}} (\underline{X}) = A\underline{X}$$

can be used to simulate a draw $\underline{S}$ with the correlation structure required.

See (Ripley 1987) page 98 for more information.

The computational load may be reduced by removing those elements of the simulation with a very small contribution. In the example introduced in Chapter 5, typically 50% of the rows of A have a maximum value of less than

10^{-10}. Using a reduced rank estimate (i.e., removing those rows with a very small effect) significantly reduces the computational time with a very small effect on the final simulation result. A further advantage of this approach is that problems due to rounding errors on very small eigenvalues making them appear negative, which, since a covariance matrix is positive semi definite and symmetric is impossible, can be reduced.

For small problems full data may be simulated simultaneously across all parts of the system which keeps the total number of simulations required reasonably small. The covariance of data simultaneously simulated across the entire system will be positive definite even for small numbers of simulations.

2.7.1 Accuracy of simulations

Expectations, variances and covariances estimated using simulation will always be approximations and this will have an impact on the estimates made. A simple estimate of the errors in the covariance structure may be made by considering the variance of the estimates for the elements known by symmetry to be the same. Putting an upper bound on the error gives an indication of how large simulations must be to obtain a given level of accuracy.

The Bayes linear adjusted expectation is

$$E_D(\phi) = E(\phi) + cov(\phi, D)var(D)^{-1}(D - E(D))$$

Estimation errors may occur in $E(\phi)$, $cov(\phi, D)$, $var(D)$ or $E(D)$. Errors for values based on a mean calculation (for example, $E(\phi)$ and $E(D)$) may be estimated using $var(\overline{\phi}) = \frac{var(\phi)}{n}$. Fourth order beliefs (that is the variance of the variance) are required to put confidence bounds on the variances and covariances.

Let $\theta^4 = E((m_i - \mu)^4)$ where μ is the expected value of m_i and m_i is a single realisation, then since each realisation is independent:

$$var(var(m_i)) = var(\frac{1}{n}\sum_1^n (m_i - \mu)^2) = \frac{1}{n^2}\sum_1^n var((m_i - \mu)^2)$$

Using $var(a) = E(a^2) - (E(a))^2$

$$var(var(m_i)) = \frac{1}{n^2}\sum_1^n \left(E(((m_i - \mu)^2)^2) - (E((m_i - \mu)^2))^2\right)$$

$$= \frac{1}{n^2}\sum_1^n (E((m_i - \mu)^4) - \sigma^4) = \frac{1}{n^2}\sum_1^n (\theta^4 - \sigma^4) = \frac{1}{n}(\theta^4 - \sigma^4)$$

In principal $cov(x, D)var(D)^{-1}$ can now be bounded. It can also be seen that to improve the estimate by a factor of κ, n must be increased by a factor of κ^2.

Most wall thickness data is only recorded to one decimal place and this can be used to suggest what level of simulation accuracy is required. Errors due to calculations which are significantly smaller than this source of error will not cause concern.

2.8 The existing solutions

Wall thickness measurements are, currently throughout Shell, widely dealt with in one of two possible ways.

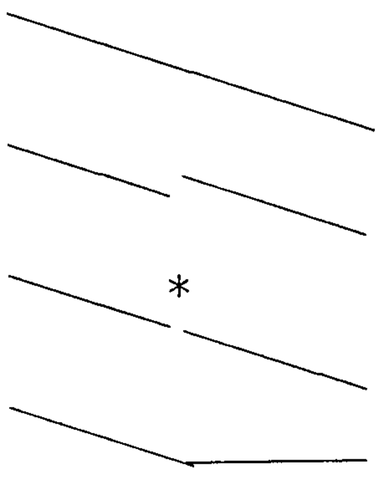
When only summary statistics are observed a common approach is to check the observation made against a predetermined criterion (for example minimum allowed wall thickness) and a decision then taken on whether it is safe to continue using the system. Often no further analysis of this data will be made. Indeed, the observed values are often not even recorded so subsequent more advanced analysis is impossible. No direct consideration is given to (i) the areas not inspected, or (ii) the current corrosion rate.

Alternatively, observed data within Shell is sometimes modelled with a multiprocess dynamic linear model (DLM) based on (Smith and West 1983). This model is only appropriate where data are measured at the same fixed locations known every time. However, the restrictions of the multiprocess

DLM are not always well understood by the engineers using the multiprocess software available. Attempts to use this approach when regional minima are observed and when observations are not always recorded at the same locations emphasises the demand for a more appropriate model.

The application of the multiprocess DLM used by Shell is explained in detail in (Hoeve, den Heijer, and Festen 1998) and (Hoeve, Jonathan, and den Heijer 2000), and an example is given in (Hoeve and Lewis 1999). (Quintana and West 1987) also used the multiprocess DLM for modelling financial time series.

The multiprocess DLM is based on four distinct possibilities for a DLM:



A straight line

A straight line with a jump (high system level noise)

A straight line with an outlier (high observation noise)

A straight line with a change in slope

(high system slope noise)

The multiprocess model is best suited to long stable time series where observations can, in some sense, be thought of as falling into one of the four categories previously described. However, each of these four possible

models requires suitable prior probabilities or weights as well as transition probabilities between each of the models. These probabilities will often be difficult for the user to specify accurately. More fundamentally the lack of a spatial element means that no information is available about any locations except those explicitly measured and all observations must be made at the same locations whenever data is observed. Including a spatial element within the multiprocess model would be possible but the resulting model would be extremely complex and ill suited to making inferences about good inspection scheme design. The lack of a spatial aspect to the model is a particular problem when data are missing.

If appropriate, the standard DLM may be similarly constrained to effectively switch between these four models by specifying the probability distribution for ν_{st}, $\omega_{\theta st}$ and $\omega_{\beta st}$ to have a high weight in their tails. This may be considered a particularly suitable model if an influential external event may or may not occur. If a system is being considered where underlying system and observation changes are likely to fall within a continuum, a suitable heavy tailed unimodal distribution may be considered a more suitable uncertainty model.

In summary, the multiprocess DLM allows forecasting of future corrosion

at measurement sites but has no spatial element and so (i) requires for every inspection, data to be measured at the same locations and (ii) gives no information about locations intermediate to the measurement points. It is now felt by experts using the multiprocess model, that for most applications such a model is unnecessarily over parameterised given the limited data, and that a spatial model coupled with diagnostic warnings to identify unexpected observations and manual interventions to recalibrate model parameters when necessary, would better represent the process.

Shell view the introduction of a rigorous framework for modelling and the rational basis it allows for future improvements in modelling as the major impact of DLM modelling.

Chapter 3

Furnace application

In this chapter an illustrative example is discussed. The example concerns the modelling of corrosion in a furnace which has historically had observation made of 128 spatial points over seven time points. The experts' prior beliefs and belief elicitation are also discussed. Having established a model, diagnostics are used to assess parameter choice and model fit. Due to the current lack of motivation to record accurately observations made, available examples are very limited and much of the motivation for this work is from the judgements and concerns of experts rather than the example itself.

In the first section of this chapter the furnace used in this example is introduced, while in section two the data collected on the furnace since 1982 are discussed. In section three prior beliefs about the system are discussed and a suitable spatial correlation introduced. In section four a range of diagnostic

tests that can be used to check that the model fits the data are introduced and these are applied in section five. In section six correlation within the observation noise is modelled and its effect on the forecast considered.

3.1 Motivating example

Corrosion occurs in many situations and it is often important to know the extent of the corrosion at any particular location. Early identification of corrosion can help in the initiation of preventative measures and accurate long-term prediction is important for forecasting the remaining life of the system. Within the oil industry, corrosion affects furnaces, pipelines, storage tanks, valves, nozzles and many other systems. The model presented is potentially suitable for all these applications.

The example described concerns the modelling of corrosion in a full scale industrial furnace. The furnace is located in the lubricating oil production facility at a large European oil refinery. The lubricating oil production facility takes long residue from the refinery's crude distiller unit, separating and purifying them, via a series of distillation columns, de-asphalting and de-waxing units, into lubricating oils with specific viscosities, such as spindle oil and machine oils. Following initial separation of lighter fractions from the

long residue, the furnace under consideration is used to heat the remaining hydrocarbons prior to further separation. An outline description of the furnace follows:

This furnace has four sides (N,E,S,W) with one coil of 8 pipes on each side. For each coil (identified by the first number in Figure 3.1) the hydrocarbon ascends tube 1, descends tube 2, ..., descends tube 8. In Figure 3.1, entrances to the coils are located in the NW and SE corners (identified by the arrows). Each pipe is 21 feet (6.3m) tall.

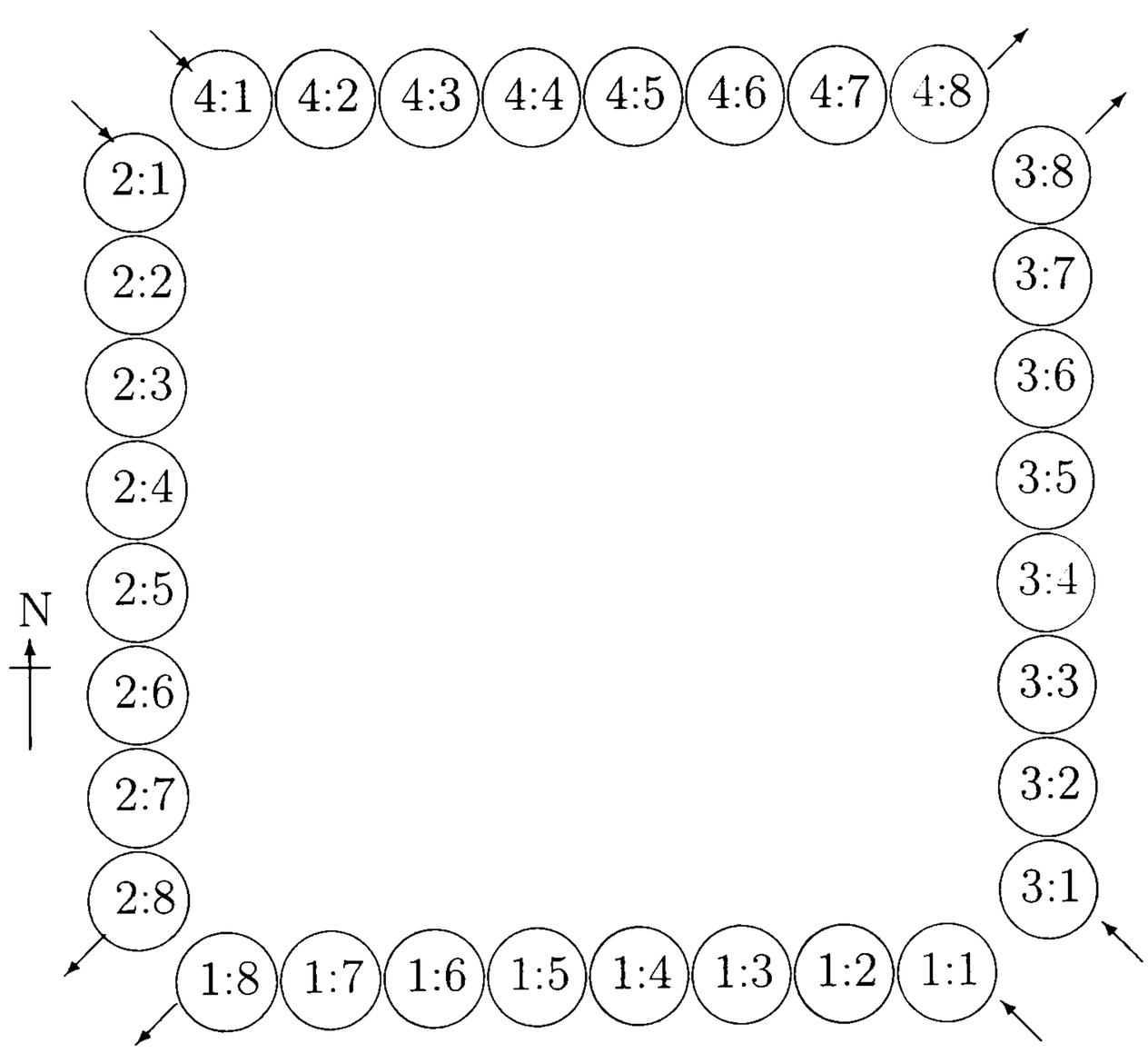


Figure 3.1: Descriptive diagram of the furnace

Inspection of the furnace involves major disruption, if not complete shut-down, of the production process at considerable economic cost. For this reason, inspections tend to coincide with full shut-downs of the production facility for major scheduled maintenance and improvements. Typically, such shut-downs occur once every 3-5 years. The main issues for inspection of the furnace are the detection of corrosion in the (refractories and) heating tubes, and estimation of the rate of corrosion. A related concern is optimising the value of inspection resources, balancing inspection intensity against precision of inferences.

The data are presently modelled with the multiprocess dynamic linear model discussed in Section 2.8. As previously discussed, experts now believe a spatio-temporal model would better represent the process.

3.2 Historical data

In this section the available data for the furnace described in Section 3.1 are discussed. The full data are include in appendix 1.

Vessel wall thickness was measured at heights 3', 7', 12' and 18' on every pipe in August 1983, October 1985, November 1986, March 1988, June 1990, July 1993 and September 1996. Approximately 10% of the data are missing -

mainly at two particular points in time. For the purposes of this example, the furnace is considered to have come on line at the start of 1982 (time 0) so the first wall thickness data were recorded 20 months later. All data are recorded in millimetres to one decimal place except at time 46 when it is rounded to the nearest $\frac{1}{2}$ mm. The data are ultrasonic thickness readings, made using a compression probe and spot readings. It seems likely that some of the observations actually reflect a temporary transient thickness possibly caused by interference when measurements are made. In particular the change in corrosion trend which occurred around 59 months shown in Figure 3.2 cannot be attributed to any particular factor by the experts. Following common practice the transient thickness is used as a surrogate for true wall thickness where this additional error will form a large part of the system uncertainty.

Visual inspection of the data reveals one outlier corresponding to coil 3, tube 2, and height 12' at time 75 when a measurement of 1.4mm was recorded. This is clearly a recording error of some description. In particular, the regular pattern of missing data suggests no inspection was made at this location at this time, were the wall thickness really only 1.4mm the furnace would not have survived another 100 months service and if a location were found with such an alarming wall thickness a major investigation would have

been carried out. This value is now treated as missing.

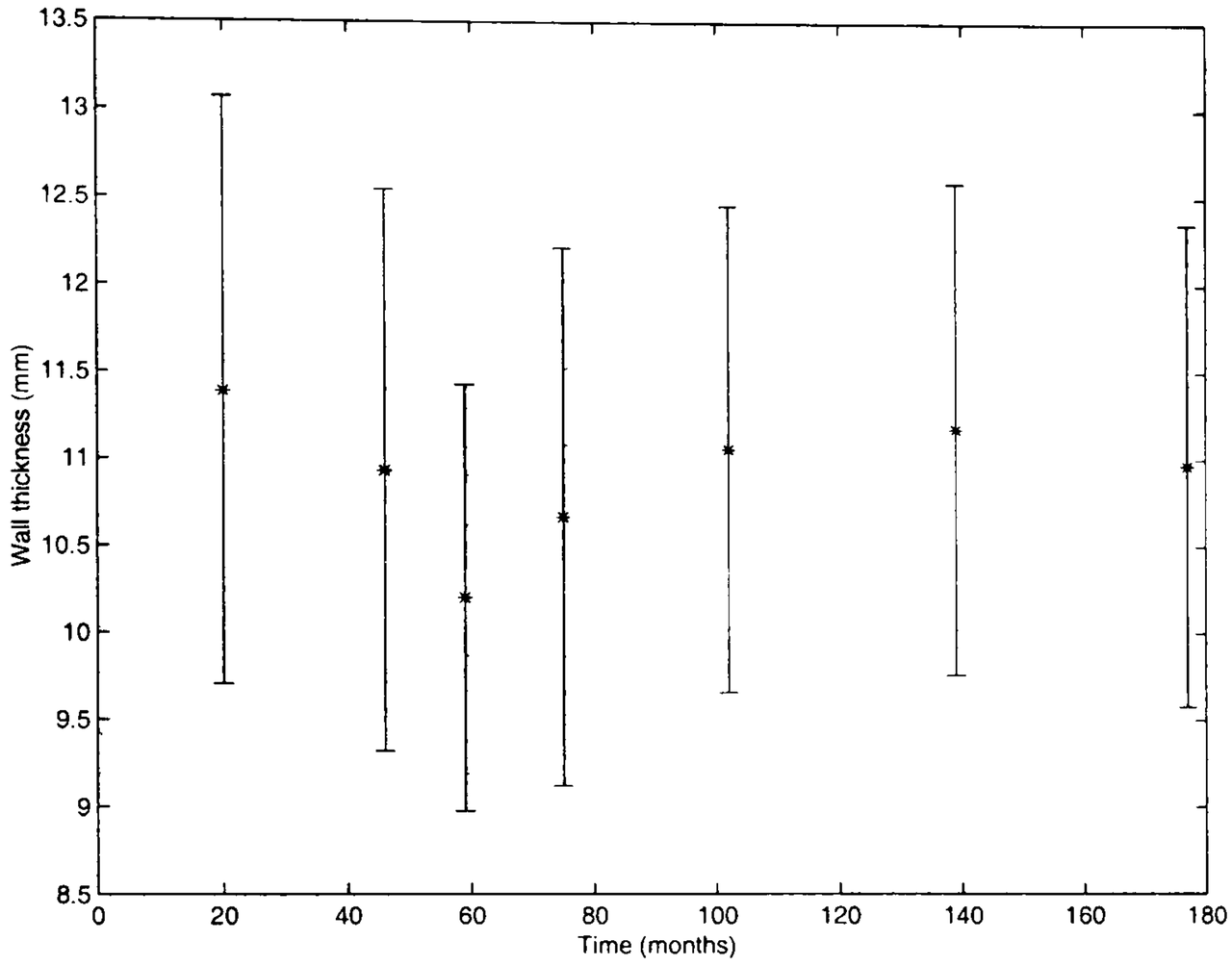


Figure 3.2: Mean wall thickness ± 3 data s.d.

In Figure 3.2 the mean wall thickness at each time is plotted with whiskers showing three data deviations. Standard deviations of the means are more difficult to calculate because the observations are correlated. The change in corrosion trend which occurred around 59 months shown in Figure 3.2 cannot be attributed to any particular factor by the experts. However, experts state that this pattern is quite unusual and that the underlying wall thickness of a furnace cannot increase so it seems likely that there was some calibration error for that inspection.

Initial mean wall thickness is 11.39mm and final mean wall thickness is

10.96mm.

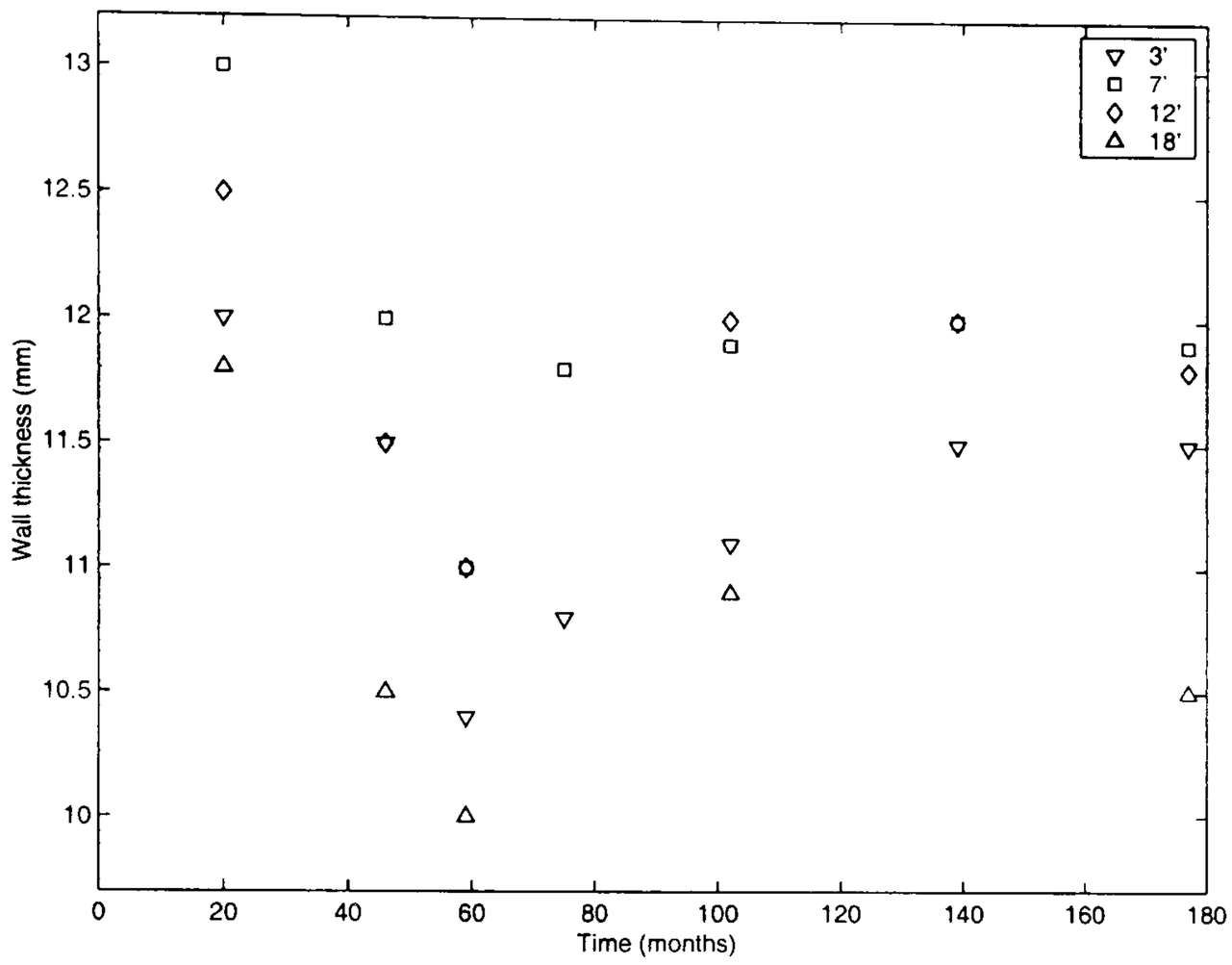


Figure 3.3: Wall thickness for pipe 1

It can be seen in Figure 3.3 that the wall thicknesses at different locations are correlated. In this case particular concerns may be held about the pipe above height 18.

Figure 3.4 shows every measurement at height 7` in coil 1. Wall thicknesses for different pipes generally increase and decrease simultaneously, for example, at time 20 pipes ordered by wall thickness are 1, 2, 3 & 7, 8, 4 & 6 and 5, and at the last measurement they are ordered 1, 2, 4 & 7, 3, 5, 6 and 8, showing less rearrangement than might be expected. As expected, this supports the beliefs of the system experts that remaining wall thicknesses

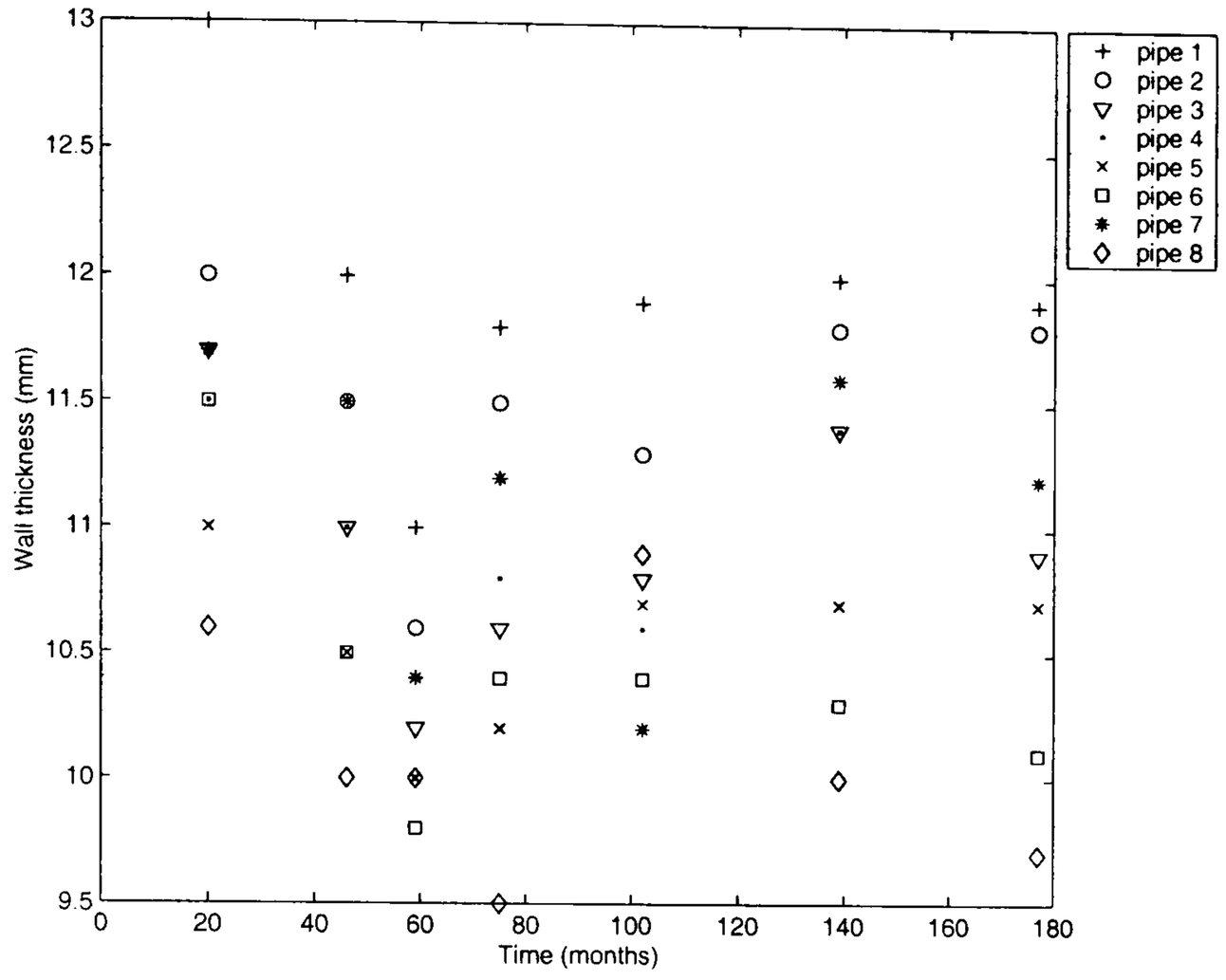


Figure 3.4: Wall thickness for height 7' across coil 1

are correlated across pipes. For this particular data set correlations between adjacent pipes are of the order of 0.7.

Experts believe the corrosion will be about 0.3mm per annum. Initial corrosion is in the range 0.2mm per annum with a maximum mean corrosion (preceding the inspection at time 59) of about 0.7mm per annum. Net corrosion over the whole period is approximately 0.025mm per annum. The standard deviation of the mean corrosion rates is about 0.5mm, but, within the model, the variance of the mean corrosion rate can be set to be lower than this value, since experts believe the corrosion rate is positive and also believe it to be reasonably constant.

Experts believe current corrosion rate is an unbiased estimator of future corrosion rate, if conditions remain constant, and it is reasonable to assume that the system will deteriorate continuously while the furnace is in use, so the corrosion will be modelled with a locally linear trend.

3.3 Prior beliefs

In this section the process used for choosing the parameters for the spatio-temporal model used in this example are discussed.

3.3.1 General considerations

Shell (the industrial collaborator for this project) consider it good practice to base inspection regimes on experience with several vessels of similar nature which have run under similar operating conditions for substantial periods of time (see (Terpstra, Ritchie, and den Heijer 2002) for further details).

In general, model fitting based on careful analysis of data collected on such similar systems, in combination with careful elicitation of expert beliefs is recommended; the general process of elicitation for complex spatial systems is discussed in detail in (Craig, Goldstein, Seheult, and Smith 1998). This may raise challenging statistical problems but is unavoidable if reliable and efficient inspection is to be carried out.

3.3.2 Parameter choice for the example

For the purpose of illustration in the current example, parameter values considered plausible by expert assessment were chosen to be consistent with the observed data from times one to five. The model is then assessed against subsequent times. In practical applications, data sets for other similar furnaces should be exploited, to improve the assessment of the covariance structure.

The manufacturing specification of the vessel gives an initial wall thickness of 11.5 mm which is consistent with the observed means. Initial corrosion is set to be $\frac{0.3}{12}$ per month. System experts initially expressed very strong beliefs that the observation errors are uncorrelated (subsequent analysis persuaded them that their confidence might have been misplaced). The observed data for this particular example especially in relation to the 59^{th} month appear to go against this belief. Modelling correlation within the observational errors may allow more accurate representation of this particular data set. In Section 3.6 the effect of having correlated observation noise on the inferences made is further discussed. However, introducing correlations across observation errors does not appear (see Section 3.6) to make much difference to the forecasts. For the purposes of this example, we shall proceed by assuming that the observation errors are indeed uncorrelated.

Observation noise can be estimated relatively accurately using repeated measurements at one location while system variance is often harder to estimate. Following consultation with experts, the observation variance (V_t) is believed to be about 0.35^2.

Current practice suggests that the designer of the existing inspection regime believed it sufficient to measure 4 locations on each pipe. Assuming the original inspection scheme to be reasonably satisfactory, this may be used as a crude diagnostic check: that is, the analysis should agree that about 4 points per pipe is an appropriate number of inspections.

A covariance matrix is built for the system levels (W_θ) and then the same structure used (with a re-scaled variance) in the covariance of the system slope (W_β), and in specifying prior variance ($C_{\theta 0}$ and $C_{\beta 0}$). If more detailed belief structures were available for W_β and $C_{\theta 0}$ and $C_{\beta 0}$, these would be used instead, but this would not effect the general methodology described in this thesis. The system levels are assumed uncorrelated with the system slopes, prior levels and prior slopes. Similarly system slopes are assumed uncorrelated with prior levels and prior slopes which are in turn also assumed uncorrelated.

The system update variance ($W_{\theta s}$) is constructed from a sum of suitable

covariances, as in Section 2.9. Thus the update covariance between any two points on the furnace may be described using:

$$cov(\omega_{its}, \omega_{its'}) = \Upsilon_i \left(Ae^{-\alpha|h_s - h_{s'}|^{\Phi_i}} + Be^{-\gamma|p_s - p_{s'}|^{\Phi_i}} \right.$$

$$\left. + Ce^{-\delta|c_s - c_{s'}|^{\Phi_i}} + De^{-\epsilon|d_s - d_{s'}|^{\Phi_i}} \right) \tag{3.1}$$

where h_j = height, p_j = pipe, c_j = coil, d_j = inside pipe distance around the coil for location j, and all other values are suitable constants. Setting $\Phi_i = 1$ is considered by experts to give a suitable model.

The slope update uncertainty $W_{\beta s}$ and the prior slope uncertainty $C_{\beta 0 s}$ are both set to be a tenth of the uncertainty assigned to level update uncertainty $W_{\theta s}$ and prior level uncertainty $C_{\theta 0 s}$ respectively, however the correlation structure is left unchanged.

Following consultation with experts the following values were considered plausible for this example. $\alpha = \gamma = \delta = \epsilon = 0.003$ and in the overall update variance parameter $\Upsilon_\theta = 0.00001$ (W_θ) and $\Upsilon_\beta = 0.000001$ (W_β) and in the overall initial variance parameter $\Upsilon_\theta = 0.05$ and $\Upsilon_\beta = 0.005$ ($C_{\theta 0}$ and $C_{\beta 0}$).

Eliciting beliefs about spatial covariances is difficult. Experts agreed with the following statement: 'correlation due to 2 heights differing by k feet' will be 2 times stronger than 'correlation due to 2 pipes differing by k pipes', and

4 times stronger than 'correlation due to 2 coils differing by k coils' (coil 4 is next to coil 1), but will have the same correlation as 'correlation due to 2 points as hydrocarbon flows differing by k feet'. This statement is interpreted to give values for $A = 1$, $B = 0.5$, $C = 0.25$ and $D = 1$ in equation 3.1. This gives a minimum update correlation between any two points of the system over one month of approximately 0.66, and a typical correlation between top and bottom of a pipe of 0.96.

A ceiling for the update slope variance can be estimated using the requirement on the update slope to remain negative with a high probability. Initial corrosion rate is ≈ 0.02. If at least 3 standard deviations are required between this value and 0, the standard deviation must be ≤ 0.005 and the variance $\leq 2.5 \times 10^{-5}$.

3.4 Diagnostics

Diagnostics may be used to check if the data observed is at odds with the model or the prior beliefs held. In this section three general diagnostics which are applied in Section 3.5 are discussed.

3.4.1 Introduction

To calculate a simple diagnostic for a single observation x with expectation $E(X)$ and variance $var(X)$, the first step is to calculate the standardised observation $S(x)$ and the discrepancy $Dis(x)$:

$$S(x) = \frac{x - E(X)}{\sqrt{var(X)}}$$

$$Dis(x) = (S(x))^2 = \frac{(x - E(X))^2}{var(X)}$$

Since $E(Dis(x)) = 1$, a value far from 1 suggests either that $E(X)$ or $var(X)$ might be misspecified or that there might be a problem with the observation. For example, if x is normally distributed 95% of the standardised observations $S(x)$ would be expected to lie between approximately -2 and 2. With sufficient data, many discrepancies can be calculated and unusual discrepancies can be identified using for instance, the three sigma rule (see (Pukelsheim 1994)).

3.4.2 Successive differencing

Successive differencing can be used on models with mean zero error terms, for example, the polynomial growth model. This is simplest in the constant model case (no β equation in equation 2.1), where only a single difference is

needed.

$$\begin{aligned} \theta_t &= \theta_{t-1} + \omega_t, & \omega_t &\sim < 0, W > \\ y_t &= \theta_t + \nu_t, & \nu_t &\sim < 0. V > \end{aligned} \tag{3.2}$$

Then $\Delta(y_t) = y_t - y_{t-1} = \theta_t + \nu_t - \theta_{t-1} - \nu_{t-1} = \omega_t + \nu_t - \nu_{t-1}$ and $var(\Delta y_t) = W + 2V$. This may be compared to the mean of the squared first order data differences $mean_t(\Delta(y_t)^2)$.

For a polynomial model of order n it is necessary to difference $n+1$ times. The data difference for the linear growth model may be explicitly considered where

$$\begin{aligned} \theta_t &= \theta_{t-1} + \beta_{t-1} + \omega_{\theta t}, & \omega_{\theta t} &\sim < 0, W_\theta > \\ \beta_t &= \beta_{t-1} + \omega_{\beta t}, & \omega_{\beta t} &\sim < 0, W_\beta > \\ y_t &= \theta_t + \nu_t, & \nu_t &\sim < 0, V > \end{aligned} \tag{3.3}$$

The calculation involved for the univariate and the multivariate data difference are identical. In the multivariate case y_t, θ_t, β_t, $\omega_{\theta t}$, $\omega_{\beta t}$ and ν_t are vectors and their variances W_θ, W_β, $W_{\theta\beta}$ and V are matrices, where $W_{\theta\beta} = cov(\omega_{\theta t}, \omega_{\beta t})$ (which has been in general assumed $= 0$).

$$\Delta(\Delta(y_t)) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

$$\begin{aligned}
&= \theta_t + \nu_t - 2\theta_{t-1} - 2\nu_{t-1} + \theta_{t-2} + \nu_{t-2} \\
&= \theta_{t-1} + \beta_{t-1} + \omega_{\theta t} + \nu_t - 2\theta_{t-1} - 2\nu_{t-1} + \theta_{t-2} + \nu_{t-2} \\
&= -\theta_{t-2} - \beta_{t-2} - \omega_{\theta t-1} + \beta_{t-2} + \omega_{\beta t-1} + \omega_{\theta t} + \nu_t - 2\nu_{t-1} + \theta_{t-2} + \nu_{t-2} \\
&= -\omega_{\theta t-1} + \omega_{\beta t-1} + \omega_{\theta t} + \nu_t - 2\nu_{t-1} + \nu_{t-2}
\end{aligned}$$

hence

$$\begin{aligned}
var(\Delta(\Delta(y_t))) &= var(\omega_{\beta t-1} - \omega_{\theta t-1}) + var(\omega_{\theta t}) + var(\nu_t) + 4var(\nu_{t-1}) + var(\nu_{t-2}) \\
&= W_\beta + 2W_\theta - 2W_{\theta\beta} + 6V
\end{aligned}$$

This variance may be compared with the mean of the squared second differences of the data $mean_t(\Delta(\Delta(y_t))^2)$. For more complex models where other sources of additional independent variation are known to exist (e.g a seasonal model) this value will be a lower bound.

Missing data

If one spatial data point is missing, one simple approach in the multivariate problem is to completely ignore the relevant rows and columns of the variance matrix. However, if all the data for a particular time point are missing the problem cannot be so easily resolved.

Suppose data are available at times t, $t+k$ and $t+n$, where $k < n$. The double difference is then:

$$
\begin{aligned}
(y_{t+n} - y_{t+k}) - (y_{t+k} - y_t) &= y_{t+n} - 2y_{t+k} + y_t \\
&= \theta_{t+n} + \nu_{t+n} - 2\theta_{t+k} - 2\nu_{t+k} + \theta_t + \nu_t \\
&= \theta_{t+n-1} + \beta_{t+n-1} + \omega_{\theta t+n} + \nu_{t+n} \\
&\quad -2\theta_{t+k-1} - 2\beta_{t+k-1} - 2\omega_{\theta t+k} - 2\nu_{t+k} + \theta_t + \nu_t \\
&= \theta_t + \sum_{i=1}^{n} [\beta_{t+i-1} + \omega_{\theta t+i}] + \nu_{t+n} \\
&\quad -2\theta_t - 2\sum_{i=1}^{k} [\beta_{t+i-1} + \omega_{\theta t+i}] - 2\nu_{t+k} + \theta_t + \nu_t
\end{aligned}
$$

Writing as $\beta_{t+k} = \beta_t + \sum_{i=1}^{k} \omega_{\beta t+i}$, gives

$$\sum_{i=0}^{n-1}\beta_{t+i} - 2\sum_{i=0}^{k-1}\beta_{t+i} \quad = \quad (n-2k)\beta_t + \sum_{i=1}^{n-1}(n-i)\omega_{\beta t+i} - 2\sum_{i=1}^{k-1}(k-i)\omega_{\beta t+i}$$

So

$$\begin{aligned}
&(y_{t+n} - y_{t+k}) - (y_{t+k} - y_t) \\
&\quad = \quad \theta_t + \sum_{i=0}^{n-1}\omega_{\theta t+i+1} + \nu_{t+n} - 2\theta_t - 2\sum_{i=0}^{k-1}\omega_{\theta t+i+1} - 2\nu_{t+k} \\
&\qquad +\theta_t + \nu_t + (n-2k)\beta_t + \sum_{i=1}^{n-1}(n-i)\omega_{\beta t+i} - 2\sum_{i=1}^{k-1}(k-i)\omega_{\beta t+i}
\end{aligned}$$

Hence the expectation is now $(n-2k)\beta_t$ which, if the points are evenly spaced, is identically zero as expected. The variance of this quantity may then be calculated:

$$
\begin{aligned}
& var\left((y_{t+n} - y_{t+k}) - (y_{t+k} - y_t)\right) \\
= \; & 6V + \left(var\left[\sum_{i=1}^{n} (\omega_{\theta t+i} + (n-i)\omega_{\beta t+i}) \right.\right. \\
& \left.\left. -2\sum_{i=1}^{k} (\omega_{\theta t+i} + (k-i)\omega_{\beta t+i}) + (n-2k)\beta_t \right] \right) \\
= \; & 6V + (n-2k)^2 var(\beta_t) + \left(var\left[-\sum_{i=1}^{k-1} (\omega_{\theta t+i} - (n-2k+i)\omega_{\beta t+i}) \right.\right. \\
& \left.\left. +(n-k)\omega_{\beta t+k} - \omega_{\theta t+k} + \sum_{i=k+1}^{n-1} (\omega_{\theta t+i} + (n-i)\omega_{\beta t+i}) + \omega_{\theta t+n} \right] \right)
\end{aligned}
$$

but

$$
\begin{aligned}
& var\left(-\sum_{i=1}^{k-1} (\omega_{\theta t+i} - (n-2k+i)\omega_{\beta t+i}) \right. \\
& \left. +(n-k)\omega_{\beta t+k} - \omega_{\theta t+k} + \sum_{i=k+1}^{n-1} (\omega_{\theta t+i} + (n-i)\omega_{\beta t+i}) \right) \\
= \; & W_\theta(k-1) + W_\beta \sum_{i=1}^{k-1} (n-2k+i)^2 + W_\beta(n-k)^2 + W_\theta \\
& +(n-k-1)^2 W_\theta + \sum_{i=k+1}^{n-1} (n-i)^2 W_\beta + \sum_{i=1}^{k-1} (n-2k+i) W_{\theta\beta} \\
& + \sum_{i=k+1}^{n-1} (n-i) W_{\theta\beta}
\end{aligned}
$$

and since

$$\sum_{i=1}^{k-1}(n-2k+i)W_{\theta\beta}+\sum_{i=k+1}^{n-1}(n-i)W_{\theta\beta}$$

$$= W_{\theta\beta}(\frac{(k-1)(k+2(n-2k))+(n-k-1)(n-k)}{2})$$

it can be seen that

$$var\left((y_{t+n}-y_{t+k})-(y_{t+k}-y_t)\right)$$

$$= 6V+(n-2k)^2var(\beta_t)+W_\theta(k-1)$$

$$+W_\beta(\frac{n^3}{3}-\frac{n^2}{2}-2nk^2+\frac{n}{6}+2nk-2k^2+2k^3)+W_\theta$$

$$+W_\theta(n-k-1)^2+W_{\theta\beta}(\frac{(k-1)(k+2(n-2k))+(n-k-1)(n-k)}{2})$$

$$= 6V+(n-2k)^2var(\beta_t)+W_\theta(k+(n-k-1)^2)+$$

$$W_{\theta\beta}(\frac{(k-1)(k+2(n-2k))+(n-k-1)(n-k)}{2}) \qquad (3.4)$$

$$+W_\beta(\frac{n^3}{3}-\frac{n^2}{2}-2nk^2+\frac{n}{6}+2nk-2k^2+2k^3)$$

If the data are evenly spaced in time the mean of the differences may be calculated and it should lie closer to the expected value than the single values. If the data are unevenly spaced the differences must be corrected to remove the effect of the trend term $(n-2k)\beta_t$. The mean of the squared corrected

data differences may then be compared to the expected value calculated using the weights calculated above.

3.4.3 Univariate forecast error

The standardised prediction error at time t is

$$s_t = \frac{y_t - E[y_t|D_{t-1}]}{\sqrt{var(y_t|D_{t-1})}} = \frac{y_t - f_t}{\sqrt{Q_t}} \tag{3.5}$$

A histogram of the standardised prediction errors should then look similar to the normal curve if the errors are Gaussian. Heavy tails suggest the prior variances are too small. For small sample sizes alternative distributions should be considered in place of the normal distribution. A probability plot can be used in addition to the histogram (see (Chambers, Cleveland, Kleiner, and Tukey 1983)).

3.4.4 Multivariate forecast error

The multivariate forecast may also be standardised for comparison with its expected value. When dealing with multivariate data the multivariate nature of the problem can of course be ignored and the errors taken for each point and considered independently but this gives no diagnostic assessment of the correlations.

Consider first the one step ahead multivariate forecast.

$$Dis(y_t) = \left((y_t - E(y_t|D_{t-1}))'var(y_t|D_{t-1})^{-1}(y_t - E(y_t|D_{t-1}))\right)$$

If $var(y_t|D_{t-1})$ is not full rank then a generalised inverse may be used to invert it, however, we consider only the full rank case.

The expected value of this discrepancy given D_{t-1} may be calculated by taking the trace (of this scalar) and using $TraceAB = TraceBA$

$$\begin{aligned}
&E\left[\left((y_t - E(y_t|D_{t-1}))'var(y_t|D_{t-1})^{-1}(y_t - E(y_t|D_{t-1}))\right)|D_{t-1}\right] \\
&= E\left[trace\left((y_t - E(y_t|D_{t-1}))'var(y_t|D_{t-1})^{-1}(y_t - E(y_t|D_{t-1}))\right)|D_{t-1}\right] \\
&= trace\left(var(y_t|D_{t-1})^{-1}E\left[(y_t - E(y_t|D_{t-1}))(y_t - E(y_t|D_{t-1}))'|D_{t-1}\right]\right) \\
&= \left[trace\left(var(y_t|D_{t-1})^{-1}var(y_t|D_{t-1})\right)\right] = r
\end{aligned}$$

the rank of $var(y_t)$.

If $((y_t - E(y_t|D_{t-1}))'var(y_t|D_{t-1})^{-1}(y_t - E(y_t|D_{t-1}))) \gg r$ this indicates the observation variances or the system variances may be too small. The discrepancy may be normalised as $Dr(y_t) = \frac{Dis(y_t)}{rank(var(y_t))}$ to have expectation 1.

Forecasts based only on the prior beliefs may also be considered. Defining $\underline{Y} = (\underline{y_1'}, \ldots, \underline{y_t'})'$ and defining

$$size(Y) = (Y - E(Y|D_0))'var(Y|D_0)^{-1}(Y - E(Y|D_0)) \qquad (3.6)$$

and then following earlier calculations $E\left(size(Y)|D_0\right) = rank(var(Y|D_0))$ if $var(Y|D_0)$ is full rank. Observation forecasts and their variance are calculated using equations 2.6 and 2.7.

In the multivariate case, vector Y is composed from a set of vectors $(\underline{y_1'}, \ldots, \underline{y_t'})'$. This leads to a very large variance matrix which in turn will often have very small eigenvalues. Even a small unexpected change (due to the data) in the direction of a small eigenvalue will often equate to a very large number of standard deviations. Often these unexpected changes will be due to a combination of rounding error and noise rather than a major outlier.

If $size(Y) \gg rank(var(Y|D_0))$ the prior variances may be too low.

3.4.5 Diagnostic variance

Diagnostics are most useful if diagnostic variances are evaluated as well as diagnostic expectations. Variances may be calculated directly or via simulation. Direct calculation involves fourth order expectations. Often fourth order expectations will either require further direct specification or will be derived from the probability distributions of these quantities.

Variance of the forecast error

Given that the univariate discrepancy $\frac{(X-E(X))^2}{var(X)}$ (see section 3.4.3) has been calculated with expectation one, the variance $var\left(\frac{(X-E(X))^2}{var(X)}\right)$ may be used to calibrate how far the calculated discrepancy is from its expected value. Supposing $X \sim N(\mu, \sigma^2)$ then normal distribution third and fourth moments may be used to evaluate this variance by conditioning on σ^2. The kurtosis of the curve may be used when evaluating this quantity. The kurtosis measures the degree of peakedness of a probability density function. In particular for the normal distribution $\frac{E(X^4)}{\sigma^4} = 3$. Refocusing around zero, that is using $X \sim N(0, \sigma^2)$

$$var\left(\frac{X^2}{\sigma^2}\middle|\ \sigma^2\right) = \frac{E(X^4|\sigma^2)}{\sigma^4} - \left(\frac{E(X^2|\sigma^2)}{\sigma^2}\right)^2 = 2$$

Similar calculations are required for the multivariate case. However, alternative approaches such as using a general ratio of 2^{nd} to 4^{th} moments and an algebraic computer package may practically be necessary.

Data difference

The variance of the data difference for a linear trend model ($E((\Delta^2 y_t - E(\Delta^2 y_t))^2) = var(\Delta^2 y_t) = W_\beta + 2W_\theta - 2W_{\theta\beta} + 6V$) is now considered.

To calculate this, expectations and variances for W_β, $2W_\theta$, $2W_{\theta\beta}$ and $6V$

are also needed, unless it is assumed that $\Delta^2 y_t$ is normal. If $\Delta^2 y_t \sim N(\mu, \sigma)$ then

$$var(((\Delta^2 y_t) - E(\Delta^2 y))^2 | \sigma^2) = 2\Sigma^4 \tag{3.7}$$

from the previous calculation. However, accurately eliciting expert opinion about σ^2 may be challenging.

An alternative approach is to simulate data repeatedly consistent with the prior belief structure (see Section 2.7) and apply the diagnostic to this simulated data. The variance of the diagnostic can then be found and used to estimate the number of standardised discrepancies between the data diagnostic and its expected value. Inaccuracies in the error structure may give rise to further inaccuracies when using this approach.

3.4.6 Alternative estimates for uncertainty

A potential approach when exact calculations are difficult is the three sigma rule (see (Pukelsheim 1994)), where for a unimodal distribution

$$p(|\frac{y - E(y)}{\sqrt{var(y)}}| \leq 3) \geq 0.95$$

An alternative which may be used to give a more general credible interval when the distribution is not unimodal is Tchebychev's inequality: $P(|X -$

$E(X)| \geq a) \leq \frac{var(X)}{a^2}$. Rearranging:

$$P(-a \leq X - E(X) \leq a) \geq 1 - \frac{var(X)}{a^2}$$

If Y_t is formed from independent Gaussian quantities, $Dis(Y_t)$ has expectation r and variance $2r$, so letting $X = \frac{Dis(Y_t)-r}{\sqrt{2r}}$ gives $E(X) = 0$ and $var(X) = 1$ so:

$$P(-a \leq \frac{Dis(Y_t) - r}{\sqrt{2r}} \leq a) \geq 1 - \frac{1}{a^2}$$

$$P(-\sqrt{2ra^2} + r \leq Dis(Y_t) \leq \sqrt{2ra^2} + r) \geq 1 - \frac{1}{a^2}$$

Setting $a = \sqrt{20}$ to give a 95% conservative credible interval gives:

$$P(-\sqrt{40r} + r \leq Dis(Y_t) \leq \sqrt{40r} + r) \geq 0.95$$

or equivalently:

$$P(-\sqrt{\frac{40}{r}} + 1 \leq Dr(Y_t) \leq \sqrt{\frac{40}{r}} + 1) \geq 0.95$$

which shows that confidence is gained as the rank increases. Substituting $r = 1$ demonstrates the conservativeness of these bounds.

3.4.7 Further diagnostics

Many more diagnostic tests are possible. In particular, each of the diagnostics considered may be repeated from a posterior view point by considering

adjusted expectation and variance in place of the data variance.

The size ratio for the adjustment of B by D may also be calculated: $Sr_D(B) = \frac{Size_D(B)}{E(Size_D(B))}$, where

$$Size_D(B) = (E_D(B) - E(B))'(var(E_D(B)))^{-1}(E_D(B) - E(B))$$

with expectation $E(Sr_D(B)) = 1$. Small values may suggest exaggerated prior uncertainty, while large values suggest a surprising change in beliefs.

Further exploitation of the eigen-structure of the variance matrices is also possible. The orthogonality of the eigenvectors allows uncorrelated effects to be considered. Using the orthogonality of the eigenvectors and their expected effects (the eigenvalues), uncorrelated linear combinations of the data which are unusual may be identified.

3.4.8 Estimating variances from the data

By considering a range of successive differences estimates can be made for V, W_θ, W_β and $W_{\theta\beta}$. In most cases it will not be appropriate to calculate these four unknown values from exactly four data differences. A better solution would be update beliefs about V and W in a Bayesian setting using for example the Bayes linear update equations (equations 2.2 and 2.3). For more details of such updates see (Wilkinson 1997).

In practice estimates for V and W ($\hat{V}$ and $\hat{W}$) must be made over many time updates. However, over long time series, it may not be sensible to assume that V and W are constant in time. Further complications may occur if differences cannot be considered independent because they were partially based on the same observation.

Further simplifications are possible to make these calculations more tractable. In particular, because of the short time series and the missing data only a point estimate is calculated, that is only $W_{\theta s}$ is estimated from the data. Assuming the estimate for V to be accurate, if a fixed ratio is assumed between W_θ and W_β, then W_θ and W_β can be more easily estimated directly from the data. In the example a ratio of 1:10 is taken (see Section 3.3.2). Using a similar ratio between the prior slope uncertainty and the update uncertainty (1:500) the variance of the expectation of a second order data difference $(n-k)^2 var(\beta_t)$ (see equation 3.4) becomes

$$(n-k)^2(tW_\beta + C_{\beta 0}) = (n-k)^2 tW_\theta(\frac{1}{10} + 500) \tag{3.8}$$

Substituting equation 3.8, $W_\beta = \frac{W_\theta}{10}$ and $W_{\theta\beta} = 0$ into equation 3.4.

$$
\begin{aligned}
& var(y_{t+n} - 2y_{t+k} + y_t) \\
= \; & 6V + (n-k)^2 var(\beta_t) \\
& + W_\theta \left((n + (n-k-1)^2) + \frac{n^3}{30} - \frac{n^2}{20} - \frac{nk^2}{5} + \frac{n}{60} + \frac{nk - k^2 + k^3}{5} \right) \\
= \; & 0.735 + W_\theta ((n + (n-k-1)^2) + \frac{n^3}{30} - \frac{n^2}{20} \\
& - \frac{nk^2}{5} + \frac{n}{60} + \frac{nk - k^2 + k^3}{5} + (n-k)t(\frac{1}{10} + 500))
\end{aligned}
$$

Hence the expected double data difference is expressed with a single unknown value W_θ which may then be estimated.

t	k	n	$mean_s(\Delta(\Delta(y_{ts})) - (n-2k)\beta_t)^2$	Estimate for W (10^{-5})
20	46	59	2.1733	2.5035
20	46	102	1.9150	0.1204
20	46	177	8.0251	0.0916
20	59	102	3.4355	0.3664
20	59	177	12.2417	0.1721
20	102	177	1.0720	0.0374
46	59	102	2.0844	0.2212
46	59	177	9.3425	0.1307
46	102	177	1.7258	0.0599
59	102	177	3.3969	0.1177

Table 3.1: Double data difference and estimate for W for different t, k, and n

The full range of combinations of n, k and t are considered in Table 3.1 for the full data sets observed. Most of the estimates for $W_{\theta s}$ are smaller than the prior value specified (10^{-5}). However, errors may be due to problems with

prior slope and level uncertainty specification or with the corrosion rate.

3.5 Validation of model

In this section the model fit is checked, some of the diagnostic tests described in the previous section are applied as is a sensitivity analysis to changes in belief.

3.5.1 A simple check

A simple check that the model is fitting the data can be made by plotting the system mean for each time and location as well as the data and 3 standard deviation confidence bounds.

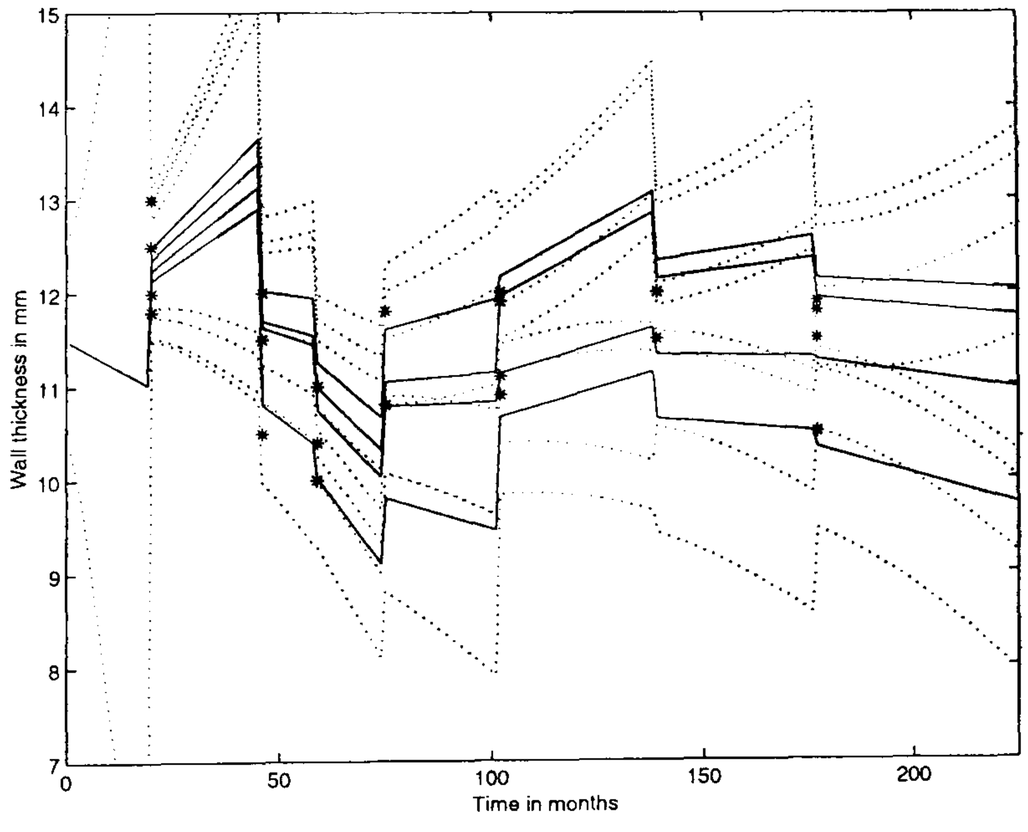


Figure 3.5: Data and forecasts at the measurement points of pipe 1, coil 1

Figure 3.5 is fairly typical of the plots obtained by doing this. '*' indicates

an observation and a solid line the one step ahead point forecast with its three standard deviation credible interval indicated by the dotted lines. Most data (22 out of 24) fall within the 3 standard deviation credible intervals specified for the pipe. This again is typical of all 32 pipes.

3.5.2 Application of diagnostics

The three diagnostic tests are applied at the time of the fifth inspection having fitted the model on the first five data points.

The differences between the expected data difference $var(y_{t+n} - 2(y_{t+k}) + y_t)$ and the actual corrected data difference $mean_t\{[\Delta(\Delta(y_t)) - (n-2k)\beta_t]^2\}$ (as described in Section 3.4.2) are in the range [-1.5, 3.1] with a mean of 1.6.

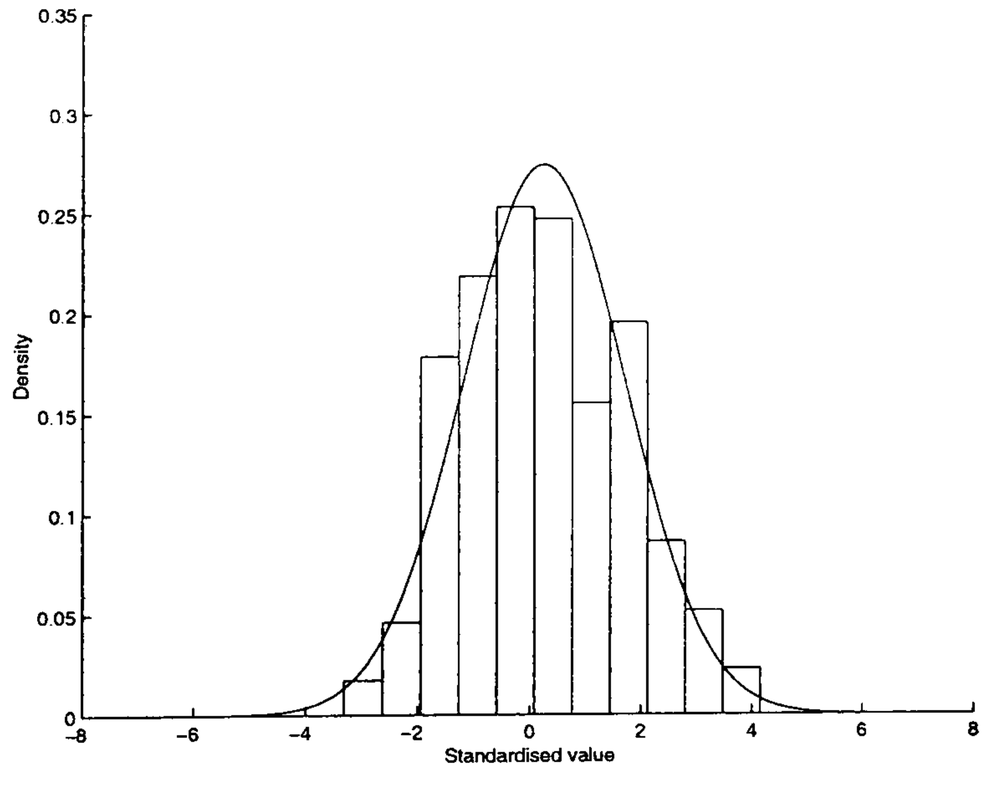


Figure 3.6: Univariate forecast error density at the time of the fifth inspection

Figure 3.6 shows the univariate forecast error density (see equation 3.5

and description in Section 3.4.3) which shows no cause for concern. The light tails indicate model variances may be slightly too large. The bias of the mean might be attributed to the correlation between the points effectively reducing the sample size.

The model correlation structure is exploited to remove the correlation between the points in the multivariate forecast error (see equation 3.6 and description in Section 3.4.4), where there is an expected value of 640 and an observed value of 780 suggesting prior variances may be slightly too small.

The same diagnostics can be applied following the final inspection in September 1996. Data differencing has range [-2.6, 2.0] with mean 0.0 which does not indicate any cause for concern. Figure 3.7 shows the univariate forecast error which follows the normal curve reasonably closely. The multivariate forecast error has an observed value of approximately 1061 with an expected value of 1024 which despite not having a variance estimate does not appear to indicate any cause for concern.

3.5.3 Sensitivity analysis

The sensitivity of the model to changes in prior beliefs and update variances can be important. The extent to which the data can distinguish between changes in beliefs may be examined by changing these parameters and then

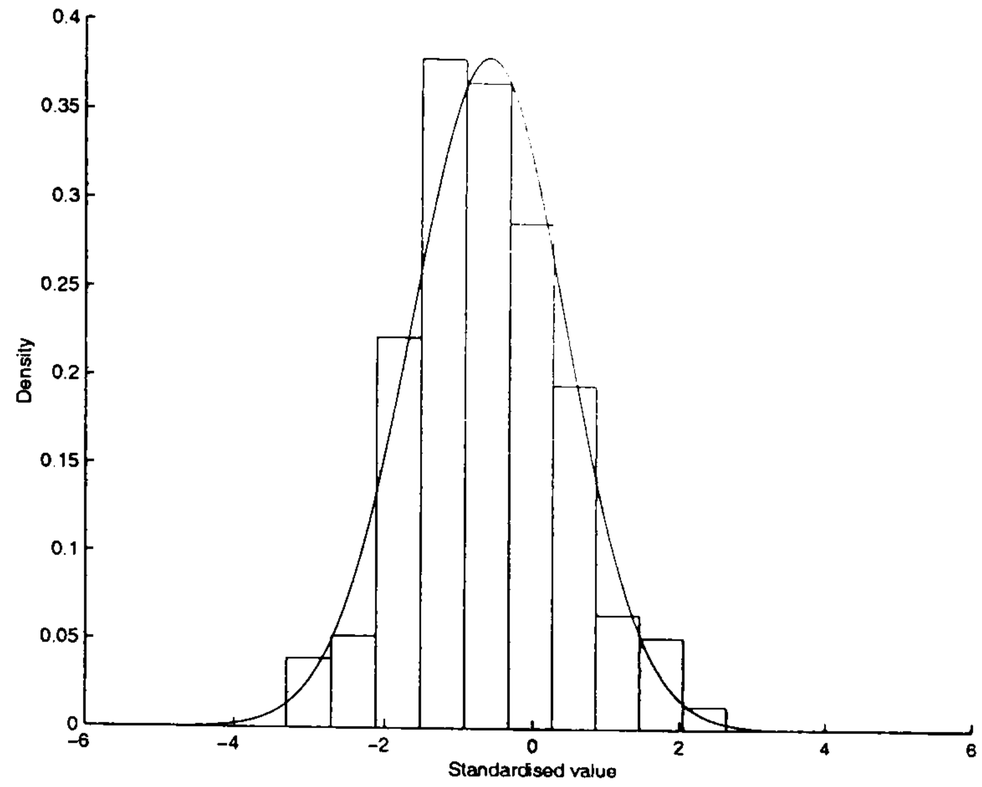


Figure 3.7: One step ahead forecast error density time 7

re-running the diagnostics. It is difficult to makes changes of equivalent magnitude for the three different aspects considered. Working relative to the original values a 40% change is considered in the observation variance and a three fold increase and decrease for the system variance and correlation. See Table 3.2.

Description	Data differencing	Multivariate forecast error (Expected=1024)
original	[-2.6, 2.0], mean = 0.0	1061.2
↑ (+40%) observation variance V	[-2.5,2.0], mean = 0.0	562.3
↓ (-40%) observation variance V	[-2.9,2.0], mean = 0.0	1928.3
↑ (×3) system variance W_θ	[-2.8, 2.2], mean = 0.2	1744.1
↓ (/3) system variance W_θ	[-2.4, 2.2], mean = 0.2	965.8
↑ (×3) system correlation α, β, γ & δ	[-2.6, 2.0], mean = 0.0	1012.2
↓ (/3) system correlation α, β, γ & δ	[-2.6, 2.0], mean = 0.0	1106.4

Table 3.2: Sensitivity to model parameter change

Univariate forecast error histograms give no cause for concern. Data differencing is fairly consistent and does not raise any diagnostic concerns for the changes being considered. Looking at the multivariate forecast error, the observation variance is the most important of the variables to elicit accurately - fortunately it can also be estimated relatively easily using repeated measurements. System correlation is probably the hardest component to estimate and to elicit beliefs about. Unfortunately, it is also relatively difficult to detect an effect with any of the three diagnostics considered because it is not particularly sensitive to these types of diagnostic check. It is therefore important to carry out a further sensitivity analysis of the chosen inspection regime to changes in the system correlation to allow a further check on possible problems that may be caused by under or over estimating correlations.

3.6 The effect of correlated noise

In this section the effect of correlation noise on the predictions made is further discussed. The noise is considered to have two components: a fixed correlation across all locations (with standard deviation 0.25) and an IID noise term (with standard deviation 0.20). That is:

$$\begin{aligned} cov(\nu_s, \nu_{s'}) &= 0.25 && \text{if } s \neq s' \\ cov(\nu_s, \nu_{s'}) &= 0.25 + 0.20 && \text{if } s = s' \end{aligned}$$

The variance of any point (0.32^2) is very similar to variance of IID noise case previously considered (0.35^2).

Figure 3.8 shows the system mean for each time and location as well as the data and 3 standard deviation confidence bounds for pipe 1, coil 1.

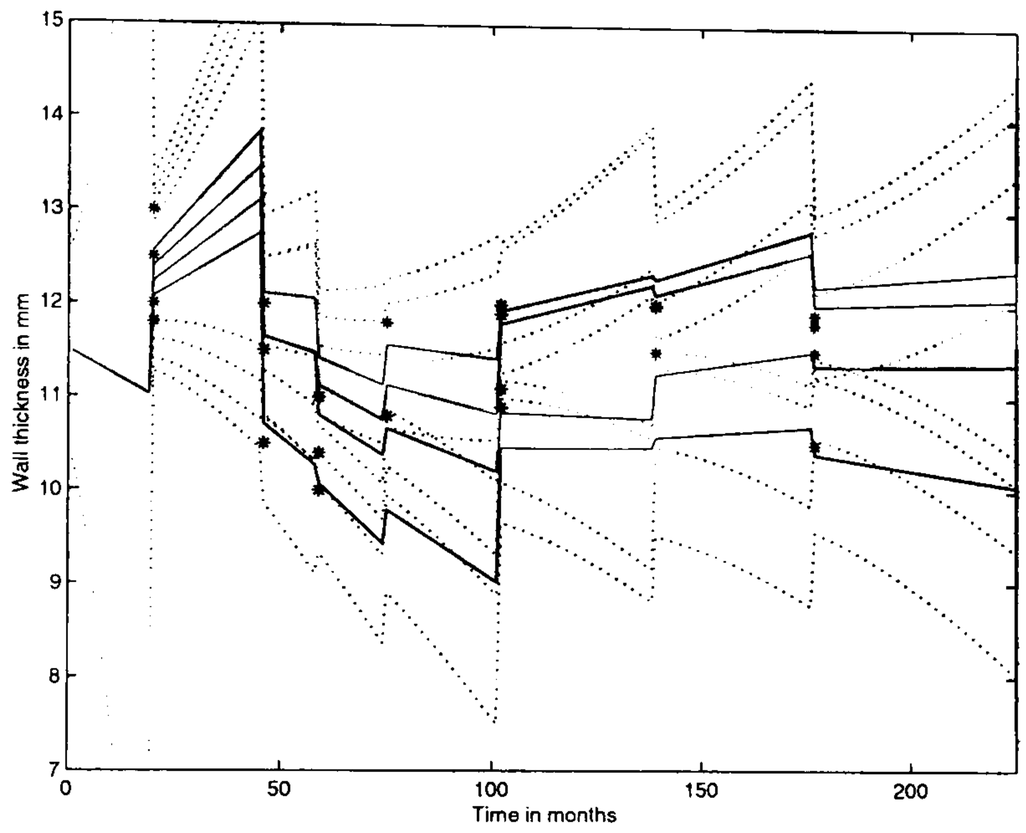


Figure 3.8: Data and forecasts at the measurement points of pipe 1, coil 1

Figure 3.8 is fairly typical of the plots obtained by doing this. As in Figure 3.5 with which this plot is directly comparable, '*' indicates an observation and a solid line the one step ahead point forecast with its three standard deviation credible interval indicated by the dotted lines. The two plots are similar with Figure 3.8 showing a higher level of damping on the trend term as would be expected.

No significant effects on the model are found when correlation is consid-

ered within the observation noise for this example. For the purposes of this example we proceed using independent uncorrelated noise.

3.7 Concluding remarks

These measures have been described to show how a diagnostic analysis might be carried out for models typical of those to which our inspection methodology will be applied. It is emphasised that the inspection methodology described in subsequent chapters is general. The examples are only used for motivation and illustration.

Chapter 4

Future inspections

In the first section of this chapter some potential inspection schemes and several ways of comparing them are discussed. In the second section a range of candidate inspection schemes are introduced. In the third section the change in cost when the number of inspections are changed is considered, while in the fourth section the considerable benefits of being able to inspect anywhere rather than being restricted to existing inspection sites are demonstrated. In the fifth section the risk of system failure in the future is also calculated and then a demonstration made of how simulation can be used to assess how accurately any given inspection regime estimates this risk in comparison with 100% inspection. In the final section future inspection intervals are discussed.

4.1 Inspection criteria

Historically at Shell, inspections have often been made at fixed locations. Potential increases in efficiency from alternative inspection locations are considered. Using the spatial model, the effect of inspection at any set of points on the system can be analysed. Variance at a location will be lower if that point has been inspected, but the variance will also be influenced by the location of other inspections (both current and previous). If the uncertainty about a location is too high that point or a collection of points with sufficiently high cross-correlation must be inspected.

A formal method is required for deciding where future inspections should be made. This is done by setting one or more constraints and then optimising. There are many possibilities for these constraints and a few intuitive choices are considered.

For each point of the system (s) there is a minimum required wall thickness c_s (≥ 0) above which the system level is expected to stay. The inspection at time t should balance the probability of the wall thickness falling below this minimum required value, the inspection cost and the interval of k time units until the next inspection is required due to the uncertainties becoming unacceptably large at some point on the system. The inspection interval

should be as large as possible consistent with acceptable reductions in uncertainty, but often it will be overruled by some other constraint, for example. the next scheduled shut down or safety legislation.

Two scenarios requiring attention are high uncertainty and low remnant life. Increased inspection will help the former but do little for the latter. Observations do not change the true life of the system, only the uncertainties about it. The aim is to minimise the uncertainty about the expected life of the entire system. As a multivariate calculation this is most easily handled using simulation. Ideally, for every possible inspection regime, nested simulations would be run where, first, many sets of data for an initial inspection would be simulated and following every initial inspection further simulations would be run to ascertain the maximum risk prior to the next inspection. The most efficient scheme with sufficiently low risk would then be used for the true inspection. Running this nested scheme over many possible inspection schemes would be very computationally demanding, so we need to develop tractable approximations to such an approach.

A useful inspection will reduce the uncertainty at and around the inspection location. This reduction in variance may be used alongside a heuristic argument about how near the wall thickness is to the minimum allowed

value to avoid the high computational load associated with the simulation approach described above. Two variance and two heuristic criteria are used in this chapter to quantify the results of different inspection schemes and thus allow direct comparisons, see Table 4.1. These criteria cover many intuitive possibilities and can be maximised over possible sampling schemes. Therefore these criteria are used to identify designs with good heuristic properties and full simulation strategies are reserved for confirming that the chosen design does offer an acceptable basis for sampling inspection. Many further alternative criteria are also possible.

A good criterion will allow a fair comparison between different inspection schemes. Furthermore, the criterion must be evaluated before an inspection at time t, and should assess the effect this inspection will have on the uncertainty at time $t + k$. Before an inspection at time t the available data on which to plan future inspections will be D_{t-1} or equivalently if the previous inspection was at time $t - k'$, $D_{t-k'}$ (since no data was observed in the meantime). Since the variance of the underlying values depends only on the observation locations and not the actual values observed it can be easily incorporated into a criterion. The expected underlying wall thickness at time $t + k$ is independent of the inspection scheme but dependent on the observed

values and hence can only be conditioned on D_{t-1}.

The first criterion considered is the maximum forecast variance of the underlying wall thicknesses over the entire system at time t. given all observations up to time $t-1$, D_{t-1}. High variance is uninformative and so this value should as small as possible. A closely related criterion is the mean forecast variance of the underlying wall thicknesses over the entire system. Again this should be small and has the advantage of avoiding all resources being used in a single area of very high uncertainty rather than bringing uncertainty down more generally across the system. The third and fourth criteria are based on a heuristic value: the standardised distance between expectations based on current data and the minimum allowed wall thickness (c_s). Letting s index all spatial locations across the system, the heuristic criterion is defined as $h(s,t,k) = \frac{E(\theta_{s(t+k)}|D_{t-1})-c_s}{\sqrt{var(\theta_{s(t+k)}|D_t)}}$. This heuristic value should be large across the entire system. If it is small at any point then it is not possible to be confident that the critical minimum wall thickness will not be passed. Two aspects of this heuristic, its mean value and its minimum value are considered, both over the entire system. The expectation of the data at time $t+k$ will not depend on the locations of the inspections made at time t hence the numerator is conditioned on D_{t-1}. The variance of beliefs about

the underlying wall thickness at time $t+k$ ($var(\theta_{s(t+k)}|D_t)$) will depend on the locations of observations made at time t but not on the actual observed values.

1	maximum variance	$\max_s(var(\theta_{st}\|D_{t-1}))$
2	mean variance	$mean_s(var(\theta_{st}\|D_{t-1}))$
3	mean heuristic	$mean_s(h(s,t,k))$
4	minimum heuristic	$\min_s(h(s,t,k))$

Table 4.1: Some possible criteria

A good inspection scheme will have a low maximum and mean variance and a high mean and minimum heuristic.

Attention is focused mainly on the heuristic criteria, because the variance criteria, while easier to visualise, are best suited when the wall thickness is far from the minimum allowed wall thickness. This, however, is the same as maximising the minimum (or mean) precision of any point. If a point is far from the critical level, significant changes in variance will dominate the precision calculation, so there are few advantages over the heuristic approach. A combination of all four criteria is possible.

4.2 Candidate inspection schemes

In this section a range of candidate inspection schemes are considered.

In the terms of the example considered, interest is in good location of

inspections at the time of the sixth inspection. Prior beliefs are based on having updated using all available data up to and including the fifth inspection while inferences are required at the time of the seventh inspection. The fifth inspection was made in June 1990, the sixth in July 1993 and the seventh in September 1996. These correspond to months 102, 139 and 177 in the model used in this thesis.

To find the optimal inspection scheme for a given budget, every possible affordable combination of measurement points would have to be considered. In addition to the high computational load the end result might have a higher implementation cost per point inspected, as the inspector would have to locate every inspection point individually rather than following a simple pattern. Instead the improvements to be had from a range of possible basic sampling strategies that can be implemented using the spatial model are first considered. Schemes may either inspect regularly across the system or, in the case of the risk-based scheme, inspect in accordance with the prior beliefs. The schemes considered are summarised in Table 4.2 and illustrated in Figure 4.1.

The 'all' scheme inspects every point and will give the largest reduction in uncertainty that can be achieved for any inspection plan, and can thus

Scheme	Description	Number of observations
All	Heights 1', 2', ..., 21' on every pipe	672
Nothing	No inspection	0
Existing	Heights 3', 7', 12' and 18' on every pipe	128
3 intermediate	Heights 1', 10', and 21'	96
4 intermediate	Heights 5', 10', 15' and 20'	128
5 intermediate	Heights 1', 5', 10', 15' and 20'	160
Alternating	Even pipes as 'Existing' and odd pipes as '4 intermediate'	128
Risk-based	Categorise each pipe as 'low risk' (50%) or 'high risk' (50%) and then apply '3 intermediate' or '5 intermediate' respectively.	128

Table 4.2: Some possible inspection schemes

be used as a bench mark. Conversely the 'nothing' scheme allows comparison with the uncertainty when no inspections are made. Candidate schemes may also be compared to the 'existing' inspection scheme which was actually used to collect the first five sets of data. The intermediate schemes inspect points on every pipe midway between the existing inspection locations. The '3 intermediate' scheme includes inspection of both ends of the pipes, (the highest uncertainty identified by this model before the sixth inspection is at the ends of the pipe). The '4 intermediate' scheme targets the four areas of highest uncertainty while the '5 intermediate' scheme has one inspection point in each of the gaps left by the 'existing' inspection scheme. The alternating scheme tries to make use of the spatial correlation across pipes by

alternating the 'existing' and '4 intermediate' schemes. The alternating option may be further extended by swapping the schemes over for subsequent inspections. In this example, the risk-based scheme categorises as high risk any pipes which are believed at some location to be within 3.5 standard deviations of the critical value. The total number of points inspected at each inspection for the 'alternating' and 'risk based' schemes remains fixed at 128.

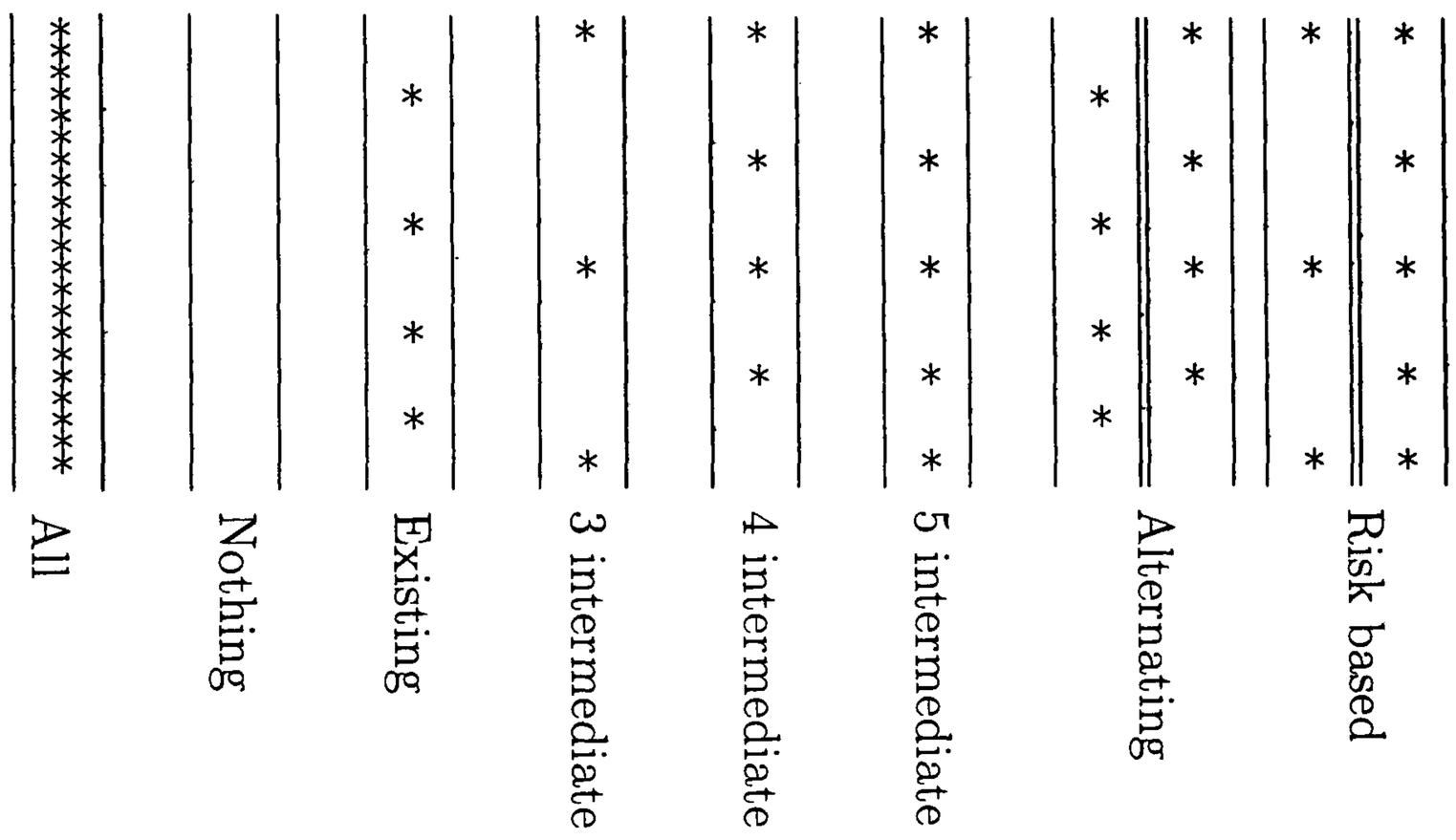


Figure 4.1: Observation locations

As discussed in Section 3.3.2, for the purposes of this example the last data observation was at time 102, an inspection is to be made at time $t = 139$ and an inference made at time $t + k = 177$ ($k = 38$), and a critical wall thickness (c_s) of 6mm.

Of the schemes making less than 128 point inspections only the risk-based

Scheme	maximum variance	mean variance	mean heuristic	minimum heuristic
All	0.53	0.50	13.35	9.97
Nothing	2.30	1.56	4.58	2.64
Existing	2.17	1.33	6.22	2.80
3 intermediate	1.87	1.21	6.21	2.96
4 intermediate	2.11	1.08	6.92	2.73
5 intermediate	1.32	1.00	7.36	4.11
Alternating	2.05	1.08	7.24	2.82
Risk-based	1.55	1.07	6.83	4.10

Table 4.3: Different criteria versus different schemes

scheme gives much information about the 'minimum heuristic' (see Table 4.3). Because the '5 intermediate' scheme targets the 5 major spatial areas of uncertainty between the traditional inspection locations the 'minimum heuristic' shows a marked improvement over the four point schemes. None of the candidate schemes are anything like equivalent to a full inspection. The '3 intermediate' scheme is generally better than the existing inspection scheme and has a considerably lower inspection burden.

4.3 How many measurements

Doing a full search over a given number of inspections per pipe to find the optimal set of inspection locations is computationally very intensive. A step-wise delete (or add) process can be used instead with a very much reduced computational load. Starting with a full inspection across the entire furnace,

the effect of removing all inspections from each level in turn is considered and then an assessment made about which set of inspections is needed least according to the heuristic. The inspections at this level can then be removed and the process repeated until inspections are only being made at the desired number of heights. Similarly a stepwise add process may also be used.

In Figure 4.2 the expected value of the heuristic is plotted against the number of heights inspected. Having a very low number of inspections appears ineffective at reducing the maximum uncertainty and having a few uninspected points dramatically affects results. High confidence can only be achieved by 100% inspection. Using the ordering determined by the minimum heuristic the mean heuristic is also plotted. The slope looks concave but this could be a feature of applying the stepwise delete to the minimum heuristic and not to the mean heuristic. The jumps in level in the minimum heuristic line may be a consequence of the existing fixed point inspection scheme leaving five key areas of uncertainty.

The pipe heights are, from least important, to most important:

7, 18, 12, 3, 20, 16, 9, 14, 5, 11, 2, 19, 17, 8, 13, 6, 4, 10, 15, 1, 21.

The ends of the pipe are most critical and the four locations inspected in the first five inspections at heights 3, 7, 12 and 18 least critical. Points

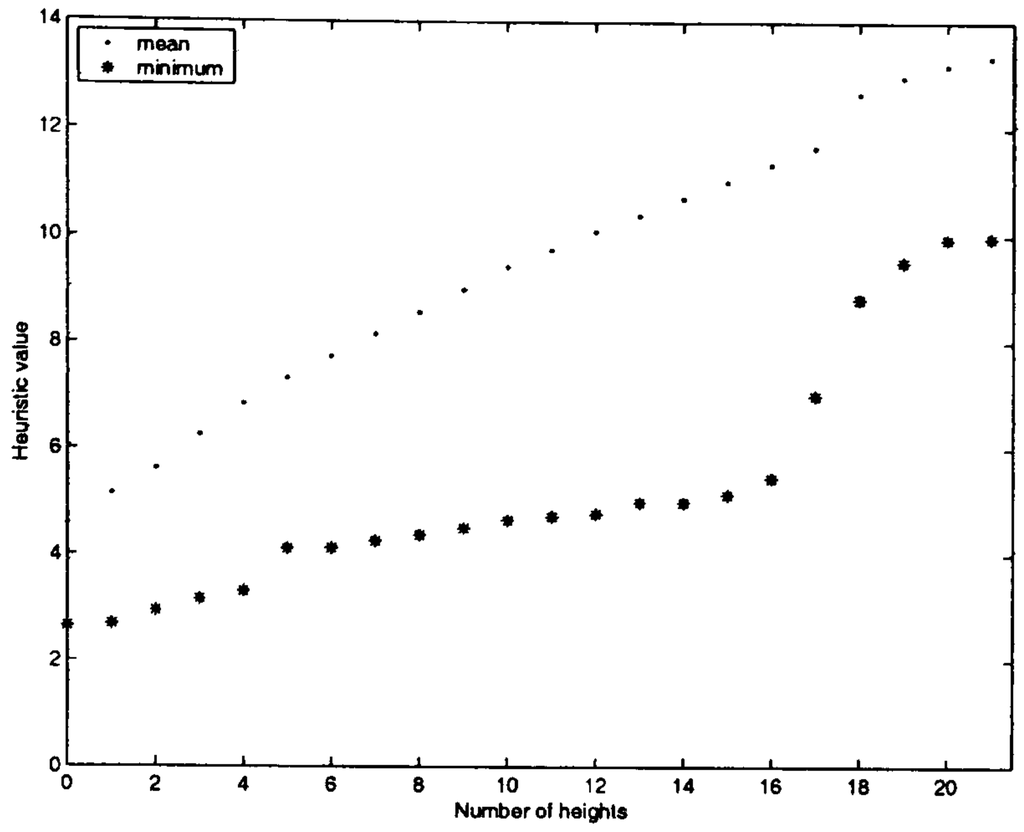


Figure 4.2: Heuristic value using stepwise delete

intermediate to these existing inspections also have a high priority. Points adjacent to those not yet removed by the stepwise delete also have a low inspection priority. It seems likely that heights 4 and 6 remaining undeleted for so long may be a consequence of using a stepwise delete rather than a better search algorithm. Further improvements may be achieved by combining stepwise deletions with stepwise additions. However, in this case, because of the regular nature of the previous inspections, significant improvement seems unlikely.

Using stepwise deletions four recommended points 1, 10, 15 and 21 are obtained with a minimum heuristic value of 3.31, which is a marked improvement over the original suggestions described in Table 4.3. Were resources

available it would seem sensible to make a fifth inspection at height 4' (or perhaps 5') to exploit the jump in the heuristic value curve. This possibility is investigated later.

Figure 4.3 shows the variance for different numbers of heights inspected. Heights are selected for removal from the inspection scheme using a stepwise delete based on the maximum and the mean variance criteria. As with the heuristic criteria value the mean variance is almost linear while the maximum variance has various steps and full inspection is much superior to almost full inspection.

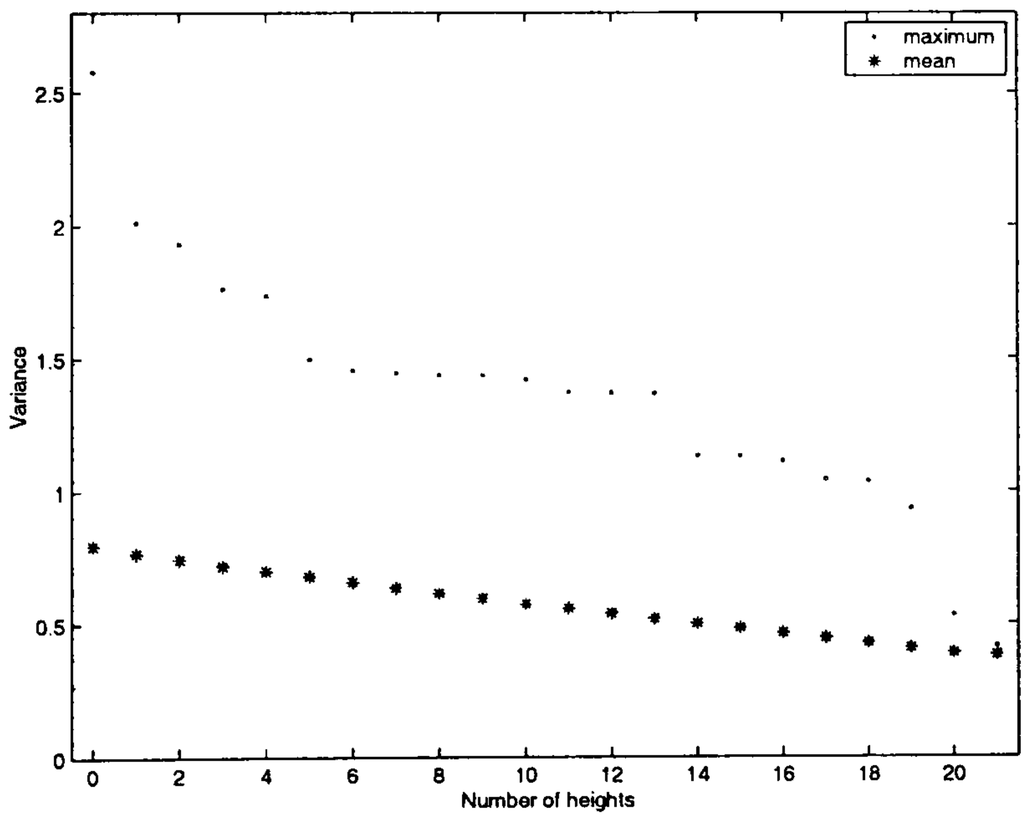


Figure 4.3: Variance using stepwise delete

4.4 More advanced inspection regimes

The optimal inspection scheme will comprise of measurements at locations across the entire system and not just at restricted heights across the system. The resulting scheme may be practically difficult to implement and there may be restrictions on where measurements can be made but these concerns are put aside for the present. Considering a more advanced inspection scheme allows the expected value of the scheme to be calculated as well as the cost of the inspections. This scheme may then be fairly compared against alternatives.

An approximation to the optimal scheme, is to sequentially add (or delete) single points according to the criteria. The criterion value is checked at every point and an inspection made at the point with maximum impact on the heuristic. Subsequent points are then selected based upon the updated beliefs. Figure 4.4 shows the increase in the minimum and mean heuristic values as the number of inspections increases using a sequential add on the minimum heuristic until there is an average of four inspections on each pipe. Devising an inspection scheme using the stepwise add is computationally efficient but likely to produce a scheme far from optimal. Investigation of alternative and potentially better inspection schemes is potentially worth

while.

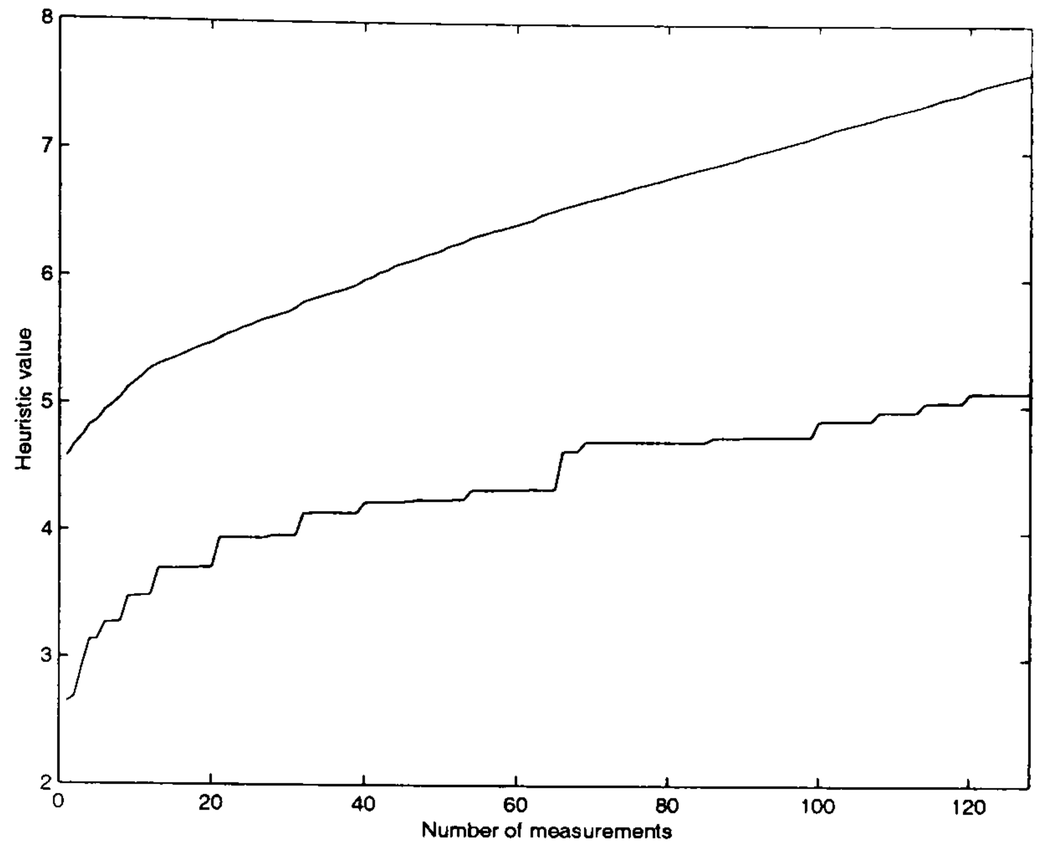


Figure 4.4: Number of points inspected against heuristic value
Upper line - mean heuristic, lower line - minimum heuristic

The mean heuristic increases continuously with the first inspections being most important. The minimum heuristic increases somewhat less smoothly. This may be compounded as a result of using the stepwise add, but may also be due to a certain intensity of inspection covering entire areas of similar structure. For example, suppose many of the pipes have an area with a very similar level of uncertainty resulting from the regular structure of the existing inspection scheme and that these areas are almost uncorrelated. If this set of areas are the worst part of the system, every one of these areas will need an inspection to increase the uncertainty of the overall worst point.

Comparing inspection based on these 128 inspection locations (that is, an average of four per pipe) to the best scheme found with fixed inspection heights on every pipe, the minimum heuristic value is much higher (5.11 rather than 3.31). To get this level of accuracy using fixed height inspections 15 inspections would need to be made on every pipe. The mean heuristic is also improved (7.63 rather than 6.85) which would require six fixed height inspections per pipe. To achieve the same minimum heuristic accuracy as the 128 fixed height inspections only nine carefully positioned inspections need be made.

In Figure 4.5 the heights of the points added by the stepwise add are sequentially plotted. A clustering of points around the areas where prior uncertainty was highest - heights 1, 5, 10, 15 and 21, can clearly be seen. The number of inspections in each coil, in each pipe and generally throughout the system has also been considered but no obvious patterns were found (see Figures 4.6 and 4.7).

4.5 Assessing the risk

The objective when making inspections is to allow the furnace to be run for as long as possible and be stopped just before the wall thickness at any point

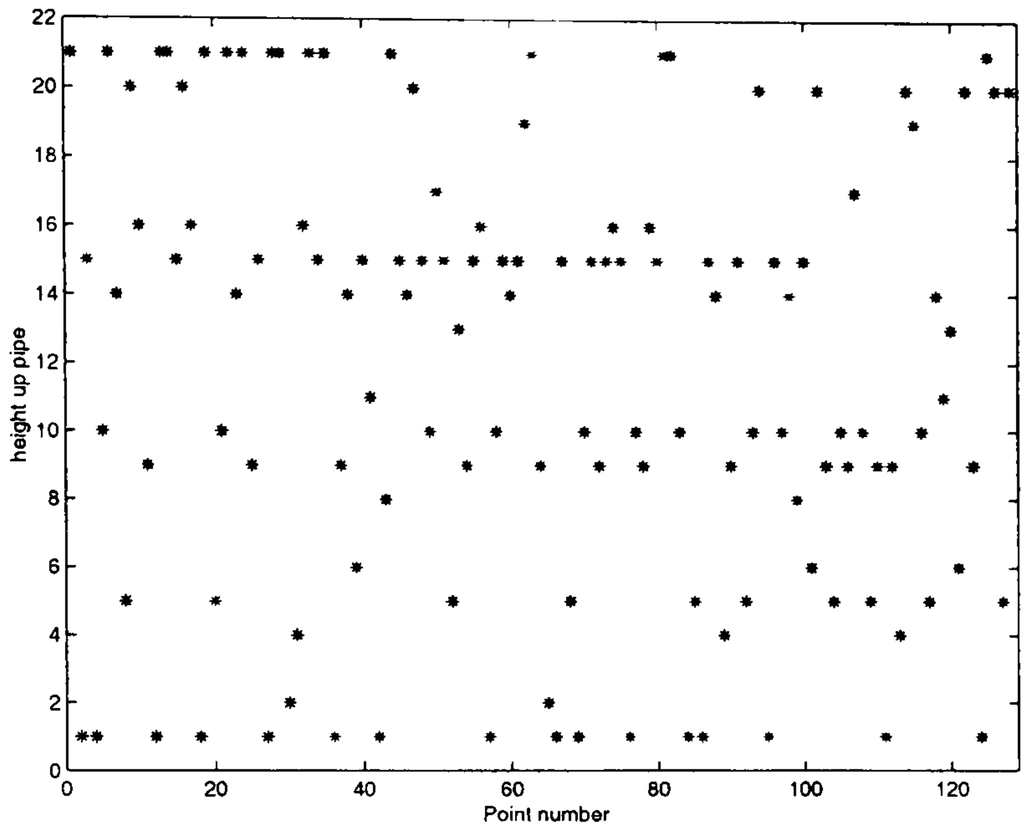


Figure 4.5: Optimal inspection heights using criterion

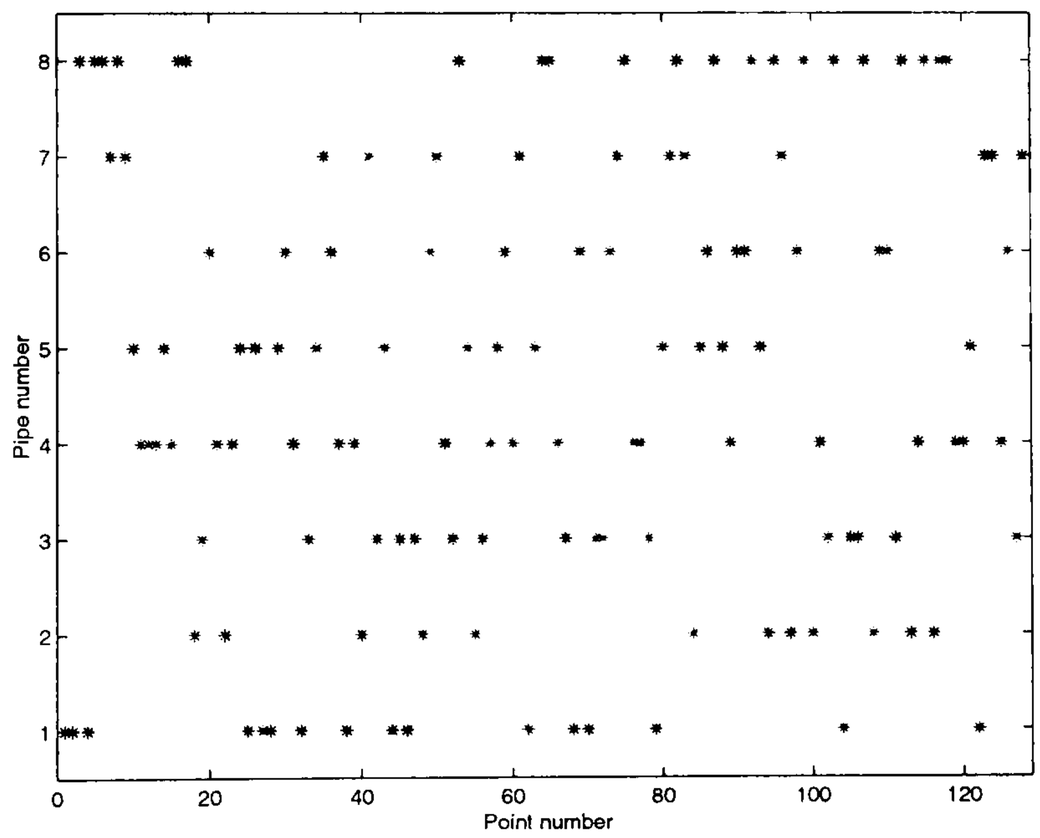


Figure 4.6: Optimal inspection pipes using criterion

becomes too thin. Because there is considerable uncertainty about the time of the crisis point a safety margin is likely to be required, depending on the consequences of overshooting the minimum wall thickness.

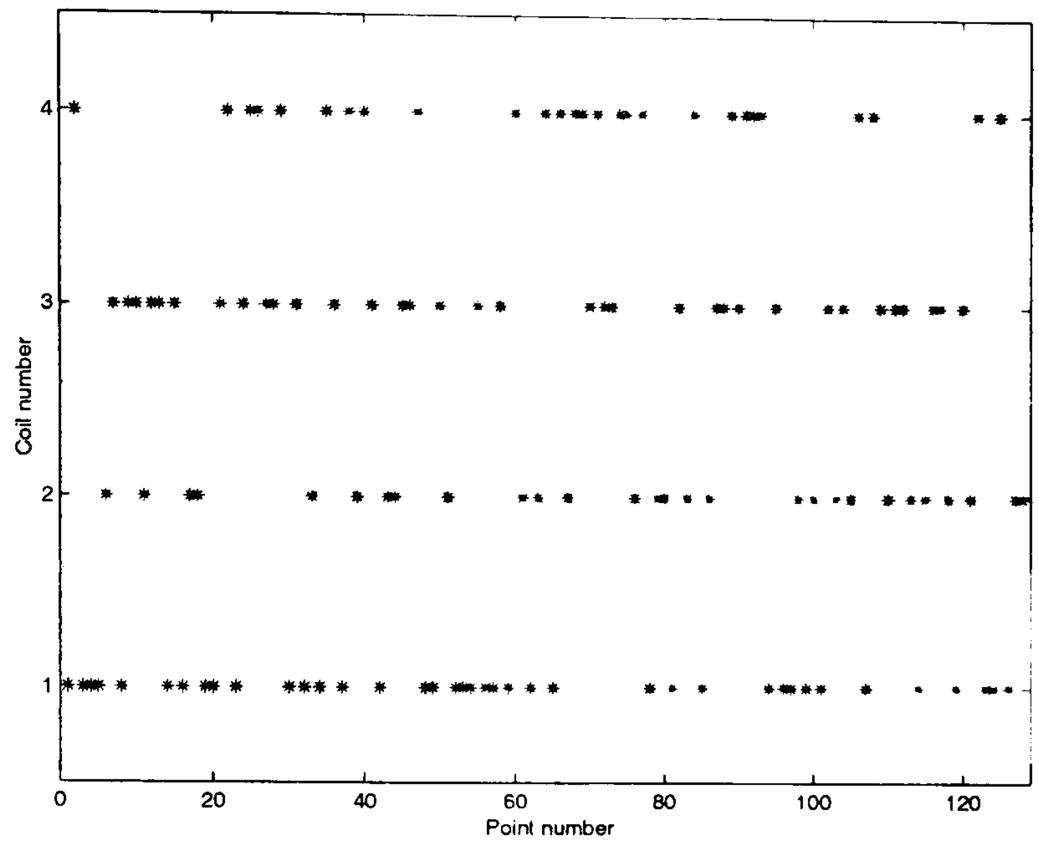


Figure 4.7: Optimal inspection coils using criterion

In a multivariate model, calculating the probability that the worst point of the system will pass the crisis point is difficult. By simulating data sets corresponding to current beliefs for the end point of the forecast, the probability that the wall thickness will fall below the critical minimum value for any point of the system may be found.

10 000 data sets are simulated covering the entire system corresponding to beliefs at the time of the seventh inspection, given all data up to and including the fifth inspection. This simulation is based on the known covariance between every point of the underlying system at the time of the seventh inspection having updated for all observations made up to and including the time of the fifth inspection. The minimum wall thickness is below the critical

value of 6mm in about 7% of cases. From the simulation study the mean wall thickness of the worst point of the system is found to be about 7.8mm with a standard deviation of 1.14mm. The minimum wall thickness measurement actually observed and recorded at the time of the seventh inspection was 9.7mm. This large difference between the expected minimum and the observed minimum emphasises the risk of ignoring the un-monitored intermediate points of the system, as is common practice at present in Shell and more generally when a spatial model is not used.

4.5.1 Comparing inspection schemes

The inspection of all pipes according to a variety of inspection schemes is compared with the inference that would be made following a full inspection. The same underlying simulated data is used for every candidate scheme in each comparison.

First, given prior beliefs and all data from inspections 1 to 5, beliefs about the system at the time of the sixth inspection are updated. A full set of data is then simulated according to the updated beliefs at the time of the sixth inspection. Each inspection scheme is then considered in turn so the conclusions according to each scheme are directly comparable since they are based on the same set of simulated data. A hypothetical sixth inspection

is made according to the chosen scheme and then in the light of this set of observations beliefs are updated about what should be expected at the time of the seventh inspection. A large number of realisations (3000) of the entire system at the time of the seventh inspection are then simulated from the updated beliefs for the system at this time. The proportion of simulations that have a minimum wall thickness below the critical level of 6mm can then be calculated. If this proportion is one then the inspection made suggests the system is certain to fail by the time of the seventh inspection. Conversely, if every simulation has a minimum wall thickness over 6mm then the inspection scheme suggests that the system is almost certainly safe up to the time of the seventh inspection. By comparing with the proportions from other inspection schemes with updated beliefs based on the same simulated data for the sixth inspection a fair comparison may be made. This whole process may then be repeated for different simulations of data at the time of the sixth inspection. An inspection scheme whose predictions are generally at odds with those made following 100% inspection may be considered an undesirable inspection scheme. Conversely, an inspection scheme which generally gives predictions close to the predictions made following 100% inspection may, given the restrictions imposed, be considered a relatively good inspection

scheme. If 100% inspection is not sufficiently informative it is likely that the inspection interval is too long.

Attention is focused on seven candidate inspection schemes with 377 realisations of data at the time of the sixth inspection in each case. Four schemes based on the points identified by the stepwise add process in Section 4.4 are considered. Specifically, the most important 16, 32 and 64 points as well as the full 128 points are considered. The '4 intermediate' and the '5 intermediate' schemes from Table 4.2 and the existing inspection scheme are also brought into the comparison. The existing scheme is expected to compare unfavourably to the alternative schemes, while the '4 intermediate' and '5 intermediate' schemes are expected to fall between the 16 point and 128 point schemes.

Figure 4.8 shows the correlations between the forecast minimum wall thickness at the time of the seventh inspection generated following the previous description, over 377 simulations for each of the inspection schemes under consideration. The correlation between the '5 intermediate' scheme (160 inspection points in total) and full inspection is only slightly greater than the correlation between the 32 point scheme and full inspection (0.900 and 0.908 respectively). The existing scheme has a much lower correlation

with full inspection than any of the alternative candidate schemes.

Figure 4.9 shows the correlation between the estimated probabilities of system failure for the different inspection schemes. The existing inspection scheme has a correlation of about 0.3 with all other schemes and is not very informative so was left out of the plot. The 128 point scheme has the strongest correlation with full inspection. The 16, 32, 64 and 128 point schemes have a correlation with full inspection which increases as more points are observed. The '4 intermediate' and '5 intermediate' schemes are very similar to each other and have about the same correlation with full inspection as the 64 point scheme.

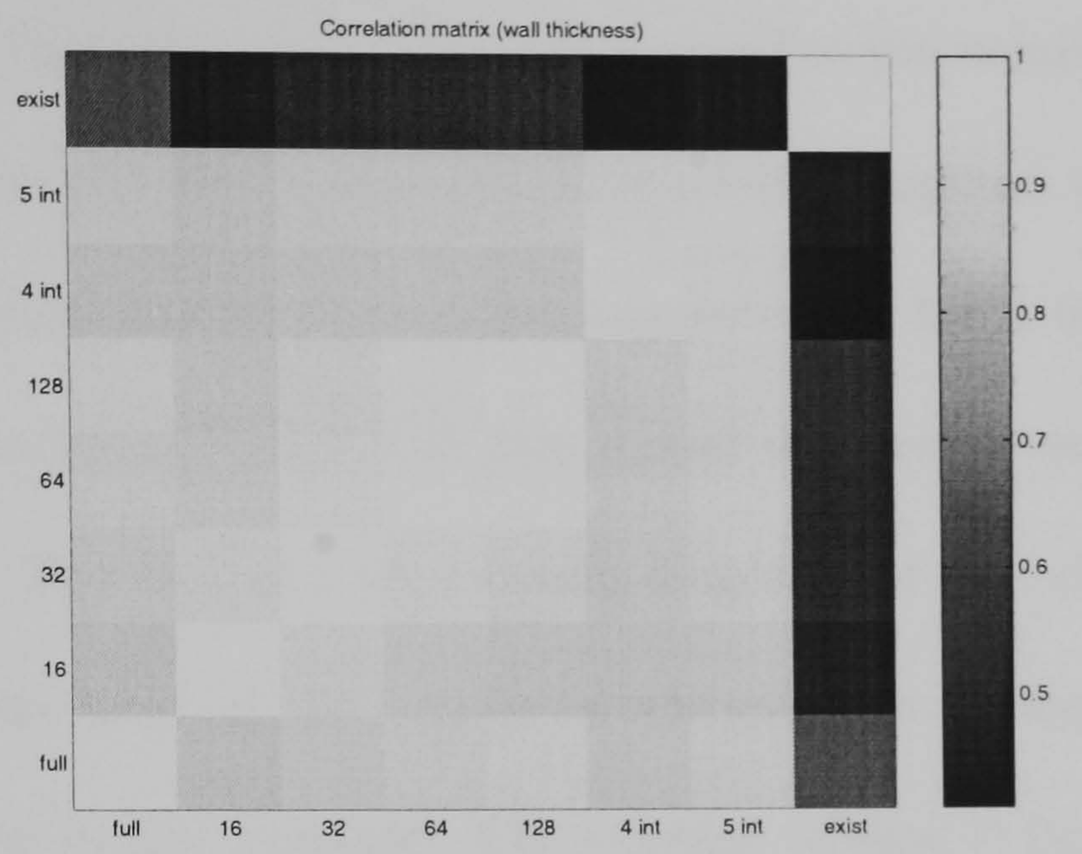


Figure 4.8: Correlations between scheme forecasts

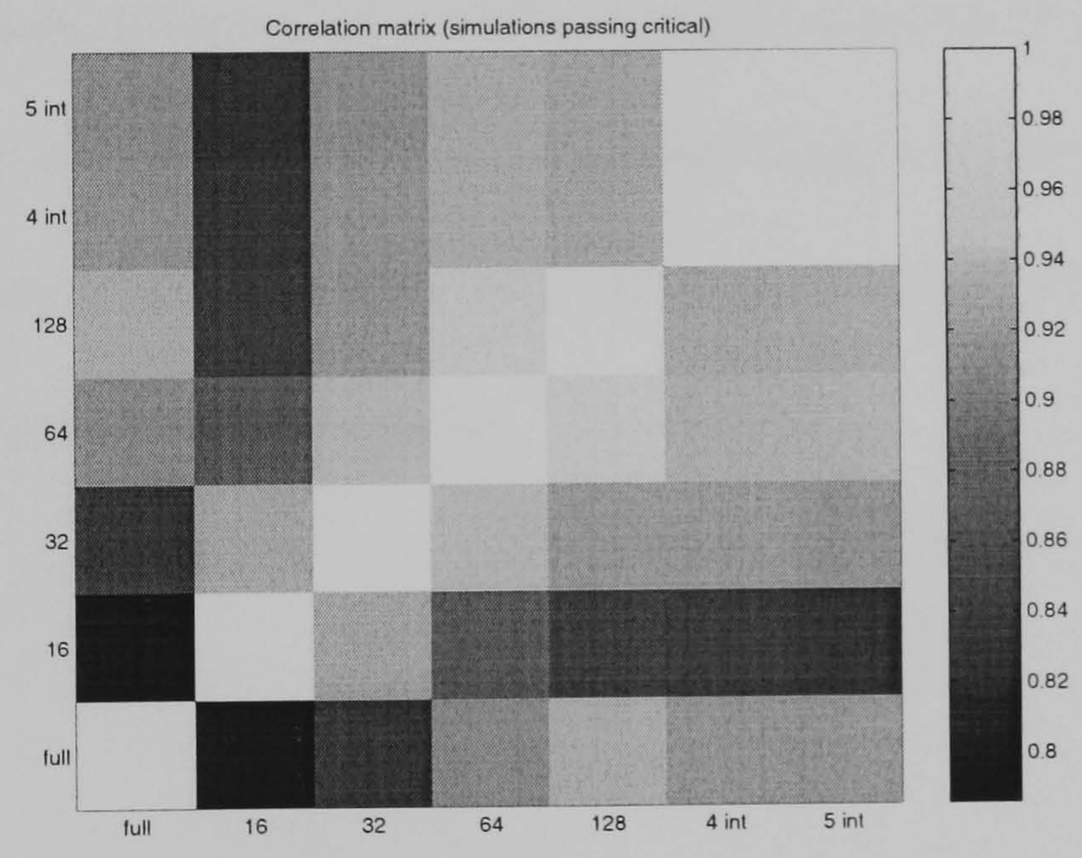


Figure 4.9: Correlations between scheme failure proportions

Figure 4.10 shows the probability of passing the minimum wall thickness suggested by each of the seven candidate schemes against the probability of passing the minimum wall thickness suggested by 100% inspection. The same data is shown in Figure 4.11 but spread out over several plots to aid interpretation. The lack of a very strong diagonal for any of the candidate schemes suggests none of the candidate schemes is a strong alternative to 100% inspection in this example. The existing scheme is furthest from the diagonal and shows little sign of a trend in that direction. The 16, 32, 64 and 128 point schemes all show a reasonable diagonal trend which gets stronger as more points are observed. The '4 intermediate' and '5 intermediate' schemes also show a reasonable diagonal trend although it is difficult to distinguish significant differences between them.

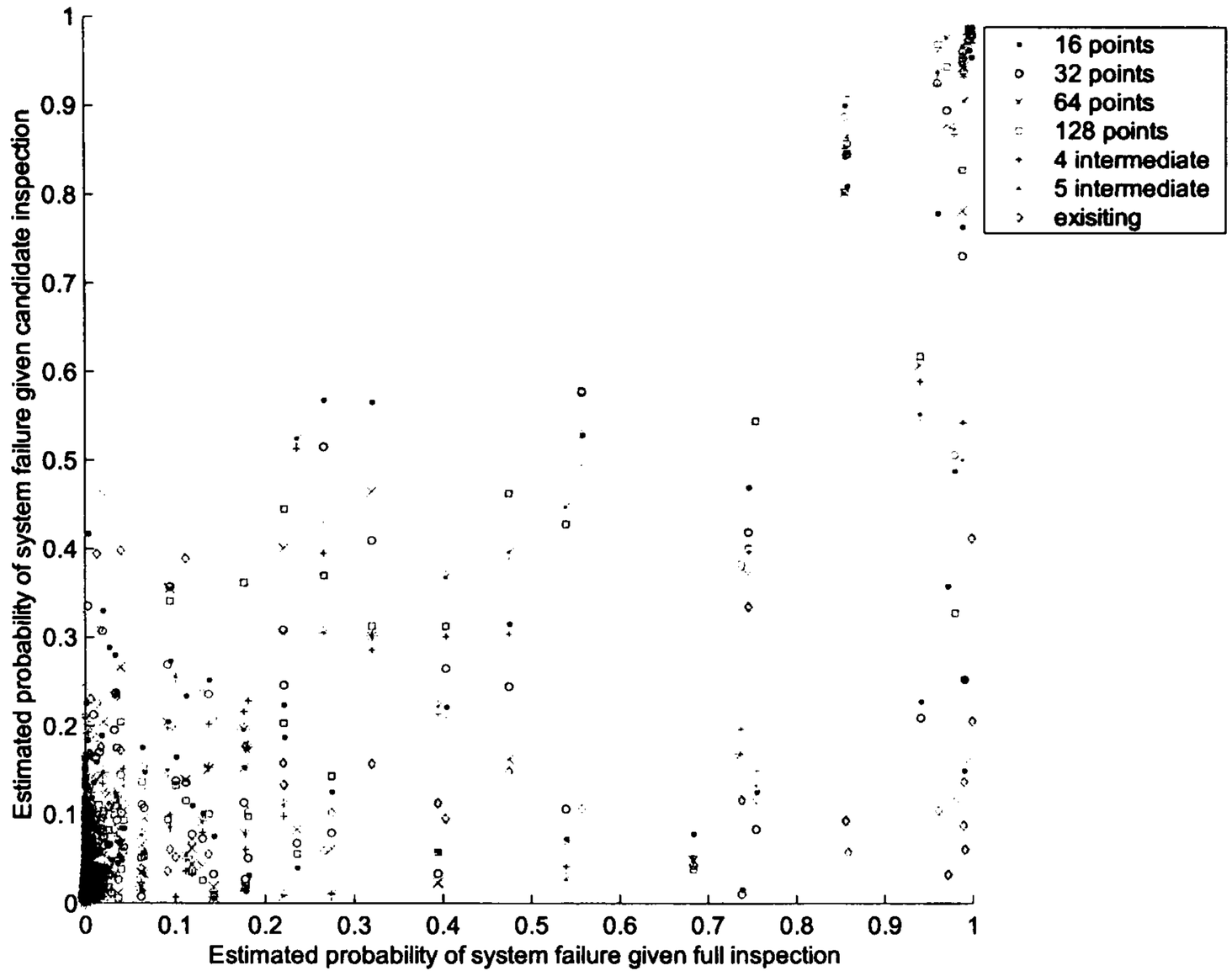


Figure 4.10: Proportions in risk for full inspection against candidate schemes

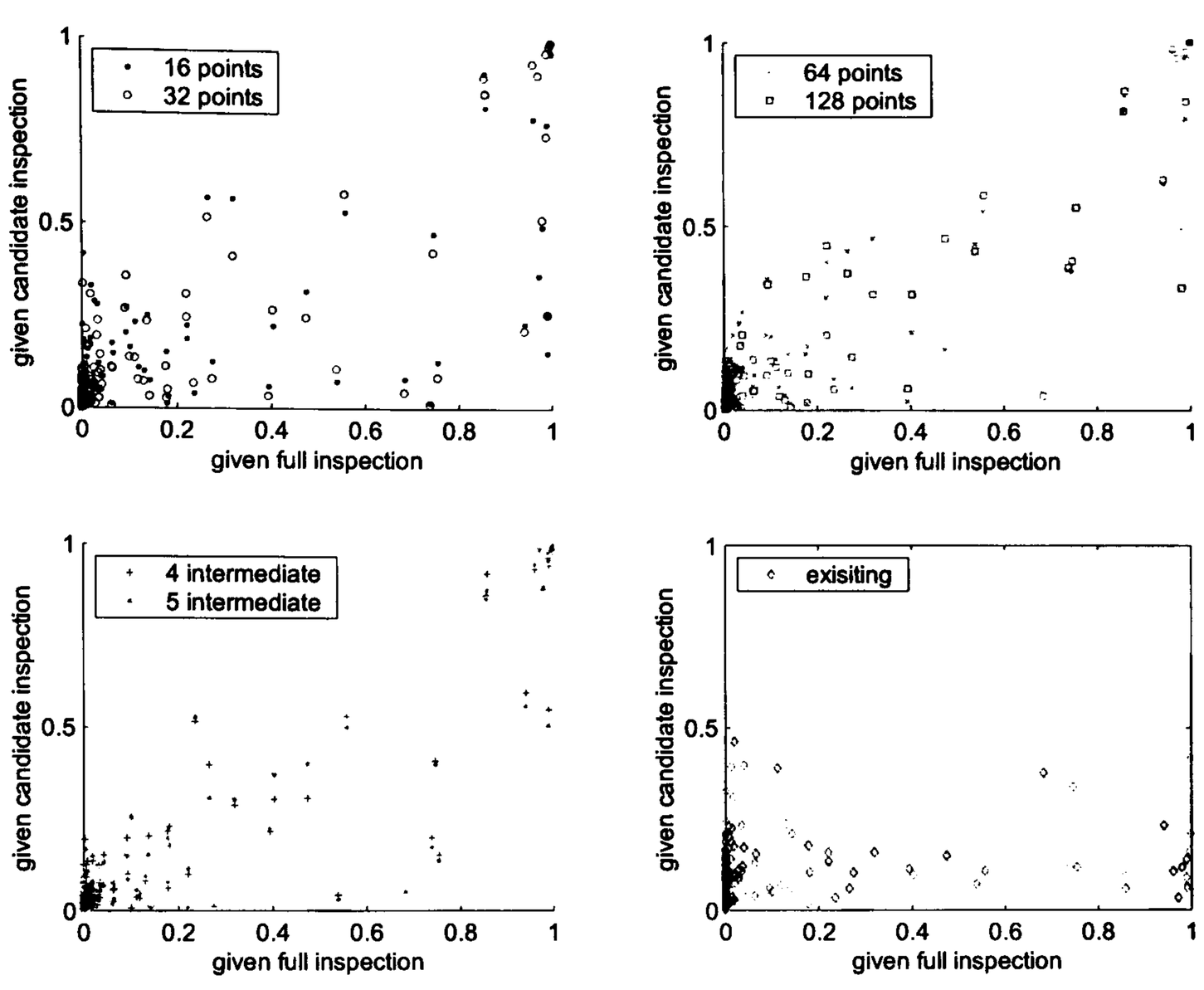


Figure 4.11: Proportions in risk for full inspection against candidate schemes

Each plot may be split into four (unequal) quadrants at the value of the maximum risk that is considered acceptable - for example, 0.05. Points in the bottom left quadrant indicate neither the candidate scheme nor 100% inspection suggest the minimum wall thickness criteria is likely to be reached. Points in the top left quadrant indicate that full inspection suggests the system is safe but the candidate scheme disagrees. Points in the top right quadrant indicate neither inspection regime believes the state of the system to be satisfactory, while points in the bottom right quadrant indicate the full inspection has found a high risk which the candidate scheme has failed to identify. The cost of a false positive and of a false negative can then be used to value each inspection regime. Table 4.4 shows the frequencies in each of the 4 quadrants around 0.05 for these 8 schemes. The quadrants in each matrix correspond to the quadrants described above. As before, the 16, 32, 64 and 128 point schemes get better as the number of points increases. The '4 intermediate' and '5 intermediate' schemes are very similar to each other. The existing scheme identifies much more risk than a full inspection classifying about 50% of simulations as passing the critical wall thickness. All the other schemes classify a similar number of situations (between 20 and 45) as not giving cause of concern where full inspection has identified a

problem. The existing scheme flags a much higher number of simulations as having a problem where none exists than any of the candidate schemes. A natural extension of this approach is to split the plot into more areas allowing a more detailed consideration of misclassifications.

Scheme	Count		Scheme	Count	
16 point	45	36	4 intermediate	20	33
	291	5		316	8
32 point	33	35	5 intermediate	20	33
	303	6		316	8
64 point	27	35	existing	150	36
	309	6		186	5
128 point	22	36	full inspection	0	41
	314	5		336	0

Table 4.4: Frequencies for the 4 quadrants about 0.05 for different schemes

In addition to the plots and tables above various numerical comparisons are also possible. An uninformative inspection scheme will give no information about future wall thickness so will not effect predictions hence the variance of predictions will be close to zero. An informative inspection scheme will correctly make many extreme forecasts (system will almost certainly fail / almost certainly not fail) and thus have high variance. Hence, a larger variance for the estimated probability of system failure for a sampling scheme is an indication of a better inspection scheme. Table 4.5 shows the mean predicted minimum value (over the system) and its variance for all the inspection

schemes calculated from the 377 realisations of the system. The mean prediction becomes higher as the inspection schemes become more extensive because points at lower risk and further from the critical wall thickness are included in the inspection. The variance of the proportion of simulations with critical results for each scheme show the behaviour expected with a higher variance for the schemes previously identified as likely to produce better results. The mean deviation from the diagonal in Figure 4.10 allows a further easy comparison between schemes. This is calculated as the square root of the mean squared difference between the predictions made based on full inspection and predictions based on the candidate inspection scheme. Again the results reflect the previous discussions.

Scheme	mean prediction	variance of mean	variance of proportion	mean deviation from diagonal
full	7.901	0.057	0.037	NA
16	7.840	0.044	0.020	0.124
32	7.874	0.047	0.024	0.108
64	7.887	0.051	0.027	0.084
128	7.910	0.052	0.029	0.072
4 intermediate	7.903	0.048	0.027	0.082
5 intermediate	7.910	0.049	0.027	0.082
Existing	7.771	0.036	0.005	0.190

Table 4.5: Comparison between schemes

None of the candidate inspection schemes is exceptionally good at esti-

mating the probability of system failure. However, the existing inspection scheme can be easily improved upon. It seems likely that the prior distribution has too much variance for these schemes as they stand and that either a much fuller inspection is required or the next inspection should be brought forward. In a case study of a particular system a full sensitivity analysis to changes in such constraints and parameters should be carried out and suitable changes considered where high sensitivity is found.

Study of plots and general calculations like those described above for a range of different designs allows an inspection plan which balances cost against acceptable levels of risk to be chosen. The final decision about which sampling scheme is most suitable must be left to the experts.

4.6 Determining the maximum inspection interval

Within Shell, the time of the next inspection is often calculated based on the predicted remnant life of the system. This is determined using an interval factor [0,1], where more critical items have a smaller interval factor. The inspection interval is then the product of this interval factor and the predicted remnant life. This inspection interval may be overruled and in particular

shortened by other considerations such as convenient times to make future inspections. The difficult part of this approach is the calculation of the remnant life. In this example, prior to the sixth inspection, the underlying wall thickness θ is increasing (that is $\beta > 0$) across all locations. This makes comparison with forecasts at individual points of expected time until system failure uninformative. An alternative approach to determining the inspection interval is through the specification of a 'maximum risk', a typically small maximum allowed probability of system failure.

The probability of system failure may be calculated by simulation. Isolated consideration of a single time point may not be appropriate and ideally the cumulative effect should be considered. One possible way of simulating cumulative data is to simultaneously simulate data across all time points. However, this will often have a very high computational load. An alternative is to simulate data immediately following the inspection and then simulate one step ahead updates. The minimum wall thickness at each time (up to some limit) may then be evaluated. A possible limit for the inspection interval could be based on the time that the expected value of any point of the system hits the critical wall thickness. By repeating this simulation the probability of system failure at any given time may be estimated. The inspection

scheme used will not affect the time of system failure only the accuracy of its prediction.

Figure 4.12 shows the probability of system failure between the sixth and the seventh inspection based on 1000 simulations. This is independent of any inspection that is made. In each simulation data is simulated at the time of the sixth inspection and updated sequentially for subsequent time points. Failure at any time for a particular simulation automatically made all subsequent times also failures. Figure 4.13 shows probabilities of hitting critical values of 8, 7, 6, 5, and 4 mm based on 1000 simulations over 200 time steps starting after the sixth inspection. The probability of some point of the system reaching a given wall thickness appears to increase quickly and then as would be expected tends towards 1 as further time steps are considered.

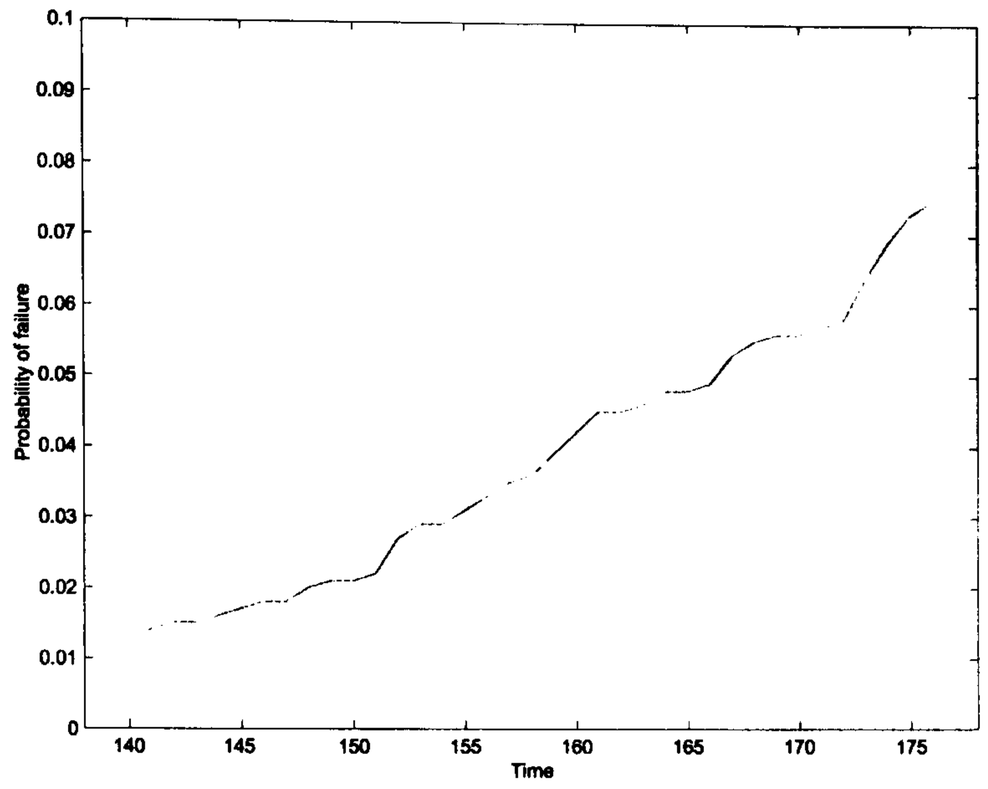


Figure 4.12: Probability of system failure through time

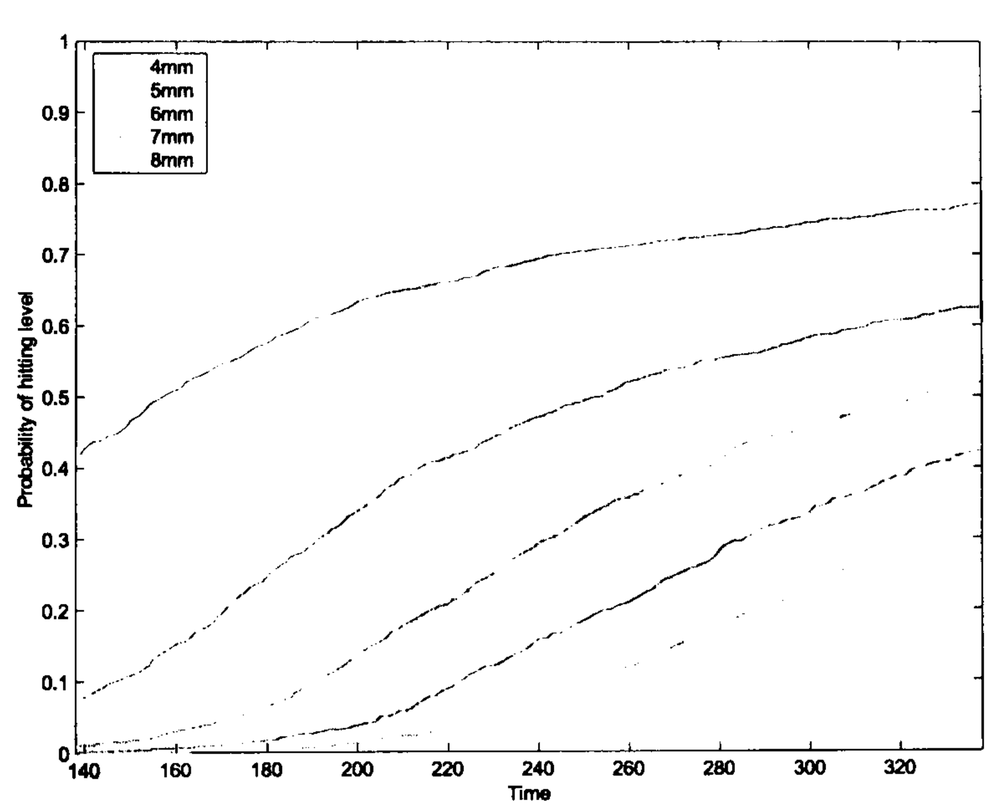


Figure 4.13: Long term probability of system failure

Chapter 5

Minimum data

Many modern inspection tools allow non-intrusive inspections to be made. These have many advantages over visual inspections both economically and for safety reasons. Many non-intrusive inspection techniques allow a region to be quickly surveyed while the inspection device records only the minimum wall thickness observed in that region. In this chapter the key issues that must be addressed when using spatio-temporal regional minima data are discussed. We also discuss the use of spatio-temporal DLM's for designing efficient inspection schemes for industrial systems. Specifically, systems which are corroding, where inspection data is aggregated and transformed will be described. This chapter also concerns the benefits of careful modelling and analysis of spatio-temporal phenomena such as wall thickness.

In section one, a motivating example is described - storage tanks at a

large petrochemical facility used to store various solvents. In section two the key aspects of model fit for this example are discussed. In section three the possible applications of spatially transformed observations within a DLM framework are discussed. In section four errors due to approximating a continuum with a finite grid are investigated. Some of the main problems with making spatio-temporal inferences using data recorded for general inference only are also described.

5.1 Motivating example

The work is motivated and illustrated by two cylindrical storage tanks. The first of these tanks is used to store caustic soda (10% NaOH - a strong alkaline commonly used in the manufacture of other chemicals). The second tank is used to store crude acetone (also known as propanone - widely used in industry as a solvent for numerous organic substances). Caustic soda and acetone both have relatively low corrosion properties, hence the tanks do not need frequent internal inspection and have a long expected life.

The tanks are built on raised concrete foundations and are approximately 30' (10 m) tall with a diameter of approximately 20' (7 m). The sides of the tanks are constructed from 36 plates welded together in the form of

6 rings (strakes) each about 5' high. Each strake is constructed from 6 identical parts (hexicircles) which are welded together. The top and bottom are similarly constructed from numerous plates welded together. There are also two manways (inspection hatches) located close to the floor, an earthing conductor and a set of stairs leading to the roof of the tank. These fixed points can be used to calibrate spatial locations but are not thought to alter significantly the risk associated with any particular parts of the tanks. The tanks were commissioned in the 1940s and are still in use at the current time. The work is not a case study of either of these tanks. However, the common physical aspects are used to motivate and illustrate the approach.

It is known that the lower part of these tanks will be more prone to internal corrosion than other parts of the tanks, because of the buildup of sludge, and so more inspections have been carried out on this part of the tanks. In the following example, only the lower part of the first strake is modelled while it is assumed that all other parts of the tank will be less seriously corroded; however, the methods applied could easily be extended to the full tank.

There have been four inspections of each tank. Each inspection measures the wall thickness at a range of locations to assess vessel integrity at the

time of the inspection and in the future. The earliest inspection (26/6/1970) for tank 3005 has 32 measurements in the first strake with a total of 29 further measurements in the second and subsequent strakes. Wall thickness measurements of higher strakes are generally made along the line of the stairs - avoiding the need for scaffolding. The second inspection (2/3/1979) is in the form of three equally spaced inspection rings on the lower strake, each with 18 regularly spaced measurements and a further 31 measurements in the second and subsequent strakes. The third inspection (30/5/1991) provides no spatial information although 19 measurements appear to have been taken in the first strake and 14 in subsequent strakes. The most recent inspection (May 2002) has 12 minima from different regions towards the bottom of the first strake. These regions were approximately 1' by 2' and were equally spaced around the tank.

The earliest inspection (2/5/1969) for tank 3006 has 28 measurements in the first strake with a total of 15 further measurements in subsequent strakes. The second inspection (March 1980) is in the form of three equally spaced rings on the lower strake each with 18 regularly spaced measurements and a further 32 measurements in the second and subsequent strakes. The third inspection (9/7/1992) has four regional minima and maxima, while the

most recent inspection (May 2002) has twelve minima from different regions towards the bottom of the first strake.

Corrosion can be either internal or external. External corrosion is mainly due to rain and can be seen as small patches of external rust. This is found using regular external visual inspections and is then patched externally as necessary and painted over. The resulting wall may be thicker than might otherwise be expected. The spatial details of the external repairs carried out are not available. However all the repairs are small (less then 2" high) and located just above the floor. Internal corrosion will generally be more serious and require more substantial repairs - no repairs have been carried out following internal corrosion on either tank.

Particular problems with these data include:

1. No record of the tank's design specification perhaps due to the long time scale involved.

2. A lack of detailed information about the inspection methods used especially for the early inspections.

3. Considerable evolution of the inspection techniques used.

4. Lack of accurate spatial information for spatial data.

5. Irregularly spaced observations.

6. High observational noise.

The data for the two tanks being modelled are summarised in Table 5.1 and Table 5.2, and plotted in Figure 5.1. While observations on tank 3005 appear to show little other than observation noise, observations on tank 3006 suggest some corrosion most markedly in the mean and maximum measurements.

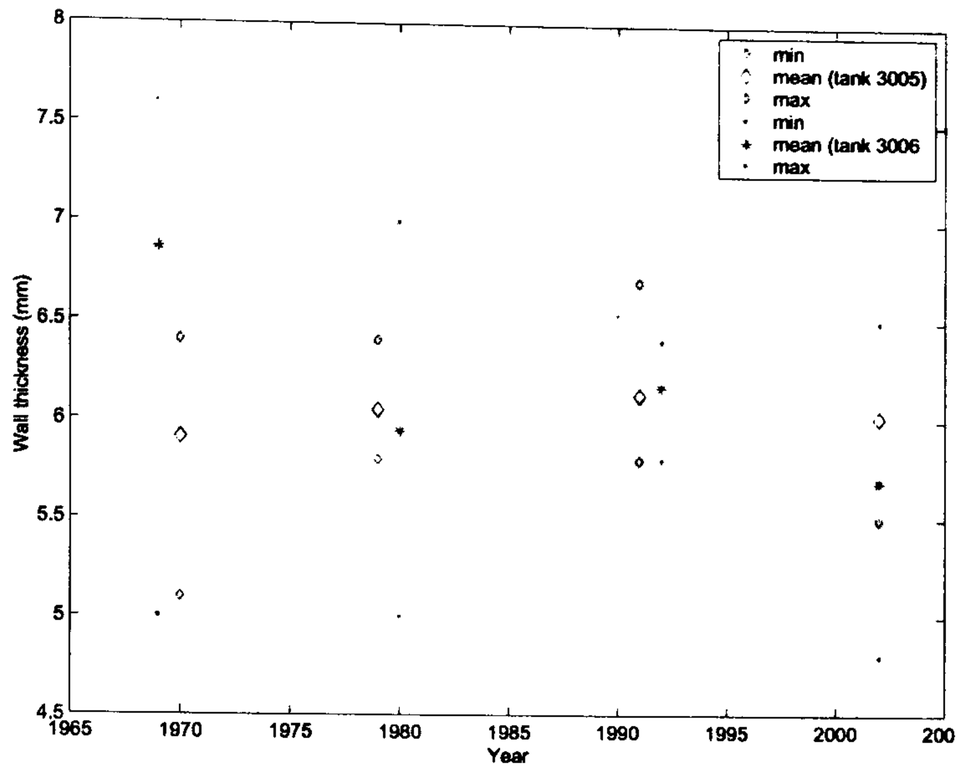


Figure 5.1: Wall thickness summary

Date	Minimum	Mean	Maximum
1970	5.1	5.91	6.4
1979	5.8	6.05	6.4
1991	5.8	6.13	6.7
2002	5.5	6.02	6.8

Table 5.1: Data summary for tank 3005

Date	Minimum	Mean	Maximum
1969	5.0	6.87	7.6
1980	5.0	5.94	7.0
1992	5.8	6.17	6.4
2002	4.8	5.69	6.5

Table 5.2: Data summary for tank 3006

These summaries (Figure 5.1 and Tables 5.1 and 5.2) clearly illustrate many of the problems with these data sets including the limited temporal range, and the effects of changing inspection techniques and non constant observation variances.

A particular problem with these data sets is the lack of spatial information for tank 3005 in the 1991 inspection. If strong assumptions are made about the heights of the inspections and about the points having been recorded in order this leaves only uncertainty about the start point and the direction of inspection. This gives 16 possible permutations. One option is to take the mean of the 16 sets of posterior beliefs to form global posterior

beliefs. Using this approach, a further problem arises because of the eight horizontal locations inspected around the tank, five of these had only two measurements while the remaining three had three measurements. This results in a different variance matrix following each update potentially giving problems associated with averaging variance matrices. A simpler alternative is to update the model using just non-spatial summary statistics from the data. Using these data, given the very limited background information and the extensive changes in inspection technology, the resulting inference would be largely uninformative for any form of analysis used.

These problems emphasise the importance of having a flexible model which separates out the fine-scale description of the spatio-temporal development of the system from the details of how the system's integrity is checked. Current expert beliefs are therefore used to motivate and construct a general model which is sufficiently flexible to meet the modelling and inspection requirements. In this thesis, future inspection schemes are optimised based on current expert opinion about these and other similar tanks and current inspection technologies. The two tanks discussed in this section are only used to provide motivation for the illustrative example which considers the life cycle of a hypothetical new tank.

5.2 An illustrative model

There are many possible correlation structures for the system update variance matrix discussed in Section 2.6. Prior beliefs for θ and β may be based on any valid correlation structure. For simplicity in this example the same correlation structure (see equation 5.1) is used between the system levels ($W_{\theta s}$), the system slopes ($W_{\beta s}$), the prior levels ($C_{\theta 0 s}$) and the prior slopes ($C_{\beta 0 s}$). If more detailed belief structures were elicited for each of the four parts, they would be used instead. In this example the update correlation structure is described by

$$cov(\omega_{\kappa t s}, \omega_{\kappa t s'}) = W_{\kappa s s'} = \tau_\kappa e^{-c_\kappa d_{(s,s')}} \qquad \kappa = \theta, \beta \qquad (5.1)$$

where τ_κ is a measure of temporal strength, $c_\kappa > 0$ is a relative scaling parameter controlling spatial correlation and $d_{(s,s')}$ is a measure of distance between s and s'. The prior covariance is similarly described by

$$cov(\theta_{\kappa 0 s}, \theta_{\kappa 0 s'}) = C_{\kappa 0 s s'} = \tau_{\kappa 0} e^{-c_{\kappa 0} d_{(s,s')}} \qquad \kappa = \theta, \beta \qquad (5.2)$$

where $\tau_{\kappa 0}$ is again a measure of temporal strength and $c_{\kappa 0} > 0$ is a relative scaling parameter controlling spatial correlation. The four structures $W_{\theta s}$, $W_{\beta s}$, $C_{\theta 0 s}$ and $C_{\beta 0 s}$ each then have their own variances controlled by τ_θ, τ_β, $\tau_{\theta 0}$ and $\tau_{\beta 0}$ respectively. Possible ratios between level and slope variance are

discussed in Section 7.4.1. It is assumed that there is no correlation between $\omega_{\theta ts}$, $\omega_{\beta ts}$, θ_{0s} and β_{0s} given $W_{\theta ts}$, $W_{\beta ts}$, $C_{\theta 0s}$ and $C_{\beta 0s}$. Using the model, means, variances and covariances can be calculated for any locations at the current time and in the future.

5.2.1 Model parameters

The parameters used in the model are briefly described. In subsequent sections the sensitivity of the model to changes in these parameters is discussed.

This system is typical of a monitoring based system where significant risk due to corrosion is unlikely and inspection is primarily for reassurance. Hence, within this example, only variance measures are considered. The variance is not affected by the mean wall thicknesses or the corrosion rate. In an application, the initial wall thicknesses and corrosion rate can be combined with a minimum allowed wall thickness to give the expected life of the system.

Observation error (noise), ν_s, is modelled as independently and identically distributed noise in line with common thinking. Correlation between the observation errors could easily be included in the structure if required. The effect of different correlations is considered in Section 7.4.2.

The covariance between any two points (p_i, h_i) and (p_j, h_j) where $h_i =$ height of point i, $p_i =$ clockwise circumference distance from north of point

i and P = the total circumference in the tank may be described using:

$$cov\left(\omega_{\kappa t(p_i,h_i)}, \omega_{\kappa t(p_j,h_j)}\right) = \tau_\kappa e^{-c_\kappa \sqrt{(min\{|p_i-p_j|,||p_i-p_j|-P|\})^2+a^2(h_i-h_j)^2}} \qquad \kappa = \theta, \beta \tag{5.3}$$

and similarly for the prior beliefs with parameters $\tau_{\kappa 0}$ and $c_{\kappa 0}$.

Different beliefs about horizontal and vertical correlation strengths may be reflected by re-scaling the distance metric by changing parameter a in equation 5.3. Following consultation with experts for this example $a = 2\sqrt{2}$. The two key aspects of the problem, the spatial correlation and the temporal correlation, can be controlled by two parameters in equation 5.3: spatial strength (controlled by c_κ) and update variance (controlled by τ_κ). For the purposes of this example, the prior system level variance, $\tau_{\theta 0}$, can be fixed at 10 and the other parameters scaled accordingly using different values of τ_κ.

5.3 Observing minima

Minima data arising from underlying spatio-temporal DLM's is not a well studied problem. The following chapters of this thesis are concerned with the incorporation of such data into Bayesian inspection schemes.

Simple DLM updating can be used where an observation is a true value plus random noise (see Section 2.1). For more complex transforms of the

underlying values, a more sophisticated analysis is required. The main issues that must be addressed when considering observations in the form of regional minima are considered in this section. In the example, a sample that covers only a small fraction of the system will be used as a basis for establishing the entire system's integrity using the spatial correlation between points.

As can be shown by simulation, the distribution of the minimum of n independent N(0,1) draws is skew. Figure 5.2 shows the approximate distribution of the minimum of 10 independent N(0,1) draws based on 10 000 simulations, although in this example the exact distribution my also be calculated.

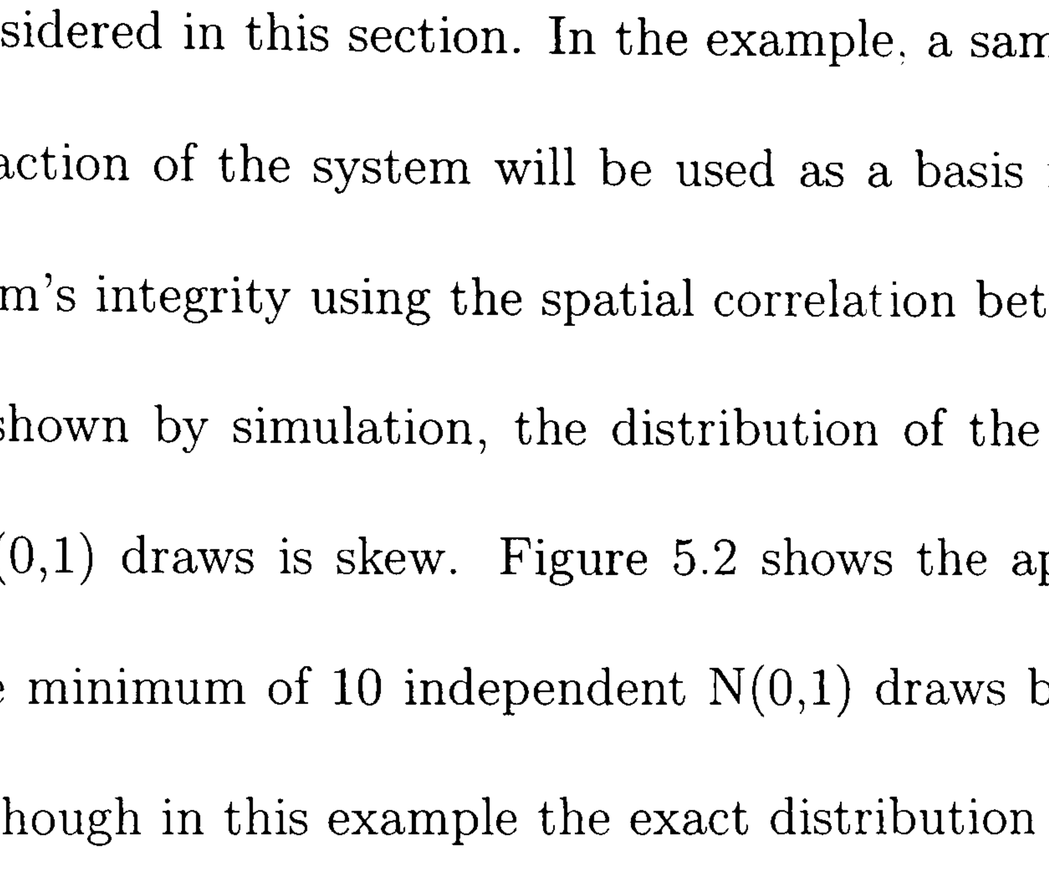


Figure 5.2: Density histogram

The severity of the skewness between correlated data depends on the level

of correlation between the data points. Analytical distributions and partial asymptotic results can be obtained for minima of uncorrelated systems.

In the independent case with n points, the probability of the minimum data point being in the range $[a, b]$ may be calculated as

$$\sum_{i=1}^{n} \int_{a}^{b} p_{x_i}(y) \prod_{j=1,\neq i}^{n} p(x_j \geq y) dy$$

without the complex conditioning necessary when correlated distributions are considered. Results are less tractable for correlated distributions.

The true wall thickness is continuous and this continuum is modelled with a finite set of correlated points. Higher correlations reduce the number of points required to gain a given confidence level in the summary observations. Greater inspection will be required when there is a lower correlation across the system. The accuracy of the approximation made when modelling a continuous system with a discrete grid will increase as a higher density of points is modelled. For any set of parameter values the sensitivity to changes in the density of points should be investigated.

Prior beliefs about the underlying model may be harder to elicit directly from experts when historical data consist of minima because of the difference between what is observed and the underlying model. Simulation may be used to help an expert to check that their beliefs about the underlying model give

realisations in line with their experience.

The lower part of the tank is modelled with n_s equally spaced points horizontally around the tank, n_h equally spaced points vertically in the lower part of the tank and n_t equally spaced time points.

5.3.1 More about minimum

If the minimum of a positively correlated spatio-temporal random process over a grid where there is no observation noise and no trend term is considered, the minimum will be expected to fall between the following two extremes: if correlation is very high, there is effectively a single point random walk which will be on average expected to maintain its initial starting value. If instead all points on the grid are independent then the minimum on average will be expected to become smaller for every subsequent time iteration and, moreover, this change will be expected to become smaller at every step. Intuitively, the more points there are close to the previous minimum, the more the minimum will be expected to decrease at the next iteration, hence an earlier jump will have the biggest decrease in minimum (since following a jump the points will be expected to have higher spread and hence the current minimum point to be more isolated).

These properties can be easily demonstrated using simulation. In Fig-

ure 5.3 the true wall thicknesses of ten independent points are considered over 1000 time points. Starting at time zero with a wall thickness of zero and no prior uncertainty, each of the ten points in turn takes 1000 $N(0,1)$ steps. In Figure 5.3 for the first few iterations many lines can be seen close to the minimum line, while for iterations in the subsequent period up to time 500, particular points can be seen to represent the minima for long series of sequential iterations. Where the minimum line at a particular time is isolated from the other nine lines that line can be seen to remain the minimum. In these circumstances the distribution of the one step ahead minimum will just be $N(0,1)$. Between times 500 and 1000 many lines are close to the minimum and the distribution of future minimum will be less clear.

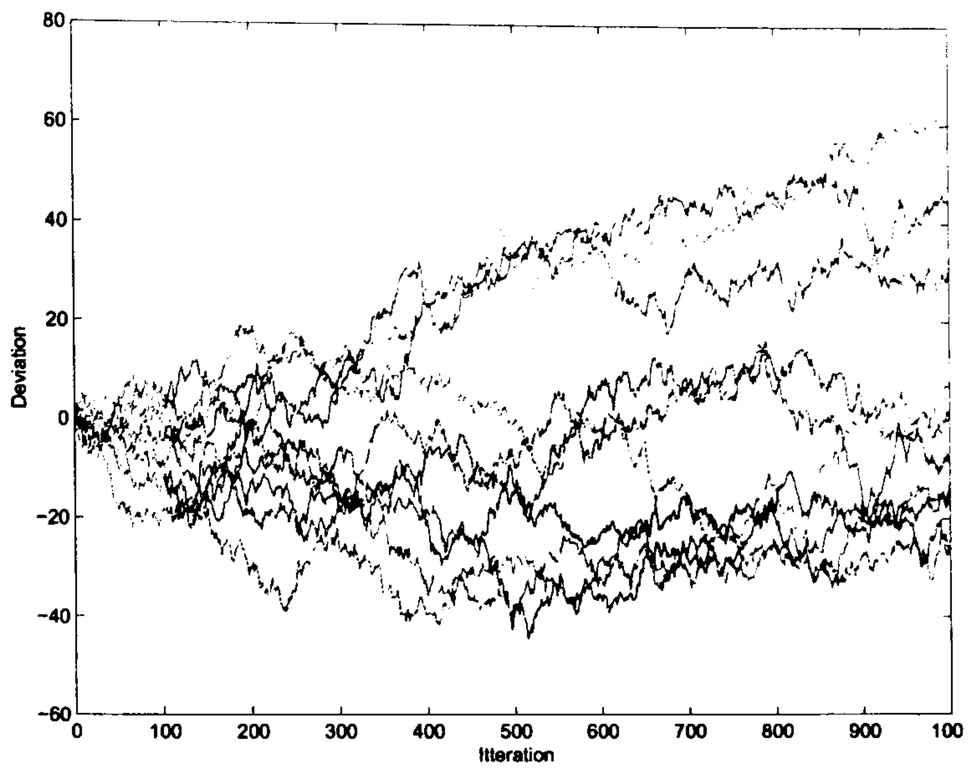


Figure 5.3: Paths plot

Figure 5.4 shows the density of 100 000 independent sequential realisations over fifteen time points. Starting at time zero with a wall thickness of zero and no prior uncertainty, each of the 100 000 points in turn takes fifteen $N(0,1)$ steps. Hence, Figure 5.4 demonstrates the type of distribution that might be expected for a single point.

Figure 5.5 shows the density of the minima where each minimum is over 100 IID points and calculations are based on 1000 simulations. Starting at time zero with a wall thickness of zero and no prior uncertainty, each of the 100 points takes fifteen $N(0,1)$ steps. This is repeated for 1000 minima and hence the density estimated. The minimum can be clearly be seen in each subsequent time to be decreasing but by a smaller amount for each iteration. Hence, Figure 5.5 shows the type of distribution that might be expected for a regional minimum.

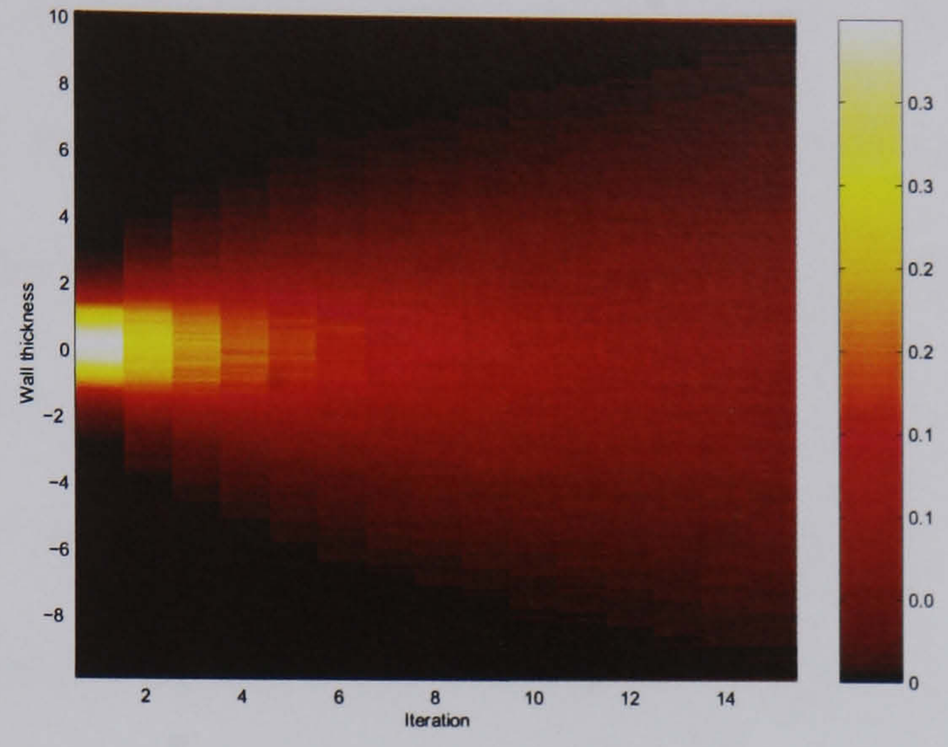


Figure 5.4: Density plot (Single point)

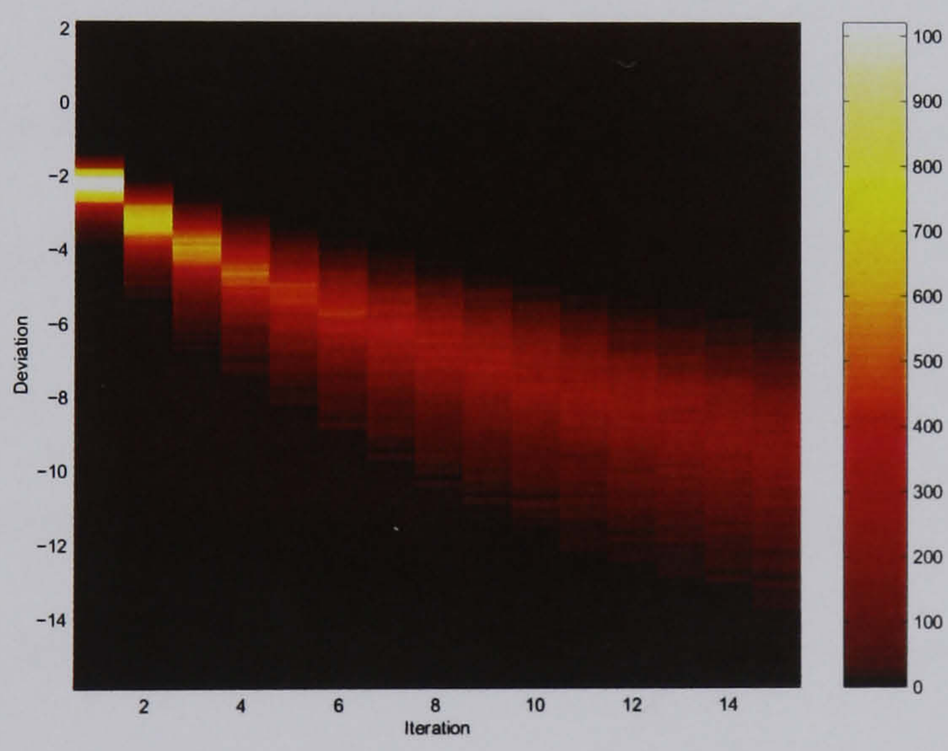


Figure 5.5: Density plot (regional minima)

Transformed observations will generally not be Markov. For example, if global minima are observed these will not be Markov. If there are multiple points bunched around the observed minimum value at time t then the forecast will be different to the situation where the minimum at time t is a relative outlier and it is likely that the same physical point will again be the minimum at time $t+1$ even if that single point increases. In this example following an observation of the minima at time t additional knowledge would change beliefs about the future.

5.3.2 Observational noise

Observation noise may take a range of forms. A simple possibility is for the underlying minima to be observed with observation noise. As an alternative, we consider a regional minimum to be based on the underlying values at each point with its own observation noise, that is on y_{ts} where s covers the region of interest. Observation noise in the context of this example is further discussed in Section 7.4.2.

5.4 Approximating a continuum with a finite grid

Whenever a continuous surface is approximated by a finite grid errors are likely to occur. In this section the effects of different grid intensities are investigated.

The major constraint on the approximating grid is the increase in computational load as the number of points used in the grid increases. With a sufficiently strong spatial correlation a coarse grid will accurately estimate regional and global minima while with a low spatial correlation a much higher grid density will be required.

An observed regional minimum is defined as the minimum of a subsection of the grid at a particular time which includes point by point IID observational noise. The observed regional minimum at time t with bottom left location $s = (p_i, h_i)$ may be written as $\mu_{yts} = min_{i,j} y_{t(i,j)}$ where

$$i \in p_i, p_i + 1, \ldots, p_i + k_{p_i} - 1$$

and

$$j \in h_i, h_i + 1, \ldots, h_i + k_{h_i} - 1$$

where y_{ts} is defined in equation 2.5, and the subsection of the grid consid-

ered has k_{p_i} horizontal points and k_{h_i} vertical points. Without observational noise it becomes an underlying regional minimum. The underlying regional minimum at time t and bottom left location s may similarly be written as $\mu_{\theta ts} = min_{i,j}\theta_{t(i,j)}$. Using the covariance structure defined in equation 5.3 with temporal strength $\tau_\kappa = 1$ and spatial strength $c_\kappa = \frac{0.001}{2}$ a 10 by 20 grid is considered. Data may be simulated over the entire grid and the minimum value calculated based on different intensities of observation, based upon the same data for each inspection scheme. The mean of these minima over a set of simulations may be compared allowing the effect of different intensities of observation to be seen over a single time point.

Nine different intensities of observation are considered in Figure 5.6 and further described in Table 5.3.

Scheme	Description	mean minimum
A	10 by 20 grid	-0.1819
B	10 by 10 grid	-0.1706
C	10 by 7 grid	-0.1686
D	10 by 5 grid	-0.1536
E	10 by 3 grid	-0.1514
F	5 by 10 grid	-0.1541
G	5 by 5 grid	-0.1365
H	3 by 3 grid	-0.1367
I	4 corners plus middle	-0.1229

Table 5.3: Brief description of schemes

The third column of Table 5.3 shows the mean area minimum value of

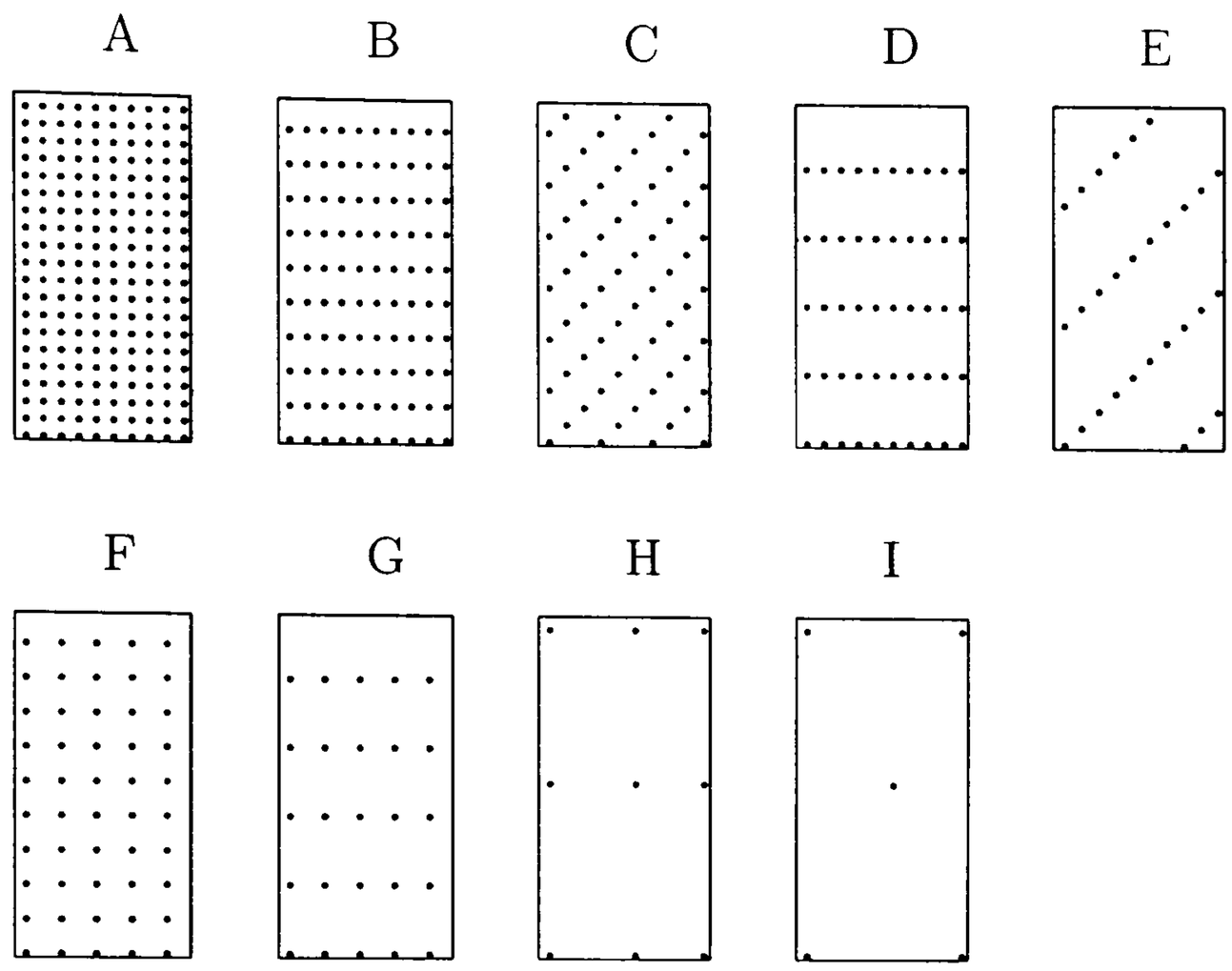


Figure 5.6: Intensities of observation

each scheme calculated using 40 000 simulations of the region. There is a clear trend that making more inspections decreases the expected minimum. Each mean has a standard deviation of about 0.002 indicating further that the differences between the schemes are significant.

The effect over time may also be considered. The prior level variance $\tau_{\theta 0}$, and prior slope variance $\tau_{\beta 0}$ are set at 10 and 0.001 respectively. The update slope variance $\tau_{\beta} = 0.001$ while the update level variance $\tau_{\theta} = 1$. Prior means for the level and slope are both zero. The choice of these parameter values are further discussed in Chapter 7. Table 5.4 shows the mean area minimum value of each scheme for each time calculated over 40 000 simulations. The

Scheme	1	2	3	4	5	6	7	8
A	-0.7764	-0.9839	-1.1205	-1.2543	-1.3552	-1.4752	-1.5805	-1.6665
B	-0.7267	-0.9215	-1.0488	-1.1748	-1.2677	-1.3811	-1.4785	-1.5575
C	-0.7178	-0.9117	-1.0381	-1.1619	-1.2531	-1.3640	-1.4616	-1.5397
D	-0.6496	-0.8265	-0.9395	-1.0516	-1.1322	-1.2360	-1.3239	-1.3938
E	-0.6391	-0.8139	-0.9249	-1.0351	-1.1138	-1.2155	-1.3026	-1.3711
F	-0.6530	-0.8306	-0.9438	-1.0575	-1.1385	-1.2415	-1.3293	-1.3995
G	-0.5739	-0.7333	-0.8317	-0.9315	-0.9993	-1.0922	-1.1707	-1.2301
H	-0.5737	-0.7357	-0.8343	-0.9333	-1.0024	-1.0935	-1.1714	-1.2322
I	-0.5114	-0.6598	-0.7456	-0.8348	-0.8939	-0.9746	-1.0445	-1.0987

Table 5.4: Mean minimums for times 1 to 8 for schemes A to I

ratios between the schemes remain very similar.

The extensive changes in minimum value between the schemes suggest that for this spatial correlation strength regional minima values suggested by any of these simulation densities may be too optimistic. In an application to any particular problem once design choices are made a high intensity simulation may be run to obtain accurate forecasts.

Chapter 6

Bayes linear forecasts based on minima observations

In this thesis we aim to develop efficient inspection schemes for reducing uncertainty about system corrosion with minimum inspection cost. To identify efficient inspection schemes, many candidate schemes must be compared, where both the number of times at which to inspect, and also the number of locations to inspect at each time point may be varied. A full Bayes pre-posterior evaluation of the reduction in uncertainty associated with any one inspection scheme based on data in the form of regional minima would be enormously computationally intensive. Where a range of design choices are to be considered, it is important to use an approach which can be evaluated reasonably quickly for each possible design and the Bayes linear approach is well suited to such problems.

Analysing design choice problems based on transformed data is challenging because transformed observations are often associated with distributions which are analytically intractable. The correlations between observed (transformed) values and transformations of the system values of particular interest can be calculated using simulation. Beliefs about these system-based transformations are then updated using a Bayes linear framework. Critical issues, such as identifying when and where future inspections should be made, and questions about the confidence in the system's integrity in the future can be addressed within this framework.

In the first section of this chapter two different approaches to updating the model when observations are a function of the underlying values are considered. In section two we discuss an alternative approach to simulating the underlying system especially suited to large problems with many spatial points. In section three the Bayes linear updates based on full data, regional minima data and global minimum data are compared to a full Bayesian update based on full data.

6.1 Updating beliefs

In this section the use of transformed observation values at time t to forecast future transformed output (for example, regional minima) at time $t + k$ is discussed.

Using a fully Bayesian update for minima data would be very computationally demanding. The Bayes linear approach uses the simple updating equations (2.2 and 2.3) to calculate the adjusted expectation and the adjusted variance. In particular the adjusted variance depends only on the locations of the observations and not on the observed values and so can be used as the basis of a tractable design methodology.

Using the Bayes linear approach, the forecast values for future observations and any underlying system parameters (for example true regional minima) are adjusted directly following an observation. The covariance between all the observed data values and all other system quantities of interest at all times must be calculated (see Section 2.7). There is no requirement to have the same method of data collection or the same measurement variance at each time point. However, the transformations and the variances for all times must be known.

Two distinct approaches to this problem are considered:

Forecasting through the underlying model

Beliefs about all aspects of the future state of the system are updated through the underlying model. An observation at time t will be a transform of the underlying values (θ_t), for example the regional minima after IID observation noise. First, following an observation at time t, the underlying beliefs at time t (θ_t) are updated using this observation ($f_t(\theta_t)$) and all preceding observations. This update is made by simulating from the underlying model to find the distribution of the transformed observations. The covariance between the observed values ($f_t(\theta_t)$) and the underlying model (θ_t) can then be estimated. Beliefs about the underlying model at time t can then be updated in the light of the transformed observation using a Bayes linear adjustment. The underlying model may then be stepped forward to time $t+k$ based on the updated beliefs at time t. Inferences ($f_{t+k}(\theta_{t+k})$) can then be made by first simulating from the underlying system at time $t+k$ and then calculating the data for any function of interest (for example the underlying global minimum at time $t+k$). The distribution of any functions of interest may then be estimated based on a sufficiently large number of simulations. This direct approach is illustrated in Figure 6.1.

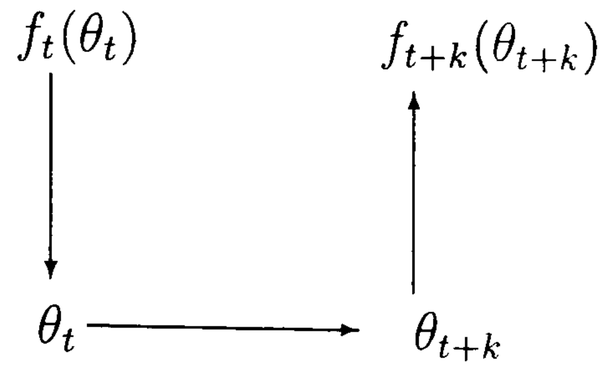


Figure 6.1: Forecasting through the underlying model

Forecasting directly

Forecasts are made directly from the observation to the parameters over which inferences are required (for example the global minima without observation noise).

The covariance between $f_t(\theta_t)$ and $f_{t+k}(\theta_{t+k})$ can then be calculated either directly or using simulation. The direct approach will only be possible in some simple cases. The simulation approach can be used more generally. Underlying system data is simulated for times t and $t+k$ (θ_t and θ_{t+k}). This may either be done using the covariance between them calculated directly using matrix 2.4 or using a series of simulated underlying data. When using the simulation approach, underlying data is first simulated at time t and then k sequential updates of the underlying model are simulated and added on giving data at times t and $t+k$. Transforms of the underlying data can then be made to get a simulated realisation of the observation at time t

and the quantity of interest at time $t + k$. A suitable update, for example equations 2.2 and 2.3, can then be used to adjust the forecast mean and variance at time $t + k$ given the observations up to time t.

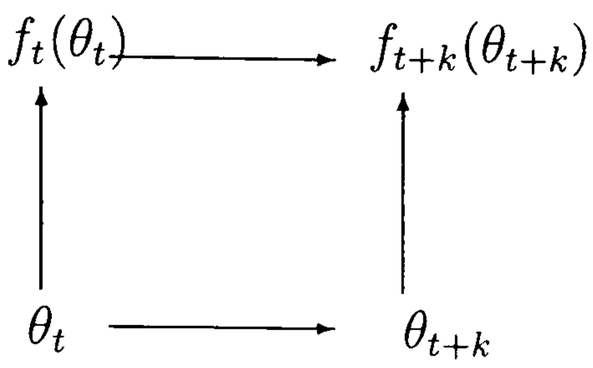


Figure 6.2: Forecasting directly

The direct approach might be expected to be more accurate than the indirect approach the only approximations being due to simulation error. The indirect approach, however, has the advantage of updating the underlying model which allows fast and efficient updating in the future. The most suitable approach will depend on the situation. Simulation can be used to evaluate the merits of the different techniques. If point data (y_{ts}) were observed both direct and indirect methods would be expected to produce equivalent predictions. In this thesis attention is on short time series and inference over transformations of the underlying system values. Hence, attention is focused on the direct approach because a higher accuracy is expected and high computational efficiency for long temporal systems is not an important concern.

6.2 Further simulation approaches

Section 2.7 contains a general discussion about simulation of entire systems. In this section further use of simulation to find more difficult covariance structures is discussed.

6.2.1 The partial simulation approach

If the problem is too large for direct calculation of the covariance matrix from simulated data, the symmetry of the problem may be exploited. In particular, a full Bayesian update would be computationally intractable for large problems, however, knowledge of the covariances is sufficient when using a Bayes linear update. If we have uniform beliefs about uncertainty across the entire vessel and in the one-step-ahead forecast, the covariance between any pair of minima which are the same spatial distance apart at fixed times will be the same. That is, $g(d, t, t') = cov(\mu_{\theta ts}, \mu_{\theta t's'})$ where d is the distance between s and s', and similarly for $cov(\mu_{yts}, \mu_{yt's'})$ and $cov(\mu_{yts}, \mu_{\theta t's'})$. A single calculation of this covariance can then be used for all observed regional minima or underlying regional minima between two regions at time t and t' that are d spatial units apart. This can then be repeated for all d, t and $t' \geq t$ and hence the full covariance structure found. While this is fast and efficient

for large problems (e.g. if all 640 possible minima must be considered) due to simulation error the resulting covariance matrix may not be positive definite.

For smaller problems where full data may be simulated simultaneously across all parts of the system the inferred correlation structure will be positive definite. In the partial simulation approach the calculation of the covariance between two minima via simulation results in errors in the covariance structure possibly resulting in the covariance structure not being positive definite. Increasing the number of simulations helps to reduce this problem. Positive definiteness is equivalent to all eigenvalues being positive. Negative eigenvalues can only be due to rounding errors and will be relatively small for sufficiently large simulations. Covariance matrices may be reconstructed with the negative eigenvalues removed. Comparing the matrices generated using 2000 and 20 000 simulations to that generated using 200 000 simulations relative max/min error is about 10% and 5% respectively while the mean value is fairly accurate for all simulation sizes. The critical number of simulations is best assessed in consideration with its effect on the final decision given the parameter values. Reconstructing the matrix from its eigen-components having removed any negative eigenvalues destroys the symmetry between equally spaced points. By averaging over all rearrange-

ments that maintain equivalent estimates in equivalent positions a positive definite covariance matrix where equally spaced points have equal covariance may be constructed. For example, a four by four positive definite matrix has a symmetric estimate

$$a = \begin{pmatrix} a_{11(0)} & a_{21(1)} & a_{31(2)} & a_{41(1)} \\ a_{12(1)} & a_{22(0)} & a_{32(1)} & a_{42(2)} \\ a_{13(2)} & a_{23(1)} & a_{33(0)} & a_{43(1)} \\ a_{14(1)} & a_{24(2)} & a_{34(1)} & a_{44(0)} \end{pmatrix}$$

$$\frac{1}{8}\left\{ \begin{pmatrix} a_{11(0)} & a_{21(1)} & a_{31(2)} & a_{41(1)} \\ a_{12(1)} & a_{22(0)} & a_{32(1)} & a_{42(2)} \\ a_{13(2)} & a_{23(1)} & a_{33(0)} & a_{43(1)} \\ a_{14(1)} & a_{24(2)} & a_{34(1)} & a_{44(0)} \end{pmatrix} + \begin{pmatrix} a_{44(0)} & a_{14(1)} & a_{24(2)} & a_{34(1)} \\ a_{41(1)} & a_{11(0)} & a_{21(1)} & a_{31(2)} \\ a_{42(2)} & a_{12(1)} & a_{22(0)} & a_{32(1)} \\ a_{43(1)} & a_{13(2)} & a_{23(1)} & a_{33(0)} \end{pmatrix} \right.$$

$$+ \begin{pmatrix} a_{33(0)} & a_{43(1)} & a_{13(2)} & a_{23(1)} \\ a_{34(1)} & a_{44(0)} & a_{14(1)} & a_{24(2)} \\ a_{31(2)} & a_{41(1)} & a_{11(0)} & a_{21(1)} \\ a_{32(1)} & a_{42(2)} & a_{12(1)} & a_{22(0)} \end{pmatrix} + \begin{pmatrix} a_{22(0)} & a_{32(1)} & a_{42(2)} & a_{12(1)} \\ a_{23(1)} & a_{33(0)} & a_{43(1)} & a_{13(2)} \\ a_{24(2)} & a_{34(1)} & a_{44(0)} & a_{14(1)} \\ a_{21(1)} & a_{31(2)} & a_{41(1)} & a_{11(0)} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{44(0)} & a_{34(1)} & a_{24(2)} & a_{14(1)} \\ a_{43(1)} & a_{33(0)} & a_{23(1)} & a_{13(2)} \\ a_{42(2)} & a_{32(1)} & a_{22(0)} & a_{12(1)} \\ a_{41(1)} & a_{31(2)} & a_{21(1)} & a_{11(0)} \end{pmatrix} + \begin{pmatrix} a_{11(0)} & a_{41(1)} & a_{31(2)} & a_{21(1)} \\ a_{14(1)} & a_{44(0)} & a_{34(1)} & a_{24(2)} \\ a_{13(2)} & a_{43(1)} & a_{33(0)} & a_{23(1)} \\ a_{12(1)} & a_{42(2)} & a_{32(1)} & a_{22(0)} \end{pmatrix}$$

$$+ \left. \begin{pmatrix} a_{22(0)} & a_{12(1)} & a_{42(2)} & a_{32(1)} \\ a_{21(1)} & a_{11(0)} & a_{41(1)} & a_{31(2)} \\ a_{24(2)} & a_{14(1)} & a_{44(0)} & a_{34(1)} \\ a_{23(1)} & a_{13(2)} & a_{43(1)} & a_{33(0)} \end{pmatrix} + \begin{pmatrix} a_{33(0)} & a_{23(1)} & a_{13(2)} & a_{43(1)} \\ a_{32(1)} & a_{22(0)} & a_{12(1)} & a_{42(2)} \\ a_{31(2)} & a_{21(1)} & a_{11(0)} & a_{41(1)} \\ a_{34(1)} & a_{24(2)} & a_{14(1)} & a_{44(0)} \end{pmatrix} \right\}$$

A possible problem with removing the negative eigenvalues is for example, if the off diagonal terms become larger than the diagonal terms. As an alternative, positive definiteness may also be guaranteed if the diagonal is sufficiently large. Increasing the values on the diagonal has the effect of increasing the variance of each point and may be considered a less risky procedure since it reduces the spatial correlations.

6.2.2 Data with its spatial structure removed

A particular extension to the simulations discussed earlier and particularly relevant to the current data set is when the spatial location of collected data is unknown. Simulation may be used to calculate covariances by removing the spatial association between the underlying data and the observations made. One possible approach is to calculate the covariance structure between ranked observations and future uncertainties. Beliefs about future uncertainties may then be updated to reflect the observed data (after it also has been ranked).

6.3 Comparing full Bayesian modelling to the Bayes linear update

Bayes linear calculations allow much faster and less computationally intensive calculation than full Bayesian calculations. In this section the Bayes linear approach with a direct forecast when regional and global minima are observed is compared to the full Bayes model when full observations are made to see how much information is lost. A simple example is considered where calculations are not too computationally demanding. A full inspection at times $t = 1, \ldots, 4$ and a subsequent full Bayesian update of beliefs at time $t = 6$ is used to compare the minimum variance achievable in the prediction

of the global minimum, and compared to the prediction error when either full data, regional minima or global minima are observed at times $t = 1, \ldots, 4$ and updates of beliefs at time $t = 6$ are made via a Bayes linear approximation.

Because of the high computational load associated with full Bayesian calculations attention is restricted to a small problem. A four by four grid of 16 points is considered with four regions based on the four quadrants, see Figure 6.3.

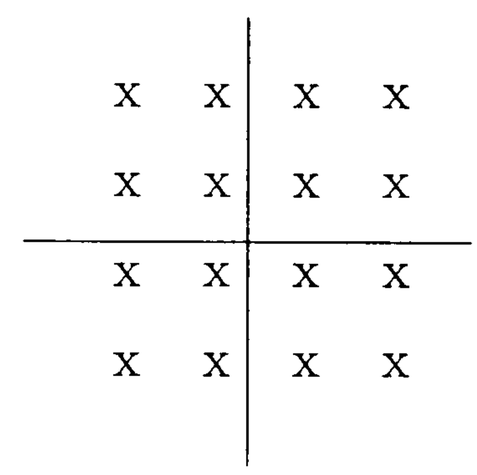


Figure 6.3: Simple example

A linear trend is assumed for the underlying model (see equation 2.1). Prior level and slope values are not important because attention is focused on the forecast variances not means. No prior uncertainty about the level and slope is assumed (that is, prior uncertainty = 0). Update covariances follow $W_\theta(s, s') = \tau e^{-cd}$ where $d_{(s,s')}$ is a straight line measure of distance between the two locations s and s'. Level variance and slope variance have the same correlation structure with $W_\beta = rW_\theta$. Observational errors V are

IID $N(0, \nu^2)$.

For this theoretical example set $\tau = 0.05$, $c = \sqrt{0.5}$, $\nu^2 = 0.01$ and $r = 1/4$.

Figure 6.4 shows the resulting covariance structure between the 16 points, the 4 regional minima and the global minimum at time one for these parameters before observing any data.

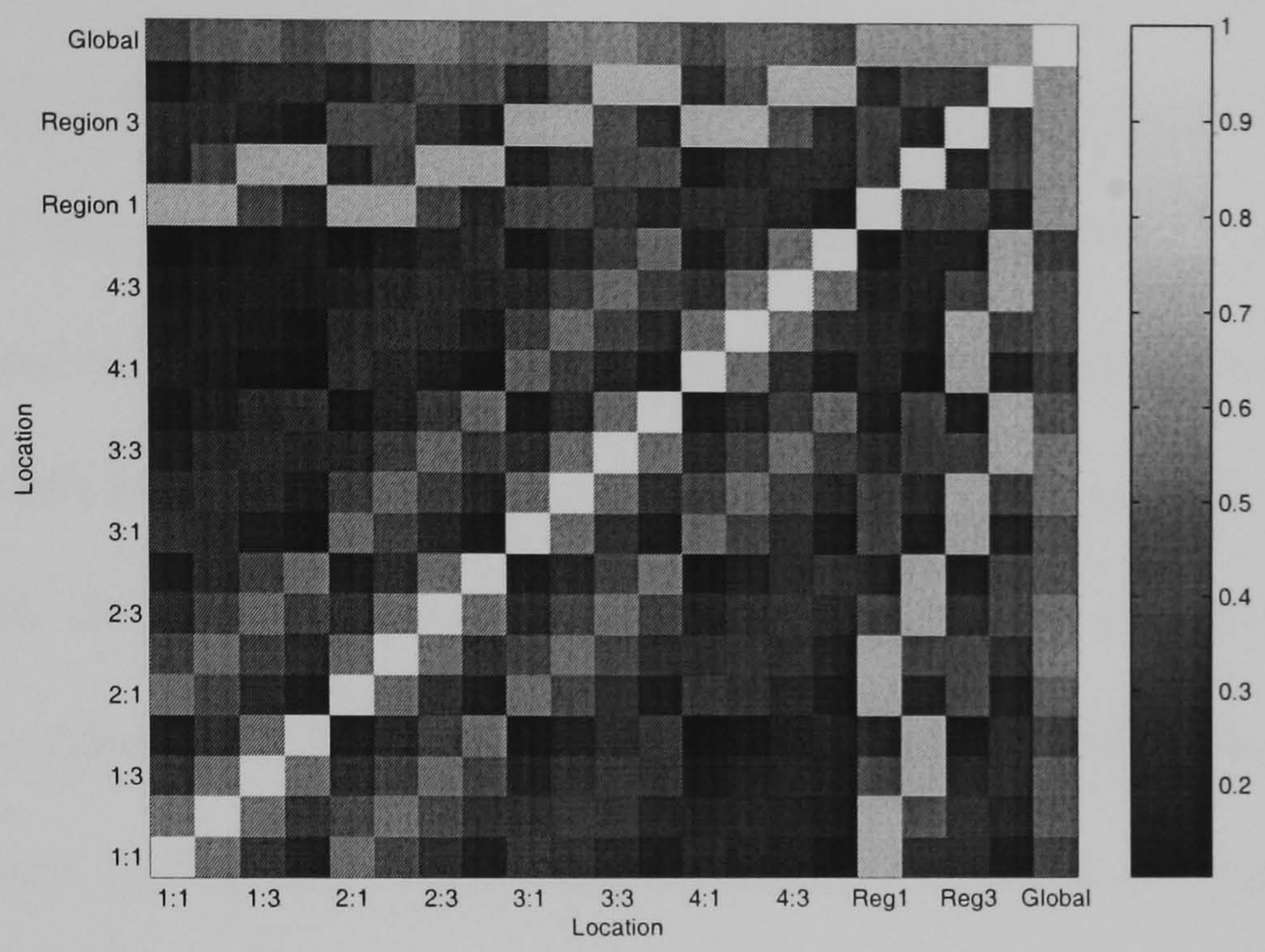


Figure 6.4: Correlation at time 1

Data is simulated sequentially over the 16 spatial points for the six time points. Each sequential step is based on a draw from the level update variance ($W_{\theta s}$) and a further draw from the slope update variance ($W_{\beta s}$). The four regional minima and the global minimum are then calculated from the spatial points for each time. The covariance of this data is then calculated over the 50 000 simulations made. Following an observation of simulated data up to time four, these beliefs must be updated. When using a Bayes linear update having observed full data, regional minima data or global minima data, belief updates follow directly from equations 2.2 and 2.3. When making a full Bayesian update based on full point data up to time four the closed form updates may be used to update beliefs about the underlying system values at time six (or equivalently a Bayes linear update). Simulating from the resulting multivariate normal distribution for the underlying system values repeatedly allows beliefs about the global minimum at time six to be calculated without recourse to any simplifications. The update summaries are based on 4000 simulations and each of these contains a further 10 000 simulations to make the fully Bayesian update of beliefs about the global minimum at time six.

In Figure 6.5 the global minimum for 1000 simulations of the 4 by 4 grid

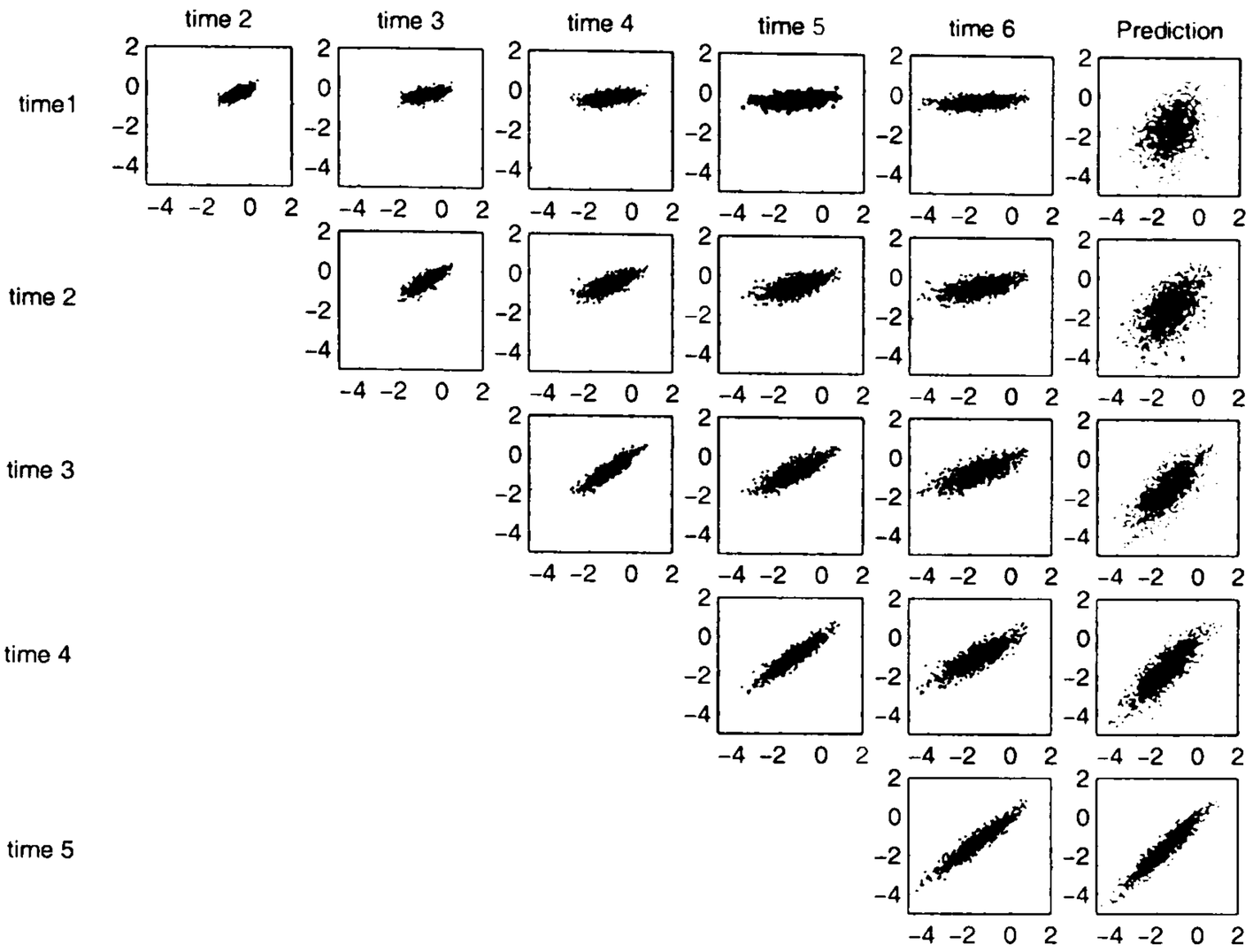


Figure 6.5: Matrix of simulations

are plotted over the six time points considered. The linear association between the observation at time four and the observation at time six makes the Bayes linear approach well suited. The one-step-ahead linearity becomes more pronounced in later forecasts. The sixth column shows the linear forecast at time six based on least squares using all data up to the given time. Linearity is much in evidence.

Applying a full Bayesian approach to the minima observations would be more computationally demanding and would naturally yield no better predictions than making full observations. Hence for example, if regional

minima allow accurate prediction using a Bayes linear approximation, then a full update based upon regional observations is unlikely to be significantly better.

b, ν^2		Full Observation with full Bayes update	Full Observation	Regional Minima only	Global Minima only
			Bayes linear update		
$\sqrt{0.5}$, 0.01	Forecast var	0.3937	0.4041	0.4085	0.4168
	Increase in var		2.64%	3.76%	5.87%

Table 6.1: Changes in variance for different schemes

Table 6.1 shows how doing a full Bayesian update in this example reduces the variance of the global minimum at time six when full data is observed by 2.64% more than a Bayes linear model. Such a small decrease in variance may be felt not to warrant the vast increase in computational resource required especially as the problems get larger and more complex. When making a Bayes linear adjustment of forecast variance using only the observed regional minima of the four quadrants or just the global minima for times one to four slightly larger variances about the forecast global minimum at time six result. For this example there is only a 3.76% increase in variance associated with observing regional minima rather than full data and doing a full Bayesian update.

The sensitivity of the model may be assessed by changing the input pa-

rameters. The parameters of the model may be changed to have a higher observational variance and lower spatial correlation, for example, $\tau = 0.05$, $c = \sqrt{4}$, $\nu^2 = 0.1$ and $r = \frac{1}{4}$. The parameters may also be changed to have a very low observational variance and high spatial correlation, for example, $\tau = 0.05$, $c = \sqrt{0.1}$, $\nu^2 = 0.0001$ and $r = \frac{1}{4}$. The result using these two sets are shown in Table 6.2. Again the change in variance is fairly small suggesting the Bayes linear approach is a reasonable approximation to a full update for this example. In these cases, when making a Bayes linear adjustment of forecast variance using only the observed regional minima or the global minima slightly larger variances about the forecast global minimum again result at time six. As expected higher spatial correlations and lower observation noise result in more confident forecasts.

b, ν^2		Full Observation with full Bayes update	Full Observation	Regional Minima only	Global Minima only
			Bayes linear update		
$\sqrt{(4)}$, 0.1	Forecast var	0.3732	0.4019	0.4069	0.4138
	Increase in var		7.69%	9.03%	10.88%
$\sqrt{(0.1)}$, 0.0001	Forecast var	0.4268	0.4283	0.4338	0.4415
	Increase in var		0.35%	1.64%	3.44%

Table 6.2: Sensitivity of changes in variance

Chapter 7

Finding good inspection schemes

In this chapter possible criteria are considered which allow comparison between different minima inspection schemes. Several possible ways of reducing the number of candidate schemes are also discussed. A range of parameter choices are considered and some of the resulting schemes compared. Some of the candidate spatial inspection schemes are then compared to a non-spatial fixed point inspection scheme with comparable inspection burden. In the final section the use of ranked regional minimum data for prediction of the global minimum is discussed.

In section one of this chapter regional minima are described within the context of the tanks introduced in Chapter 5. In section two criteria which may be used to compare different inspection schemes based on reducing un-

certainty are introduced, while in section three a range of search strategies for finding good inspection schemes out of the many that are possible are considered. In section four we apply the criteria and search strategies to the example and investigate the effect of observational noise. In the fifth section a wide range of inspection schemes are compared for different parameter values, while in section six the spatial inspection scheme is compared to the non-spatial equivalent over the same range of parameter values. In section seven we discuss predicting the global minimum from the regional minima observation.

7.1 Region minima in this example

For the example considered in this thesis where interest is in design choices a low intensity scheme must be considered. Scheme G (see Chapter 5) is used with $n_s = 640$, $n_h = 5$ and $n_t = 8$ (see Section 5.3 for more details). A regional minimum is thus defined as the minimum of a 5 (horizontal) by 5 (vertical) subsection of the grid at a particular time. This gives a maximum of 640 possible different regional minima over the tank for any inspection. The observed regional minimum at time t and bottom left location $s = (p, 1)$ may be written as $\mu_{yts} = min_{i,j} y_{t(i,j)}$ where $i = p, p+1, \ldots p+4$ and $j =$

$1, 2, \ldots, 5$ where as before y_{ts} is defined in (2.5). Similarly the underlying regional minimum at time t and bottom left location s may be written as $\mu_{\theta ts} = min_{i,j}\theta_{t(i,j)}$.

The effect on the estimated regional minimum value of increasing the density of points in each region should be carefully considered. An acceptable approximation of the continuous surface should not be sensitive to small changes in this density. This is discussed in more detail in Section 5.4.

The Bayes linear approach uses the simple updating equations (introduced in Section 2.2) to calculate the adjusted expectation

$$E_{D_t}(\phi) = E(\phi) + cov(\phi, D_t)var(D_t)^{\dagger}(D_t - E(D_t)) \tag{7.1}$$

and the adjusted variance

$$var_{D_t}(\phi) = var(\phi) - cov(\phi, D_t)var(D_t)^{\dagger}cov(D_t, \phi) \tag{7.2}$$

where $\dagger$ is a generalised inverse and ϕ is a vector whose values it is intended to learn about via D_t the vector of all observed data up to time t. In this application D_t is a vector of observed regional minima $\mu_{yt_is_j}$ while ϕ is a vector of underlying regional minima $\mu_{\theta t_i s_j}$. In particular the adjusted variance $var_{D_t}(\phi)$ depends only on the observation locations and not on the observed values themselves and so can be used as the basis of a tractable

design methodology.

For the level of detail described in this problem, a full simulation of data across the entire observed and underlying systems is practically possible but very slow. For this example only 64 regional minima are considered evenly spaced horizontally around the tank. See Figure 7.1. This means only half of the horizontal points on the grid are considered in sets of five, with five points observed and five points unobserved. These 64 regional minima are assumed to have horizontal left most points at 1, 11, ..., 631 and are respectively indexed by $\psi_1, \psi_2, \ldots, \psi_{64}$ which form a set Ψ. Henceforth only the observed regional minima $\mu_{yt\psi_i}$ and underlying regional minima $\mu_{\theta t\psi_j}$ where $\psi_i, \psi_j \in \Psi$ are considered. This smaller number of minima also reduces the computational load of subsequent searches to find the optimal set of inspection locations (see Section 7.3). These searches are very demanding when there is a very large number of possible combinations to consider. Once design choices have been made, these intermediate regions should be analysed within the underlying model, if computationally possible, to check for unacceptably high posterior variance (see section 7.6.2).

For this example we simulate the correlation structure required in order to evaluate equations 7.1 and 7.2 as follows. For each simulation a set of initial

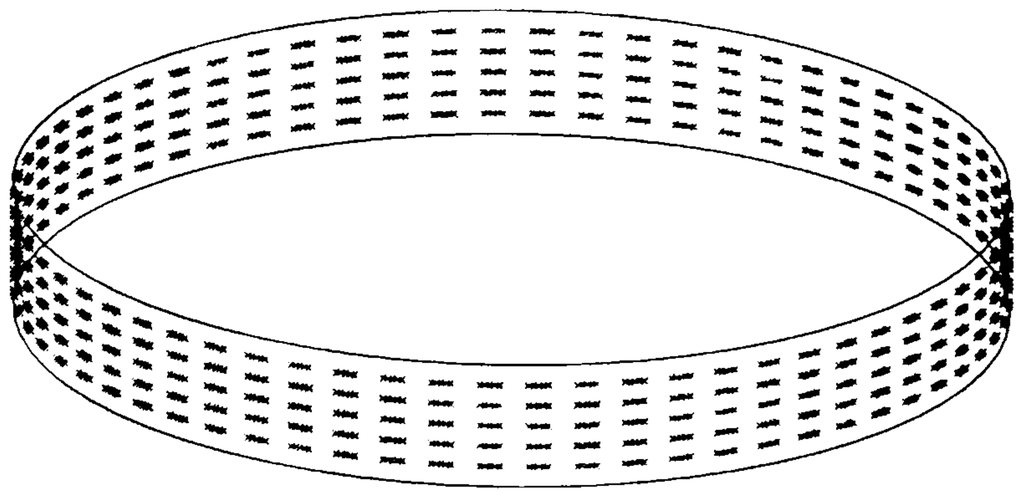


Figure 7.1: 64 regions of 25 points considered within the model

wall thicknesses and corrosion rates are simulated in accordance with prior beliefs θ_{0s} and β_{0s}. Level and slope updates ($\omega_{\theta ts}$ and $\omega_{\beta ts}$) are then simulated for times 1 to 8. The prior and update simulations are then combined to get true wall thicknesses θ_{ts} at the 1600 locations modelled over the eight times. The 64 underlying regional minimum wall thicknesses ($\mu_{\theta t\psi_i}$) and the 64 observed regional minimum wall thicknesses ($\mu_{yt\psi_j}$) are calculated for each time (the latter by adding the IID observation noise ν_{ts} to the 1600 underlying wall thicknesses). Finally the correlation structure between the underlying regional minima and observed regional minima wall thicknesses across times 1 to 8 are calculated over 40 000 simulations of this data. The covariance structure between the 64 minima at time eight is shown in Figure 7.2.

The simulation error is most easily estimated by considering the 64 regions at a single time where, before any data is observed they should all have the same variance. The variance of this variance is due to the simulation error.

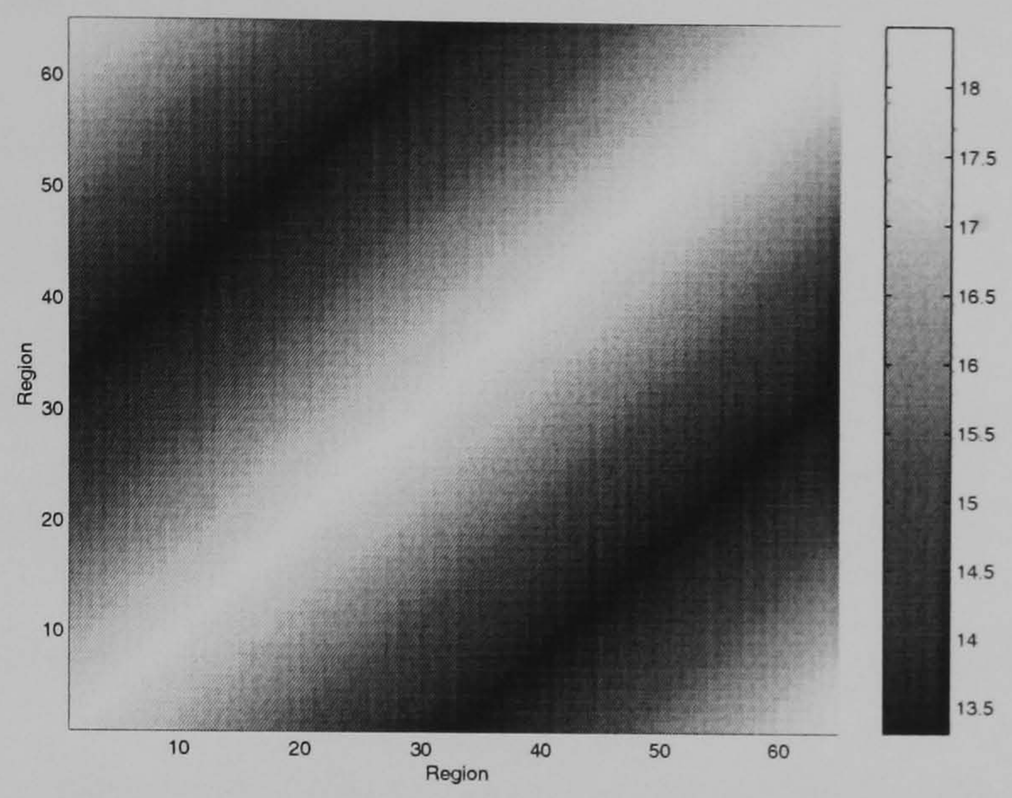


Figure 7.2: Covariance between the 64 regions at time 8

Figure 7.3 shows the variance for each of the 64 regions for times one to nine for the underlying (the lower set of lines) and the observed variance (the upper set of lines).

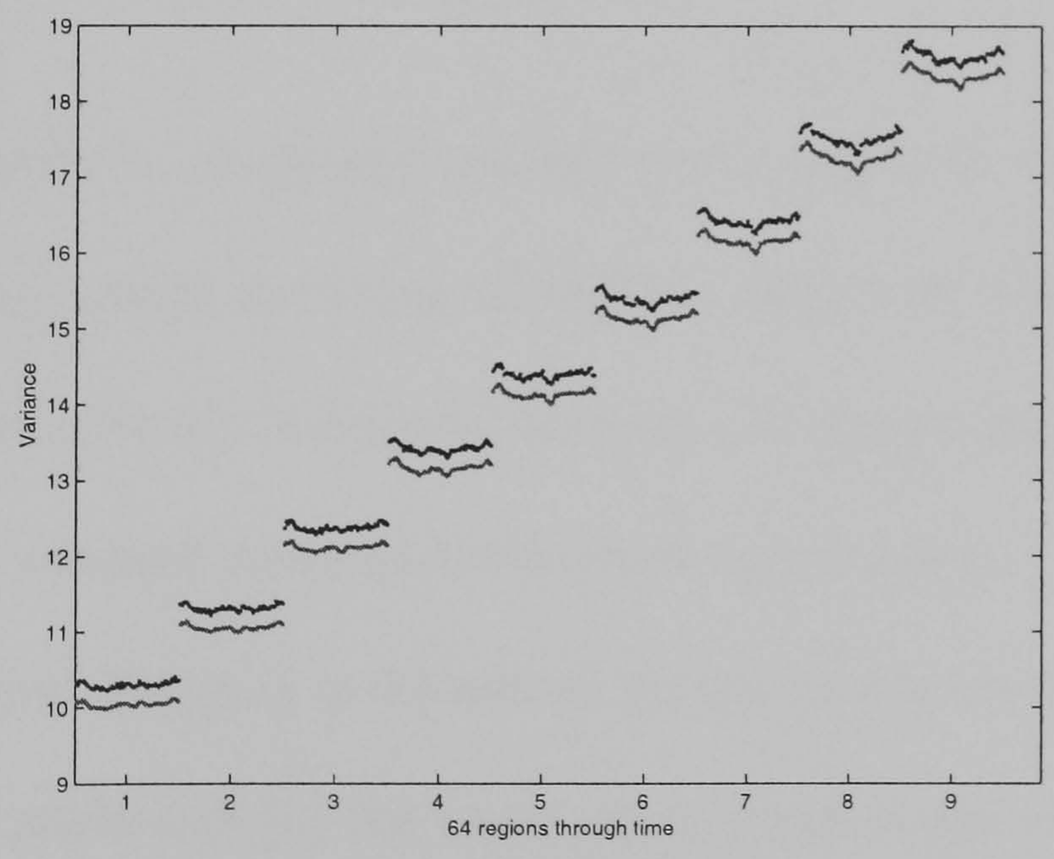


Figure 7.3: Variance of the 64 regions through time

The standard deviation of the measures for a particular time point for the observed regional minima range from 0.33 to 0.93 with a general tendency to be larger for later time points. This level of variance may be considered quite large and for a full application further increasing the number of simulations may be necessary to reduce this variance.

The variation in the covariances that are known to be equivalent is of the order of 0.1%, hence the strong diagonal lines in Figure 7.2. A possible refinement would be using a mean of these theoretically equal values following a similar approach to that suggested in Section 6.2.1. In addition to reducing variance this would also create the correct symmetric structure.

7.2 Criteria for comparison

Having calculated the covariance between every observed regional minimum and underlying regional minimum of interest across all times, the effect of making any particular set of inspections on the uncertainties across the entire system may be assessed using updates based on equations 7.1 and 7.2.

The inspection design Δ is defined by (i) the vector m_t of times at which inspections are made and (ii) the locations at which inspections are made at time t, $\psi_{t_1}, \ldots, \psi_{t_{m_{nt}}}$, where $t_1, \ldots, t_{m_{nt}}$ are m_{nt} choices from the index set

$[1, \ldots, 64]$. In this example, m_{nt} is restricted to be the same value m_n for each time at which inspections are made.

If two inspections are made at each of four different times where there are a total of 64 possible locations for each inspection, the total number of different inspection schemes is approximately $(64^2)^4$. Extending this to 8 time points gives approximately 7.9×10^{28} distinct inspection schemes to be considered. Efficient ways of reducing the number of possible alternatives must be considered.

The simplifying restriction is made that, if m_n regional minima inspections are made at time t, then the m_n inspections are equally spaced around the tank. This gives rotational symmetry. In addition to being an intuitively natural simplification, the symmetry drastically reduces the number of possible locations that need to be considered for possible inspections.

It is necessary to maintain a certain level of confidence in the system's integrity at all times. An inspection scheme should be selected which will reduce predictive variance across the whole vessel, so that at each stage informed decisions can be made based on the state of the whole system. The expected regional minimum wall thickness will be constant through space prior to any observations being made. That is $E_{D_0}(\mu_{\theta t \psi_i})$ does not depend

on ψ_i. If this is not the case, or following inspection, beliefs about underlying wall thicknesses should influence inspection locations. A reasonable example criterion for selecting inspection points is now considered where expected regional minimum wall thickness is constant through space. Many other criteria are possible and would be evaluated in a similar fashion. Let T be the time of the last inspection. The criterion considered is the average of the maximum adjusted variances of the underlying regional minimum wall thicknesses from all the locations modelled, Ψ, over all times up to the time of the one step ahead forecast $T+1$ for a given inspection design Δ which is denoted $F(\Delta)$, where

$$F(\Delta) = mean_{t<T+1}\{max_{\psi_i \in \Psi} var_{D_{t-1}}(\mu_{\theta t \psi_i})\}$$

where $\mu_{\theta t \psi_i}$ is the underlying regional minima at time t and location ψ_i and D_t is all observed data up to time t. $var_{D_{t-1}}(\mu_{\theta t \psi_i})$ is found using equation 7.2. More sophisticated criteria might for example consider a weighted mean where weights are time dependent.

The inspection design Δ will be restricted by the times when inspections are allowed (m_t) and the number of locations inspected (m_n) when an inspection is made. For any choice of m_n and m_t, the optimal inspection scheme $\Delta^*(m_n, m_t)$ will then minimise the criterion $F(\Delta)$ over all designs Δ with

the given values m_n, m_t and rotational symmetry at each time point.

7.3 Possible search strategies

Having restricted attention to equally spaced inspections, for small problems a full search can be made. In larger problems, further load reducing strategies are required.

Exploiting initial symmetry

The symmetry of the prior beliefs may be exploited to reduce the number of inspections that must be made. It can immediately be seen that having the same initial beliefs for every point in space allows the location of the first set of equally spaced inspections to be arbitrarily fixed. The rotational symmetry of the inspections may be further exploited. Attention is restricted to the space between any two of the equally spaced inspections made at time one, all other intervals being taken to be rotationally symmetric. Some plausible heuristics for reducing the potentially enormous search space are now considered. A good inspection scheme may be expected to be, in some sense, evenly spaced both temporally and spatially. The ranges suggested in this section are illustrated in Figure 7.4 where $*$ denotes an inspection and a line indicates candidate inspection locations where curved ends are inclusive.

Suppose there are ζ potential observation locations between each pair of adjacent inspections made at the time of the first inspection. Subsequent inspections may be considered in relation to the range $[1, \zeta]$. It may be assumed by symmetry that the second inspection is less than or equal to $\frac{\zeta}{2}$. Were this second inspection below $\frac{\zeta}{3}$ there would be an unexpectedly larger gap above and a subsequent inspection would be expected in the range $[\frac{\zeta}{3}, \frac{2\zeta}{3}]$. going against the principal of even spatial and temporal distribution. Hence, the second inspection is expected to be in the range $[\frac{\zeta}{3}, \frac{\zeta}{2}]$. Given a second inspection in this range, the third inspection may be expected to be between $\frac{2\zeta}{3}$ and ζ. If the inspection at time two is exactly at $\frac{\zeta}{2}$ this follows from symmetry. Alternatively, if the second inspection is at $\frac{\zeta}{3}$ the third inspection would be expected to be at $\frac{2\zeta}{3}$ or closer to the original inspection at ζ. If the second inspection is to the right of $\frac{\zeta}{3}$ the third inspection is expected to be to the right of $\frac{2\zeta}{3}$. Given the first three inspections are in the above ranges the fourth inspection may then be expected to be less than $\frac{\zeta}{2}$. Restricting the fifth inspection in general is more difficult and will have less effect.

For problems with a large number of time points this intuitive approach will probably not provide sufficient rationalisation.

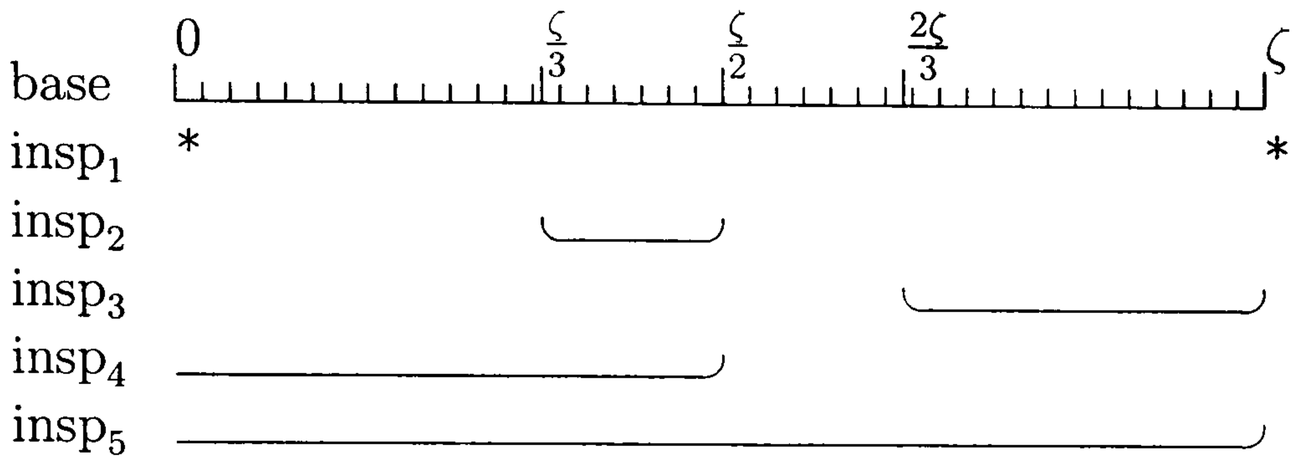


Figure 7.4: Potential inspection locations

'Tentacle' search

An alternative possibility is a 'tentacle' search from likely solutions. Given an initial solution, small changes to the locations inspected at each time are considered, repeating the search with the improved solution until convergence occurs. Allowing eight points to increase, decrease or stay the same gives about 2000 possible scenarios for each iteration and a significant number of iterations may be required to converge. There is a risk that a local minimum may be found rather than the global minimum of interest.

Coarser grid

A further alternative is initially only to assess some of the possible inspection location combinations before focussing in more detail on the locations with the best preliminary results. Considering only every r^{th} location reduces

the load by a factor of approximately r^{t-1} over t time points. A 'tentacle search' can then be made around any possible minima. For many optimisation problems there is either a significant risk of missing an optimal solution (r is large) or the resulting search space still being too large. For example, in the problem where there are 32 locations to choose between over 8 time points, using a 1 in 7 grid still leaves over 65 000 possible choices while significantly restricting possible inspection locations.

k step look ahead

Instead of simultaneously choosing inspection sites across all time points simultaneously, the temporal aspect of the model may be exploited. All possible inspections between times 2 and $k+1$ may be considered. The best of these candidate schemes is then used to fix the inspection location at time 2. All inspection locations between times 3 and $k+2$ are then considered and the best of these candidate schemes used to fix the inspection location at time 3. This may then be repeated until all times have been optimised over. The larger k is the better this scheme will be, but also the greater the computational load. An increase in time points will only increase computational load linearly. This type of k step ahead search procedure will be most effective when temporal correlation is not too strong since this

reduces the effect of older inspections.

These methods can and should be combined to produce an efficient search for a good inspection scheme. In the following section a full search is used where reasonably possible, and otherwise a 2 step ahead search is applied to a 50% grid followed by a tentacle search until a local minimum is found.

7.4 Application to inspection planning

In this section, the effect of the covariance and the prior belief specification on the effectiveness of the inspection scheme is considered. A system is developed to provide decision support when assessing how often and at how many locations inspections should be made. Information is presented in a graphical format allowing easy comparison between possible schemes and an assessment of model sensitivity.

7.4.1 Correlation between minima

Maximum temporal correlation is achieved when there is little update uncertainty. Under these circumstances the prior slope variance ($C_{\beta 0}$) will dominate forecast variance for sufficiently large t. Within the framework considered, prior slope variance ($C_{\beta 0}$) is a fixed proportion of prior level variance ($C_{\theta 0}$). The prior level variance is fixed at 10 (see Section 5.2.1) and hence

this ratio is critical if high temporal correlations are to be considered.

For any $W_{\theta s}$, fixing the ratio of $W_{\theta s}$ to $W_{\beta s}$ at 10:1, the temporal correlation between an observed regional minimum at time 1 ($\mu_{y1\psi_i}$) and the true regional minimum over the same area at time 8 ($\mu_{\theta 8\psi_i}$) has a maximum correlation strength of about 0.4. That is, even with $W_{\theta s} \approx 0$ the prior uncertainty about the corrosion rate is critical for accurate forecasting of future wall thicknesses. Increasing the ratio to 100:1 the maximum correlation reaches about 0.8 while at ratio of 1000:1 the correlation can reach almost 1. Small temporal correlations will still occur when update uncertainty is large. Henceforth a ratio of 1000:1 is assumed to allow a full range of correlations.

In Table 7.1 some quite extreme parameters are considered reflecting very large, moderate and very little temporal change and similarly very strong, moderate and very little spatial correlation. Large and small temporal and spatial correlations can be seen to occur. The first value in each pair in Table 7.1 is the approximate spatial correlation between a regional minimum observation of one area ($\mu_{y1\psi_i}$) and the true regional minimum of the region on the opposite side of the tank ($\mu_{y1\psi_{(i+32)}}$) at time 1. The second value of each pair is the approximate temporal correlation between a regional minimum observation at time 1 ($\mu_{y1\psi_i}$) and the true regional minimum at the same

location at time 8 ($\mu_{\theta 8\psi_i}$). Both values are based upon 1000 simulations. Every location has IID observation noise with a standard deviation of 1.0. c_κ and τ_θ are as defined in Section 5.2.

Spatial (c_κ) Temporal (τ_θ)	0.0001	0.001	0.01
100	0.96, 0.13	0.71, 0.10	0.00, 0.09
1	0.96, 0.73	0.73, 0.74	0.05, 0.70
0.001	0.95, 0.99	0.72, 0.98	0.07, 0.98

Table 7.1: Temporal and spatial correlation of minima (extreme values)

The effects of finer changes in knowledge level based around the middle value of Table 7.1 are now considered. With a temporal strength (τ_θ) of 1 and a spatial strength (c_κ) of 0.001 these beliefs result in a spatial correlation at time 1 between a regional minimum observation of one region ($\mu_{y1\psi_i}$) and the true regional minimum of the region on the opposite side of the tank ($\mu_{y1\psi_{(i+32)}}$) of about 0.7 and a temporal correlation between an observed regional minimum at time 1 ($\mu_{y1\psi_i}$) and the true regional minima at time 8 ($\mu_{y8\psi_i}$) also of about 0.7 . Different pairs of spatial and temporal parameters are now taken around these values to explore the sensitivity of the inspection plan to different strengths of spatial and temporal correlation. As shown in Table 7.2, temporal parameter values (τ_θ) of 5 and 0.3 and spatial parameter values (c_κ) of 0.0025 and 0.0003 give temporal and spatial correlations (as

defined above) of approximately 0.5 and 0.9. These correlations (ρ) were chosen to approximate a three by three grid of ρ^2 values based on approximately 0.25, 0.5 and 0.75 which are now considered for both temporal and spatial components. The first value of each pair in Table 7.2 is the spatial correlation while the second value in each pair is the temporal correlation both described above. IID observation noise has a standard deviation of 1.0.

Spatial (c_κ) Temporal (τ_θ)	0.0025	0.001	0.0003
5	0.45, 0.44	0.71, 0.43	0.89, 0.44
1	0.44, 0.72	0.71, 0.73	0.89, 0.73
0.3	0.44, 0.87	0.72, 0.88	0.89, 0.87

Table 7.2: Temporal and spatial correlation of minima

7.4.2 The effect of observation noise

For many inspection tools, the parameters governing independent observation noise will be quite well known since they can be easily assessed using repeated measurements. Dependence within the noise is likely to be less well modelled. Using simulation, for this example, the correlation structure in Table 7.2 is found to be quite robust to changes in the noise variance that are less than 1.0 for both correlated and independent noise.

Independent and identically distributed noise is first considered added to

the data values before regional minima are assessed. In Table 7.3 a wide range of observation noise levels are considered where the spatial coefficient is 0.001 and the temporal coefficient is 1 (following the discussion in Section 7.4.1). As can be seen the correlation structure is quite robust to IID observation noise on the observations. A very similar pattern may be seen for the other sets of spatial and temporal parameters previously considered in Table 7.2.

Noise (s.d.)	Temporal (1), Spatial (0.001)
0.0001	0.74, 0.72
0.001	0.74, 0.72
0.01	0.74, 0.72
0.1	0.74, 0.72
1.0	0.73, 0.70
10.0	0.08, 0.04

Table 7.3: Temporal and spatial correlation, IID noise

The effect on the spatial and temporal correlation of a having a correlated observation noise may also be considered. In Table 7.4 a correlation of one is considered. This equates to having a single observation noise for each region observed. A slightly more marked effect can be seen than for the IID noise but for noise with a standard deviation of less than 1 the correlations remain reasonably robust.

Observation noise either IID or strongly correlated with variance greater than one very quickly reduced the correlations.

Noise (s.d.)	Temporal (1), Spatial (0.001)
0.0001	0.74,0.72
0.001	0.74,0.72
0.01	0.74,0.71
0.1	0.73,0.71
1.0	0.68,0.66
10.0	0.42,0.40

Table 7.4: Temporal and spatial correlation, correlated noise

While the observational noise is less than one it has little effect on the correlations, so, any forecasts made will also be reasonably robust to changes in the observational noise. Observation noise with a standard deviation of 1.0 is relatively large compared to the standard deviation of the general temporal component (0.3, 1 or 5) but less so compared to the prior level uncertainty (10).

For this example an IID observation noise of 1 is used on the assumption that lower levels of observation noise would result in stronger conclusions based on lower variances.

7.5 Comparison between inspection schemes

Suppose that for a given inspection scheme there are m_n equally spaced observed regional minima ($\mu_{yt_i\psi_j}$) during each inspection. Different numbers of inspections are considered; $m_n = 1, 2, 4, 8$ and 16. Having chosen the

first inspection, a range of possible locations for subsequent inspections may then be considered. For each inspection regime, the adjusted variance of each underlying regional minimum may be assessed at each time point. If, at any time point, sufficient corrosion is discovered to raise doubts about the future integrity of the vessel, then it will be necessary to repair the vessel. or, at least increase the frequency of inspection around the region causing concern.

Five different inspection schemes are considered (see Table 7.5) with m_n observations made at each time.

Scheme (number of times inspected)	Times inspected (m_t)
a(1)	1
b(2)	1,7
c(3)	1,4,7
d(4)	1,3,5,7
e(7)	1,2,3,4,5,6,7

Table 7.5: Different inspection schemes

The proportion of the initial uncertainty remaining (given prior knowledge D_0) following observation of the data may be calculated. Taking this proportion away from one gives the proportion of variance removed by the best inspection scheme for each pair (m_n, m_t). This may be written as

$$P(m_n, m_t) = 1 - \frac{F(\Delta^*(m_n, m_t))}{mean_{t \leq T+1, \psi_i \in \Psi} var_{D_0}(\mu_{\theta t \psi_1})}$$

where T is the time of the last inspection.

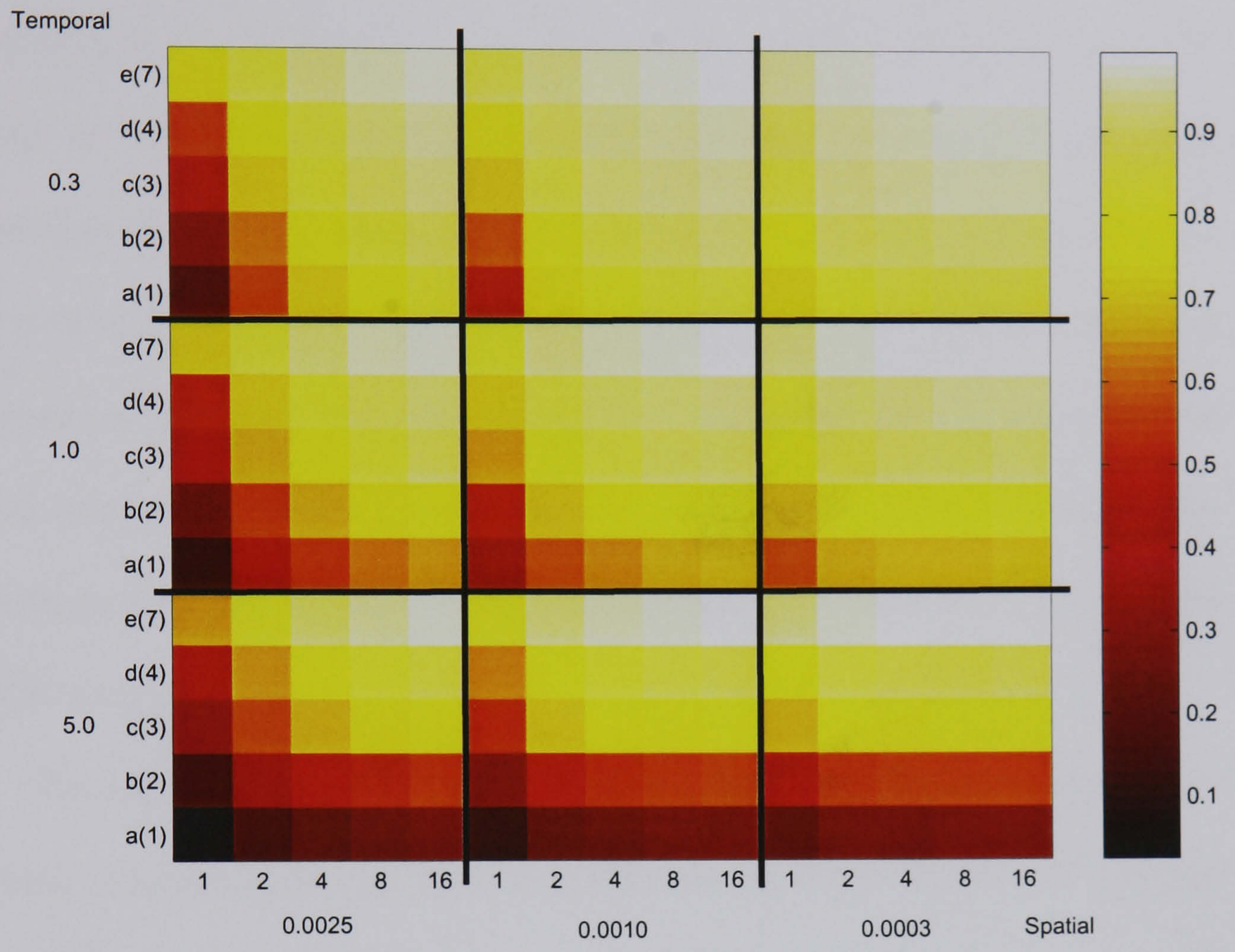


Figure 7.5: Matrix of local simulations

Figure 7.5 shows, for each m_n, set of inspection times (m_t) and set of values of τ_θ and c_κ in Table 7.2 the value of $P(m_n, m_t)$ for design $\Delta^*(m_n, m_t)$. Near white indicates an informative inspection procedure while near black indicates an uninformative inspection procedure. The middle sub-grid corresponds to spatial and temporal parameters of 1 and 0.001 respectively. Each square in this grid then corresponds to a given number of inspections at each time point (on the horizontal axis) and inspections made at times indicated by the vertical axis. Locations are chosen as in Section 7.3 to optimise $F(\Delta^*(m_n, m_t))$.

The top right sub-grid has the maximum spatial and temporal correlation which is reflected by the high proportion of the available information that is gained by any inspection. Conversely the bottom left sub-grid has the minimum spatial and temporal correlation which is reflected by the much lower proportion of the available information that is gained even with many more inspections in space and time.

The increase in information following increased temporal or spatial inspections is shown by the larger values in the top right of each subgraph for all parameter values. As would be expected, it can also be seen that as the temporal dependence increases the effect of having more inspections in time

decreases. Similarly, for the spatial component, as the spatial dependence increases the effect of having more inspections in space decreases. Using these results it is possible to compare the benefit of extra spatial inspections to the benefit of extra temporal inspections.

In general observing twice as many locations at a given inspection time will be cheaper than making the original number of inspections at two different times because of the reduced set-up costs. An inspection cost function may be used to make a comparison between costs and gain in accuracy for different inspection schemes. In its simplest form this function will include a cost for each regional minimum observed and a fixed cost for making one or more inspections at a given time. The scheme that gains maximum information for a fixed cost or a given amount of information at minimum cost can then be found, and pictures such as Figure 7.5 can be used to explore the sensitivity of the design choice to the correlations in the model.

7.6 How much improvement

Inspection locations have been selected based upon reducing system variances. These spatial schemes may be compared to the simpler non-spatial schemes with the same number of inspections and fixed inspection locations

through time (required for non spatial models) to see how much additional information is gained. The benefits for the nine pairs of parameters (introduced in Table 7.2) when an inspection is made at every time interval (scheme $e(7)$) with $m_n = 1$, 2, 4, 8, or 16 inspection points are considered.

Values of $F(\Delta^*)$ are calculated using a spatial(.) and a non spatial scheme($*$) for the nine sets of parameters. The non spatial scheme makes inspections at different times at the same set of locations. The results are shown in Figure 7.6. For each set of parameters inspections of 1, 2, 4, 8 and 16 regions are considered (horizontal axis). Increasing the number of points inspected decreases the uncertainty for all sets of parameter values.

The values of $F(\Delta^*)$ for the non spatial scheme may be up to 30 times higher than the equivalent values for the spatial scheme. Figure 7.7 shows this relative comparison.

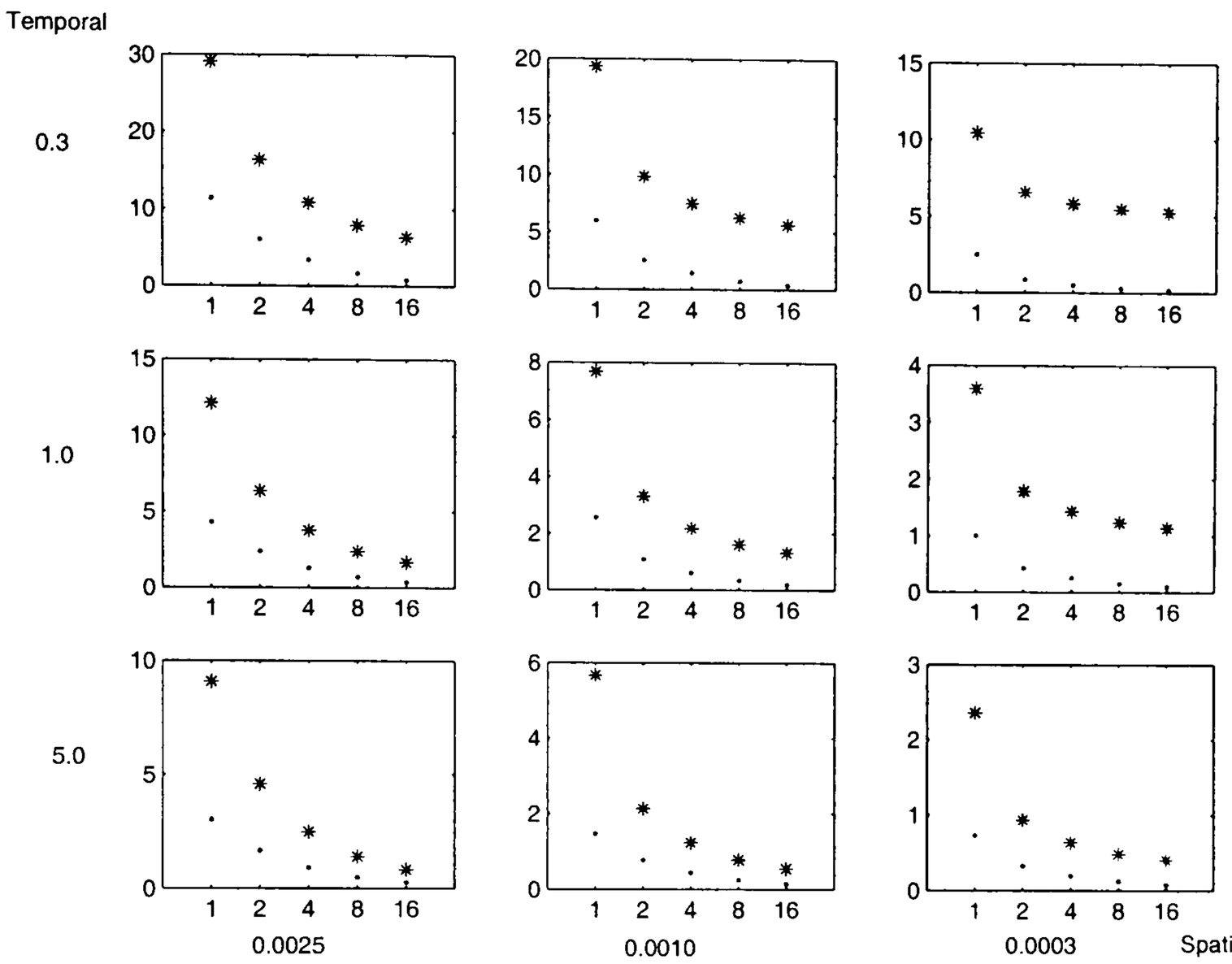


Figure 7.6: Values of $F(\Delta^*)$ using a spatial(.) and a non spatial scheme(*)

The large values to the right of Figure 7.7 indicate significant information gains. The trend up to the right shows that as more points are inspected having variable inspection locations becomes relatively more important. Lower temporal effects generally make using a spatial inspection scheme more important while, conversely, lower spatial strength generally reduces the benefits of having a spatial inspection scheme as would be expected.

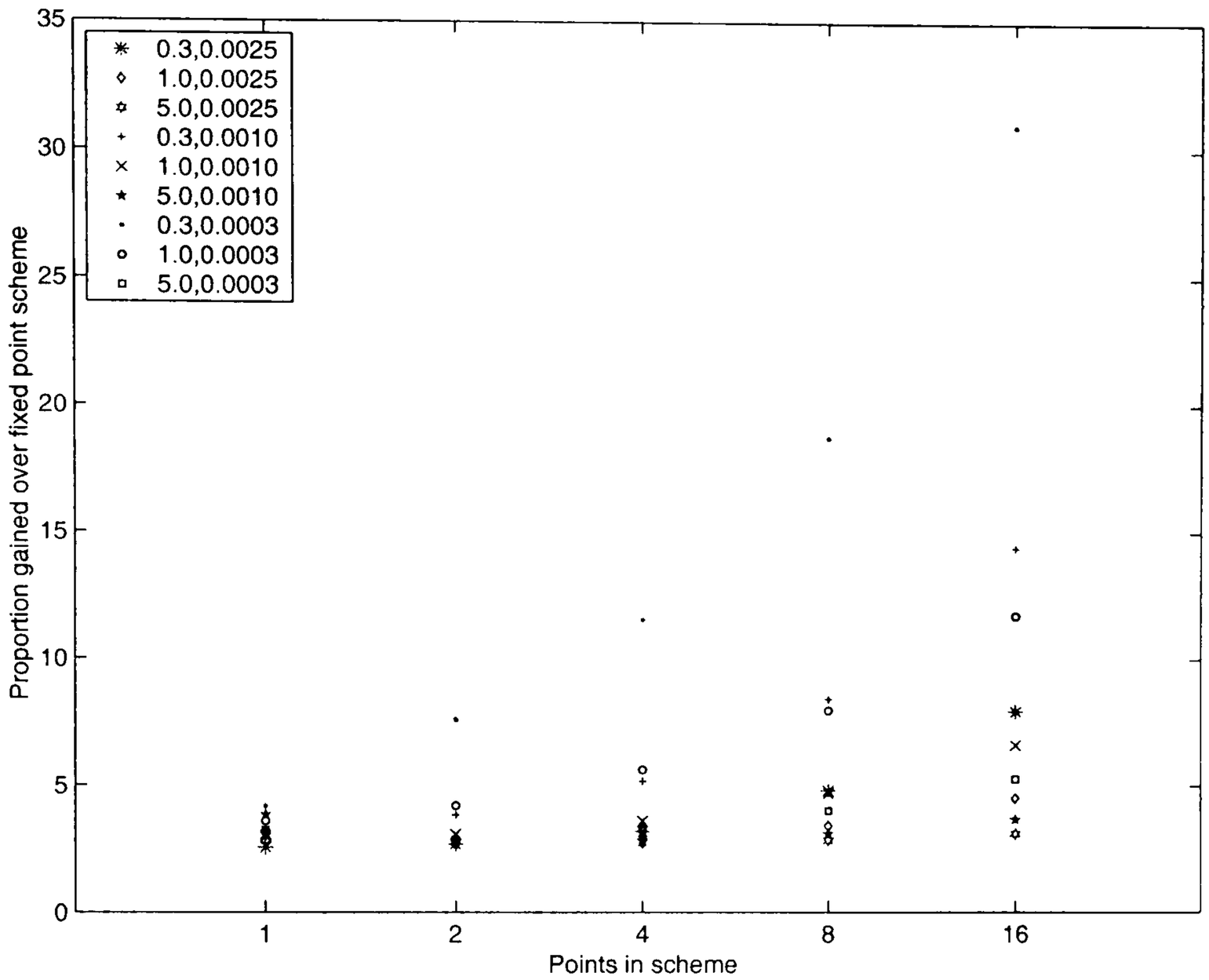


Figure 7.7: Gain from spatially informative sampling

7.6.1 Robustness

Having identified a good inspection design for a given uncertainty specification, the effect of changes in prior uncertainty on the preferred inspection scheme should be assessed. If feasible changes in the prior beliefs governing the model do not significantly change the preferred inspection locations for a given inspection budget, then the model is robust to changes in parameters. A sensitive model will require consideration of the additional risk associated with plausible changes in parameter values and may suggest a more extensive

inspection is required.

7.6.2 Intermediate regions

As discussed in Section 7.1, once design choices have been made, the intermediate regions should be analysed within the model. We focus on the optimal scheme when $\tau_\theta = 1$, $c_\kappa = 0.001$, inspections are made at times 1, 4 and 7, and when each time an inspection is made four regional minima are observed ($m_n = 4$). In this example $F(\Delta^*(m_n, m_t))$ goes from 2.13 in the 64 regional minima system to 3.57 for the equivalent calculation in the 128 regional minima system. Similarly $P(m_n, m_t)$ goes from 0.89 in the 64 point calculation to 0.75 in the 128 point calculation. The optimal design (based upon $F(\Delta^*(m_n, m_t))$) for the 128 regional minima system has a value of 3.46 and $P(m_n, m_t) = 0.77$. This suggests that in this case the 64 regional minima provided a good choice of inspection scheme, but, all regions should be considered when estimating the confidence interval for the expected life of the system. Further extending this work may require more extensive use of the partial simulation approach suggested in Section 6.2.1.

7.7 Prediction of global minimum

The life of the system is essentially defined by the global minimum. Accurate estimates of future global minima will help predict the remaining life of the system. The expectation and the variance of the global minimum may be adjusted following an inspection. A particular sampling scheme is considered based upon ranked regional observations.

For a particular inspection scheme, the prior covariances between 'the observed regional minima after they have been ranked' ($\mu_{yt\psi_{i_1}} \geq \mu_{yt\psi_{i_2}} \geq \ldots \geq \mu_{yt\psi_{i_{m_n}}}$) and the 'underlying global minimum at times 1 to $T+1$' ($g_{\theta 1}, \ldots, g_{\theta T+1}$) are calculated, where T is the time of the last inspection, and regions ψ_{i_j} are used to force the condition $\mu_{yt\psi_{i_j}} \geq \mu_{yt\psi_{i_{j+1}}}$. We let D^*_{t-1} be all the ranked data up to time $t-1$, where data for a particular time is ranked independently of data at other times. We assume for the ranked regional minima at time t $\mu^*_{yt1} \geq \mu^*_{ytm_n}$. The covariance between μ^*_{ytj} and $g_{\theta t}$ may be calculated directly from the data previously simulated to find the covariance structure between the regional minima (see Section 7.1). Each set of inspection locations has a different ranking for each set of simulated data and so needs a separate covariance calculation. The covariance structure can then be adjusted (using equation 7.2) to evaluate the effect of a set of

regional minima observations on beliefs about the global minimum.

For every candidate inspection scheme, it is necessary to refer to the original simulated data set with the result that calculations to decide between different inspection schemes can be much slower than adjustments for regional minima data (see Section 7.1) where all covariances can be calculated upfront. Consequentially, optimal design searches for inspection schemes based upon this global minimum criterion may become very computationally demanding especially where large simulated data sets are considered. For this example, attention is focused on a particular scheme.

The particular scheme considered has temporal covariance parameter $\tau_\theta = 1$, spatial covariance parameter $c_\kappa = 0.001$ and $m_n = 4$ equally spaced inspections made at times 1, 4 and 7. Four equally spaced regional minima are thus observed at three different times. A candidate set of locations were found based on a small subset of the data (1000 cases). There are 16 candidate inspection locations on each quadrant of the tank and within each quadrant inspections are made at candidate locations 1, 7 and 14 (at times 1, 4 and 7 respectively). This candidate set of locations was found to reduce the mean one step ahead global system variance following adjustment by the ranked observations the most over all possible schemes. That is, measure-

ments at locations 1, 7 and 14 at times 1, 4 and 7 respectively minimised :

$$F^*(\Delta) = mean_{t<T+1} var_{D^*_{t-1}}(g_{\theta t})$$

The covariance is now found between D^*_t and the set of global system minima through time $(g_{\theta 1}, \ldots, g_{\theta T})$. In this example, this means that for each set of four regional minima observed there are now four different covariances with the underlying global minimum at any given time. These four different covariances should be better able to predict the global minimum than four estimates of one covariance in the unranked case. In this example on average over the eight times approximately 6% of the prior variance about the one step ahead global system minimum remains following adjustment by the ranked observations - see equation (7.3).

$$P^*(m_n, m_t) = 1 - \frac{mean_{t<T+1} var_{D^*_{t-1}}(g_{\theta t})}{mean_{t<T+1} var_{D_0}(g_{\theta t})} \tag{7.3}$$

Using the same data but not ranking leaves approximately 50% more uncertainty about the global system minimum at each time considered.

Previously (see Section 7.5), based on these parameters and applying this inspection scheme to reduce uncertainty across the whole system rather than just about the using the methods previously described, $P(m_n, m_t) = 0.89$.

The development of these results in time may also be considered. Figure 7.8 shows the standard deviation of the global minimum for each of the eight time points considered following the adjustment for data at times one (first adjustment), one and four (second adjustment) and one, four and seven (third adjustment).

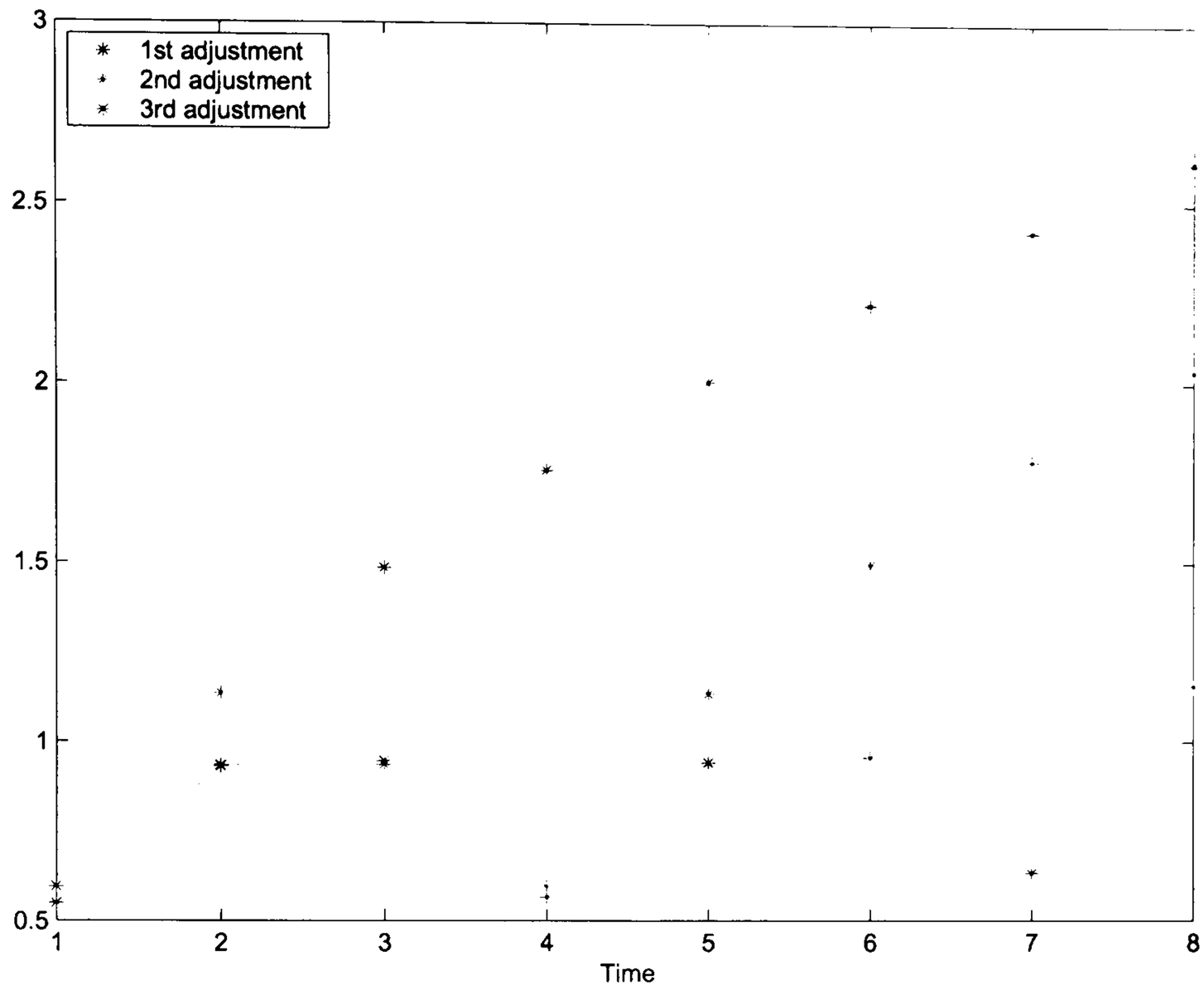


Figure 7.8: Standard deviation of the global minimum through time following each adjustment

The adjusted variance through time may also be calculated for the other sets of parameters considered in Table 7.2. The same inspection scheme is assumed, that is $m_n = 4$ equally spaced inspections made at times 1, 4 and 7 and measurements at locations 1, 7 and 14 on each quadrant. Figure 7.9 shows the standard deviation of the global minimum through time for different parameter values following adjustment for data observations at times one, four and seven. For these parameter values high temporal change is the key factor influencing uncertainty between inspections. Strong spatial correlations do reduce the uncertainty but to a much lesser extent.

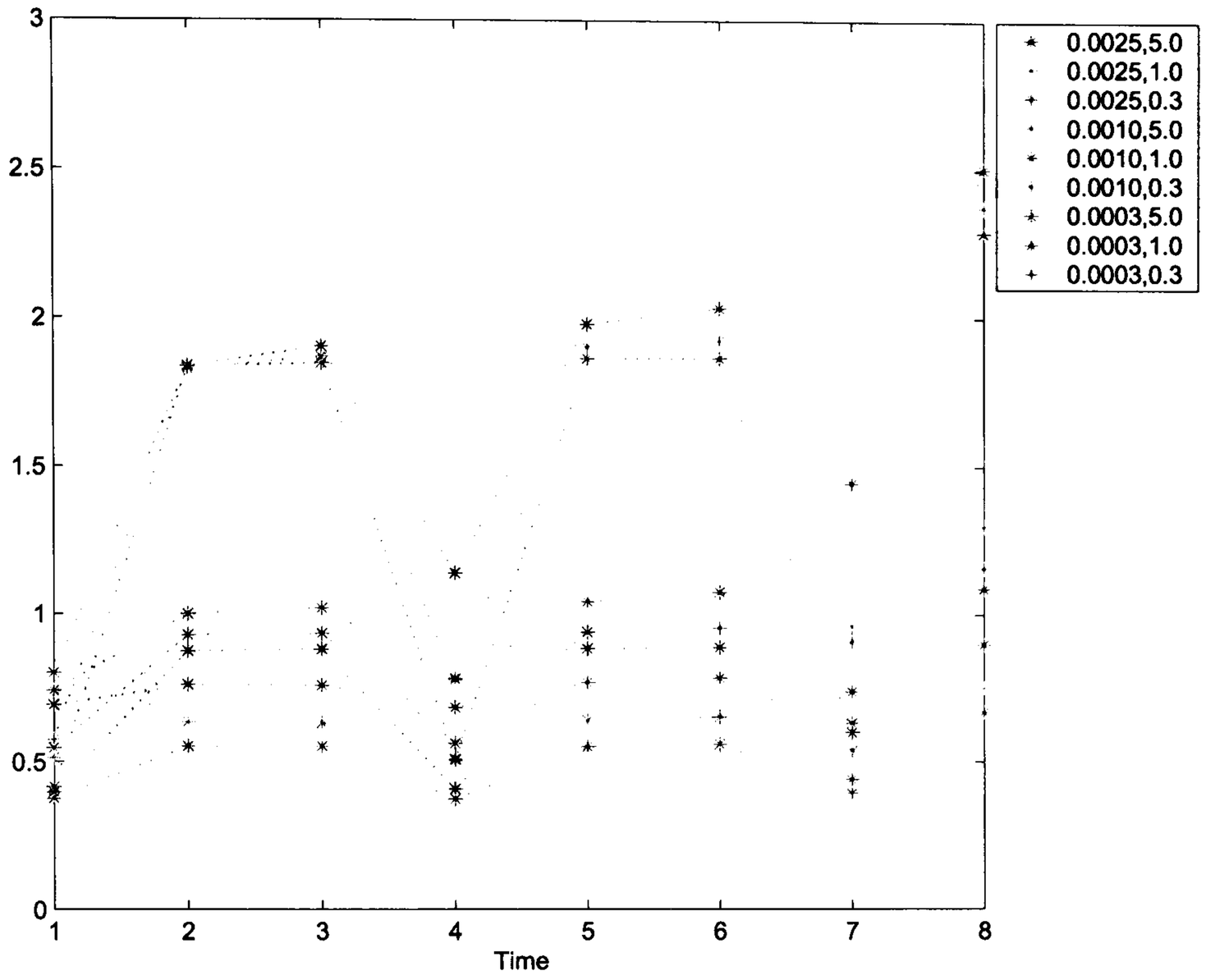


Figure 7.9: Standard deviation of the global minimum through time for different parameter values

7.8 Discussion of global minima

The plans discussed in this chapter are based on minimising future uncertainty across the entire system. This is a sensible approach before data are observed or if observed data do not show corrosion, hence, all areas of the system pose equal risk. If information about corrosion and wall thickness becomes available either generally across the system or as a result of inspections this should form part of the criteria used. In general, if there are two locations with the same variance, it is preferable to gain information about the location where the wall is believed to be at greatest risk. A similar heuristic criterion to that introduced in Section 4.1 may be considered.

Discussion

Efficient inspection for maintenance of complex industrial systems subject to degradation effects, such as corrosion has been discussed. With appropriate statistical modelling, the potential for better use of inspection resources to improve the quality of inferences has been demonstrated. A Bayesian spatio-temporal dynamic linear model has been developed suitable for a wide range of problems, such as the modelling of the spread of the corrosion when monitoring wall thickness, where the inspection method used may collect either point data or transformed data.

The Bayesian spatio-temporal dynamic linear model used has three main advantages. Firstly, it allows incorporation of expert judgements into the inspection planning in a natural and powerful way. Secondly, the model allows the maximum flexibility in the choice of inspection times and locations. Finally, the model is sufficiently tractable to allow us to obtain, at least approximately, optimal inspection procedures, and to obtain, by simu-

lation, an assessment of the risk associated with these procedures. Efficient inclusion of transformed observations which are difficult to use directly to update beliefs are allowed within the standard DLM setup. This allows efficient designs to be constructed. Furthermore, it also allows comparison between different sampling schemes. For many of the methods we have discussed computational load increases quadratically compared to model size. We have discussed possible ways of reducing this computational burden and hence allowing application to larger problems.

The results of existing inspections are often badly recorded, so that subsequent inferences are very difficult to make accurately. It has been shown that detailed recording of inspection locations coupled with a suitable spatio-temporal model may substantially reduce uncertainty following an inspection.

Many of the developments made in this thesis could be used in a decision support role during the inspection of complex industrial systems.

The model considers changes in corrosion rate as a random walk. This provides a simple and plausible model for the corrosion rates. It has been assumed for these examples that the corrosion rate change can be modelled with a single variance for all times. However, a changing variance can be handled similarly, for example if suggested by assessment of related data

sets. Sudden changes in corrosion rate as might be experienced if there is a significant change in operating conditions may be best handled with an intervention. In systems where there is shielding of some description this will reduce the wall loss while the shielding remains intact. A two process multivariate model would be more realistic in these circumstances where the model could switch from one corrosion rate to the other.

We have assumed fixed values for all our variances and hyper-parameters, a good model fit relying on expert judgement, supplementary data on similar systems and our diagnostic checks. From a fully Bayesian perspective these values will also be unknown and have uncertainties attached to them. This can be especially relevant in time series models over long periods where the variances are more likely to change over the course of the model. Modelling the problem with these additional uncertainties would require the use of more sophisticated simulation models. Using full MCMC methods to identify optimal inspection regimes, would be extremely computationally demanding. A simple approach would be to model the uncertainty in the hyper-parameters while updating beliefs given all existing data but to assume these values set at their current expected values while considering future inspection schemes, using the methodology that we have described. A further possibility would

be to use the approach that we have described, for fixed choices of the hyper-parameters, to restrict the number of possible inspection plans that need to be considered. These plans could then be compared within a fully Bayesian framework, for example using the simulation approach described in (Muller 1998).

A natural extension to this work is to incorporate decision analysis, balancing utilities for the consequences of different forms of unexpected system failure against the cost utility for making inspections. This would allow an assessment of the optimal expenditure on inspection, incorporating an assessment of the cost of the constraints, for example having to inspect when a system is shut down for maintenance.

Bibliography

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Prentice Hall.

ASTM Standard G16. Standard guide for applying statistics to analysis of corrosion data.

Barão, M. I. and J. A. Tawn (1999). Extremal analysis of short series with outliers: sea levels and athletics records. *Applied Statistics 48*(4), 469–487.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1997). On the analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*.

Chambers, J., W. Cleveland, B. Kleiner, and P. Tukey (1983). *Graphical Methods for Data Analysis*. Wadsworth.

Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: an application to structural design. *Applied Statistics 43*(1),

1–48.

Coles, S. G. and J. A. Tawn (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics 45*(4), 463–478.

Craig, P. S., M. Goldstein, A. H. Seheult, and J. A. Smith (1998). Constructing partial prior specification for models of complex physical systems. *The Statistician 47*(1), 37–53.

Cressie, N. and H. C. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*(448), 1330–1340.

den Heijer, C. and M. Festen (1996). Inspection of oil products and gas lift flow lines, Oman PDO. Technical report, Shell.

Dunn, M. (1997). Guideline for the use of risk assessment methods. Technical report, Shell.

Evensen, G. (1994). Seqential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*.

Gabrielli, C., F. Huet, M. Keddam, and R. Oltra (1990). A review of the probabilistic aspects of localised corrosion. *Science*, 266–278.

Goldstein, M. (1999). Bayes linear analysis. *In: Encyclopaedia of Statistical Sciences, Update Volume 3*, 29–34. S. Kotz, C.B. Read and D.L. Banks(eds.). Wiley, New York.

Heemink, A. W. and A. J. Segers (2002). Modelling and prediction of environmental data in space and time using Kalman filtering. *Stochastic Environmental Research and Risk Assessment 16*, 225–240.

Hoeve, A., K. den Heijer, and M. Festen (1998). Trending of ultrasonic wall thickness measurements in time. Technical report, Shell.

Hoeve, A. and K. Lewis (1999). Analysis of NARL H-1401 inspection results using TST. Technical report, Shell.

Hoeve, F., P. Jonathan, and K. den Heijer (2000). Multivariate dynamic linear modelling of an industrial furnace reactor. Technical report, Shell.

Huang, H.-C. and N. Cressie (1995). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis 22*, 159–175.

Julier, S. J. and J. K. Uhlmann (1997). A new extension of the Kalman filter to nonlinear systems. *Sugnal Processing, Sensor Fusion and Target*

Recognition CI, SPIE Proceedings 3068, 182–193.

Kalman (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 35–45.

Laycock, P., R. Cottis, and P. Scarf (1990). Extrapolation of extreme pit depths in space and time. *Journal of Electrochemical Society 137.*

Little, J. D., M. Goldstein, P. Jonathan, and C. den Heijer (to appear 2004). Spatio-temporal modelling of corrosion in an industrial furnace. *Applied Stochastic Models in Business and Industry.*

Mardia, K. V., C. Goodall, E. J. Redfern, and F. J. Alsonso (1998). The kriged Kalman filter. *Test 7*, 217–285.

Muller, P. (1998). Simulation based optimal design. *Bayesian Statistics 6.*

Muñoz, A., V. Carey, J. P. Schouten, M. Segal, and B. Rosner (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics 48*, 733–742.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *Journal of the American Statistical Association 88*, 390–399.

Oehlert, G. W. (1996). Shrinking a wet deposition network. *Atmospheric Environment 30(8)*, 1347–1357.

Pukelsheim, F. (1994). The three sigma rule. *American Statistician 48*(2). 88–91.

Quintana, J. M. and M. West (1987). An analysis of international exchange rates using multivariate DLM's. *The Statistican 36*, 275–281.

Ripley, B. D. (1981). *Spatial Statistics.* Wiley.

Ripley, B. D. (1987). *Stochastic Simulation.* Wiley.

Sansó, B. and L. Guenni (2000). A nonstationary multisite model for rainfall. *Jornal of the American Statistical Association 95*(452), 1089–1100.

Schipper, C. M. and J. J. M. van Nisselroij (1998). Non-intrusive inspection demonstration on NAM-Tjuchem vessel v-11 (part 1). Technical report.

Shaddick, G. and J. Wakefield (2002). Modelling daily multivariate pollutant data at multiple sites. *Applied Statistics 51*(3), 351–372.

Shibata, T., M. Akashi, K. Ikematsu, and H. Tsuge (1998). Estimating localized-corrosion maximum depths by extreme - value analysis. *Corrosion Engineering*.

Smith, A. and M. West (1983). Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics*, 867 –878.

Stroud, J. R., P. Muller, and B. Sansó (Nov 2001). Dynamic model for

spatiotemporal data,. *Journal of the Royal Statistical Society 63*, 673-689.

Tallin, A. and M. Conley (1994). Assessing inspection results using Bayes' Theorem. *Proceedings of the 1994 Improving Reliability in Petroleum Refineries and Chemical Plants Conference*..

Terpstra, S. (2002). Shell expro - Definition of S-NII Model. Technical report, Shell.

Terpstra, S. and D. Ritchie (2000). Implementation of non intrusive inspection on three vessels on Shell Expro's Brent Bravo offshore platform. Technical report, Shell.

Terpstra, S., D. Ritchie, and C. den Heijer (2002). Non intrusive inspection manual applied to topside facilities. Technical report, Shell.

UK Health And Safety Executive (2002). Guidelines for use of statistics for analysis of sample inspection of corrosion.

Wan, E. A. and R. van der Merwe (2000). The unscented Kalman filter for nonlinear estimation.
http://cslu.ece.ogi.edu/publications/ps/UPF_CSLU_talk.pdf

Weissman, I. (1978). Estimation of parameters and large quantiles based

on the k largest observations. *Journal of the American Statistical Association 73*(364), 812–815.

Welch, G. and G. Bishop (2003). An introduction to the Kalman filter. Technical Report TR95-041, Department of Computer Science. University of North Carolina - Chapel Hill.

West, M. and J. Harrison (1996). *Bayesian Forecasting and Dynamic Models*. Spring-Verlag.

Whittaker, J. (1990). *Graphical Models in Applied Mathematical Statistics*. Wiley.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association 96*(454), 382–397.

Wilke, C. K. and N. Cressie (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrica 86*, 815–829.

Wilkinson, D. J. (1997). Bayes linear variance adjustment for locally linear DLMs. *Journal of Forecasting 16*(5), 329–342.

Appendix 1: Data

Two data sets have been considered in this thesis.

Furnace F4102

The first column is the coil number, the second column the pipe number and the third column the height in feet of the inspection. Missing values are indicated by a -99. All measurements are in mm.

```
data=[   1,1,3,12.0,11.5,10.4,10.8,11.1,11.5,11.5,-99,
    1,1,7,13.0,12 ,11.0,11.8,11.9,12.0,11.9,-99,
    1,1,12,12.5,11.5,11.0,-99,12.0,12.0,11.8,-99,
    1,1,18,11.8,10.5 ,10.0,-99,10.9,-99,10.5,-99,
    1,2,3,11.9,12.0,10.6 ,11.0,11.2,11.3,11.1,-99,
    1,2,7,12.0,11.5,10.6 ,11.5,11.3,11.8,11.8,-99,
    1,2,12,12.3,11.0,10.5,-99,11.3,11.4,11.1,-99,
    1,2,18,12.0,10.5 ,10.5,-99,11.7,-99,10.8,-99,
    1,3,3,11.7,11.0,10.0,10.5,10.7,11.2,11.0,-99,
    1,3,7,11.7,11.0,10.2,10.6,10.8,11.4,10.9,-99,
    1,3,12,12.4,11.5,10.5,-99,11.6,11.8,11.6,-99,
    1,3,18,12.7,10.5 ,10.5,-99,11.1,-99,11.2,-99,
    1,4,3,11.2,11.0,10.2,11.0,11.1,11.2,11.0,-99,
    1,4,7,11.5,11.0,10 ,10.8,10.6,11.4,11.2,-99,
    1,4,12,12.2,10.5 ,10.2,-99,11.0,11.5,11.0,-99,
    1,4,18,12.4,11.5,10.8,-99,11.0,-99,11.4,-99,
    1,5,3,12.0,11.5,10.5,10.8,10.4,11.6,11.3,-99,
    1,5,7,11.0,10.5,10.0,10.2,10.7,10.7,10.7,-99,
    1,5,12,11.5,11.0,10,-99,11.4,10.7,10.5,-99,
    1,5,18,11.1,11.0,10.0,-99,10.7,-99,10.2,-99,
```

```
1,6,3,10.4,10.5,9.6,10.0,10.2,10.8,10.2,-99,
1,6,7,11.5,10.5,9.8,10.4,10.4,10.3,10.1,-99,
1,6,12,11.6,11.0,9.8 ,-99,11.2,11.3,11.6,-99,
1,6,18,11.4,10.5 ,9.8,-99,11.7,-99,10.9,-99,
1,7,3,11.8,11.5,10.8,11.0,11.7,11.5,11.4,-99,
1,7,7,11.7,11.5,10.4 ,11.2,10.2,11.6,11.2,-99,
1,7,12,11.6,11.0,10.5,-99,11.7,10.9,10.9,-99,
1,7,18,12.0,11.5,10.5,-99,11.0,-99,11.0,-99,
1,8,3,10.2,11.0,10.0,9.5,9.8,10.3,10.4,-99,
1,8,7,10.6,10.0,10.0,9.5,10.9,10.0,9.7,-99,
1,8,12,10.4,9.5 ,10.4,-99,10.2,10.0,10.0,-99,
1,8,18,10.5,10.0,9.4,-99,10.9,-99,11.4,-99,
2,1,3,11.2,11.0,10.2,10.7,11.1,10.3,11.2,-99,
2,1,7,12.7,12.0,10.8,11.0,11.4,11.5,11.1,-99,
2,1,12,11.0,11.5,10.2 ,-99,11.6,10.7,11.0,-99,
2,1,18,10.5,11.0,10.2,-99,11.2,-99,11.1,-99,
2,2,3,10.8,10.5,9.6,10.0,10.2,10.7,10.3,-99,
2,2,7,11.2,10.5,10.8,11.5,11.3,11.9,11.4,-99,
2,2,12,11.3,11.5,10 ,-99,11.7,11.5,11.0,-99,
2,2,18,11.4,11.5,10.5,-99,11.2,-99,10.6,-99,
2,3,3,11.0,10.5,10.2,10.5,10.6,11.4,11.0,-99,
2,3,7,11.0,10.5,10.2,10.0,11.4 ,11.3,10.8,-99,
2,3,12,10.5,10.0,10.2,-99,11.3,10.9,10.8,-99,
2,3,18,11.9,10 ,-99,-99,11.1,-99,11.1,-99,
2,4,3,10.2,10.0,10.0,10.0,10.7,10.3,10.3,-99,
2,4,7,10.8,10.0,9.8,10.0,10.1,11.5,10.4,-99,
2,4,12,11.0,10.0,9.8,-99,11.0,10.5,10.3,-99,
2,4,18,11.8,10.5 ,10.2,-99,10.8,-99,10.2,-99,
2,5,3,11.0,11.5,10.2 ,11.0,11.2,11.0,10.9,-99,
2,5,7,10.2,10.0,9.6,9.8,10.4,10.6,10.4,-99,
2,5,12,11.2,11.0,9.8 ,-99,11.3,10.9,11.0,-99,
2,5,18,11.0,11.0,10.5,-99,11.0,-99,10.5,-99,
2,6,3,10.4,11.0,10.2,10.5,11.0,11.3,10.6,-99,
2,6,7,11.4,10.5,10.0,10.8,10.9,11.5,10.8,-99,
2,6,12,11.9,10.5 ,9.8,-99,11.2,11.2,10.9,-99,
2,6,18,12.3,11.5,11.0,-99,11.8,-99,11.6,-99,
2,7,3,11.2,11.5,10 ,10.5,11.1,11.6,11.0,-99,
2,7,7,11.0,11.0,10.2,10.2,10.7,11.8,10.7,-99,
```

```
2,7,12,11.6,11.0,10.5,-99,10.5,11.2,10.9,-99,
2,7,18,12.2,10.5 ,10.2,-99,11.2,-99,10.5,-99,
2,8,3,11.2,10.5,10.3,10.5,11.1,10.4,10.9,-99,
2,8,7,12.0,10.5 ,10.2,10.6,10.9,11.2,11.0,-99,
2,8,12,11.2,10.5,10.2,-99,11.1,10.7,10.6,-99,
2,8,18,10.5,10.5,9.6,-99,10.4,-99,11.8,-99,
3,1,3,11.4,11.0,10.0,10.6,10.3,10.7,11.1,-99,
3,1,7,11.7,10.5 ,9.6,10.8,11.4,10.8,11.4,-99,
3,1,12,11.7,11.0,9.6 ,-99,11.1,10.6,10.4,-99,
3,1,18,12.3,11 ,10.0,-99,11.3,-99,11.1,-99,
3,2,3,11.9,11.0,-99,11.2,11.5,10.8,11.6,-99,
3,2,7,12.0,12.0,10.5,11.2,11.5,11.9,11.3,-99,
3,2,12,11.8,11.0,9.6,-99 ,11.5 ,11.2,11.4,-99,
3,2,18,11.9,11 ,10 ,-99,11.3,-99,11.1,-99,
3,3,3,11.5,11.5,11.0,11.0,11.1,11.5,11.2,-99,
3,3,7,11.7,11.5,10.2 ,10.5,10.7,11.0,11.4,-99,
3,3,12,11.7,11.0,10.2,-99,11.2,11.0,11.0,-99,
3,3,18,11.5,11.5,10.2 ,-99,11.0,-99,10.5,-99,
3,4,3,10.9,11.0,10.2,10.0,10.4,10.9,10.3,-99,
3,4,7,11.5,10.5 ,9.8,10.2,11.4,11.0,11.8,-99,
3,4,12,11.8,11.5,9.5 ,-99,10.8,11.5,10.7,-99,
3,4,18,11.2,11.0,10.5,-99,10.7,-99,11.3,-99,
3,5,3,11.5,12.0,11.2,10.5,10.2,11.9 ,11.0,-99,
3,5,7,11.3,11.0,10 ,11.2,11.7,11.4,11.7,-99,
3,5,12,11.2,11.0,9.8 ,-99,10.7,10.8,10.8,-99,
3,5,18,11.2,11.5,10.4,-99,10.5,-99,11.0,-99,
3,6,3,11.5,11.5,10.2 ,10.0,10.5,11.3,10.4,-99,
3,6,7,11.0,11.0,10.4,10.5,11.3,11.5,10.7,-99,
3,6,12,12.2,11.0,10.6,-99,10.9,11.2,10.6,-99,
3,6,18,11.8,11 ,10 ,-99,11.9,-99,11.4,-99,
3,7,3,10.8,11.5,10.8,10.5,10.5,11.6,11.0,-99,
3,7,7,11.0,11.5,11.2,10.5,11.1,11.8,10.5,-99,
3,7,12,11.0,10.0,10.4,-99,10.5,11.2,9.8,-99,
3,7,18,10.8,10.5,9.8,-99,10.7,-99,11.2,-99,
3,8,3,11.5,10 ,10.4,10.5,10.6,10.5,11.2,-99,
3,8,7,11.2,10.0,9.6,11.0,11.4,11.0,11.0,-99,
3,8,12,11.0,11.5,10.4,-99,11.3,11.4,10.9,-99,
3,8,18,11.7,11.0,10.2,-99,11.7,-99,10.5,-99,
```

```
4,1,3,10.8,11.5,10.0,10.2,10.5,11.5,10.9,-99,
4,1,7,10.9,11.5,9.6,10.6,10.6,11.5,10.4,-99,
4,1,12,11.0,10.5,9.6 ,-99,11.2,10.9,11.0,-99,
4,1,18,10.5,10.5,10.0,-99,11.6,-99,10.9,-99,
4,2,3,11.4,11.5,10 ,10.5,10.3,11.5,10.6,-99,
4,2,7,11.5,11.5,10.5 ,11.5,11.3,11.2,12.1,-99,
4,2,12,11.2,11.0,9.6 ,-99,11.1,11.5,11.3,-99,
4,2,18,10.8,10.5,10.0,-99,11.5,-99,10.8,-99,
4,3,3,11.7,11.5,11.0,11.5,11.4,10.8,11.3,-99,
4,3,7,11.3,11.5,10.2 ,10.5,11.0,11.3,10.6,-99,
4,3,12,11.7,11.0,10.2,-99,11.3,11.4,10.9,-99,
4,3,18,10.7,11.0,10.2,-99,11.4,-99,11.3,-99,
4,4,3,11.0,10.5,10.2,10.5,10.1,11.4,11.2,-99,
4,4,7,11.2,10.5,9.8,11.0,11.6,11.0,11.3,-99,
4,4,12,11.2,11.0,9.5,-99,10.9,10.6,10.7,-99,
4,4,18,11.0,11.0,10.5,-99,12.0,-99,11.4,-99,
4,5,3,12.0,11.0,11.2,11.5,11.3,10.4,11.7,-99,
4,5,7,11.5,11.0,10,11.2,11.2,11.8,11.4,-99,
4,5,12,10.6,11.0,9.8,-99,11.1,11.4,10.9,-99,
4,5,18,11.5,10.5 ,9.8,-99,11.1,-99,11.2,-99,
4,6,3,11.5,10,10.2,11.0,10.7,11.5,11.3,-99,
4,6,7,11.1,11.0,10.5,11.2,11.7,11.4,11.1,-99,
4,6,12,11.2,11.5,10.6,-99,11.4,11.2,10.8,-99,
4,6,18,11.3,11.0,10 ,-99,11.7,-99,10.9,-99,
4,7,3,12.1,11.0,10.8,11.0,10.7,11.8,11.6,-99,
4,7,7,11.7,10.5,11.2,11.5,11.5,11.0,11.9,-99,
4,7,12,11.0,12 ,10.4,-99,11.7,10.3,10.6,-99,
4,7,18,11.0,10.0,9.8,-99,11.5,-99,11.6,-99,
4,8,3,11.0,11.5,10.4,10.0,10.6,11.2,10.2,-99,
4,8,7,11.5,10.5 ,9.6,10.6,10.8,11.8,10.6,-99,
4,8,12,11.0,10.5,10.4,-99,11.2,11.6,11.2,-99,
4,8,18,11.4,11.0,10.2,-99,11.5,-99,10.5,-99,];
```

Tanks 3005 and 3006

Data for the first two inspections for both tanks is included with the estimated location coordinates based on a 600 by 100 grid. The full data is also included in it original form.

Original

Inspection History Summary.

Functional Location: G011-SOLV-T3006

Plant: 45002

Equipment Type: TX

Equipment Number: T3006

Brief Item Description: Crude DMK

Nominal Tank 5.0mm – 6 strakes high

Date	Summary
	Date of Construction unknown.
Nov-1956	First recorded inspection. External – in order Internal – Floor plating had needle type corrosion with shell welds lightly corroded for first 18" from floor. Two (2) vertical seam welds pin holed in two locations in 2nd course pinned closed.
No date	U/T – Shell 1st strake 6.6/7.6 remainder 4.8/6.0mm
Jan – 1979	Internal: Walls – Beneath a heavy scale build up plates generally pitted with a band of corrosion 100mm wide, 300mm from floor slightly heavier but immeasurable. Floor – No change from previous examination. Roof – as viewed from floor – in order External: Shell to annular CUD losses up to 1.3mm deep x 450mm high. Nozzles in order sound to hammer test. U/T – Shell 1st strake 5.2/6.5 & remainder 5.5. Floor – 6.2/6.5. Roof – 5.5mm.
Feb – 1979	Heavy leakage occurred from sump to floor weld. Sump renewed.
Mar – 1980	No significant change from previous examination
No date	U/T – Shell 1st strake 4.7/7.0 remainder 4.6/5.2mm. Roof 4.6/5.0 with Crown 6.2mm thk
Jul – 1992	Internal: Walls – Lightly roughened beneath product deposit weld profiles good. Floor – Needle gun prepared with exposed plates roughened/corroded 0.25mm max. Roof – as viewed from floor – in order External: Shell to annular local CUD losses up to 3.5mm deep beneath the stairway. Weld overlay undertaken this outage. Nozzles in order sound to hammer test. U/T: Shell 1st strake 5.8/6.4 & remainder 4.3/5.0mm. Floor – 4.0/6.5. Roof 4.5/5.0 with Crown 6.0/6.1mm Note: All stair tread capping renewed this outage
MAY'2002	S2 EXTERNAL: Action note: 96000219719 raised to de-scale & paint shell to floor which hadn't been painted since repairs undertaken in Jul'92 plus exposed areas of roof. AGREED ACTION NOTE COMPLETION DATE: DEC'2002 U/T: Shell 1st strake 4.8/6.5 & Remainder N/A this outage. Roof – 4.5/6.0. Crown – 5.8/6.5. REF: T3005-0002-02-Utdated: 12/06/2002

Page 1 G011-SOLV-T3006 INSPECTION HISTORY SUMMARY

Original

Inspection History Summary.

Functional Location: G011-SOLV-T3005

Plant: 45002

Equipment Type: TX

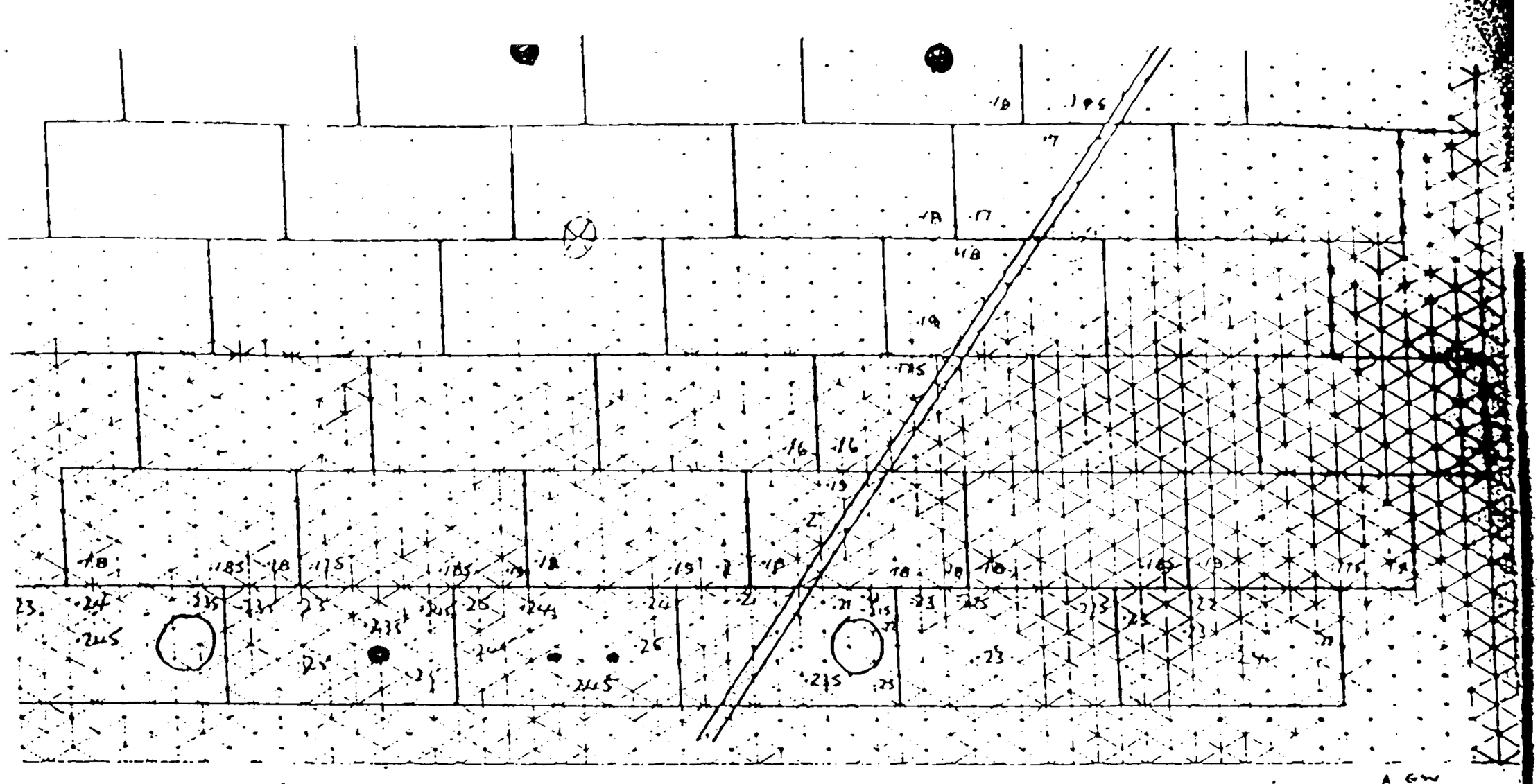
Equipment Number: T3005

Brief Item Description: Caustic Soda Tank (10% Na OH)

Nominal Tank 4.76mm – 6 strakes high

Date	Summary
	Date of Construction unknown.
Jan – 1953	First recorded inspection External – in order Internal – Plating in order with welds slightly corroded U/T – Shell 5.8 Floor 6.35mm
Jan – 1959	Internal – Floor and lower 2 strakes slightly etched with isolated patches of heavy corrosion. Welds similarly attacked.
June – 1970	U/T survey shell. 1st strake – 5.3/6.35. Remainder 4.0/5.0
Jan – 1979	Internal – walls – free from significant defects Floor – good order Roof – as viewed from floor – in order External – In order sound to hammer test U/T – 1st strake 6.3/6.5. Remainder 5.2/5.8. Floor 6.3/6.7mm. Roof – 4.9/5.2
May – 1991	U/T – 1st strake 5.9/6.5. Remainder 4.5/5.2. Roof – 5.4/6.2
Mar – 1992	External – Shell to annular CUD losses up to 3.0mm deep requiring weld overlay. Redundant instrument level leg and nozzles removed. Roof – Isolated areas of scabbing with 0.5/1.5mm less – descaled and painted (stair tread renewal this outage) Internal – Floor, shell and roof in good order – scaffold access to roof being afforded this outage. U/T – Shell 1st strake 6.0/6.4 & 4.9/5.2. Floor – 5.5 to 6. Roof – 4.9/5.2. Crown – 6.0/6.1 Civil attention to bitumen sealant required.
MAY'2002	S2 EXTERNAL: Action note. 96000219717 raised to de-scale & paint shell to those which hadn't been painted since repairs undertaken in Mar'92 plus exposed areas of roof. AGREED ACTION NOTE COMPLETION DATE: DEC'2002 U/T: Shell 1st strake 5.5/6.8 & Remainder N/A this outage. Roof – 5.5/6.5. Crown – 6.0/6.5. REF: T3005-0001-02-Utdated: 12/06/2002

PAGE 1 G011-SOLV-T3005 INSPECTION HISTORY SUMMARY



B.I.X LIMITED - ON STREAM INSPECTION DIVISION

: SHELL CHEMICALS UK. Ltd.

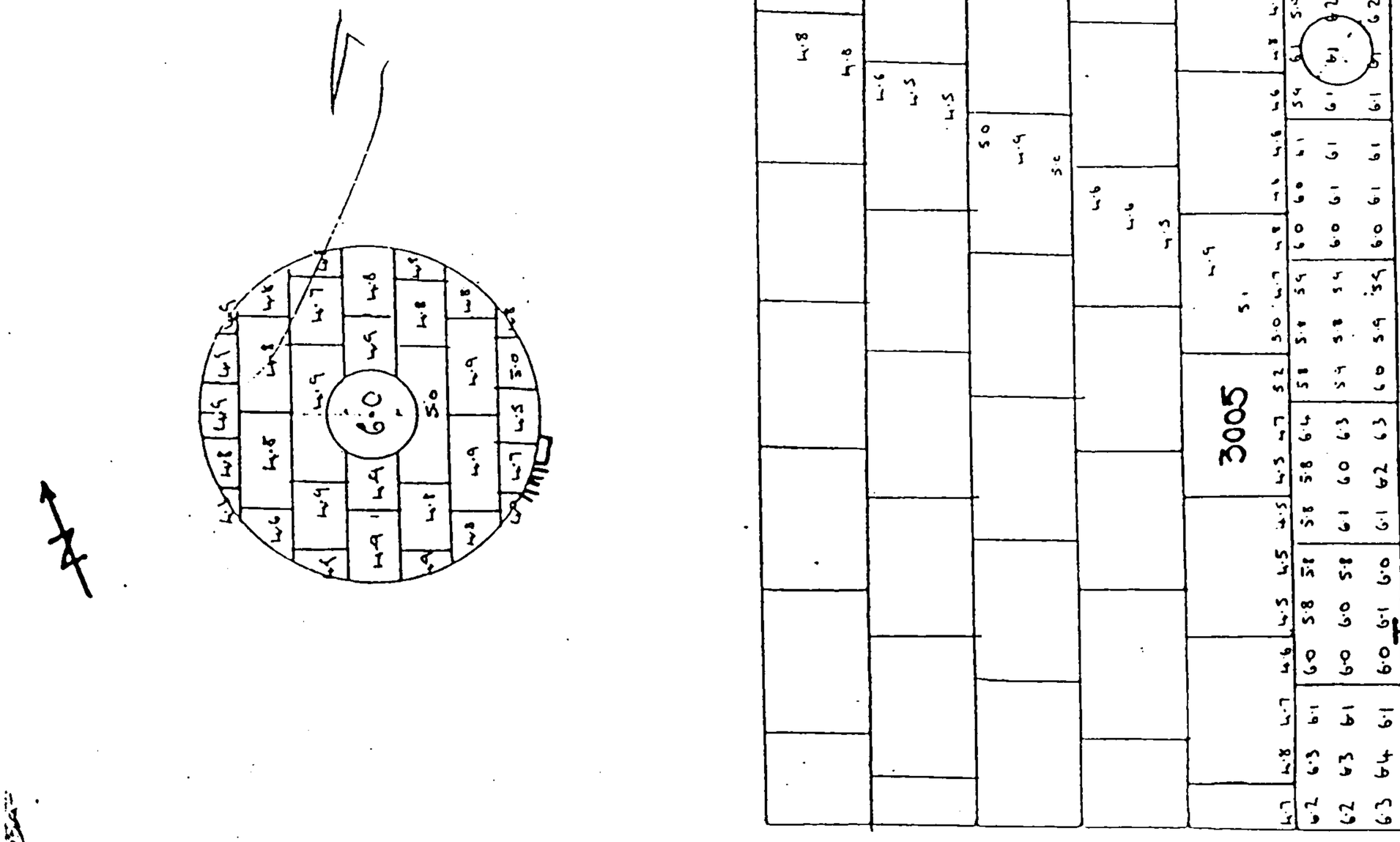
ion: STANLOW.

: L.C.P TANKAGE

Unit TANK 3005

Process Line IN SERVICE

SC. UK
Drg No 436264
SN. 24 or 25

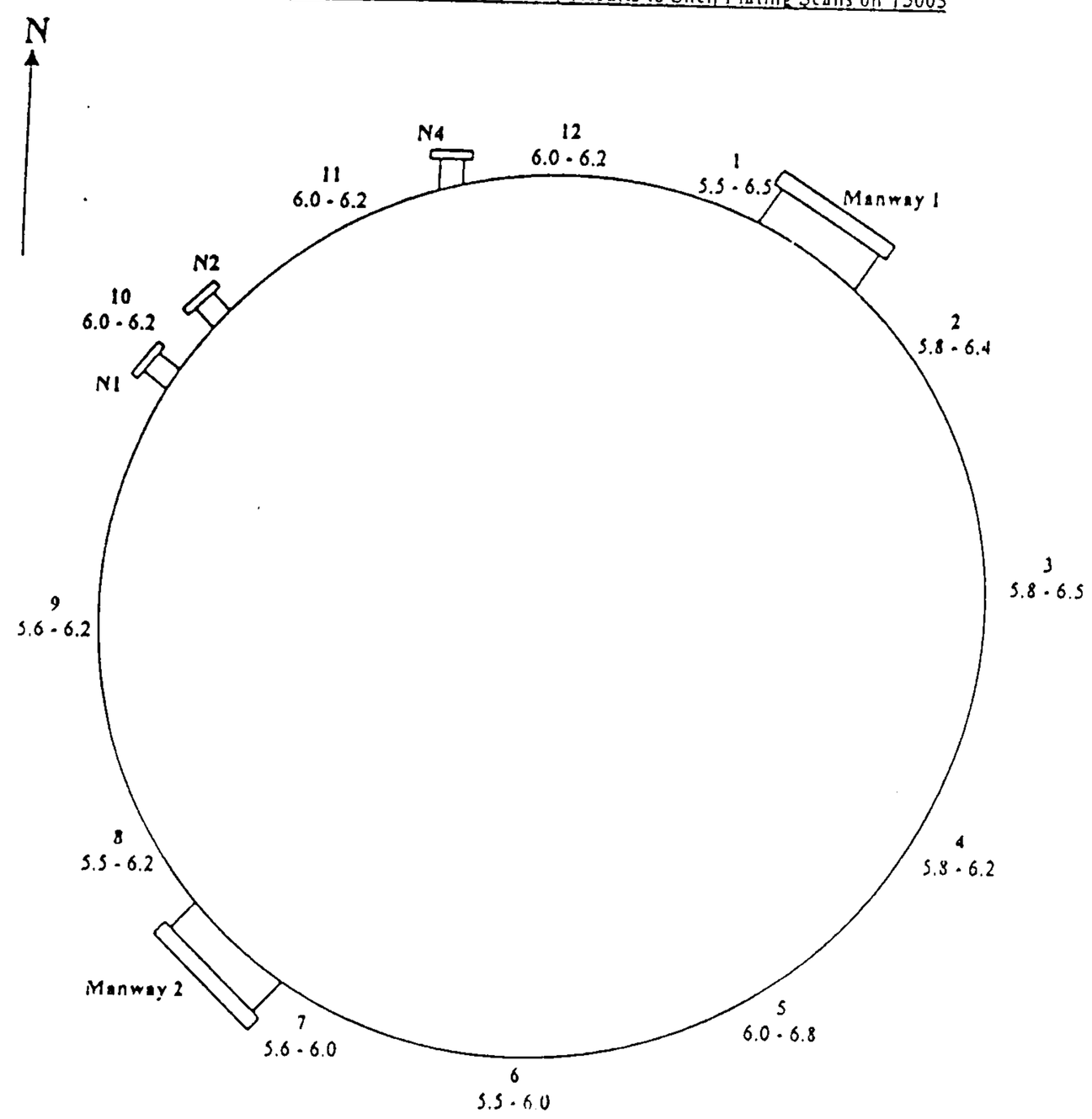


ULTRASONIC TEST REPORT

Equipment Number:GO11-TJ005
Report Number: GO11-SOLVENTS-TJ005-0001-02-UT
Page Number 2

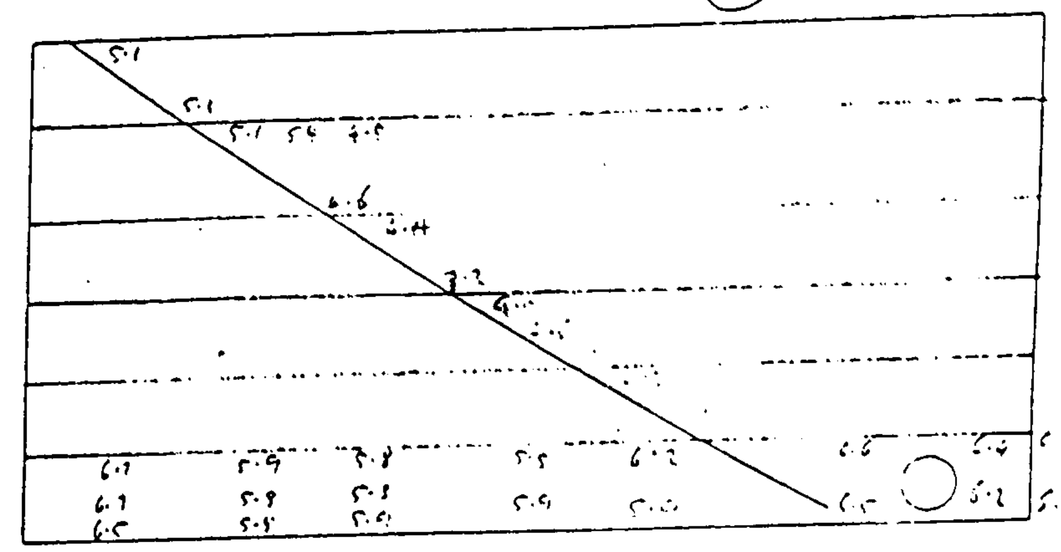
Northern

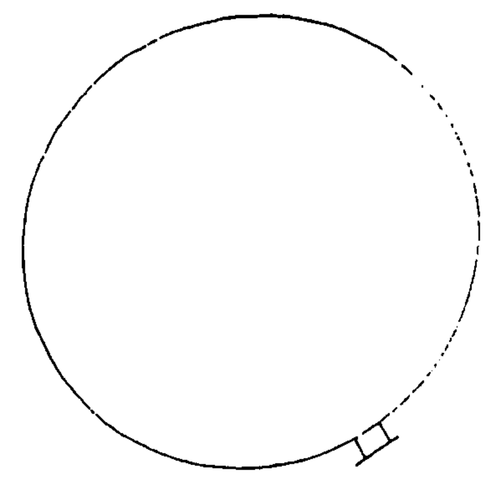
Ultrasonic Thickness Survey Results to Shell Plating Scans on TJ005



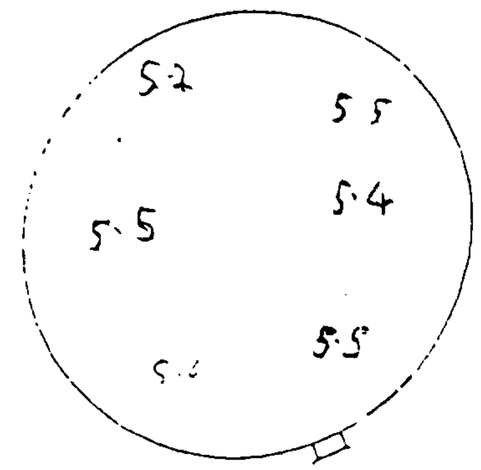
ll Recorded Readings in mm
nly the Minimum and Maximum Readings Recorded

SOLVENTS TANKAGE
TANK NUMBER 3005
DATE 30/5/91





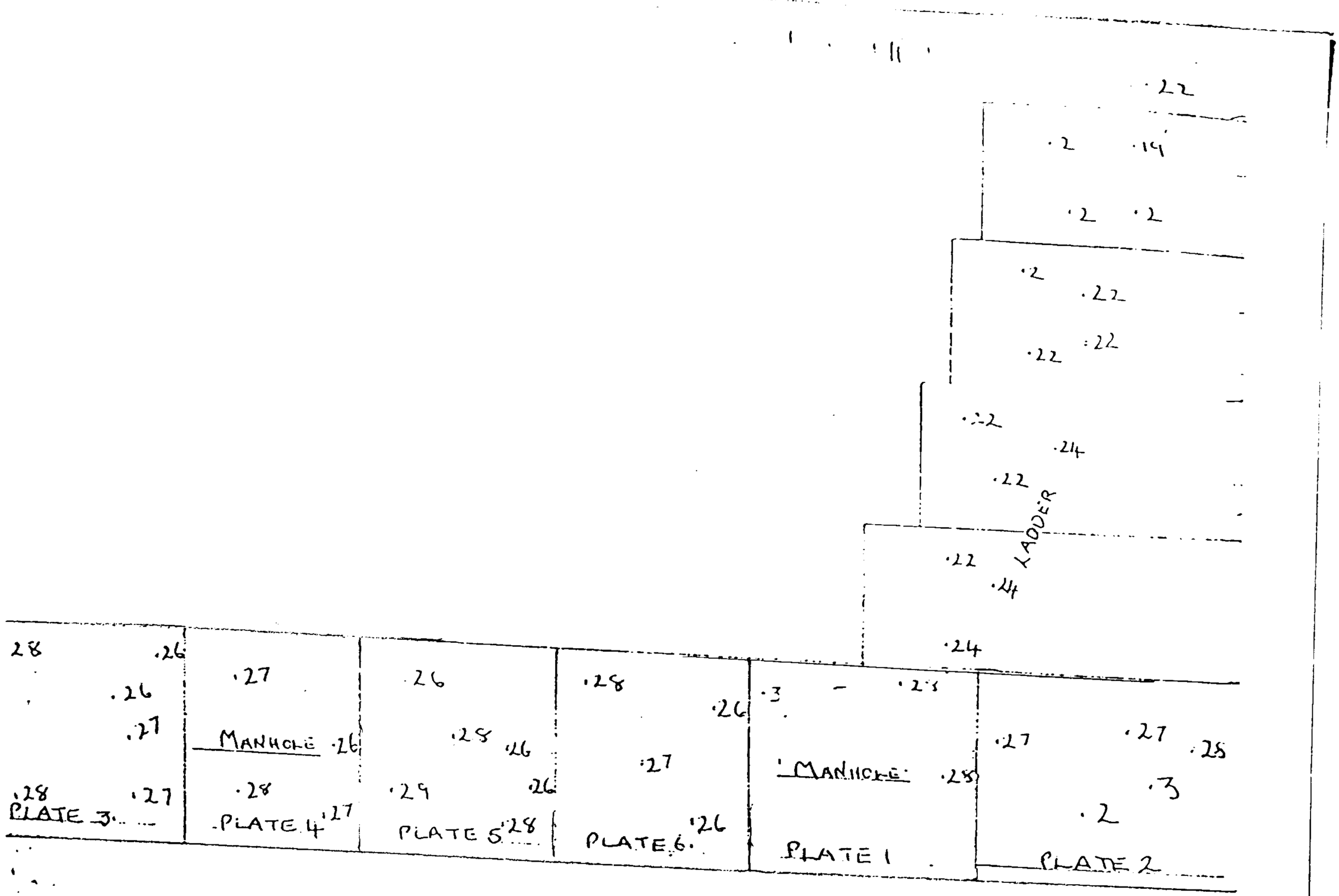
FLOOR

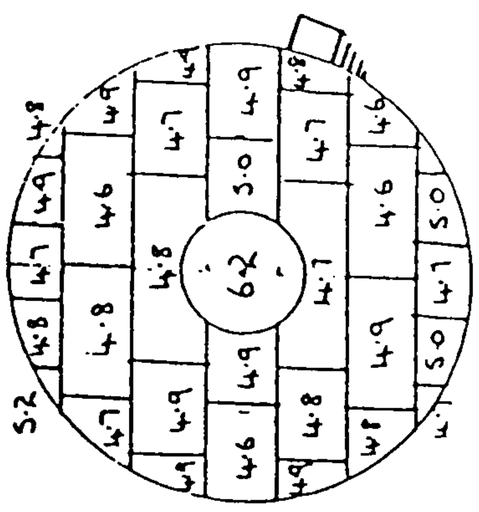


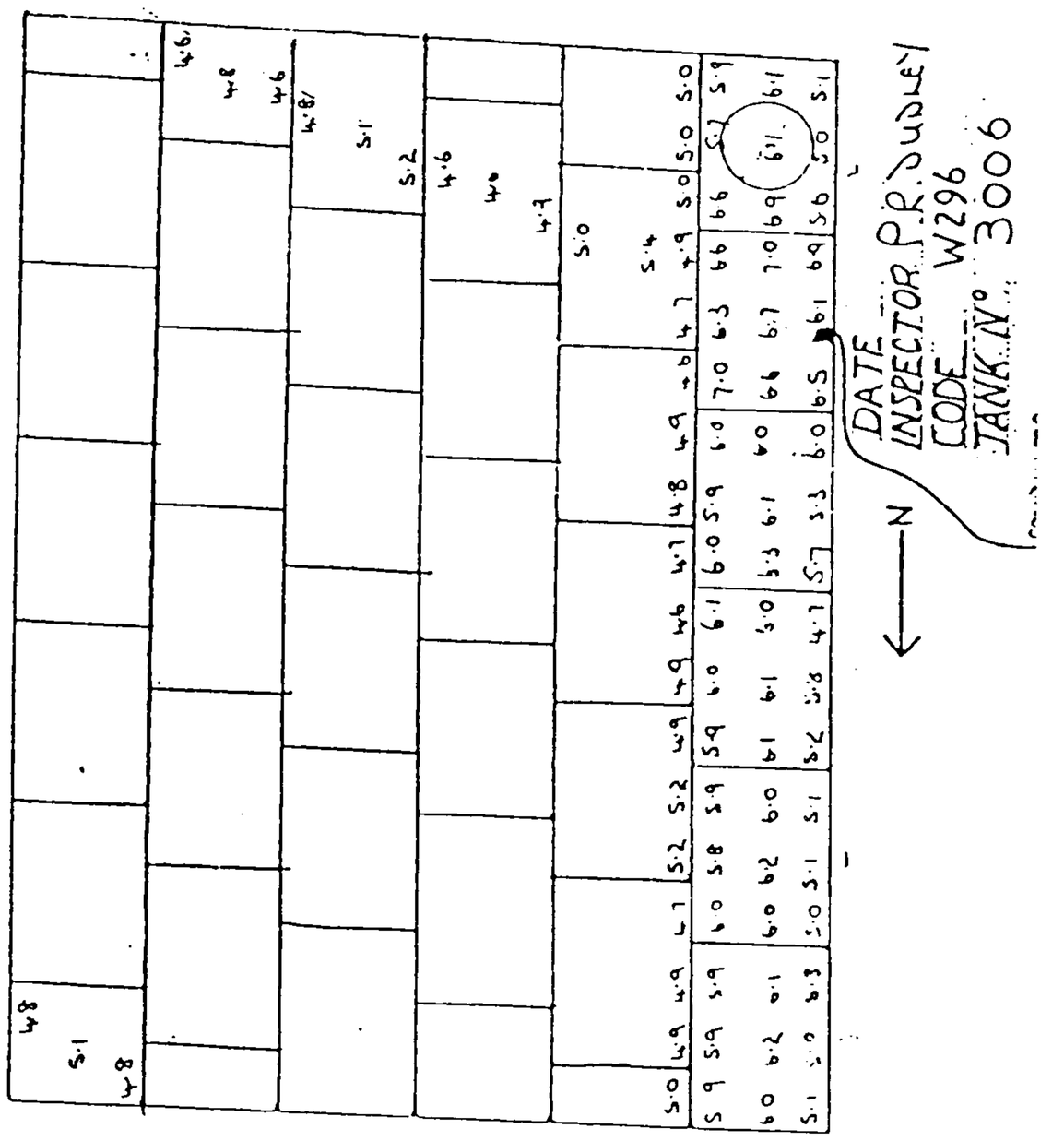
ROOF

NOMINAL THICKNESS 3/16"
SERVICE 10% Na OH
FREQUENCY 144 MONTHS.

Northern NDT Services Ltd. Unit6 Portside Business Park Portside North,Ellesmere Port, Wirral CH65-2HQ
Tel 0151 3734800 Fax 0151 3734801 e-mail ...@northernndt.co.uk


LADDER
MANHOLE
MANHOLE
PLATE 1
PLATE 2
PLATE 3
PLATE 4
PLATE 5
PLATE 6


N
62

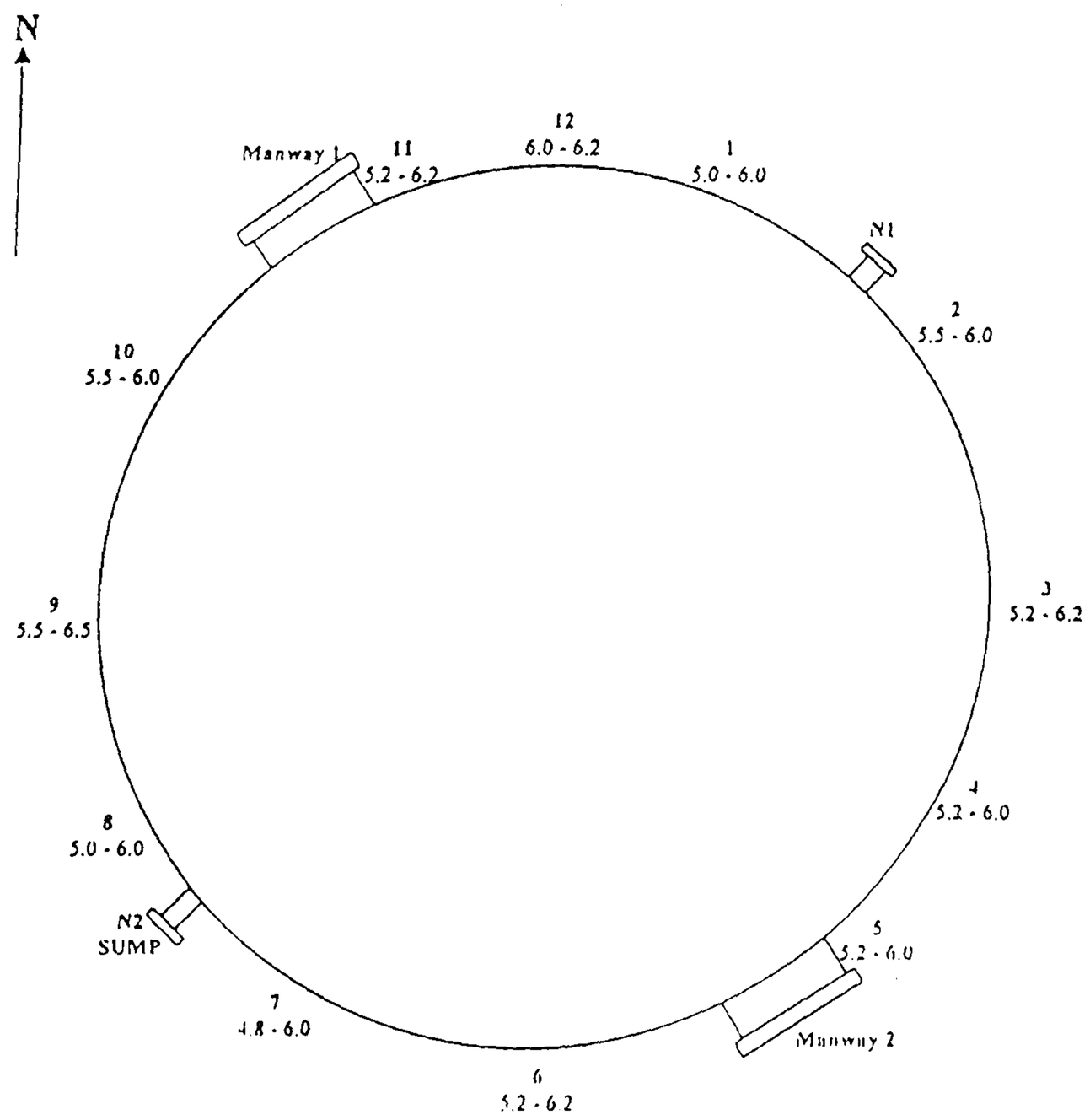

N
DATE
INSPECTOR P.R. DUDLEY
CODE W296
TANK N° 3006

ULTRASONIC TEST REPORT

Equipment Number:GO11-T3006
Report Number: GO11-SOLVENTS-T3005-0002-02-UT
Page Number 2

Northern

Ultrasonic Thickness Survey Results to Shell Plating Scans on T3006



All Recorded Readings in mm
Only the Minimum and Maximum Readings Recorded

Northern NDT Services Ltd. Unit6 Pontside Business Park Pontside North,Ellesmere Port, Wirral CH65-2HQ

NDT 171 REV 0

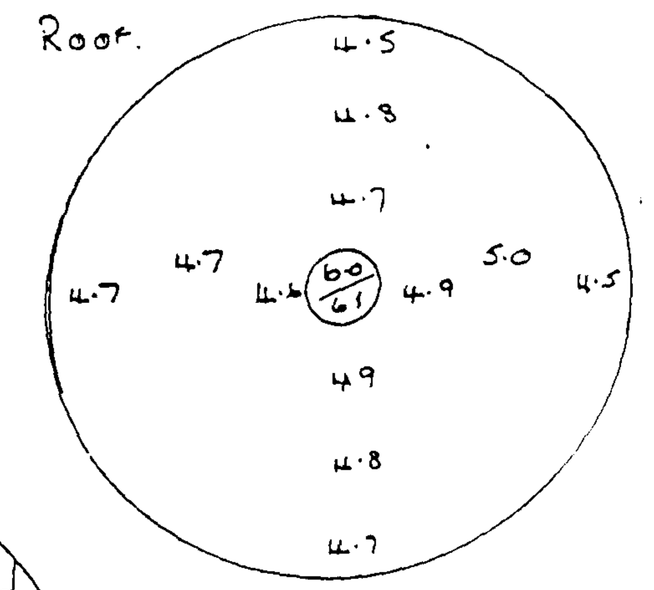
N.D.T. REPORT

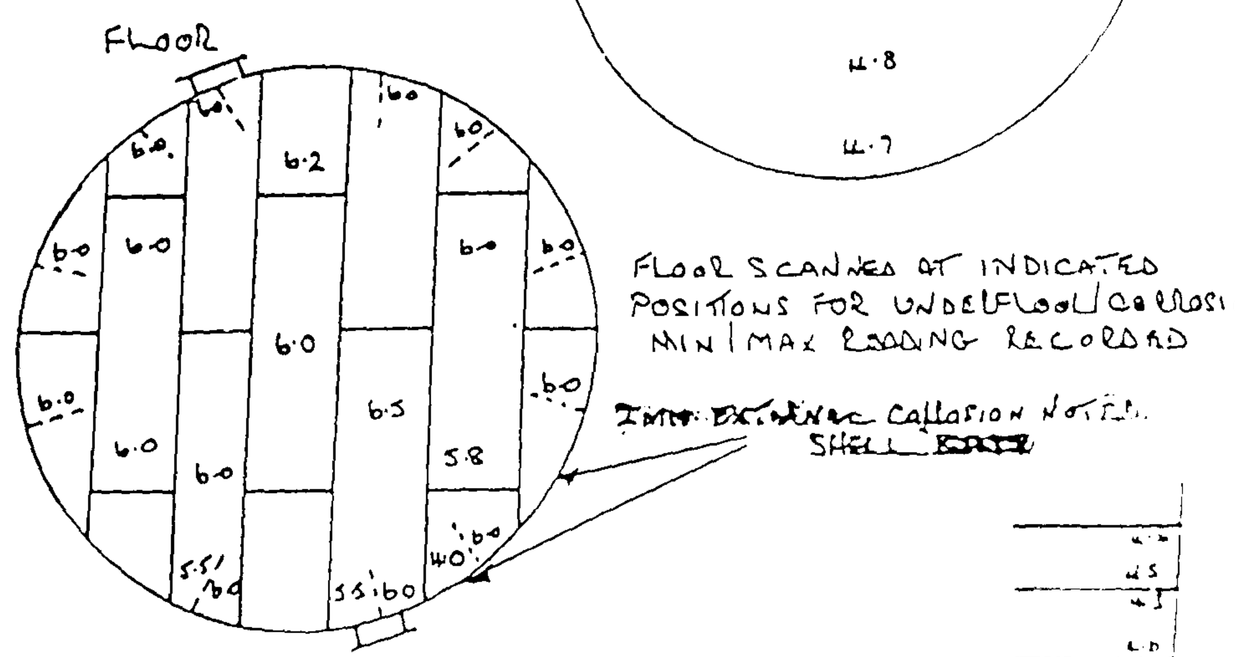
REVA 01-08-91

AREA SOLVENTS RAD/UT/MPI/DP RECORD NO. 283-92

F.A.O. K KING EIS 24.06

TANK 3006
U.T SURVEY





FLOOR SCANNED AT INDICATED POSITIONS FOR UNDERFLOOR CORROSION MIN/MAX READING RECORDED

~~INTERNAL EXTERNAL~~ CAVITATION NOTED ON SHELL ~~PLATE~~

Authorised Person

T Bamatt

Date 9.7.92

```
% tank 3005    % x,y,mm
data70=[307 82  5.842
     326 84  6.096
     326 52  6.223
     373 86  5.969
     407 82  5.969
     433 86  5.842
     430 32  6.35
     460 64  5.969
     480 23  6.35
     486 82  6.223
     505 86  6.35
     508 48  6.223
     530 82  6.096
     549 18  6.223
     584 86  6.096
     579 50  6.35
     27  89  5.334
     55  20  5.969
     88  18  5.842
     72  82  5.334
     85  92  5.08
     85  80  5.461
     93  64  5.588
     107 87  5.842
     127 87  5.715
     136 42  5.842
     180 82  5.969
     203 73  5.842
     222 89  5.588
     219 64  5.842
     251 41  6.096
     290 55  5.588]; %32 rows
% x,y,mm
data79=[410 90  6.2
     410 50  6.2
     410 10  6.3
     450 90  6.3
     450 50  6.3
     450 10  6.4
     490 90  6.1
     490 50  6.1
     490 10  6.1
     510 90  6
     510 50  6
     510 10  6
     550 90  5.8
     550 50  6
     550 10  6.1
     590 90  5.8
     590 50  5.8
     590 10  6
     10  90  5.8
     10  50  6.1
     10  10  6.1
     50  90  5.8
     50  50  6
     50  10  6.2
     90  90  6.4
     90  50  6.3
     90  10  6.3
     110 90  5.8
     110 50  5.9
     110 10  6
     150 90  5.8
     150 50  5.8
     150 10  5.9
     190 90  5.9
     190 50  5.9
     190 10  5.9
     210 90  6
     210 50  6
     210 10  6
     250 90  6
     250 50  6.1
     250 10  6.1
```

```
290 90  6.1
290 50  6.1
290 10  6.1
310 90  5.9
310 50  6.1
310 10  6.1
350 90  6.1
350 50  6.1
350 10  6.1
390 90  5.9
390 50  6.2
390 10  6.2];  %54 rows

% x,y,mm
data91=[37   10    6.7
     37   50    6.7
     37   90    6.5
     112   10    5.9
     112   50    5.8
     112   90    5.8
     187   10    5.8
     187   50    5.8
     187   90    5.9
     262   10    5.8
     262   50    -99
     262   90    5.9
     337   10    6.2
     337   50    -99
     337   90    5.9
     412   10    6.6
     412   50    -99
     412   90    6.5
     487   10    6.4
     487   50    -99
     487   90    6.2
     562   10    6
     562   50   -99
     562   90    6.1];
```

```
% x,y,min mm, max mm
data2002=[ 25   30    5.5    6.5
      75   30    5.8    6.4
     125   30    5.8    6.5
     175   30    5.8    6.2
     225   30    6    6.8
     275   30    5.5    6
     325   30    5.6    6
     375   30    5.5    6.2
     425   30    5.6    6.2
     475   30    6    6.2
     525   30    6    6.2
     575   30    6    6.2];

%tank 3006  % x,y,mm
data69=[205 86  7.112
    212 19  7.112
    271 48  6.858
    274 21  6.858
    263 68  6.604
    285 86  6.604
    329 78  6.858
    384 40  6.604
    331 26  7.112
    378 17  6.858
    422 90  6.604
    448 61  7.112
    475 57  6.604
    416 32  7.366
    489 38  6.604
    473 18  7.112
    516 86  7.112
    582 75  6.604
    545 48  6.858
    572 22  6.604
    7   84  7.62
```

```
67  92  7.112
88  50  7.112
107 67  6.858
157 74  6.858
182 67  7.112
    142 34  5.08
    168 49  7.62]; %28 rows
% x,y,mm
data80=[    210 90  5.9
    210 50  6
    210 10  5.1
    250 90  5.9
    250 50  6.2
    250 10  5
    290 90  5.9
    290 50  6.1
    290 10  5.8
    310 90  6
    310 50  6
    310 10  5
    350 90  5.8
    350 50  6.2
    350 10  5.1
    390 90  5.9
    390 50  6
    390 10  5.1
    410 90  5.9
    410 50  6.1
    410 10  5.2
    450 90  6
    450 50  6.1
    450 10  5.8
    490 90  6.1
    490 50  6
    490 10  4.7
    510 90  6
    510 50  6.3
    510 10  5.7
    550 90  5.9
    550 50  6.1
    550 10  5.3
    590 90  6
    590 50  6
    590 10  6
    10  90  7
    10  50  6.6
    10  10  6.5
    50  90  6.3
    50  50  6.7
    50  10  6.1
    90  90  6.6
    90  50  7
    90  10  6.9
    110 90  6.6
    110 50  6.9
    110 10  5.6
    150 90  5.7
    150 50  6.1
    150 10  5
    190 90  5.9
    190 50  6.1
    190 10  5.1];  %54 rows

% x,y,min mm, maxmm
data92v=[0,30,5.8,6.3;
    150,30,6.3,6.4;
    300,30,6.2,6.2;
    450,30,6,6.2]; %4 rows

% x,y,min mm, max mm
data2002v=[0,  30, 5,      6
     50, 30, 5.5,    6
    100, 30, 5.2,    6.2
    150, 30, 5.2,    6
    200, 30, 5.2,    6
    250, 30, 5.2,    6.2
```

```
300, 30, 4.8,    6
350, 30, 5,      6
400, 30, 5.5,    6.5
450, 30, 5.5,    6
500, 30, 5.2,    6.2
550, 30, 6,      6.2]; % 12 rows
```

Appendix 2: Matlab algorithms

This appendix contains some of the most important algorithms used in calculations within this thesis. Long lines have been wrapped and are indicated by ... at either end of the break.

General closed form DLM updates

General closed form updates can be made using the following algorithm.

```
function [m,C,a,R,e,Q,y,f]=DLMupdate(m,C,y,G,F,V,W,delta);
% DLM update program. All notion as West and Harrison.
% Set delta=1 for non discounted models.
a=G*m; % forecast mean
if nargin==7        %if no value for delta, no discounting assumed.
    delta=1;
end


R=G*C*G'/delta+W; % forecast variance

%%% make adjustments to y, F and V corresponding to missing data
% copy F, y and V matrices to speed up processing
ly=length(y);
count=0;
for i=1:ly;
    if y(i)~=-99
        count=count+1;
    end
end
```

```
[aaa,bbb]=size(F);
F2=zeros(aaa,count);
y2=zeros(count,1);
V2=zeros(count,count);
j=0; FF=F; yy=y; VV=V;
for i=1:ly
    if yy(i)~=-99
        F2(:,i-j)=FF(:,i);
        y2(i-j)=yy(i);
        jk=0;
        for k=i:ly
            if yy(k)~=-99
                V2(k-jk-j,i-j)=VV(k,i);
                V2(i-j,k-jk-j)=VV(i,k);
            else
                jk=jk+1;
            end
        end
    else
        j=j+1;
    end
end

y=y2;
V=V2;
F=F2;

if 1==isempty(y) % if no data is observed
    F=[];
end
if 1~=isempty(F) % if some data is observed
    f=F'*a; % forecast point means
    Q=V+F'*R*F; % observation variance
    Q=(Q+Q')/2; % Remove rounding errors by exploiting symmetry
A=R*F*(Q^-1);
[len,wid]=size(y);
if wid>len
    y=y';
```

```
    end
   e=y-f;  % errors
    m=a+A*e; % adjusted system mean
    C=R-A*Q*A'; % adjusted system variance
else % if no data is observed
    m=a;
    C=R;
    Q=[];
    e=[];
    f=[];
end
```

Comparison between a full Bayes update and a Bayes linear update

Two key programs are used in this comparison

```
function posterbase3
%define W and V
loc=sqrt(.25);
%% calculate W
ovr=sqrt(0.05);
W=zeros(32,32);   %System noise
%overallW=ovr^2; % coef for W, may want slope part as well
slopeW=1/4;
%localW=loc;   % coef for W, may want slope part as well
noi=sqrt(0.01);
time=6;
runs=4000;
for i=1:4   % must consider correlation between any of n^2 points
    for j=1:4
        for ii=1:4
            for jj=1:4
                d=sqrt((i-ii)^2+(j-jj)^2);
                W(2*((i-1)*4+j)-1,2*((ii-1)*4+jj)-1)=ovr^2*(exp(-loc*(d)));
                W(2*((i-1)*4+j),2*((ii-1)*4+jj))=slopeW*ovr^2*(exp(-loc*(d)));
            end
        end
    end
end

V=noi^2*eye(16,16);   % Obs noise diagonal matrix
[evecW,evalW]=eig(W);
eW=evecW*sqrt(abs(evalW)); % calculate transform for simulation
slope=zeros(16,1);
data=zeros(21*7,50000);
for run=1:50000 % simulate and find minimums
    for i=2:7
        x=eW*randn(32,1);
        slope=slope+x(2:2:end);
```

```
x2=x(1:2:end)+noi*randn(16,1);
data(21*(i-1)+1:21*(i-1)+16,run)=data(21*(i-2)+1:21*(i-2)+16,run)+...
...+ x2+slope;
data(21*(i-1)+17,run)=min(data(21*(i-1)+[1 2 5 6],run));
        data(21*(i-1)+18,run)=min(data(21*(i-1)+[3 4 7 8],run));
        data(21*(i-1)+19,run)=min(data(21*(i-1)+[9 10 13 14],run));
        data(21*(i-1)+20,run)=min(data(21*(i-1)+[11 12 15 16],run));
        data(21*(i-1)+21,run)=min(data(21*(i-1)+1:21*(i-1)+16,run));
    end
    slope=zeros(16,1);
end

fullercov=cov(data(:,1:50000)'); % calculate variance
EX=mean(data')';
[evecfull,evalfull]=eig(fullercov(22:end,22:end));
efull=evecfull*sqrt(abs(evalfull)); % calculate transform for simulation

% run realisations
EX=EX(22:end);
fullercov=fullercov(22:end,22:end);
for run=1:1000
    if rem(run,100)==1; disp(run); end
    [true1(run,:),true2(run,:),true3(run,:),true4(run,:),true5(run,:),...
    ...true(run,:),ansreg(run,:),ansglobal(run,:),ansfull(run,:),...
    ...ansgold(run,:)]=posterrunner3(EX,fullercov,efull,W,V);
end

% display results
disp(['parameters loc, ovr and noi for below, ' ,num2str(loc),' ,'...
 ...,num2str(ovr), ' , ' ,num2str(noi)]) disp('error true,glob,
reg, full, gold')
[0,mean(ansglobal(:,1)-true),mean(ansreg(:,1)-true),mean(ansfull(:,1)-true),
mean(ansgold(:,1)-true)]
[0,(ansglobal(:,1)-true)'*(ansglobal(:,1)-true),(ansreg(:,1)-true)'*...
...(ansreg(:,1)-true)
,(ansfull(:,1)-true)'*(ansfull(:,1)-true),(ansgold(:,1)-true)'*...
...(ansgold(:,1)-true)]
disp('s.d. true,glob,reg, full, gold')
```

```
[var(true(:,1)),mean(ansglobal(:,2)),mean(ansreg(:,2)),mean(ansfull(:,2)),
mean(ansgold(:,2))]
vv=([ansglobal(:,1),ansreg(:,1),ansfull(:,1),ansgold(:,1)]-true*[1 1 1 1])'*...
...[ansglobal(:,1)-true,ansreg(:,1)-true,ansfull(:,1)-true,ansgold(:,1)-true]/(ru
[sqrt(var(true(:,1))),sqrt(var(ansglobal(:,1))),sqrt(var(ansreg(:,1))),...
...sqrt(var(ansfull(:,1))),sqrt(var(ansgold(:,1)))]

save true.mat true true1 true2 true3 true4 true5

ttt=[true1,true2,true3,true4,true5,true];
clf
for i=1:6
    for j=i+1:6
        if j~=i
            subplot(5,5,5*(i-1)+j-1)
            plot(ttt(:,j),ttt(:,i),'.')
        end
    end
end
```

The subroutine called towards the end of this program is below. Using the previously calculated results a single data realisation is simulated and an analysis made.

```
function [true1,true2,true3,true4,true5,true,ansreg,...
...ansglobal,ansfull,ansgold]
=posterrunner3(esim,fullercov,efull,W,V);

% calculate simulation structure for W
[evecW,evalW]=eig(W);
eW=evecW*sqrt(abs(evalW));
slope=zeros(16,1);
data=zeros(21*7,1);
% simulate a data set over 7 time points
for i=2:7
    x=eW*randn(32,1);
    slope=slope+x(2:2:end);
    x2=x(1:2:end)+V*randn(16,1);
```

```
data(21*(i-1)+1:21*(i-1)+16,1)=data(21*(i-2)+1:21*(i-2)+16,1)+ x2+slope;
data(21*(i-1)+17,1)=min(data(21*(i-1)+[1 2 5 6],1));
data(21*(i-1)+18,1)=min(data(21*(i-1)+[3 4 7 8],1));
data(21*(i-1)+19,1)=min(data(21*(i-1)+[9 10 13 14],1));
    data(21*(i-1)+20,1)=min(data(21*(i-1)+[11 12 15 16],1));
    data(21*(i-1)+21,1)=min(data(21*(i-1)+1:21*(i-1)+16,1));
end

% save true results
true1=data(21*1);
true2=data(21*2);
true3=data(21*3);
true4=data(21*4);
true5=data(21*5);
true=data(end);

%%%% analyse
% regional minima
varX=fullercov;
EX=esim;
for t=1:4
    which=[(t-1)*21+16+1:t*21]; % define points to be observed
    varD=varX(which,which); % variance of data
    covXD=varX(:,which); % covariance between data and everything else
    EDX=blEDX(EX,covXD,varD,data(which),EX(which));  % adjusted expectation
    varDX=blvarDX(varX,covXD,varD); % Adjusted variance
    EX=EDX;     % update expectations
    varX=varDX; % variance/ covariance of everything else
    if t==4
        ansreg=[EDX(end),sqrt((varDX(end,end)))];
    end
end
ansreg1=ansreg;

% global minimum
varX=fullercov;
EX=esim;
for t=1:4
```

```
which=[t*21-1]; % define points to be observed
varD=varX(which,which); % variance of data
covXD=varX(:,which); % covariance between data and everything else
EDX=blEDX(EX,covXD,varD,data(which),EX(which));  % adjusted expectation
    varDX=blvarDX(varX,covXD,varD); % Adjusted variance
    EX=EDX;     % update expectations
    varX=varDX; % variance/ covariance of everything else
    if t==4
        ansglobal=[EDX(end),sqrt(varDX(end,end))];
    end
end

% full observation (Bayes linear adjustment)
varX=fullercov;
EX=esim;
for t=1:4
    which=[(t-1)*21+1:t*21]; % define points to be observed
    varD=varX(which,which); % variance of data
    covXD=varX(:,which); % covariance between data and everything else
    EDX=blEDX(EX,covXD,varD,data(which),EX(which));  % adjusted expectation
    varDX=blvarDX(varX,covXD,varD); % Adjusted variance
    EX=EDX;     % update expectations
    varX=varDX; % variance/ covariance of everything else
   if t==4
        ansfull=[EDX(end),sqrt((varDX(end,end)))];
    end
end


% full observation with full Bayesian update
varX=fullercov;
EX=esim;
for t=1:4 % update linear beliefs
    which=[(t-1)*21+1:t*21]; % define points to be observed
    varD=varX(which,which); % variance of data
    covXD=varX(:,which); % covariance between data and everything else
    EDX=blEDX(EX,covXD,varD,data(which),EX(which));  % adjusted expectation
    varDX=blvarDX(varX,covXD,varD); % Adjusted variance
```

```
    EX=EDX;      % update expectations
    varX=varDX; % variance/ covariance of everything else
end
runs=10000; % simulated from linear beliefs to establish nonlinear results
datasim=zeros(21,runs);
[evecvarDX,evalvarDX]=eig(varDX(21*5+1:21*5+16,21*5+1:21*5+16));
evardxsqrteval=evecvarDX*sqrt(max(evalvarDX,0));
for run=1:runs;
    datasim(1:16,run)=EX(21*5+1:21*5+16)+evardxsqrteval*...
    ...randn(size(evalvarDX,2),1);
    [datasim(17,run),a1]=min(datasim([1,2,5,6],run));
    [datasim(18,run),b1]=min(datasim([3,4,7,8],run));
    [datasim(19,run),c1]=min(datasim([9,10,13,14],run));
    [datasim(20,run),d1]=min(datasim([11,12,15,16],run));
    [datasim(21,run),e1]=min(datasim([17,18,19,20],run));
end
esim3=mean(datasim')';
varDX=cov(datasim');
%disp('esim')
%esim
ansgold=[esim3(21),sqrt((varDX(21,21)))];
```

Comparing inspection schemes

This section includes the three computer programs used when comparing inspection schemes based on the data in appendix 1.

```
function papersim8emaker2

times=[0+20,26+20,39+20,55+20,82+20,119+20,157+20,205+20]

data2=-99*ones(672,12);
for i=1:4*8 % put each row if data in right place.
    data2((i-1)*21+[3,7,12,18],5:12)=data((i-1)*4+1:i*4,4:11);
end
for i=1:4   %coils
    data2(21*8*(i-1)+1:21*8*i,1)=i;
end
```

```
for i=1:4*8 %pipes
    data2(21*(i-1)+1:21*i,2)=rem(i-1,8)+1;
end
for i=1:4*8*21 % height
    data2(i,3)=rem(i-1,21)+1;
end

for i=1:size(data2,1); % distance
    p=data2(i,2); h=data2(i,3); c=data2(i,1);
    data2(i,4)=21*8*2*floor(c/2)+(2*rem(c,2)-1)*(21*2*floor(p/2)+...
    ...(2*rem(p,2)-1)*h+rem(p-1,2));
end

m=zeros(672*2,1);
for i=1:672
    m(2*i-1,1)=11.5;
    m(2*i)=-0.3/12;
end
m0=m;

ratiols=0.1;
A=1;  % height
a=0.003; % 0.003 neighbourhood strength
B=0.5;  % pipes
b=0.003;  % 0.003 neighbourhood strength
KAPPA=0.25;  % Coils
kappa=0.003;   % 0.003 neighbourhood strength
D=1;        % distance
delta=0.003;   % 0.003 neighbourhood strength
obs=0.35^2;

all=0.00001; % overall variance
Cvar=0.05; % overall initial variance
t2=zeros(672*2,672*2);

for i=1:672
    for j=1:672
        tl=(A*exp(-a*abs(data2(i,3)-data2(j,3)))+B*exp(-b*abs(data2(i,2)-...
```

```
            ...data2(j,2)))+KAPPA*exp(-kappa*abs(data2(i,1)-data2(j,1)))+...
            ...D*exp(-delta*abs(data2(i,4)-data2(j,4))); % +E e^{-es}
            t2(2*i-1:2*i,2*j-1:2*j)=[tl,0;0,ratiols*tl];
        end
    end
    W=all*t2;
    C=Cvar*t2;
    C0=C;
    V=obs*eye(672,672);

    G=eye(1344);
    F=zeros(1344,672);
    for i=1:672
        F(2*i-1,i)=1;
        G(2*i-1,2*i)=1;
    end

    delta=1;
    tt=1;
    disp('go')
    for t=1:max(times)
        if times(tt)==t
            y=data2(:,tt+4);
            disp(times(tt));
            tt=tt+1;
        else
            y=-99*ones(size(data2,1),1);
        end
        [m,C,a,R,e,Q,y,f]=DLMupdate(m,C,y,G,F,V,W,1);
        if t==138
            m138=m;
            C138=C;
        end
    end
    save t2.mat t2
    save m138.mat m138
    save C138.mat C138
    y=-99*ones(21*8*4,1);
```

```
[m139,C139,a,R,e,Q,y,f]=DLMupdate(m138,C138,y,G,F,V,W,1); %observe data

[eVec139,eVal139]=eig(C139);
e139=eVec139*sqrt(eVal139);
save e139.mat e139

disp('here')
simdata139=m139+e139*randn(21*8*4*2,1)+randn(21*8*4*2,1)*obs;

if 1==1;
    % 4 obs at
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
        temp=[1;10;15;21];
        y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
    end
    [m4,C4,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m4,C4,a,R,e,Q,y,f]=DLMupdate(m4,C4,y,G,F,V,W,1); %observe data
    end
    [eVec1774,eVal1774]=eig(C4);
    e1774=eVec1774*sqrt(eVal1774);
    save e1774.mat e1774
    disp('4 done')
end

if 1==1
    % 5 obs at
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
        temp=[1;5;10;15;21];
        y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
    end
    [m5,C5,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m5,C5,a,R,e,Q,y,f]=DLMupdate(m5,C5,y,G,F,V,W,1); %observe data
```

```
end
[eVec1775,eVal1775]=eig(C5);
e1775=eVec1775*sqrt(eVal1775);
save e1775.mat e1775
    disp('5 done')
end

temp2=[21,505,162,  1,157,336,476,152,482,436 240,400,420,105,414,167,331,...
...190, 63,110,409 546, 77,441,513,603,337,357,609,107,403, 16 231, 99,651,...
...442, 72,518,237,540,473, 43,260 189,393,350,566, 36,115,479,246, 47,160,...
... 93 372, 58, 64,430,120,581,309, 19,273,660,149 568,225,509,610,346,561,...
...387,456,646,666,232 577, 51,184,267,147,504,304,526, 89,274,498 434,571,...
...450,624,656,598, 41,484,141, 31,287 155,204, 69,398,324,341,220,555,164,...
...535,446 282,379,492,193, 83,334,367,488,329, 74,412 258,524,135,127,588,...
...125,215,314];
if 1==1;
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
        temp=temp2;
        y(temp)=simdata139(2*temp-1);
    end
    [m128,C128,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m128,C128,a,R,e,Q,y,f]=DLMupdate(m128,C128,y,G,F,V,W,1); %observe data
    end
    [eVec177128,eVal177128]=eig(C128);
    e177128=eVec177128*sqrt(eVal177128);
    save e177128.mat e177128
    disp('128 done')
end

if 1==1
    y=-99*ones(21*8*4,1);
for j=1:8*4;%1:21*8*4
    temp=temp2;
    y(temp)=simdata139(2*temp-1);
end
```

```
[m25,C25,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m25,C25,a,R,e,Q,y,f]=DLMupdate(m25,C25,y,G,F,V,W,1); %observe data
end
    [eVec17725,eVal17725]=eig(C25);
    e17725=eVec17725*sqrt(eVal17725);
    save e17725.mat e17725
    disp('25 done')
end


if 1==1
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
        temp=temp2;
        temp=temp(1:16);
        y(temp)=simdata139(2*temp-1);
    end
    [m16,C16,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m16,C16,a,R,e,Q,y,f]=DLMupdate(m16,C16,y,G,F,V,W,1); %observe data
    end
    [eVec17716,eVal17716]=eig(C16);
    e17716=eVec17716*sqrt(eVal17716);
    save e17716.mat e17716
    disp('16 done')
end


if 1==1
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
    temp=temp2;
    temp=temp(1:64);
    y(temp)=simdata139(2*temp-1);
end
```

```
[m64,C64,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m64,C64,a,R,e,Q,y,f]=DLMupdate(m64,C64,y,G,F,V,W,1); %observe data
    end
    [eVec17764,eVal17764]=eig(C64);
    e17764=eVec17764*sqrt(eVal17764);
    save e17764.mat e17764
    disp('64 done')
end



if 1==1
    y=-99*ones(21*8*4,1);
    for j=1:8*4;%1:21*8*4
        temp=temp2;
        temp=temp(1:32);
        y(temp)=simdata139(2*temp-1);
    end
    [m32,C32,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m32,C32,a,R,e,Q,y,f]=DLMupdate(m32,C32,y,G,F,V,W,1); %observe data
    end
    [eVec17732,eVal17732]=eig(C32);
    e17732=eVec17732*sqrt(eVal17732);
    save e17732.mat e17732
    disp('32 done')
end


if 1==1
    % 4 orig at
    y=-99*ones(21*8*4,1);
    disp('y done')
    for j=1:8*4;%1:21*8*4
        temp=[3,6,12,18];
```

```
    y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
end
[m4orig,C4orig,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
    for t=1:38
        [m4orig,C4orig,a,R,e,Q,y,f]=DLMupdate(m4orig,C4orig,y,G,F,V,W,1); %observ
    end
    [eVec1774orig,eVal1774orig]=eig(C4orig);
    e1774orig=eVec1774orig*sqrt(eVal1774orig);
    save e1774orig.mat e1774orig
    disp('4 orig done')
end


if 1==1
    y=simdata139(1:2:end);
    [m21,C21,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
    y=-99*ones(21*8*4,1);
    for t=1:38
        [m21,C21,a,R,e,Q,y,f]=DLMupdate(m21,C21,y,G,F,V,W,1); %observe data
    end
    [eVec17721,eVal17721]=eig(C21);
    e17721=eVec17721*sqrt(eVal17721);
    save e17721.mat e17721
    disp('21 done')
end
```

Having done the once only calculation of the simulation matrices the following program runs the simulations. Many versions of this program may be run simultaneously.

```
function papersim25shell
randn('state',sum(100*clock))
load t2.mat t2
obs=0.35^2;

all=0.00001; % overall variance
```

```
Cvar=0.05; % overall initial variance

W=all*t2;
C=Cvar*t2;
V=obs*eye(672,672);

G=eye(1344);
F=zeros(1344,672);
for i=1:672
    F(2*i-1,i)=1;
    G(2*i-1,2*i)=1;
end

load m138.mat m138
load C138.mat C138
y=-99*ones(21*8*4,1);
[m139,C139,a,R,e,Q,y,f]=DLMupdate(m138,C138,y,G,F,V,W,1); %observe data
load e139.mat e139
firstsim=500;
secondsim=3000;
min1774=zeros(firstsim*secondsim,3);
min1774orig=zeros(firstsim*secondsim,3);
min1775=zeros(firstsim*secondsim,3);
min17721=zeros(firstsim*secondsim,3);
min177128=zeros(firstsim*secondsim,3);
min17716=zeros(firstsim*secondsim,3);
min17732=zeros(firstsim*secondsim,3);
min17764=zeros(firstsim*secondsim,3);

load e1774.mat e1774
load e1774orig.mat e1774orig
load e1775.mat e1775
load e17721.mat e17721
load e177128.mat e177128
load e17716.mat e17716
load e17732.mat e17732
load e17764.mat e17764
```

```
temp2=[21,505,162,  1,157,336,476,152,482,436 240,400,420,105,414,167,331,...
...190, 63,110,409 546, 77,441,513,603,337,357,609,107,403, 16 231, 99,651,...
...442, 72,518,237,540,473, 43,260 189,393,350,566, 36,115,479,246, 47,160,...
... 93 372, 58, 64,430,120,581,309, 19,273,660,149 568,225,509,610,346,561,...
...387,456,646,666,232 577, 51,184,267,147,504,304,526, 89,274,498 434,571,...
...450,624,656,598, 41,484,141, 31,287 155,204, 69,398,324,341,220,555,164,...
...535,446 282,379,492,193, 83,334,367,488,329, 74,412 258,524,135,127,588,...
...125,215,314];

for i=1:firstsim;
    i
    simdata139=m139+e139*randn(21*8*4*2,1)+randn(21*8*4*2,1)*obs;
    if 1==1
        % 4 obs at
        y=-99*ones(21*8*4,1);
        for j=1:8*4;%1:21*8*4
            temp=[1;10;15;21];
            y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
        end
        [m4,C4,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
        y=-99*ones(21*8*4,1);
        for t=1:38
            [m4,C4,a,R,e,Q,y,f]=DLMupdate(m4,C4,y,G,F,V,W,1); %observe data
        end
        disp('4 done')
        % 4 obs at
        y=-99*ones(21*8*4,1);
        for j=1:8*4;%1:21*8*4
            temp=[3;7;12;18];
            y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
        end
        [m4orig,C4orig,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe da
        y=-99*ones(21*8*4,1);
        for t=1:38
            [m4orig,C4orig,a,R,e,Q,y,f]=DLMupdate(m4orig,C4orig,y,G,F,V,W,1); %ob
        end
        disp('4 orig done')
```

```
% 5 obs at
y=-99*ones(21*8*4,1);
for j=1:8*4;%1:21*8*4
    temp=[1;5;10;15;21];
    y(21*(j-1)+temp)=simdata139(2*(21*(j-1)+temp)-1);
end
[m5,C5,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m5,C5,a5,R5,e5,Q5,y,f5]=DLMupdate(m5,C5,y,G,F,V,W,1); %observe data
end

disp('5 done')

y=-99*ones(21*8*4,1);
temp=temp2;
y(temp)=simdata139(2*temp-1);
[m128,C128,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
    for t=1:38
        [m128,C128,a,R,e,Q,y,f]=DLMupdate(m128,C128,y,G,F,V,W,1); %observe da
    end
    disp('128 done')
end


y=simdata139(1:2:end);
[m21,C21,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m21,C21,a,R,e,Q,y,f]=DLMupdate(m21,C21,y,G,F,V,W,1); %observe data
end
disp('21 done')

y=-99*ones(21*8*4,1);
temp=temp2;
temp=temp(1:16);
y(temp)=simdata139(2*temp-1);
```

```
[m16,C16,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m16,C16,a,R,e,Q,y,f]=DLMupdate(m16,C16,y,G,F,V,W,1); %observe data
end
disp('16 done')


y=-99*ones(21*8*4,1);
temp=temp2;
temp=temp(1:32);
y(temp)=simdata139(2*temp-1);
[m32,C32,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m32,C32,a,R,e,Q,y,f]=DLMupdate(m32,C32,y,G,F,V,W,1); %observe data
end
disp('32 done')


y=-99*ones(21*8*4,1);
temp=temp2;
temp=temp(1:64);
y(temp)=simdata139(2*temp-1);
[m64,C64,a,R,e,Q,y,f]=DLMupdate(m139,C139,y,G,F,V,W,1); %observe data
y=-99*ones(21*8*4,1);
for t=1:38
    [m64,C64,a,R,e,Q,y,f]=DLMupdate(m64,C64,y,G,F,V,W,1); %observe data
end
disp('64 done')


simdata139=zeros(2*21*8*4,secondsim);
for j=1:secondsim
rr=randn(21*8*4*2,1);
if 1==1
    simdata1774=m4+e1774*rr;
    min1774((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
```

```
...min(simdata1774(1:2:end))];
if min(simdata1774(1:2:end))<6;
    min1774((i-1)*secondsim+j,3)=1;
end

simdata1774orig=m4orig+e1774orig*rr;
min1774orig((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata1774orig(1:2:end))];
if min(simdata1774orig(1:2:end))<6;
    min1774orig((i-1)*secondsim+j,3)=1;
end

simdata1775=m5+e1775*rr;
min1775((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata1775(1:2:end))];
if min(simdata1775(1:2:end))<6;
    min1775((i-1)*secondsim+j,3)=1;
end

    simdata177128=m128+e177128*rr;
    min177128((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
    ...min(simdata177128(1:2:end))];
    if min(simdata177128(1:2:end))<6;
        min177128((i-1)*secondsim+j,3)=1;
    end
end
simdata17716=m16+e17716*rr;
min17716((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata17716(1:2:end))];
if min(simdata17716(1:2:end))<6;
    min17716((i-1)*secondsim+j,3)=1;
end

simdata17732=m32+e17732*rr;
min17732((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata17732(1:2:end))];
if min(simdata17732(1:2:end))<6;
    min17732((i-1)*secondsim+j,3)=1;
```

```
end

simdata17764=m64+e17764*rr;
min17764((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata17764(1:2:end))];
if min(simdata17764(1:2:end))<6;
    min17764((i-1)*secondsim+j,3)=1;
end


simdata17721=m21+e17721*rr;
min17721((i-1)*secondsim+j,1:2)=[(i-1)*secondsim+j,...
...min(simdata17721(1:2:end))];
if min(simdata17721(1:2:end))<6;
    min17721((i-1)*secondsim+j,3)=1;
end


if rem(j,20)==0
            save min1774shell.mat min1774
            save min1774origshell.mat min1774orig
            save min1775shell.mat min1775
            save min17721shell.mat min17721
            save min177128shell.mat min177128
            save min17716shell.mat min17716
            save min17732shell.mat min17732
            save min17764shell.mat min17764
        end
    end
end
save min1774shell.mat min1774
save min1774origshell.mat min1774orig
save min1775shell.mat min1775
save min177128shell.mat min177128
save min17721shell.mat min17721
save min17716shell.mat min17716
save min17732shell.mat min17732
save min17764shell.mat min17764
```

Results from many copies of the above program run simultaneously. The following program compiles results from two of programs but may easily be extended.

```
function z=papercompiler25;
if 1==1
    load min17721hob2.mat
    load min1774hob2.mat
    load min17716hob2.mat
    load min17732hob2.mat
    load min17764hob2.mat
    load min1775hob2.mat
    load min177128hob2.mat
    load min1774orighob2.mat
    for i=0:3000:max(min1775(:,1))-1;
        z1(i/3000+1,:)=mean([min17716(i+1:3000+i,[2,3]),min17732(i+1:3000+i,...
        ...[2,3]),min17764(i+1:3000+i,[2,3]),min17721(i+1:3000+i,[2,3]),...
        ...min1774(i+1:3000+i,[2,3]),min1775(i+1:3000+i,[2,3]),min177128...
    ...(i+1:3000+i,[2,3]),min1774orig(i+1:3000+i,[2,3])]);
end
load min17721cau.mat
load min1774cau.mat
load min17716cau.mat
load min17732cau.mat
load min17764cau.mat
load min1775cau.mat
load min177128cau.mat
load min1774origcau.mat
for i=0:3000:max(min1775(:,1))-1;
    z2(i/3000+1,:)=mean([min17716(i+1:3000+i,[2,3]),min17732(i+1:3000+i,...
    ...[2,3]),min17764(i+1:3000+i,[2,3]),min17721(i+1:3000+i,[2,3]),...
    ...min1774(i+1:3000+i,[2,3]),min1775(i+1:3000+i,[2,3]),...
    ...min177128(i+1:3000+i,[2,3]),min1774orig(i+1:3000+i,[2,3])]);
    end
end
disp('min17716, min17732,min17764,min17721,min1774,min1775,min177128,min1774orig'
z=[z1;z2]
```

```
clf
hold on
if 1==1
    plot(z(:,8),z(:,2),'.k')
    plot(z(:,8),z(:,4),'ob')
    plot(z(:,8),z(:,6),'xg')
    plot(z(:,8),z(:,14),'sm')
    plot(z(:,8),z(:,10),'+r')
    plot(z(:,8),z(:,12),'*c')
end
plot(z(:,8),z(:,16),'dg')
legend('min17716', 'min17732','min17764','min177128','min1774',...
...'min1775','min1774orig',-1)
legend('16 points', '32 points','64 points','128 points','4 intermediate',...
...'5 intermediate','exisiting',-1)
%legend('min177128','min1774orig',-1)
[(z(:,8)-z(:,2))'*(z(:,8)-z(:,2))
    (z(:,8)-z(:,4))'*(z(:,8)-z(:,4))
    (z(:,8)-z(:,6))'*(z(:,8)-z(:,6))
    (z(:,8)-z(:,10))'*(z(:,8)-z(:,10))
    (z(:,8)-z(:,12))'*(z(:,8)-z(:,12))
    (z(:,8)-z(:,14))'*(z(:,8)-z(:,14))
    (z(:,8)-z(:,16))'*(z(:,8)-z(:,16))]
xlabel('True risk')
ylabel('Observed risk')
```