Methods for enhancing system dynamics modelling: state-space models, data-driven structural validation & discrete-event simulation

Mark Andrew Bell, B.Sc.(Hons.), M.Sc.

Submitted for the degree of Doctor of Philosophy at Lancaster University.

September 2015

STOR-i
excellence with impact

Abstract

System dynamics (SD) simulation models are differential equation models that often contain a complex network of relationships between variables. These models are widely used, but have a number of limitations. SD models cannot represent individual entities, or model the stochastic behaviour of these individuals. In addition, model parameters are often not observable and so values of these are based on expert opinion, rather than being derived directly from historical data. This thesis aims to address these limitations and hence enhance system dynamics modelling. This research is undertaken in the context of SD models from a major telecommunications provider.

In the first part of the thesis we investigate the advantages of adding a discrete-event simulation model to an existing SD model, to form a hybrid model. There are few examples of previous attempts to build models of this type and we therefore provide an account of the approach used and its potential for larger models. Results demonstrate the advantages of the hybrid's ability to track individuals and represent stochastic variation.

In the second part of the thesis we investigate data-driven methods to validate model assumptions and estimate model parameters from historical data. This com-

mences with use of regression based methods to assess core structural assumptions of the organisation's SD model. This is a complex, highly nonlinear model used by the organisation for service delivery. We then attempt to estimate the parameters of this model, using a modified version of an existing approach based on state-space modelling and Kalman filtering, known as FIMLOF. One such modification, is the use of the unscented Kalman filter for nonlinear systems. After successfully estimating parameters in simulation studies, we attempt to calibrate the model for 59 geographical regions. Results demonstrate the success of our estimated parameters compared to the organisation's default parameters in replicating historical data.

Acknowledgements

I would firstly like to thank my supervisors, Dr Matt Nunes, Dr Dave Worthington and Professor Peter Neal. I am honoured to have had the opportunity to work with all of them. Their guidance, support and encouragement have been invaluable. A special thank you to Matt for his guidance and close supervision throughout chapters 4-6, which has contributed to me developing a passion for, and pursuing, further research in the area.

A sincere thanks also to Dr Kjeld Jensen at BT – for the support, helpful comments and also for providing me with the opportunity to gain insights into research at a large organisation. A special thank you to Dr Chris Nemeth, for numerous helpful discussions, in particular on state-space modelling and Bayesian inference. Thanks also to Drs Dennis Prangle, Roger Brooks and Stephan Onggo for further helpful discussions. This final version of the thesis has been improved thanks to the helpful comments of my viva examiners, Professor Michael Pidd and Dr Kostas Triantafyllopoulos.

My PhD has been supported by the EPSRC funded Statistics and Operational Research (STOR-i) doctoral training centre, and by BT Research and Innovation. I am grateful to STOR-i for providing a stimulating research environment, and for the

additional training opportunities that have allowed me to develop my research skills. In particular, I would like to thank Professors Jon Tawn and Idris Eckley for their guidance and support.

I have met many wonderful people who have made my time in Lancaster enjoyable. In particular, I would to like thank my friends Drs Chris Nemeth, Matt Nunes, Sarah Taylor and Shreena Patel. We've had some great times.

Finally, I would like to thank my parents for their constant support and encouragement, not only during my PhD, but in all of my academic pursuits. Words cannot express how much I appreciate it. Most important of all, I am eternally grateful to my partner, Lorna. Her patience and unyielding support has made all of this possible. Thank you.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Mark Andrew Bell

Contents

List of Figures

List of Tables

List of Abbreviations

BT	British Telecommunications plc
OR	Operational Research
SD	System Dynamics
DES	Discrete-Event Simulation
RFT	Right First Time
MA	Moving Average
CI	Confidence Interval
KF	Kalman Filter
DLM	Dynamic Linear Model
FIMLOF	Fully Integrated Maximum Likelihood via Optimal Filtering
MRO	Model Reference Optimisation
MAPE	Mean Absolute Percentage Error
OLS	Ordinary Least Squares
UKF	Unscented Kalman Filter
RMSE	Root Mean Squared Error

Chapter 1

Introduction

This thesis investigates the feasibility and value of implementing recent developments in operational research and statistics, in order to enhance System Dynamics (SD) models. This thesis is split into three sections. Each of these sections aims to add value to SD modelling by addressing one of the limitations of SD models below.

1. The low–detailed approach of SD means that individual entities in the system cannot be represented. In addition, there is no means of representing the stochastic variation in the behaviour of these individuals.

2. A lack of statistical rigour often exists in the validation of these models.

3. Model parameters may not be observable and so values of these are often based on system knowledge/estimates, rather than being derived directly from historical data.

This research is supported by BT Research and Innovation. Although the approaches presented throughout this thesis are applied to models and data that are from Open-

reach – a separate company that is owned by the BT Group – for simplicity we refer to our industrial partner as BT. BT make extensive use of a large SD model (referred to as their *hydraulics model*) to plan and manage their service delivery processes. Results from the hydraulics model include prediction of key performance metrics. These enable BT to project how well their system will perform over a period of up to 12 months ahead. Consequently the output of the model influences key decisions within the organisation.

This thesis seeks to address the three limitations of SD models described above to determine approaches for improving SD models in practice, using the BT model as a guide to the type of real limitations that exist in SD models, and to demonstrate and evaluate the potential solutions that recent developments from OR and statistics can offer. The main chapters of this thesis (chapters 2-6) are split into three sections. Each of these sections investigates a different approach that aims to address one of the three SD modelling limitations described above. Since this thesis consists of three distinct sections, there are separate, self-contained literature reviews within each section. We detail below these three sections of the thesis.

Chapter 2: System dynamics & discrete-event hybrid models

This section aims to address SD modelling limitation (1) by adding a discrete-event simulation (DES) model to an existing SD model. DES is an alternative modelling approach to SD that offers a completely different perspective of the system. Whereas SD modelling requires a more distant view of the system in order to model large

systems over long time scales, DES is a much more 'zoomed-in' approach capable of capturing greater detail. As such, there is reason to believe that the best of both worlds can be achieved by combining the two approaches in a hybrid model. This chapter investigates the added benefits of adding a DES model to an existing SD model to form a hybrid.

The SD model used is a simplified version of BT's hydraulics model. A DES model that is capable of individually representing large numbers of BT engineers is added to the SD model. The DES part is shown to provide added capabilities to the standalone SD model. The hybrid model's ability to represent stochastic variation and track individual jobs through the system allows important performance measures to be calculated. SD models cannot track individuals (e.g. jobs or engineers) through the system and so previously, BT estimated one of these performance measures via a regression relationship with SD model output. When using the hybrid model, the measure can be estimated directly and does not need to rely on the regression relationship.

Results demonstrate the effects of stochastic variation on the performance measure; i.e. more variation results in poorer performance. Results also demonstrate the advantages of using the hybrid model to estimate the performance measure directly. In addition, the weaknesses of the current regression approach used by BT and the potential for misleading results are demonstrated.

Chapter 3: Data-driven structural validation of SD models

This section aims to address SD modelling limitation (2) by using data-driven techniques to strengthen existing structural validation tests of SD models. The core assumptions of the BT hydraulics model are evaluated from analysis of historical time series data.

Within the model, during periods of increased demand, feedback mechanisms aim to replicate management decisions and increase workforce numbers. The increased workforce numbers are then able to reduce the job queue back to a manageable level. The analysis aims to investigate whether the time series data revealed any evidence of this feedback. Also of interest is determining whether this increase is linear, and if the feedback operates in the same way across the different geographical regions of the UK.

The exploratory analysis within this section is based on a traditional regression approach. Overall, this consists of the process below.

- Defining increased demand periods: Methods are used to treat the raw time series and detect increases in demand above a defined threshold.
- Tracking system changes: Having identified these increased demand intervals, explanatory variables are defined to track changes in the time series. A suitable response variable measuring changes in the engineer numbers is defined to measure the system's response to the increased demand.
- Regression: Assuming that a response from the system could be observed, there

is little prior knowledge of the type of response that could be expected. Therefore linear regression is chosen as the method to provide an initial indication of the nature of any response to demand increases.

Results provide strong evidence that a response to increases in demand was observable and depended on the characteristics of the increase. However, contrary to current assumptions at BT, results also suggest that the response provided by different geographical regions may not be the same.

This section can also be viewed as laying the foundations prior to chapters 4-6, where we attempt to estimate the parameters of the hydraulics model. If the data had revealed insufficient evidence of a response to increased demand, estimating the parameters of an SD model with a questionable structure would potentially be of little value.

Chapters 4-6: Using State-space methods to calibrate the SD model

This section aims to address SD modelling limitation (3) by estimating the parameters of a SD model from historical time series data. BT analysts have a high degree of confidence in the structure and core assumptions of the hydraulics model. However there is less confidence in the model parameters. At BT, as is often the case with SD modelling, model parameters are based on expert knowledge/estimates and are not derived directly from data, since data for these parameters is often not observable. Currently, the SD model is calibrated by hand using a cumbersome trial and error procedure involving repeated runs of the model and comparing the output to

historical data - until the output from the model is a sufficiently good match. This process can be time-consuming and there is no guarantee that the parameters found are optimal.

This section, throughout chapters 4-6, presents a modified version of an existing approach to automatically calibrate the hydraulics model. Model parameters are estimated directly from time series data. This approach is based on representing the system as a state-space model and using filtering techniques to estimate the underlying state of the time series from what is assumed to be noisy data. Chapters 4 and 5 test the approach on simulated data, initially on simplified versions of the hydraulics model before progressing to the full version of the model. In chapter 6 we attempt to calibrate the hydraulics model to historical time series data for 59 geographical regions of the UK.

Chapter 4 introduces the theory behind the approach and presents the literature review for this section. This covers the origins of the method and its developments and applications in recent years. We present a considerably simplified version of the hydraulics model which in terms of the key variables is a linear representation of the system. Simulation studies demonstrate the effectiveness of our approach when attempting to estimate the known parameters of simulated data with added noise. Results also reveal that our modifications to the existing algorithm improve the accuracy of parameter estimates.

Chapter 5 introduces two versions of the hydraulics model, both of which are nonlinear. The first model is only slightly more complex than the model in chapter 4 but the second model is effectively the full version of the model as used by BT. This is a considerable leaps in terms of complexity and it is also highly nonlinear. These two nonlinear versions of the hydraulics model require adjustments to our approach. A different filtering technique is required to estimate the underlying state of the time series from noisy data. For both versions of the hydraulics model, simulation studies demonstrate that when our selected filter is incorporated into our algorithm for parameter estimation, we are able to estimate the known parameters from simulated data with added noise.

In chapter 6, we apply our algorithm to historical BT time series data for 59 geographical regions. In doing so we aim to calibrate the hydraulics model for each region and fit the model's output to match the historical data. We describe some further adjustments that are required to the algorithm. Results of using our parameter estimates are compared to two sets of BT parameters. In this application, although limitations of our approach are exposed, a practical solution that can be implemented by BT is shown to be more successful at calibrating the model than the BT parameters.

Chapter 2

Constructing hybrid simulation models using system dynamics and discrete-event methodologies

2.1 Introduction

System dynamics (SD) simulation models are widely used in a variety of settings, from modelling organisations (Forrester, 1968a) to economic systems (Radzicki et al., 2004). However, the SD approach is not without its limitations. Its low–detailed approach means that individual entities in the system cannot be modelled and there is no means of representing the stochastic behaviour of these individuals. Discrete-event simulation (DES) models do not possess such limitations. Hence combining SD and DES methods in a hybrid model can overcome these limitations.

The objective of this chapter is to review the recent literature on SD/DES hybrid simulation methods and to investigate the effects of applying one such method for adding a DES model to BT's (SD) hydraulics model, in terms of both model development and the extra insights that it is capable of producing.

This chapter is organised as follows. Section 2.2 presents a literature review. Section 2.3 presents the SD model that will be used to construct the SD/DES hybrid model and describes the application to BT's systems. Section 2.4 presents details of the approach for constructing the hybrid model. Section 2.5 presents results from the hybrid model and demonstrates the benefits of this over the standalone SD model for BT. Section 2.6 presents the discussion.

2.2 Literature review

In this section we present the literature review. Section 2.2.1 introduces SD and DES. Section 2.2.2 compares the two approaches and discusses the limitations of each. Section 2.2.3 describes previous attempts at constructing SD/DES hybrid models and discusses the challenges involved.

2.2.1 System dynamics and discrete-event simulation

In operational research, simulation modelling involves constructing a computer model to represent what is usually a complex system (e.g. a hospital department or supply chain), in order to increase understanding of the system and enable experimentation.

Benefits of such models are that few assumptions are required and the results can be understood by non-experts (Robinson, 2004). Two of the most commonly used simulation techniques are system dynamics (SD) and discrete–event simulation (DES).

SD was first devised by Jay Forrester (1960, 1961, 1968b) in his work on 'Industrial Dynamics'. The technique is governed by the idea that the dynamic behaviour that is exhibited by a system is caused solely by the structure present in that system (Morecroft and Robinson, 2005). SD models the problem from a bird's eye perspective; from the top all the way down by treating populations as homogeneous and taking aggregations and averages, rather than focussing on details such as the behaviour of individual entities (e.g. patients or manufacturing parts) in the system. In addition, most SD models are entirely deterministic. SD is usually used to answer questions at a strategic rather than operational level, for example to understand an organisation's overall strategy.

In constructing SD models, software packages such as Vensim (2010), enable the user to create stock and flow diagrams. These are used to both visually represent the structure of the model and to program the equations that determine the relationships between each of the variables. An example of such a diagram, representing a word of mouth model taken from Sterman (2000), is shown in Figure 2.2.1. Stocks and flows are effectively sources and sinks which determine the flow of entities through the system. 'Potential adopters' and 'adopters' are stocks in the model that can store entities. 'New adopters' is a flow representing the transition of entities where poten-

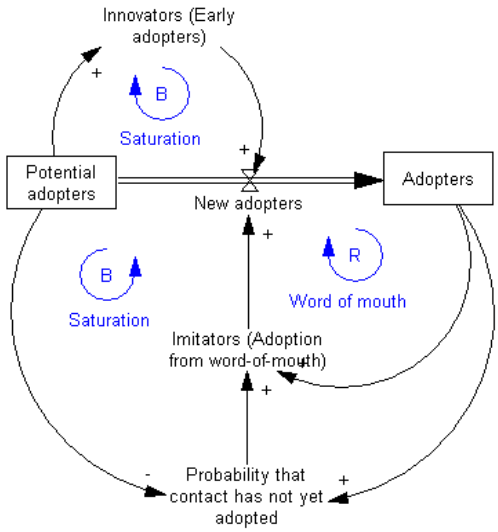

Figure 2.2.1: Stock and flow diagram of word of mouth model from Sterman (2000).

tial adopters become adopters. This flow of entities (and hence the variables in a SD model) is continuous, 'like a fluid, flowing through a system of reservoirs or tanks connected by pipes', as Brailsford and Hilton (2001) observe. The equations that define this flow are first order differential equations.

Central to the SD approach is the idea that feedback from the system affects the dynamic behaviour of the system over time. In Figure 2.2.1, feedback loops are denoted by blue arrows with 'B' and 'R' denoting a balanced and a reinforcing loop respectively. The type of feedback loop present can have significant consequences on the overall behaviour of the system. A balanced loop ensures that the system regulates itself whereas a reinforcing loop leads to uncontrolled growth (Brailsford and Hilton, 2001). See Sterman (2000) and Morecroft (2007) for thorough background texts on both qualitative and quantitative aspects of SD modelling.

DES has a completely different approach to SD. Rather than taking a bird's eye view, DES 'zooms in' on the system and attempts to capture a high level of detail. Populations are treated as heterogeneous and there is a high degree of flexibility for modelling these differences within populations. Individual entities are tracked through the system as they proceed to various activities, between these activities they wait in queues (Robinson, 2004). Generally, a software package such as Witness (2012) is used and provides a visual representation of the entities' passage through the system.

DES models stochastic variability within the system (e.g. waiting times and arrival times) using probability distributions that are determined carefully to ensure they represent the behaviour of the real system – hence large amounts of data can sometimes be required. Due to the increased detail of DES models, they are generally used at operational level (e.g. a hospital department), to model a smaller population than SD models over a shorter timescale. DES models the passing of time by moving the simulation clock forward only when the next event occurs – which usually means that intervals between events are irregular. Robinson (2004), Pidd (2004) and Law and Kelton (2000) are all good background texts on DES.

DES models are used in a wide variety of applications. In healthcare for example, numerous DES models have been constructed for A & E departments (Ferrin et al., 2007; Fletcher et al., 2007) and outpatient clinics (Wijewickrama and Takakuwa, 2005; Guo et al., 2004), but also for operating rooms (Ferrin et al., 2004) and pharmacies (Wong et al., 2003). More generally, other popular applications of DES models are

supply chains (Windisch et al., 2015), maintenance processes (Alrabghi and Tiwari, 2016), manufacturing (Detty and Yingling, 2000) and logistics (Cheng and Duran, 2004). Brailsford and Hilton (2001) describe DES as 'arguably the most widely used OR technique in practice'.

2.2.2 Comparison of SD and DES

A number of previous works compare the SD/DES approaches. In attempting to answer the question of when each approach should be used, Brailsford and Hilton (2001) provide two detailed example models; Hilton's is a SD model for cardiac surgery and Brailsford's is a DES model for AIDS. These examples illustrate the types of system that are suitable for each modelling approach. In addition a general guide is provided for when to use each approach. The authors also argue that when deciding between which approach to use, the model's *purpose* is more important than the type of system being modelled. The authors highlight a hospital outpatient clinic as an example and claim the following. SD would be used if interest was in the interaction between neighbouring departments and if there were a large number of homogeneous patients – i.e. to model the system as a whole. DES would be used if the clinic had low interaction with other departments and there were a smaller number of heterogeneous patients – i.e. to model a part of the system in detail.

Chahal and Eldabi (2008) offer a similar view when observing the difference between 'detail complexity' and 'dynamic complexity' as follows. Detail complexity arises from the complex interactions in the system and the more distant view from SD simply

cannot capture the detail – hence DES is more appropriate. Dynamic complexity is captured well by SD but the detailed view provided from DES models "lack the global vision".

Morecroft and Robinson (2005) also consider the question of when each approach should be used, by developing a SD and DES model of the same system, a fishery, and comparing results. It was found that both models offered 'plausible explanations' for the behaviour of the system. The authors suggest that neither method has any overall advantage over the other and that developing both types of models for a specific system can be beneficial.

It is clear that although each approach has its strengths, each also has limitations. In addition to lacking the global vision, other limitations of DES are the need to run the model multiple times – often causing long run times, the high data requirements and the lack of ability to sufficiently represent the feedback loops present in a system (Viana et al., 2014). The main limitations of SD are that its low detailed approach cannot capture the 'detail complexity' as it cannot represent individual entities in the system. In addition, there is no way of representing the stochastic variation in the behaviour of these individuals.

It seems apparent that where one approach's limitations are exposed the other is able compensate. Chahal and Eldabi (2008) and Morecroft and Robinson (2005) go a step further and argue that the limitations of the approaches actually complement

each other. To overcome the limitations of using each approach separately, there is a clear need to use SD/DES hybrid models. Pidd (2012) observes that in practice, solving a problem often requires use of multiple methods and argues that combining other methods with DES 'should be the norm'. Brailsford et al. (2010) point out that the software to combine the two approaches is available, it's now the 'conceptual philosophy and practical methodology' that need to be developed. More generally, Kotiadis and Mingers (2006) explore the theoretical challenges associated with combining different modelling paradigms and argue that overcoming these challenges is achievable.

2.2.3 SD/DES hybrid models

The use of SD/DES hybrid simulation models has grown in recent years. Heath et al. (2011) provide an overview of the different simulation paradigms and also discuss cross-paradigm modelling and its challenges. A number of works have demonstrated the added benefits of constructing hybrid models such as Brailsford et al. (2010), Alvanchi et al. (2011) and Viana et al. (2014). Brailsford et al. (2014) dedicates a significant proportion of its material to the details of how both techniques can be combined and discusses the conceptual and practical challenges of doing so. Two chapters in particular, Pidd (2014) and Borschev (2014), explore the technical details of these practical challenges.

At the conceptual level, the question of which technique will be used to represent which part of the system needs to be considered carefully by the modeller. Chahal

and Eldabi (2008) define three formats for combining SD and DES as follows. *Hierarchical*; where two distinct models (SD at strategic level and DES at operational level) pass information between them. *Process environment*; which also consists of two distinct models passing information between them, but the distinction is not 'strategic and operational'. This time DES models the 'process' (a part of the system) and is contained within the SD model, which models the surrounding relationships and interactions. *Integrated*; where there is a single hybrid model; DES and SD are each used to capture elements of the system but there is no longer any separation between these parts of the model. For the integrated format, Brailsford et al. (2010) argues that this 'Holy Grail' has yet to be achieved.

Morgan et al. (2017) proposes a 'toolkit' of 6 designs for mixing SD and DES methods and a set of questions for modellers to ask when deciding which design to use. However, only the following 4 designs actually link SD and DES. *Sequential*, the use of one approach to identify the need for (and also to inform) the use of the other – e.g. Brailsford et al. (2004); *Enrichment*, an aspect of one approach is transferred into a model using the other approach – e.g. SD remains the core method and is enriched by the inclusion of discrete events; *Interaction*, the 2 models run independently but periodically stop and exchange data at fixed time steps – this includes both *hierarchical* and *process environment* formats above; *Integration*, as for the *integrated* format above.

Brailsford et al. (2010) present two healthcare case studies that each use the *pro-*

cess environment format and *interaction* design above in a hybrid approach. Both models use DES to represent a part of the system in detail and SD to model the surrounding interactions. The authors argue that it would have been extremely difficult to model either of the problems without a hybrid approach – and also that the *process environment* and *hierarchical* formats above (and hence the *interaction* design) are not true hybrid models since there are essentially two separate models passing data to each other. One of the case studies of Brailsford et al. (2010) is a Chlamydia infection model. The details of this are presented in much greater detail in (Viana et al., 2014).

The practical challenges faced in constructing a model that adequately represents the different paradigms must be overcome. In his book chapter, Pidd (2014) points out that linking SD and DES requires the modeller to take into account the differences of each approach in three areas: time handling, causes of variation in system behaviour and degrees of aggregation. He details the contrast in the approaches and we summarise his points below.

- *Time handling*: SD models use a 'time slicing' approach, where the simulation clock moves forward using equal time intervals of length dt. On the other hand, after each event in a DES model, the simulation clock skips ahead to the next event and the length of intervals between each event can vary. Pidd explains that a hybrid model must be able to incorporate these differences in time handling as follows: 'either both models must briefly cease operation in order to

exchange data, or the discrete-event model must schedule some regular, time-sliced events so that interaction can occur with the system dynamics model at those regular events'. Borshchev (2014) explains that the latter approach is adopted by the multi-method software AnyLogic (The AnyLogic Company, 2016).

- *Causes of variation in system behaviour*: The use of probability distributions in DES models means that each simulation run is effectively a sampling experiment – hence in order to form meaningful conclusions, multiple runs of a DES model are required. Some researchers have devised automated procedures for making decisions about the required number of model runs, see for example Hoad et al. (2011). On the other hand, as Pidd points out, since SD models are based on the assumption that 'system structure leads to system behaviour', in general they only require a single run. Pidd states that the modeller must be careful when interpreting the results of a hybrid model since 'the two different types of variation can lead to a factorial explosion when attempting to understand the results'.

- *Degrees of aggregation*: DES models are 'atomistic', since 'system behaviour is a result of interactions between individual entities and the resources they use'. In a SD model, he describes the variables are 'quasi-continuous', since the stocks and flows only change at each time slice dt and are held constant in between these. As SD models do not capture the behaviour of individuals, and instead concentrate on how their aggregated variables change over time,

> he describes SD models as 'quasi-continuous aggregations', in which 'variation in behaviour occurs as a result of relationships between the variables included'. Pidd highlights the emphasis that the hybrid model should place on discrete entities or continuous variables as an issue that must be resolved.

For the issue of *degrees of aggregation*, Pidd (2014) states that it is 'relatively straightforward' to use the value of a continuous variable to trigger a discrete event. This is one of a number of possibilities for linking parts of SD and DES models. In his book chapter, Borshchev (2014) thoroughly discusses the technical aspects of linking SD and DES (and also agent-based) models and provides numerous examples of models with links from SD to DES, from DES to SD, and links both ways.

One approach for coding a hybrid model is to code the SD and DES models in separate environments and then use a method for linking the two models. This is the approach adopted in Brailsford et al. (2010) and Viana et al. (2014). In constructing each hybrid model, the SD and DES models are built separately in a dedicated software package – Vensim (2010) for SD and Simul8 (SIMUL8 Corporation, 2015) for DES – and the link between them is automated using VBA. Zulkepli et al. (2012) use the same packages for their SD and DES models in a hybrid approach applied to healthcare, though the link between the models is not automated. Abduaziz et al. (2014) also produce a hybrid model using dedicated packages – iThink (ISEE systems, 2012) for SD and Arena (Rockwell Automation, 2016) for DES – applied to the automotive industry. These are examples of the *interaction* design above.

Another approach is to work entirely within a DES environment such as Witness (2012). Ziegler et al. (2000) show that elements of the SD modelling approach can be incorporated within a DES framework. Ziegler (2006) points out the behavioural features that need to be supported in a DES framework when incorporating both SD and DES parts and the interfaces between them. Heath et al. (2011) and Viana et al.(2014) point out that many DES software packages can represent continuous variables and can 'therefore be adapted to provide the underlying structures of SD models'. An advantage of using DES software over linking separate SD and DES environments is that there is no need to devise a method that automates the linking of the models, such as the additional VBA code required in Viana et al. (2014). As the models are likely to exchange information at regular intervals, additional code to link separate modelling environments would be required to run regularly. Hence instead including all code within the DES environment, and not requiring any additional code to link the models, may also improve the model's run times.

Another option for coding the hybrid model is to use dedicated multi-method software such as AnyLogic. AnyLogic can produce DES and SD models (and also agent-based models) in one environment with nearly all the main features of individual SD and DES packages (Heath et al., 2011). Pruckner and German (2013) use AnyLogic to produce a hybrid model for electricity generation systems. Mazaeda et al. (2012) use EcosimPro (EA Internacional, 2016) to produce a hybrid model for sugar manufacturing.

According to the arguments in Borshchev (2014), in which the flexibility of the AnyLogic software is demonstrated, use of such multi-discipline software for our hybrid model may be easier to implement than in a dedicated DES environment. However, we propose to construct the hybrid model in the DES package Witness. We have no requirement for agent-based methods and in addition, there appear to be few examples in the literature of SD/DES hybrid models built in a dedicated DES environment, so it is hoped that the approach used in this chapter may serve as a guide for future researchers.

It is worth highlighting that Heath et al. (2011) point out that when using multi-discipline software such as AnyLogic or DES software such as Witness, these packages 'remain essentially either a DES environment with some continuous features, or a SD environment with some discrete or stochastic features.' They proceed to argue that there is currently no 'genuinely hybrid modelling methodology that combines the characteristic features of both DES and SD'.

Wynn et al. (2012) use DES to model information flows within BT with feedback loops provided by SD methods in a hybrid approach, using Cambridge Advanced Modeller (Wynn et al., 2010) software. The feedback examines the current system performance relative to targets and adjusts workforce levels according to the size of the backlog of jobs. This feedback is similar in nature to the feedback provided by the SD part of our hybrid model, a simplified version of the hydraulics model, introduced

in section 2.3.

2.3 BT Application

In this section we present the BT hydraulics model. We explain how it is used within the organisation and also outline a key limitation of the model. We present a simplified version of this SD model. This version will be used to construct the SD/DES hybrid model in section 2.4.

Access to the UK's growing telecommunications network is essential for organisations, businesses and public services. The network has a complex structure, so faults can occur frequently and for many reasons – e.g. due to weather conditions. When a fault occurs it is important that repair work is performed as soon as possible. BT (through Openreach), is solely responsible for repair and maintenance of the vast majority of the UK's network. To perform this work effectively they must ensure that at any time they have sufficient workforce numbers available to meet demands – which can vary considerably. The SD model that assists with this is known as the hydraulics model. It is important that BT understands the impact of changes in demand on the system. The hydraulics model helps them to be sufficiently aware of their own work flows to ensure that when an increase in demand occurs, measures that are required to complete the additional jobs fast enough are put into place to ensure customer satisfaction.

The rationale behind the hydraulics model and their choice of the SD approach is

described in detail in the book chapter Jensen et al. (2013); written by the BT scientists who devised and developed the model. The SD model that we use in this chapter is a simplified version of the full hydraulics model. As we see in chapter 5 where the full hydraulics model is introduced, the full model contains a number of similar additional feedback mechanisms compared to the simplified model. These feedback mechanisms control other aspects of the system such as the engineers' overtime and 'shrinkage' (time on training or leave), again attempting to replicate the decisions of management. Hence, although the SD model used in this chapter is a simplified version of the hydraulics model, it nonetheless represents the key aspects of the full model's behaviour and enables us to learn more about the full model. We also note that the simplified model represents repair jobs only and not other job types such as 'provision' (installation) jobs. A stock and flow diagram of this simplified model is shown in Figure 2.3.1.

From Figure 2.3.1, we see that the simplified model is effectively a single feedback mechanism that adjusts workforce numbers in response to changes in demand. The nature of this feedback mechanism has been carefully modelled to replicate the management decisions if such a situation were to be observed in the real system. For example when an increase in demand is observed in the model, the increase of the engineer numbers, and the speed of this increase, is designed to match that of how management decisions would determine such increases in the real system.

There are three stocks in the model; the 'backlog' (job queue), 'capacity deployed'

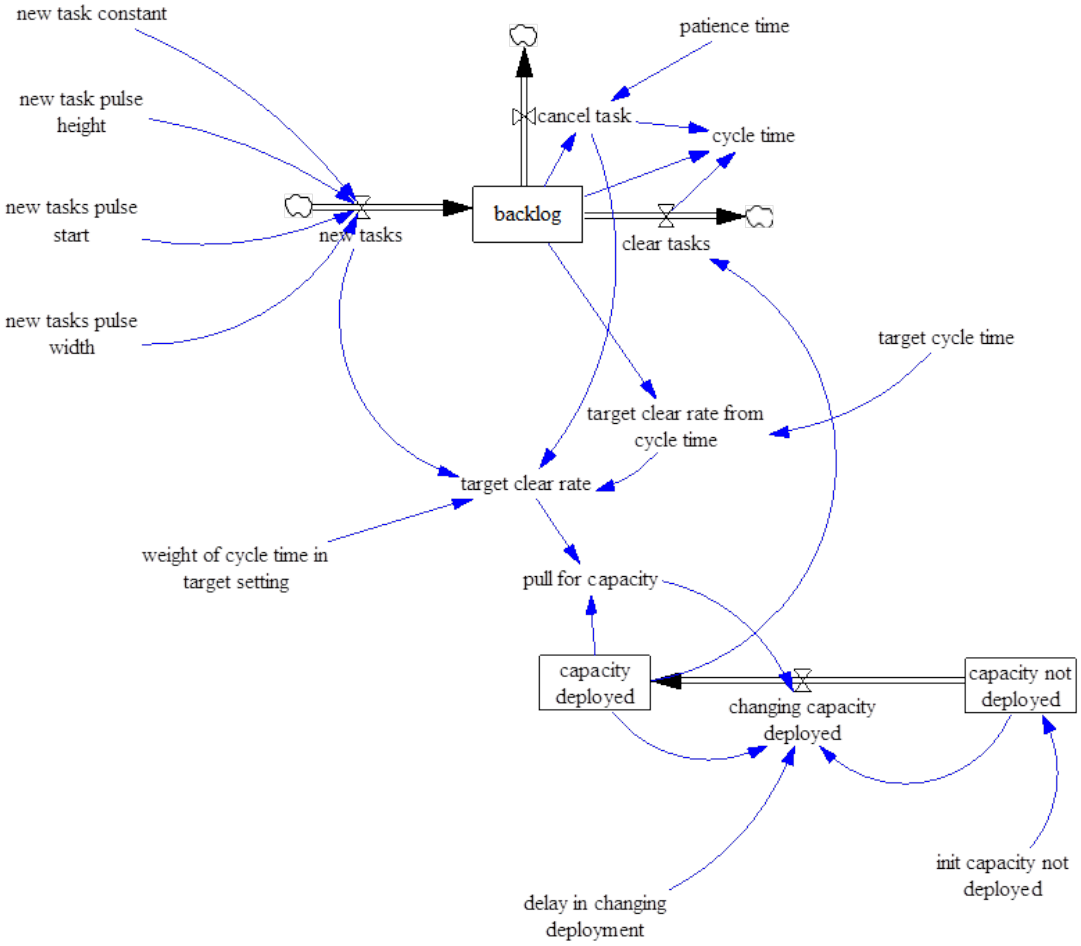

Figure 2.3.1: Stock and flow diagram of the simplified BT hydraulics model.

and 'capacity not deployed'. The model consists of two separate flows of entities; jobs and engineers. In the upper part of the model, jobs enter the system to the backlog and then are either cancelled or cleared (completed). In the lower part, maintenance engineers flow between the two stocks 'capacity deployed' and 'capacity not deployed'. The 'capacity deployed' stock represents engineers that are available to work on repair jobs. The 'capacity not deployed' stock represents engineers that are not currently available (e.g. on leave or training etc.), but are present in the system. Hence, engineers in 'capacity not deployed' can be transferred across to 'capacity deployed' when

the demand for engineers becomes great enough. This is determined by the feedback mechanism within the model.

The nature of this feedback is determined by the three model parameters; the 'target cycle time', 'weight of cycle time in target setting' and 'delay in changing deployment'. For full details of these see chapter 4. This demand for engineers essentially involves comparing the queuing time of jobs in the backlog to the 'target cycle time' (TCT) – which is the time that BT wish to complete each job within. If queuing times are large relative to the TCT, the system pulls engineers towards the capacity deployed stock, increasing the engineer numbers to deal with the extra jobs in the queue. If queuing times small relative to the TCT, the system pushes engineers back towards capacity not deployed. There is a delay in this process represented in the model by the 'delay in changing deployment' model parameter. In reality, the process of transferring engineers, reassigning work and arranging all the required meetings etc. all takes time.

In Figure 2.4.1, the variable 'cycle time' is the time that each job spends in the system from first being reported, to being completed by the engineers. That is,

$$\text{cycle time} = \text{queue time} + \text{engineer completion time}, \tag{2.3.1}$$

where the engineer completion time is the service time. A key performance measure that is of particular interest to BT is referred to as 'right first time' (RFT). This is the percentage of repair jobs that are completed correctly at the first attempt within 2 days, hence RFT is calculated using job cycle times. Accurate forecasts of

RFT are crucial for an organisation such as BT, where customer satisfaction is crucial.

The limitations of the SD approach described in section 2.2.2 are apparent in the hydraulics model. Ideally, to model RFT, individual jobs would be tracked through the system and their cycle times measured. It would be simple to identify the percentage of jobs that fail to meet the 2 day target. However, the low-detailed approach of SD means that the hydraulics model can't track individual jobs through the system and so cannot model RFT directly.

The current approach for modelling RFT is to use a regression relationship between RFT and average cycle time. This has been derived from historical data. Weekly estimates of the average cycle time are output from the model. These are calculated using Little's law (Little, 1961); see chapter 3 for more on this. For each week, the regression relationship is applied to the average cycle time to obtain an estimate of the RFT. In section 2.5 we demonstrate that relying on the regression relationship has the potential for misleading results. We also show the advantages of using a hybrid approach that can model RFT directly since individual jobs are tracked through the system.

2.4 Methodology

This section presents the details of our approach for constructing the hybrid model. Section 2.4.1 presents an overview of the approach, while sections 2.4.2 – 2.4.4 describe

the details of each of the three stages in the process. Section 2.4.5 describes some of the challenges faced when constructing hybrid models in a DES environment.

2.4.1 Overview of approach

As we explained in section 2.3, the simplified hydraulics model contains feedback loops that replicate the decisions of management to transfer engineers to where they're required – i.e. to either 'deployed' or 'not deployed'. We propose to build a hybrid model that retains this feedback. In order to represent individual jobs and engineers in the system, we propose to model a part of the system in detail using DES. In this part of the model, individual jobs will enter the system, wait in the queue and proceed to one of the engineers before leaving the system. The tracking of individual jobs will enable the cycle time of each job to be measured and hence the RFT to be modelled directly. The SD and DES models will then be linked to enable the passing of information at regular intervals dt.

As we saw in section 2.2.3, using DES to model a part of the system in detail and SD to model the surrounding interactions, with the models passing information regularly, means that we are using the *process environment* format of Chahal and Eldabi (2008) and the *interaction* design of Morgan et al. (2017).

2.4.2 Stage 1: System dynamics model in Witness

The first stage of the hybrid modelling process is to build the SD model entirely within the Witness environment. To ensure that the model is operating correctly, the output of this model is carefully validated against the output from the same model produced in the SD software Vensim, before proceeding to stage 2.

Since SD variables are 'quasi-continuous', only changing at each time slice *dt* and remaining constant in between these (Pidd, 2014), the SD model equations are effectively difference equations that are updated at each *dt* interval. As we explained in section 2.2.3, Pidd (2014) suggests that one way to handle simulated time in a hybrid approach is for the DES model to schedule regular events at each *dt* interval, so that interaction can occur with the SD model. One way to achieve this in Witness is to set up 'pseudo' entities that arrive at a 'pseudo' activity at the start of each *dt* interval. The activity contains a block of code that is activated upon the arrival of each entity. This block of code contains the difference equations of the SD model – and hence these difference equations update at the start of each *dt* interval.

It is worth highlighting that when the SD part of the model updates in this way, i.e. when the SD model's difference equations update via this 'pseudo' updating process, this does not progress the simulation clock in the DES environment. Hence we can update the SD model at the start of each *dt* interval without advancing the simulation clock. In later stages when we add the DES parts of the model, the DES model

can update throughout the *dt* interval, with no concerns about overlapping updates.

When coding the difference equations of the SD model in a DES environment, it is important to determine the order of which the SD variables should update. Even for the simple BT model in Figure 2.3.1, it is not immediately apparent which variables should update first. In a SD package such as Vensim, at each *dt* interval, the stocks are updated first, followed by the flows and other variables. Essentially, the greater the separation between a variable and its 'furthest away' stock (in terms of how many variables lie in between) that influences its update, the later in the order that variable should be updated. For example the 'changing capacity deployed' variable in Figure 2.3.1 requires the update of the variable 'pull for capacity', which itself requires the update of 'target clear rate', which requires the update of 'target clear rate from cycle time' and the 'backlog' stock. In this way it is clear that the variable 'changing capacity deployed' must be updated last at each *dt* after all other variables have been updated. When building a SD model in dedicated packages such as Vensim, this issue is automatically taken into account.

2.4.3 Stage 2: Intermediate model – SD to DES one way

In stage 2, the DES model is added to the existing SD model in Witness from stage 1. However, at stage 2 we restrict ourselves to constructing an 'intermediate' model that only passes information one way, from SD to DES. This allows us to check that the DES part of the model is operating correctly when receiving information from the SD model.

The part of the system that is modelled in detail using DES is shown in the stock and flow diagram of Figure 2.4.1. This part of the SD model represents the 'flow' of jobs as they enter the system through the 'new tasks' rate, enter the backlog stock, then leave the system either through the cancel or clear rate – a proportion of all jobs are cancelled according to the parameters of the SD model. DES will represent jobs, the backlog and engineers as follows. Each job will be represented as an individual entity that passes through the system. The backlog will be represented as a queue where jobs must wait until there is an available engineer to work on them. Each engineer will be represented by an activity that jobs arrive at and then leave after a certain interval – which represents the job completion time.

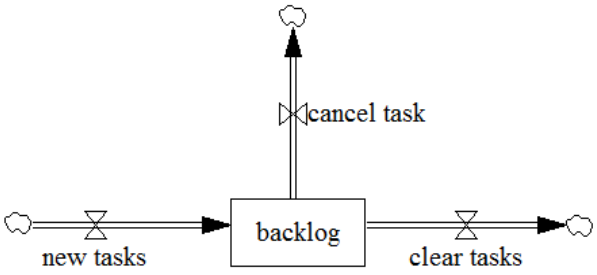

Figure 2.4.1: The focus of the DES part of the model.

In this way each job enters the system according to a specified inter–arrival time. The job then waits in the queue, either for as long as is necessary for an engineer activity to be available, or until it is cancelled. If the job is not cancelled it proceeds to an available engineer. After spending a certain amount of time at the engineer activity, the job then leaves the system. The tracking of individual jobs in this way enables

the cycle time of each job to be measured and hence the RFT to be modelled directly.

The SD and DES parts are linked in two ways. The first link involves the modelling of job arrivals. The SD model defines the rate 'new tasks' over the interval dt as the average number of job arrivals per dt interval. Denote this rate by λ. For each dt interval, we code the DES model to use the value of λ to set the inter–arrival times for the job entities as they enter the system. The rate λ can be converted to an inter–arrival time of $1/\lambda$ within each dt interval. This will result in jobs arriving at regular intervals. One approach for introducing stochastic variability in the inter–arrival times Y in a DES model is as follows: set $Y \sim \text{Exp}(1/\lambda)$ (Robinson, 2004). Hence on average there are λ job arrivals per dt interval with irregular inter–arrival times. Examples of the conversion of rates from the SD part of the model, to become the parameters for distributions of inter–arrival times in a DES part of the model, can be found in Viana et al. (2014) and Borshchev (2014).

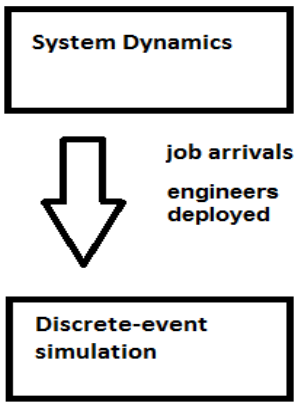

Figure 2.4.2: Stage 2 model structure.

The second link between the SD and DES parts of the model involves the number of available engineers. The SD model defines 'capacity deployed' as the number of available engineers over each *dt* interval. As the DES model represents each engineer as an activity, the DES model must represent all available engineers in the model as an activity – both those that are available ('capacity deployed' in the SD model) and those that are not available ('capacity not deployed') in the SD model. Hence to ensure that the correct number of available engineers are working on jobs throughout each *dt* interval, the DES model 'switches off' certain engineers that are not available. At the start of each *dt* interval, the SD model passes the value for 'capacity deployed' over to the DES model. The DES model uses this value to determine how many engineers should be 'switched off' for that *dt* interval. Any engineers that are not 'switched off' are available to complete jobs. A diagram of the one-way sharing of information between the two parts of the model is shown in Figure 2.4.2.

2.4.4 Stage 3: Full hybrid simulation model

In stage 3 we link the DES model back to the SD model to form the hybrid model. The hybrid includes links both ways; from SD to DES and from DES to SD. The link from DES back to SD is achieved as follows. At the end of each interval *dt*, the DES model has fully updated. The current length of the job queue in the DES model is fed back into the SD model by using this value to represent the 'backlog' stock in the SD model's difference equations. Hence, the SD model updates using the length of the job queue that was determined by the previous update of the DES part of the model. We use the number of entities in the DES queue to update a stock in the SD

model. Borshchev (2014) includes an example of using the number of entities in a DES queue to switch a SD rate on or off. Communication between the two models is now operating in both directions, as we see in Figure 2.4.3.

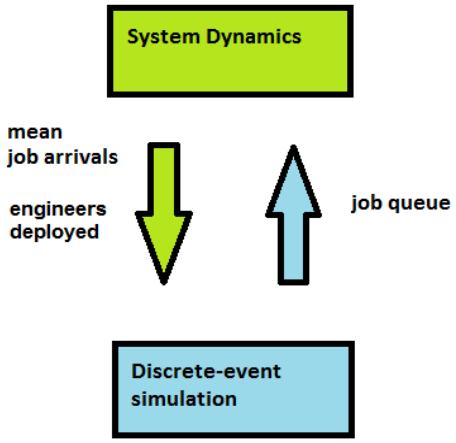

Figure 2.4.3: Stage 3 model structure.

The updating process of the stage 3 model is as follows. At the start of each new *dt* interval, the SD model's difference equations update, using the current value of the job queue from the DES model – that was updated over the previous *dt* interval – to represent the 'backlog' stock. Then, throughout the duration of the new *dt* interval, the DES model updates using the SD variables 'new tasks' and 'capacity deployed', to set the inter–arrival times of jobs and available engineer numbers respectively. At the start of the next *dt* interval, the SD model's difference equations update again, and so on.

A hybrid model such as this has the advantages of both the feedback provided by the SD model so that it can replicate the decisions of management to adjust workforce

numbers, and also the detail provided from the DES model to represent individual jobs and engineers, enabling us to model RFT directly.

2.4.5 Challenges of hybrid modelling in DES software

In this section we describe some of the challenges that are faced when constructing hybrid models in dedicated DES software such as Witness. These challenges involve the ordering of updating of continuous variables and discrete elements within the hybrid model. These were revealed when the links between the SD and DES parts of the model were included.

The order of updating *within* each part of the hybrid model, i.e. within each of the SD and DES parts, can be important. Although the two models update separately, for certain aspects of BT's system, we require each part of the two models to update in the same order. For example in the SD model, the stocks are updated at the start of the SD updating process and as part of this, cancelled jobs (a percentage of the jobs in the backlog) are removed from the backlog stock – hence this percentage of cancelled jobs are removed at the start of the SD model update. Therefore the DES model must also remove cancelled jobs from the backlog queue at the start of its updating interval – to ensure that an appropriate number of jobs are cancelled. Although this requirement was for BT's system only, aspects of other systems may require such considerations.

The order of updating *across* each part of the hybrid model is also important. We do

not want the updates of each part of the model to overlap. For the updating process of the hybrid, we want the SD part to update at each dt interval – updating the SD difference equations requires no passage of time in the DES environment – and then the DES part to update within each dt interval. A timeline of this process is shown in Figure 2.4.4, where $dt = 1$.

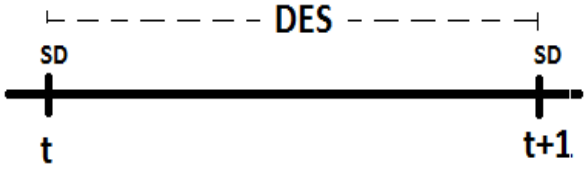

Figure 2.4.4: Desired timeline of hybrid model updating process for each dt interval.

A problem can arise when we have a DES event scheduled at the same time as the SD model update – at the end of a dt interval, e.g. at time t or $t + 1$ in Figure 2.4.4. When this occurs, DES software such as Witness examines the priority of activities to determine which event to perform next. If these priorities are equal then it examines the order in which the events were created. However, on some occasions the SD model updates before the DES part has finished updating. Even after making adjustments to the priorities of events, this issue was still apparent for certain updating intervals. On these occasions, the SD part updated prior to the final job arrival or job completion in the DES model.

2.5 Results

In this section we use the simplified version of the BT hydraulics model within the SD/DES hybrid approach outlined in section 2.4, to demonstrate the potential benefits of hybrid modelling in the context of the hydraulics model. We investigate the behaviour of the hybrid model under a simple artificial scenario – a temporary stepped increase in the average job arrivals per day. The effects of stochastic variation on system performance are investigated by adjusting standard deviations for engineer job completion times (service times), and examining the resulting cycle times and RFT. In addition, we demonstrate that when not modelling RFT directly, relying on a regression relationship to model RFT can be misleading.

The model runs for 500 days in total. On all days except days 70-79, average job arrivals are set at 900 per day. On days 70-79, this increases to 1200. Days 70-79 were chosen for the stepped increase, since this gives the system sufficient time to return to steady state, both prior to the stepped increase, and after the stepped increase before the simulation terminates.

A key decision is the length of the updating interval dt for the SD part of the model. Setting dt too large means the model's response to changes in the system is too slow – i.e. the feedback in the SD model is not able to respond fast enough to the increases in job arrivals. Setting dt too small increases computational time unnecessarily. A range of values were investigated ranging from $dt = 1$ day down to $dt = 1/16$ day.

By visual inspection of the time series and comparing model run times, $dt = 1/4$ day was found to be a suitable compromise.

We introduce stochastic variability into the hybrid model as follows. Inter-arrival times Y_i of jobs entering the system are modelled as $Y_i \sim \text{Exp}(1/\lambda)$, with $\lambda = 225$. Hence on average, for all days except days 70-79, there are 225 job arrivals per dt interval (1/4 day) and so 900 arrivals each day. The Lognormal distribution is often selected to model service times due to the right-skew being a good representation of the actual data (Robinson, 2004). After consultation with BT, the completion times X_i of jobs by the engineers are modelled as $X_i \sim$ Lognormal (μ, σ), where $\mathbb{E}[X] = \frac{1}{3}$ days, since on average each engineer completes 3 jobs per day and hence $\frac{3}{4}$ jobs per dt interval. To investigate the effects of stochastic variability we use the following standard deviation values: s.d.$[X] = 0.1$, 0.2, 0.3 and 0.5. In total 400 engineers are present in the model, with only 300 of these initially deployed.

As we highlighted in section 2.2.3, Pidd (2014) points out that multiple runs of a DES model are required and also that the modeller must be careful when interpreting the results of a hybrid model. As such, for each value of s.d.$[X]$, 25 model runs are performed in Witness, each with different pseudo random number streams. For each variable in the model, an average is taken across all these runs at each updating interval dt. Note that for the settings of this experiment, and after consultation with BT, the threshold of RFT was increased to around 2.5 days (rather than 2 days) in order to obtain realistic RFT values.

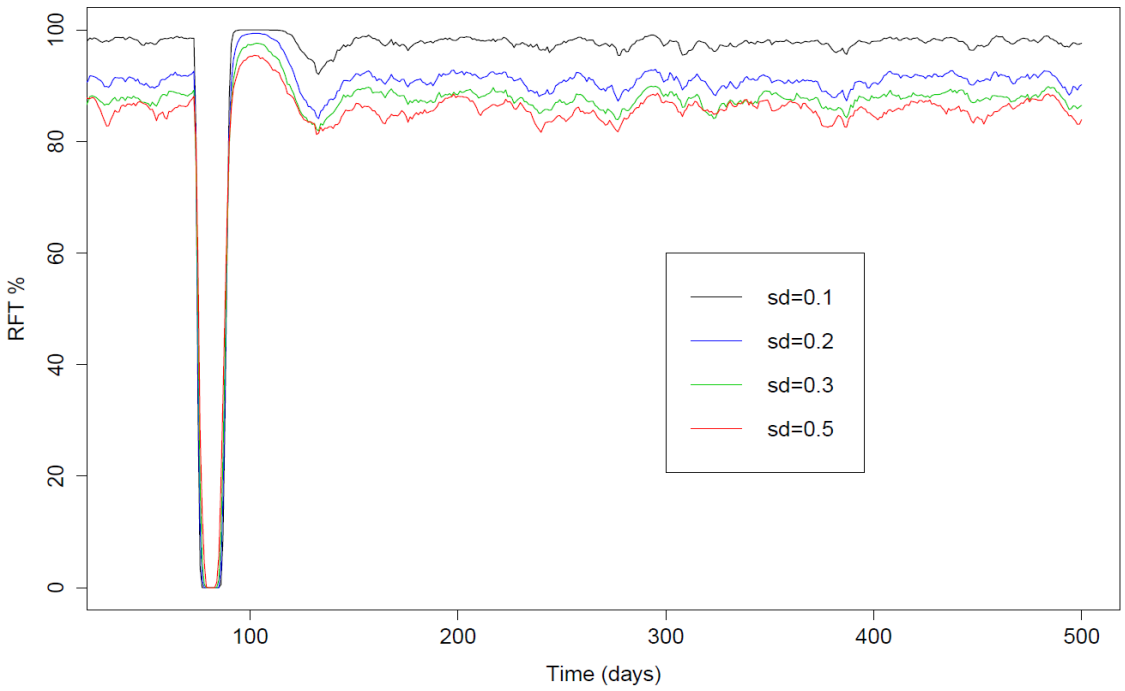

Figure 2.5.1: Effects of stochastic variation on RFT.

Recall that cycle time = queue time + engineer completion time. Figure 2.5.1 shows the averages across the 25 runs of the RFT values at each interval dt, for each of the 4 values of s.d.$[X]$. The sudden decrease in the RFT around day 70 is caused by the stepped increase in job arrivals. However it is the steady state sections of the RFT time series of Figure 2.5.1 that are most interesting. These show that increasing s.d.$[X]$, decreases the RFT when the system is in steady state. In other words, increasing the variability of engineer completion times has reduced system performance, as a greater number of jobs 'fail' to meet the completion target. In fact, this drop in performance is greater than 10%, when comparing RFT's resulting from s.d.$[X]$ = 0.1 and 0.5. The effect of this variability does has limitations. Use of larger values

of s.d.$[X]$ were also investigated and increases beyond s.d.$[X] \approx 0.7$ resulted in little change in RFT.

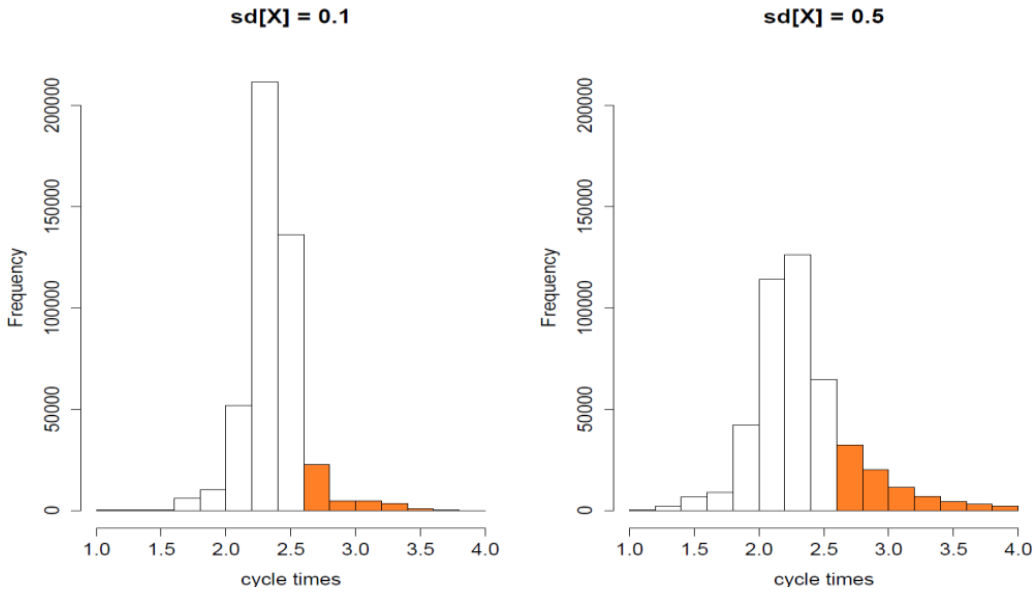

Figure 2.5.2: Effect of stochastic variation on cycle time distributions.

Figure 2.5.2 presents histograms of the distributions of cycle times for the highest and lowest values of s.d.$[X]$; 0.1 and 0.5. We can see that for s.d.$[X] = 0.5$, the distribution has heavier tails. In particular, it is the heavier right tail that causes a greater proportion of jobs to 'fail' to meet the completion target, hence the lower corresponding RFT value. The jobs that fail to meet the target are coloured in orange.

For single–server systems, the Pollaczek-Khintchine formula (Pollaczek, 1930; Khintchine, 1932), shows that the mean queueing time (and hence the mean cycle time) is influenced by the standard deviation of the service time (increasing standard deviation increases the mean queueing time); see for example Harrison and Patel (1992).

For multi–server systems, empirical results, see for example Hillier and Yu (1981), show that this also holds, but with the scale of this influence gradually decreasing as the number of servers increases. This decrease continues until there is no influence when we have infinite servers. Hence, the results of this section are not unique to our selection of the lognormal distribution for modelling service times. Similar results, i.e. the effects of increasing s.d.$[X]$, would have been observed for other reasonable choices of service time distributions.

As we explained in section 2.3, the hydraulics model cannot estimate the RFT directly. Currently at BT, the average cycle time from the SD model is used to estimate RFT, via a linear regression relationship between the two variables, derived from historical data. As we have seen, adjusting the values of s.d.$[X]$ influences the tails of the distribution of cycle times and hence the RFT. Determining the influence of s.d.$[X]$ on the daily average cycle time for our system is less straightforward.

Our hybrid model does not contain an infinite number of servers (engineers), but does contain a large number; between 300 and 400 depending on how many are active. This means that for our system, according to the Pollaczek-Khintchine formula and empirical results above, s.d.$[X]$ should have an (albeit reduced) influence on the daily average cycle time. Figure 2.5.3 shows the time series of the average daily cycle times from the hybrid, averaged over the 25 runs. For the range of s.d.$[X]$ values used here, it is not clear whether s.d.$[X]$ is influencing the average cycle time in our system.

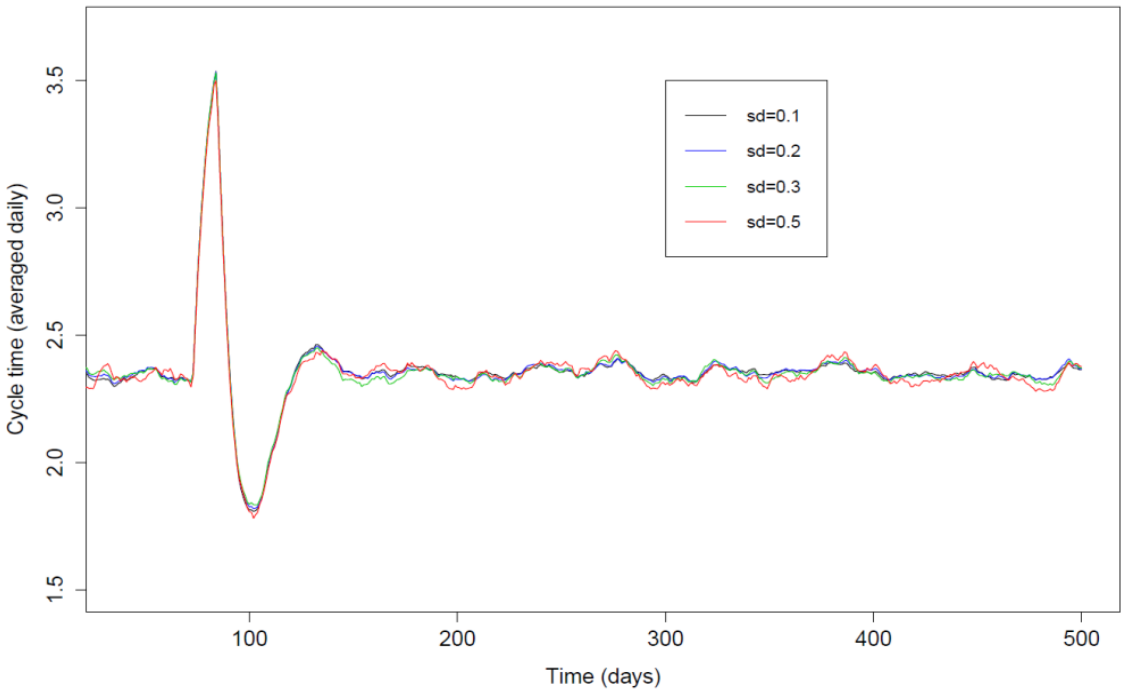

Figure 2.5.3: Effects of stochastic variation on daily average cycle time.

What is clear is that the influence of s.d.$[X]$ on RFT is not the same as the influence of s.d.$[X]$ on the average cycle time. In other words, if we increase s.d.$[X]$ to an extent that results in an average cycle time increase of 1%, the tail probabilities of the cycle time distribution will increase by more than 1%. Since the regression approach used at BT assumes a fixed relationship between RFT and average cycle time, this also assumes one of the following: either that s.d.$[X]$ is constant over time (which is not true according to system experts at BT) or that changes in s.d.$[X]$ have an equal effect on both average cycle time and RFT. Since our results show that the latter of these is not the case, our results suggest that relying solely on the regression relationship to model RFT may not always be reliable.

In order to investigate this, we compare RFT values output directly from the hybrid model with RFT estimates obtained using regression. Our hybrid model is based on the simplified version of the hydraulics model, hence it is a smaller-scale version of the full system. We therefore must use the same regression model structure as BT's, but estimate the parameters separately for our scaled–down system. If ct_i is the average cycle time of day i, then rft_i, the RFT for day i, is estimated at BT as in (2.5.1).

$$rft_i \quad = \quad \alpha + \beta\, ct_i + \epsilon_i \tag{2.5.1}$$

BT assume that this regression relationship is fixed – regardless of the variability in cycle times. We investigate the use of two values of s.d.$[X]$ (0.1 and 0.5) in the hybrid model and run the model for $i = 1, \ldots, 500$ days. Denote the resulting time series for these 500 days as CT and RFT, for ct_i and rft_i respectively. We then fit a regression model to the hybrid's simulated data for series RFT and CT, for each value of s.d.$[X]$. These models are denoted by $M_{0.1}$ and $M_{0.5}$ respectively. The estimated parameters of each of these models, $\hat{\alpha}$ and $\hat{\beta}$, can be considered to be the 'true' regression parameters for this scaled–down system.

In order to assess the use of the regression relationship in estimating RFT, we imagine that like BT, we have no RFT data directly from a hybrid model and must rely entirely on a regression approach based on historical data using a single set of fixed parameters. We apply the regression relationships of each of the models $M_{0.1}$ and

$M_{0.5}$, to the corresponding CT data using s.d.$[X] = 0.1$ or 0.5, to obtain two sets of RFT estimates, denoted by $\hat{RFT}_{0.1}$ and $\hat{RFT}_{0.5}$ respectively. Since BT use a fixed regression relationship, we can demonstrate the approach's potential for misleading results as follows. If we take the CT data for the s.d.$[X] = 0.5$ runs from the hybrid, but rather than using the regression relationship of model $M_{0.5}$ to estimate the RFT values, we instead use regression model $M_{0.1}$, Figure 2.5.4 shows the resulting RFT estimates.

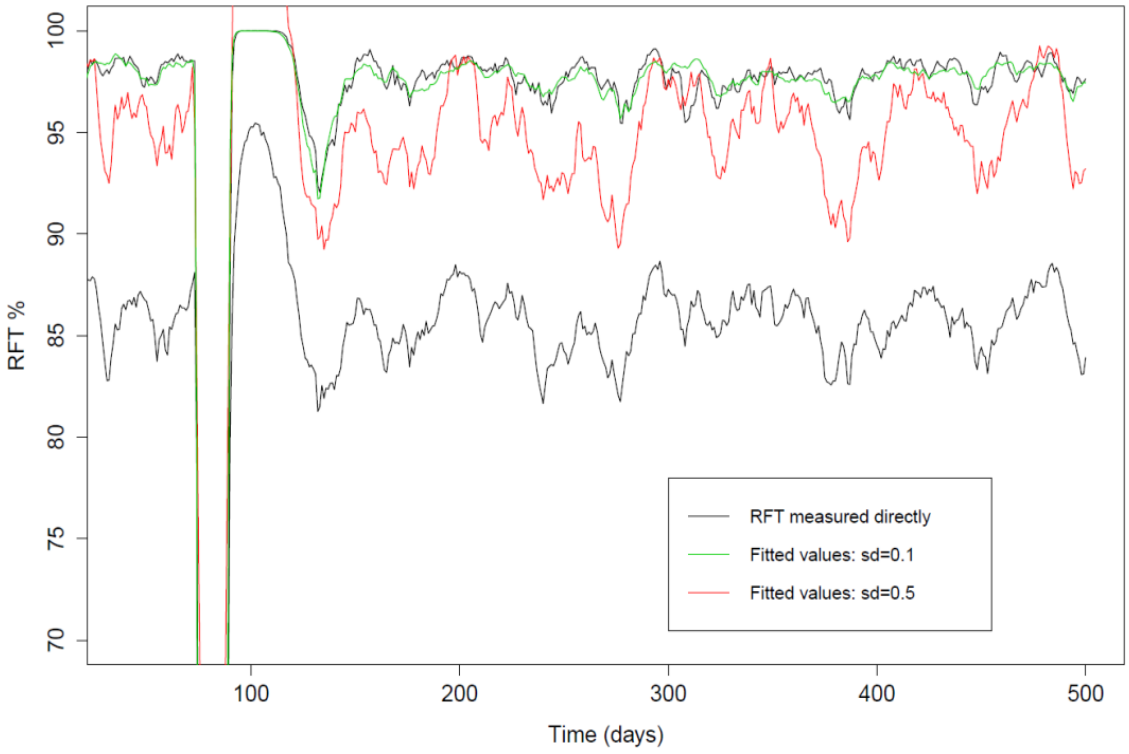

Figure 2.5.4: Misleading RFT estimates from the regression approach.

The black lines of Figure 2.5.4 are the directly measured RFT series from the hybrid model, for runs with s.d.$[X] = 0.1$ (upper line) and s.d.$[X] = 0.5$ (lower line). The green line is $\hat{RFT}_{0.1}$, the estimated RFT series resulting from the correct regression

model $M_{0.1}$, whose parameters were estimated using CT data with s.d.$[X] = 0.1$. As one would expect, $\hat{RFT}_{0.1}$ values lie close to the directly measured RFT values from the hybrid. The red line is not $\hat{RFT}_{0.5}$. Instead the red line is the estimated RFT series using the (incorrect) regression model $M_{0.1}$ rather than $M_{0.5}$, using the CT data with s.d.$[X] = 0.5$. These values lie far away from the corresponding directly measured RFT series resulting from s.d.$[X] = 0.5$ (lower black line) and instead lie closer to the RFT series resulting from s.d.$[X] = 0.1$. So by assuming a fixed regression relationship, and hence assuming that changes in s.d.$[X]$ have an equal effect on both average cycle time and RFT, two sets of CT data with different s.d.$[X]$ values have resulted incorrectly in similar RFT estimates. This example demonstrates that if s.d.$[X]$ changes over time, relying on a fixed regression relationship can lead to misleading results.

2.6 Discussion

In this section we present a discussion. We begin by discussing the benefits of a hybrid model to BT. We then reflect on our approach for coding the model within a dedicated DES environment. We conclude with a discussion of the contribution of this chapter to the literature.

The results from the hybrid model using the simple artificial scenario for job arrivals have demonstrated some of the benefits of a hybrid approach to BT. Results have revealed the effect of the standard deviation of service times on cycle time dis-

tributions and RFT – increasing the standard deviation results in decreased system performance in steady state. As we explain in section 2.5, this effect is well known in queuing theory. However, the hybrid model has enabled us to understand the size of this effect, by allowing us to directly model the important performance measure RFT. BT previously had no such insights into the behaviour of RFT. In addition, the hybrid model has enabled us to demonstrate the potential for misleading results when relying solely on the regression relationship to estimate RFT.

The work of this chapter demonstrates that a simple SD/DES hybrid model can be constructed in dedicated DES software such as Witness, although a requirement is that this software must offer a programming language facility in order to incorporate the SD model's difference equations. However, as section 2.4.5 reveals, hybrid modelling in a dedicated DES environment poses a number of challenges.

Perhaps the most important of these is ensuring the correct order of updating across each part of the hybrid model. As we explained in section 2.4.5, an issue can occur when DES events and SD model updates are scheduled simultaneously– at the end of each *dt* interval. For some *dt* intervals, in the DES model, one fewer job arrives or is completed than is required, before the SD model updates again. Although not significant enough to impact the results of section 2.5, this is problematic and for more complex models may become a more serious issue. Ultimately, incorporating SD difference equations within a DES environment has limited the control we have over the order of events that are due to occur simultaneously.

It is also worth highlighting that overcoming some of the challenges faced in using DES software, was not straightforward and sometimes required a fairly lengthy investigation. These challenges were effectively caused by the software assuming (naturally) that the modeller is using a DES approach. Two alternative approaches for constructing hybrid models were discussed in section 2.2.3. These were the use of separate SD and DES software with some method for linking the models, and the use of dedicated multi-discipline software such as AnyLogic.

If using separate SD and DES software, although separate checks of the SD and DES models would still be required, it could be taken for granted that each model is operating as a SD or DES model should, within its dedicated software. Hence there would be no issues regarding the ordering of updating of continuous variables and discrete elements within the hybrid model, such as those reported for a DES environment in section 2.4.5. Hence no additional tests would be required in order to check that each part of the model is updating appropriately. The modeller would have confidence that at the end of each *dt* interval, the DES part updates fully prior to the SD part updating. If automating the link between the two models, as in Viana et al., (2014), although not straightforward, this process is potentially less time–consuming than overcoming the challenges faced when using a DES environment.

We can make similar observations regarding the use of multi-discipline software such as AnyLogic. This environment is dedicated to the use of combining multiple meth-

ods such as SD and DES. Borshchev (2014) demonstrates with detailed examples the highly specialised solutions that are offered in the AnyLogic software for linking the two methods. Therefore, overall we can observe that it is likely that both of these alternative approaches would be easier to apply in a hybrid modelling approach and would require a less lengthy testing process.

The hybrid model has advantages over both standalone SD and DES models, for the BT application, but also more generally, by combining key attributes of each approach. Unlike SD models, the hybrid can track individual entities in the system and so can model individual service times and hence key performance measures. In addition, the effects of stochastic variation on the behaviour of individual entities can be thoroughly explored. Unlike DES models, the hybrid incorporates a complex feedback structure from an SD model that has been validated in a major organisation so that it can replicate the decisions of management to adjust workforce numbers. DES models can themselves incorporate feedback mechanisms, however this is usually of a more 'discrete' nature – such as drafting extra engineers if the backlog reaches a critical level or certain engineers checking the queue size after completing each job and going off duty if not required.

We believe that this chapter adds to the literature by providing another example of the benefits of using a hybrid modelling approach, over a standalone SD or DES model. Although the hybrid model is scaled–down and relatively simple, it demonstrates that a more complex scaled–up version may provide considerable benefits to

a large organisation like BT. This chapter also demonstrates some of the challenges that need to be overcome by researchers considering hybrid modelling within a DES environment. The hybrid model constructed in this chapter however, demonstrates that solutions can be found.

Chapter 3

Data-driven techniques for structural validation of system dynamics models

3.1 Introduction

3.1.1 Background

This chapter presents an approach for validating specific aspects of the structure of a system dynamic model. We present a data-driven approach that can be used to strengthen existing structural validation tests of SD models. We demonstrate the insights provided from this approach using the example of the BT hydraulics model. An exploratory analysis of historical time series data is presented that aims to investigate three core assumptions of the hydraulics model. Fundamental to the feedback

loops within the hydraulics model is the assumption that BT respond to increased demand by increasing engineer numbers, as we explained in chapter 2. The nature of this increase in workforce is assumed to be linear. In addition, different geographical regions are assumed to exhibit the same behaviour.

Validation of the first of these core assumptions, that workforce numbers increase as demand increases, is also necessary step prior to chapters 4 - 6, where we attempt to estimate the parameters of the hydraulics model. This is because our method for parameter estimation relies on the structure of the hydraulics model being an accurate representation of BT's system. This core assumption of how the organisation responds to increases in demand is a key part of the model's structure.

More specifically, the following are the main assumptions of how the model responds to changes in demand:

1. The organisation increases workforce numbers for job flows experiencing periods of increased demand.

2. This increase in workforce is approximately linear.

3. Behaviour of the system across different regions is assumed uniform.

The objective of this chapter is to investigate these three hydraulics model assumptions using historical time series data from BT. One would expect assumption (1) to be an essential strategy for any organisation attempting to perform effective service

delivery. This is the core assumption that we refer to above. As part of our analysis we will also investigate assumption (2). As the dataset contains time series for 74 geographical regions across Great Britain, the analysis can investigate assumption (3) and determine system behaviour at a more detailed regional level.

The BT historical time series data are effectively queuing data providing information for job arrivals, queue size, job completions and workforce numbers. Investigating how the system responds to increased demand requires a particular focus on certain intervals within the time series. These are the periods where high demand is observed. Throughout, we refer to these periods as 'spike' periods. However it is worth noting that these increased demand periods are not always sharp sudden increases and may be steady increases observed over a couple of weeks. When the job arrivals increase, the job queue increases, placing extra strain on the system as current performance begins to fall away from the targets. These are the periods during which an observable response from the system, i.e. to increase workforce numbers, can be considered most likely to occur. The analysis of this chapter attempts to determine if (and to what extent) workforce numbers are increased as the organisation makes decisions that determine its response to the increased demand.

3.1.2 Validation of system dynamics models

In this section we explain some of the standard tests that are used for validating a SD model. Although these are usually subject to much iteration, Barlas (1996) defines the three main stages in the validation of a SD model as the following:

- *Direct structure tests*: These determine if the structure of the model is comparable to knowledge about the structure of the real system. This stage in the validation process is performed before the model has been coded and involves various stakeholders involved in the project. These tests assess the validity of individual model equations and are divided into two types; empirical and theoretical. These are summarised well in Barlas (1994). Empirical tests compare model equations to knowledge of relationships in the real system. Theoretical tests compare model equations with general knowledge of the system from the literature. See also Barlas (1989a), Barlas (1989b) and Forrester & Senge (1980) for more details. Barlas (1989a) points out that the order of the validation process is essential; the direct structure tests must come first. The reason given by the author is that unless we have confidence in the model's structure, there is no point in proceeding with the behaviour tests below as a model with serious structural flaws can still produce accurate behaviour with sufficient 'parameter tuning'. This view on the ordering of tests reflects the general view in the literature, see for example Forrester & Senge (1980), Qudrat-Ullah (2011) and Barlas (1996). Barlas (1989b) defines these as 'strong' tests as they directly assess the model's structure, but points out that their weakness is their qualitative nature meaning that they can be difficult to communicate.

- *Structure-oriented behaviour tests*: These are tests that use model output to assess the model's structure indirectly, by applying certain behavioural tests. For example, the *extreme-condition* test assigns extreme values to certain model pa-

rameters and compares the model's behaviour to observed or expected behaviour of the real system. Barlas (1989a) points out that although the characteristics of SD models mean that standard statistical tests are unsuitable, a process consisting of only qualitative methods such as direct structure tests is insufficient for model validation. Barlas (1989b) argues that this stage of tests are the most important as they are 'behaviour tests that can provide some structural information'. Only when the model has passed the first two validation stages is it possible to proceed to behaviour pattern testing.

- *Behaviour pattern tests*: These involve comparing behaviour patterns from the model output to those of the observed data. Barlas (1989a) suggests a 6-stage process incorporating a number of quantitative methods that analyse time patterns rather than individual data points. These include comparisons of trend, period, mean and amplitude variation, phase lag detection and calculation of an overall summary measure which is to be used strictly as a reporting tool after the model has passed all previous tests. Modifications of this process are outlined in Barlas (1994). Several tests discussed in Forrester & Senge (1980) also fall into this category. Sterman (1984) describes an overall summary statistic based on the historical fit of the model to calculate the behaviour discrepancy in terms of bias, variance and covariance. Behaviour testing is generally considered to be weaker than structural tests. Barlas (1989a) warns that if these tests are used without structural tests, 'spurious behaviour accuracy' (where structural errors are present in the model but sufficient 'parameter tuning' has been performed)

cannot be separated from true behaviour validity where the model has an accurate structure and behaviour. However the author points out the usefulness of these tests for models where the structure has been shown to be accurate; a poor performance in these tests shows that some parameters or exogenous variables are not represented correctly. A strong performance increases the user's confidence in the model.

The hydraulics model has already undergone a formal validation process by specialists at BT. This involved tests from all three stages of the validation process outlined above. In this chapter we present a data-driven approach that can be used to strengthen existing structural validation tests of SD models. We claim that this data-driven approach can be considered as an addition to the direct structure tests described earlier. Although this type of test is not normally data-driven, we argue that this type of direct structure test can offer additional insights in the validation process. In this chapter this is demonstrated by assessing the three assumptions of the BT hydraulics model highlighted in section 3.1.1, through analysis of historical time series data from the system.

The structure of this chapter is as follows. Section 3.2 gives an overview of the methods used in the analysis and describes the historical time series data. Section 3.3 explains how spike periods are detected. Section 3.4 describes how system changes are measured within spike periods. Section 3.5 presents details of the algorithm that is used for the analysis, with the results presented in section 3.6. Section 3.7 presents

the discussion.

3.2 Methodology

3.2.1 Overview

In this section we present an overview of the methods used in this analysis. We begin by summarising below the three main stages of the analysis.

1. Spike detection: In order to investigate system behaviour around 'spikes' in job arrivals, criteria are defined for what constitutes a spike and how far above normal fluctuations job arrivals must be. For details see section 3.3.

2. Definition of variables: Having defined spike periods of interest, key variables are defined that measure changes in system behaviour. For details see section 3.4.

3. Regression: Using the variables defined above, the important factors that affect the system's response to the spike are determined. This allows us to investigate the validity of assumptions (1)-(3) in section 3.1.1. Results of these regression models are presented in section 3.6.

In section 3.5 we provide full details of the algorithm that is used in the analysis of this chapter. The regression models of step (3) include 4 explanatory variables X_1, X_2, X_3, X_4 and a single response variable Y. We give a brief description of these here.

- X_1 *Additional job arrivals*: Number of additional repair jobs that arrive within the spike period.

- X_2 *Reserve repair capacity*: The system's ability to increase workforce numbers if required for repair jobs.

- X_3 *Repair tension*: The current performance of the system relative to targets.

- X_4 *Reserve provision capacity*: The system's ability to increase workforce numbers if required for provision jobs (e.g. the installation of telephone lines or broadband equipment).

- Y *Additional engineers*: Numbers of additional engineers that are utilised within the spike period.

For each detected spike, linear regression models are used in an attempt to use the variables $X_1, ..., X_4$ to explain the behaviour of the region's response Y. For each spike i, we therefore model the region's additional engineers as:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + \epsilon_i,$$

where β_i, i = 1,...,4 represent the estimates of the coefficients of the variables $X_1, ..., X_4$.

3.2.2 Regional time series data

In this section we describe the BT regional time series data. The data consists of 74 sets of 8 daily time series, each set representing a geographical region of Great Britain. The approximately 20 different BT job flows have been aggregated into only 2; repair and provision jobs. For each region, 4 of the time series describe the behaviour of the repair job flow, while the remaining 4 describe the behaviour of the provision job flow. We describe below the daily time series for each region. Note that these 4 series

are present for both repair and provision jobs - hence we have 8 time series for each region.

- job arrivals: number of jobs that arrived on that day.
- job queue size: the size of the job queue at the *start* of that day.
- job completions: number of jobs completed on that day.
- engineers: number of engineers working that day.

Within each region, all 8 series are of the same length but lengths of series differ between regions, ranging from 750 days up to 903 days. The only exception to this is a single region from the South-East of England encompassing the towns of Aldershot, Guildford and Haslemere which has a length of only 60 days due to missing data. This region was excluded from the analysis, leaving 73 regional sets remaining. All series commence on July 1st 2010, with the longest series expiring on December 19th 2012.

3.3 Spike detection

In this section we describe the process of detecting spike periods. This is made more challenging due to a seasonal pattern present in the data. Of the two job flows, repair and provision, the repair jobs are of highest priority to BT, since repairing the network is more urgent than installations of equipment for new customers. Hence repair job arrivals drive the system changes for each region, as the region attempts to maintain its job queue within performance targets. Hence the spike detection methods of this

section are applied to the repair job arrivals for each region.

A popular class of techniques for identifying changes in time series are known as change point detection methods, see for example Chen and Gupta (2011). These detect changes in the mean or variance of a time series at a particular point in time. However, the analysis of this chapter investigates core assumptions of BT's hydraulics model. This model responds to changes in the system observed over a period of time, usually a matter of weeks. Hence, our approach for spike detection focusses on gradual, longer-term changes in system behaviour, such as an increase in trend, over periods of up to 21 days. Hence for our analysis change point detection methods are not appropriate.

For each regional set of time series, 3 of the 4 series (job arrivals, job completions and engineer numbers), for both the repair and provision job flows, show a strong weekly seasonal pattern, with values decreasing considerably for weekend days. An example of this seasonality over a 10 week period for a region's repair job arrivals is shown in Figure 3.3.1.

In order to detect spikes, in sections 3.3.1 - 3.3.3 we present three alternative methods and define how a spike is detected for each. Each method has a different approach for treating the seasonality present in the data.

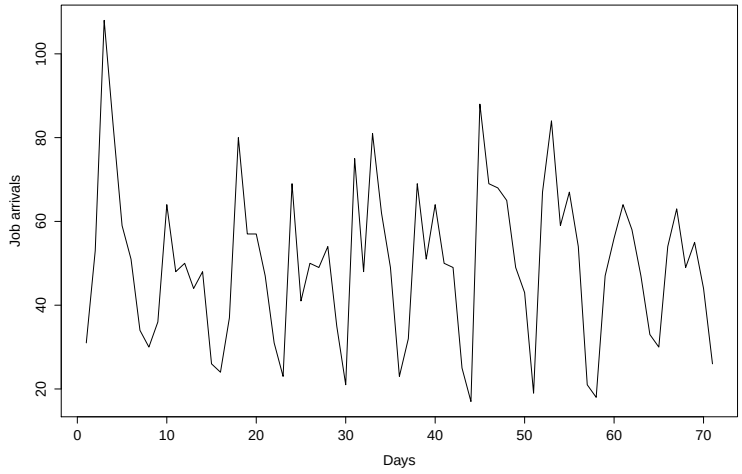

Figure 3.3.1: Example of weekly seasonality in the job arrivals.

3.3.1 Simple moving average

A simple method for smoothing a time series is to take a centred 7 day moving average (MA). This smooths out the seasonal effect and provides an estimate of the trend. In calculating the average values, observations are given equal weighting.

For spike detection, let a_t be the MA on day t. Then a spike is defined as any period where the MA increases above a threshold of 25% within 7 days, i.e. any a_t such that:

$$\max(a_t, ..., a_{t+7}) \geq 1.2\, a_t.$$

The value of the series at a_t is effectively used as a 'base' value and any increase beyond the threshold within 7 days is classified as a spike. In other words, a_t provides a 'counter-factual', i.e. the expected value of the series on day t if there was no spike. Use of the MA in this way, means that the detection of any significant changes in

trend, indicates that a spike in job arrivals has occurred.

3.3.2 Holt-Winters method

In this section we present an alternative method for providing a 'counter-factual'. MA estimates the trend of a series, but ignores any seasonality exhibited by the data. Hence using MA on a series that exhibits seasonality produces an estimate of the trend that may be corrupted by the seasonality in the data. As the BT data exhibits both a trend and seasonality, we investigate the use of the Holt-Winters method. This models level, trend and seasonality using separate components.

This method lies in the category of exponential smoothing techniques. Charles Holt's classic paper, Holt (1957), on exponentially weighted moving averages initially appeared in 1957, see Holt (2004) for an updated version. This technique extends to time series that also exhibit seasonality, where it is known as the Holt-Winters method (Winters, 1960). The notation below which uses the additive treatment of the seasonal component is similar to that used in Hyndman et al. (2008). The authors point out that the multiplicative treatment is more common than the additive, however this requires non-zero values in the time series. This was not always the case for the BT data, mainly due to behaviour on Sundays and Bank Holidays.

Like any exponential smoothing method, the method outlined here is an algorithm that produces a point forecast (Hyndman et al., 2008). For job arrivals series $y_1, y_2, ..., y_n$ with a seasonal period of 7, the one step ahead point forecast using the

additive Holt-Winters method is $\hat{y}_{t+1|t} = l_t + b_t + s_{t-6}$, where l_t is the level term, b_t is the trend term and s_t is the seasonal term, with each defined on day t as follows:

$$\begin{aligned} l_t &= \alpha(y_t - s_{t-7}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-7}. \end{aligned}$$

α, β and γ are the smoothing parameters for the level, trend and seasonal components respectively. These are defined such that $0 < \alpha, \beta, \gamma < 1$. The point forecasts $\hat{y}_{t+1|t}$ derived in this way represent a weighted moving average of past observations with weights decreasing exponentially (Hyndman et al., 2008).

We estimate these smoothing parameters by maximum likelihood, in order to minimise the squared prediction error of the 1 step ahead forecasts; $\sum_{t=1}^{n} e_t^2$, where $e_t = y_t - \hat{y}_{t|t-1}$. We use the maximum likelihood estimates, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, to calculate the fitted values $\hat{y}_t$ of the series for each day t. These fitted values $\hat{y}_t$ can be broken down into fitted values for the individual level $\hat{l}_t$, trend $\hat{b}_t$ and seasonal components $\hat{s}_t$, since:

$$\hat{y}_t = \hat{l}_t + \hat{b}_t + \hat{s}_t.$$

For the purposes of spike detection, our interest is not in the seasonal component, rather it is in estimating the trend in the series. Therefore after obtaining the overall fitted values $\hat{y}_t$, we subtract the fitted values for the seasonal component $\hat{s}_t$ from the overall fitted values as follows:

$$\tilde{y}_t = \hat{y}_t - \hat{s}_t = \hat{l}_t + \hat{b}_t.$$

Whereas MA's estimate of trend may be corrupted as it ignores the seasonality in the data, by estimating the seasonal component in this way via Holt-Winters and then removing it from the fitted values, we can estimate the trend in the data, without this estimate being corrupted by the seasonality. Hence we select this approach as the default for spike detection.

This filtered series $\tilde{y}_t : t = 1, ..., n$, also provides a smoothed representation of the original series $y_1, y_2, ..., y_n$, as we see in Figure 3.3.2. This compares the filtered series with MA, applied to the arrivals for an example region. For the purposes of spike detection, let $a_t = \tilde{y}_t$ be the filtered series on day t. A spike is defined as for MA in section 3.3.1, but the threshold is decreased to 20% to ensure that both methods detect similar numbers of spikes for analysis. Hence a spike is defined as any period where the filtered series increases 20% or more within 7 days.

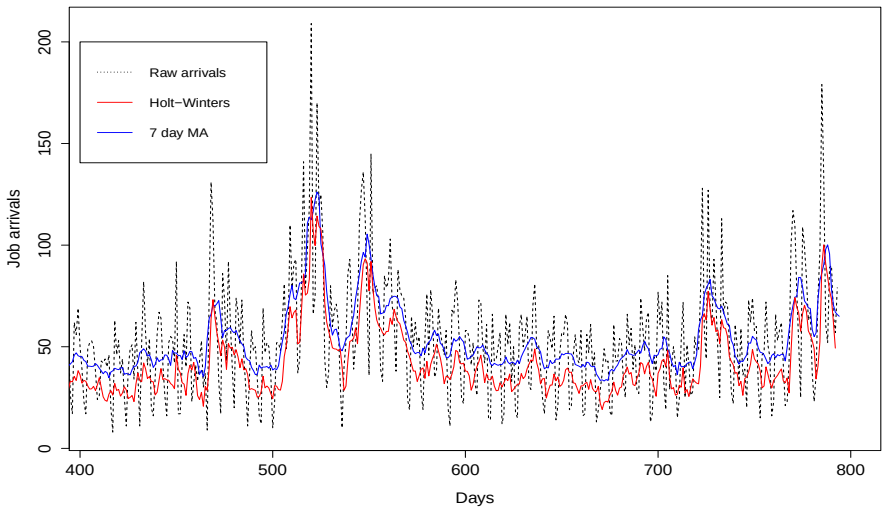

Figure 3.3.2: Holt-Winters and 7 day MA.

3.3.3 Seasonal differencing

MA and (our use of) Holt-Winters filtering, result in smoothed series that provide an estimate of the trend. However, a smoothed job arrivals series means that we may potentially be unable to detect any short-term increases. We investigate the use of an alternative to these methods, seasonal differencing; see for example Box et. al (2015). In the previous section we use Holt-Winters to model the seasonality and then remove it. Seasonal differencing removes the seasonality (and often the trend also), rather than modelling it.

To perform seasonal differencing on the BT data, the raw series are differenced for each weekday. We define the resulting series as d_t, the differenced series on day t. This differenced series in many respects resembles a stationary series. A spike is defined as any increase in d_t resulting in a value greater than or equal to 1.8 standard deviations of the raw series for job arrivals $y_1, ..., y_n$. That is, any d_t such that

$$d_t \geq 1.8 \text{ s.d.}\,(y_1, ..., y_n).$$

3.4 Definition of variables

In this section we explain how system changes are measured within the spike intervals of the time series. In doing so we calculate explanatory variables $X_1, ..., X_4$ and response variable Y for the regression modelling.

The spike detection approaches of section 3.3 presented the use of 3 methods for

treating the raw time series data. When defining variables, we must be consistent, and use the same method that was used to treat the data. E.g. if Holt-Winters filtering is used to treat the data, this method is also used in the definition of the variables. Since the 3 methods treat the data differently, using one method to detect a spike and then another to define the variables, may not sufficiently reveal any system changes. The only exception to this is in the definition of the variable X_3 which requires use of smoothed series as we explain below.

All variables defined in this section are calculated within a certain period of the spike arriving. From here on this is referred to as the *spike interval*. The length of this spike interval would seem to be an important issue. On the one hand, it must be sufficiently long to track system changes, for example giving the region a 'fair' chance to respond to the increased demand. On the other hand it should not be so long that it is detecting trends in time series rather than short-term effects/responses to a spike. A default interval of 14 days is chosen, with alternatives of 7 and 21 days also investigated. A multiple of 7 days must be used for the interval length due to the seasonality in the data.

In sections 3.4.1 - 3.4.4 we define the 4 explanatory variables $X_1, ..., X_4$ and response variable Y for the regression modelling. Section 3.4.5 describes some additional variables that were investigated.

3.4.1 X_1: Additional job arrivals

This variable attempts to measure how many additional jobs arrived within the spike interval.

(a) Holt-Winters and MA spike detection: Let h_t represent the smoothed job arrivals series on day t, using either Holt-Winters or MA methods (although for MA this was denoted by a_t in section 3.3.1). If a spike is detected on this day t and the default spike interval length of 14 days is used, then this interval covers the period (t, t+13); i.e. the day of the spike and the following 13 days. Figure 3.4.1 shows an example of the behaviour of h_t during a spike interval. The interval is contained by the two vertical red dashed lines. $X_{1,t}$ is calculated by measuring values of h_t in

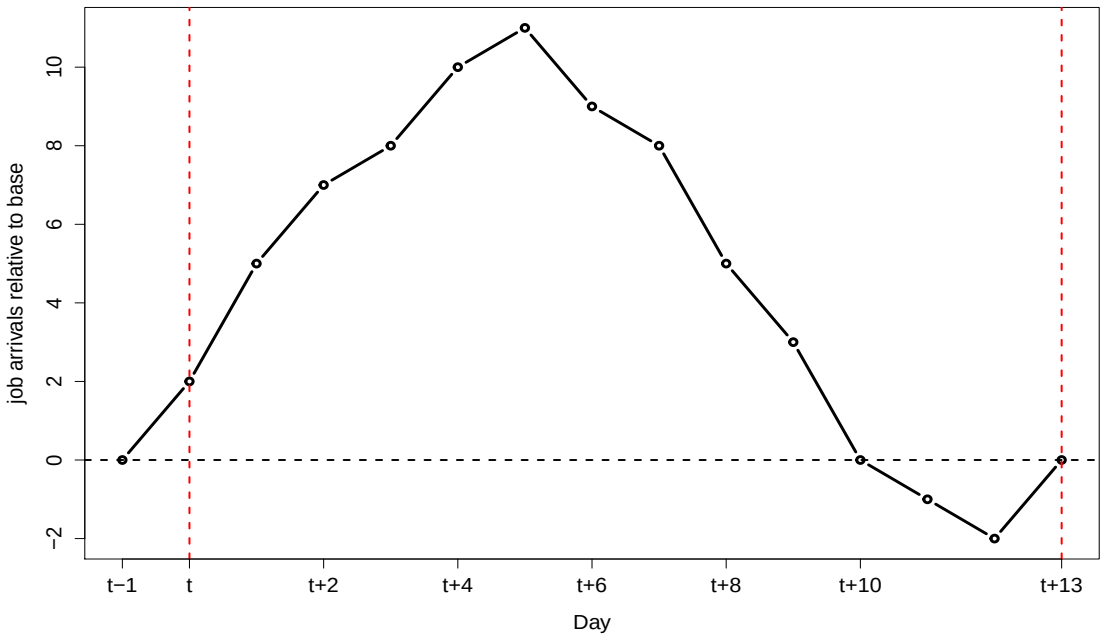

Figure 3.4.1: Variable definition for smoothed series.

this spike interval relative to a 'base' measure of h_t on day $t-1$; h_{t-1}. This base measure is the value of the smoothed series on the day prior to the spike arriving, represented by the horizontal black dashed line in Figure 3.4.1. For each day of the spike interval, the additional arrivals are calculated as that day's arrivals minus the base arrivals. An example of these daily differences between h_t and the base h_{t-1} is plotted in Figure 3.4.1. $X_{1,t}$ is the mean of these differences and hence is calculated as follows:

$$X_{1,t} = \frac{1}{14}\Sigma_{i=0}^{13}(h_{t+i} - h_{t-1})$$

As is the case in Figure 3.4.1, not all these differences will be positive; for some days within the spike interval h_t may lie below the base value. To ensure that no important information is missed, negative differences were also taken into account. This provides a complete record of the behaviour of the series throughout this interval. There is a risk that measuring system changes of interest such as $X_{1,t}$ in this way may become obscured by highly negative differences later in the 14 day interval. However if this is the case it is likely to be exposed by investigating use of shorter interval lengths. The results presented in section 3.6 reveal the effects of using alternative interval lengths of 7 and 21 days.

(b) Differencing spike detection: Let y_t be the raw job arrivals (for repair jobs) on day t, as was defined in section 3.3. Rather than using a single base value, we now calculate the difference between the raw arrivals series in the spike interval $(y_t, ..., y_{t+13})$ and a base value *for each of the previous 7 days* prior to the spike

arriving, $(y_{t-7}, ..., y_{t-1})$.

$$X_{1,t} = \frac{1}{7}\Sigma_{i=0}^{6}(y_{t+i} - y_{t+i-7}) + \frac{1}{7}\Sigma_{i=7}^{13}(y_{t+i} - y_{t+i-14})$$

So that for example each of the two Mondays within the spike interval is differenced with the latest Monday prior to the spike arriving. The first summation calculates the difference between arrivals on days 1-7 of the spike interval and the base values while the second summation calculates the differences for days 8-14.

3.4.2 X_2: Reserve repair capacity and X_4: Reserve provision capacity

These variables attempt to measure the system's ability to increase workforce numbers if required. Throughout this section we provide the definition for variable X_2. Note however that the definition of X_4 is exactly the same as that for X_2, except that all the time series involved in the definition are the corresponding series for provision jobs, rather than repair jobs.

(a) Holt-Winters and MA spike detection: Let s_t be the smoothed repair engineer numbers on day t and m_t be the previous 28 days' average on this day. Calculate $X_{2,t}$ as follows:

$$X_{2,t} = m_t - s_t.$$

Thus $X_{2,t}$ provides an estimate of the current engineer numbers relative to the previous month, by comparing short and longer-term averages. If the value of s_t when the spike arrives is considerably lower than m_t, we may reasonably expect that it is possible for the system to increase workforce numbers if required. If however s_t is greater than m_t, we may expect limitations in the system's ability to respond.

(b) Differencing spike detection: Let r_t be the raw engineer numbers on day t. Values for each day of the week prior to the spike arriving $(r_{t-7}, ..., r_{t-1})$ are compared with their average values over the previous 4 weeks, m_t. Hence this time we calculate m_t as follows:

$$m_t = \frac{1}{4}\Sigma_{i=1}^{4} r_{t-(7i-1)} = \frac{1}{4}(r_{t-6} + r_{t-13} + r_{t-20} + r_{t-27}).$$

For each day of the week prior to the spike arriving, the average difference between the raw arrivals r_t and the previous 4 week average m_t is calculated by the variable $X_{2,t}$ and takes the form below:

$$X_{2,t} = \frac{1}{7}\Sigma_{i=1}^{7}(m_t - r_{t-i})$$

So for example one part of the summation involves comparing the Monday's value in the week prior to the spike with the average of the previous 4 Mondays.

3.4.3 X_3: Repair tension

Defining this variable requires knowledge of the cycle time (CT). The cycle time is the time measured from when the repair job is first reported to BT, up until the time

the job is completed; i.e. the time that it takes for a job to pass through the system. This variable was not present in the regional dataset and was therefore estimated for each day t using Little's law on the smoothed series for the job queue (q_t) and job completions (c_t). See Little (1961) for the original proof of Little's law and Little (2011) for the updated 50th anniversary edition.

Little's law states that the average cycle time is equal to the size of the job queue divided by the rate of job completions. This formula is exact in a steady state system where job arrivals into the queue are equal to the number of departures from the system; the job completions. This was not always the case for the dataset under analysis. However Little's law is often used as an approximation to this relationship, e.g. see Sterman (2000). The BT organisation also use this method extensively as an approximation. Since we have daily series, Little's law was used for each day t to estimate a daily average cycle time for the repair job flow as follows:

$$CT_t = \frac{q_t}{c_t}.$$

Tension (T_t) for repair jobs on day t is a measure of how well the system is performing relative to targets. This is based on the estimate of the average cycle time and also the target cycle time, which is known to be around 2 days. The value of this target implies that the organisation seeks to complete all repair jobs within 2 days of their arrival. Tension compares current performance with the target, so that if a spike arrives on day t:

$$X_{3,t} = T_t = \frac{CT_t}{CT_{tar}} = \frac{CT_t}{2},$$

where CT_t is an estimate of the average cycle time on day t and CT_{tar} is the target cycle time. A tension value of 1 or less means that the performance of the system is meeting the targets. As tension increases above 1, the system performance diminishes further away from the targets and the need for increased workforce numbers increases. This variable can be considered to be a measure of the system's repair job performance when the spike arrives.

For all of the spike detection methods used, $X_{3,t}$ was calculated with the method described above using smoothed series. There was no equivalent variable based on differencing methods, since smoothed series were required to estimate the average cycle times. For MA and differencing spike detection, the series used to calculate $X_{3,t}$, q_t and c_t, were smoothed using moving averages. When Holt-Winters was used for spike detection, this was also used to smooth q_t and c_t.

3.4.4 Y: Additional engineer numbers

This variable attempts to measure how many additional engineer numbers were utilised within the spike interval. The definition of this is the same as for X_1, the additional job arrivals; the only difference is that the smoothed series h_t is now the smoothed repair engineer numbers rather than the job arrivals. Effectively we use h_{t-1}; the day prior to the spike as a base value and measure the average of any changes in this series within the spike interval.

3.4.5 Additional variables

The 4 explanatory variables $X_1, ..., X_4$ described above are those that were found to be significant (at worst at the 5% level) in the regression models in section 3.6. However, in total during the analysis, use of a number of additional variables was investigated. We describe some of these variables here.

- Duration of spike: number of consecutive days that the spike criteria were satisfied.
- Peak of spike: maximum of the smoothed series within the spike interval.
- Provision tension.
- Job queue and cycle time increases.
- 'Speed' of response: the relative speed of increase in engineer numbers, e.g. by calculating steepest gradients in the smoothed series.

3.5 Algorithm details

In this section we present the algorithm used to analyse the BT data. As we saw in section 3.4, values of our variables Y, X_1, X_2, X_3, X_4, are calculated using intervals of time series. We fit regression models to these variables in section 3.6. Hence later in this section, we also investigate any time-dependence within these variables.

The spike detection methods and variable definitions of sections 3.3 and 3.4 respec-

tively require subjective choices. These include the following, with the default selections given in parentheses:

- Spike interval length (14 days).
- Treatment of raw data (Holt-Winters).
- Spike detection threshold (20%).

The default selections are used for the national–level model of section 3.6.1, and the area–level models of section 3.6.2. In order to investigate the sensitivity of results to these choices, in section 3.6.3 we compare results for the national–level model using alternative selections.

We present the algorithm below. This assumes that the model choices are the default selections above.

1. For each geographical region r where $r = 1, \ldots, 73$:
   (a) Treat repair job arrivals series using Holt-Winters filtering.
   (b) Detect spike s when job arrivals of region r exceeds the defined threshold of 20%.
   (c) For each spike s, detected in region r, perform the following:
      - Calculate values for variables $Y_{r,s}$, $X_{1_{r,s}}$, $X_{2_{r,s}}$, $X_{3_{r,s}}$, $X_{4_{r,s}}$ using the relevant time series within the 14 day spike interval.

- To ensure that no spike intervals overlap, remove spikes lying within 14 days of other spikes by deleting the spike(s) with the smallest corresponding value of variable X_1. X_1 effectively measures how many 'additional' job arrivals result from the spike. E.g. for spikes s_1 and s_2 detected in region r, remove spike s_1 if $X_{1_{r,s_1}} < X_{1_{r,s_2}}$. Otherwise remove spike s_2.

2. National–level model: A single regression model is formed using concatenated data, collected from all spikes across all regions in a national-level model. The results of this model are presented in section 3.6.1. In this model, for each spike s, the additional engineers are modelled as follows:

$$Y_s = \beta_0 + \beta_1 X_{1,s} + \beta_2 X_{2,s} + \beta_3 X_{3,s} + \beta_4 X_{4,s} + \epsilon_s,$$

where $s = 1, ..., n$; with n being the total number of spikes across all regions in the dataset.

3. Area–level models: Separate area–level regression models are formed for each of the 9 geographical areas (groups of regions defined by BT's regional mapping) using concatenated data from all spikes across regions located within the geographical area. The results of these models are presented in section 3.6.2. For each spike s_a, detected within area a, the additional engineers are modelled as follows:

$$Y_{s_a} = \beta_0 + \beta_1 X_{1,s_a} + \beta_2 X_{2,s_a} + \beta_3 X_{3,s_a} + \beta_4 X_{4,s_a} + \epsilon_{s_a},$$

where $s_a = 1, ..., n_a$; with n_a being the total number of spikes across all regions

within area a.

As we see in the algorithm, our regression variables $Y, X_1, ..., X_4$ are defined using isolated intervals of the time series for each region. An example of these isolated intervals is presented in Figure 3.5.1 for an example region; Aylesbury. This shows the time series for both the smoothed repair engineer numbers (upper plot) and the variable Y (lower plot) – which is defined using the engineer numbers series, as we see in section 3.4.4.

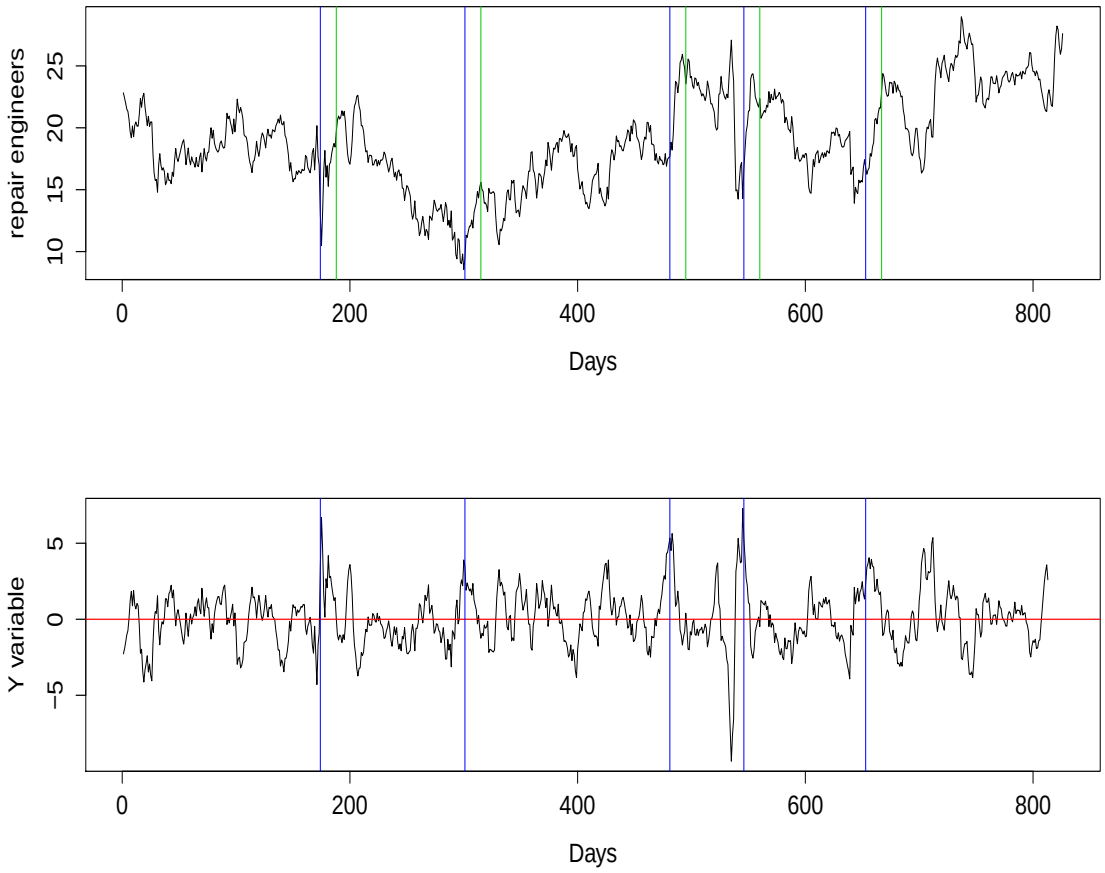

Figure 3.5.1: Time series of Y variable for Aylesbury region.

The vertical blue lines mark locations in both series where a spike has been detected in the corresponding job arrivals. The vertical green lines mark days of the repair engineers series that are 14 days after each spike has been detected – i.e. at the end of the *spike interval*. The value of variable Y at each spike location (intersected by the blue lines), is defined using only data within the spike intervals of the repair engineer series – the 14 day intervals starting at each blue line and ending at each green line in Figure 3.5.1. As we see in section 3.4, a similar process is used for the explanatory variables, $X_1, ..., X_4$, which are also defined using data from relevant time series that lie within these isolated spike intervals.

As we see in steps (2) and (3) of the algorithm, the spike data for each region is concatenated, either across all regions, or across groups of regions. For each of the regression variables, we refer to the indexing of these concatenated values as the *spike index*, which is best explained with an example. If the algorithm for the national–level model detects 10 spikes in region 1; $s_{1,1}, ..., s_{1,10}$, 11 spikes in region 2; $s_{2,1}, ..., s_{2,11}$ and also 8 spikes in the final 73rd region; $s_{73,1}, ..., s_{73,8}$, the spike index would concatenate data as follows: $(s_{1,1}, ..., s_{1,10})$, $(s_{2,1}, ..., s_{2,11})$, $..., (s_{73,1}, ..., s_{73,8})$. Spike data for the remaining regions would be entered in sequential order between regions 2 and 73. Hence the spikes of geographically neighbouring regions are adjacent in the spike index. In this way, data for each of the regression variables consists of spike data collected over all 73 regional time series. Since for each region, spike data is collected using isolated intervals in time, the spacing of the data over time within each of the regression variables is highly variable. Hence, adjacent data in the spike index is not

necessarily – and is usually not – adjacent in time.

In Figure 3.5.2 we present autocorrelation plots of the regression variables $Y, X_1, ..., X_4$ of the national–level model. The spike data within each of these variables are ordered by the spike index. These plots show the autocorrelations between spike data of each regression variable that lie close in the spike index – i.e. the autocorrelations at lower lags are between spike data obtained from the same or neighbouring regions. Due to the partially sequential nature of the spike index involved in the formulation of these variables, some notable autocorrelations exist, particularly at low lags. This is partly because spike data from the same region is likely to exhibit similar properties.

3.6 Results

In this section we present the results of the analysis. We begin by presenting results from the national-level model, analysing spikes across all 73 regions in section 3.6.1, in order to understand overall system behaviour. Section 3.6.2 presents results from 9 area-level models in order investigate regional differences in system behaviour. In section 3.6.3 we investigate the robustness of results to the subjective choices outlined in section 3.5.

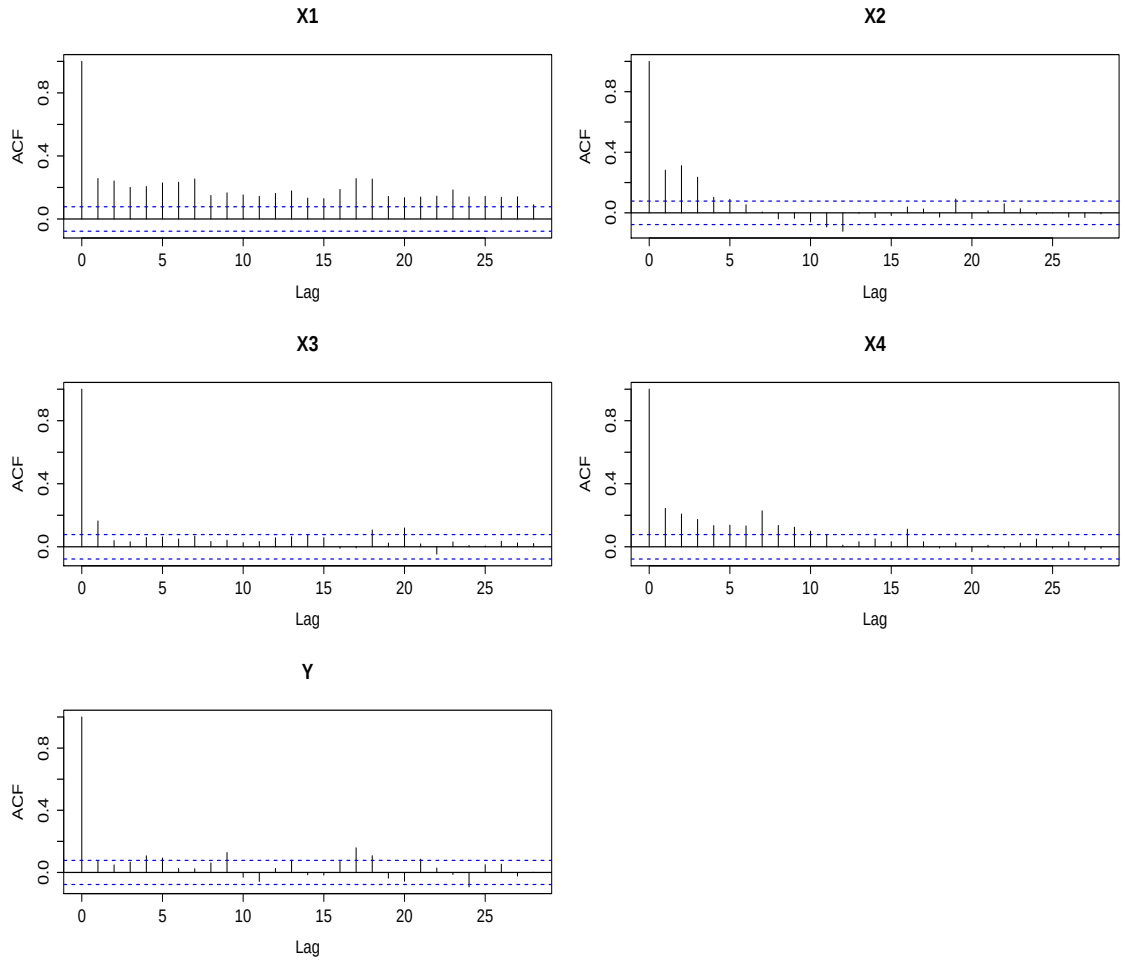

Figure 3.5.2: Autocorrelations within the spike index of the regression variables.

3.6.1 National-level model

In this section we present results from the national-level model, analysing spikes detected across all 73 regions in order to understand overall system behaviour across the whole of Great Britain.

For the national–level model, 639 spikes are detected in total. Initially Y was regressed on each of $X_1, ..., X_4$ in turn, with the variable from the model resulting in the lowest AIC (which was X_1) being retained in the model. This was repeated until

the addition of another variable no longer improved the model; i.e. the AIC no longer decreased. The details of this stepwise regression are shown in table 3.6.1. *Note that the model described in each row of the table includes both the variable for that row, in addition to any variable(s) in rows above.*

Model	Explanatory variables	AIC	Adjusted R^2
1	X_1: Extra job arrivals	3889	0.26
2	X_2: Reserve repair workforce	3680	0.47
3	X_3: Repair tension	3639	0.50
4	X_4: Reserve provision workforce	3635	0.51

Table 3.6.1: Model selection: stepwise regression.

From table 3.6.1 it is clear that in terms of explaining the behaviour of response variable Y, X_1 and X_2 are the most important variables. This implies that in a region's response to a spike, the key factors in understanding the number of additional engineer numbers utilised is the size of the spike (extra job arrivals) and whether the region *is able* to respond (repair reserve capacity). X_3 and X_4 are less important in explaining the response but are retained in the model since their additions improve AIC – which penalises over-fitting. Hence model 4 is used throughout this section. Figure 3.6.1 plots the explanatory variables against the response variable Y. For X_1, X_2 and X_4 in particular, an approximately linear relationship can be observed. Hence a linear model is an appropriate choice.

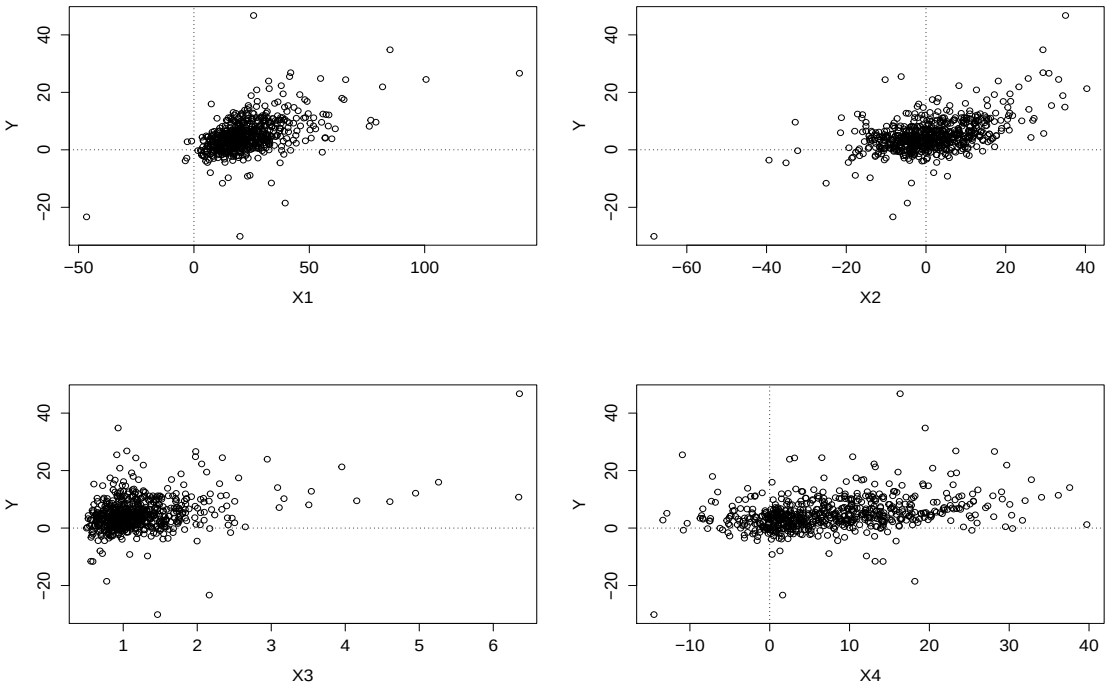

Figure 3.6.1: Explanatory variables vs response variable.

Table 3.6.2 provides the estimates, 95% confidence intervals and p-values from the regression model. The estimates of all four β coefficients are positive. That is, a spike consisting of more job arrivals, a greater reserve repair workforce capacity, a larger repair tension and a greater reserve repair workforce capacity are all associated with a response from the region resulting in an increased number of engineer numbers. Quadratic and cubic terms of the explanatory variables were also tested in the regression model but were not found to be significant. Estimates and confidence intervals for X_3 are provided for an increase of 0.1, rather than 1. This is because the standard deviation of the observations of this variable is around 10% of that of the other three explanatory variables.

Explanatory variable	Unit	Increase	Estimate	95% CI	p-value
X_1: job arrivals	jobs/day	1	0.18	(0.16,0.20)	$< 10^{-16}$
X_2: reserve repair	eng/day	1	0.22	(0.19,0.26)	$< 10^{-16}$
X_3: rep tension	NA	0.1	0.18	(0.13,0.23)	$< 10^{-10}$
X_4: reserve prov	eng/day	1	0.05	(0.01,0.09)	0.01

Table 3.6.2: Regression results for model 4.

The estimate of the coefficient of X_4 suggests that a region with a greater reserve provision workforce capacity is likely to have a slightly greater response in terms of engineer numbers. This is not as intuitive as the estimates of coefficients of X_1, X_2, X_3. A possible explanation is that a region's response consists in the main of overtime and also to a lesser extent the transfer of workforce; e.g. from provision to repair jobs. If a region has a greater reserve provision workforce when a spike arrives, then it may also have a greater *total* workforce. This would mean that more engineers are available for overtime and for transfers to other jobs such as repairs. We can also observe from table 3.6.1 that X_4 is the least important of the variables in terms of explaining the behaviour of Y. The estimate of its coefficient also has by far the largest p-value (0.01), as we see in table 3.6.2.

It is sensible to check that the explanatory variables do not exhibit multicollinearity. Table 3.6.3 shows the correlations between the 4 explanatory variables. The correla-

	X_1	X_2	X_3	X_4
X_1	1.00	0.13	0.08	0.26
X_2	0.13	1.00	0.24	0.24
X_3	0.08	0.24	1.00	0.15
X_4	0.26	0.24	0.15	1.00

Table 3.6.3: Correlations between the explanatory variables.

tions are all relatively low and hence do not cause any concerns for the national-level regression model.

The diagnostic plots from model 4 in table 3.6.1 are shown in Figure 3.6.2. Upper left of this figure are the observations of Y plotted against the fitted values from the model. Aside from a small number of outliers and a slight 'fanning out' of points observed for higher values, the fitted values appear to be lying close to the observations, with no clear asymmetry in spread of points around the 'observed = fitted' line. In other words the model is fitting the data reasonably well.

Lower right of Figure 3.6.2 plots the residuals versus the spike index. Figure 3.6.3, presents a larger version of this plot for the standardised residuals, complete with vertical blue lines dividing the spike index into the 9 geographical areas. Recall from section 3.5, that due to our methods for obtaining and concatenating spike data, the x-axis here is *not* time. We can see that there are a number of outliers and in addition,

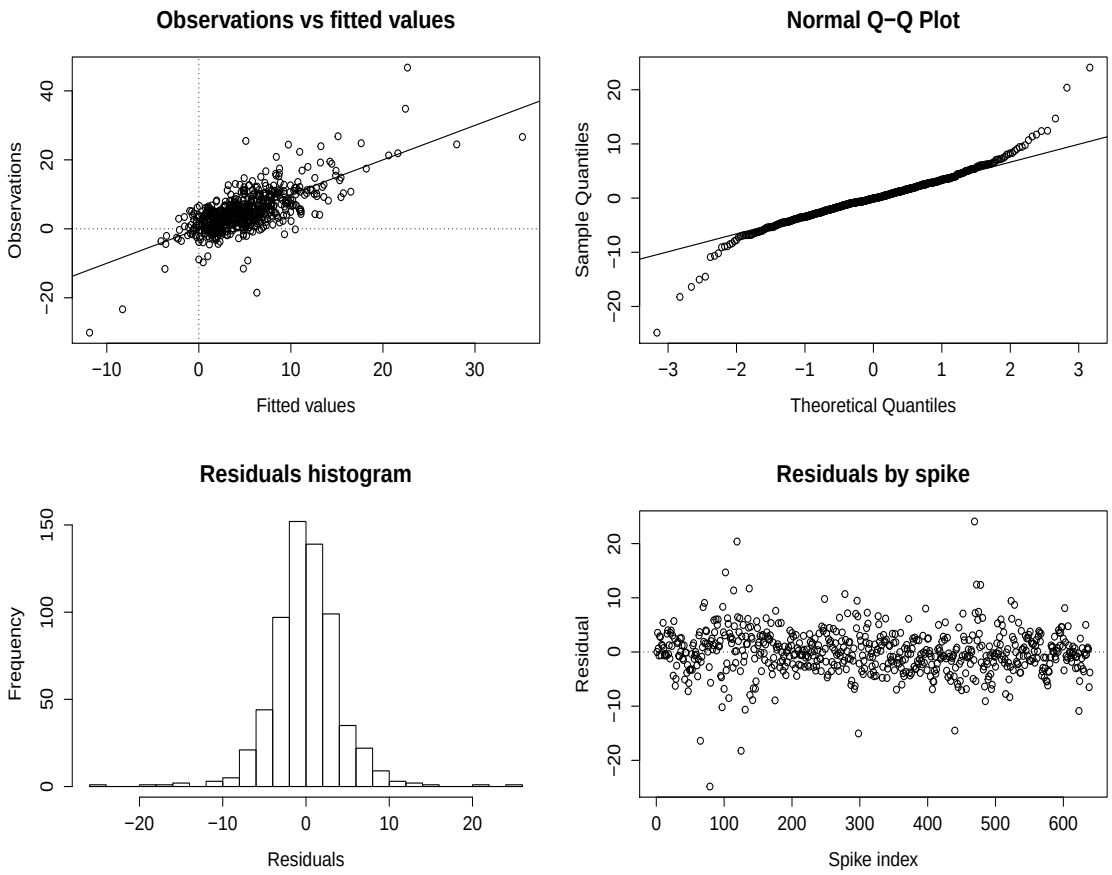

Figure 3.6.2: Regression diagnostics from model 4.

these outliers seem to appear in clusters. This is particularly evident for areas 1 and 7 – when counting from the left.

Since spike data for neighbouring regions lies close together in the spike index, the clusters of residual outliers in Figure 3.6.3 correspond to spike data from the same or neighbouring regions. This suggests that the national model is fitting some regions (and hence some geographical areas), better than others. As we will see, this is supported by results in section 3.6.2 – the national model is attempting to capture overall relationships between variables, but these relationships change across different

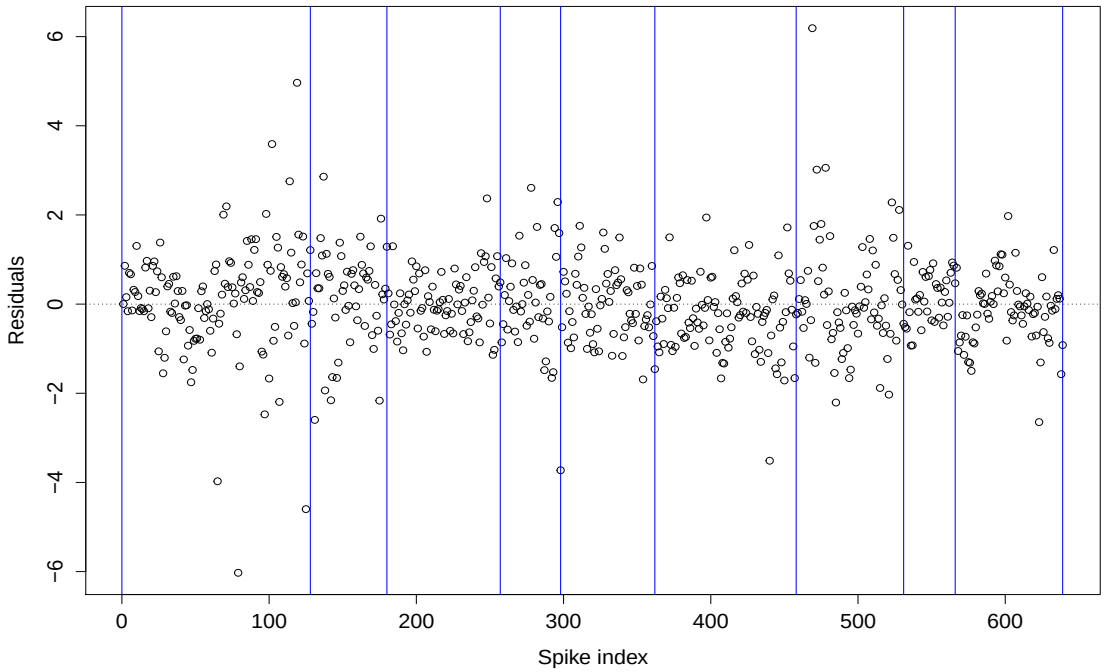

Figure 3.6.3: Standardised residuals ordered by the spike index.

geographical areas. This effect is also apparent in Figure 3.6.4 – which shows the autocorrelations of the residuals – with some low (but significant) correlations at lower lags. In addition, some correlations are also significant at higher lags, although these values are low and lie close to the 95% confidence interval.

Another assumption of linear regression is that the residuals have constant variance. Figure 3.6.5 plots residuals against explanatory variables. As the explanatory variables increase, there is no clear increase in variance of the residuals.

As we have already highlighted, the residuals of the national model have a number of outliers. These are revealed in both the Q-Q plot of Figure 3.6.2 and also Fig-

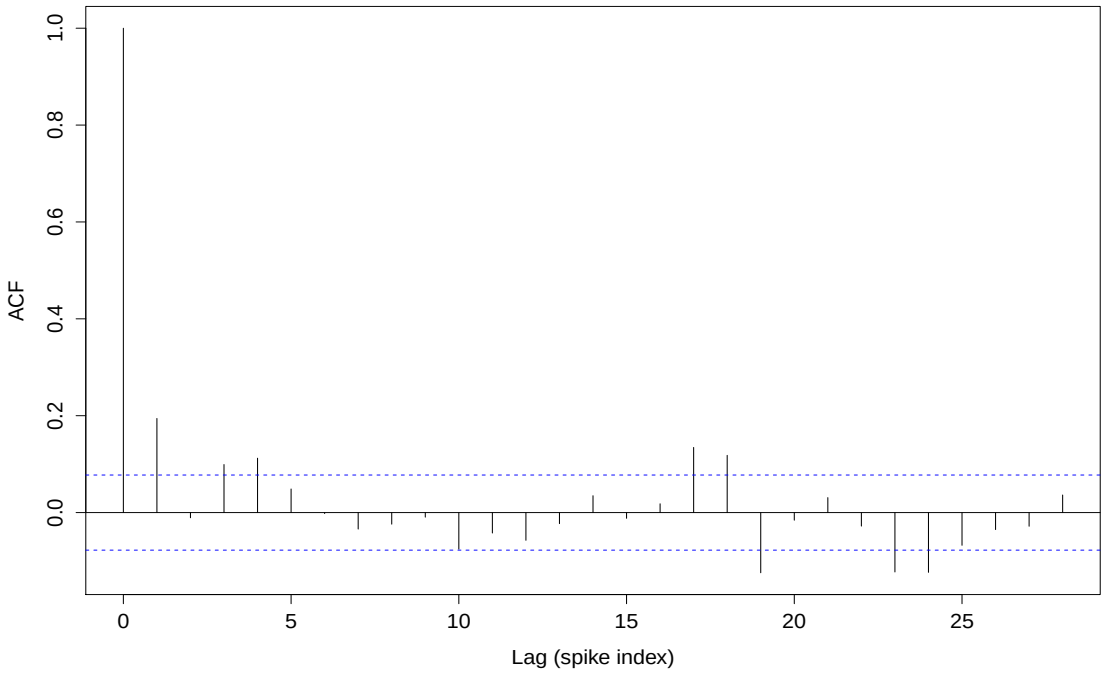

Figure 3.6.4: Autocorrelations of residuals

ure 3.6.3. The majority of the outliers are from geographical area 1; Scotland. A visual inspection of job arrivals for Scottish regions reveals generally more 'spiky' behaviour, compared to other geographical areas. More spikes per region are detected in Scotland (21.3) than any other area. This value is nearly triple the average for the other 8 areas (7.75). The behaviour of Scotland is well known amongst BT scientists. Certain parts of Scotland are known for having in general more extreme weather conditions than England and Wales. Extreme weather such as high winds or heavy rain increases the number of faults that occur in the network – hence leading to more spikes in the job arrivals series.

Another regression assumption is that the residuals are normally distributed. We

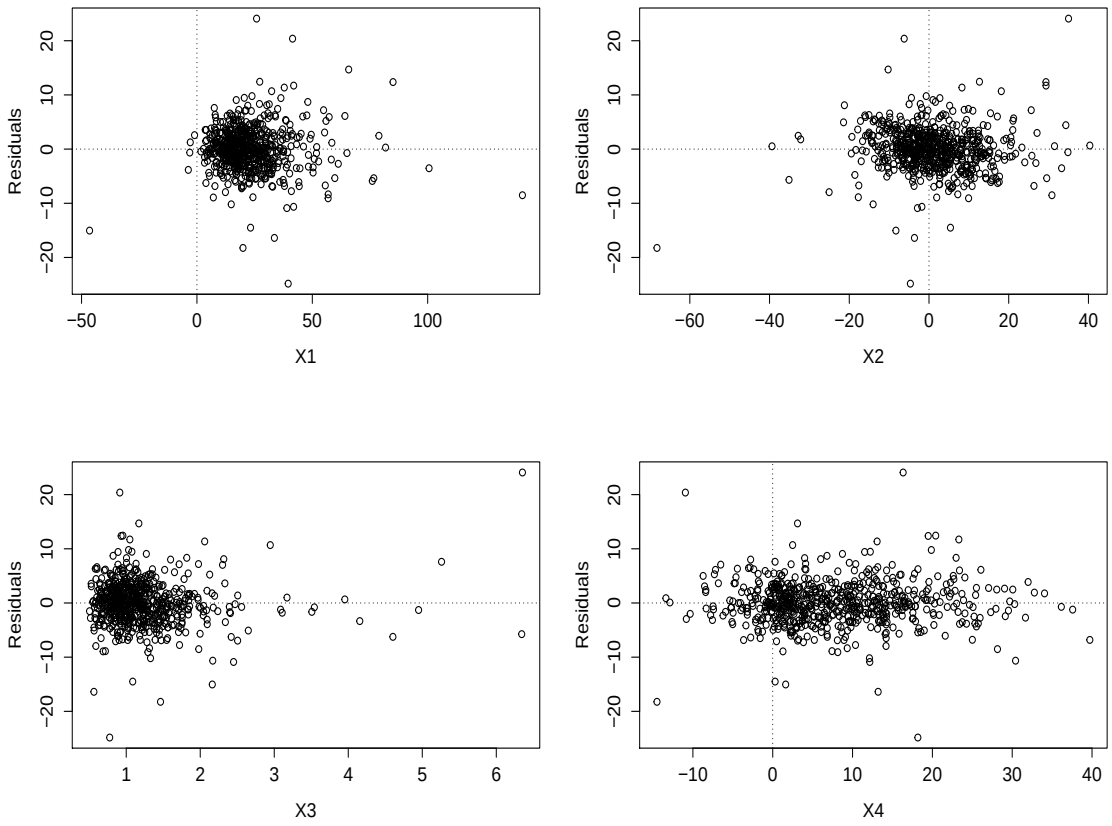

Figure 3.6.5: Explanatory variables against residuals.

can use the Shapiro–Wilk test (Shapiro and Wilk, 1965) to investigate the normality of the residuals. The resulting p-value of this test is $p < \alpha = 0.05$, meaning that we reject the null hypothesis (that these residuals are Gaussian) at the 95% confidence level. The Normal Q-Q plot of the residuals from the regression in the upper right of Figure 3.6.2 raises further concerns about the normality of the residuals. A likely reason for these results is the large number of outliers highlighted above.

To ensure that the outliers identified in Figure 3.6.3 are not excessively influencing the results, robust regression was performed and compared with the existing results.

Huber & Ronchetti (1981) and Hampel et al. (1986) discuss a common method of robust regression using *M-estimation* to minimise the objective function. Fitting is performed using iterated re-weighted least squares (IWLS). Three common estimators were investigated, defined by Huber, Hampel and an additional bisquare estimator. Use of Huber's estimator is effectively a convex optimisation problem, and the results are shown in table 3.6.4. Use of Hampel and bisquare estimators lead to similar re-

Explanatory variables	Estimate	95% CI (Robust)	95% CI (Model 4)
X_1: job arrivals	0.17 (0.18)	(0.15,0.18)	(0.16,0.20)
X_2: reserve repair	0.19 (0.22)	(0.16,0.21)	(0.19,0.26)
X_3: rep tension	0.13 (0.18)	(0.09,0.18)	(0.13,0.23)
X_4: reserve prov	0.07 (0.05)	(0.04,0.10)	(0.01,0.09)

Table 3.6.4: Robust regression results (Huber estimator).

sults. The second column from the left of this table shows the β estimate using robust regression with the previous estimate in parentheses. Although all four explanatory variables remain significant, the β estimates are now slightly different. Although perhaps a more notable difference for X_3 can be observed, the results are largely similar across the 4 variables. For X_1 and X_2, the two most important variables in the model, this similarity is particularly clear. This provides reassurance and suggests that that the regression results are not driven by the outliers identified in Figure 3.6.3.

3.6.2 Area-level models

In this section we focus on investigating system behaviour at area-level. This will enable a better understanding of the differences in behaviour at a more localised level. We have 639 spikes in total and 73 different geographical regions; meaning that on average there are less than 9 spikes detected per region. This is insufficient data to form a regression model for each region. However, these 73 regions can be divided into 9 geographical areas, so that each area includes a sufficient number of detected spikes to construct a regression model with sufficient data for inference. Each geographical area is a group of regions based on BT's regional mapping. Table 3.6.5 presents the results of fitting the regression model of section 3.6.1 (model 4 of table 3.6.1), separately to spike data from each of the 9 geographical areas.

The left most column of table 3.6.5 lists the 9 geographical areas across Great Britain. For each area, estimates of the regression coefficients are provided with a 95% confidence interval (CI) in parentheses. Line 10 shows the results obtained for the national model of section 3.6.1. For clarity, all values in the table except for spike numbers – i.e. values of β estimates and CI's – have been multiplied by 100. For example when analysing the spikes from the Scotland area, the estimate of β_1 for X_1 is 0.18, with a 95% CI of (0.13,0.24).

Although virtually all the CI's from the 9 area-level models have some overlap with the national model CI's, the majority of the β estimates from the 9 area-level models,

Area	spikes	β_1	β_2	β_3	β_4
Scotland	128	18 (13,24)	27 (18,36)	05 (-08,17)	-02 (-13,09)
North-East	52	19 (08,29)	29 (19,39)	31 (13,49)	-06 (-20,08)
North-West	77	17 (07,27)	17 (08,26)	15 (02,28)	13 (06,20)
N Wales/Mid	41	36 (25,48)	08 (-09,25)	23 (04,41)	08 (-15,30)
S Wales/Mid	64	12 (01,22)	33 (24,41)	03 (-10,17)	15 (04,26)
South-West	96	20 (15,26)	19 (10,27)	23 (03,43)	03 (-06,13)
South-East	73	17 (09,25)	22 (12,32)	51 (35,68)	14 (02,27)
London	35	13 (00,25)	25 (10,39)	-22 (-65,22)	00 (-13,12)
East Anglia	73	08 (-02,19)	09 (02,15)	09 (-11,30)	05 (-04,15)
National model	639	18 (16,20)	22 (19,26)	18 (13,23)	05 (01,09)

Table 3.6.5: Regression coefficient estimates for the 9 geographical areas (x100).

22 out of 36, fall outside of the corresponding national model CI's. In addition there are many cases where the area-level CI is considerably different to the national model CI. These area-level estimates alone are not sufficient to prove that we have significant regional differences. However they do raise the question as to what extent these differences are present, as one may reasonably expect these estimates to lie closer together (and more of them to lie within the main CI), if we are to believe that similar system behaviour exists across each of the regions. In addition, when performing the Shapiro–Wilk test on the residuals from each of the 9 models, results for 8 of the 9 models do not reject the null hypothesis (that the residuals are Gaussian) at the 95%

confidence level. These results provide some reassurance regarding the assumption of Gaussian residuals for the 9 models.

However it is worth highlighting that the inference that we can draw from these area-level results is limited. These results are based on considerably less data than the national model and hence less confidence can be placed in the β estimates, the CI's (which are considerably wider) and also the results of Shapiro–Wilk tests of normality.

3.6.3 Robustness of results

Due to the nature of this analysis, a number of subjective methods were required during spike detection and the tracking of system changes for variable definitions. We present alternative selections in this section and compare results to the national model results. Table 3.6.6 shows the effects on the regression coefficients and 95% CI's of some alternative selections, with the defaults used in the main analysis in bold font. Again these values have been multiplied by 100. The upper 3 rows compares methods for treating the seasonality and detecting spikes. The middle 3 rows compares use of different spike detection thresholds. The lower 3 rows compares different lengths of spike intervals. Each of the three sets of rows compares different selections, with the other two selections taking the default choice. For example, row 2 of table 3.6.6 uses moving averages to estimate the trend, a detection threshold of 20% and an interval length of 14 days.

All the estimates in the table lie within 0.1 of the default selections. The exception

	β_1	β_2	β_3	β_4	spikes
Treat data:					
Holt-W	**18 (16,20)**	**22 (19,26)**	**18 (13,23)**	**05 (01,09)**	**639**
Mov ave	13 (11,16)	45 (39,50)	25 (18,33)	-05 (-09,00)	698
Diff	14 (12,16)	38 (31,45)	28 (17,38)	-03* (-03,-02)	698
Threshold:					
15%	20 (18,22)	19 (16,21)	20 (16,25)	04 (00,07)	997
20%	**18 (16,20)**	**22 (19,26)**	**18 (13,23)**	**05 (01,09)**	**639**
25%	19 (16,21)	22 (18,26)	10 (03,18)	03 (-02,07)	446
Interval:					
7	15 (12,17)	16 (12,19)	21 (16,26)	08 (04,11)	639
14	**18 (16,20)**	**22 (19,26)**	**18 (13,23)**	**05 (01,09)**	**639**
21	26 (23,28)	24 (21,27)	18 (12,23)	00* (-04,04)	639

Table 3.6.6: Robustness of results (x100).

are values of β_2 in the upper 3 rows, where there are more considerable variations – in particular when using moving averages β_2 is around twice as large as when using Holt-Winters; 0.45 compared to 0.22. However since both estimates are positive the overall effect is the same; increases in X_2 still cause a greater response. It is the magnitude of this effect that has increased. In addition, although estimates of β_4 in the upper 3 rows lie with 0.1 of the default, the estimates using moving average and differencing are negative, unlike the default. Combined with the fact that X_4 had the

highest p-value and was the least important of the 4 variables in the model, this raises questions as to how much faith can be placed in the results involving X_4. Overall, these estimates can be seen to be fairly robust to these selections. Although values of certain coefficients change under different selections, the overall conclusions from the results of the national model are largely unchanged.

3.7 Discussion

The analysis of this chapter has attempted to determine whether key assumptions of the hydraulics model are validated by BT's historical time series data. We have found fairly strong evidence for assumption (1); that in general a measurable response to a spike is observed. In addition, there is evidence to support assumption (2); that this response is linear. Approximately linear relationships were observed between the explanatory variables and the response. In addition, during testing of the regression models, quadratic and cubic terms were included in the models but were not found to improve the model fits. Results however raise questions over assumption (3); whether this response is the same across different regions.

The national model revealed some interesting relationships in terms of how the system behaves as a whole across all the 73 regions. However certain regression assumptions, such as the requirement that the residuals are independent and Gaussian, do not strictly hold. Although there are a number of outliers, the robust regression results provide reassurance that outliers are not unduly influencing the results of the regres-

sion. Results suggest that the national-level model is fitting some geographical areas better than others. A limitation of the analysis was that subjective methods were required in the processes of spike detection and variable definitions. However, section 3.6.3 demonstrated that the national model results are fairly robust to some of these selections, with the overall conclusions largely unchanged.

Results of the area-level models suggest that the responses of each geographical area to increases in demand may not be the same. These models violated fewer assumptions than the national model, such as normality of the residuals. However the inference that we can draw is limited, since these area–level results are based on considerably less data (i.e. fewer spikes) than the national model. Hence less confidence can be placed in the tests and results of these models. This highlights a more general limitation of our approach. To obtain meaningful results from the regression model, we need a sufficient amount of spike data. This means that we cannot form conclusions separately for individual regions as the number of spikes detected in each is insufficient.

At this exploratory stage we are not able to make further claims regarding these potential differences in the behaviour of different regions. There is some evidence to support these claims, but these claims can only be substantiated by further investigation. We explore these regional differences through an alternative approach presented in chapters 4 - 6 where we attempt to calibrate the hydraulics model across geographical regions. Our method relies on the structure of the hydraulics model being an accurate representation of system behaviour. As such, the validity of assumption (1),

a core assumption of the model's feedback mechanism, is of particular significance for the approach used in these chapters.

This chapter has demonstrated that it is possible to investigate key assumptions of a SD model from historical time series data using regression methods. We have shown that such an analysis can be used to strengthen existing structural validation tests of SD models. We also suggest that this approach can be considered as an additional test to the group of direct structure tests described in section 3.1. More generally, this chapter demonstrates the insights that can be obtained from using data-driven methods to validate the structure of a SD model.

Chapter 4

Linear state-space models and the Kalman filter

4.1 Introduction

In this chapter we present an approach for estimating the parameters of system dynamics (SD) models from time series data. This is tested on simulated data from a simplified version of the hydraulics model. Unlike chapters 2 and 3 which presented self-contained research contributions, this chapter and the following two are related by their use of an approach to estimate the parameters of different versions of the hydraulics model. After commencing with the simplified version of the hydraulics model in this chapter, we progress in chapter 5 to applying the approach to simulated data from more complex versions of the hydraulics model – including the full version. We then attempt to estimate the hydraulics model parameters for different geographical regions using BT's historical time series data in chapter 6.

This introductory section is structured as follows. In section 4.1.1 we introduce the background for this research and its motivation from the perspective of an organisation such as BT that wishes to calibrate their models as accurately as possible. Section 4.1.2 describes the existing methods for estimating the parameters of SD models and explains the reasons for our chosen approach.

4.1.1 Background

The hydraulics model, used by BT to model their workflows and evaluate their service delivery, was introduced in section 2.3. The rationale behind this model and their choice of the SD approach is described in detail in the book chapter Jensen et al. (2013); written by the BT scientists who devised and developed the model. The authors state that the use of the word 'hydraulics' is used as a metaphor in order to describe the methodology used in the model "due to its analogy with reservoirs, flows and pressures in fluid mechanics".

BT analysts have a great deal of confidence in the structure and core assumptions of the hydraulics model. Chapter 3 demonstrated that BT's historical time series data for the 74 geographical regions of Great Britain supports a key assumption of the model; that BT increase workforce numbers during periods of increased demand. However analysts have considerably less confidence in the values of the hydraulics model parameters.

In a complex dynamic system, parameters are often not observable and therefore historical data cannot be collected to determine sensible values. When the hydraulics model was originally formulated, a combination of system knowledge/opinion (from relevant stakeholders) and best guesses were used to devise parameter values; i.e. they were not derived directly from data. This perspective of modelling approach within SD has been termed the 'classical' school of thought, see section 4.1.2. Although the model is used to represent 59 geographical regions across the UK, these parameters were originally assumed to be the same and a "default" set of parameters were used across the different regions.

More recently, there has been considerable interest at BT in calibrating the model, i.e. determining a set of parameters that are a good match to historical time series, for each geographical region. At BT, the current approach for calibrating the model involves using a cumbersome trial and error procedure that requires repeated runs of the model for different values of the parameters. We can demonstrate this as follows. If we denote the hydraulics model by *Hyd*, its model parameters as $\underline{\theta}$, exogenous variables as U and the output time series produced from the model as $\underline{y}$, we see that due to the nature of SD modelling, the hydraulics model is effectively a deterministic function of the parameters $\underline{\theta}$ and U as follows:

$$\underline{\theta} \Rightarrow Hyd\,(\underline{\theta}, U) \Rightarrow \underline{y}.$$

The current process at BT for calibrating the hydraulics model involves repeatedly running the model for different values of the parameters $\underline{\theta}$ and each time comparing

the output time series $\underline{y}$ to historical data. This process is repeated until $\underline{y}$ is a sufficiently good match to the historical data, with each set of parameters $\underline{\theta}$ selected with the aim of improving this match. This school of thought has been termed the 'hand calibration' school, see section 4.1.2. This procedure can be time-consuming and there is no guarantee that the parameters found are optimal. This is an issue particularly when there are a large number of datasets, as is the case for BT's 59 geographical regions.

BT analysts are aware of sometimes considerable differences in behaviour between these regions, so relying on a default set of parameters is likely to limit the accuracy of the model. Chapter 3 highlighted potential weaknesses in assuming that system behaviour is the same across all regions by demonstrating differences in the response to increased demand. This chapter, and the following two, aim to go a step further and quantify these differences through the hydraulics model parameters. If the model can be accurately calibrated for each region, BT can understand their systems at a local regional level and system performance can be improved.

In this chapter we improve an existing method that uses a state-space approach to automate the procedure of estimating parameters of SD models using historical time series data from the system under investigation. Effectively the approach attempts to solve the inverse problem by working backwards from historical data $\underline{y}^*$ in order to

obtain estimates $\underline{\hat{\theta}}$ of the hydraulics model parameters $\underline{\theta}$:

$$\underline{\hat{\theta}} \Leftarrow ... \Leftarrow \underline{y}^*$$

The approach relies on the assumption that the structure of the hydraulics model is a reasonable representation of the system - and hence that historical data $\underline{y}^*$ has a similar structure to $\underline{y}$ which is defined as above; the time series which would have been output from the hydraulics model using parameters $\underline{\theta}$. Starting with the historical time series data $\underline{y}^*$, we assume that this has the same structure as the hydraulics model output $\underline{y}$ but is corrupted by some Gaussian noise; i.e $\underline{y}^* = \underline{y} + \underline{\epsilon}$ with $\underline{\epsilon} \sim N(0, R)$, for some covariance matrix R. In other words we represent the system as a state-space model, as we explain in section 4.2.1. In state-space terms, this means assuming that each of the historical time series $\underline{y}^*$ has a latent state $\underline{y}$ which cannot be observed directly, but matches the structure of the SD model.

Since SD models are based on differential equations, to represent this in practice, we first discretise the model to form difference equations. These equations are re-arranged to represent the structure of each of the latent variables in the state-space model. We can estimate the value of these latent variables at each time series observation using a method known as the Kalman filter (Kalman, 1960), described in section 4.2.3. Terms calculated from the Kalman filter also enable the calculation of the log-likelihood at each time series observation, allowing the calculation of an overall log likelihood for the data, given a set of model parameters. Optimisation algorithms can then be used in a maximum likelihood approach to determine the best

set of parameters for the SD model to best represent the historical data.

4.1.2 Estimating parameters of system dynamics models

In this section we describe existing methods for estimating the parameters of SD models and explain why we selected our chosen approach. However, the question of whether it is even necessary to estimate the parameters of a SD model has been a divisive issue amongst system dynamicists for years. Radzicki et al. (2004) presents an excellent exposition of the two 'schools of thought' amongst modellers and suggests a third:

1. The 'classical' school, as termed in Peterson (2003), do not place any importance on estimating parameters and fitting SD models to historical data. Instead, Radzicki et al. (2004) point out their belief that "it is the stock-flow-feedback loop structure of a system, and not its particular parameter values, that determines its behaviour." Their belief is that the structure is so important that when it is properly represented in the model, parameter values can be increased/decreased at least 10% without significantly changing the model behaviour - hence estimating these parameters is considered unimportant. Radzicki et al. (2004) refer readers to Legasto and Macariello (1980) and Forrester (1980a; 1980b). To estimate the parameters in the model, the modeller must obtain information at the required level. This could involve interviewing the system's decision makers or measuring delays in the system (Sterman, 2000). Due to the lack of scientific rigour in estimating the parameters, this school has

been criticised by a number of sources, particularly from the field of economics (Nordhaus, 1973; Radzicki et al., 2004).

2. The 'statistically inclined' school, termed by Richardson (1981), place a high importance on comparing simulated output with historical time series data. Unlike the other two schools, these methods use statistical rigour to estimate parameters. An appropriate objective function is selected and optimisation algorithms search for the best solution from the parameter space. A number of approaches have been successful in estimating SD model parameters from historical time series data. Chen et al. (2011) contains an excellent review of these approaches. One such approach, explained in sections 4.2.5 and 4.2.6, will be the focus of the next 3 chapters.

3. The 'hand calibration' school, which Radzicki et al. (2004) suggest offers a compromise between the other two. Hand calibration involves an iterative process of simulating the SD model, comparing the output to historical time series and adjusting the parameters to improve the fit. Lyneis and Pugh (1996) give a good exposition of this and highlight the following flaws:

    - The method relies more on the skills of the individual modeller, rather than a well-defined set of steps; as Radzicki et al. (2004) point out it's "more of an art than a science".
    - The results may not be replicable, as different modellers are likely to find different estimates of parameters.

- When the 'best' set of parameters are found, there is no guarantee that these parameters are optimal.

The method introduced in sections 4.2.5 and 4.2.6 does not suffer from these limitations.

Within the statistically inclined school, Dangerfield and Roberts (1996) present the two uses of optimisation for SD models.

- Policy optimisation to improve performance: There is often a necessity to set the model parameters to optimise a particular variable, for example to maximise profit. The authors explain that in this context, optimisation determines how parameters guiding workforce numbers (hiring & firing) and inventory control should be set to keep costs down. Chen et al. (2011) includes a thorough review of the different approaches that exist for policy optimisation. Many of the approaches detailed can also be used for model calibration.

- Optimisation to fit data: This involves estimating the model parameters to produce output that best matches historical time series data. This process is sometimes called model calibration (Oliva, 2002). Dangerfield and Roberts point out that a reasonable fit to historical data can be useful as a means of reinforcing confidence in the model for clients.

For estimating the parameters of the BT hydraulics model, it is the latter of these that is our focus. When attempting to calibrate a SD model, there are two sub-schools within the statistically inclined school for how best to achieve this:

- 'Fully integrated maximum likelihood via optimal filtering' (FIMLOF): This was devised by Schweppe (1973) and uses Kalman filtering together with his representation of the log-likelihood in (4.2.14) in a maximum likelihood approach for parameter estimation. This is our chosen method and we introduce this in more detail in sections 4.2.5 and 4.2.6.

- Model reference optimisation (MRO): Estimating model parameters is specified as an optimisation problem, adjusting parameters to minimise some function of the difference between simulated and historical data (Oliva, 2002). For more details see Lyneis and Pugh (1996). This method does not rely on any filtering approaches like FIMLOF.

When using optimisation techniques for SD models, Coyle (1996; 1999) highlights the importance of selecting an appropriate objective function and states that a poor choice could be 'truly disastrous'. As the objective function for MRO is a function of the errors (between simulated and historical data), for this objective function to be suitable, we must have sufficient confidence in the historical data and for example a high belief that it is not distorted by noise. Although the use of Kalman filtering in FIMLOF is more computationally intensive, an important advantage of the FIMLOF approach is that it is applicable to data that is known to be corrupted by noise (Peterson, 1976). Indeed, this is exactly the type of data that the Kalman filter is designed for. The historical BT time series data analysed in chapter 6 is believed by system experts to have been corrupted by noise.

For this chapter we restrict ourselves to linear systems. A commonly highlighted limitation of FIMLOF is that for nonlinear systems there is a need to linearise the system (Radzicki et al., 2004; Oliva 2002; Dangerfield & Roberts 1996). In chapter 5 where we proceed to nonlinear systems, we employ a sophisticated modern Kalman filtering technique known as the Unscented Kalman filter, an extension to the classical Kalman filter, proposed by Julier et al. (2000). This is capable of accurately approximating a nonlinear system. We show that the use of this method largely addresses this limitation.

An additional advantage of FIMLOF is that it involves optimising a log-likelihood. Although in practice computational difficulties may exist, in theory with an appropriate optimisation technique we are guaranteed to optimise the likelihood surface. There is no such guarantee when a function of the errors is the objective function as in MRO. There are a number of examples in the literature where FIMLOF has successfully estimated SD model parameters, even for complex nonlinear models (Peterson, 1975; 1976; Ryzhenkov, 2002; Radzicki et al., 2004).

The objective of this chapter is to explain the background and details of the FIMLOF approach and demonstrate its effectiveness at estimating the known parameters of simulated data. In this chapter, we use data simulated from the simplified version of the hydraulics model. This was the model introduced in chapter 2 and from here on is referred to as hydraulics model 1. Section 4.2 contains the literature review and relevant background theory. Section 4.3 explains how we apply FIMLOF to estimate

the parameters of hydraulics model 1. Section 4.4 is a simulation study presenting the results of estimating the known parameters of hydraulics model 1 from noisy simulated data. Section 4.5 presents the discussion.

4.2 Literature review

Chapters 2 and 3 introduced background on SD modelling. In this section we present details of the FIMLOF method for estimating SD model parameters from historical time series data. We begin by introducing the necessary background theory on state-space models, Gaussian state-space models and the Kalman filter in sections 4.2.1 - 4.2.3 respectively. In section 4.2.4 we evaluate some popular techniques for estimating parameters of state-space models. The origins and early research using FIMLOF are then presented in section 4.2.5. The modern applications of FIMLOF are presented in section 4.2.6.

4.2.1 State-space models

In time series modelling, a vector of data points $\mathbf{x} = x_1, ..., x_T$ explains the behaviour of a population of random variables $\mathbf{X}$ over time. Hence for data that is continuous and observed at regular (e.g. daily) intervals, for each interval at time t, observation x_t describes the behaviour of X_t. There are occasions however when it may not be possible to observe the series $\mathbf{x}$ directly. For example, observations $\mathbf{y} = y_1, ..., y_T$ may be corrupted by noise and it may therefore be necessary to infer the behaviour of x_t from y_t for each interval at time t. One approach for dealing with this problem is to

represent the system as a state-space model. The structure of this type of modelling approach can be represented in Figure 4.2.1. The noisy series of observations $\mathbf{y}$ is

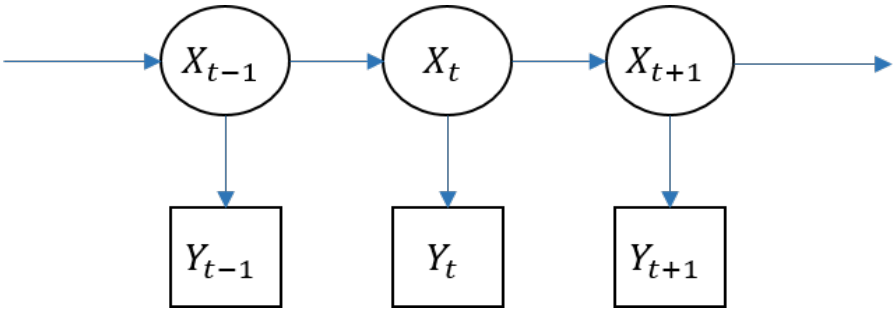

Figure 4.2.1: State-space model structure.

assumed to be dependent on a series of unobservable states $\mathbf{x}$. These unobservable states $\mathbf{x}$ are sometimes called latent states. For each interval at time t, the updating structure of noisy observation y_t and the underlying state x_t takes the following form:

$$x_t = f(x_{t-1}, u_t) + \epsilon_{1,t} \tag{4.2.1}$$

$$y_t = g(x_t, u_t) + \epsilon_{2,t} \tag{4.2.2}$$

where u_t is an exogenous variable or control input at time t. The *state equation* (4.2.1), describes a Markov process representing the updating structure of the state x_t as some function f of the previous state x_{t-1} and current value of the exogenous variable u_t, with the addition of some noise $\epsilon_{1,t}$. The *observation equation* (4.2.2), assumes that the observation y_t is some function g of the current state x_t and exogenous variable u_t, plus some noise $\epsilon_{2,t}$. For a general state-space model we make no assumptions on the distributions of the noise, $\epsilon_{1,t}$ and $\epsilon_{2,t}$. However, for the modelling approach in this chapter these are both assumed to be zero mean white noise processes. For a more thorough explanation of state-space models, see Durbin and Koopman (2001).

4.2.2 Gaussian state-space models

Gaussian state-space models are a special case of the general state-space model of (4.2.1)-(4.2.2). This class of models assume that the noise of both the state and observation equations has a Gaussian distribution. Again observations y_t are assumed to be dependent on unobservable state x_t at each update at time t. In the next three chapters, we focus on a particular set of Gaussian state-space models where the state and observation equations are defined as follows:

$$x_t = f(x_{t-1}, u_t) + \epsilon_{1,t} \tag{4.2.3}$$

$$y_t = x_t + \epsilon_{2,t}. \tag{4.2.4}$$

The state equation (4.2.3) has the same representation as (4.2.1), but note that the observation equation (4.2.4) has now dropped the g function of (4.2.2); y_t is simply the state x_t plus some noise. We drop this function for simplicity as it is not required in the systems that are studied as we show in section 4.3.3 when we form a Gaussian state-space model from hydraulics model 1. The noise terms have a Gaussian distribution as follows: $\underline{\epsilon}_1 \sim \mathcal{N}(0, Q)$ and $\underline{\epsilon}_2 \sim \mathcal{N}(0, R)$, with Q representing the state covariance and R the observational covariance.

The nature of the *state update function* $f(.)$ in (4.2.3) has important consequences. If $f(.)$ is a linear function, then we have a special case, the linear Gaussian state-space model. For this class of models, inference on observations $\mathbf{y}$ to understand the behaviour of the underlying states $\mathbf{x}$ is a much simpler task.

Linear Gaussian state-space models

For the general state-space model representation in section 4.2.1, if updating functions f and g in (4.2.1)-(4.2.2) are linear and we have $\underline{\epsilon}_1 \sim \mathcal{N}(0, Q)$, $\underline{\epsilon}_2 \sim \mathcal{N}(0, R)$ for some Q and R, then we have a linear Gaussian state-space model. These are sometimes referred to as dynamic linear models or DLMs. See Petris et al. (2009) for more details and an excellent introduction to this class of models, including methods for parameter estimation. West & Harrison (1999) also provides a thorough background text with details of the wide variety of models in this class.

The state-space models shown so far have been univariate with one series of observations $\mathbf{y}$ and a single underlying vector of states $\mathbf{x}$. Of course all these models extend to the multivariate case, where we have p different series of observations and q underlying states. Throughout this chapter and chapters 5 and 6, we study state-space models where $p = q$. That is, we have p different series of observations, each with its own underlying state, so that at time t, x_t and y_t have the same dimension. We can also have r different exogenous variables, though for the system studied in this chapter we have only one, so that u_t is a scalar. Due to the linear nature of the equations in a DLM, we can express the state and observation equations as follows:

$$x_t = Fx_{t-1} + Hu_t + \epsilon_{1,t} \tag{4.2.5}$$

$$y_t = Gx_t + \epsilon_{2,t}, \tag{4.2.6}$$

where again we have $\underline{\epsilon}_1 \sim \mathcal{N}(0, Q)$, $\underline{\epsilon}_2 \sim \mathcal{N}(0, R)$ for some Q and R. F, G, and H are constant or time-varying (p x p) matrices that preserve the updating structure of

linear functions f and g. Q and R are time-invariant (p x p) matrices that specify the covariance structure of the noise between the p different states and observations respectively. Estimating the underlying states $\mathbf{x}$ of a linear Gaussian state-space model is straightforward thanks to the celebrated Kalman filter which we now describe.

4.2.3 The Kalman filter

This is used to estimate the underlying states $\mathbf{x}$ of a linear Gaussian state-space model, such as that represented in (4.2.5)-(4.2.6). This was devised by Rudolf Kalman (1960) and it can be shown that the Kalman filter estimate of the state is optimal in terms of minimising the mean squared error. The Kalman filter is effectively a succession of Gaussian distributions, the mean and covariance of which are calculated at each recursive update. These Gaussian distributions update in a Bayesian system consisting of a two-step process; sometimes known as the *prediction* and *correction* steps. We use similar notation to Petris et al. (2009), where the reader is directed for proofs.

For the linear Gaussian state-space model of (4.2.5)-(4.2.6), in which the matrices F, G, H, Q, R are time-invariant, let $X_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$. Then the following statements are true.

(a) *Prediction step:* The predictive distribution $X_t|y_{1:t-1} \sim \mathcal{N}(a_t, V_t)$ where

$$a_t = \mathbb{E}(X_t|y_{1:t-1}) = Fm_{t-1} + Hu_t, \tag{4.2.7}$$

$$V_t = \mathrm{Var}(X_t|y_{1:t-1}) = FC_{t-1}F^T + Q. \tag{4.2.8}$$

(b) Predictive distribution $Y_t|y_{1:t-1} \sim \mathcal{N}(f_t, W_t)$ where

$$f_t = \mathbb{E}(Y_t|y_{1:t-1}) = Ga_t, \tag{4.2.9}$$

$$W_t = \text{Var}(Y_t|y_{1:t-1}) = GV_tG^T + R. \tag{4.2.10}$$

(c) *Correction step:* Filtering distribution $X_t|y_{1:t} \sim \mathcal{N}(m_t, C_t)$ where

$$m_t = \mathbb{E}(X_t|y_{1:t}) = a_t + V_tG^TW_t^{-1}e_t = a_t + K_te_t, \tag{4.2.11}$$

$$C_t = \text{Var}(X_t|y_{1:t}) = V_t - V_tG^TW_t^{-1}GV_t, \tag{4.2.12}$$

where $e_t = Y_t - f_t$ is the forecast error, and $K_t = V_tG^TW_t^{-1}$ is known as the Kalman gain.

The prediction step uses information from matrices F, H and Q in the state update equation (4.2.5). The correction step however uses information from matrices G and R from the observation update equation (4.2.6), but is also influenced by the observations $\mathbf{y}$. The extent of this influence is determined by the Kalman gain. We see this effect in (4.2.11), which as Petris et al. (2009) point out, shows that "filter mean m_t equals the prediction mean a_t plus a correction depending on how much the new observation differs from its prediction." Peterson (1976) provides a good description of the Kalman gain's effect on the updating process by highlighting two extremes. At each update, one option is to rely entirely on the updating process outlined in the underlying state-space model, totally ignoring the observations $\mathbf{y}$. Peterson refers to this as naive simulation (NS). In contrast to this, another option is to ignore the state-space model and rely totally on the observations; this is effectively OLS.

The Kalman gain is influenced by noise covariance matrices Q and R and more specifically, their ratio – sometimes referred to as the *signal to noise ratio*. Effectively if $Q < R$, the Kalman gain will cause the correction step of the Kalman filter to update nearer to NS – i.e. more faith is placed in the underlying state-space model than the observations. If $Q > R$ more faith is placed in the observations than the state-space model. See Petris et al. (2009) for a more thorough discussion of this effect. The Kalman gain determines the optimal location of the correction step by finding a point somewhere between NS and OLS that minimises $E[||x_t - m_t||^2]$ where x_t is the true underlying state at time t. This is demonstrated in Figure 4.2.2.

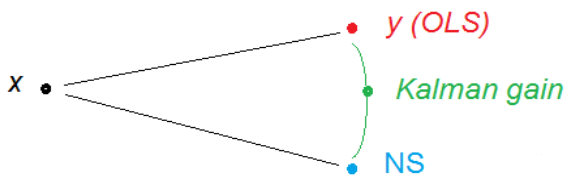

Figure 4.2.2: The Kalman gain's influence on the update.

With p sets of observations and states, F, G, V_t, W_t, C_t, K_t and covariance matrices Q and R are (p x p) matrices. The diagonal elements of Q and R represent the variance terms of the noise for each series. Hence for a multivariate state-space model, we may have different signal to noise ratios for each series.

Terms from the Kalman filter recursions can also be used to calculate the log-likelihood

at each update. If we have n random vectors $Y_1, ..., Y_n$ and their distributions depend on unknown parameter θ, let $p(y_1, ..., y_n; \theta)$ be the joint density of the observations for a particular value of θ. For a linear Gaussian state-space model we can express this joint density as

$$p(y_1, ..., y_n; \theta) = \prod_{t=1}^{n} p(y_t | y_{1:t-1}; \theta). \tag{4.2.13}$$

Since the densities on the RHS of (4.2.13) are Gaussian we can express the log-likelihood as

$$l(\theta) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \Sigma_{t=1}^{n} (\log |W_t| + e_t^T W_t^{-1} e_t), \tag{4.2.14}$$

where p is the number of series and n is the number of observations in each series. This representation was first devised by Schweppe (1965). The log-likelihood of (4.2.14) can be numerically maximised to find $\hat{\theta}$, the maximum likelihood estimate (MLE) of θ.

A number of packages exist for implementation of the Kalman filter in the R (2013) programming language. These are well documented by (Tusell, 2011). Among these are *dlm*, the R package that accompanies Petris et al. (2009) and *dse* (Gilbert, 2013). *dlm* provides a user-friendly way of creating dynamic linear models and implementing the Kalman filter, though one limitation is its inability to represent exogenous variables, unlike *dse* which can. Calculating the log-likelihood as in (4.2.14) involves the inversion of the matrix W_t. This can lead to numerical errors, especially when applying the Kalman filter to complex systems with large matrices. To circumvent this issue, an approximation to the log-likelihood devised by Anderson and Moore

(1979) can be used which avoids inverting W_t:

$$l(\theta) \approx -\frac{np}{2}\log(2\pi) - \frac{1}{2}\Sigma_{t=1}^{n}(\log|W_t| + 1). \tag{4.2.15}$$

This is also the default method for calculating the log-likelihood in the *dse* R package. We show in section 4.4, that using the likelihood in this form can have advantages as many computational issues are avoided.

4.2.4 Parameter estimation of state-space models

Now that we have an expression for the log-likelihood, in this section we evaluate some commonly used methods for parameter estimation of state-space models. We then turn our attention to the FIMLOF method in sections 4.2.5 and 4.2.6.

The EM algorithm is a popular algorithm for maximum likelihood estimation. It was introduced by Dempster et al. (1977) and early developments were made by Shumway and Stoffer (1982) and Watson and Engle (1983). The algorithm is an iterative process based on the assumption that at iteration $j+1$ there exists a parameter estimate θ_j from the previous iteration. Each iteration has two steps; an E-step (expectation) and an M-step (maximisation). The E-step calculates the expected value of the log-likelihood function, with respect to the conditional distribution of X given observations $y_{1:T}$ under the current estimate of the parameter θ_j. In the context of state-space modelling this is as follows:

$$Q(\theta_j, \theta) = \mathbb{E}[\log\, p(X_{1:T}, y_{1:T}|\theta)|y_{1:T}, \theta_j] = \int \log\, p(x_{1:T}, y_{1:T}|\theta)\; p(x_{1:T}|y_{1:T}, \theta_j)\; dx_{1:T}.$$

The M-step finds the parameter that maximises this expectation with respect to θ:

$$\theta_{j+1} = \arg \max_{\theta \in \Theta} Q(\theta_j, \theta).$$

The algorithm has been found to be particularly effective in cases where data are incomplete.

Although the algorithm usually has a fast convergence in its early stages, its rate of convergence near the maximum is often considerably slower than numerical maximisation (NM) techniques (Watson and Engle, 1983). Shumway and Stoffer (2006) also point out that the EM algorithm has generally been found slower than the use of some NM techniques such as quasi-Newton methods. It is also well known that the EM algorithm is only able to find the local optima, and as such the algorithm is not robust to the choice of starting values - see for example Barber (2012). Some authors, such as Macdonald (2014), have also questioned whether the additional computational effort of the EM algorithm is worthwhile and argued that it should not necessarily be the default approach over NM, with good numerical optimisers now freely available. Macdonald presents seven different models that demonstrate advantages of NM over the EM algorithm.

An alternative to maximum likelihood estimation is to use a Bayesian approach. For a thorough exposition of Bayesian estimation of state-space models, see Petris et al. (2009), sections 4.2 - 5.4. In a Bayesian approach, inference on $X_{0:T}$ and parameter

θ, given observations $y_{1:T}$, is expressed through their joint posterior density:

$$p(X_{0:T}, \theta | y_{1:T}) = p(X_{0:T} | \theta, y_{1:T})\ p(\theta | y_{1:T}). \tag{4.2.16}$$

This is based upon Bayes theorem (Bayes, 1763). A conjugate prior, when combined with the likelihood to form the posterior distribution in a Bayesian analysis, results in a posterior from the same family of distributions as the prior. The simplest technique using a Bayesian approach here would be to sequentially update using conjugate priors in a conjugate Bayesian analysis. This is possible for a few special cases where the posterior distribution of (4.2.16) can be expressed in closed form. For example, if the observational variance, denoted by R in (4.2.6), is assumed unknown and constant - with all other parameters known - this can be estimated using a conjugate Bayesian analysis based on gamma prior/posterior distributions for the precision parameter $\phi = 1/R$; see West and Harrison (1997), section 4.5. Petris et al. (2009), section 4.3.1, presents a similar approach for estimating a scale parameter, denoted by σ^2, when state and observational variances Q and R are both known to be multiples of σ^2.

In chapters 5 and 6, we attempt to estimate the parameters of state-space models representing highly nonlinear systems. In addition, our estimation process will not be restricted to variance parameters, but also parameters contained within matrices F, G and H of (4.2.5)-(4.2.6). We are not able to estimate these parameters using a conjugate Bayesian analysis, as the joint posterior distribution of (4.2.16) can not be expressed in closed form. Hence, this approach would not be suitable for these systems. In general, for Bayesian estimation of state-space models, Monte Carlo methods

are required to draw a sample from the posterior distribution of interest (Petris et al., 2009).

One such approach that is popular is to use a Gibbs sampler (Geman and Geman, 1984), which is a special case of Markov Chain Monte Carlo (MCMC). This involves draws in turn from the conditional distributions $p(\theta|X_{0:T}, y_{1:T})$ and $p(X_{0:T}|\theta, y_{1:T})$ to obtain samples from the joint posterior distribution. However, as we will see later in this chapter - and especially in chapter 5 where we progress to nonlinear models - the parametric forms of matrices F, G and H of (4.2.5)-(4.2.6), suggest that implementing a Gibbs sampler will not be possible without substantial restrictions on the model. This suggests that more general MCMC algorithms would be a more appropriate choice for a Bayesian approach to parameter estimation for our systems.

Alternative algorithms to MCMC are used when online inference is required. MCMC algorithms are of limited use here, since each time a new observation becomes available, an entirely new Markov chain must be simulated, which causes a linear increase in the computational cost. For online inference, a popular approach which also decreases computational cost compared to MCMC, is to use sequential Monte Carlo methods - more commonly known as particle filters, when applied to state space models. These are based on importance sampling techniques. See Gordon et al. (1993) for an example of a simple and popular particle filter. Online inference however, is beyond the scope of this thesis as we restrict ourselves to offline estimation. In addition, all of these algorithms, MCMC in particular, are computationally intensive and as such,

will not be pursued here.

We now turn our attention to the FIMLOF method. Section 4.2.5 describes the origins of the method and some early examples. Section 4.2.6 outlines its development over time and provides some modern examples of its application.

4.2.5 Early examples of FIMLOF

As shown in section 4.2.3, Schweppe (1965) was the first to represent the likelihood in the form of (4.2.14). This showed how to derive the likelihood for a linear Gaussian state-space model using terms from the Kalman filter recursions, in the context of engineering and control systems. This representation of the likelihood was used in Schweppe (1973) and Peterson & Schweppe (1975), to devise the FIMLOF method. Peterson & Schweppe (1975) apply FIMLOF to estimate parameters of SD models. A summary of the theory and key results from the thesis can be found in Peterson (1976). Using Schweppe's representation of the likelihood in (4.2.14), they use Powell's method, (Powell, 1964), to find the local optima of the likelihood surface. Powell's method does not require the objective function to be differentiable.

Peterson & Schweppe (1975) includes two simulation studies and an application to fuel demand data. There are few examples of FIMLOF simulation studies in the literature where known parameters are estimated and as such, we present details of those two studies in this section and attempt to highlight their limitations. The most important of these limitations is that due to the limited computational power that was available

at the time, both simulation studies only estimate a single set of parameters. In our simulation studies in sections 4.4 and 5.3.5, we perform Monte Carlo experiments to give a more thorough exposition of the FIMLOF algorithm.

Simulation study 1: The first study is a simple state-space model representing a single underlying state as an autoregressive process:

$$x_t = sx_{t-1} + \epsilon_{1,t} \tag{4.2.17}$$

$$y_t = x_t + \epsilon_{2,t}, \tag{4.2.18}$$

with $s = 0.75$ and $\underline{\epsilon}_1 \sim \mathcal{N}(0, q)$, $\underline{\epsilon}_2 \sim \mathcal{N}(0, r)$ with $q = r = 1$. The authors define a distinction between *structural* parameters such as s in (4.2.17) or those found in matrices F, G and H in (4.2.5)-(4.2.6); and *variance* parameters such as q and r above or those that form the diagonals of Q and R in (4.2.5)-(4.2.6). From here on we use the same terms to distinguish between the two types of parameters. In (4.2.17)-(4.2.18), three parameters, a single structural parameter s and two variance parameters q and r are estimated for simulated data of $n = 100$ and 1000. s is estimated well for both run lengths, but the variance parameters are estimated less well, especially with a run length of 100. The authors state that they have found this to be the case generally; maximum likelihood estimates of structural parameters are more accurate than variance parameters. Wider confidence intervals for variance parameters than structural parameters is further evidence of this.

Simulation study 2: The second simulation study involved a more complex realis-

tic model of a firm described in (Forrester, 1968a). The state-space model formed consisted of 9 state and 7 observation equations with 13 structural parameters (and no variance parameters) for estimation. The exact conditions of the experiment were chosen to match that of Senge (1974) that estimated parameters using OLS. Peterson & Schweppe (1975) showed that estimates using FIMLOF were a considerable improvement on Senge's OLS estimates and were able to estimate the 13 parameters to a high degree of accuracy. Peterson (1976) also includes details of this particular simulation study.

These two simulation studies were limited by the computational power available at the time. As such, there are some limitations to these studies. Firstly, as we mentioned at the start of this section, only one set of parameter estimates are given for each of the two studies. If these estimates are the global optima of the likelihood surface, then this is not a problem. In an optimisation problem, the choice of starting values (initial guesses for the parameters) that are input to the algorithm should not affect the output – in other words the output should be robust to the starting values. However, Peterson and Schweppe use the Powell method for optimisation which can only find the local optima (Vierhaus et al. 2014). As we show in section 4.4, when using FIMLOF, methods such as the Powell method are not robust to starting values. We provide full details of simulation studies in sections 4.4 and 5.3.5, giving a full exposition of all parameter estimates from multiple starting values. We also explain some of the difficulties involved in an attempt to assist the interested researcher in this area. For optimising the likelihood surface, instead of the Powell method we use the

Nelder-Mead simplex method (Nelder & Mead, 1965). Although this also finds the local optima and as such is not robust to starting values, in section 4.4 we show that this is an improvement over the Powell method, both in terms of speed and accuracy.

Another issue with Peterson and Schweppe's simulation studies are the variance parameters. These were estimated in the first study, but this is effectively a simple 'toy' model. The variance parameters were not estimated in the second study. This simplifies the problem, and the likelihood surface, considerably. When the variance parameters are known, the number of unknown parameters contained within the Kalman gain at each update is decreased considerably, especially for a complex (7 x 9) system with 7 observation equations and 9 state equations. Even if assuming zero values for the off-diagonal terms in the covariance matrices Q and R, there would still be 9 variance parameters in Q and 7 variance parameters in R forming the diagonals of these matrices. Eberlein and Wang (1985) highlight the importance of noise terms in FIMLOF as follows: "In order to implement a FIMLOF estimator, the way in which noise enters the system must be carefully specified. ... not only which equations the noise enters but what the characteristics of the noise are. Changing the specified noise characteristics can have rather profound effects on the resulting estimate. This is because the Kalman gain ... is strongly influenced by the characteristics of noise entering the equations." In our simulation studies in sections 4.4 and 5.3.5, all the variance parameters, the diagonals of covariance matrices Q and R, are estimated as parameters. This considerably increases the complexity of the model and therefore the likelihood surface, however this avoids making inaccurate assumptions regard-

ing the variance terms. Although Radzicki et al. (2004) also highlights the detailed knowledge of the system's noise that is required for FIMLOF, in the modern applications of FIMLOF discussed in section 4.2.6, the covariance matrices Q and R are given relatively little attention. In contrast, our simulation studies demonstrate the importance of the terms within these matrices.

Peterson and Schweppe (1975) then applied FIMLOF to a SD model for fuel demand in the United States. The historical data was cross-sectional; for each of the 49 states time series of 5 observations were available – all believed to have same underlying structural and variance model parameters. The resulting state-space model had 3 state and 3 observation equations. The state equations were linear, though the observation equations were nonlinear. 20 parameters were estimated; 14 structural and 6 variance parameters. FIMLOF parameter estimates were compared to the weighted least squares (WLS) estimates of Boughman & Joskow (1975) by examining the log-likelihood values. The FIMLOF estimates were found to improve on the WLS estimates by more than a factor of 2 in terms of the log-likelihood. However there were no comparisons of the SD model output (using the estimated parameters) with the historical data. Therefore no indication was provided on the performance of the parameter estimates in terms of the resulting prediction error.

Instead of calculating prediction error, a series of tests were proposed by Peterson & Schweppe for validation of the state-space model based on the estimated parameters. In these tests, the estimated parameters were used in the Kalman filter and

residuals inspected. It was argued that when the parameters used in the Kalman filter represent the 'true' model, the normalised predicted residuals of the filter should be a white process, with constant unit variance and a Gaussian distribution. It was argued that since these properties of the residual process are not used directly in maximising the log-likelihood, they provide an independent test of model validity. The tests were said to be sensitive to small errors in the model specification and the authors state that even when the model parameters selected are from the global maximum of the likelihood surface, the residuals are not guaranteed to pass. The tests also included use of Durbin-Watson statistics and tests on the correlation matrices. The final state-space model for the fuel demand data performed well in these tests. Any correlations in the residuals were shown to be caused by the cross-sectional nature of the data.

For the fuel demand model variance parameters were not estimated outright. Instead they were assumed proportional to the initial values of the relevant states or observations. That is, the variance of the underlying state i that is the diagonal term (i, i) of covariance matrix Q was assumed proportional to the initial value of state i, $x_{0,i}$: $\sigma^2_{Q,i} = \alpha_i \, x_{0,i}$, for $i = 1, 2, 3$. Similarly, the variance of observation j that is the diagonal term (j, j) of R was assumed proportional to the initial value of observation j, $y_{0,j}$: $\sigma^2_{R,j} = \beta_j \, y_{0,j}$, for $j = 1, 2, 3$. The 6 constants of proportionality $\underline{\alpha}$ and $\underline{\beta}$ were estimated as unknown parameters. Initial values of these were found by visual inspection of sample standard deviations of the data. This can be seen as an attempt to inform the model on how to scale the variance parameters, when searching the parameter space.

4.2.6 Modern applications of FIMLOF

Although there are additional early examples of the application of FIMLOF, such as Moore and Schweppe (1973) and Mehra and Tyler (1973), which apply FIMLOF to nuclear power plants and engineering systems respectively, there are relatively few examples of FIMLOF's use over the 40 or so years since its conception. After showing such early promise, it is perhaps difficult to understand why FIMLOF (and also MRO) did not become more widely used. Perhaps the main reason for this was that at the time (and maybe right up to present day), the majority of SD modellers belonged to the classical school of thought. Not only did they attach little importance to the estimation of model parameters, they also distrusted use of quantitative data, not only in this way, but in SD modelling in general (Graham, 2002).

Another reason likely concerns the computational intensity of the methods; both for FIMLOF and MRO. Although much less of an issue nowadays, in the 1970's where computer power was considerably less, this was a major issue. In the estimation of SD model parameters, the importance of computational power, and the difficulty of the task overall, is demonstrated well in Dangerfield and Roberts (1996) – even with statistically rigorous methods such as FIMLOF and MRO. The authors provide a helpful guide for overcoming many of the practical issues and also highlight the difficulty of the problem with a simple example. A 30 parameter model in which each parameter can take a range of 10 discrete values (which is a very conservative estimate) has 10^{30} parameter combinations. The authors point out that a computer calculating 1 million

of these per second would take 3.17 x 10^6 years to work out all combinations. For a complex, nonlinear system with a continuous parameter space, it is not difficult to imagine the enormous size of the parameter space and consequently the difficulty of the task in hand.

It is perhaps not surprising that there have been attempts to simplify this problem to make the task more achievable. Eberlein (1986) presents two such approaches. The first is to approximate the Kalman gain by estimating the state and observational covariance matrices (Q and R). Rather than optimising the likelihood, the Kalman filter, with the estimated covariance matrices and Kalman gain is run on the data. Errors are inspected and optimisation involves minimising the square error loss. The second method involves breaking a complex SD model up into simpler sub-models in a process termed 'sectorization' which simplifies the problem. Although these are shown to be successful on a simple inflation model, there are few additional examples of these methods being applied.

There are some notable examples of more modern applications of FIMLOF. The popular simulation software Vensim (2010), has a feature that enables FIMLOF to be applied to historical data. As we explained in 4.2.3, the Kalman filter gives an optimal estimate of the underlying state for a linear Gaussian system. However, when the system is nonlinear the optimality no longer holds (Durbin and Koopman, 2001) and Schweppe's representation of the log-likelihood in (4.2.14) becomes an approximation (Peterson, 1976). Despite this, Peterson & Schweppe (1975) and Peterson

(1976) applied FIMLOF to nonlinear systems and used the standard Kalman filter in 4.2.3 to estimate the underlying state of the system. For nonlinear systems, Vensim uses the extended Kalman filter (Jazwinski, 1970; Bar-Shalom et al., 2001) instead of the standard Kalman filter and this has been shown to be a more accurate approximation. In chapter 5 we introduce an alternative Kalman filtering technique that has been shown to be more accurate than the extended Kalman filter for estimating the state of nonlinear systems. A commonly highlighted weakness of FIMLOF is the requirement that the system is linear (Dangerfield & Roberts, 1996; Radzicki et al., 2004). We show in chapter 5 that when using an appropriate Kalman filter this requirement is now less of an issue. The remainder of this section is dedicated to two examples of recent applications of FIMLOF to complex, nonlinear systems. These are of interest for comparisons with the BT system.

In the first example Ryzhenkov (2002) applies FIMLOF (within Vensim) to a complex nonlinear SD model for economic data. The economic system under inspection is (7 x 8) with 7 observation equations and 8 (nonlinear) state equations. The total number of observations, N = 34. Although it is not entirely clear exactly how many structural parameters are estimated, there appear to be approximately 20. In addition 15 variance parameters are also estimated. High measurement errors are assumed, since there is said to be much uncertainty about the initial values of the state of system, $\underline{x}_0$. Some of these initial states and covariances are also estimated as additional parameters. Powell's method is used for optimisation.

The state-space model formed from the SD model is referred to as the stochastic dynamic model. When assessing the performance of FIMLOF in estimating the model parameters, the focus is on comparing output of the stochastic dynamic model with the historical time series data. This is *not* the equivalent of using the estimated structural parameters in the deterministic SD model because of the noise terms in the stochastic dynamic model. The author admits that when comparing this model to historical data, the state and observational noise are used as exogenous inputs. This suggests that the state and observational noise series are obtained by running the extended Kalman filter on the historical data, using the estimated structural and variance parameters to obtain the noise terms. These noise terms are then used as exogenous inputs for the stochastic dynamic model. The stochastic dynamic model is a good match to historical data, with low prediction error for most of the series. However, using noise terms in this way to improve the match between simulations and historical data could be argued in some ways as cheating. The purpose of estimating parameters of a SD model is to enable the model, with the use of *only* the estimated structural parameters and any exogenous variables, to represent the behaviour of historical time series as accurately as possible. This is how we assess the performance of our parameter estimates for the historical BT regional time series in chapter 6.

In the second example, Radzicki et al. (2004) formed a SD model based on Harrod's (1939) economic growth model. He added new parts to the model to represent managerial expectations. This nonlinear model consisted of 5 stocks with 9 parameters and 4 initial values to be estimated. There was no mention of any variance

parameters. It is not clear whether the variance terms were also estimated as parameters or if fixed values were used. FIMLOF was applied to historical economic data to estimate the parameters and initial values. When using these estimated parameters in the SD model, the output resulted in an impressive match to historical time series data. However, in terms of prior understanding of economic systems, some parameter estimates were deemed to be unreasonable, (i.e. when expected to be larger than others they were smaller – and vice versa). However, the author admits that "experimentation with the model uncovered FIMLOF runs that yielded much more reasonable parameter estimates", but these corresponded to poorer fits when matching SD model output to historical time series. However it is not clear exactly what this "experimentation" was. 95% confidence intervals are provided for some, but not all estimated parameters. As Radzicki et al. (2004) explain, FIMLOF was "not able to zero-in on a particular value due to integration and round-off errors encountered during simulation ... [these errors] are unavoidable when simulating continuous-time dynamical systems on a digital computer".

The author states that system dynamicists don't take the view of defining a model as valid/invalid and instead subject the model to a large number of tests. Confidence in the model is said to increase with the number of tests passed. There is no mention of Peterson's validation tests using the Kalman filter residuals. Instead, a helpful detailed list of the tests that a modeller should use to validate a systems dynamics model is presented. Similar references are given for these validation tests as we provide in chapter 3. One category of these tests, termed *tests of overall model behaviour*, effec-

tively describe how well output from the model matches historical data. The modified Harrod SD model is subjected to this category of tests. It is argued that the model passes all these tests since it "nicely mimics some of the observed behaviour" from the real system. However doubts are raised as to whether the "model's parameter values are consistent with the relevant descriptive and numerical knowledge of the actual system". This is a validation test included within the category of *structural assessment tests*. This test could have been passed with more "reasonable" parameter estimates, but the author used the parameter estimates from FIMLOF that resulted in better time series fits. This demonstrates an interesting conflict in this application of FIMLOF. Should the modeller use more realistic parameter estimates (in terms of prior system knowledge) that do not fit the historical data as well? Or should the 'best' parameters be chosen in terms of fitting the data? Using FIMLOF forces the user towards the latter approach. In situations where such a choice is required, the model is going to fail at least some validation tests.

Both Ryzhenkov (2002) and Radzicki et al. (2004) present only one set of parameter estimates and both use the Vensim simulation package for implementing FIMLOF. The Vensim help files state that the default method for optimisation is the Powell method, but do not give details of any other method(s). Ryzhenkov includes the output file from Vensim which explicitly states that the Powell method was used for optimisation. Although it is not stated explicitly whether implementing FIMLOF in Vensim is robust to the starting values or not, the Vensim help files provide a clue. They state that when multiple starting values for the optimisation are selected

as an option, the optimisation will continue and not stop until the user cancels it. At this point, the optimiser "shuts down cleanly, writing out the best values found" to the output file. This implies that multiple sets of estimates are found from the optimisation. In other words, that the method in Vensim is *not* robust to starting values. Indeed Ryzhenkov (2002) used multiple starting values that were set at random for each parameter in the optimisation algorithm. He highlights the difficulty of estimating the parameters and explains this process when implementing in Vensim as follows: "Typically maximising the log likelihood function by a hill climbing algorithm with random multiple starts cannot be finished". He argues that when terminating the process prematurely, the modeller relies "not only on logic but intuition as well". He adds, "therefore to find a genuine optimal solution is hardly possible in practice"; highlighting the difficulty of implementing the approach as the Powell method can only find the local optima. The simulation studies we present in sections 4.4 and 5.3.5 are also not robust to the starting values. However we provide full details of the parameter estimates produced to give a more thorough exposition of the method, rather than providing a single set of estimates.

We introduce the full version of the BT hydraulics model in chapter 5 and estimate its parameters to match historical data in chapter 6. We argue that this model is considerably more complex and nonlinear than the models discussed in this section.

4.3 Applying FIMLOF to hydraulics model 1

In this section we explain the details of how the FIMLOF method was applied to hydraulics model 1, the simplified version of the hydraulics model, to estimate its parameters. In section 4.3.1 we highlight the key aspects of hydraulics model 1 from the perspective of applying FIMLOF; the parameters to be estimated and the stocks of the model. In section 4.3.2 we lay the foundations of representing the structure of hydraulics model 1 as a state-space model by substituting out the auxiliary variables in the model to represent hydraulics model 1 as two Markovian difference equations. A state-space model is formed from these difference equations in section 4.3.3. We then outline the FIMLOF algorithm in section 4.3.4.

4.3.1 Hydraulics model 1 parameters

A reminder of the stock and flow diagram of this model is provided in Figure 4.3.1. Hydraulics model 1 is effectively a deterministic function of the model parameters $\underline{\theta}$ and the exogenous time series U; the job arrivals (or 'new tasks' in Figure 4.3.1). Hence the model simulates output time series $\underline{y}$ according to the following structure:

$$\underline{\theta}, U \Rightarrow Hyd\left(\underline{\theta}, U\right) \Rightarrow \underline{y}$$

The job arrivals series is classed as an exogenous variable because the behaviour of this series is not itself of interest to us – rather it is the effect of this series on the system. As such, U can be considered to be an additional input to the model. $\underline{\theta}$ consists of 3 model parameters as follows: $\underline{\theta} = (\theta_d, \theta_w, \theta_t)$. Each of these is time-invariant in hydraulics model 1 and hence these are the (structural) model parameters that we

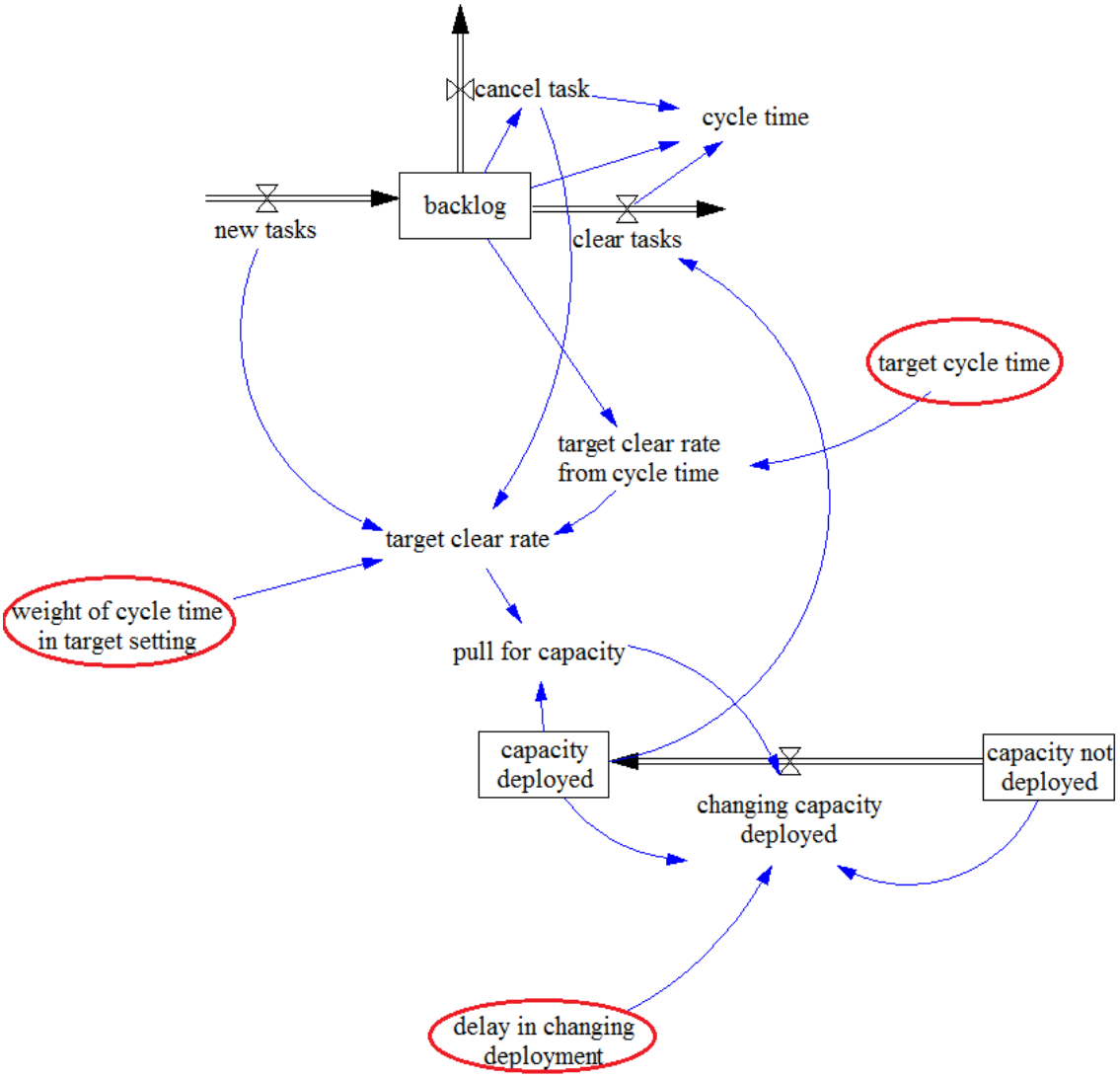

Figure 4.3.1: Stock and flow diagram of hydraulics model 1.

wish to estimate. Each of these three model parameters is described below and circled in red in Figure 4.3.1.

- θ_d = 'delay in changing capacity' in Figure 4.3.1. When the system experiences an increase in demand and more engineers are required, this is the delay (in days) associated with transferring workforce due to the time taken for rescheduling work and the necessary meetings etc.. In Figure 4.3.1, this transfer is represented

by moving engineers from 'capacity not deployed' to 'capacity deployed'.

- θ_w = 'weight of cycle time in target setting' in Figure 4.3.1. The 'target clear rate' in the model (the target for how fast jobs are completed) is set using a weighting of the job arrivals and the cycle time (or wait time). θ_w determines the weighting that is attached to the cycle time.

- θ_t = 'target cycle time' in Figure 4.3.1. This is the target time (in days) that BT aim to complete all repair jobs within.

In SD terms, the model output $\underline{y}$ corresponds to the stocks of the model. See chapter 2 for a full explanation of SD modelling. For hydraulics model 1, $\underline{y}$ consists of two daily time series; the 'backlog' and the 'capacity deployed'. The backlog is the length of the job queue. This is a key performance metric in the model. The capacity deployed corresponds to the number of people deployed on jobs each day. This can be seen as a measure of the system's response. In Figure 4.3.1 the third stock, 'capacity not deployed', is simply a constant minus the capacity deployed stock and therefore for the sake of simplicity we are not interested in the output of this stock.

4.3.2 Deriving hydraulics model 1 difference equations

In order to implement the FIMLOF method we must represent the structure of hydraulics model 1 as a state-space model. The first step towards this is to represent hydraulics model 1 as Markovian difference equations. The previous contributions to the literature described in sections 4.2.5 and 4.2.6, that attempt to use FIMLOF to estimate the parameters of SD models, provided *either* the underlying SD model

equations or the state-space model - but not both. In this section we provide a step by step guide for the conversion of a SD model into Markovian difference equations. This process involves the following:

1. Grouping the SD model variables into stocks, exogenous variables and auxiliary variables.

2. Determining the order in which auxiliary variables update in the SD model.

3. Performing substitutions to form Markovian difference equations.

In section 4.3.3, a state-space model is constructed from these difference equations. This process of the formulation of a state-space model from a SD model is a novel contribution to the area. It is worth highlighting that the technique described in this section can be applied to *any* SD model.

Like any SD model, variables within the hydraulics model are continuous, with differential equations capturing the relationships between them (Sterman, 2000). In practice however, during computation these differential equations are discretised to form difference equations. These difference equations update regularly in time, with a suitably small interval dt selected by the modeller.

The variables within SD models can be divided into 3 categories: stocks, auxiliary variables and exogenous variables. In Figure 4.3.1 we have already highlighted the stocks, exogenous variable and the model parameters; auxiliary variables are simply

any other variables. When a SD model updates, it is the stocks that drive the updating process. To understand this, we must examine the difference equations at each dt update. We examine the updating structure of a model with $dt = 1$ that updates from time $t-1$ to time t. At time t, the stocks are the first variables in the model to update, using the previous values of stocks and auxiliary variables at $t-1$. The parameters $\underline{\theta}$ and the new value (at time t) of the exogenous variable(s) may also be used as follows:

$$\text{stocks}(t) = \text{Hydraulics (stocks}(t-1), \text{ auxiliary variables}(t-1), \text{ U}(t), \underline{\theta}).$$

For example in Figure 4.3.1, the backlog and the capacity deployed are the first variables to update for time t. After these stocks have updated at time t, the auxiliary variables are then updated *using the updated value of the stocks*. The order of the updating process amongst the auxiliary variables is determined by how close they lie to the stocks in the model structure – shown in the stock and flow diagram. For example, in Figure 4.3.1, 'target clear rate from cycle time' updates before 'target clear rate' since it is closer to the backlog stock. At the conclusion of the updating process for time t, all auxiliary variables will have updated either from the updated stocks at time t, or from other auxiliary variables that have updated first at time t as they are closer to the stocks in the model structure:

$$\text{auxiliary variables}(t) = \text{Hyd(stocks}(t), \text{ auxiliary variables}(t), \text{ U}(t), \underline{\theta}).$$

From this updating process, we can observe that the auxiliary variables can be seen

to act as intermediate variables between stocks – assisting in their update at each dt.

Clearly the difference equations for the stocks and auxiliary variables have the Markov property. However, there is a simpler, more concise way of representing hydraulics model 1. The nature of this updating process means that by substitution, it is possible to simplify the representation of the difference equations of a SD model, removing the auxiliary variables entirely. This reorganising process results in each stock at time t being represented only in terms of other stocks at $t-1$, some function $g()$ of the model parameters $\underline{\theta}$ and the exogenous variable(s) U as follows:

$$\text{stocks}(t) = \text{Hydraulics}\ (\text{stocks}(t-1),\ \text{U}(t),\ g(\underline{\theta})).$$

In order to obtain this more concise representation of hydraulics model 1, we present the full difference equations for the model and proceed to show how these substitutions are implemented. Table 4.3.1 gives the notation used for the hydraulics model 1 variables. The equations of the stocks of hydraulics model 1 take the following form:

$$b_t = b_{t-1} + n_{t-1} - c_{t-1} \tag{4.3.1}$$

$$c_t = c_{t-1} + cc_{t-1}. \tag{4.3.2}$$

(4.3.1) requires no substitutions, since the backlog is already expressed in terms of only stocks and the exogenous variable, the new tasks. This is not the case for (4.3.2), since capacity deployed is expressed in terms of an auxiliary variable; changing capacity

Symbol	hydraulics model 1 parameter
b	backlog
c	capacity deployed
n	new tasks
cn	capacity not deployed
cm	maximum capacity deployed
tct	target clear rate from cycle time
cl	cleared jobs
ct	cycle time
tcr	target clear rate
pf	pull factor
cc	changing capacity deployed

Table 4.3.1: Notation for hydraulics model 1 variables.

deployed. The full set of difference equations for hydraulics model 1 are shown below.

$$
\begin{aligned}
cn_t &= cm - c_t \\
tct_t &= \frac{b_t}{\theta_t} \\
cl_t &= c_t \\
ct_t &= \frac{b_t}{cl_t} \\
tcr_t &= \theta_w(tct_t) + (1-\theta_w)(n_t) \\
pf_t &= \frac{tcr_t}{c_t} \\
cc_t &= \frac{c_t}{\theta_d}(pf_t - 1)
\end{aligned}
$$

In order to obtain the difference equation for capacity deployed in the required form, we must perform substitutions as follows:

$$
\begin{aligned}
cc_t &= \frac{c_t}{\theta_d}(pf_t - 1) && (4.3.3)\\
&= \frac{c_t}{\theta_d}[(\frac{tcr_t}{c_t}) - 1] && (4.3.4)\\
&= \frac{c_t}{\theta_d}[(\frac{\theta_w(tct_t) + (1-\theta_w)(n_t)}{c_t}) - 1] && (4.3.5)\\
&= \frac{c_t}{\theta_d}[(\frac{\theta_w(\frac{b_t}{\theta_t}) + (1-\theta_w)(n_t)}{c_t}) - 1] && (4.3.6)
\end{aligned}
$$

(4.3.3)-(4.3.6) are obtained by substituting in the terms for pf_t, tcr_t and tct_t respectively. (4.3.6) now represents the changing capacity deployed variable in the required form; that is in terms of only stocks, the model parameters $\underline{\theta}$ and the exogenous variable. We can use this representation to obtain the required form for (4.3.2) by substituting in cc_t and rearranging as follows:

$$
\begin{aligned}
c_t &= c_{t-1} + cc_t\\
&= c_{t-1} + \frac{c_{t-1}}{\theta_d}[(\frac{\theta_w(\frac{b_{t-1}}{\theta_t}) + (1-\theta_w)(n_t)}{c_{t-1}}) - 1]\\
&= c_{t-1} + \frac{c_{t-1}}{\theta_d}(\frac{\theta_w(\frac{b_{t-1}}{\theta_t}) + (1-\theta_w)(n_t)}{c_{t-1}}) - \frac{c_{t-1}}{\theta_d}\\
&= c_{t-1} + \frac{\theta_w(\frac{b_{t-1}}{\theta_t}) + (1-\theta_w)(n_t)}{\theta_d} - \frac{c_{t-1}}{\theta_d}\\
&= (\frac{\theta_w}{\theta_d\theta_t})\, b_{t-1} + (1 - \frac{1}{\theta_d})\, c_{t-1} + (\frac{1-\theta_w}{\theta_d})\, n_t\\
&= \phi_1\, b_{t-1} + \phi_2\, c_{t-1} + \phi_3\, n_t.
\end{aligned}
$$

ϕ_1, ϕ_2 and ϕ_3 are defined as follows:

$$\phi_1 = g_1(\underline{\theta}) = \frac{\theta_w}{\theta_d \theta_t} \tag{4.3.7}$$

$$\phi_2 = g_2(\underline{\theta}) = 1 - \frac{1}{\theta_d} \tag{4.3.8}$$

$$\phi_3 = g_3(\underline{\theta}) = \frac{1 - \theta_w}{\theta_d}. \tag{4.3.9}$$

Hence, substituting out the auxiliary variables of hydraulics model 1 in this way produces the following difference equations:

backlog(t+1) = backlog(t) - capacity deployed(t) + job arrivals(t)

cap deployed(t+1) = ϕ_1 backlog(t) + ϕ_2 cap deployed1(t) + ϕ_3 job arrivals(t).

If we take dt to equal 1, i.e. the model updates once for each day, the equation for the backlog is intuitive; today's queue is equal to yesterday's queue minus the people deployed (and hence the jobs completed – since in this simple model each engineer completes one job per day) plus the new job arrivals. The equation for the capacity deployed is more complex and is determined by a set of parameters ϕ_1, ϕ_2 and ϕ_3 which are simple functions of $\underline{\theta}$. The difference equations formed above, along with the $\underline{\phi}$ parameters, preserve the exact structure of hydraulics model 1. By substituting out the auxiliary variables we have simplified the representation considerably, yet preserved the Markov property of the equations. Representing the system in this way provides us with a framework for constructing a state-space model in the next section.

4.3.3 Deriving the state-space model equations

The difference equations formed in the previous section are a concise way of representing hydraulics model 1's structure. To implement FIMLOF we need to construct a state-space model from these equations.

In chapter 2 we explained that hydraulics model 1 is a much simplified version of the full model and represents a single feedback loop only. However, if hydraulics model 1 was a more complex model representing a real system, it would be hoped that these difference equations would represent the structure of the real system well. In practice however, even for a very accurate model, this would not be an exact representation; some noise is likely to be present. Under this assumption, we obtain the following equations by adding noise terms to the difference equations of hydraulics model 1:

$$b_{t+1} = b_t - c_t + n_t + v_{1,t} \tag{4.3.10}$$

$$c_{t+1} = \phi_1 b_t + \phi_2 c_t + \phi_3 n_t + v_{2,t} \tag{4.3.11}$$

where b represents the backlog, c the capacity deployed and n the new job arrivals. $\underline{v}_1$ and $\underline{v}_2$, the noise associated with updating the backlog and capacity deployed respectively, represent any inaccuracies in the fit of the hydraulics model difference equations to the real system.

In state-space terms, we assume that these relationships represent the underlying state of the system; a state that is not directly observable. In other words, (4.3.10)-

(4.3.11) are the state equations of our state-space model. To implement FIMLOF, we further assume that the noise is Gaussian and hence $\underline{v}_1 \sim N(0, \sigma^2_{Q_1})$, $\underline{v}_2 \sim N(0, \sigma^2_{Q_2})$.

We can complete the state-space model by defining the observation equations. If we assume that the observable data has some additional observational noise (for example caused by measurement error), then the observable data can be represented as follows:

$$b^*_t = b_t + w_{1,t} \tag{4.3.12}$$

$$c^*_t = c_t + w_{2,t}, \tag{4.3.13}$$

where b^* is the observable data for the backlog, c^* is the observable data for capacity deployed and we again assume Gaussian noise as follows: $\underline{w}_1 \sim N(0, \sigma^2_{R_1})$, $\underline{w}_2 \sim N(0, \sigma^2_{R_2})$. It is also worth noting that (4.3.10)-(4.3.13) are all linear equations. Therefore, under this representation for (4.3.10)-(4.3.13), we now have a linear Gaussian state-space model, or dynamic linear model (DLM). As we explained in section 4.2.3, the assumption of Gaussian noise allows the use of the Kalman filter to estimate the underlying states b_t, c_t from observations at time t, b^*_t and c^*_t. The linear nature of (4.3.10)-(4.3.13) also mean that the Kalman filter will give optimal estimates of the states in terms of mean squared error and therefore the representation of the log-likelihood in (4.2.14) will be exact. We can represent the DLM in (4.3.10)-(4.3.13) in matrix form as follows:

$$\underline{y}_t = G\underline{x}_t + \underline{w}_t \tag{4.3.14}$$

$$\underline{x}_t = F\underline{x}_{t-1} + H\underline{u}_t + \underline{v}_t, \tag{4.3.15}$$

where

$$\underline{y}_t = \begin{bmatrix} b_t^* \\ c_t^* \end{bmatrix}, \ \underline{x}_t = \begin{bmatrix} b_t \\ c_t \end{bmatrix}, \ G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \underline{w}_t \sim N(0, R), \ R = \begin{bmatrix} \sigma_{R_1}^2 & 0 \\ 0 & \sigma_{R_2}^2 \end{bmatrix},$$

and

$$\underline{u}_t = \begin{bmatrix} n_t \\ n_t \end{bmatrix}, F = \begin{bmatrix} 1 & -1 \\ \phi_1 & \phi_2 \end{bmatrix}, \ H = \begin{bmatrix} 1 & 0 \\ 0 & \phi_3 \end{bmatrix}, \ \underline{v}_t \sim N(0, Q), \ Q = \begin{bmatrix} \sigma_{Q_1}^2 & 0 \\ 0 & \sigma_{Q_2}^2 \end{bmatrix}.$$

For each update at time t, the vectors $\underline{y}_t$, $\underline{u}_t$ and matrices F, G, H, Q and R are used in the Kalman filter equations (4.2.7)-(4.2.12) to estimate the latent states $\underline{x}_t$.

The off-diagonal terms of matrices Q and R are assumed to be zero. Assuming a zero covariance structure in both the state and observational noise for backlog and capacity deployed is not a problem in the simulation studies in section 4.4 and chapter 5, where we know that the simulated data possesses this structure. This structure simplifies the problem, and the number of parameters that are to be estimated. However, in chapter 6 when we progress to BT time series data, this assumption must be re-evaluated.

4.3.4 Modified FIMLOF algorithm 1

Now that we have constructed the state-space model using the structure of hydraulics model 1, we now have the necessary terms, associated with (4.3.14)-(4.3.15), to present

an algorithm for implementing the FIMLOF method.

The matrices Q and R, associated with (4.3.14)-(4.3.15), represent the system and observational covariances respectively. The diagonal elements of these; $\sigma^2_{Q_1}$, $\sigma^2_{Q_2}$ and $\sigma^2_{R_1}$, $\sigma^2_{R_2}$, correspond to the variance of noise terms v_1, v_2 and w_1, w_2 respectively from (4.3.10)-(4.3.13). In the simulation study in section 4.4, noisy data will be artificially simulated within both the state and observation equations, so that $\sigma^2_{Q_1}$, $\sigma^2_{Q_2}$, $\sigma^2_{R_1}$ and $\sigma^2_{R_2}$ will all be non zero. Therefore these terms will be treated as additional parameters that will also be estimated.

The algorithm that we present in this section also includes two modifications to the standard FIMLOF algorithm devised in Schweppe (1973) and Peterson & Schweppe (1975). These were discovered through experimentation and were found to improve performance in terms of the accuracy of estimated parameters. The first of these is to use the approximation of the log-likelihood in (4.2.15) instead of the exact representation in (4.2.14). The second of these is to use the Nelder-Mead simplex method (Nelder and Mead, 1964) rather than Powell's method for the optimisation. The results demonstrating the improved performance of these two modifications are presented in section 4.4.6.

For noisy observable time series data $\underline{y}$ for the backlog and capacity deployed, we estimate the structural parameters $\underline{\theta} = (\theta_d, \theta_w, \theta_t)$ and the variance parameters $\underline{\sigma}^2 = (\sigma^2_{Q_1}, \sigma^2_{Q_2}, \sigma^2_{R_1}, \sigma^2_{R_2})$ of the state-space model in (4.3.14)-(4.3.15) using the algorithm

below. From here on we refer to this as algorithm 1.

1. The Nelder-Mead simplex method selects a candidate set of parameters $(\underline{\theta}, \underline{\sigma}^2)$, (or in the first iteration the user selects suitable starting values $\underline{\theta} = \underline{\theta}_0$ and $\underline{\sigma}^2 = \underline{\sigma}_0^2$).

2. $\underline{\phi} = \underline{g}(\underline{\theta})$ is calculated using (4.3.7)-(4.3.9).

3. Kalman filter proceeds along time series $\underline{y}$ and for each update at time t:

   - The $\underline{\phi}$ parameters in matrices F and H and the variance parameters $\underline{\sigma}^2$ influence the Kalman filter's estimate of latent states $\underline{x}_t$.
   - Terms from these calculations are used to calculate the log-likelihood, using the approximation of (4.2.15), for the candidate set of parameters, $\log(L(\underline{\theta}, \underline{\sigma}^2), t)$.

4. After the Kalman filter has calculated the log-likelihood for each t, an overall log-likelihood for the time series is calculated by summing the individual log-likelihood terms: $\log(L(\underline{\theta}, \underline{\sigma}^2)) = \Sigma_{t=1}^{T} \log(L(\underline{\theta}, \underline{\sigma}^2, t))$.

5. The optimisation algorithm repeats the process until a local optimum is found and maximum likelihood estimates $\hat{\underline{\phi}}$ and $\hat{\underline{\sigma}}^2$ are determined.

Having obtained the maximum likelihood estimates (MLEs) $\hat{\underline{\phi}}$, it is straightforward to calculate the MLEs for structural parameters $\hat{\underline{\theta}} = (\hat{\theta}_d, \hat{\theta}_w, \hat{\theta}_t)$. We simply use the inverse of the relevant $g()$ function in (4.3.7)-(4.3.9). This is possible since the maximum likelihood estimates are invariant to such reparameterisations (Pawitan, 2001).

The algorithm above explains how we get from the noisy data $\underline{y}$ to the structural parameters $\hat{\underline{\theta}}$ for hydraulics model 1. The logic of this process can be summarised as follows:

$$\hat{\underline{\theta}} \Leftarrow g^{-1}(\hat{\underline{\phi}}) \Leftarrow \hat{\underline{\phi}} \Leftarrow FIMLOF \Leftarrow \underline{y}$$

In the simulation study in section 4.4, artificial noisy data $\underline{y}$ is simulated from hydraulics model 1 with known structural and variance parameters. This algorithm is used for the estimation of these parameters and its performance is assessed.

4.4 Simulation study 1

In this section we assess the performance of algorithm 1 in estimating the parameters of hydraulics model 1. The objectives of this study are explained in section 4.4.1. Section 4.4.2 presents the details of the experiments, while section 4.4.3 explains the details of the job arrivals used to simulate data from hydraulics model 1. Section 4.4.4 verifies the Gaussian assumptions of the simulated data that are required for Kalman filtering. Section 4.4.5 presents the results of the study and details the amount of noise added to the simulated data for each experiment. Section 4.4.6 describes some adjustments that were made to the standard FIMLOF algorithm that improved performance.

4.4.1 Objectives of study

In this section we present results of a simulation study aimed at assessing the performance of algorithm 1 in estimating the parameters of hydraulics model 1 from noisy simulated data. In chapter 6 we apply a modified version of this algorithm to estimate the hydraulics model parameters from historical BT time series data in order to calibrate the hydraulics model at a regional level. Although this modified algorithm is based on a more complex version of the hydraulics model, and uses an adjusted version of the Kalman filter for nonlinear systems, the underlying process is the same as in this study. Whilst it is highly unlikely that this BT data will be represented exactly by the relevant hydraulics model, successful estimation of parameters from noisy simulated data can be seen as a first step towards this application to BT data. It can also be seen as validation that, for a simplified model at least, it is possible to successfully estimate the parameters of this type of SD model.

The objectives of the simulation study are to determine the following:

- Performance of algorithm 1 in estimating the structural parameters $\underline{\theta}$ of noisy data from hydraulics model 1
- The effects of system noise ($\sigma^2_{Q_1}$, $\sigma^2_{Q_2}$) and observational noise ($\sigma^2_{R_1}$, $\sigma^2_{R_2}$) on the accuracy of the structural parameter estimates $\underline{\hat{\theta}}$

In sections 4.2.5 and 4.2.6 we highlighted the relatively little attention given to the variance parameters in previous examples of FIMLOF. This is in spite of a number of authors explaining the importance of modelling the noise appropriately when using

FIMLOF, see for example Eberlein and Wang, (1986). In chapter 6 it is highly likely that our model will not be an exact representation of each regional system, i.e. some noise will be present. It will thus be useful at this stage to have some idea of the effects that the different types of noise have on the accuracy of our estimates. There is also the question of a noise threshold, i.e. whether there is some limit when noise is added beyond which parameter estimation becomes impractical.

4.4.2 Experiment details

For each dataset, 7 parameters in total will be estimated; the 3 structural hydraulics model parameters $\underline{\theta} = (\theta_d, \theta_w, \theta_t)$ and the 4 variance parameters $\sigma^2_{Q_1}$, $\sigma^2_{Q_2}$ and $\sigma^2_{R_1}$, $\sigma^2_{R_2}$ – though it is the structural parameters that are the focus of the study. For each experiment, we simulate 100 sets of noisy data from hydraulics model 1, each of length $n = 500$ days. Variation is created across these datasets by simulating stochastic job arrivals. For each of the 100 datasets, 5 sets of starting values are input to the optimisation algorithm – resulting in 500 sets of parameter estimates for each experiment. These estimates are compared to the true parameter values. The effects of adding different levels of noise to the data are also investigated.

Across all the experiments, we set $\underline{\theta}$ to the default hydraulics model 1 parameters. These are the values that, if the model was sufficiently complex to represent the full system, BT analysts believe should be used: $\underline{\theta} = (\theta_d, \theta_w, \theta_t) = (3, 0.2, 2)$. Such a model assumes the following:

- The delay associated with changing workforce numbers is 3 days
- When determining the 'target clear rate' in the model, 20% weighting is given to the cycle time and 80% to the job arrivals
- The target wait time for all jobs is within 2 days

Across all the experiments, when updating hydraulics model 1 we set $dt = 1$. The unit of time for the model is in days, so with $dt = 1$ the variables within the model update once per day – producing daily time series for the backlog and capacity deployed. A smaller dt value would update the model more often and the variables in the model would be closer to being theoretically continuous, see chapter 2. This would more accurately represent the behaviour of the model. However, when we apply our approach to BT regional time series data in chapter 6, we have only daily series. The real system will of course be continuously updating – but we will only have access to data updated once per day. Our state space method requires data for each dt update. So if in the real system we have daily time series data, we cannot update the state-space model more than once per day. Therefore for this simulation study, we select dt to be the smallest possible interval where we have access to time series data from the real system; $dt = 1$ day.

Peterson & Schweppe (1975) and Peterson (1976) found that the accuracy of parameter estimation with FIMLOF is improved with longer time series, by investigating series of lengths 100 and 1000. The BT data presented in chapter 6 has a length of nearly 1000. However, rather than opting for this length, we simulated datasets of

length 500 in order to create a more challenging set of experiments.

As we explain in sections 4.2.5 and 4.2.6, current implementations of FIMLOF are not entirely robust to the starting values input to the optimisation routine. That is, the selection of starting values may affect the parameter estimates produced. As part of our experiments, we investigate this by testing 5 distinct sets of starting values for each of the 100 datasets. Sets 1-5 (denoted by columns s1 - s5), are shown in table 4.4.1. The values for each parameter across the 5 sets are spaced at regular intervals. Set 1 takes the lowest values, which then increase up to set 5 which has the highest values. Set 3 takes the median of these values - these are also equal to the true values of the model parameters. The starting values of the 4 variance parameters $\sigma_{Q_1}, \sigma_{Q_2}, \sigma_{R_1}, \sigma_{R_2}$ are set around 0 at e^{-10}, despite the parameters taking sometimes relatively high values as we see later. We do this to test the algorithm's ability to estimate parameters from noisy data – when have no initial idea of the scale of that noise.

Parameter	s1	s2	s3	s4	s5
θ_d	2	2.5	3	3.5	4
θ_w	0.1	0.15	0.2	0.25	0.3
θ_t	1	1.5	2	2.5	3

Table 4.4.1: Starting values for optimisation.

For the optimisation routine, certain parameters can be constrained to reduce the field of search. This is one of a number of recommendations that Dangerfield and Roberts (1996) present as strategies for making the search of the large parameter space more practical. Since $\phi_1 = \frac{\theta_w}{\theta_d \theta_t}$ from (4.3.7) and the $\underline{\theta}$ parameters have the constraints $0 \leq \theta_w \leq 1$ and $\theta_d, \theta_w \geq 0$, we can deduce that ϕ_1 is constrained as follows: $0 \leq \phi_1 \leq 1$. In addition, since $\sigma^2_{Q_1}$, $\sigma^2_{Q_2}$, $\sigma^2_{R_1}$ and $\sigma^2_{R_2}$ represent variance terms, these must be non negative. Therefore these are constrained using exponential functions, for example $\sigma^2_{Q_1} = \exp(\rho^2_{Q_1})$.

4.4.3 Job arrivals

In chapter 2, the response of hydraulics model 1 to an artificial scenario, a single stepped increase in job arrivals, was examined. This response consisted of increasing workforce numbers (capacity deployed) because of the increasing backlog. In order for this step in job arrivals to ellicit a response from the model, a sufficiently large step was required. In other words the job arrivals series was required to possess a sufficient change in demand for a response to be observed from the hydraulics model. The step chosen was a 33% increase for 10 days. In this simulation study, two of the parameters in particular, θ_d and θ_w, determine the dynamics of how the model responds to changes in demand. Therefore it is essential that sufficient changes in demand are present in the job arrivals in order to ellicit a response from hydraulics model 1. We present a small study here that was used to determine which job arrivals should be used in the main experiments detailed in section 4.4.5.

In this study, experiments were set up according to the details in section 4.4.2. No noise was added to the data simulated from hydraulics model 1. Stepped increases were introduced at certain periods along the job arrivals series U. The five scenarios below were simulated, increasing in levels of changes in demand for U. Except for periods of the time series stated otherwise, each realisation of series U is simulated from $U_t \sim N(\mu, \sigma^2)$ with $\mu = 930$ and $\sigma = 5$.

1. Steady state: no steps.

2. Single step: for days 300-350 increase μ to 1200.

3. Double step: for days 150-200 increase μ to 975, days 350-400 increase μ to 1050, otherwise $\mu = 900$.

4. Multiple steps: where μ takes the following values for each 50 day interval, starting with the interval from days 0-50 up to days 450-500. $\mu \in$ (900, 1000, 900, 950, 900, 975, 900, 1025, 950, 900).

5. High activity system: where μ takes the following values for the same 50 day intervals as (4). $\mu \in$ (850, 975, 800, 1200, 1050, 900, 1100, 1200, 875, 930).

Time series for job arrivals scenarios 3 and 5 are shown in Figure 4.4.1, taking the upper and lower plots respectively. Figure 4.4.2 shows the results of the study. Each row shows the histograms for each job arrivals scenario. The histograms in each column represent one of the 3 structural parameters; θ_d the left column, θ_w the centre and θ_t on the right. Each histogram shows the 500 estimates for that column's parameter for that row's job arrivals scenario. The vertical red line on each histogram is the true

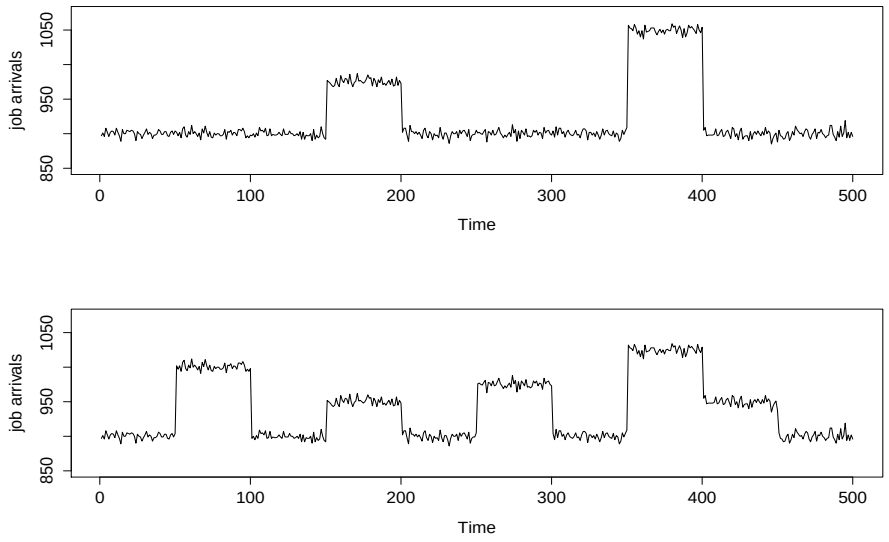

Figure 4.4.1: Job arrivals scenarios 3 and 5.

parameter value.

With the steady state scenario (no spikes) on the first row, there is a noticeable drop in performance compared to the other 4 scenarios, especially for the parameter θ_d, with a wider spread of parameter estimates. With change in demand set at a single step or above, the change in performance is less obvious. Therefore it makes sense to use a scenario with at least a single step. Conversely, it also makes sense to use a scenario that does not have unrealistically high levels of changes in demand. The high activity scenario 5, the lower plot in Figure 4.4.1, is likely to suffer from this. Indeed, this is confirmed when comparing this series with the job arrivals in the BT historical data in chapter 6. Scenario 3, the 2 step job arrivals scenario, was selected for the main experiments presented in section 4.4.5. This was seen as a compromise

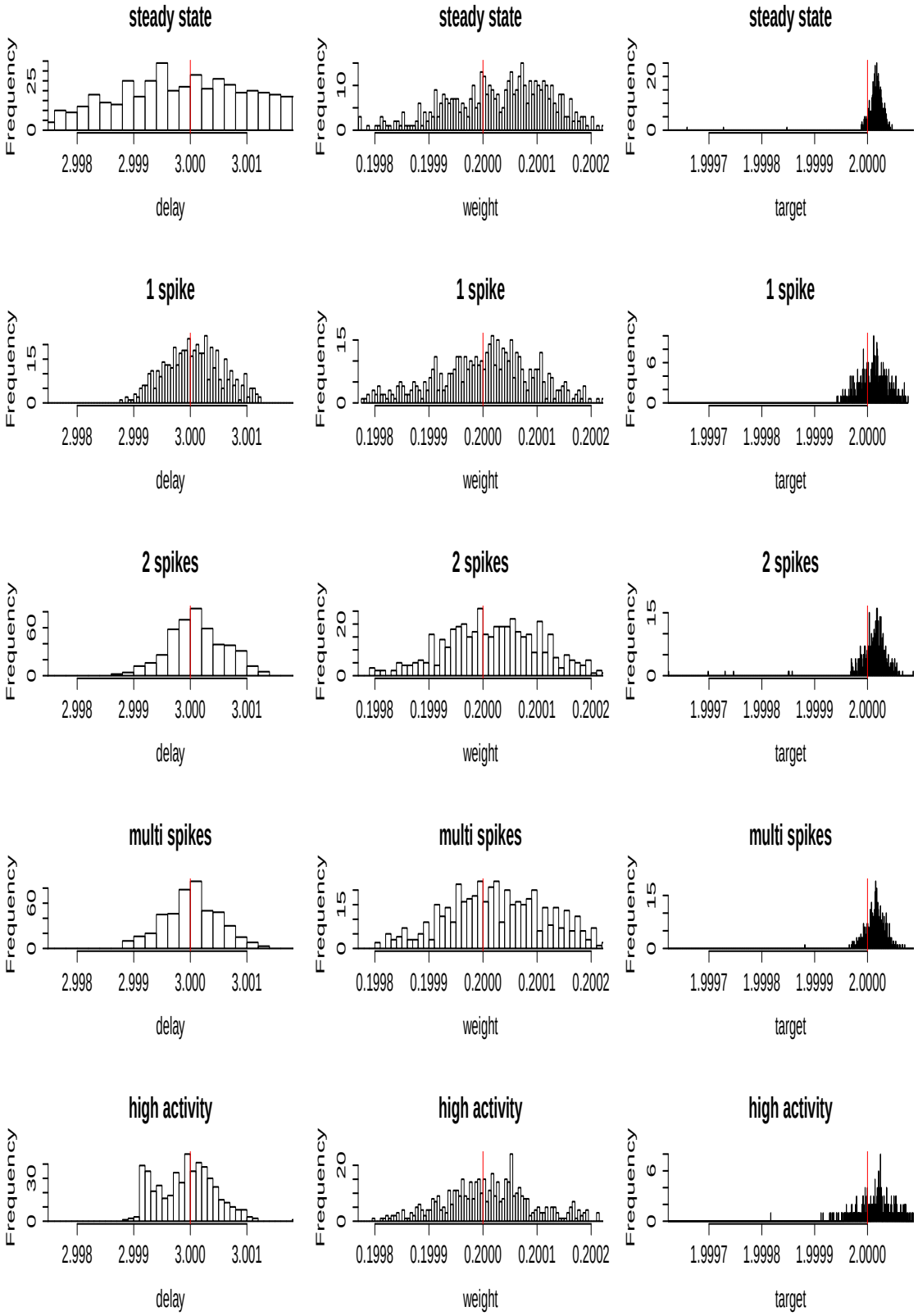

Figure 4.4.2: Effect of job arrival patterns on accuracy of estimates.

between possessing sufficient changes in demand to ellicit a response from hydraulics model 1 whilst not being unrealistically variable. In section 4.4.4, where we verify the Gaussian requirements of the simulated data for Kalman filtering, we also present example time series for the backlog and capacity deployed, simulated from hydraulics model 1 using job arrivals scenario 3.

4.4.4 Verifying Gaussian assumptions for Kalman filtering

In the experiments of section 4.4.5, we estimate the parameters of data simulated from the state-space model of (4.3.10)-(4.3.13), which is based on hydraulics model 1. We already know from (4.3.10)-(4.3.13) that we have a linear system, and that both the state and observation errors are Gaussian. In this section we demonstrate that this simulated data is suitable for Kalman filtering - which is applied within the FIMLOF algorithm.

We begin by observing that the time series data $\underline{y}$ simulated from hydraulics model 1 consists of non-integer values. That is, although the model effectively represents counts of jobs in the backlog and counts of deployed engineers in the system, the data simulated from the model are not counts and take non-integer values. This is the case for SD modelling in general where variables are treated as continuous. To verify the Gaussian nature of the errors, we simulate time series from hydraulics model 1 using job arrivals scenario 3 (the upper plot in Figure 4.4.1) as the exogenous variable and one of the sets of model parameters $\underline{\theta}$ that are used in section 4.4.5 - we select the same parameter values as for experiment 1. That is, $\underline{\theta} = (\theta_d, \theta_w, \theta_t) = (3, 0.2, 2)$

with state noise $\sigma_Q = 1$ and observational noise $\sigma_R = 10$ and set the length of the time series $n = 500$ days. One realisation of the time series produced from hydraulics model 1 for the backlog and capacity deployed is shown in Figure 4.4.3.

In Figure 4.4.3 we see that each of the stepped increases in the job arrivals of scenario 3 (of Figure 4.4.1) causes a slightly delayed step in the backlog, as the uncompleted jobs cause the queue size to increase. A step in capacity deployed, delayed slightly further, can be observed. This can be viewed as hydraulics model 1 replicating the behaviour of management to increase the workforce numbers to deal with the extra jobs. In other words, aside from an initial period of around 30 days where the backlog and capacity deployed series have not yet reached steady state, we can see that the trend present in these two series is caused by the trend in the job arrivals; the two temporary stepped increases. If the system was in steady state, i.e. if the job arrivals had no trend - which would rarely be the case in a complex system such as BT's - the raw backlog and capacity deployed series would be Gaussian distributed. To assess the suitability of our data for Kalman filtering, we remove the trend from the backlog and capacity deployed series and inspect the residuals.

We remove the trend by taking a centred 3 point moving average of each series and inspecting the residuals when subtracting the moving average series from the raw series. Histograms and Q-Q plots of these residuals are presented in Figure 4.4.4. These demonstrate that when the trend has been removed, the backlog and capacity deployed series are Gaussian and as such, are appropriate for Kalman filtering. It is

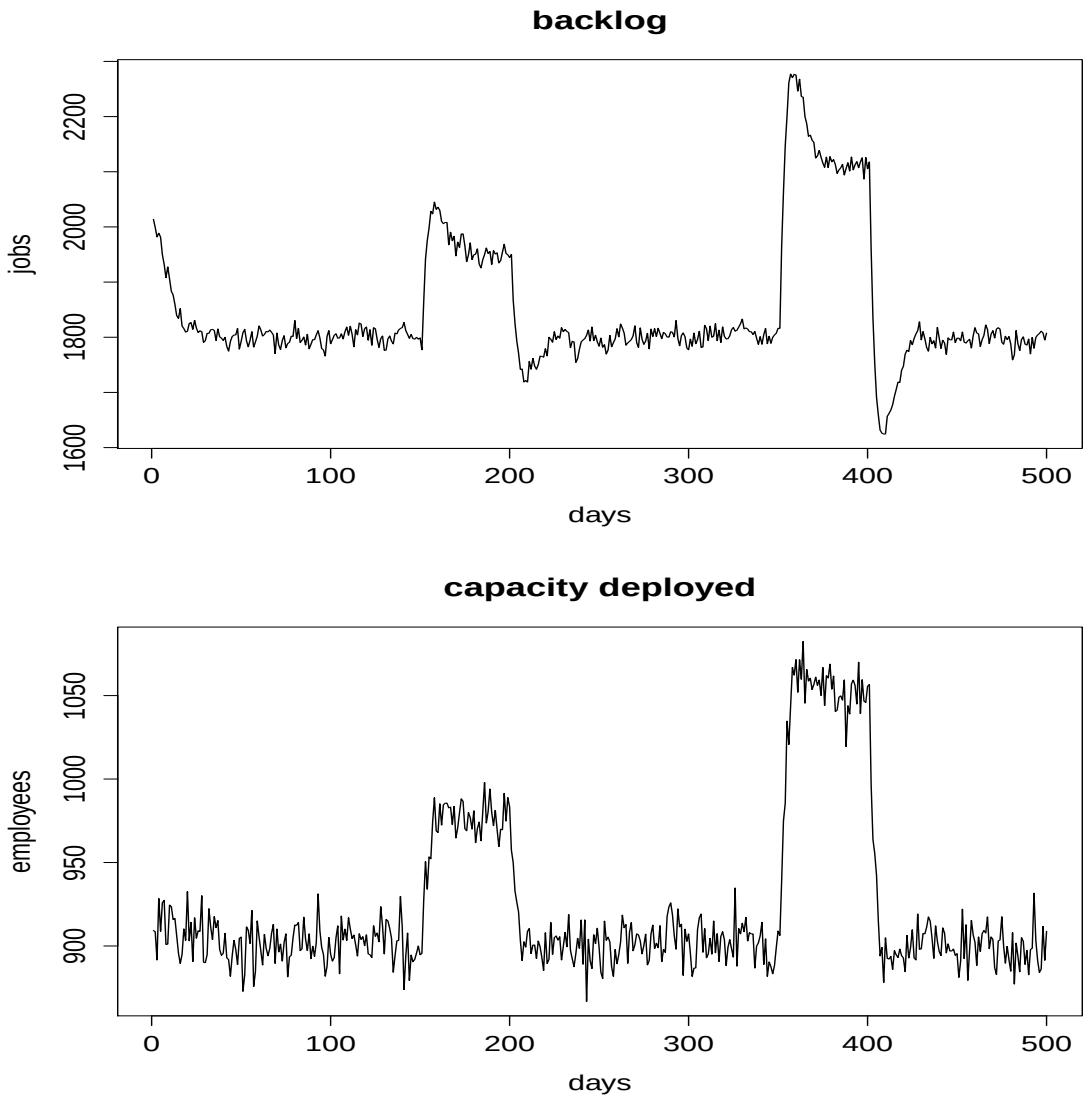

Figure 4.4.3: Hydraulics model 1 output time series using job arrivals scenario 3.

worth highlighting that the data simulated from hydraulics models 2 and 3, introduced in chapter 5, after removal of the trend, can similarly be shown to be Gaussian.

4.4.5 Results

In this section we present the the results of the simulation study. Both state noise and observational noise are added to simulated data from hydraulics model 1. The effect of these noise terms on the accuracy of the structural parameter estimates $\underline{\hat{\theta}}$ is investigated. We present the results of two experiments. Experiment 1 fixes the state noise and varies the amount of observational noise, while experiment 2 fixes the observational noise and varies the amount of state noise. Both experiments will test the performance in terms of the accuracy of parameter estimation, while comparing the results of each will determine the individual effect of state and observational noise.

For experiment 1, we fix both the observation noise terms $\underline{w}_1$ and $\underline{w}_2$ of (4.3.12)-(4.3.13) to have a standard deviation of 10. That is, $\underline{w}_1 \sim N(0, \sigma^2_{R_1})$, $\underline{w}_2 \sim N(0, \sigma^2_{R_2})$ where $\sigma_{R_1} = \sigma_{R_2} = 10$. This means that white noise processes $\underline{w}_1$ and $\underline{w}_2$ have a variance of $10^2 = 100$. This variance is around 5% of the mean of the simulated backlog series and 10% of the mean of the capacity deployed series, as can be seen in Figure 4.4.3. We then vary the state noise terms, σ_{Q_1} and σ_{Q_2}, using the following values: 1, 5, 10, 20, 50, 80. These higher value terms are included to assess the possible drop in performance when high levels of noise are present. Table 4.4.2 shows the six parameter sets used to add noise to the simulated data in experiment 1.

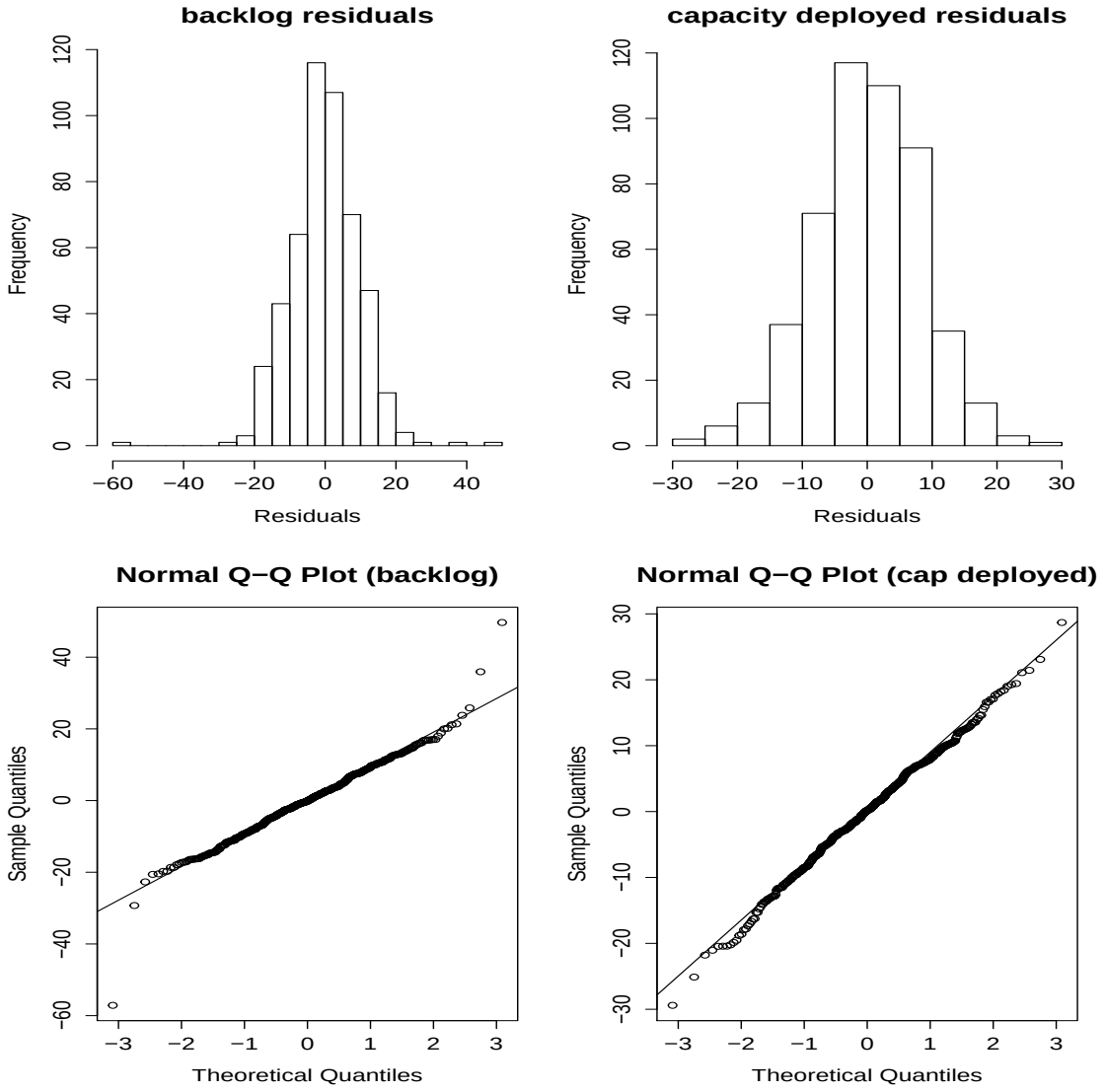

Figure 4.4.4: Histograms and Q-Q plots of residuals.

Histograms of the parameter estimates from experiment 1 are shown in Figure 4.4.5.

	Expt 1		Expt 2	
Parameter set	σ_Q	σ_R	σ_Q	σ_R
1	1	10	10	1
2	5	10	10	5
3	10	10	10	10
4	20	10	10	20
5	50	10	10	50
6	80	10	10	80

Table 4.4.2: Parameter sets for the noise terms of experiments 1 and 2.

These take a similar form to the histograms in Figure 4.4.2, with each row showing the histograms for each noise parameter set and each column again representing one of the 3 hydraulics model 1 structural parameters; θ_d the left column, θ_w the centre and θ_t on the right. Each histogram shows the 500 parameter estimates for that column's structural parameter for that row's noise parameter set.

In experiment 2 we effectively do the opposite, fixing the state noise terms and varying the observational noise. We now fix both the state noise terms $\underline{v}_1$ and $\underline{v}_2$ of (4.3.10)-(4.3.11) to have a standard deviation of 10. That is, $\underline{v}_1 \sim N(0, \sigma^2_{Q_1})$, $\underline{v}_2 \sim N(0, \sigma^2_{Q_2})$ where $\sigma_{Q_1} = \sigma_{Q_2} = 10$. We then vary the observational noise terms, σ_{R_1} and σ_{R_2}, using the same values as for the state noise terms in experiment 1; 1, 5, 10, 20, 50,

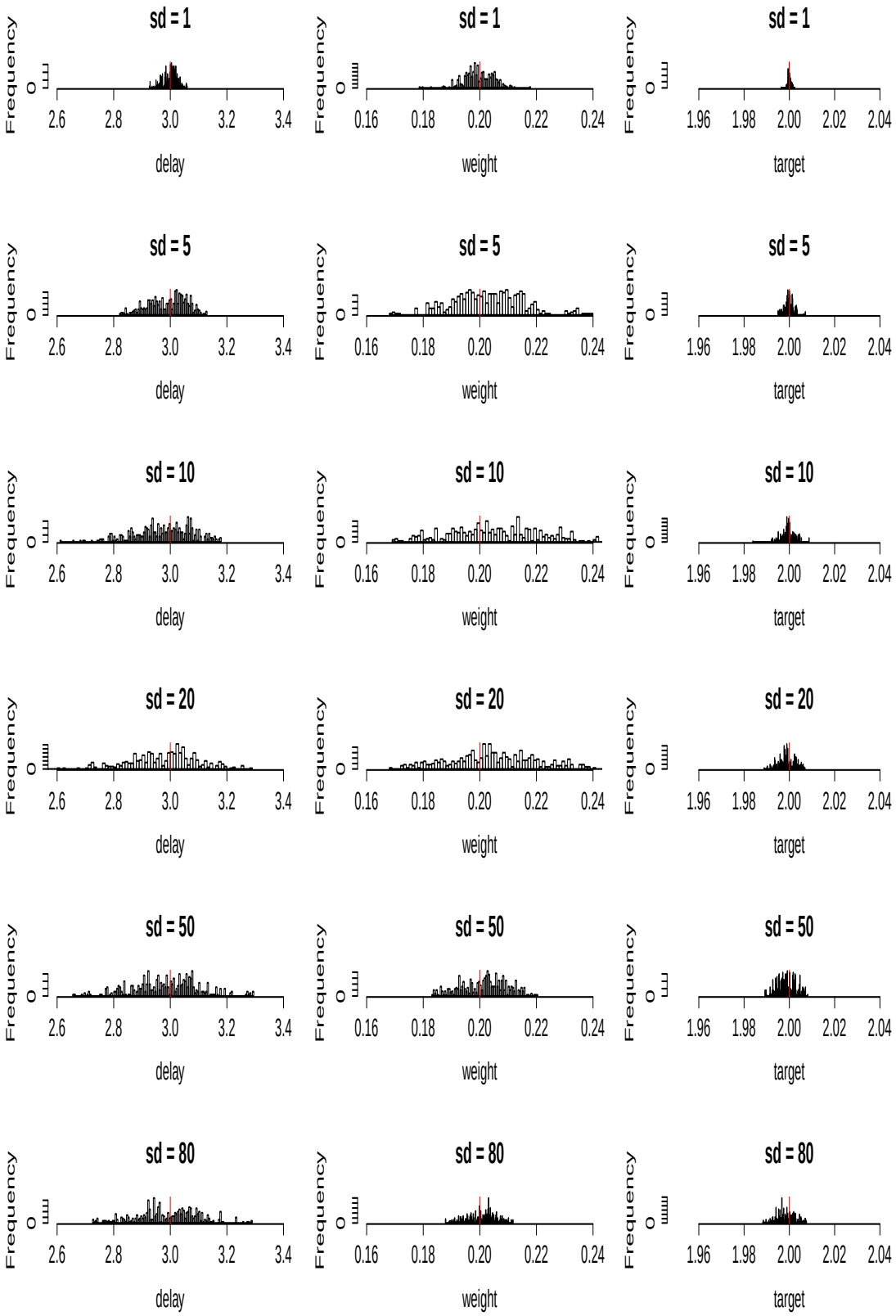

Figure 4.4.5: Experiment 1 histograms of parameter estimates.

80. These 6 parameter sets used to add noise to the simulated data in experiment 2 are shown in table 4.4.2. Histograms of the parameter estimates for experiment 2 are shown in Figure 4.4.6.

To compare the results from experiments 1 and 2, we examine the relative difference of each of the parameter estimates compared to the true value. Let $\hat{\theta}_i$ be the estimate of true parameter θ_i, for $i = d, w, t$. Then $\tau_i = \frac{|\hat{\theta}_i - \theta_i|}{\theta_i}$, for $i = d, w, t$, is the absolute value of the relative difference between $\hat{\theta}_i$ and θ_i. $\overline{\tau}_i = 100 \times \frac{1}{n} \Sigma_1^n \tau_i$, where $n = 500$, represents the mean absolute percentage error (MAPE). $\overline{\tau}_i$ is calculated for $i = d, w, t$, for each different setting of system and observational noise, for experiments 1 and 2. These values of $\overline{\tau}_i$ are shown in table 4.4.3.

	Expt 1 ($\sigma_R = 10$)				Expt 2 ($\sigma_Q = 10$)			
Parameter set	σ_Q	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$	σ_R	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	1	1.9	0.4	0.1	1	1.2	0.2	0.0
2	5	5.6	1.0	0.2	5	5.0	0.9	0.2
3	10	8.3	1.4	0.2	10	8.3	1.4	0.2
4	20	9.6	1.4	0.3	20	11.5	1.9	0.3
5	50	9.2	0.7	0.3	50	14.2	2.6	4.6
6	80	8.8	0.4	0.4	80	17.1	3.5	0.4

Table 4.4.3: MAPE of parameter estimates for experiments 1 and 2.

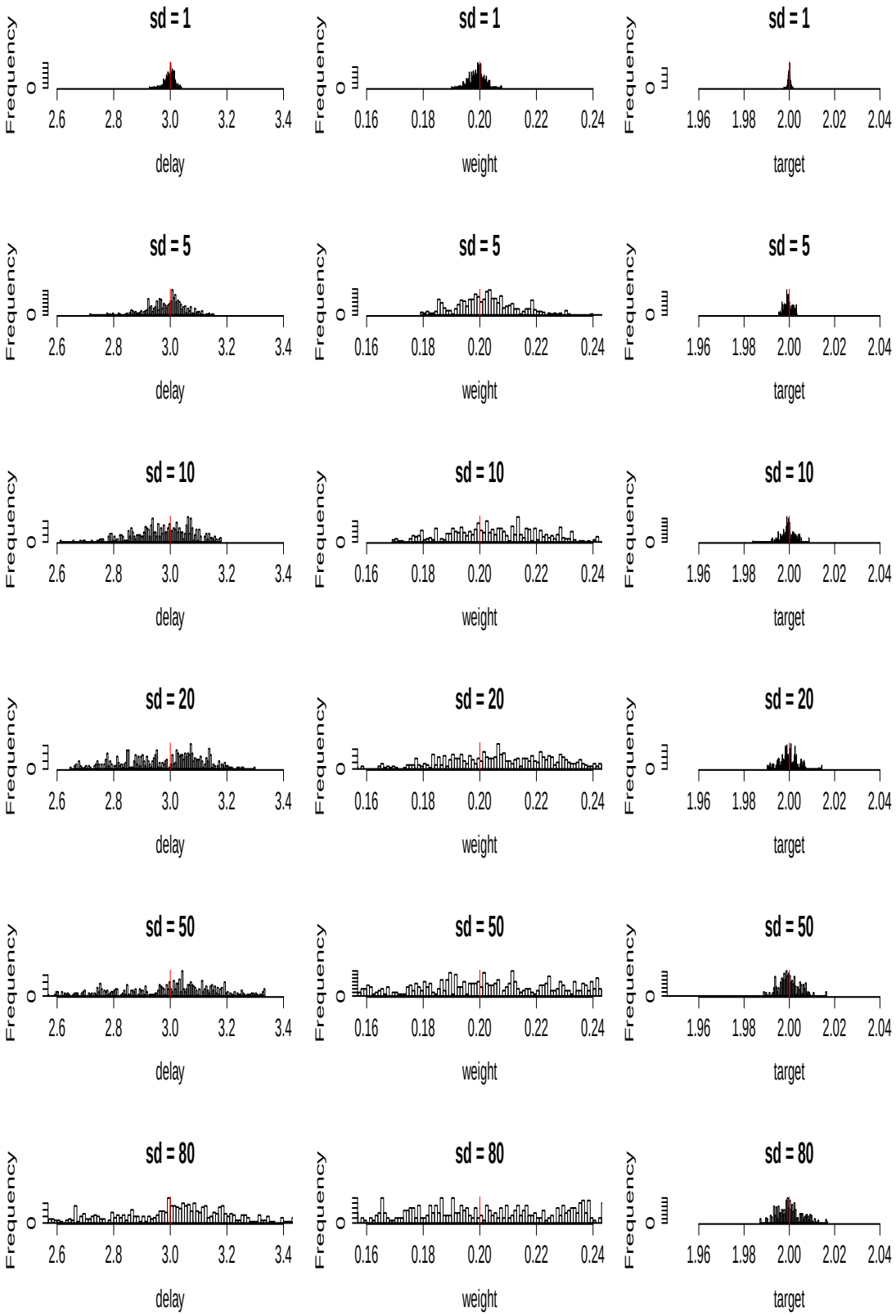

Figure 4.4.6: Experiment 2 histograms of parameter estimates.

From table 4.4.3 and the histograms in Figures 4.4.5 - 4.4.6, the following are immediately apparent:

1. With lower and mid-range noise parameters, parameter estimation is to a high degree of accuracy.

2. For each experiment, increasing the noise added to the data results in generally less accurate parameter estimates.

3. An interesting exception to observation (2) are parameter sets 5 and 6 of experiment 1 for parameters θ_d and θ_w. In other words parameter estimation is more robust to added state noise $\underline{\sigma}_Q$ than added observational noise $\underline{\sigma}_R$.

4. For both experiments, θ_t is the most accurately estimated parameter, followed by θ_w and then θ_d.

Observation (1) can be reasoned as follows. For parameter sets 1-3, θ_d is the most poorly estimated parameter with the least accuracy observed when using noise parameter set 3, where $\sigma_Q = \sigma_R = 10$. Even with these noise parameters, the MAPE of estimates for θ_d is 8.3%. All other parameter estimates from using noise parameter sets 1-3 have MAPE values well below this. In experiment 1, even with higher amounts of state noise added with parameter sets 4-6, the MAPE of all estimates remain within 10%. The same cannot be said for experiment 2 where for parameter sets 4-6, the MAPE for θ_d increases above 10%.

Observation (2) is not surprising; as we add more noise to the simulated data we

obtain less accurate parameter estimates. With greater amounts of added noise, the structure of the data is corrupted more, so that it represents the hydraulics model to a lesser extent. Although the Kalman filter is designed for noisy Gaussian data, its accuracy in estimating the latent state from noisy data lessens as the amount of noise increases. When this estimate of the latent state – which is in this case the true output from hydraulics model 1 – is less accurate, estimates of the structural parameters $\underline{\theta}$ are likely to be less accurate.

Observation (3), the greater robustness of estimates to higher state noise than observational noise is interesting. A closer inspection of table 4.4.3 reveals the following:

- For parameter sets 1 & 2 with lower amounts of added noise, experiment 2 produces the more accurate parameter estimates. With these parameter sets, σ_Q is fixed at 10 while σ_R is 1 or 5.

- The opposite is observed for parameter sets 4-6 with greater amounts of added noise; experiment 1 produces the more accurate estimates. With these parameter sets, σ_R is fixed at 10 while σ_Q is 20, 50 or 80.

Across parameter sets 1-6, there is one common aspect of the noise parameters that result in these more accurate estimates; the σ_Q/σ_R ratio. The σ_Q/σ_R ratio is often called the *signal to noise* ratio in the engineering literature. Petris et al. (2009) contains a good introduction and examples of this important aspect of state-space modelling. In these experiments, the best parameter estimates result when σ_Q is larger than σ_R; i.e. a larger signal to noise ratio exists in the noise added to the

data. Although there is a general pattern in the results of less accurate parameter estimates from data with greater amounts of added noise, the signal to noise ratio is a likely explanation for the difference in success between experiments 1 and 2 at estimating parameters from data with higher levels of added noise; parameter sets 5 and 6. Whilst we observe a considerable drop in performance in experiment 2 for parameter sets 5 and 6 - even for the normally reliable estimate of θ_t, the opposite is observed for experiment 1 and the accuracy of estimates improves. The signal to noise ratios for parameter sets 5 and 6 are 5 and 8 for experiment 1 and $\frac{1}{5}$ and $\frac{1}{8}$ for experiment 2. For experiment 1, the effect of the added noise which would normally result in less accurate estimates, is overridden by the large signal to noise ratios. The small signal to noise ratio of experiment 2 is not able to override the effect of the added noise. For parameter sets 1 & 2 with lower amounts of added noise, the larger signal to noise ratio also explains the improved performance of experiment 2. This has signal to ratios of 10 and 2 compared to $\frac{1}{10}$ and $\frac{1}{2}$ for experiment 1.

To explain observation (4) we must discuss the concept of *identification* of parameters. This is the ability of the data to distinguish between different model parameters. Eberlein and Wang (1985) include a good discussion of this issue for econometric models. The authors highlight an example of two model parameters that always appear in the SD model multiplying one another; then doubling one and halving the other would have no effect on model output. The two parameters individually would not be identified. Although that is not the case here, identification of parameters is still an issue. θ_t is the most accurately estimated parameter because it is the most easily

identified. The value of this parameter, the target cycle time, determines the capacity deployed levels relative to the backlog. In other words, given the backlog it determines how fast the jobs need to be completed. It is straightforward to estimate this directly from data: $\theta_t \approx$ backlog/capacity deployed. As such, the shape of the data (trend or steps in job arrivals) is almost irrelevant in estimating this parameter. The opposite is true however for parameters θ_d and θ_w which certainly cannot be estimated in a simple manner directly from the data. Information that may reveal the values of these parameters is only provided when the system is responding to changes. In Figure 4.4.3, changes occur in the system in response to steps in job arrivals to either increase workforce numbers (when the steps begin) or decrease numbers (after the step). This occurs on only 4 occasions across the entire time series. Therefore, there is considerably less information available to estimate these parameters θ_d or θ_w than there is for θ_t. Peterson and Schweppe (1975) and Eberlein and Wang (1985) both describe examples of models where identification of parameters has proved to be a serious issue.

As we explain in sections 4.2.5 and 4.2.6, the standard FIMLOF approach in the literature finds the local minima rather then the global minima and hence is not robust to the choice of starting values. This is still the case when applying our modifications to the FIMLOF algorithm for this simulation study. In the study, for each of the 100 datasets, parameter estimates from 5 different sets of starting values were calculated. As expected, for many of the 100 datasets, slightly different parameter estimates resulted from choosing different starting values. However, it is worth highlighting that when comparing parameter estimates from each of the 5 sets of starting values, there

is only a marginal change in accuracy between them. In other words, although the values of the parameter estimates are not robust to choice of start values, the accuracy of those estimates is robust to that choice, at least for the 5 sets used in this study.

Another important aspect of the results is the choice of starting values that were used for the variance parameters σ_{Q_1}, σ_{Q_2}, σ_{R_1} and σ_{R_2} in the optimisation routine. It is worth highlighting that in the simulation study, although we were adding in some cases considerable noise to the data, we did not inform the optimisation routine of this (via the starting values) and deliberately used values close to zero of e^{-10}. In other words, although we assumed that we knew the correct noise structure (the diagonal nature of the covariance matrices), we made no assumptions regarding the scale of this noise by using starting values that were close to zero. The reasoning for this is that in chapter 6, although assumptions will again be made regarding the noise structure of the historical BT time series, we will have no prior knowledge of the scale of this noise. It was therefore considered reasonable at this stage to investigate the effects of this noise on performance *without prior knowledge of the scale* of this noise.

It is worthwhile comparing the results to determine any changes in performance when the algorithm is given the correct noise structure in the starting values of the optimisation routine. We choose parameter sets 3-5 from experiment 1 in table 4.4.2, so that $\sigma_R = 10$ is fixed and σ_Q takes values of 10, 20 and 50. However, unlike experiment 1 where 'naive' starting values were set at $\sigma_{Q_0} = \sigma_{R_0} = \exp(-10)$, this time we set 'informed' start values at the true parameter value by squaring the relevant standard

deviation parameter. This means we set $\sigma_{Q_0} = \sigma_Q^2$ and $\sigma_{R_0} = \sigma_R^2$. These results are shown in table 4.4.4. It is clear that although there is a slight improvement in performance for parameter set 3 when using informed start values, performance actually drops for parameter sets 4 and 5. Understanding this drop in performance when using informed starting values is not immediately apparent. Even for this simplified model, when estimating both the structural and variance parameters the 7 parameter likelihood surface is likely to be extremely complex. The Nelder-Mead simplex method used here, although shown to be an improvement on the Powell method in section 4.4.6, can only find the local optima. It is possible that when using informed variance start values for parameter sets 4 and 5, the optimisation routine gets stuck in a local optima.

Although for this study the effect of using naive starting values for variance parameters actually improved performance, this issue is a concern. It is not desirable for the starting values used to estimate variance parameters to affect the accuracy of the resulting estimates of the variance, but more serious is that these variance starting values can actually affect the accuracy of estimates for the structural parameters $\underline{\theta}$, as table 4.4.4 demonstrates. However, what is promising overall at this stage is that experiments 1 and 2 show that the structural parameters $\underline{\theta}$ can be estimated to a good degree of accuracy with no prior knowledge of the scale of the added noise. This is an important step towards applying this method to the historical BT time series in chapter 6.

		Expt 1 ($\sigma_R = 10$)			
Parameter set	σ_Q	Start values $\sigma_{Q_0}, \sigma_{R_0}$	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
3	10	Naive: e^{-10}	8.3	1.4	0.2
		Informed: σ_Q^2, σ_R^2	8.2	1.4	0.2
4	20	Naive: e^{-10}	9.6	1.4	0.3
		Informed: σ_Q^2, σ_R^2	10.9	1.4	0.3
5	50	Naive: e^{-10}	9.2	0.7	0.3
		Informed: σ_Q^2, σ_R^2	12.6	0.8	0.3

Table 4.4.4: MAPE of parameter estimates for experiment 1 using naive and informed starting values for the variance parameters in the optimisation routine.

4.4.6 Modifications to FIMLOF

This section describes how two of our modifications to the FIMLOF algorithm resulted in improved performance in terms of the accuracy of parameter estimates. The details of these modifications are described and we present the results of comparing these modifications with the standard FIMLOF settings.

As we explain in section 4.2.5, the standard FIMLOF algorithm was devised in the 1970's by Schweppe (1973) and Peterson & Schweppe (1975). In the years since there have been relatively few examples of works that have incorporated the FIMLOF algorithm and possible reasons for this were discussed in section 4.2.6. The only two recent examples of applications of the FIMLOF algorithm, Ryzhenkov (2002) and

Radzicki *et al.* (2004), were described in section 4.2.6. These however did not develop the FIMLOF algorithm and simply applied it 'off-the-shelf' within the Vensim (2010) SD simulation package.

Through experimentation, we discovered two modifications to the standard FIMLOF algorithm devised in Schweppe (1973) and Peterson & Schweppe (1975) that resulted in improvements to performance, in terms of the accuracy of estimated parameters. In order to estimate the parameters as accurately as possible, these modifications were incorporated in the main experiments in section 4.4.5.

Firstly, the use of different representations of the log-likelihood calculation was investigated. Performance using the approximation of the log-likelihood of (4.2.15) within the FIMLOF algorithm was compared to the exact representation of (4.2.14). The setup of an experiment to compare the two was the same as the experiments described in the main results in section 4.4.5. The only difference here was that no state noise σ_Q was added to the simulated data from hydraulics model 1. Six different levels of observational noise σ_R were added. The standard deviations of this added noise again took the values 1, 5, 10, 20, 50 and 80. The results of this experiment are presented in table 4.4.5.

Contrary to what may be expected, results show that estimates obtained using the approximation of the log-likelihood are consistently superior than the exact representation, for all three parameters. This is particularly noticeable for parameter sets 1

		Approximate			Exact		
Parameter set	σ_R	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	1	0.2	0.1	0.0	72.2	21.9	2.4
2	5	0.8	0.2	0.0	16.9	9.7	0.2
3	10	1.6	0.4	0.0	2.2	0.8	0.1
4	20	3.5	0.9	0.1	4.5	1.6	0.1
5	50	8.9	1.7	0.2	16.3	3.6	0.2
6	80	26.0	5.4	0.3	41.4	9.3	0.3

Table 4.4.5: MAPE of parameter estimates using the exact and approximate forms of the log-likelihood calculation within FIMLOF.

and 2 with low levels of added noise, where the MAPE for the exact representation is considerably higher. The poor performance of the exact representation for low levels of noise is due to a large number of parameter estimates out of the 500 that are consistently over or underestimated. The algorithm appears to become stuck in a local minima, sometimes far from the true parameter value.

The approximation in (4.2.15) does not involve the inversion of matrices. It is possible that by avoiding this operation, an accumulation of numerical errors is avoided that actually distort the accuracy of the likelihood greater than it is improved by using an exact, rather than approximate, representation. It is worth highlighting that while this applies for the simulation study in this section, it does not apply for the full

version of the hydraulics model introduced in chapter 5 where we switch to using the exact form in (4.2.14).

The use of alternative optimisation algorithms was also investigated. One such method was the Nelder-Mead simplex method (Nelder & Mead, 1965). The performance of using this within FIMLOF was compared to Powell's method. The setup of an experiment to compare the two was the same as the log-likelihood experiment above, again with no state noise σ_Q added to the simulated data from hydraulics model 1, and the same six levels of added observational noise σ_R. The results of this experiment are presented in table 4.4.6.

		Nelder-Mead			Powell's method		
Parameter set	σ_R	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	1	0.2	0.1	0.0	4.6	2.2	0.2
2	5	0.8	0.2	0.0	3.4	2.3	0.2
3	10	1.6	0.4	0.0	2.9	2.0	0.8
4	20	3.5	0.9	0.1	8.1	2.1	0.5
5	50	8.9	1.7	0.2	10.7	3.0	0.9
6	80	26.0	5.4	0.3	20.1	4.2	0.5

Table 4.4.6: MAPE of parameter estimates using Nelder-Mead and Powell's method within FIMLOF.

Results show that with the exception of parameter set 6, estimates obtained using the Nelder-Mead method are consistently superior, for all three parameters. The results for parameter sets 1–4 in particular show that the Nelder-Mead method offers a considerable improvement over Powell's method. Note also that computations for the Nelder-Mead method were on average around 30% faster than Powell's method. The results shown here both use the approximation to the log-likelihood of (4.2.15). Note that use of Powell's method using the exact log-likelihood of (4.2.14) resulted in a further drop in performance.

4.5 Discussion

In this chapter we have presented an improved version of an existing approach for estimating the parameters of linear Gaussian state-space models, formed from the structure of a simplified SD model. Gaussian noise was added to simulated data from the SD model and the effect of this noise on the accuracy of estimates was studied. Results showed that despite having no prior information on the values of the variance parameters, the methods demonstrated success in estimating the parameters to a good degree of accuracy, especially for lower amounts of added noise. The signal to noise ratio of the added noise was considered to be an important factor in the ability of the method to estimate parameters.

A process for the formulation of a state-space model from a SD model has also been presented. This consisted of grouping the SD variables, determining the order in which

each should update and then performing substitutions to form Markovian difference equations; the addition of state and observational noise terms forms the state-space model. The step by step guide that we provide for this process is a novel contribution to the area. Although this was demonstrated for a simple SD model, the process extends to more complex SD models, as we see in chapter 5.

As we explain in sections 4.2.5 and 4.2.6, the FIMLOF method has had relatively little testing on simulated data with known model parameters in the literature. The simulation study in section 4.4 makes a claim to contribute towards this area. Although the analyses of Ryzhenkov (2002) and Radzicki et al. (2004) are applications to historical data and not simulation studies, they present only a single set of estimates – the 'best' estimates produced from the Vensim (1994) package using random multiple start values for the optimisation. Likewise, the simulation studies in Peterson & Schweppe (1975) and Peterson (1976) present only one set of estimates. In our simulation studies, we perform Monte Carlo experiments and hence explore more thoroughly the performance of the FIMLOF algorithm and reveal the effects of the starting values.

The results of the simulation study can be taken as an important first step towards estimating the parameters of the BT time series data in chapter 6. Although using a simplified version of the SD model, we have demonstrated success in estimating its parameters without any prior knowledge of the scale of the added noise. In chapter 5 we take a further step by advancing to more complex versions of the hydraulics model

and conduct more simulation studies.

Chapter 5

Nonlinear state-space models and the Unscented Kalman filter

5.1 Introduction

The modified FIMLOF method detailed in chapter 4 is limited to linear Gaussian state-space models. Hydraulics model 1, described in chapter 4, is an extremely simple version of the full system dynamics model used by BT. The difference equations formed from this model in section 4.3.2 were linear in nature. This is not generally the case. System dynamics models that are used in practice are usually highly nonlinear (Sterman, 2000). In this chapter we present two versions of the hydraulics model that produce nonlinear difference equations, and therefore form nonlinear state-space models. An existing modified version of the Kalman filter for nonlinear systems, known as the unscented Kalman filter, is applied within the FIMLOF algorithm. The objective of this chapter is to present further modifications to the FIMLOF algorithm in

chapter 4, extending the method to more realistic, nonlinear systems.

Below we briefly describe the two more complex BT models; hydraulics models 2 and 3. Full details of these are provided in sections 5.3.1 and 5.4.1 respectively.

- Hydraulics model 2: This model is similar to hydraulics model 1 in chapter 4, except for a few simple adjustments. These are made to accommodate a weekly profile exhibited by the exogenous variable, the job arrivals, and effectively consist of smoothing certain variables. Although simple in nature, these adjustments are sufficient to cause the model's difference equations to become nonlinear. Estimating the parameters of this model effectively tests our modified algorithm on a model that is only marginally more complex than hydraulics model 1, but is nonlinear.

- Hydraulics model 3: This is a considerably more complex version of the hydraulics model. Although non-repair jobs such as installation jobs have been removed, this is otherwise the full version of the model. As such, it is capable of representing the real BT system. Estimating the parameters of this model effectively tests our modified algorithm on a model that is considerably more complex than hydraulics models 1 and 2 and which contains far more variables. It also has a much higher degree of nonlinearity than hydraulics model 2. Success with this model can be seen as the final step before applying our method to historical BT time series data in chapter 6.

This chapter is organised as follows. Section 5.2 introduces the unscented Kalman filter and the advantages of this over other techniques for nonlinear systems. Section 5.3 explains the details of hydraulics model 2 and describes further modifications to the FIMLOF algorithm that incorporate the unscented Kalman filter. A simulation study presents results of estimating the known parameters of hydraulics model 2 from simulated data. Section 5.4 has a similar structure to section 5.3, but provides details of hydraulics model 3 and a simulation study aimed at estimating its parameters. Section 5.5 presents the discussion.

5.2 Unscented Kalman filter

5.2.1 Nonlinear filtering

A more general form of the Gaussian state-space model introduced in (4.2.3)-(4.2.4), is shown in (5.2.1)-(5.2.2), where $\underline{y}$ are the vector of observations, $\underline{x}$ are the latent states and $\underline{u}$ are the exogenous variable observations. Noise terms have a Gaussian distribution as follows: $\underline{\epsilon}_1 \sim \mathcal{N}(0, Q)$ and $\underline{\epsilon}_2 \sim \mathcal{N}(0, R)$. $f(.)$ is the state update function and $g(.)$ the observation update function. At time t we have the following representation:

$$x_t = f(x_{t-1}, u_t) + \epsilon_{1,t}, \tag{5.2.1}$$

$$y_t = g(x_t, u_t) + \epsilon_{2,t}. \tag{5.2.2}$$

When (one or more) of functions f and g are nonlinear, the Kalman filter is unable to find the optimal estimate of latent states $\underline{x}$ and can only find an approximation.

Other methods have been developed, effectively including modifications to the standard Kalman filter, that while not exact, are able to find a more accurate approximation of the latent state. One such method is the extended Kalman filter (EKF) (Jazwinski, 1970). This attempts to find a linear approximation of the state and observation equations and then apply the Kalman filter to the linearised model. See chapter 10 of Durbin & Koopman (2001) for more details. However, the EKF has been shown to give poor performance for systems that are highly nonlinear. Shiryaev et al. (2002) demonstrates the weaknesses of using the EKF in the context of estimating system dynamics model parameters. The authors point out that the linearisation of the EKF does not always provide a good approximation of the nonlinear model. They also state that increased dimension of a system is another factor in reducing performance.

A popular alternative to the EKF is the unscented Kalman filter (UKF) (Julier & Uhlmann, 1997). Rather than attempting to approximate a nonlinear function, the UKF works on the principle that it is easier to approximate a probability distribution. This is achieved by deterministically selecting a set of points that when transformed through the nonlinear functions are able to give a Gaussian approximation of the mean and covariance of the filtered distribution. This is called the unscented transform (Julier and Uhlmann, 1997). The EKF is based on a first order approximation. However, Wan & van der Merwe (2000) show that the UKF is a third order approximation and as such, this more accurately represents the filtered distribution. Because of this, the UKF has been shown to outperform the EKF in many applications (Julier

& Uhlmann 1997; Wan & van der Merwe, 2000; Durbin & Koopman, 2001). In addition, the EKF requires a matrix of partial derivatives (Jacobians) to be calculated at each step. This is not required for UKF which is argued to be conceptually simple and straightforward to apply (Julier and Uhlmann, 2004).

As explained in section 4.2.6, Ryzhenkov (2002) and Radzicki et al. (2004), via the Vensim system dynamics simulation package, use the EKF when applying FIMLOF to nonlinear systems. Motivated by the argument above, for the nonlinear systems in this chapter (hydraulics models 2 and 3) and chapter 6, we choose to use the UKF, rather than the EKF. We describe the details of the unscented transform in section 5.2.2 and proceed to describe how it is incorporated in the UKF.

5.2.2 Unscented Kalman filter

Unscented transform

We begin this section by explaining the concept of the unscented transform before proceeding to describe how it is used within the UKF. The unscented transform is based on the idea that it is easier to approximate a Gaussian distribution than it is to approximate an arbitrary nonlinear function or transformation (Uhlmann, 1994). A diagram of the approach, taken from Julier and Uhlmann (1997), is shown in Figure 5.2.1. Below we present a description of the unscented transform, using the same notation as Julier and Uhlmann (1997), where the reader is directed for full details. For further expositions, see also Wan & van der Merwe (2000) and Julier & Uhlmann

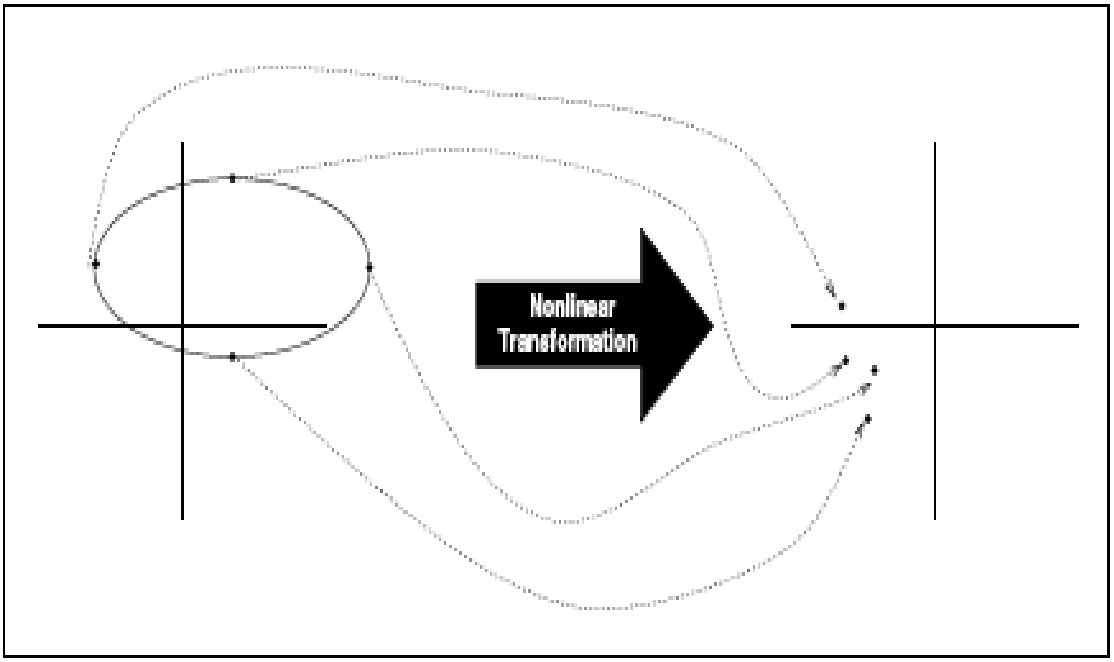

Figure 5.2.1: The unscented transform, taken from Julier & Uhlmann (1997).

(2004).

Suppose we have a random variable $\mathbf{x}$ with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}_x$. Random variable $\mathbf{y}$ is related to $\mathbf{x}$ through the nonlinear function $\mathbf{y} = \mathbf{f}(\mathbf{x})$. The objective of the transform is to calculate the mean and covariance of $\mathbf{y}$, $\bar{\mathbf{y}}$ and $\mathbf{P}_y$. A set of points called *sigma points* are deterministically selected so that their sample mean is $\bar{\mathbf{x}}$ and sample covariance is $\mathbf{P}_x$. The nonlinear function $\mathbf{f}$ is applied to each point and the mean $\bar{\mathbf{y}}$ and covariance $\mathbf{P}_y$ of these transformed points are calculated. Effectively the unscented transform works by examining the changes in the mean and covariance of the random variable after the nonlinear transformation.

If $\mathbf{x}$ has n dimensions, it is approximated by $2n + 1$ sigma points $\underline{\mathcal{X}}$. These are selected as follows:

$$\begin{aligned}
\mathcal{X}_0 &= \bar{\mathbf{x}} \\
\mathcal{X}_i &= \bar{\mathbf{x}} + \left(\sqrt{(n+\lambda)\mathbf{P}_x)}\right)_i \text{ where } i = 1, ..., n \\
\mathcal{X}_{i+n} &= \bar{\mathbf{x}} - \left(\sqrt{(n+\lambda)\mathbf{P}_x)}\right)_i \text{ where } i = n+1, ..., 2n
\end{aligned}$$

with each sigma point $\mathcal{X}_i$ having an associated weight W_i

$$\begin{aligned}
W_0 &= \frac{\lambda}{n+\lambda} \\
W_i &= \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) \\
W_{i+n} &= \frac{1}{2(n+\lambda)}
\end{aligned}$$

where $\lambda = \alpha^2(n+\kappa) - n$. Note that this representation of the sigma points is a modern update, given for example in Wan & van der Merwe (2000). The authors point out that α determines the spread of the sigma points around $\bar{\mathbf{x}}$ and has a default of 10^{-3}. κ is a secondary scaling parameter usually set to zero, whereas β is used to incorporate prior knowledge of the distribution of $\mathbf{x}$; $\beta = 2$ is said to be optimal for Gaussian distributions (Wan & van der Merwe, 2000). The unscented transform works by using these sigma points as follows:

1. Transform each sigma point through the nonlinear function: $\mathcal{Y}_i = \mathbf{f}(\mathcal{X}_i)$.

2. Calculate the mean of the new distribution using the weighted average of the transformed sigma points: $\bar{\mathbf{y}} = \Sigma_{i=0}^{2n} W_i \mathcal{Y}_i$.

3. Calculate the covariance of the new distribution using the weighted outer product of the transformed sigma points: $\mathbf{P}_y = \Sigma_{i=0}^{2n} W_i (\mathcal{Y}_i - \bar{\mathbf{y}})(\mathcal{Y}_i - \bar{\mathbf{y}})^T$.

Unscented Kalman filter algorithm

When both the noise terms of the nonlinear state-space model are additive, as in (5.2.1)-(5.2.2), the full UKF algorithm can be simplified slightly. We present this algorithm here, as can be found in Haykin et al. (2001) chapter 7 - written by Wan & van der Merwe. For terms of the UKF that correspond to terms of the standard Kalman filter in section 4.2.3, we use similar notation for consistency.

The initial state $\mathbf{x}_0$ is a random vector with known mean $\mathbf{m}_0 = \mathbb{E}[\mathbf{x}_0]$ and covariance $\mathbf{C}_0 = \mathbb{E}[(\mathbf{x}_0 - \mathbf{m}_0)(\mathbf{x}_0 - \mathbf{m}_0)^T]$. For $t = 1, \ldots, \infty$:

1. Calculate sigma points:

$$\mathcal{X}_{t-1} = \left[\mathbf{m}_{t-1} \quad \mathbf{m}_{t-1} + \gamma\sqrt{\mathbf{C}_{t-1}} \quad \mathbf{m}_{t-1} - \gamma\sqrt{\mathbf{C}_{t-1}}\right] \tag{5.2.3}$$

   Note that γ is defined as $\gamma = \sqrt{n + \lambda}$.

2. Time update:

$$\begin{aligned}
\mathcal{X}_{t|t-1} &= f\left[\mathcal{X}_{t-1}, \mathbf{u}_{t-1}\right] && (5.2.4)\\
\mathbf{a}_t &= \sum_{i=0}^{2n} W_i \mathcal{X}_{i,t|t-1} && (5.2.5)\\
\mathbf{V}_t &= \sum_{i=0}^{2n} W_i \left[\mathcal{X}_{i,t|t-1} - \mathbf{a}_t\right]\left[\mathcal{X}_{i,t|t-1} - \mathbf{a}_t\right]^T + Q && (5.2.6)\\
\mathcal{Y}_{t|t-1} &= g\left[\mathcal{X}_{t|t-1}\right] && (5.2.7)\\
\mathbf{f}_t &= \sum_{i=0}^{2n} W_i \mathcal{Y}_{i,t|t-1} && (5.2.8)
\end{aligned}$$

3. Measurement update:

$$\mathbf{W}_t = \sum_{i=0}^{2n} W_i \left[\mathcal{Y}_{i,t|t-1} - \mathbf{f}_t\right] \left[\mathcal{Y}_{i,t|t-1} - \mathbf{f}_t\right]^T + R \tag{5.2.9}$$

$$\mathbf{C}_{x_t y_t} = \sum_{i=0}^{2n} W_i \left[\mathcal{X}_{i,t|t-1} - \mathbf{a}_t\right] \left[\mathcal{Y}_{i,t|t-1} - \mathbf{f}_t\right]^T \tag{5.2.10}$$

$$\mathcal{K}_t = \mathbf{C}_{x_t y_t} \mathbf{W}_t^{-1} \tag{5.2.11}$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathcal{K}_t(\mathbf{y}_t - \mathbf{f}_t) \tag{5.2.12}$$

$$\mathbf{C}_t = \mathbf{V}_t - \mathcal{K}_t \mathbf{W}_t \mathcal{K}_t^T \tag{5.2.13}$$

For the UKF, the unscented transform is effectively incorporated whenever a nonlinear function is present in the state or observation equation. Therefore, if functions $f(.)$ and $g(.)$ of the state-space model in (5.2.1)-(5.2.2) are both nonlinear, i.e. we have nonlinear state and observation equations, the unscented transform is performed in both the prediction and correction steps of the filter. However, later in this chapter, when we form state-space models from hydraulics models 2 and 3, nonlinear functions will be present in the state equations only. That is, for (5.2.1)-(5.2.2), $f(.)$ will be nonlinear and $g(.)$ will be linear. In fact, $g(.)$ will be the identity function. In these cases, we can simplify the UKF algorithm further and skip steps (5.2.7)-(5.2.8), since $\mathcal{Y}_{t|t-1} = \mathcal{X}_{t|t-1}$ and $\mathbf{f}_t = \mathbf{a}_t$.

5.3 Applying FIMLOF to hydraulics model 2

In this section we incorporate the unscented Kalman filter into a modified FIMLOF algorithm and attempt parameter estimation on hydraulics model 2. Section 5.3.1 presents the details of hydraulics model 2 and highlights the key differences between

this and hydraulics model 1. We derive the difference equations for hydraulics model 2 in section 5.3.2 and form the resulting nonlinear state-space model in section 5.3.3. Details of modified FIMLOF algorithm 2 are provided in 5.3.4 and results of the simulation study in section 5.3.5. In section 5.3.6 we investigate the accuracy of FIMLOF algorithm 2 and section 5.3.7 evaluates the assumption of diagonal covariance matrices.

5.3.1 Hydraulics model 2 details

Hydraulics model 2 has a similar structure to hydraulics model 1 and as such, can be represented using the same stock and flow diagram in Figure 4.3.1. However there are two key differences between this model and hydraulics model 1. The first of these is that certain variables use values that are smoothed over the previous 7 days. This is to enable the model to respond appropriately to job arrivals that exhibit a weekly profile. This profile was presented in chapter 3 and is typical of the pattern of repair job arrivals to the BT system. This profile consists of a repeating sequence of 7 constants that describe the expected relative numbers of job arrivals for each day of the week. Without the adjustment for this profile, the model would respond to the weekly profile in the job arrivals, rather than a trend in their behaviour. All difference equations in the model are the same as for model 1, except for the following three

which replace the terms b_t, c_t and n_t with $s(b_t)$, $s(c_t)$ and $s(n_t)$:

$$\begin{aligned} tct_t &= \frac{s(b_t)}{\theta_t} \\ tcr_t &= \theta_w(tct_t) + (1-\theta_w)(s(n_t)) \\ pf_t &= \frac{tcr_t}{s(c_t)}. \end{aligned}$$

The same notation as for hydraulics model 1 in table 4.3.1 is used. The function $s()$ denotes smoothing over the previous 7 days. For example, smoothing the backlog would take the form $s(b_t) = \frac{1}{7}\sum_{j=t-6}^{t} b_j$. The second difference between this model and hydraulics model 1 is how the cleared jobs are modelled. We see from the difference equations of hydraulics model 1 that the cleared jobs are simply the capacity deployed: $cl_t = c_t$. For model 2, the weekly profile is incorporated into this equation as follows:

$$cl_t = wp_t\, c_t,$$

where wp_t is the weekly profile on day t. This effectively gives the model a simple version of a roster profile, so that it can attempt to approximately match the job arrivals for that day when clearing jobs.

5.3.2 Derivation of difference equations

In order to implement the FIMLOF method we must represent the structure of hydraulics model 2 as a state-space model. We follow the novel process, outlined for hydraulics model 1 in section 4.3.2, in order to formulate a state-space model from a SD model. This consists of grouping the SD variables, determining the order in which

each should update and then performing substitutions to form Markovian difference equations. The addition of state and observational noise terms forms the state-space model in section 5.3.3. We extend this process here to a nonlinear system.

As the structure of hydraulics model 2 is similar to model 1, we already know how to group the variables into stocks, exogenous variables and auxiliary variables. From section 4.3.2, we also know the order in which the auxiliary variables update within the model. All that remains is to perform the necessary substitutions in order to represent the stocks in the model as Markovian difference equations. As in section 4.3.2, this again requires the stocks to be represented in terms of only the stocks, model parameters $\underline{\theta}$ and exogenous variables. Since section 5.3.1 revealed differences between the model equations of hydraulics models 1 and 2, we cannot use the same difference equations as for model 1 and must again substitute out the auxiliary variables in order to obtain the difference equations for model 2.

The equations of the stocks for hydraulics model 2 are similar to those for hydraulics model 1 and take the following form:

$$b_t = b_{t-1} + n_t - wp_t\, c_{t-1} \tag{5.3.1}$$

$$c_t = c_{t-1} + cc_{t-1}. \tag{5.3.2}$$

(5.3.1) now includes the weekly profile term, wp. This changes its value for each day of the week and as such, can be considered as an additional exogenous variable. Hence, like for model 1, (5.3.1) requires no substitutions, since the backlog is already

expressed in the required form. (5.3.2) is unchanged from model 1 as we see in its equivalent in (4.3.2). The changing capacity auxiliary variable cc must again be substituted out. This is achieved as follows:

$$
\begin{aligned}
cc_t &= \frac{c_t}{\theta_d}(pf_t - 1) && (5.3.3)\\
&= \frac{c_t}{\theta_d}\left[\left(\frac{tcr_t}{s(c_t)}\right) - 1\right] && (5.3.4)\\
&= \frac{c_t}{\theta_d}\left[\left(\frac{\theta_w(tct_t) + (1-\theta_w)(s(n_t))}{s(c_t)}\right) - 1\right] && (5.3.5)\\
&= \frac{c_t}{\theta_d}\left[\left(\frac{\theta_w(\frac{s(b_t)}{\theta_t}) + (1-\theta_w)(s(n_t))}{s(c_t)}\right) - 1\right] && (5.3.6)
\end{aligned}
$$

(5.3.3)-(5.3.6) are obtained by substituting in the hydraulics model 2 terms for pf_t, tcr_t and tct_t respectively. (5.3.6) now represents cc in the required form. We substitute this form of cc into (5.3.2) and rearrange as follows:

$$
\begin{aligned}
c_t &= c_{t-1} + cc_{t-1}\\
&= c_{t-1} + \frac{c_{t-1}}{\theta_d}\left[\left(\frac{\theta_w(\frac{s(b_{t-1})}{\theta_t}) + (1-\theta_w)(s(n_t))}{s(c_{t-1})}\right) - 1\right]\\
&= c_{t-1} + \frac{c_{t-1}}{\theta_d}\left(\frac{\theta_w(\frac{s(b_{t-1})}{\theta_t}) + (1-\theta_w)(s(n_t))}{s(c_{t-1})}\right) - \frac{c_{t-1}}{\theta_d}\\
&= c_{t-1} + \frac{c_{t-1}}{s(c_{t-1})}\left(\frac{\theta_w(\frac{s(b_{t-1})}{\theta_t}) + (1-\theta_w)(s(n_t))}{\theta_d}\right) - \frac{c_{t-1}}{\theta_d}\\
&= c_{t-1} + \frac{c_{t-1}}{s(c_{t-1})}\left[\left(\frac{\theta_w}{\theta_d\theta_t}\right)\, s(b_{t-1}) + \left(\frac{1-\theta_w}{\theta_d}\right)\, s(n_t)\right] - \frac{c_{t-1}}{\theta_d}\\
&= c_{t-1}\left(1 - \frac{1}{\theta_d} + \frac{1}{s(c_{t-1})}\left[\left(\frac{\theta_w}{\theta_d\theta_t}\right)\, s(b_{t-1}) + \left(\frac{1-\theta_w}{\theta_d}\right)\, s(n_t)\right]\right)\\
&= c_{t-1}(\phi_2 + \frac{1}{s(c_{t-1})}[\phi_1\, s(b_{t-1}) + \phi_3\, s(n_t)])\\
&= f_c(b_{t-1}, c_{t-1}, n_t, \underline{\phi}),
\end{aligned}
$$

where ϕ_1, ϕ_2 and ϕ_3 are as defined in (4.3.7)-(4.3.9). The function f_c represents the updating function for the capacity deployed. Two things are apparent from this

representation of hydraulics model 2. Firstly, f_c is not a linear function of the stocks and exogenous variables. It is because of this nonlinearity that we will need a modified version of the Kalman filter. Secondly, the system no longer appears to be Markovian, since the smoothed terms rely on states as far back as 6 time intervals. However, on closer inspection we see that this is not actually the case, and the Markov property holds. This is because the smoothing variables $s(b_t), s(c_t)$ and $s(n_t)$ are contained within the underlying model, in this case hydraulics model 2. In the implementation of hydraulics model 2, these smoothing variables rely on additional variables to store the average of the previous 6 days of the relevant variable(s), so that when the relevant variable(s) update, only the new updated value is required to update the smoothing variable. In other words, for the backlog, the smoothing variable is calculated as $s(b_t) = \frac{1}{7}\sum_{j=t-6}^{t-1} b_j + \frac{1}{7}b_t$, where an additional variable tracks the behaviour of b_t over the previous 6 days and hence calculates $\frac{1}{7}\sum_{j=t-6}^{t-1} b_j$ at each update. The final difference equations of hydraulics model 2 are shown in (5.3.7)-(5.3.8).

$$b_t = b_{t-1} + n_t - wp_t\, c_{t-1} \tag{5.3.7}$$

$$c_t = c_{t-1}(\phi_2 + \frac{1}{s(c_{t-1})}[\phi_1\, s(b_{t-1}) + \phi_3\, s(n_t)]) \tag{5.3.8}$$

These can be used as the basis of a state-space model. This is formed in the next section.

5.3.3 Forming the state-space model

We now use difference equations (5.3.7)-(5.3.8) constructed in the previous section to form a state-space model. In (4.2.3)-(4.2.4) of section 4.2.2 we gave the general

form for a Gaussian state-space model. Due to the nonlinear nature of (5.3.8), the state-space model formed from hydraulics model 2 in this section will have a nonlinear state update function $f(.)$.

To form the state-space model we follow the same approach as for hydraulics model 1 and add Gaussian noise terms to difference equations (5.3.7)-(5.3.8). This gives us the state update equations of the model, shown in (5.3.9)-(5.3.12). As for hydraulics model 1, noise terms $\underline{v}_1, \underline{v}_2$ have a Gaussian distribution represented as follows: $\underline{v}_1 \sim N(0, \sigma^2_{Q_1})$, $\underline{v}_2 \sim N(0, \sigma^2_{Q_2})$.

$$b_t = b_{t-1} + n_t - wp_t\, c_{t-1} + v_{1,t} \tag{5.3.9}$$

$$= f_1(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) + v_{1,t} \tag{5.3.10}$$

$$c_t = c_{t-1}(\phi_2 + \frac{1}{s(c_{t-1})}[\phi_1\, s(b_{t-1}) + \phi_3\, s(n_t)]) + v_{2,t} \tag{5.3.11}$$

$$= f_2(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) + v_{2,t} \tag{5.3.12}$$

We again assume that Gaussian observation noise is present. The observation equations of the state-space model in (5.3.13)-(5.3.14) are formed by adding further noise terms.

$$b^*_t = b_t + w_{1,t} \tag{5.3.13}$$

$$c^*_t = c_t + w_{2,t} \tag{5.3.14}$$

These observational noise terms $\underline{w}_1$ and $\underline{w}_2$ are as follows: $\underline{w}_1 \sim N(0, \sigma^2_{R_1})$, $\underline{w}_2 \sim N(0, \sigma^2_{R_2})$. Covariance matrices Q and R are as before so that again we have:

$$\underline{y}_t = \begin{bmatrix} b_t^* \\ \\ c_t^* \end{bmatrix}, \ \underline{x}_t = \begin{bmatrix} b_t \\ \\ c_t \end{bmatrix}, \underline{u}_t = \begin{bmatrix} n_t \\ \\ n_t \end{bmatrix}, R = \begin{bmatrix} \sigma_{R_1}^2 & 0 \\ \\ 0 & \sigma_{R_2}^2 \end{bmatrix}, \ Q = \begin{bmatrix} \sigma_{Q_1}^2 & 0 \\ \\ 0 & \sigma_{Q_2}^2 \end{bmatrix}.$$

When the unscented Kalman filter is used, the updating functions $\underline{f}()$ for the states in (5.3.9)-(5.3.12), are represented by the difference equations (5.3.7)-(5.3.8). The equivalent update functions for the observations are simply the identity $\underline{g}(x) = x$ since (5.3.13)-(5.3.14) are not a transformation of the states and simply add noise. Hence for the simulation study in section 5.3.5, as we explain in section 5.2.2, we only require the unscented transform to be incorporated in the prediction step of the Kalman filter. The correction step can be computed as for the standard Kalman filter.

In the next section we outline the modified FIMLOF algorithm that will be used in the simulation study in section 5.3.5.

5.3.4 Modified FIMLOF algorithm 2

Having formed the state-space model using the structure of hydraulics model 2, we are now ready to present an algorithm for implementing modifications to the FIMLOF method.

For noisy observable time series data $\underline{y}$ for the backlog and capacity deployed, we estimate the structural parameters $\underline{\theta} = (\theta_d, \theta_w, \theta_t)$ and the variance parameters $\underline{\sigma}^2 = (\sigma_{Q_1}^2, \sigma_{Q_2}^2, \sigma_{R_1}^2, \sigma_{R_2}^2)$ of the state-space model in (5.3.9)-(5.3.14) using the algorithm

below. We refer to this as algorithm 2.

1. The Nelder-Mead simplex method selects a candidate set of parameters $(\underline{\theta}, \underline{\sigma}^2)$, (or in the first iteration the user selects suitable starting values $\underline{\theta} = \underline{\theta}_0$ and $\underline{\sigma}^2 = \underline{\sigma}_0^2$).

2. $\underline{\phi} = \underline{g}(\underline{\theta})$ is calculated using (4.3.7)-(4.3.9).

3. The unscented Kalman filter proceeds along time series $\underline{y}$ and for each update at time t:

   - The $\underline{\phi}$ parameters in the nonlinear state update (5.3.9) and the variance parameters $\underline{\sigma}^2$ of covariance matrices Q and R influence the filter's estimate of latent states $\underline{x}_t$.
   - Terms from these calculations are used to calculate the log-likelihood (using (4.2.15) for the approximate representation) for the candidate set of parameters, $\log(L(\underline{\theta}, \underline{\sigma}^2), t)$.

4. After the Kalman filter has calculated the log-likelihood at each discrete interval t in the series, an overall log-likelihood for the time series is calculated by summing the individual log-likelihood terms: $\log(L(\underline{\theta}, \underline{\sigma}^2)) = \Sigma_{t=1}^{T} \log(L(\underline{\theta}, \underline{\sigma}^2, t))$.

5. The Nelder-Mead method repeats the process until a local optimum is found and maximum likelihood estimates $\underline{\hat{\phi}}$ and $\underline{\hat{\sigma}}^2$ are determined.

Having obtained the maximum likelihood estimates (MLEs) $\underline{\hat{\phi}}$, we calculate the MLEs for structural parameters $\underline{\hat{\theta}} = (\hat{\theta}_d, \hat{\theta}_w, \hat{\theta}_t)$ as before using the inverse of the relevant

$g()$ function in (4.3.7)-(4.3.9). Note that the approximation for the log-likelihood of (4.2.15) is again chosen over the exact representation, due to improved parameter estimates.

In simulation study 2 in section 5.3.5, artificial noisy data $\underline{y}$ is simulated from hydraulics model 2 with known structural and variance parameters. This algorithm is used for the estimation of these parameters and its performance is assessed.

5.3.5 Simulation study 2

In this section we test the modified FIMLOF approach of algorithm 2 for estimating parameters of noisy data simulated from hydraulics model 2. This model is similar in complexity to hydraulics model 1, with the important difference of being a nonlinear system. Success in this study can be viewed as a further step towards chapter 6 when we estimate parameters of historical BT data. The objectives of the study are as follows:

- Assess performance of algorithm 2 in estimating the structural parameters $\underline{\theta}$ of noisy data from hydraulics model 2.
- The effects of system noise σ_Q and observational noise σ_R on the accuracy of the parameter estimates.

In this study, the differences in behaviour between hydraulics models 1 and 2 meant that in order to simulate realistic data from hydraulics model 2, the value of the delay parameter θ_d was changed from 3 to 5. Consequently, different starting values were

required for θ_d in the optimisation routine. The 5 sets of starting values for $\underline{\theta}$ in simulation study 2 are as shown in table 5.3.1.

Parameter	s1	s2	s3	s4	s5
θ_d	4	4.5	5	5.5	6
θ_w	0.1	0.15	0.2	0.25	0.3
θ_t	1	1.5	2	2.5	3

Table 5.3.1: Starting values for the optimisation in experiments 1 and 2.

These were the only changes from simulation study 1 in section 4.4; the study in this section follows the same format. 100 sets of data are simulated for each set of parameters and 5 sets of starting values are used in the optimisation routine; giving 500 parameter estimates. When simulating data from hydraulics model 2, run lengths were again set at 500. The same 2 step job arrivals scenario as simulation study 1 was used to simulate data and this study has the same two experiments as simulation study 1. When simulating artificial data from hydraulics model 2 and adding noise, experiment 1 fixes $\sigma_R = 10$ and adjusts $\sigma_Q = 1, 5, 10, 20, 50, 80$. In experiment 2 we fix $\sigma_Q = 10$ and adjust $\sigma_R = 1, 5, 10, 20, 50, 80$. Histograms of parameter estimates from experiments 1 and 2 are shown in Figures 5.3.1 and 5.3.2 respectively.

To compare the results from experiments 1 and 2, as for simulation study 1 we examine the relative difference of each of the parameter estimates compared to the true value

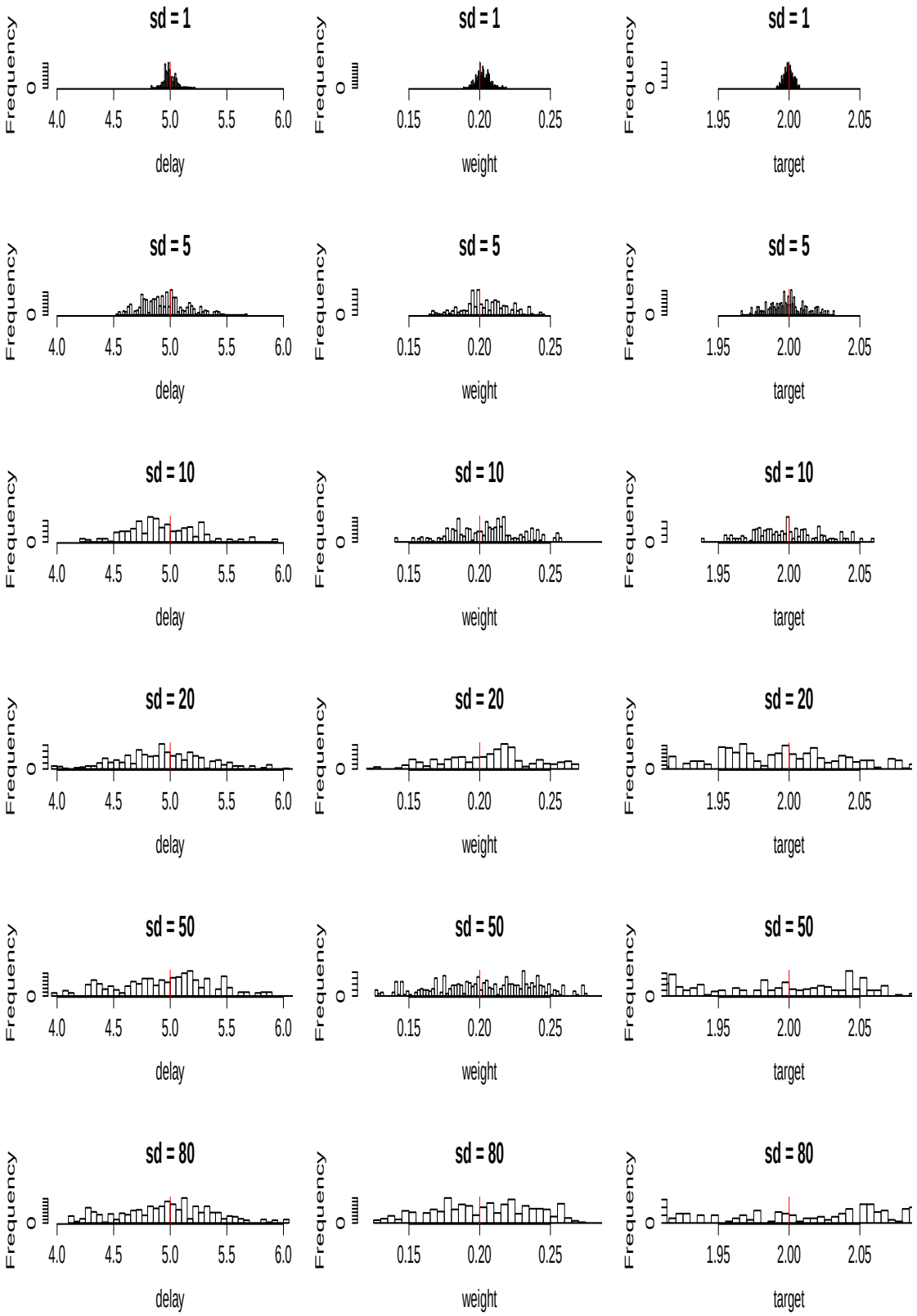

Figure 5.3.1: Experiment 1 histograms of parameter estimates.

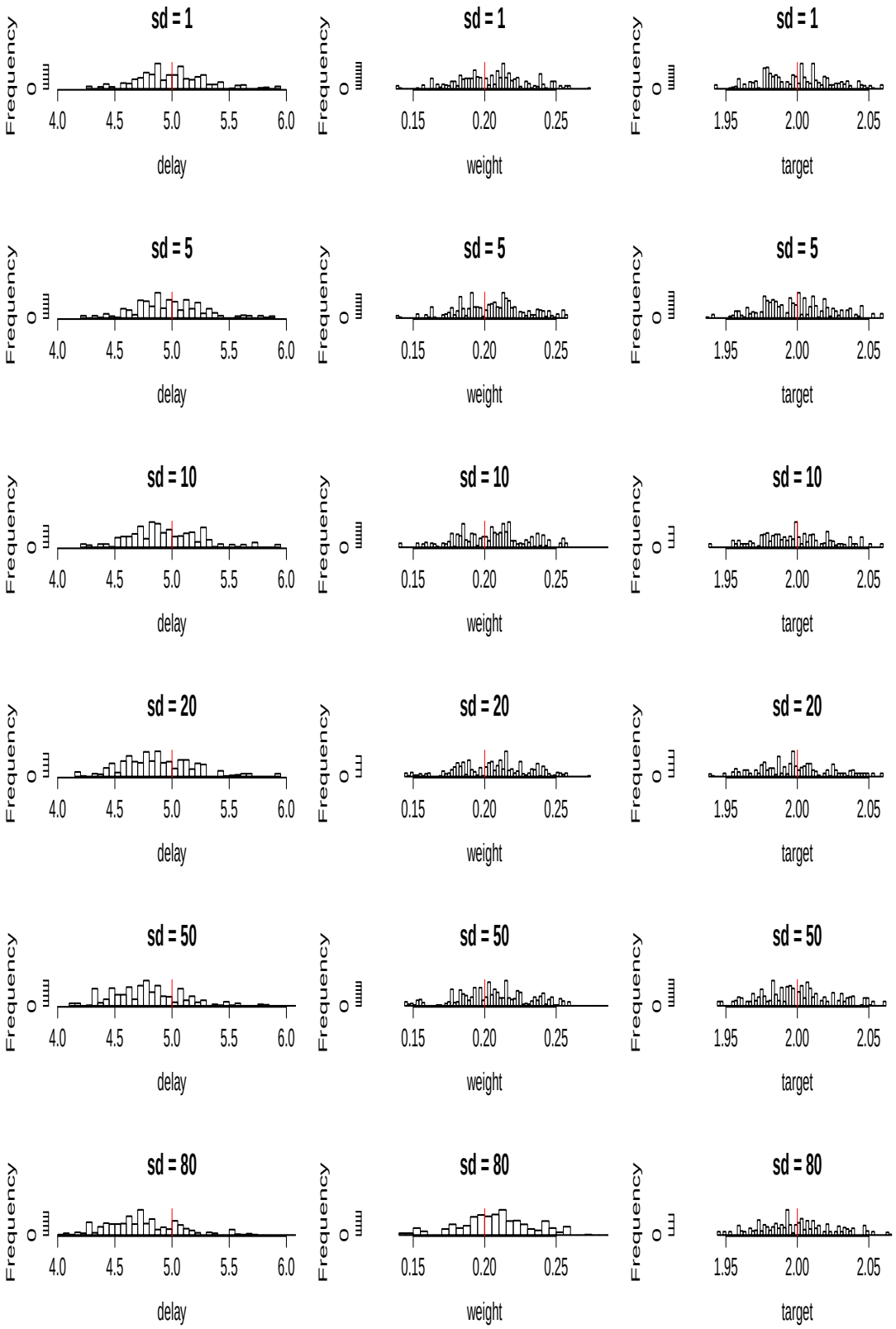

Figure 5.3.2: Experiment 2 histograms of parameter estimates.

by calculating the mean absolute percentage error (MAPE). Let $\hat{\theta}_i$ be the estimate of true parameter θ_i, for $i = d, w, t$. Then $\tau_i = \frac{|\hat{\theta}_i - \theta_i|}{\theta_i}$, for $i = d, w, t$, is the absolute value of the relative difference between $\hat{\theta}_i$ and θ_i. $\overline{\tau}_i = 100 \times \frac{1}{n} \sum_1^n \tau_i$, where $n = 500$, again represents the MAPE. The values of $\overline{\tau}_i$ for experiments 1 and 2 are shown in table 5.3.2.

	Expt 1 ($\sigma_R = 10$)				Expt 2 ($\sigma_Q = 10$)			
Parameter set	σ_Q	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$	σ_R	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	1	4.0	0.4	0.3	1	23.1	2.0	2.0
2	5	16.3	1.3	1.0	5	23.3	2.0	2.0
3	10	25.3	2.0	2.0	10	25.3	2.0	2.0
4	20	31.4	2.6	4.0	20	27.7	2.1	2.0
5	50	34.6	3.0	9.8	50	32.1	2.1	2.1
6	80	33.0	3.1	15.6	80	34.9	2.1	2.0

Table 5.3.2: MAPE of parameter estimates for experiments 1 and 2.

From table 5.3.2 and the histograms in Figures 5.3.1 - 5.3.2, the following observations are immediately apparent:

1. For each experiment, increasing the noise added to the data results in generally less accurate parameter estimates.

2. Parameter estimates are less accurate overall than for hydraulics model 1.

3. Across all parameter sets, whichever experiment has the highest σ_Q has the least accurate estimates. The only exception to this are the θ_d estimates of parameter set 6. Increasing σ_Q has a considerable effect on the accuracy of estimates whereas increasing σ_R has a relatively small effect, especially for θ_w and θ_t which hardly change. In other words parameter estimation is less robust to added state noise σ_Q than added observational noise σ_R – which is the opposite of simulation study 1.

4. θ_d is again the least accurately estimated parameter, but the difference in accuracy between θ_w and θ_t has disappeared.

As we saw in simulation study 1, observation (1) is not surprising. As more noise is added to the simulated data, its structure becomes more corrupted so that it represents hydraulics model 2 to a lesser extent. Observation (2) is a consequence of the nonlinear system exhibited by hydraulics model 2. In simulation study 1, the simulated data from hydraulics model 1 possessed a linear structure and hence the Kalman filter was able to compute an optimal estimate of the underlying state. In this study the UKF is only able to find an approximation of this underlying state. As is the case when adding greater amounts of noise, when the estimate of the latent state is less accurate, estimates of the structural parameters $\underline{\theta}$ are likely to be less accurate.

Observation (3) is perhaps the most interesting. The amount of state noise σ_Q that is added to the data appears to be the main factor in the accuracy of the resulting

parameter estimates. In experiment 1 when σ_Q is increased from 1 up to 80, the accuracy of estimates decays considerably, especially for θ_d, but also for θ_t which was consistently estimated accurately in simulation study 1. It is clear that our results are the opposite of simulation study 1 where a larger signal to noise ratio (σ_Q/σ_R) in the simulated data produced more accurate estimates. For this study, it is a smaller signal to noise ratio that produces more accurate estimates.

One possible explanation of this observation is that the nonlinear equation in the system is one of the state equations, (5.3.11). The observation equations in the state-space model (5.3.13)-(5.3.14) are both linear. As we explained previously it is therefore for the state equations of the state-space model, i.e. the prediction step, that the unscented transform is required within the UKF. When using the unscented transform, the UKF's estimate of the latent state of the noisy data is not as accurate as when using the standard Kalman filter on a linear Gaussian system – which is optimal in terms of minimising the mean squared error. Since the nonlinearity lies in the state equations, adding state noise σ_Q to these may cause the more serious effect on the accuracy of parameter estimates, compared to observational noise σ_R. Hence the approximations made by the UKF for this system are more robust to observational noise than state noise. Another explanation may be that the differences in the likelihood surfaces that are produced when applying FIMLOF to hydraulics models 1 and 2 cause the Nelder-Mead method to get stuck in a local minima – and the location of these local minima are affected differently by state and observational noise for the different models.

Unlike σ_Q, added observational noise σ_R appears to have a relatively small effect on the accuracy of parameter estimates. In experiment 2 when σ_Q is fixed, adjusting the value of σ_R up to 20 appears to have little effect on the accuracy of estimates. Even when $\sigma_R > 20$, estimates of θ_w and θ_t are largely unchanged from when $\sigma_R = 1$. These results can be viewed in two ways. On the one hand, experiment 2 provides evidence for claiming that the approach has some robustness to observational noise σ_R. However, it is more likely that the relatively unchanging results of experiment 2 can be explained by the dominant effect of σ_Q. Setting this to a relatively low value of 10 has such a significant effect on the accuracy of parameter estimates that changes in σ_R go almost unnoticed. The considerable changes in experiment 1 as σ_Q increases are further support of this. To investigate further the dominant effect of σ_Q, an additional experiment was conducted that fixed $\sigma_Q = 0$ and σ_R again varied from 1 up to 80. The results of this (experiment 3) are given in table 5.3.3. The considerable improvements in these results demonstrate further the effect of the added state noise σ_Q – and also the relative robustness of our approach to observational noise σ_R.

Observation (4) can again be explained by the concept of identification of parameters, discussed in section 4.4. It is interesting to observe the sudden drop in performance of θ_t compared to simulation study 1 where it was estimated considerably better than both θ_d and θ_w. For experiment 2 of this study, θ_t and θ_w are estimated to a similar degree of accuracy but for experiment 1, estimates of θ_t are less accurate when $\sigma_Q \geq 20$. Although hydraulics model 2 is similar to model 1, it is worth highlighting that they

	Expt 3 ($\sigma_Q = 0$)			
Parameter set	σ_R	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	1	1.6	0.3	0.1
2	5	2.3	0.4	0.2
3	10	1.1	0.2	0.1
4	20	0.7	0.2	0.0
5	50	1.5	0.2	0.1
6	80	2.5	0.3	0.1

Table 5.3.3: MAPE of parameter estimates for experiment 3.

are still entirely different systems, as their difference equations (5.3.7)-(5.3.8) demonstrate. As such, the identifiability of the parameters in each model is likely to be different.

As for simulation study 1, although the values of the estimates of the structural parameters $\underline{\theta}$ are affected by the choice of the 5 sets of starting values in the optimisation routine, the accuracy of those estimates is not affected. The same variance starting values as simulation study 1 were used in the optimisation routine, i.e. $\sigma_{Q_0} = \sigma_{R_0} = e^{-10}$, despite the sometimes high amounts of artificial noise. Like simulation study 1, although we assumed that we knew the correct noise structure (the diagonal nature of the covariance matrices), we made no assumptions regarding the scale of this noise by using starting values that were close to zero. The results obtained

are less accurate than simulation study 1, but nonetheless demonstrate the effectiveness of the approach in estimating the parameters of a nonlinear system. Although for data simulated from hydraulics model 2 our approach is less robust to added state noise σ_Q, the results for experiment 3 are particularly promising. As we explain in sections 4.2.5 and 4.2.6, a number of researchers, e.g. Eberlein and Wang (1985), have highlighted the importance of the variance terms in FIMLOF and their effect on parameter estimates. The ability to estimate parameters without prior knowledge of the scale of the added noise, as we have shown in this study and in simulation study 1, is therefore an important achievement.

Using our modified FIMLOF algorithm we have been able to estimate parameters of a nonlinear system, without prior knowledge of the noise structure. The dominant effect of added state noise σ_Q to reduce the accuracy of parameter estimates has been revealed. Without the effect of this noise in experiment 3, estimates are considerably more accurate. The progression to the nonlinear system of hydraulics model 2 from the linear system of hydraulics model 1 can be viewed as a further step towards estimating parameters of the historical BT time series in chapter 6, which is a highly nonlinear system.

In section 5.3.6, we investigate how much accuracy is lost through each of the two approximations of algorithm 2; the UKF's approximation for a nonlinear system and the parameter estimation process.

5.3.6 Investigating the accuracy of FIMLOF algorithm 2

For a linear Gaussian system, the standard Kalman filter introduced in section 4.2.3 gives an optimal estimate of the system's underlying state, when the model parameters and initial conditions are known. Hence for simulation study 1 in section 4.4, the only approximation made by modified FIMLOF algorithm 1 was in estimating the maximum likelihood estimates of parameters via the optimisation part of the algorithm - the Nelder-Mead simplex method. However, for a nonlinear Gaussian system, the UKF's estimate of a system's underlying state is not optimal. As we explain in section 5.2.1, the UKF attempts to find the closest Gaussian approximation to a probability distribution. Therefore, modified FIMLOF algorithm 2, used in simulation study 2, includes two approximations - the optimisation part of the algorithm and an additional approximation made by the UKF. In this section we investigate the accuracy of these two approximations through a simulation study.

In order to investigate the two approximations made by algorithm 2, we fit the UKF to noisy data simulated from hydraulics model 2 under two different conditions - using known parameters and using estimated parameters. The accuracy of the UKF's estimation of the latent states of the time series is determined for each of these two conditions and compared. When simulating noisy data from hydraulics model 2, we simulate three sets of time series, each using a different set of structural parameters. This is to ensure that any conclusions are robust to different sets of hydraulics model 2 parameters. These sets of parameters are presented in table 5.3.4. The variance pa-

rameters are the same for all three datasets, with the standard deviations as follows: $\sigma_Q = 10$ and $\sigma_R = 10$. As in simulation studies 1 and 2, run lengths are set at 500 days and the same 2 step job arrivals scenario is used.

Parameter set	θ_d	θ_w	θ_t
1	3	0.1	1
2	5	0.2	2
3	7	0.3	3

Table 5.3.4: Parameter sets for investigating UKF's approximation.

We describe below the process followed for each of the three sets of data simulated from hydraulics model 2, when using known and estimated parameters when fitting the UKF.

- Known parameters: The UKF is fitted to the noisy time series using the known parameters $\underline{\theta}$. The mean squared error (MSE) of the residuals is calculated. The residuals are the differences between the latent states of the noisy time series and the state estimate from the UKF - at each interval in time. When using known parameters in the UKF, we have only a single approximation in this process - the approximation made by the UKF. Hence these results are the benchmark, providing a measure of how accurately the UKF is able to approximate the nonlinear system when the parameters are known.

- Estimated parameters: 100 sets of starting values are simulated from $\mathcal{N}(\mu_i, \sigma_i)$,

with $\mu_i = \theta_i$ for $i = d, w, t$; where θ_i is the true parameter used in hydraulics model 2. $\sigma_i = 0.5,\ 0.05,\ 0.25$ for $i = d, w, t$ respectively. For example, to simulate starting values in order to estimate θ_d of parameter set 1 in table 5.3.4, we use $\mathcal{N}(3, 0.5)$. The 100 sets of starting values are used in algorithm 2 to obtain 100 sets of parameter estimates $\underline{\hat{\theta}}$. Each of these sets of parameters is then used within the UKF which is fitted to the time series, with the MSE of the residuals (defined as above) and log-likelihood being calculated. When using estimated parameters in the UKF, we have the additional approximation from the optimisation routine. Hence these results reveal how much additional accuracy is lost via our parameter estimation method.

To measure the accuracy of the estimated parameters, as for simulation studies 1 and 2 we again calculate the MAPE. Let $\hat{\theta}_i$ be the estimate of true parameter θ_i, for $i = d, w, t$. Then $\tau_i = \frac{|\hat{\theta}_i - \theta_i|}{\theta_i}$, for $i = d, w, t$, is the absolute value of the relative difference between $\hat{\theta}_i$ and θ_i. However, since $n = 100$, now we have $\overline{\tau}_i = 100 \times \frac{1}{n} \sum_1^n \tau_i = \sum_1^n \tau_i$ representing the MAPE. The values of $\overline{\tau}_i$ for the estimated parameters are shown in table 5.3.5. Across the three parameter sets, θ_t is the most accurately estimated parameter, as was the case for simulation study 2. Although we see noticeable differences in the accuracy of θ_d and θ_w if we compare these results to those from parameter set 3 of simulation study 2 (which also sets $\sigma_Q = \sigma_R = 10$), it is worth highlighting that these results are not equivalent. The results of this section are based on 100 parameter estimates rather than 500 and different starting values are used, with the simulated starting values of this section

spanning a wider range than those selected for simulation study 2.

Parameter set	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
1	5.3	15.0	0.9
2	10.7	3.1	0.8
3	9.7	9.5	0.7

Table 5.3.5: MAPE for estimated parameters.

The 100 sets of estimated parameters $\underline{\hat{\theta}}$ are used in the UKF which is fitted to the three sets of time series. The results are compared in table 5.3.6 with results from using the known parameters $\underline{\theta}$ in the UKF. Results for the estimated parameters - in rows 2, 4 and 6 - show the mean of the resulting 100 log-likelihoods, denoted by $\overline{l(\underline{\hat{\theta}})}$, and the mean of the 100 sets of MSE's of the residuals, denoted by $\overline{MSE}$. These are compared with results for the known parameters in rows 1, 3 and 5. Note that results using the known parameters show the single resulting log-likelihood $l(\underline{\theta})$ and MSE values - although the columns of table 5.3.6 are labelled as for the unknown UKF parameters.

When using estimated parameters, $\overline{MSE}$ total increases by around 45%, compared to the MSE total values resulting from using the known parameters. This increase in $\overline{MSE}$ total is entirely due to the effect of $\overline{MSE}$ backlog, which increases by around 90%. $\overline{MSE}$ capacity actually decreases by around 10%, contrary to what would be

Param set	UKF params	$\overline{MSE}$ backlog	$\overline{MSE}$ capacity	$\overline{MSE}$ total	$\overline{l(\hat{\underline{\theta}})}$
1	Known	101.6	84.0	185.6	-5364.6
	Estimated	192.0	76.3	268.3	-4760.5
2	Known	100.8	81.9	182.7	-5353.1
	Estimated	186.7	74.0	260.7	-4753.8
3	Known	100.5	81.5	182.0	-5351.3
	Estimated	186.4	73.2	259.6	-4753.8

Table 5.3.6: Comparing the use of 100 sets of estimated parameters with use of known parameters in the UKF.

expected. This slight gain in accuracy for the capacity series is more than offset by the considerable drop in accuracy for the backlog series. Therefore, in terms of the objectives specified at the start of this section, the additional loss in accuracy in terms of $\overline{MSE}$ total, resulting from using our parameter estimation process is around 45%, compared to the benchmarks set using the true parameter values. In the next section we assess our assumption of diagonal covariance matrices.

5.3.7 Evaluating the assumption of diagonal covariance matrices

In simulation studies 1 and 2 - and also in simulation study 3 in section 5.4.4 - we assume that the off-diagonal terms of covariance matrices Q and R are zero. In other words we assume that the noise series of (5.3.9) - (5.3.14), $\underline{v}_1$ and $\underline{v}_2$, are uncorre-

lated; and similarly for $\underline{w}_1$ and $\underline{w}_2$. This was the case for the noisy data generated in simulation studies 1 and 2, and will be the case for simulation study 3. However, as we will see in chapter 6, the historical BT time series data are cross-correlated and as such, are likely to have correlated noise series. It would be preferable then to include these off-diagonal terms and estimate these additional parameters in the estimation process. In this section we explain why these are omitted - both from the simulation studies and when attempting to calibrate the hydraulics model using historical BT data in chapter 6.

For all three versions of the hydraulics model, (hydraulics model 3 will be introduced in section 5.4), data was simulated using non-diagonal covariance matrices Q and R, and attempts were made to estimate all parameters. The estimation of the off-diagonal elements of Q and R was generally poor, with considerably greater MAPE values than the structural parameter estimates reported in the simulation studies. The accuracy in estimating these off-diagonal parameters decreased as the complexity of the hydraulics model increased - so that they were least accurate when estimating parameters of hydraulics model 3. To make matters worse, in many cases estimation of the diagonal elements of Q and R, the sets of variance parameters σ_Q^2 and σ_R^2, declined in accuracy when including the off-diagonal elements in the estimation process.

Including the off-diagonal parameters also increased run times, sometimes considerably. For example, when estimating parameters for hydraulics model 3 and running on the University's high performance cluster, for many sets of starting values the al-

gorithm was unable to output any estimates even after a week. It is likely that these issues with the estimation process arose due to the added complexity of the resulting likelihood surface when including off-diagonal parameters.

As we explain at the start of this section, assuming that Q and R are diagonal means that it is likely in chapter 6 that our estimation process will assume that we have diagonal covariance matrices - when that is not actually the case. For the remainder of this section we investigate the effect of this. We simulate noisy data from hydraulics model 2 using non-diagonal covariance matrices - then attempt to estimate the parameters under the assumption that these matrices are diagonal.

For this study, the structural parameters are the same as simulation study 2; $\theta_d = 5$, $\theta_w = 0.2$ and $\theta_t = 2$. The variance parameters are the same as section 5.3.6 and parameter set 3 from simulation study 2; $\sigma^2_{Q_{1,1}} = \sigma^2_{Q_{2,2}} = \sigma^2_{R_{1,1}} = \sigma^2_{R_{2,2}} = 10^2$. Run lengths are again set at 500 days and the same 2 step job arrivals scenario is used. Five values for the off-diagonal elements of covariance matrices Q and R are set as follows; $\sigma_{Q_{1,2}} = \sigma_{Q_{2,1}} = \sigma_{R_{1,2}} = \sigma_{R_{2,1}} = c$, where $c = 10, 20, 30, 40, 50$. Their corresponding correlation values ρ, along with the MAPE of parameter estimates, are presented in table 5.3.7. In obtaining these parameter estimates, the same 5 sets of starting values as simulation study 2 are used, this time on 20 sets of simulated data - giving 100 sets of parameter estimates for each value of ρ. Hence the MAPE calculation is exactly the same as section 5.3.6.

ρ	$\overline{\tau}_d$	$\overline{\tau}_w$	$\overline{\tau}_t$
0.1	7.25	15.08	0.96
0.2	6.98	15.70	0.97
0.3	6.31	14.94	1.01
0.4	6.57	14.60	1.00
0.5	6.79	14.06	1.04

Table 5.3.7: MAPE for estimated parameters from simulated data with correlated noise.

From table 5.3.7, as we increase ρ (and therefore the covariance between the backlog and capacity noise), there does not appear to be any noticeable increase in the MAPE of the parameter estimates. θ_t is an exception to this with slight increases in MAPE, however this is more than offset by decreases in MAPE for θ_w. Therefore for this system, the incorrect assumption of diagonal covariance matrices does not increase the error for estimated parameters as we increase ρ. This is reassuring for chapter 6 where the BT data is likely to possess correlated noise series.

In the next section we introduce hydraulics model 3. This is considerably more complex than hydraulics models 1 and 2 and as such, can be viewed as the final step towards estimating parameters of the historical BT time series.

5.4 Applying FIMLOF to hydraulics model 3

In this section we test our modified FIMLOF algorithm on hydraulics model 3. Although some job types have been removed, this is the full version used by BT analysts and as such, is considerably more complex than hydraulics models 1 and 2. This model contains many more variables and has a higher degree of nonlinearity. Successfully estimating the parameters of this model can be seen as the final step before applying our approach to historical BT time series data in chapter 6.

The structure of this section is as follows. We begin by explaining the details of hydraulics model 3 and the key differences between this and models 1 and 2 in section 5.4.1. We derive a state-space model from hydraulics model 3 in section 5.4.2. Section 5.4.3 presents the modified FIMLOF algorithm. Section 5.4.4 presents details of simulation study 3 which aims to assess the performance of our modified FIMLOF algorithm at estimating the parameters of hydraulics model 3. This study aims to represent the conditions of the real system as accurately as possible in preparation for the historical BT data of chapter 6.

5.4.1 Hydraulics model 3 details

Hydraulics models 1 and 2 are effectively examples of a single feedback mechanism to adjust workforce numbers according to changes in demand. In some respects, hydraulics model 3 can be viewed as containing a number of similar feedback mechanisms, though some of these have different purposes. The model is far too complex

to summarise in a stock and flow diagram here. However, Figure 5.4.1 presents a diagram constructed by BT analysts that attempts to represent the complexities of the model. Note however that this is far from a full representation. This model has 5 stocks which are summarised below.

- Backlog: the length of the job queue as before (units: jobs).
- People deployed: no. of employees working standard hours.
- Shrinkage: inactive employee time such as training, leave etc..
- Overtime (OT): no. of employees working overtime hours.
- Unused OT: employees that can be called upon for overtime if required.

From Figure 5.4.1 we can see that we now have a number of feedback mechanisms for the different deployed resources. There are also additional mechanisms to ensure that overtime and shrinkage levels are sensible. All of these mechanisms attempt to replicate management decisions within the real system to maintain these 5 stocks at sensible levels.

Hydraulics model 3 also consists of 8 exogenous variables, a considerable increase on the 1 and 2 of models 1 and 2 respectively. Effectively, the two exogenous variables of interest are the job arrivals and a new variable, the total employees in the system. These two are of interest because in chapter 6 when we examine the BT regional time series, these are the only exogenous variables that differ for each region. The remaining exogenous variables relate to roster profiles, seasonal modifiers and

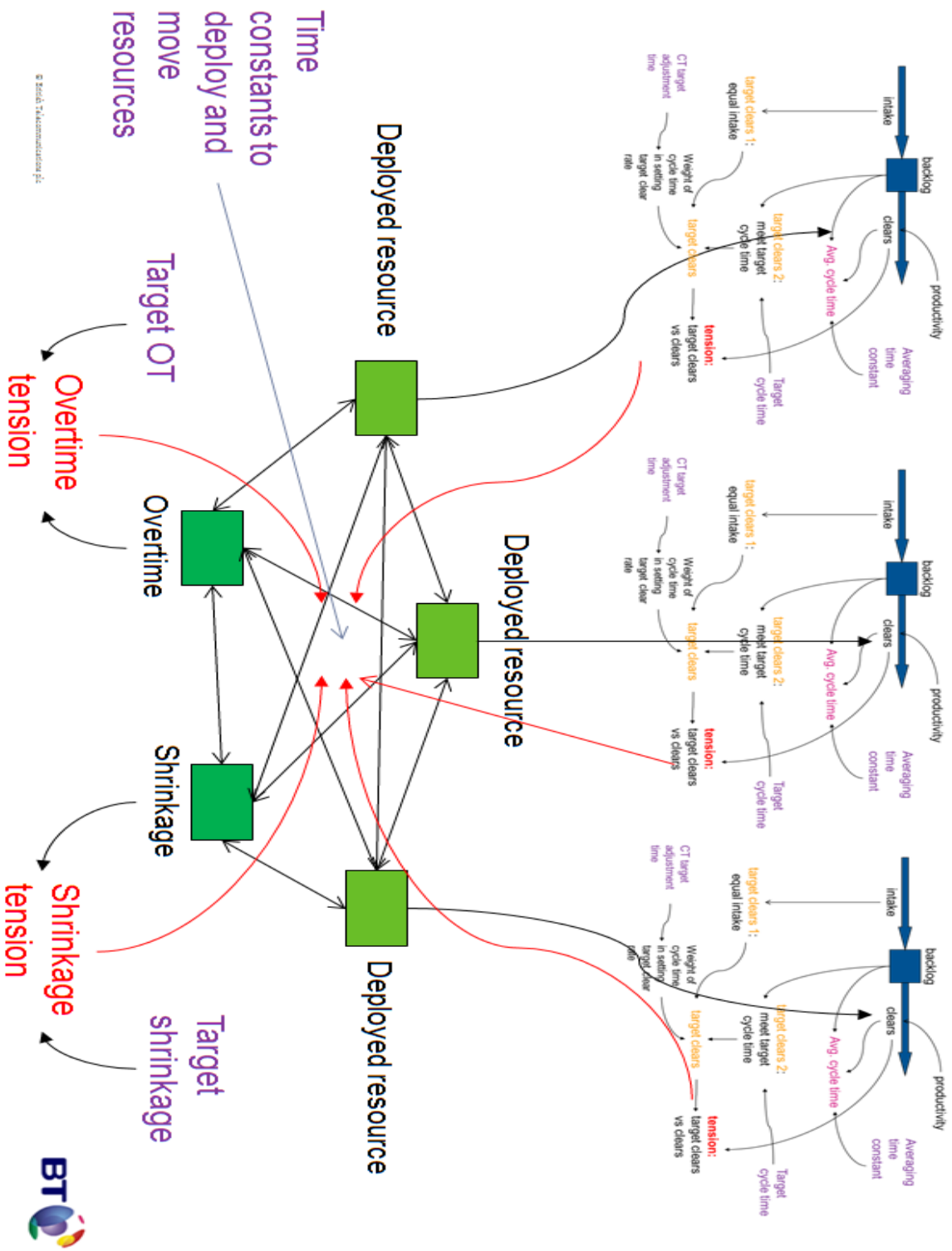

Figure 5.4.1: The complexities of hydraulics model 3, reproduced with permission from BT.

bank holiday indicators – and these are unchanged across the different regions. In total, hydraulics models 1 and 2 contained around 10 variables. For hydraulics model 3 the figure is around 100. This difference gives us some idea of the leap in complexity between model 3 and the other two.

Rather than the 3 structural parameters of hydraulics models 1 and 2, we now have 6. These are a mixture of time constants and targets and are summarised below with their units and current default values used by BT.

- θ_1, TCT (units = days, default = 2.5): target cycle time – as for models 1 and 2 the target time for jobs to be completed after entering the system.
- θ_2, TOT (units = hours per employee per day, default = 0.5/7): target overtime – since overtime hours are at a greater financial cost to BT this needs to be kept low.
- θ_3, TS (units = % of workforce, default = 20): target shrinkage – each employee requires a certain amount of leave and training.
- θ_4, TCS (units = days, default = 30): time constant for moving resources to shrinkage – there is a long delay associated with moving resources to shrinkage as a lower importance is attached to this than for example overtime.
- θ_5, TCOT (units = days, default = 4): time constant for moving resources to overtime – even in a crisis situation of extreme demand, there is a delay associated with this to represent meetings and rescheduling of jobs.

- θ_6, TCCTT (units = days, default = 5): time constant to adjust to cycle time target – there is another delay associated with adjusting standard (non overtime) employees.

We summarise the difference between the 3 versions of the hydraulics model in table 5.4.1. The bottom row of this table provides the total number of variables in each of the models. These values sum the number of stocks, auxiliary variables and exogenous variables in each of the models. This table demonstrates the considerable leap in terms of model complexity of hydraulics model 3 compared to the first two models.

	Model 1	Model 2	model 3
Stocks	2	2	5
Exogenous variables	1	2	8
Structural parameters $\underline{\theta}$	3	3	6
Total variables	11	12	100

Table 5.4.1: Comparison of hydraulics models 1, 2 and 3.

When we attempt to form difference equations for each of the stocks in the model, they are now too complex to be represented as simple functions of the other stocks and exogenous variables, as we obtained for hydraulics models 1 and 2. Instead, we can represent each stock by including necessary auxiliary variables from the model, within the difference equation. These auxiliary variables are denoted by a capital A.

Any exogenous variables are denoted by capital E. The notation for all variables is shown in table 5.4.2. Rows 1-5 represent the stocks, rows 6-14 represent the auxiliary variables and rows 15-16 represent the two most important exogenous variables.

Symbol	hydraulics model 3 parameter
b	backlog
p	people deployed
s	shrinkage
o	overtime
u	unused overtime
A_o	output (jobs completed)
A_{ts}	employees moved to shrinkage
A_{fs}	employees moved from shrinkage
A_{sp}	shrinkage pull
A_{tn}	tension
A_{ap}	adjust total people
A_{up}	unused overtime pull
A_{fu}	employees moved from unused overtime
A_{tu}	employees moved to unused overtime
E_n	new job arrivals
E_{te}	total employees

Table 5.4.2: Notation for hydraulics model 3 variables.

Difference equations for the stocks can now be represented in (5.4.1)-(5.4.5) and for the auxiliary variables in (5.4.6)-(5.4.14). For simplicity, we omit the suffix t from the auxiliary variables and exogenous variables. The two key exogenous variables E_n and E_{te}; new job arrivals and total employees respectively, are the only two that are labelled. Exogenous variables E_1 - E_5 represent roster profiles and seasonal factors. ψ_1 and ψ_2 from (5.4.6) represent constants that will not be estimated as parameters.

$$b_t = b_{t-1} + E_n - A_o = f_1(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) \tag{5.4.1}$$

$$p_t = p_{t-1} - A_{ts} + A_{fs} = f_2(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) \tag{5.4.2}$$

$$s_t = s_{t-1} + A_{ts} - A_{fs} + A_{ap} = f_3(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) \tag{5.4.3}$$

$$o_t = o_{t-1} - A_{tu} + A_{fu} = f_4(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) \tag{5.4.4}$$

$$u_t = u_{t-1} + A_{tu} - A_{fu} + A_{ap} = f_5(\underline{x}_{t-1}, \underline{u}_t, \underline{\theta}) \tag{5.4.5}$$

$$A_o = \psi_1 p_t + \psi_2 o_t \tag{5.4.6}$$

$$A_{ts} = E_1(p_t + s_t) + \frac{(A_{sp} - A_{tn})p_t}{\theta_4} \tag{5.4.7}$$

$$A_{fs} = -E_1(p_t + s_t) + \frac{(A_{tn} - A_{sp})p_t}{\theta_4} \tag{5.4.8}$$

$$A_{sp} = \left[\frac{(E_2 + \theta_3)(p_t + s_t)}{s_t + 0.01}\right]^{PF_s} \tag{5.4.9}$$

$$A_{tn} = \left[1 + \frac{S_7\left(\left[\left(\frac{b_t}{\theta_6} + \left[1 - \frac{\theta_1}{\theta_6}\right] E_3\right) E_4\right] - A_o\right)}{E_5 S_7(A_o)}\right]^{PP} \tag{5.4.10}$$

$$A_{ap} = E_{te} - p_t - s_t \tag{5.4.11}$$

$$A_{up} = \left[\frac{(1 - \frac{\theta_2}{1.23})(o_t + u_t)}{u_t}\right]^{PF_u} \tag{5.4.12}$$

$$A_{fu} = \left[\frac{(A_{tn} - A_{up})u_t}{\theta_5}\right] \tag{5.4.13}$$

$$A_{tu} = \left[\frac{(A_{up} - A_{tn})o_t}{\theta_5}\right] \tag{5.4.14}$$

These equations demonstrate the complexity of the model and its highly nonlinear nature. Despite this, some simplifications have been made to the equations. Many of them contain 'safety' functions to keep values within sensible ranges. Use of these is common in system dynamics modelling. For example, to ensure that a variable takes non-negative values, the following function is used so that the variable equals zero when negative: $\max(0,f(.))$. Many such functions are described in Sterman (2000). From the perspective of a FIMLOF approach in attempting to estimate the model's parameters, use of these functions only adds to the complexity and nonlinearities of the model.

5.4.2 Derivation of the state-space model

From the difference equations formed in the previous section, we can now form the state-space model. The process is similar to that in sections 4.3.3 and 5.3.3. The state equations are formed by adding noise terms to the difference equations; these are shown in (5.4.15)-(5.4.19). State update functions $f_1(.) - f_5(.)$ are as defined in the previous section in (5.4.1)-(5.4.5).

$$b_{t+1} = f_1(\underline{x}_t, \underline{u}_t, \underline{\theta}) + v_{1,t} \tag{5.4.15}$$

$$p_{t+1} = f_2(\underline{x}_t, \underline{u}_t, \underline{\theta}) + v_{2,t} \tag{5.4.16}$$

$$s_{t+1} = f_3(\underline{x}_t, \underline{u}_t, \underline{\theta}) + v_{3,t} \tag{5.4.17}$$

$$o_{t+1} = f_4(\underline{x}_t, \underline{u}_t, \underline{\theta}) + v_{4,t} \tag{5.4.18}$$

$$uo_{t+1} = f_5(\underline{x}_t, \underline{u}_t, \underline{\theta}) + v_{5,t} \tag{5.4.19}$$

We then proceed to add further noise terms to form the observation equations, given in (5.4.20)-(5.4.24).

$$b_t^* = b_t + w_{1,t} \tag{5.4.20}$$

$$p_t^* = p_t + w_{2,t} \tag{5.4.21}$$

$$s_t^* = s_t + w_{3,t} \tag{5.4.22}$$

$$o_t^* = o_t + w_{4,t} \tag{5.4.23}$$

$$uo_t^* = uo_t + w_{5,t} \tag{5.4.24}$$

The observation terms can be represented as $\underline{y}_t = (\underline{b}_t^*, \underline{p}_t^*, \underline{s}_t^*, \underline{o}_t^*, \underline{uo}_t^*)^T$ and the state terms as $\underline{x}_t = (\underline{b}_t, \underline{p}_t, \underline{s}_t, \underline{o}_t, \underline{uo}_t)^T$. The state noise terms $\underline{v}_i$ and observational noise terms $\underline{w}_i$ for $i = 1, ..., 5$ are assumed to be as follows: $\underline{v}_i \sim N(0, \sigma_{Q_i}^2)$ and $\underline{w}_i \sim N(0, \sigma_{R_i}^2)$. Covariance matrices Q and R then take the following form:

$$Q = \begin{bmatrix} \sigma_{Q_1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{Q_2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{Q_3}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{Q_4}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{Q_5}^2 \end{bmatrix}, R = \begin{bmatrix} \sigma_{R_1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{R_2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{R_3}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{R_4}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{R_5}^2 \end{bmatrix}.$$

In section 5.4.4 we present simulation study 3 where we attempt to estimate the parameters of this model. Note that we now have 16 parameters in total, the 6 structural parameters the structural parameters $\underline{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ and the 10 variance parameters that form the diagonals of Q and R. This is a considerable leap from the 7 parameters that were estimated in the first two simulation studies. As for simulation studies 1 and 2, we again assume a diagonal structure for covariance

matrices Q and R. This was evaluated in section 5.3.7.

5.4.3 Modified FIMLOF algorithm 3

Having formed the state-space model using the structure of hydraulics model 3, we are now ready to present an algorithm for implementing modifications to the FIMLOF method.

Firstly however, we make a change to the core of the algorithm compared to algorithms 2 and 3. For hydraulics model 3 we now discard the approximate method of (4.2.15) for calculating the log-likelihood when Kalman filtering and use the exact method of (4.2.14). It is the exact method that now brings improved performance in terms of the accuracy of parameter estimation. In section 4.4 we speculated that the approximate method performed better than the exact method as it required no inversion of matrices, therefore reducing numerical error during computation. It is likely that for a more complex and realistic system such as hydraulics model 3, the added accuracy of using the exact representation of the log-likelihood overrides any loss of accuracy of numerical errors.

For noisy observable time series data $\underline{y} = (\underline{b}^*, \underline{p}^*, \underline{s}^*, \underline{o}^*, \underline{uo}^*)^T$, we estimate the structural parameters $\underline{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ and the variance parameters $\underline{\sigma}^2 = (\sigma^2_{Q_1}$,..., $\sigma^2_{Q_5}$, $\sigma^2_{R_1}$,..., $\sigma^2_{R_5})$ of the state-space model in (5.4.15)-(5.4.24) using the algorithm below. We refer to this as algorithm 3.

1. The Nelder-Mead simplex method selects a candidate set of parameters $(\underline{\theta}, \underline{\sigma}^2)$, (or in the first iteration the user selects suitable starting values $\underline{\theta} = \underline{\theta}_0$ and $\underline{\sigma}^2 = \underline{\sigma}_0^2$).

2. The unscented Kalman filter proceeds along time series $\underline{y}$ and for each update at time t:

   - The $\underline{\theta}$ parameters in the nonlinear state update equations (5.4.15)-(5.4.19) and the variance parameters $\underline{\sigma}^2$ of covariance matrices Q and R influence the filter's estimate of latent states $\underline{x}_t = (\underline{b}_t, \underline{p}_t, \underline{s}_t, \underline{o}_t, \underline{uo}_t)^T$.

   - Terms from these calculations are used to calculate the log-likelihood for the candidate set of parameters, $\log(L(\underline{\theta}, \underline{\sigma}^2), t)$, using the exact representation of (4.2.14).

3. After the Kalman filter has calculated the log-likelihood at each discrete interval t in the series, an overall log-likelihood for the time series is calculated by summing the individual log-likelihood terms: $\log(L(\underline{\theta}, \underline{\sigma}^2)) = \Sigma_{t=1}^{T} \log(L(\underline{\theta}, \underline{\sigma}^2, t))$.

4. The Nelder-Mead method repeats the process until a local optimum is found and maximum likelihood estimates $\hat{\underline{\theta}}$ and $\hat{\underline{\sigma}}^2$ are determined.

In the simulation study in section 5.4.4, artificial noisy data $\underline{y}$ is simulated from hydraulics model 3 with known structural and variance parameters. Algorithm 3 is used for the estimation of these parameters and its performance is assessed.

5.4.4 Simulation study 3

In this section we present results of simulation study 3; aimed at assessing the performance of our modified FIMLOF algorithm when estimating the parameters of noisy simulated data. We begin by describing how the study is set up. There are a number of adjustments that are different to simulation studies 1 and 2. Some of these are by necessity due to the greater complexity of the problem. Many however are an attempt to create a study that replicates many of the conditions that will also be present when we apply the method to the historical BT data in chapter 6. This ensures that the progression from the simulation studies up to the regional data is as small as possible. Having demonstrated in the first two studies the success of the approach and observed the effects of added state and observational noise, we now shift our focus more towards creating experiments that are as realistic as possible when compared to the BT system.

As we saw in section 5.4.2, we now have 5 states and 5 observations for each t. Covariance matrices Q and R are now 5×5. This means that the computations within the UKF involve 5×5 matrices. For simulation study 1 using the standard Kalman filter, parameter estimates were computed in R in around 20 seconds on average. For simulation study 2, this increased to nearly 3 minutes. For this study, estimates were taking considerably longer; in some cases over an hour. The Rcpp package Eddelbuettel (2013) allows seamless integration between R and C++. This allows the user to use C++ functions within the R language, enabling faster computation. Effectively it provides the user with the benefits of using C++, by enabling C++ functions

to be embedded within the R code. Coding the key functions in this way dramatically speeded up the computations; parameter estimates taking on average 3 minutes.

The job arrivals for this study, unlike the previous 'two step' artificial scenarios, are a far more realistic representation of the job arrivals that we can expect in chapter 6. Job arrivals in this study were obtained by comparing simulations with the historical data and through discussions with BT analysts in an attempt to simulate series that are as realistic as possible. This involved generating an autoregressive process with $\rho = 0.3$, adding a constant to this series and giving the data a weekly profile. A characteristic of using such job arrivals – which will also be the case in chapter 6 – is that compared to the simple artificial scenario we now have a considerably greater level of activity throughout the series. This will ellicit a greater response from hydraulics model 3 – which may provide more information on some or all of the parameters $\underline{\theta}$. With the added complexities of this system and also the likelihood surface compared to the first two simulation studies, this greater response from the model may be a helpful factor.

The simulated data from hydraulics model 3 will use the same exogenous variables (e.g. roster profiles, seasonal modifiers etc.) as would be input to this model in practice at BT. To ensure that the exogenous variable for total employees is realistic, this variable for one of the BT regional series (Lancashire) was used throughout the study.

The length of the simulated time series from hydraulics model 3 is another selec-

tion that we attempt to make as realistic as possible. This was set at 500 days for the first two studies, however we now increase this to 959 days. This is the length of the BT regional time series in chapter 6. Peterson & Schweppe (1975) showed that a longer time series results in more accurate FIMLOF estimates. As such, this adjustment for realism may be an additional helpful factor in parameter estimation.

The experiments take a similar form to simulation studies 1 and 2, with 100 sets of data simulated and 5 starting values used for the optimisation, giving 500 sets of parameter estimates for each setting of added noise. Simulation study 2 showed the dominant effect of adding state noise σ_Q to the data. Hydraulics model 3 is a considerably more complex system than model 2 and is also highly nonlinear. As such, parameter estimates for this study are likely to be less accurate. Rather than the effects of added state noise σ_Q dominating the study again, we instead wish to determine the effects of observational noise σ_R on the accuracy of parameter estimates. Therefore, unlike the previous two studies, we now restrict ourselves to the case of added observational noise only. As such, we fix $\sigma_Q = 0$ and again add 6 different levels of observational noise σ_R.

Due to the nature of hydraulics model 3 it is no longer sensible to add the same amounts of noise to the data as for simulation studies 1 and 2. For example, for data simulated from hydraulics model 3 and using the job arrivals described above, the backlog series takes values well into the thousands but the shrinkage series is a percentage between 0 and 100. More importantly, the standard deviation of the shrinkage

series is a fraction of that of the backlog series. Any noise added to the data must be scaled by an appropriate *standard deviation profile* to reflect the considerably different level of variability exhibited by each series. In order to determine this profile, we examined the historical BT time series for the 59 geographical regions that will be presented in chapter 6. An average standard deviation value for each of the five series (backlog, people, shrinkage, overtime, unused overtime) across the regions was calculated. By dividing each of these by the smallest value the following standard deviation profile was determined: (200, 1, 1, 3, 4) for the five series respectively. This means that on average, the backlog standard deviation was 200 times greater than the people deployed and shrinkage series.

When adding observational noise σ_R to the data, to create six levels of added noise, this standard deviation profile was multiplied by six values of a 'multiplication factor' (mf). These values were 1, 2, 3, 5, 8 and 10. These are the six levels of adding observational noise σ_R for this study. Hence, the true values of the covariance matrix R defined in section 5.4.2 are as follows, with the diagonals $\sigma^2_{R_1}, ..., \sigma^2_{R_5}$ representing the variance of the added noise:

$$R = \begin{bmatrix} (200\,mf)^2 & 0 & 0 & 0 & 0 \\ 0 & (mf)^2 & 0 & 0 & 0 \\ 0 & 0 & (mf)^2 & 0 & 0 \\ 0 & 0 & 0 & (3\,mf)^2 & 0 \\ 0 & 0 & 0 & 0 & (4\,mf)^2 \end{bmatrix},$$

for mf = 1, 2, 3, 5, 8, 10.

As we have different parameters θ_1 - θ_6 with different default values to the first two studies, we must now define another 5 sets of starting values for the optimisation routine. In a similar pattern to the first two studies, we select sets s1 and s2 less than the true parameter values, s3 equal and sets s4 and s5 to be greater than the parameters. The spacing between values of these sets is again at regular intervals, as we see in table 5.4.3.

Parameter	s1	s2	s3	s4	s5
θ_1	1.5	2	2.5	3	3.5
θ_2	0.1/7	0.3/7	0.5/7	0.7/7	0.9/7
θ_3	0.1	0.15	0.2	0.25	0.3
θ_4	25	27.5	30	32.5	35
θ_5	3	3.5	4	4.5	5
θ_6	4	4.5	5	5.5	6

Table 5.4.3: Starting values for the optimisation routine.

We again assumed that we knew the correct noise structure – i.e. the diagonal nature of the covariance matrices Q and R. The starting values of the 5 state variance parameters σ_Q of covariance matrix Q were again set around zero; i.e. $\sigma_{Q_0} = e^{-10}$. Since there is no added state noise, i.e. $\sigma_Q = 0$, for this study these starting values

are close to the true parameter values.

As the observational noise added to the data now has a standard deviation profile (in that different amounts of noise are added to each series), this may cause the estimation of the observational variance parameters σ_R to become more challenging. Rather than setting the starting values σ_{R_0} at naive values close to zero as in the two previous studies, a more informed method was used. The starting values σ_{R_0} of these 5 σ_R parameters were set by estimating the variance directly from the 5 time series of the data. This was intended as further preparation for the BT data in chapter 6, where the standard deviation profile is apparent. The variance was estimated by fitting smoothing splines to the noisy data and examining the differences between these and the raw data. These estimates of the variances were used as starting values in the optimisation routine. Note that this process still assumes no prior knowledge of the scale of the added noise.

The two sets of 5 variance parameters σ_Q and σ_R are again constrained to be positive using the exponential function. As θ_3 represents a probability, we constrain this between 0 and 1 using the inverse of the logit function: $\text{logit}^{-1}(\theta_3) = \frac{e^{\theta_3}}{e^{\theta_3}+1}$.

The histograms in Figure 5.4.2 show estimates of parameters θ_1 - θ_3 and the histograms of Figure 5.4.3 show estimates of parameters θ_4 - θ_6. 'mf' refers to the multiplicative factor used on the standard deviation profile when adding noise. To assess the performance of parameter estimation, we again examine the relative differ-

ence of each of the parameter estimates compared to the true value by calculating the mean absolute percentage error (MAPE) using the same method as in section 5.3.5. Again, the MAPE for parameter estimate $\hat{\theta}_i$ is represented by $\overline{\tau}_i$, for $i = 1, \ldots, 6$. The values of $\overline{\tau}_i$ for this study are shown in table 5.4.4.

Parameter set	$\overline{\tau}_1$	$\overline{\tau}_2$	$\overline{\tau}_3$	$\overline{\tau}_4$	$\overline{\tau}_5$	$\overline{\tau}_6$
1	18.3	4.4	0.7	187.2	32.9	22.7
2	16.8	4.0	0.7	189.4	34.6	22.0
3	22.4	5.4	0.9	235.8	38.6	28.4
4	19.0	4.5	0.8	231.6	41.3	25.9
5	27.8	5.7	1.1	292.6	56.1	35.4
6	41.7	8.2	1.6	423.4	87.6	55.9

Table 5.4.4: MAPE of parameter estimates.

From table 5.4.4 and the histograms of Figures 5.4.2 - 5.4.3 the following observations are immediately apparent:

1. Parameter estimates are noticeably less accurate than for studies 1 and 2.

2. Increasing the added observational noise results in less accurate estimates.

3. The difference in accuracy between different parameters has widened compared to the first two studies.

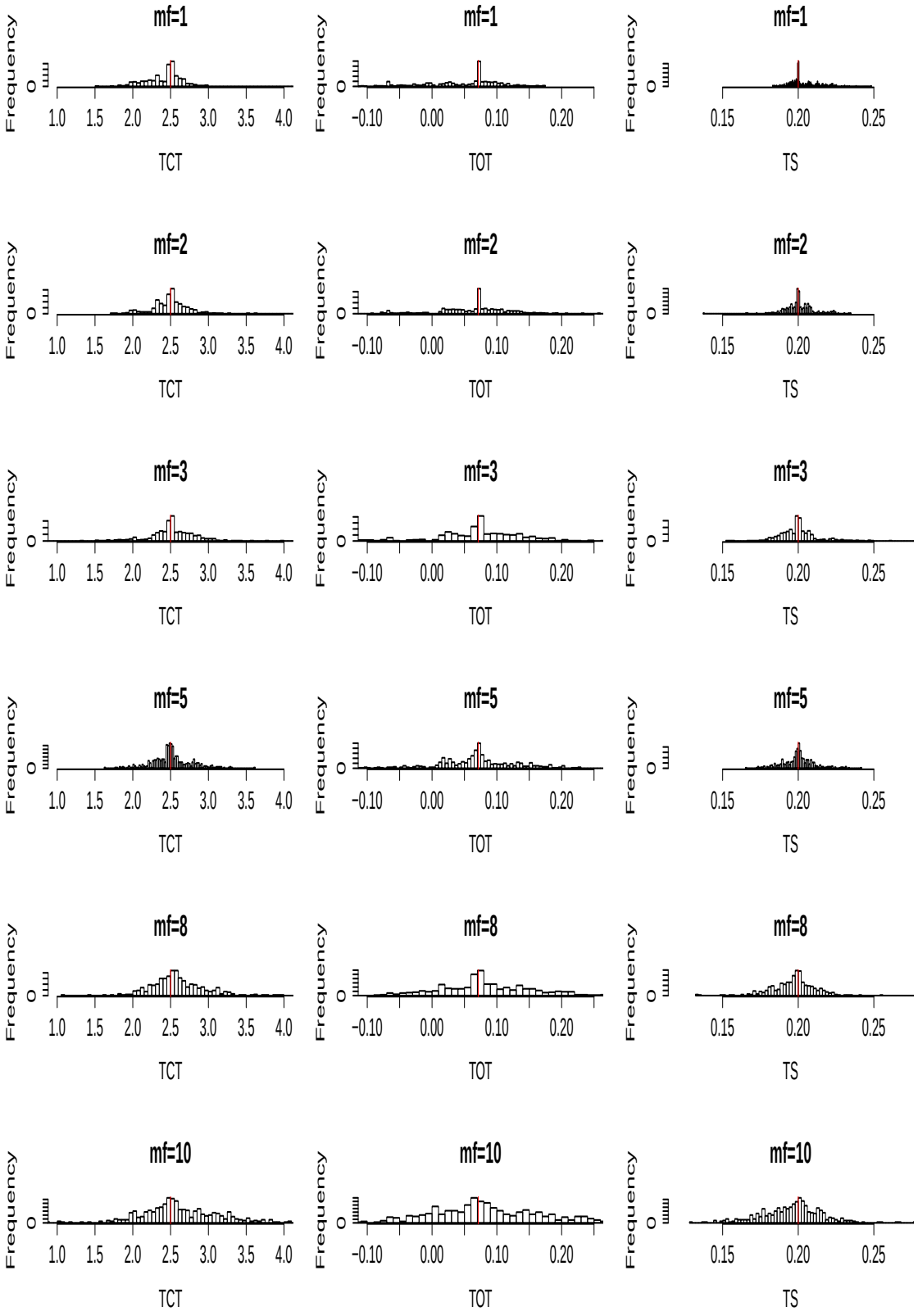

Figure 5.4.2: Simulation study 3 results for parameters θ_1 - θ_3.

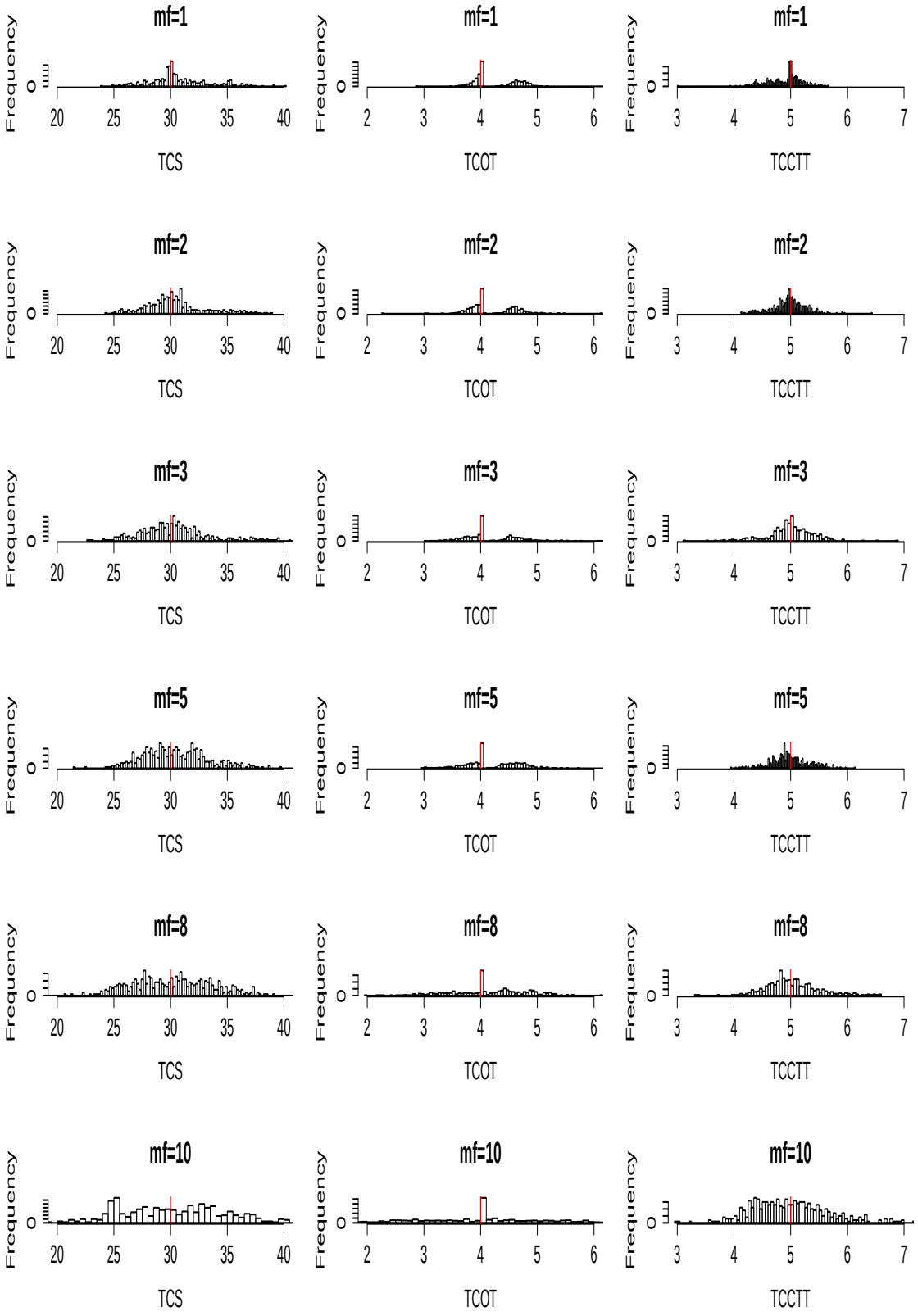

Figure 5.4.3: Simulation study 3 results for parameters θ_4 - θ_6.

There are a number of possible reasons for observation (1). Hydraulics model 3 is a far more complex and realistic system than the first two models. As such, the identifiability of some or all of the 6 parameters is likely to have decreased, causing parameter estimation to be more difficult. Hydraulics model 3 is also highly nonlinear, considerably more so than hydraulics model 2. Therefore the approximation made by the UKF is likely to be less accurate than it was for hydraulics model 2. Due to the required scaling we have added different levels of noise in this study. It is worth highlighting that for multiplicative factors of 5, 8 and 10, we have standard deviations for the added noise to the backlog of 1000, 1600 and 2000 respectively. This is a considerable leap from the first two studies. Despite this, our results demonstrate some degree of success in estimating the parameters. Another important consequence of the added complexity of hydraulics model 3 is that in this study we are estimating 16 parameters, rather than the 7 that we estimated in the first two studies. This is likely to result in a considerably more complex likelihood surface, making parameter estimation more challenging.

Observation (2), like the previous two studies is not surprising. Greater amounts of added noise corrupt the data more so that the UKF's estimate of the latent state become less accurate.

Observation (3) is an interesting exposition of the challenges associated with estimating the parameters of hydraulics model 3. The widened gap between the accuracy of estimates for different parameters, most obvious for parameters θ_3 and θ_4, is likely

caused by the differences in the identifiability of parameters. This was confirmed by BT analysts who investigated the 'importance' of each parameter by repeatedly running hydraulics model 3 for different parameter values. This included fixing the parameters and adjusting a single parameter in turn. It is not surprising to learn that the parameters that we found the hardest to estimate, θ_4 and θ_5, were found to be the least important in terms of the behaviour of the model. That is, only extreme changes in these parameters had any noticeable effect on the simulations of hydraulics model 3. This is a likely explanation for our difficulties in estimating these and the high MAPE values observed in the results.

5.5 Discussion

In this chapter, the novel process for the formulation of a state-space model from a SD model, presented in chapter 4, has been extended to more complex, nonlinear systems. Two simulation studies, studies 2 and 3, present results of applying modified FIMLOF algorithms to estimate the parameters of these nonlinear state-space models – that are based on hydraulics models 2 and 3 respectively. Results of study 2 demonstrate success in estimating the parameters, especially for lower amounts of added noise. The amount of added state noise σ_Q was shown to be an important factor in the performance of the algorithm. Results of study 3 showed a loss of accuracy compared to studies 1 and 2 due to the significantly more complex model structure and greater number of parameters, but nonetheless demonstrate some success, especially for lower amounts of added noise. The identifiability of the 6 structural parameters was also

shown to be an important issue.

In section 5.2 we highlighted the modern applications of FIMLOF in the literature that all use the extended Kalman filter for use on nonlinear systems, despite the evidence in favour of the unscented Kalman filter. There are no examples that we are aware of that have applied FIMLOF using the UKF. By applying the UKF successfully to estimate the parameters of two nonlinear systems in this chapter, we make a claim to contribute to the literature. The second of these systems, hydraulics model 3, is also more complex and nonlinear than any other models that we have seen FIMLOF applied to in the literature. This chapter demonstrates that the approach can give useful insights into a complex industrial model, the output of which is relied upon by a large multinational organisation.

Simulation study 2 provided reassurance that the FIMLOF algorithm, modified with the UKF, can be successful at estimating the parameters of a nonlinear system. Simulation study 3 was designed to recreate the conditions of the real system as far as possible. The relative success of our modified algorithm in this study is promising at this final stage. The study also exposed the importance of identification of parameters in hydraulics model 3 and the considerable differences in accuracy across the 6 parameters estimated in the model. Knowledge of which are the most challenging parameters to estimate may be useful when progressing to the real system. Therefore the results in this chapter can be viewed as a final step towards progressing to the historical BT time series data in chapter 6.

Chapter 6

Application: Parameter estimation for BT regional data

6.1 Introduction

In this chapter we apply a modified FIMLOF algorithm to estimate the parameters of hydraulics model 3 from historical BT time series data. This data represents 59 geographical regions across the UK. As system behaviour is known by BT analysts to vary across the regions, which the results of chapter 3 supported, there is interest at BT in calibrating the model for each geographical region – this is the objective of this chapter. Calibrating for each region is considered preferable to relying on a set of fixed 'default' parameters across all regions. As we explain in chapter 4, a method that automates the process of calibration is more desirable than the modeller manually calibrating the model by hand in a cumbersome trial and error procedure – which is the approach currently adopted by BT.

The results of chapters 4 and 5 have enabled the progression in this chapter to the real BT system. The test of concept on hydraulics model 1 in chapter 4 provided the foundations for this chapter as results demonstrated success of our modified FIMLOF algorithm on a simple linear system. In chapter 5, incorporating the UKF enabled progression to two nonlinear models; hydraulics models 2 and 3. The simulation study involving the latter of these was based on the full version of the hydraulics model and was designed to replicate as far as possible the real BT system. Results of this study demonstrated that we are now ready to estimate the parameters of the BT system.

The structure of the chapter is as follows. Section 6.2 presents the exploratory analysis and describes the details of the time series data. Section 6.3 explains the adjustments that were required in applying the method to this regional dataset. Section 6.4 presents the results, with the discussion in section 6.5.

6.2 Exploratory analysis

In this section we present an exploratory analysis of the BT regional dataset. This examines the variation within regions in section 6.2.2 and the variation between the different regions in section 6.2.3. We begin with an overview of the data in section 6.2.1 and explain some of the data processing requirements that were necessary for certain series.

Before proceeding we note that the data presented in this section is not the same as that presented in chapter 3. This data has a different geographical breakdown of regions and many of the time series have different units.

6.2.1 Overview of the BT regional dataset

There are 59 geographical regions in total, each representing a region within England, Scotland and Wales. The lengths of all time series for each region are 959 days, covering the period between 1st January 2012 to 16th August 2014. For each of the 59 regions, we have daily time series data for the 5 endogenous variables, the stocks of hydraulics model 3, and also for the same 8 exogenous variables as described in section 5.4.1. Unlike simulation study 3 however, the two key exogenous variables, the new job arrivals $\underline{E}_n$ and the total employees $\underline{E}_{te}$, are different for each geographical region.

As we demonstrated in section 4.4.4, although the time series $\underline{y}$ output from the three hydraulics models represent data that are effectively counts, the models' output non-integer values. We added state and observational level Gaussian noise, hence our simulated datasets were suitable for the FIMLOF approach. For BT data of this chapter, to ensure that we again have non-integer values, the units of many series change from those in our simulated data of chapters 4 and 5. Instead of counts of jobs in the backlog, jobs are now measured in units of man hours – that are required for the workforce to complete them. Similarly for the people deployed, overtime and unused overtime series, workforce is also measured in man hours. Shrinkage remains

unchanged as a percentage of workforce. This means that the BT data consists of non-integer values. Accordingly, exogenous variables such as $\underline{E}_n$ and $\underline{E}_{te}$, the new job arrivals and the total employees, are also measured in man hours. The necessary calculations for the conversions in units for all of these variables were performed by BT analysts.

To estimate the parameters of hydraulics model 3 we require daily time series data. This presented a problem for some series. We were only provided with weekly time series data for endogenous variables people deployed $\underline{p}^*$ and shrinkage $\underline{s}^*$ and the exogenous variable total employees $\underline{E}_{te}$. In order to convert these into daily series, cubic spline interpolation was used. The method chosen is described in Forsythe et al. (1977). The results of this interpolation were used as our estimate of the unknown daily time series.

The 59 series for overtime were provided in units of hours per employee per day. To ensure that the data remained in the same units as hydraulics model 3, number of employees per day, required transforming the overtime data. This involved dividing by a weekly overtime profile contained within hydraulics model 3. For certain days (e.g. Wednesday), this profile has a low value and therefore when dividing by the profile, large values can result. Consequently for many of the overtime series, some spikes can be observed at certain periods along the series as we see in Figure 6.2.1 around days 650-800. The time series for the unused overtime was derived directly from the overtime series and total employees $\underline{E}_{te}$. Therefore this series suffers from

the same issues.

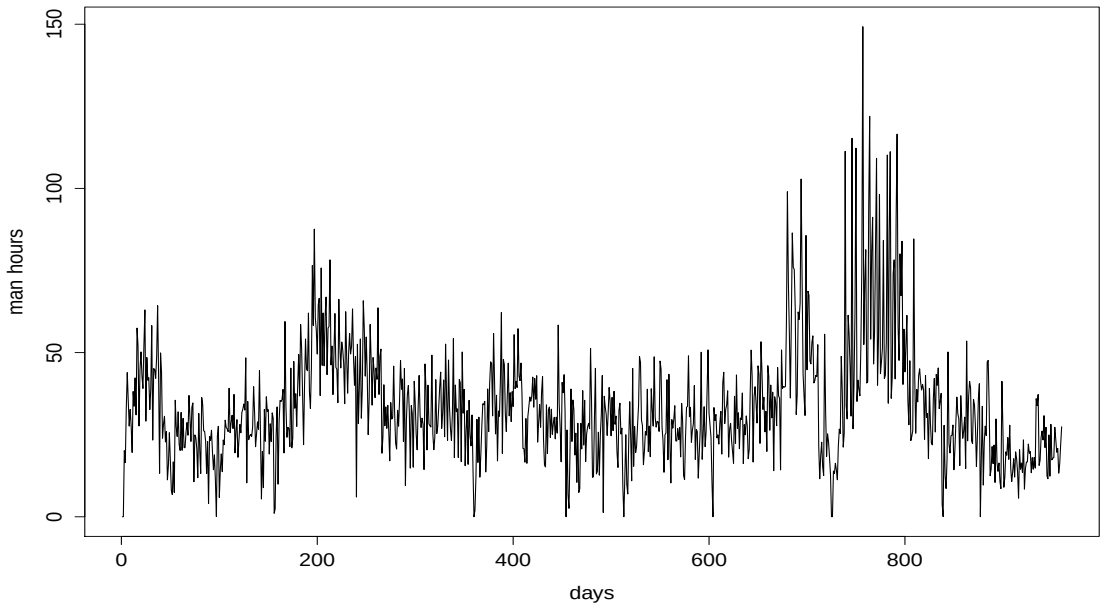

Figure 6.2.1: Overtime data corrupted by spikes.

6.2.2 Within region variation

In this section we investigate 'within region' properties of the BT data by examining correlations between different time series within the same region. Figure 6.2.2 presents boxplots of the main correlations between the series, within each of the 59 regions. On the plot, the following notation 'Bk', 'Ppl', 'Shk' and 'Ot' is used to represent the backlog, people deployed (standard workforce hours), shrinkage and overtime (overtime hours) series respectively. So that for example, the notation 'Bk:Ppl' is used to indicate a boxplot of the correlations between the backlog and people deployed series.

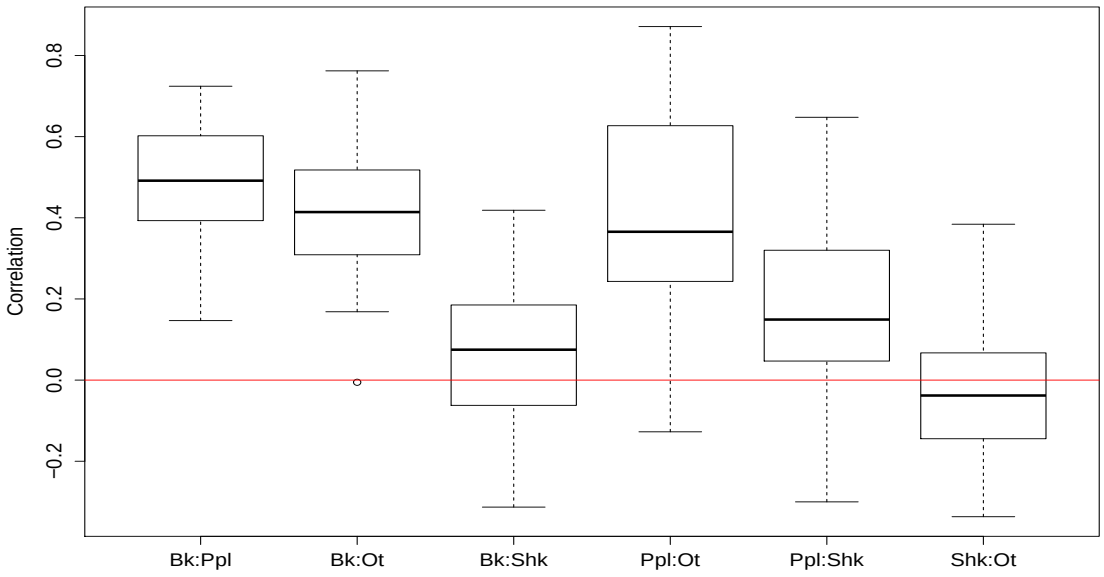

Figure 6.2.2: Boxplots of correlations between time series from the same region.

We observed from the BT dataset of chapter 3 that regions respond to increases in job arrivals (and hence increases in the backlog) by increasing people numbers. The two greatest medians in the plot (0.49 and 0.41) are for the correlations 'Bk:Ppl' and 'Bk:Ot' respectively. These positive values suggest that for this dataset also, the regions may indeed be responding to changes in the backlog by increasing people numbers and overtime. The relationships between backlog and people deployed and between backlog and overtime are key elements of a region's behaviour. These relationships determine how well the backlog is controlled, and hence how effectively the region meets its performance targets for completing jobs. We can also observe that the inter-quartile range is around 0.2 for each of these two sets of correlations – hence there is some variability between the different regions for these key relationships

within each region.

Figure 6.2.2 also reveals largely positive correlations between people deployed and overtime, though these correlations vary considerably. These positive correlations are to be expected, not only because both series have positive correlations with the backlog, but also due to the following. Each region can adjust workforce numbers by adjusting both people deployed and overtime numbers. For example in situations where the backlog must be reduced, BT analysts explained that often both the people deployed and overtime will be increased. Shrinkage correlates poorly with the backlog, people deployed and overtime. Shrinkage includes many activities such as training, leave, sickness etc. and many of these activities are not influenced by the current performance of the region.

Examining the lagged correlations reveals some interesting properties. Correlations between backlog and people deployed and also between backlog and overtime, with people deployed and overtime lagged to reveal any delayed correlation, are shown in Figure 6.2.3. These values are averaged over the 59 regions. For the lagged correlations between backlog and people deployed, the correlation increases until a lag of 16 days where it peaks, and then declines. Therefore on average, the strongest correlation between backlog and people deployed is achieved when people deployed is lagged by 16 days. This suggests that on average across the regions, the most clear response in terms of people deployed, to changes in the backlog, is most evident around day 16.

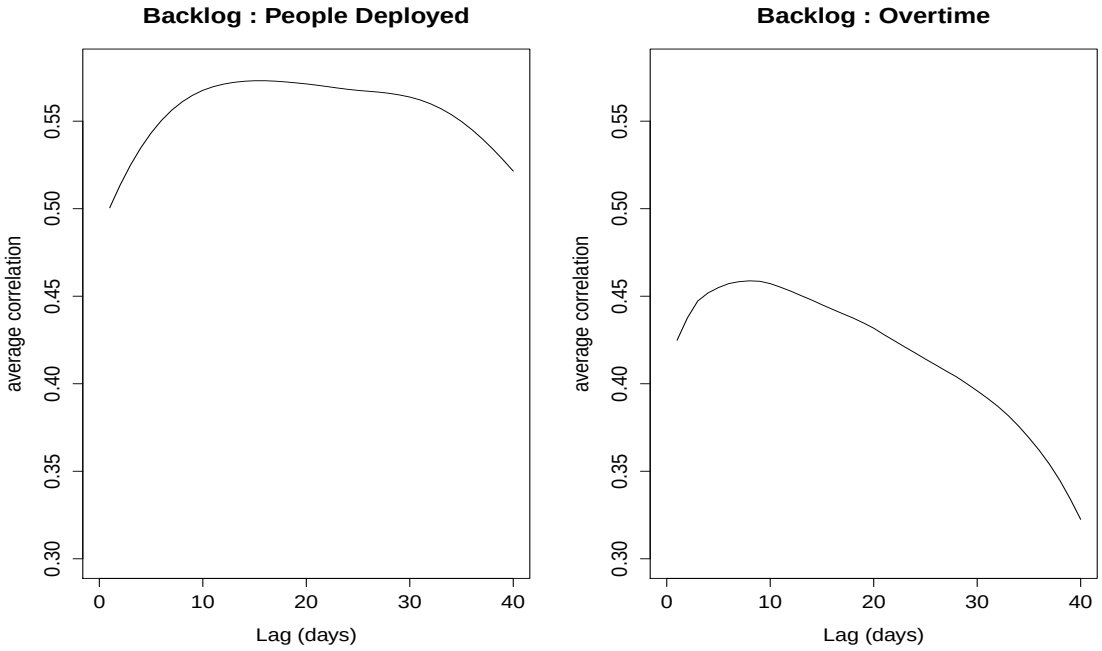

Figure 6.2.3: Lagged correlations averaged over the 59 regions.

On the other hand, the average lagged correlations between backlog and overtime reach their peak at a lag of only 8 days, increasing prior to this and declining after. So on average across the regions, the most clear overtime response to changes in the backlog is most evident around day 8. This suggests that on average, a region is able to increase overtime man hours faster than standard man hours. This is a property of the BT system that is well known amongst BT analysts. Of course, these values are purely based on lagged correlations and not isolated for example to only examine cases where there are increases in the backlog - which was studied in chapter 3. However these exploratory plots reveal some interesting differences in the rates of a region's response.

6.2.3 Between region variation

In this section we investigate 'between region' properties of the BT data by examining correlations between similar time series from different regions.

When investigating the 'between region' effects, an immediate question to ask is whether neighbouring regions are more similar to each other than regions geographically further apart. To answer this, we first group the 59 regions into 9 different areas as defined by BT. Although the regional boundaries within each area and the regional datasets are different, these are the same 9 areas as those in chapter 3; Scotland, North-East England, North-West England, North Wales/Midlands, South Wales/Midlands, South-West England, South-East England, London and East Anglia. Each of the areas contain between 5 and 9 regions.

To investigate whether neighbouring regions are more similar than regions further away, we calculate two sets of correlations; 'within area' correlations (between regions from the same area) and 'between area' correlations (between regions from different areas). These correlations are calculated for all regions for each of the backlog, people deployed, shrinkage and overtime series. This means that in total, for each series, the correlations between each region with all other regions are calculated. These are then grouped into either 'within area' or 'between area' correlations. Boxplots of these correlations are presented in Figure 6.2.4. In the plot, 'within' represents within area correlations and 'between' represents between area correlations.

For each plot there are considerably more 'between area' correlations than 'within area' correlations; 2872 compared to 320. This difference in sample sizes limits the confidence that can be placed in the comparisons, however it's clear that regions within the same area appear to be better correlated than regions of different areas. This suggests that regions that are geographically closer do exhibit more similar behaviour than regions further apart.

These correlations however should be placed in context. In addition to the correlations of Figure 6.2.4, it is important to consider similar correlations for the job arrivals. For each region, the behaviour of series such as the backlog, people deployed, shrinkage and overtime are partly driven by the nature of the job arrivals, as the region attempts to complete the jobs fast enough to meet performance targets. Similar boxplots as above are presented in Figure 6.2.5 which show the 'within area' and 'between area' correlations for the job arrivals. We can see that the job arrivals have higher correlations than any of the other series. This is especially so for 'within area' correlations, with a median correlation close to 0.9. Since the nature of the job arrivals partly drive each region's response, and considering that the performance targets are the same across all regions, we would actually expect some differences between the 'within area' and 'between area' correlations - which were shown in Figure 6.2.4. Hence the differences observed in Figure 6.2.4 are likely to be at least partly due to the differences between the 'within area' and 'between area' correlations of the job arrivals.

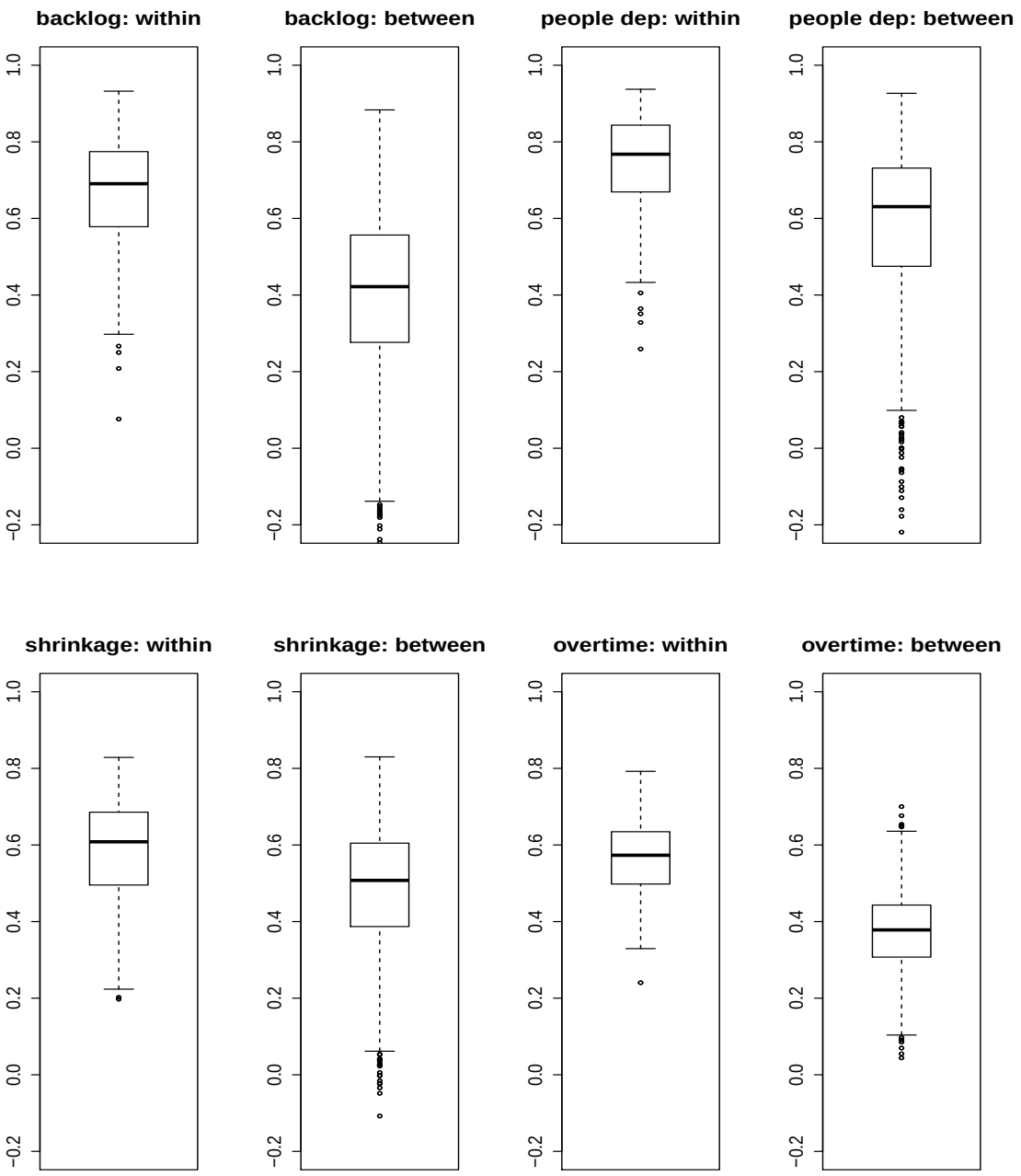

Figure 6.2.4: Comparing correlations between regions within the same area and between different areas.

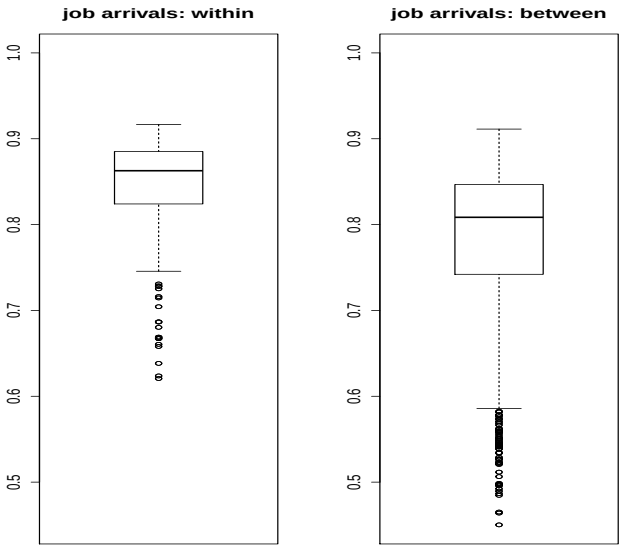

Figure 6.2.5: 'Within area' and 'between area' job arrivals' correlations.

The results in this section suggest that there is some evidence that geographically closer regions are more similar than regions further apart - although it is difficult to determine exactly how much of the differences observed are due to the 'within area' and 'between area' differences of the job arrivals. If there was more confidence in these results and clear evidence existed that geographically closer regions exhibit more similar behaviour, then in general spatial statistics methods would be a possible approach. Spatial statistics is based on the assumption that nearby entities are associated in some way. For more details see for example Diggle (2013), Ripley (2005) and Cressie (2015). However we explain in the following paragraph why we have chosen not to use a spatial model.

In practice, the different regional workforces of BT operate across the UK as follows. Each region is effectively an independent entity that must manage its own backlog by scheduling appropriate workforce engineers. BT experts with extensive system knowledge therefore instructed us to assume that each region operates independently. For example if regions A and B are neighbours and region A has a high backlog of jobs while region B has a low backlog, we can assume that region A will not be allowed to transfer extra engineers from region B and will have to deal with the high backlog independently. So although we have seen that there are higher correlations between series from nearby regions, each region must manage its backlog and workforce independently. Due to this independence of regions and the non-sharing of resources, rather than using a spatial model, we instead apply FIMLOF algorithms (modified further from those in chapter 5) for the analysis of this chapter, where we estimate the parameters of hydraulics model 3 for each of the BT regions.

6.3 Adjustments for the BT data

In this section we explain the adjustments required for estimating hydraulics model 3 parameters from the BT regional data, compared to simulation study 3 in section 5.4.4. These are divided into two sections. Section 6.3.1 outlines the necessary adjustments to the set up of experiments when using the BT dataset. Section 6.3.2 describes further modifications to the FIMLOF algorithm that resulted in improved performance.

6.3.1 Adjustments to the experiments

In this section we explain the necessary adjustments that are required to the set up of experiments when progressing to the BT dataset from simulation study 3. However it is worth highlighting that the analysis in this chapter is much like simulation study 3, albeit we are now estimating its parameters from historical time series data and not simulated data. We assume the same state-space model structure as (5.4.15)-(5.4.24) and estimate the same structural parameters $\underline{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ and variance parameters $\underline{\sigma}^2 = (\sigma^2_{Q_1}, ..., \sigma^2_{Q_5}, \sigma^2_{R_1}, ..., \sigma^2_{R_5})$.

In the simulation studies, for each experiment 100 sets of data were simulated and 5 sets of starting values were used when estimating the parameters of each dataset. This resulted in 500 sets of parameter estimates for each experiment. For the BT data we have 59 regions and hence 59 sets of data. We again use the same 5 sets of starting values for structural parameters $\underline{\theta}$ as simulation study 3. Therefore, for each experiment we have 59 x 5 = 295 parameter estimates.

It is important to distinguish between the two types of models used in section 6.4; the deterministic hydraulics model 3 described in section5.4.1, and the stochastic state-space model constructed from the equations of hydraulics model 3 in section 5.4.2. The BT parameters are sets of parameters for hydraulics model 3. This is a deterministic model as we see from the model equations of (5.4.15)-(5.4.24). Hence these BT parameters include no variance parameters and consist only of the 6 structural

parameters $\underline{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$, described in section 5.4.1. As we see in (4.2.14), calculation of the log-likelihood for our system requires values for the variance parameters $\underline{\sigma}^2 = (\sigma^2_{Q_1},..., \sigma^2_{Q_5}, \sigma^2_{R_1},..., \sigma^2_{R_5})$, the diagonal elements of the covariance matrices Q and R. Hence we cannot compute log-likelihoods when using the BT parameters.

Unlike the three simulation studies, the true parameter values $\underline{\theta}$ are of course unknown for the BT data. Hence we cannot compare our parameter estimates to the true values. Since we also cannot compute log-likelihoods when using the BT parameters, to compare our parameter estimates with BT's, we must use a different approach. Our estimates of the state-space model's structural parameters $\hat{\underline{\theta}}$ are used in the deterministic hydraulics model 3 to simulate data $\hat{\underline{y}}$. This data is compared to the BT historical data $\underline{y}$. We compare the performance of our parameter estimates $\hat{\underline{\theta}}$ with the default parameters $\underline{\theta}_{BT}$ for each region in the process outlined below:

1. Use both estimated $\hat{\underline{\theta}}$ parameters and default $\underline{\theta}_{BT}$ parameters to simulate data $\hat{\underline{y}}$ and $\underline{y}_{BT}$ from hydraulics model 3.

2. Compare output of both simulations with historical data by calculating the prediction error. We choose to calculate the root mean square error (RMSE) of the standardised residuals. Using the standardised residuals takes into account the variance of the residuals. We then compare $\widehat{RMSE}$ with $RMSE_{BT}$, the resulting RMSE's of the standardised residuals for our estimated parameters and the BT parameters respectively, which are defined as below.

$$\widehat{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\hat{y}_t - y_t}{\mathrm{sd}(\underline{\hat{y}} - \underline{y})}\right)^2}$$

$$RMSE_{BT} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_{BT,t} - y_t}{\mathrm{sd}(\underline{y}_{BT} - \underline{y})}\right)^2}$$

Throughout this chapter, unless stated otherwise, RMSE refers to the root mean squared error of the standardised residuals. We can calculate the respective RMSE values across the 59 regions to determine if our parameter estimates $\underline{\hat{\theta}}$ perform better than the defaults $\underline{\theta}_{BT}$. We can also compare performance for individual regions.

An additional set of BT parameters exist for a subset of 8 regions. During experimentation these 8 were amongst the most difficult regions to estimate parameters for that would closely replicate the behaviour of historical data. For these 8 regions, BT analysts hand calibrated hydraulics model 3 and obtained parameter estimates that were an improvement on the default parameters. In section 6.4 we compare our parameter estimates to these hand calibrated estimates.

6.3.2 FIMLOF algorithm modifications

In this section we describe further modifications to the FIMLOF algorithm, compared to algorithm 3. The modifications that we describe were found by experimentation to improve the performance of the algorithm. This improvement was measured in terms of reducing RMSE and obtaining parameter estimates that when used in hydraulics model 3 more closely match the historical data.

In order to reduce the complexity of the problem, we introduce further constraints for the optimisation. For simulation study 3 in section 5.4.4, of the structural parameters $\underline{\theta}$, only parameter θ_3 was constrained; the inverse logit function was used to constrain values between 0 and 1. However the remaining parameters $(\theta_1, \theta_2, \theta_4, \theta_5, \theta_6)$ all represent measures that must be positive. We therefore constrain these in the algorithm using the exponential function as follows: $\theta_i = \exp(\phi_i)$ where ϕ_i can be any real number, for $i = 1, 2, 4, 5, 6$.

In simulation study 3, smoothing splines were used to estimate the noise added to the data and hence set the starting values of the variance parameters in the algorithm. However as we explain in section 6.2.1, some of the time series of the BT data, the people deployed and shrinkage series in particular, are smooth series due to the interpolation that was required to convert the weekly series into daily series. The BT backlog series are also considerably smoother than the simulated series (with added Gaussian noise) which were used in the simulation studies. As such, attempting to set the starting values directly from the data by using techniques such as smoothing splines is not appropriate.

Although we have no prior information on the noise structure for the BT data, a range of starting values for the variance parameters were tested. The most successful of these introduced a relative variance profile for the time series of each region. This is based on a similar principle for the noise added to the data in simulation study 3. As we explain in the study, the 5 time series that represent the stocks; the backlog, people

deployed, shrinkage, overtime and unused overtime have different levels of variability. Consequently when adding noise to these series we wished to take this scaling into account. Similarly for the BT data, when setting the variance starting values for each geographical region, we measure the variance of each of the 5 time series, then divide these by the smallest of the values. For each region the smallest value was for the shrinkage series. We then obtain a relative variance profile for each region r denoted by V_{prof_r}. In the results presented in section 6.4, parameter estimates obtained using different sets of variance starting values are presented. All of these sets use the profile V_{prof_r}.

When BT analysts performed hand calibration on hydraulics model 3, they discovered that adjusting one term in the model that was previously kept constant (with a value of 1), could have a dramatic improvement for some regions. This term is the 'priority power', denoted by PP in (5.4.10). This equation for the tension A_{tn} at time t is effectively a ratio of the backlog and target backlog. For example if the backlog is greater than the target, then tension > 1. A tension greater than 1 indicates that targets are not being met and the model responds by increasing overtime. The priority power is the power to which the tension is raised. Hence it is effectively a measure of how desperate a particular region becomes to reduce the tension when targets are not being met. As we show in section 6.4, our results on the whole are improved by adjusting the FIMLOF algorithm to estimate PP as an additional parameter θ_{PP} for each region. When this is the case, its default value of 1 is used as a starting value in the algorithm.

In section 6.4, we apply this modified FIMLOF algorithm to the BT regional time series data and attempt to estimate the parameters of hydraulics model 3.

6.4 Results

In this section we present the results of parameter estimation on hydraulics model 3 using the regional BT time series data. Ideally, these parameter estimates would be robust to the starting values that are used in the optimisation routine. As we have seen in the simulation studies in chapters 4 and 5, this is not the case. Although many sets of starting values for the structural parameters $\underline{\theta}$ were investigated, reassuringly, changes in these were not found to significantly affect the accuracy of estimates. However, starting values for the two sets of five variance parameters, $\underline{\sigma}^2_Q = (\sigma^2_{Q_1},..., \sigma^2_{Q_5})$ and $\underline{\sigma}^2_R = (\sigma^2_{R_1},..., \sigma^2_{R_5})$, were found to be an important factor in the accuracy of parameter estimates. We denote the starting values for these two sets of parameters as $\underline{\sigma}^2_{Q_0}$ and $\underline{\sigma}^2_{R_0}$.

Consequently we split this section into three parts. In section 6.4.1 we present parameter estimates that result from using the same set of starting values for the variance parameters across all the regions. In section 6.4.2 we examine the effects of allowing the use of different variance starting values for each region. In these first two sections, our parameter estimates are compared with the default BT parameters. In section 6.4.3 the best of our estimates are compared with hand calibrated parameter estimates for a subset of 8 regions.

6.4.1 Fixed variance start values

In this section we present the parameter estimates that result from using the same set of starting values for the variance parameters ($\underline{\sigma}^2_{Q_0}$, $\underline{\sigma}^2_{R_0}$) in the modified FIMLOF algorithm, across all 59 regions. We compare two sets of parameter estimates, 'Set 1' and 'Set 2', with the BT defaults. Set 1 fixes PP while Set 2 estimates PP as an additional parameter θ_{PP}; the reasoning behind this was explained in section 6.3.2.

Set 1 and Set 2 parameter estimates use the same starting values for structural parameters in the algorithm. Although results from the same 5 sets of starting values for the structural parameters as simulation study 3 were investigated (sets s1-s5 in table 5.4.3), these were not found to be a significant factor in the accuracy of parameter estimates. As such the starting values for the structural parameters $\underline{\theta}$ are s3 in table 5.4.3, the default parameter values used in simulation study 3; (2.5, 0.5/7, 0.2, 30, 4, 5).

Set 1 and Set 2 parameter estimates also use the same starting values for the variance parameters. The preferred starting values were found by experimentation. Starting the $\underline{\sigma}^2_Q$ parameters at low values and incorporating the relative variance profile V_{prof_r} (see section 6.3.2) into starting values for the $\underline{\sigma}^2_R$ parameters was found to result in the most accurate parameter estimates. As such, the starting values for the variance parameters take the following form. The five $\underline{\sigma}^2_Q$ parameters start close to 0; $\underline{\sigma}^2_{Q_0} = e^{-10}$. Starting values for the five $\underline{\sigma}^2_R$ parameters are multiplied by the relative

variance profile as follows: $\underline{\sigma}^2_{R_0} = e^x V_{prof_r}$. x was set at 2, but it is worth noting that values of x between 2 and 3 generally produced the best estimates. In section 6.4.2 we investigate the effects of using different $\underline{\sigma}^2_{R_0}$ starting values by adjusting the value of x.

As we saw in section 5.4.1, data for the people deployed and unused overtime series are effectively derived from shrinkage and overtime respectively. Therefore our results focus on the RMSE values for the backlog, shrinkage and overtime. In terms of replicating the behaviour of historical data, BT analysts place the greatest importance on the backlog series, as the length of the queue is a key indicator of performance in any queueing system. Replicating the overtime is considered the next most important, as this characterises how the system responds to periods of high demand. The shrinkage series is considered the least important of the three.

In the process described in section 6.3.1, we compare Set 1 and Set 2 parameter estimates with the BT default parameters by using the structural parameters from each set in hydraulics model 3 and comparing the simulated series to historical data, as we describe in section 6.3.1. Boxplots of the resulting RMSE's of the standardised residuals are presented in Figure 6.4.1. Averages of these RMSE's, over the 59 regions, are shown in table 6.4.1. Set 1 fixes PP at its default value of 1, while set 2 estimates PP as an additional parameter θ_{PP} using a starting value of $PP_0 = 1$ in the optimisation routine.

Parameter Set 1 consists of 16 parameters; the 6 structural parameters $\underline{\theta} = (\theta_1, ..., \theta_6)$

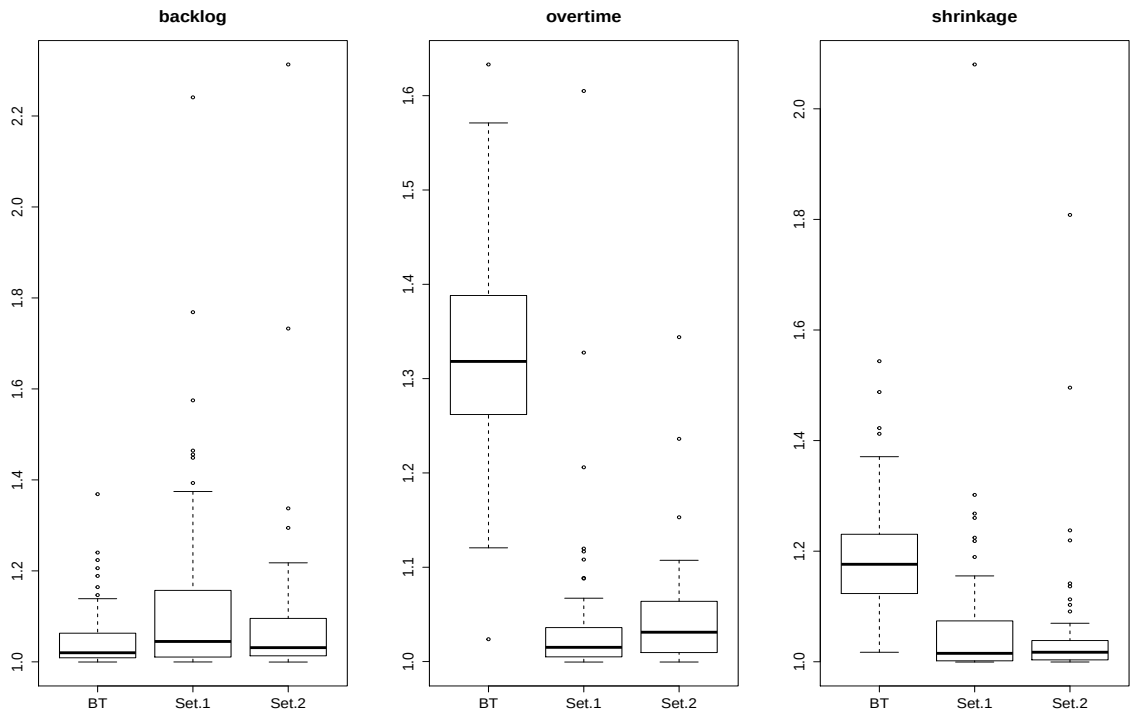

Figure 6.4.1: Boxplots of RMSE of standardised residuals.

Parameter set	backlog	overtime	shrinkage
BT defaults	1.053	1.329	1.190
Set 1	1.137	1.043	1.067
Set 2	1.093	1.047	1.054

Table 6.4.1: Average RMSE of the standardised residuals across the 59 regions.

and 10 variance parameters $\underline{\sigma}^2 = (\sigma^2_{Q_1},..., \sigma^2_{Q_5}, \sigma^2_{R_1},..., \sigma^2_{R_5})$. Set 2 consists of the same parameters as Set 1 and includes an additional structural parameter; θ_{PP}. Hence we can compare the log-likelihoods of these nested models at the MLE's using a likelihood ratio test for each individual region, see for example Cox et. al (1979). As

the difference in parameter numbers between the two models is 1, for each region the test statistic is approximately $\mathcal{X}^2$ with 1 degree of freedom. Results show that for 21 regions, the more complex model, based on Set 2 parameters, gives a significantly better fit. The estimates of θ_{PP} are equally dispersed above and below 1. In addition, the backlog is the most important series from BT's perspective as this is a key indicator of a region's performance. Table 6.4.1 and the boxplots of Figure 6.4.1 show that Set 2 parameters have an improved performance in terms of the backlog RMSE over Set 1 parameters. Hence in section 6.4.2 we continue to investigate the effects of estimating θ_{PP}.

The boxplots of Figure 6.4.1 and table 6.4.1 show that our estimated parameters, Set 1 and Set 2, do bring a significant improvement in terms of the RMSE for the overtime and shrinkage series over the BT default parameters. However, for the most important series - the backlog, our parameters have an increased RMSE and the boxplot reveals some relatively high outliers. In section 6.4.2, we show the significant improvements that result from allowing the use of different sets of $\underline{\sigma}^2_{R_0}$ starting values for each region. For the fixed $\underline{\sigma}^2_{R_0}$ starting values used in this section, we have been unable to provide hydraulics model 3 with parameters to represent the historical data as accurately as we had hoped. We devote the remainder of this section to understanding why this is the case.

It seems reasonable to begin by investigating how well hydraulics model 3 fits the data. To provide us with an initial estimate of this, we select a single region that

is a fair representation of how well the BT default parameters fit the regions on average. As the most important series is the backlog, we select a region with a backlog RMSE that equals the median backlog RMSE of all 59 regions. This region is Derby & Nottingham and the RMSE values are 1.02, 1.21, 1.03 for the backlog, overtime and shrinkage respectively. Comparing these values to the median of the overtime/shrinkage RMSE's across the 59 regions, in the boxplots of Figure 6.4.1, we see that the Derby/Nottingham RMSE's for overtime and shrinkage are considerably better than the medians. Hence this region should be one of the better performing regions using the BT default parameters. The time series from hydraulics model 3 using the BT default parameters for Derby & Nottingham are compared to the historical data in Figure 6.4.2. The black lines represent the historical data and the blue lines represent hydraulics model 3 output.

For the backlog, the most important series, many peaks and troughs of the historical data are generally either grossly exaggerated by the simulated data, or missed out altogether. It is also clear that due to the difference in the scales of the plots, the backlog is by far the poorest fitted of the three series. The fits for the overtime and shrinkage series miss out on many peaks and troughs and are not a good match in terms of shape, though the discrepancies are not as large as those for the backlog. We can also observe that the simulated overtime series is consistently lower throughout than the historical data. It seems clear that simply because the default BT parameters outperform our estimates, it does not necessarily make them good parameters. As we have seen, even regions that are 'better' fitted are not a good representation of

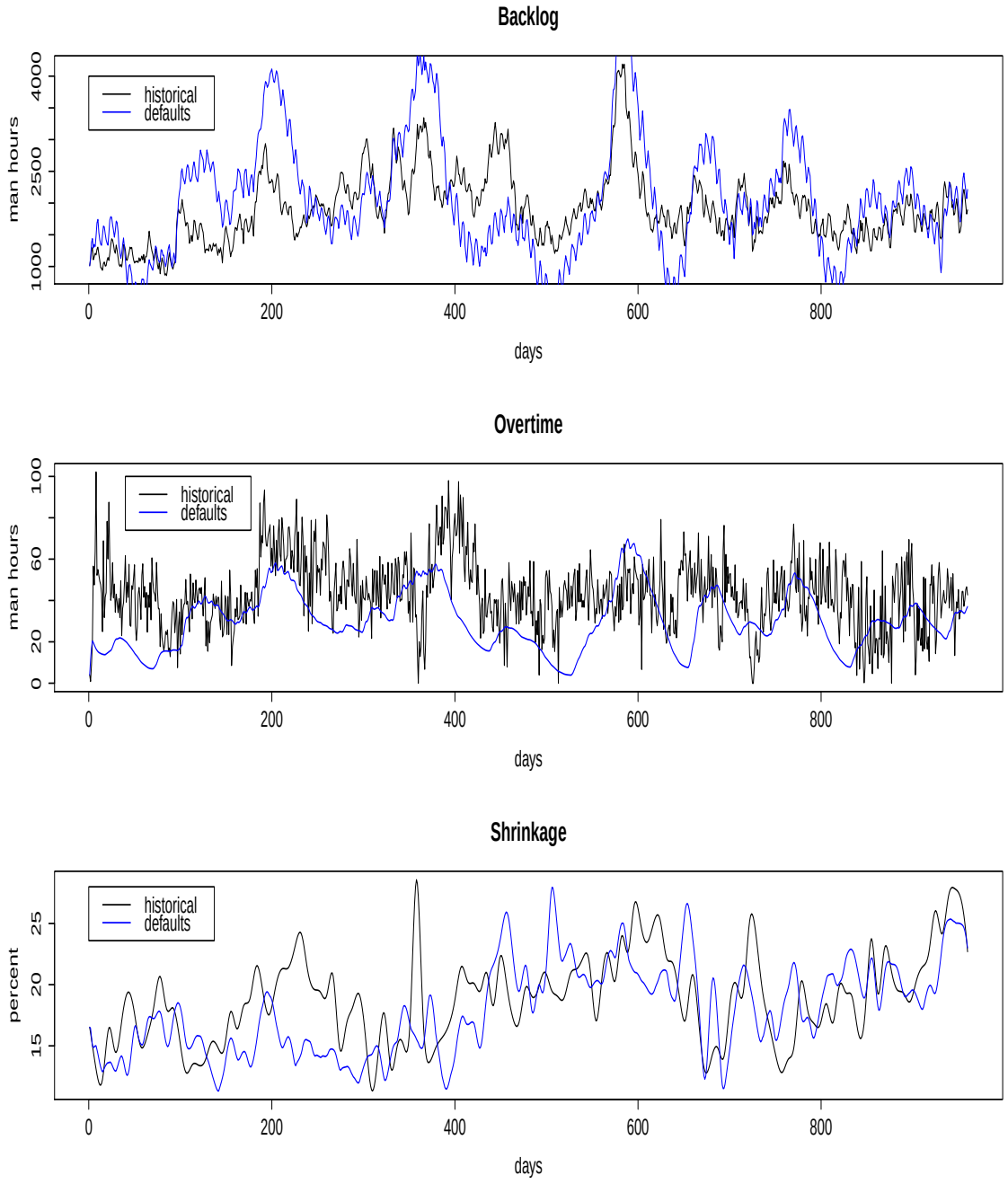

Figure 6.4.2: Simulated series from hydraulics model 3 for the Derby & Nottingham region using default BT parameters, compared to historical data.

the historical data.

One possibility is that the algorithm is repeatedly becoming stuck in local optima, leading to inaccurate parameter estimates. To investigate this, we repeated attempts at parameter estimation for a sample of the 10 regions that performed poorest in terms of RMSE, this time using carefully selected starting values that lie close together. The objective was to determine whether use of start values lying close together would result in sufficiently similar values of parameter estimates. However, results did not provide any evidence that the algorithm is becoming stuck in local optima.

We now consider the BT regional dataset. As we explained in section 6.2, the historical time series suffer from a number of issues. Firstly, we were only provided with weekly time series data for endogenous variables people deployed $\underline{p}^*$ and shrinkage $\underline{s}^*$ and also the exogenous variable total employees $\underline{E}_{te}$. Although cubic spline interpolation was used to convert these to daily series, we do not know how accurately these series represent the true historical data. Secondly, to ensure that the overtime $\underline{o}^*$ and unused overtime $\underline{uo}^*$ data remained in the same units as hydraulics model 3, units had to be changed in these series from the data we were initially provided with. This required dividing by certain weekly profiles which caused numerous unnatural spikes in the series as we demonstrated in Figure 6.2.1. In many regions these spikes were considerable. This means that of the 5 time series that represent the endogenous variables for each region, 4 of these possess flaws in that they have either been interpolated from weekly data or possess unnatural spikes. This is likely to have affected

the accuracy of our parameter estimates for each region.

In addition we can observe the following about the dataset. Like many large organisations, BT possess vast databases that record data from a number of different sources. Over time, the methods of recording data and the types of sources can change. Consequently, inconsistencies can creep into the data. Although the data underwent extensive cleaning by BT analysts prior to its use in this analysis, they admitted that it was unlikely that all of these inconsistencies were removed, leading to further question marks surrounding the data. An example of this is the exogenous series for the total employees $\underline{E}_{te}$ for each region. BT analysts admitted that the true values could potentially be 10% lower or higher than the series provided. From the difference equations of hydraulics model 3 (5.4.1)-(5.4.5), we see that total employees $\underline{E}_{te}$ affects the overall level of the shrinkage $\underline{s}^*$ and unused overtime $\underline{uo}^*$ series. Hence any inconsistencies in the $\underline{E}_{te}$ series would potentially further impact our results.

Our modified FIMLOF algorithm requires the errors to be Gaussian. However if we re-examine the time series of Figure 6.4.2, representing the historical BT data (black line) for the Derby & Nottingham region, even a simple visual inspection reveals that this is assumption is questionable. The shrinkage $\underline{s}^*$ series, due to the interpolation that was required, is smooth in appearance. The backlog $\underline{b}^*$ series is less smooth, but this is due to the presence of a slight weekly profile. The overtime $\underline{o}^*$ possesses many unnatural spikes. These characteristics are typical for each series across all the regions.

One way to investigate the assumption of Gaussian noise is by taking a centred 3-point moving average of each series, subtracting this and inspecting the residuals. Figure 6.4.3 presents Q-Q plots of these residuals for the Derby & Nottingham region. We can see that none of the plots support the assumption of Gaussian residuals. In addition, it is worth highlighting that these Q-Q plots are typical across all the regions. We can investigate more formally as follows. For each of the 59 regions we have 5 time series (each of length 959 days) representing the endogenous variables; $\underline{b}^*$, $\underline{p}^*$, $\underline{s}^*$, $\underline{o}^*$ and $\underline{uo}^*$. Hence we have 295 time series altogether. We can use the Shapiro–Wilk test (Shapiro and Wilk, 1965) to investigate the normality of the residuals for each of the the 295 series. From the resulting p-values of these tests, only 12 of the 295 p-values are greater than $\alpha = 0.05$, meaning that for 283 of the 295 series, we can reject the null hypothesis (that these residuals are Gaussian) at the 95% confidence level.

We can also examine the residuals when comparing our state-space models to the historical data. To investigate the residuals for each region, we use the region's estimated parameters (from parameter Set 1 of table 6.4.1) and exogenous variables in the state-space model, and simulate 100 sets of time series representing the 5 endogenous variables; $\underline{b}^*$, $\underline{p}^*$, $\underline{s}^*$, $\underline{o}^*$ and $\underline{uo}^*$. We then obtain 100 sets of residuals for each of the 5 time series for each region by taking the difference between these simulated series and the historical data. Across all the regions this gives us 100 x 59 = 5900 sets of residuals for each of the 5 time series. Performing the Shapiro–Wilk test on all these sets of residuals, table 6.4.2 shows the percentages where the resulting p-values can

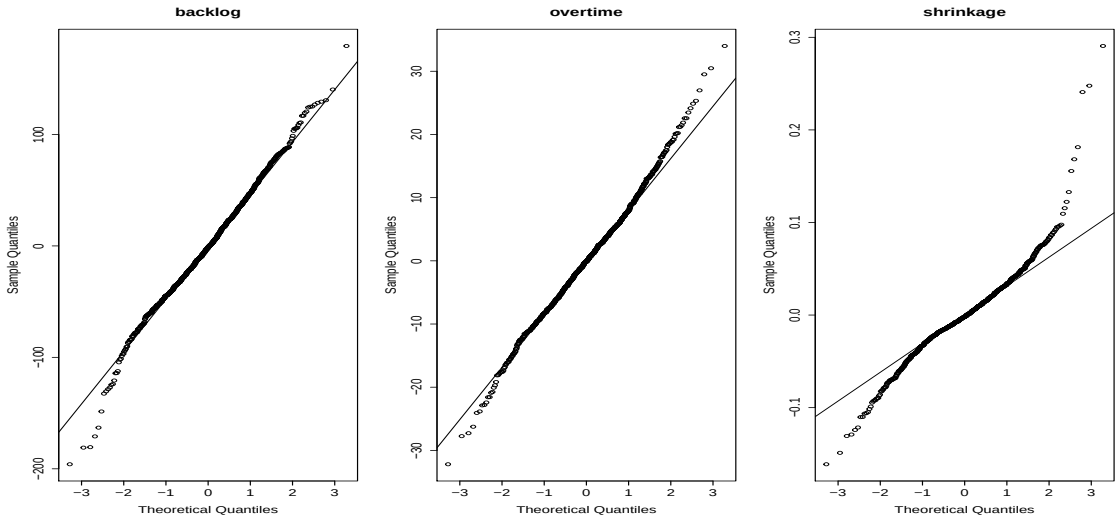

Figure 6.4.3: Q-Q plots of residuals for Derby & Nottingham region.

reject the null hypothesis (that residuals are Gaussian) at the 95% confidence level, for each of the 5 series. The high percentages of Shapiro–Wilk tests that reject the null hypothesis raise serious questions regarding the assumption of Gaussian residuals.

Time series	backlog	overtime	shrinkage	people deployed	unused overtime
Percentage	81.2	56.5	44.9	48.6	57.7

Table 6.4.2: Percentage of p-values resulting from Shapiro–Wilk tests that reject the null hypothesis at the 95% level.

We can investigate more closely by examining an individual region. Figure 6.4.4 shows the standardised residuals resulting from 1 realisation of the state-space model for the Derby & Nottingham region. In addition, the blue lines connect the standardised residuals that result from the use of the structural parameters only; that

is, using the structural parameters in the deterministic hydraulics model 3, rather than the state-space model. For this single realisation of the state-space model, when performing a Shapiro–Wilk test on the residuals for backlog, overtime and shrinkage, the resulting p-values are 0.60, 0.25 and 0.11 respectively - hence there is insufficient evidence to reject the null hypothesis. These residuals may not 'fail' the Shapiro–Wilk test of normality, however Figure 6.4.4 reveals that there are clear trends apparent in the residuals. This raises questions as to how well the model is fitting the historical data; there is some behaviour that our model is not able to explain. Although Figure 6.4.4 shows residuals for a single realisation of the state-space model for a single region, this is a typical example of the residuals observed for other realisations of the state-space models for the majority of the regions.

It is worth investigating the estimated values of the variance parameters, $\underline{\sigma}^2 = (\sigma^2_{\underline{Q}}, \sigma^2_{\underline{R}}) = (\sigma^2_{Q_1},..., \sigma^2_{Q_5}, \sigma^2_{R_1},..., \sigma^2_{R_5})$ of parameter Set 1 of table 6.4.1. We present the averages of these in table 6.4.3. It is clear that values of $\hat{\sigma}^2_{\underline{R}}$ are considerably greater than corresponding values $\hat{\sigma}^2_{\underline{Q}}$, especially for the backlog, overtime and unused overtime series. These 'signal to noise' ratios (see section 4.2.3) suggest that the state-space models place more trust in the deterministic hydraulics model 3 equations, rather than the historical data. Note that similar ratios were observed when using a variety of starting values in the algorithm, whether we set $\underline{\sigma}^2_{Q_0}$ starting values low and $\underline{\sigma}^2_{R_0}$ starts high, or vice versa.

We can observe in Figure 6.4.4 that the standardised residuals of the deterministic

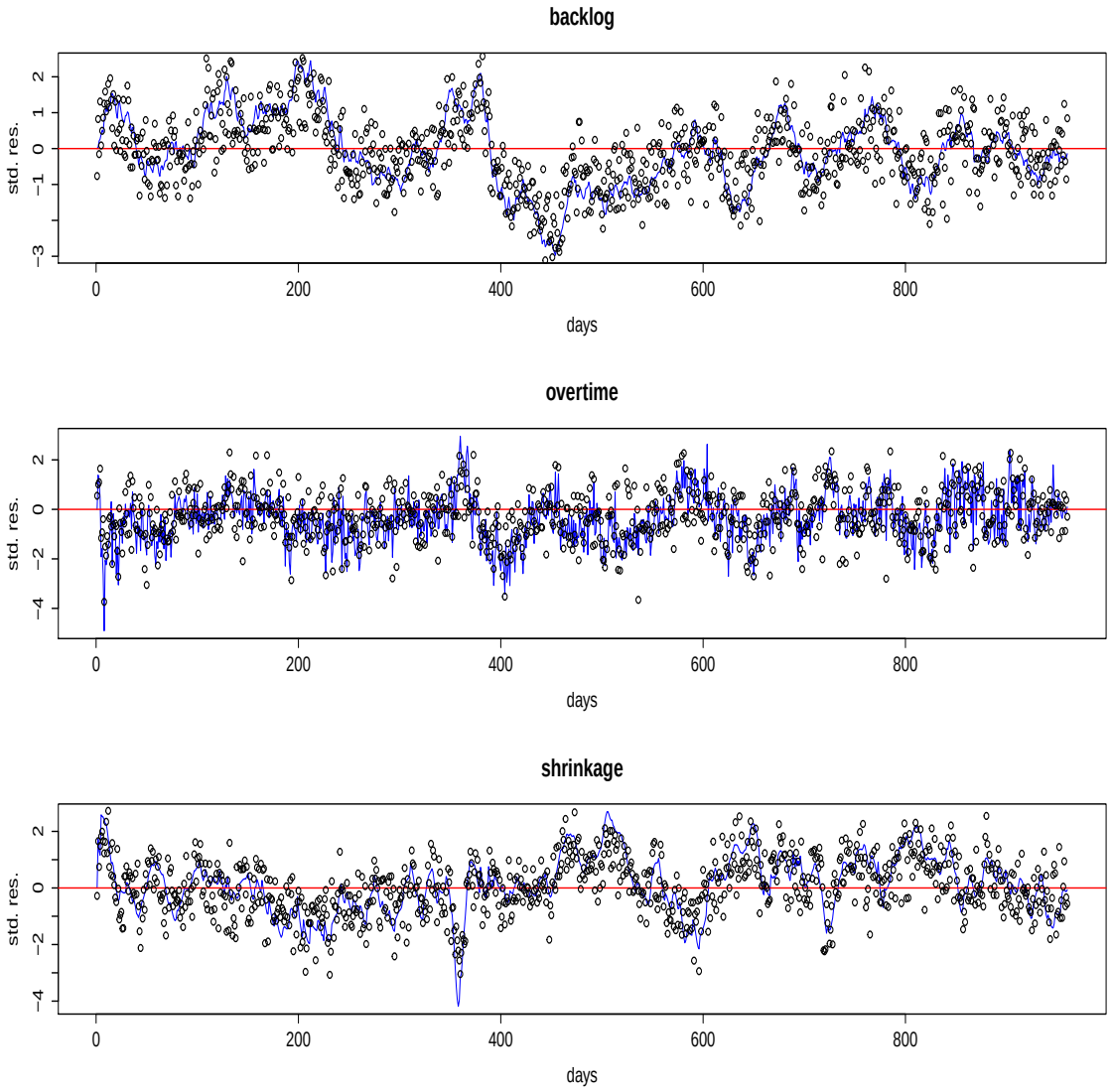

Figure 6.4.4: Standardised residuals for Derby & Nottingham region from 1 realisation of state-space model.

hydraulics model 3 - the blue lines - appear to be further away from having a Gaussian distribution than the residuals in the same figure from the state-space model. This is confirmed when we examine p-values from Shapiro-Wilk tests of these residuals across all regions for the 59 x 5 = 295 series. The null hypothesis (that the residuals

Parameters	backlog	people dep.	shrinkage	overtime	unused overtime
$\hat{\sigma}^2_{\underline{Q}}$	0.01	0.07	3.77	5.87	10.34
$\hat{\sigma}^2_{\underline{R}}$	3.45 x 10^5	7.95	8.00	396.81	484.90

Table 6.4.3: Average of estimated variance parameters for parameter Set 1.

are Gaussian) is rejected for 98.3%, 98.3%, 98.3%, 89.8% and 89.8% of the time series that represent the endogenous variables $\underline{b}^*$, $\underline{p}^*$, $\underline{s}^*$, $\underline{o}^*$, $\underline{uo}^*$ respectively. As we see from table 6.4.2, residuals from the state-space models fail the normality tests less often than residuals for the deterministic hydraulics model 3. In other words, the residuals from the state-space models (with their high-valued $\hat{\sigma}^2_{\underline{R}}$ estimates), are generally closer to having Gaussian distributions than the residuals from the deterministic hydraulics model 3.

One possibility for the high-valued $\hat{\sigma}^2_{\underline{R}}$ parameter estimates in the state-space models is that this is out of necessity in order to ensure Gaussian residuals - or residuals that are as close to being Gaussian as possible. The large $\hat{\sigma}^2_{\underline{R}}$ estimates of table 6.4.3 show that the observational noise parameters are the dominant terms. As these noise terms are Gaussian, then for sufficiently large $\hat{\sigma}^2_{\underline{R}}$ values, the residuals will similarly appear approximately Gaussian. These results add further weight to the concern that the assumption of Gaussian errors may not be suitable for this BT dataset. In section 6.5 we discuss alternative methods that may be more suitable for this dataset.

In this section we have attempted to explain why our estimated parameters are not allowing hydraulics model 3 to represent historical data as accurately as we hoped. Ideally, the variance starting values would not affect the parameters estimated from the FIMLOF algorithm – but they do. Accepting that this is the case, the question becomes the following; *despite* this issue, how can we improve the accuracy of our parameter estimates from those obtained in this section? In section 6.4.2 we investigate the effects of selecting the 'best' variance starting values for each region.

6.4.2 Adjusting variance start values

In this section we investigate whether our parameter estimates are improved by allowing use of the 'best' starting values for the variance parameters for each region.

It is reasonable to suppose that hydraulics model 3 will fit some regions better than others. As such, different values of the variance parameters would result between these regions when estimating parameters using FIMLOF. Due to the lack of robustness of the approach to the starting values used in the optimisation that we describe in the previous section, we select the 'best' set of variance starting values to use for each region. The starting values for the structural parameters remain unchanged.

In table 6.4.4, we present two sets of parameter estimates and compare these to the BT defaults. Again Set 1 involves fixing PP in the model while Set 2 estimates this as an additional parameter θ_{PP}. For each region, results for both Set 1 and Set 2 were obtained using the 'best' set of variance starting values, from a choice of 5

sets. These best sets were determined by the value of the RMSE for the backlog, the most important series in the model to fit to the data. The best sets were found by experimentation. For each of these 5 possible sets, the $\underline{\sigma}^2_{Q_0}$ starting values remain at e^{-10} while the $\underline{\sigma}^2_{R_0}$ start values took the following form: $\underline{\sigma}^2_{R_0} = e^x \, V_{prof_r}$ with x taking values of 0, 1, 2, 3 and 5.

Parameter set	backlog	overtime	shrinkage
BT defaults	1.053	1.329	1.190
Set 1	1.017	1.080	1.068
Set 2	1.009	1.037	1.019

Table 6.4.4: Average RMSE of the standardised residuals across the 59 regions.

We now see a noticeable improvement in the results. Compared to fixing the variance start values across the regions in table 6.4.1, the backlog RMSE has decreased from 1.137 to 1.017 for Set 1, and from 1.093 to 1.009 for Set 2. Both these values are now an improvement over the backlog RMSE of the BT defaults; 1.053. The overtime and shrinkage RMSE for Set 1 and Set 2 remain a considerable improvement over the defaults, with Set 2 improving further on the results for fixed variance starting values, while Set 1 show slight increases. This outcome was expected by BT analysts. The 'best' variance starting values were selected for each region based on the resulting backlog RMSE. However as the backlog is fitted more accurately and backlog RMSE reduces, other series may not fit as well.

These results are based on estimates obtained using 5 different sets of variance starting values for each region. An interesting question would be to ask what would happen if we choose from a greater number of sets? We investigate this using an additional 6 sets of starting values; giving us 11 sets altogether. The $\underline{\sigma}^2_{R_0}$ start values took the following form $\underline{\sigma}^2_{R_0} = e^x V_{prof_r}$ with x taking values of 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 and 5. The reasoning for the values of these additional 6 sets is as follows. The best starting values $\underline{\sigma}^2_{R_0}$ found previously were towards the middle of the range at $x = 2$ and 3. So rather than extending the range by using $x < 0$ or $x > 5$, use of these smaller intervals within the range was investigated. The results of selecting the best of both 5 and 11 sets of starting values $\underline{\sigma}^2_{R_0}$ for each region are shown in the boxplots of Figure 6.4.5 and table 6.4.5. Note that in the boxplots, parameter set 'S1A' denotes Set 1A parameters - that is Set 1 parameters (where PP is fixed) chosen from 5 sets of start values. Set 'S1B' denotes Set 1 parameters chosen from 11 sets of start values. Similarly for Set 2 (where PP is estimated as an additional parameter θ_{PP}) with 'S2A' and 'S2B'.

With a greater choice of variance starting values, Set 1B is a further improvement over Set 1A for all three series, in terms of the average RMSE, as we see from table 6.4.5. Similarly for Set 2B. From the boxplots we see that Set 1B and Set 2B are a considerable improvement over Sets 1A and 2A respectively, in terms of the backlog RMSE. However in terms of overtime and shrinkage, the median RMSE increases. As in section 6.4.1, improving the fit for one series, the backlog, has decreased the fit

Parameter set	Sets of $\underline{\sigma}^2_{R_0}$	backlog	overtime	shrinkage
BT defaults	-	1.053	1.329	1.190
Set 1A	5	1.017	1.080	1.068
Set 1B	11	1.003	1.056	1.034
Set 2A	5	1.009	1.037	1.019
Set 2B	11	1.003	1.036	1.017

Table 6.4.5: Average RMSE across the 59 regions – choosing from both 5 and 11 sets of start values.

for the remaining series; the overtime and shrinkage. However this is to be expected when selecting the best sets of starting values based solely on the backlog RMSE. It is also worth highlighting that the performance of the results of parameter Sets 1B and 2B are a further improvement, not only on Sets 1A and 2A, but also on the default parameters. For example the backlog RMSE for parameter Set 2B is an improvement over the BT default parameters for 58 out of 59 regions.

We now compare the performance of models with fixed PP and estimated θ_{PP}, i.e. we compare the performances of Set 1A with 2A and Set 1B with 2B. Sets 2A and 2B have lower average RMSE's than Sets 1A and 1B respectively, for all series. From the boxplots we can see that Set 2A plots are more favourable than Set 1A. Comparing boxplots for Set 1B and 2B is less clear cut, though the backlog plot, the most important series, is more favourable for Set 2B than for 1B. On balance it appears slightly

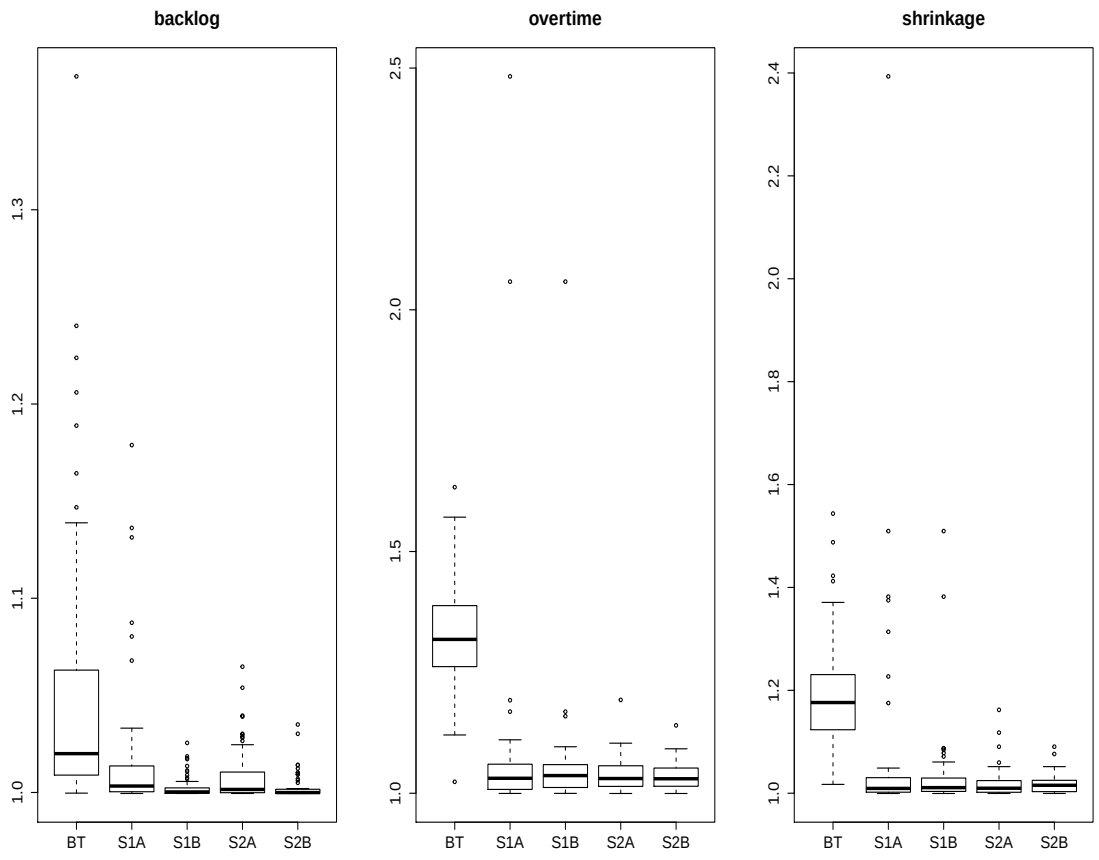

Figure 6.4.5: Boxplots of RMSE of standardised residuals.

more favourable to use parameter Set 2B over Set 1B - or in other words to estimate PP as an additional parameter θ_{PP} rather than fixing it. The improved performances of Sets 2A and 2B over Sets 1A and 1B respectively suggests that hydraulics model 3 more accurately represents the regional systems when estimating θ_{PP} for each region. In other words, setting this to its default value of 1 may be missing key aspects of the differences in behaviour across the regions.

We now demonstrate graphically the improvement resulting from choosing the 'best'

variance starting values for each region from a set of 11. In Figure 6.4.6 we use the parameters of Set 2B in hydraulics model 3 and compare the resulting time series to time series generated by the BT default parameters. Again the region selected is Derby and Nottingham. As we explained in section 6.4.1, for this region, the BT default parameters result in the median backlog RMSE but perform better than the median for overtime and shrinkage. Set 2B parameters for this region have a slightly poorer RMSE than the Set 2B median RMSE's for backlog, overtime and shrinkage; 1.001, 1.031 and 1.018 compared to 1.000, 1.029 and 1.015 respectively.

The black lines in the time series again represent the historical data and the blue lines represent hydraulics model 3 output using the default parameters. The red line represent hydraulics model 3 output using estimated parameters, Set 2B. The backlog for Set 2B parameters is a noticeable improvement, with peaks considerably less exaggerated than for the defaults. Although hydraulics model 3 using Set 2B parameters still misses certain peaks that the default parameters miss. The overtime also shows some improvement.

In this section we have demonstrated considerable improvements to the results by allowing the use of different $\underline{\sigma}^2_{R_0}$ starting values for each region. However, these results are not without their flaws. Inspecting time series for certain regions reveals some undesirable behaviour. An example of this can be seen for the region representing Reading, the time series of which are shown in Figure 6.4.7.

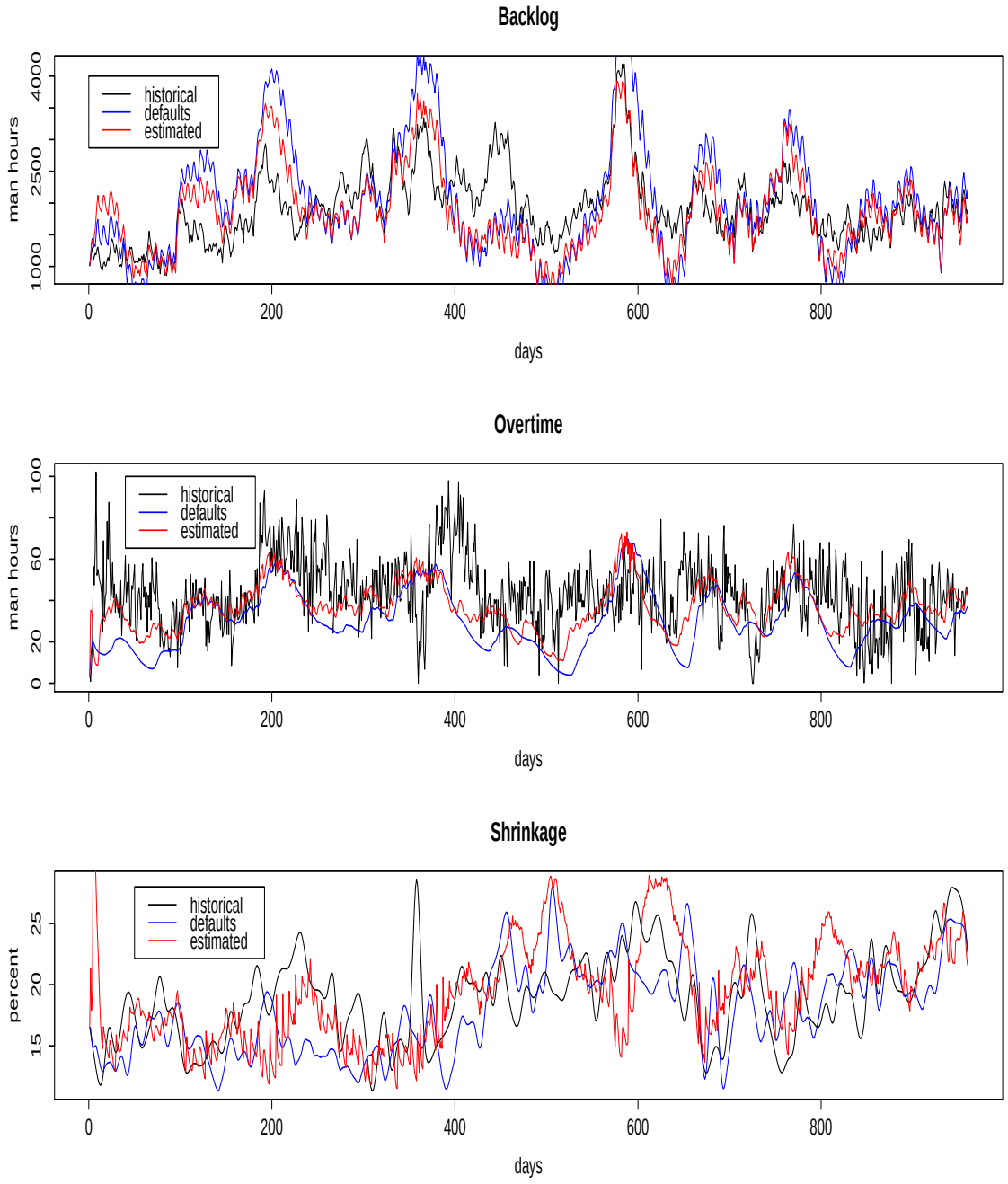

Figure 6.4.6: Comparing our estimated parameters with default BT parameters and historical data for the Derby & Nottingham region.

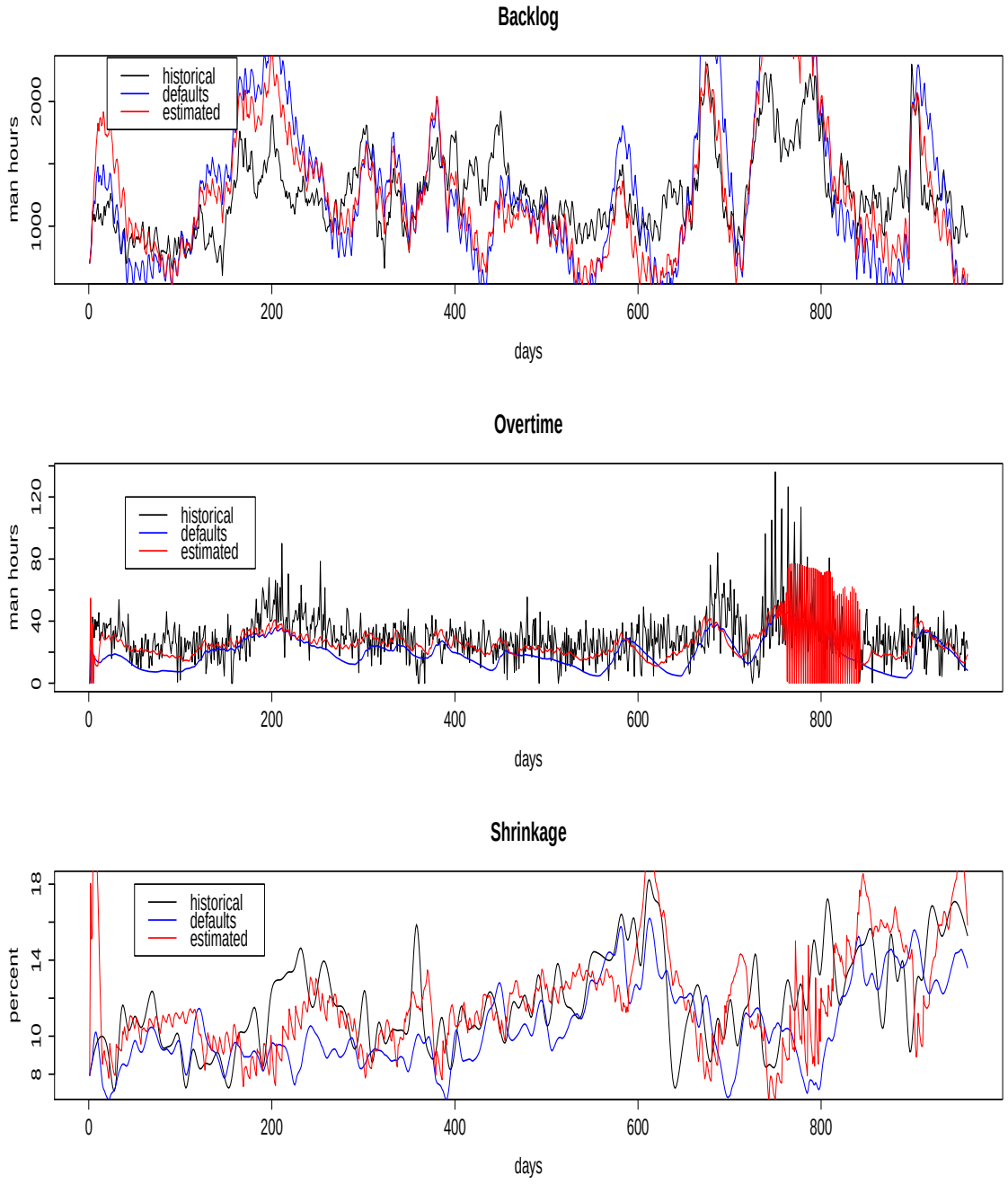

Figure 6.4.7: Strange behaviour of simulated overtime series using our parameter estimates for the Reading region.

Overall the series from our estimates are an improvement over the defaults for both backlog and overtime. However closer inspection of the overtime series resulting from our estimates between days 750-850 reveals sharp consecutive increases and decreases in the series. This is the result of some unexpected parameter estimates that from a practical point of view do not make sense. For example θ_5, the delay associated with moving employees to overtime, is estimated at 1.62 days. BT expect this value to be around 4 days. In addition, θ_1, the target cycle time, is estimated at 0.42 days rather than the expected 2.5 days. This balance of parameters means that the system is attempting to clear jobs faster than it should and adjusting overtime levels too often. These unexpected parameter estimates are also observed when fixing PP at its default values. Allowing this to be estimated as a parameter has not caused this behaviour. Also, this undesirable behaviour is not limited to the overtime series; the shrinkage in certain regions exhibits similar patterns. So, although on average our results are an improvement over the defaults, having an improved backlog RMSE for 58 out of 59 regions, the overtime and shrinkage for a small number of regions exhibit undesirable behaviour.

It is worth highlighting however that this behaviour could be prevented. The 'best' $\underline{\sigma}^2_{R_0}$ start values were selected in this section based entirely on the resulting backlog RMSE value, when comparing simulated data to historical. If instead, the best $\underline{\sigma}^2_{R_0}$ start values were selected using the RMSE values for backlog, overtime and shrinkage, it is unlikely that we would observe the undesirable behaviour of Figure 6.4.7 in the simulated data. The trade off here would be that the backlog RMSE values would

show slight increases.

In the last two sections we have compared our parameter estimates with the default BT parameters. We now compare the estimates to parameters that have been hand calibrated by BT analysts.

6.4.3 Comparing to hand calibrated parameter estimates

In this section we compare the best of our results from section 6.4.2 with a set of hand calibrated parameters. The hand calibration was performed by a Principal Research Scientist at BT with the aim of determining whether this method could find parameters that were an improvement on the defaults. The 8 most difficult regions to estimate parameters for, i.e. those with the greatest RMSE values, were selected for this. It is likely that the data for these regions is the least well fitted to hydraulics model 3 compared to other regions due to differences in behaviour across the regional systems.

The results are shown in Figure 6.4.8 and table 6.4.6. Rows 1 and 2 of table 6.4.6 represent the average RMSE values across the 8 regions calculated by comparing hydraulics model 3 output with the historical data, for the default and hand calibrated parameters respectively. In the boxplots of Figure 6.4.8 these are denoted 'BT' and 'HC' respectively. Rows 3 to 5 represent parameter set 2, where PP is estimated as an additional parameter, with different choices of sets of starting values. In row 3, a single set of variance start values $\underline{\sigma}^2_{R_0}$ are used throughout, the same as those from

section 6.4.1; $\underline{\sigma}^2_{R_0} = e^x V_{prof_r}$ with $x = 2$. For rows 4 and 5 we use parameter sets Set 2A and Set 2B of section 6.4.2, where the 'best' $\underline{\sigma}^2_{R_0}$ start values are chosen for each region from a set of 5 and 11 respectively. In the boxplots of Figure 6.4.8 these are denoted 'S2', 'S2A' and 'S2B' respectively.

Parameter set	Sets of $\underline{\sigma}^2_{R_0}$	backlog	overtime	shrinkage
BT defaults	-	1.044	1.327	1.197
BT hand calibrated	-	1.064	1.042	1.157
Set 2	-	1.316	1.079	1.098
Set 2A	5	1.006	1.054	1.011
Set 2B	11	1.001	1.042	1.013

Table 6.4.6: Average RMSE of the standardised residuals across the 8 difficult regions.

Table 6.4.6 and Figure 6.4.8 demonstrate the improvements of using our estimated parameters Set 2A and Set 2B, compared to both the BT defaults and hand calibrated parameters. In terms of the average RMSE's, both sets are a considerable improvement on the default parameters for all three series, and are an improvement on the hand-calibrated parameters in all but one case; Set 2A has a greater average overtime RMSE. Parameter Set 2B appears to perform best of all the parameter sets when inspecting the average RMSE's. Although the boxplots reveal that Set 2B is outperformed by the hand-calibrated parameters for the overtime series, Set 2B has a considerably lower mean and median RMSE for shrinkage and also backlog - the most

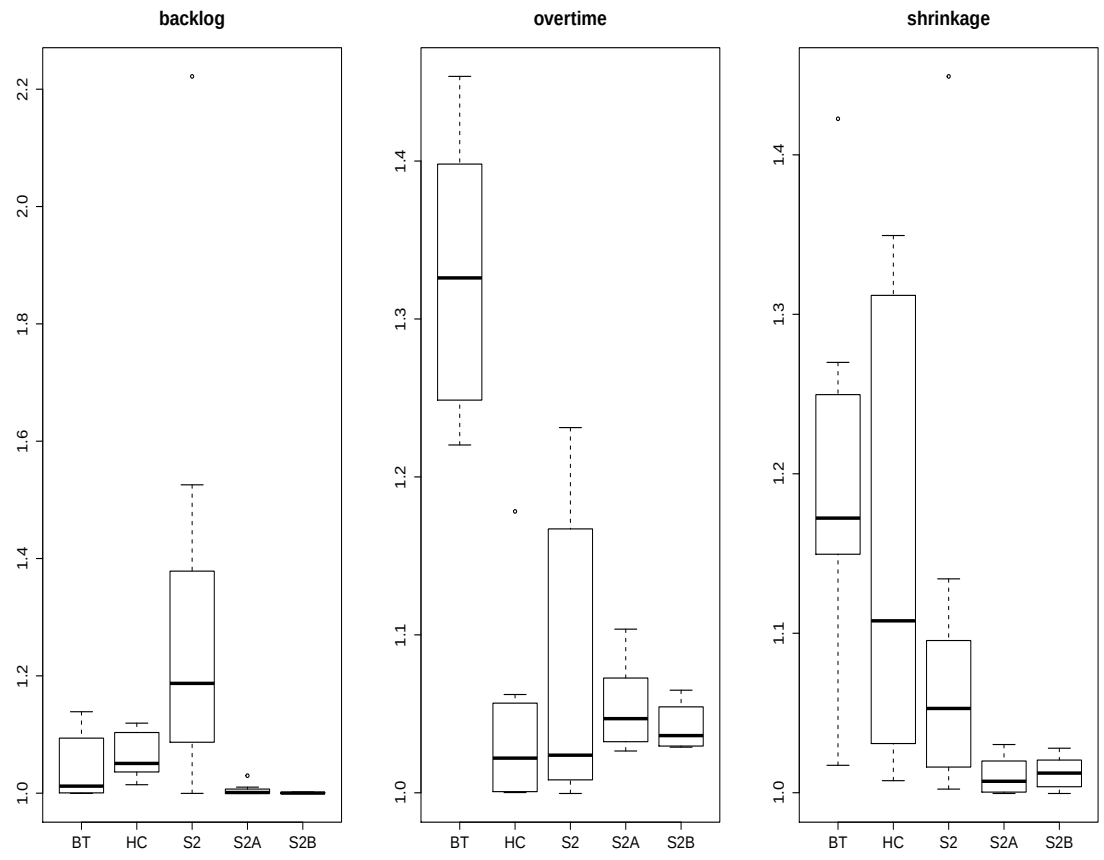

Figure 6.4.8: Boxplots of RMSE of standardised residuals.

important series. Once again, allowing more fine-tuning for each region, in terms of selecting from a wider range of $\underline{\sigma}^2_{R_0}$ start values, has improved performance. If we accept the current limitations of our approach – i.e. its lack of robustness to the starting values in the algorithm – this represents a clear improvement over the BT defaults and hand calibrated parameters for these difficult regions. This improvement is still evident for parameter Set 2A when reducing the choice of start values $\underline{\sigma}^2_{R_0}$ from 11 to 5.

It should also be noted that Set 2, with fixed $\underline{\sigma}^2_{R_0}$ start values, has the worst backlog

RMSE – considerably poorer than both the default and hand-calibrated parameters. Hence this parameter set shows an even poorer performance for these 8 difficult regions than the results for all the regions in section 6.4.1. For some of our estimated parameters, for example Set 2A, the improvements to the backlog RMSE come at a price; we slightly lose performance for the overtime series. In addition, the undesirable behaviour of the overtime and shrinkage series exposed in Figure 6.4.7 of section 6.4.2 is again apparent for at least a short interval in one of these series in 5 of the 8 regions.

As we see from table 6.4.6, the average RMSE for backlog is actually greater using hand-calibrated parameters than for the defaults. It is important to highlight that the BT hand-calibration was performed in an attempt to minimise the RMSE of the raw residuals, *not* the RMSE of the standardised residuals. However, if we pause to examine the RMSE of the raw residuals, the BT default parameters have averages of 902.24, 21.60 and 4.21 respectively for backlog, shrinkage and overtime. The BT hand-calibrated parameters are a slight improvement on the defaults for the two most important series backlog and overtime, 880.79 and 20.80 respectively, at the expense of poorer average shrinkage RMSE of 4.96. As the backlog is the most important series, if the hand-calibration had aimed to minimise the RMSE of the standardised residuals, then the average of the backlog RMSE would be an improvement over the BT defaults.

The hand-calibration was performed by a Principal Research Scientist at BT with extensive experience of modelling (and in particular system dynamics) and the BT

systems/data. This scientist also independently produced the hydraulics model. In other words, there is surely nobody who is better qualified to hand-calibrate the model for individual regions. However, whether we examine the RMSE's of the raw residuals or standardised residuals, it is clear that although the hand-calibrated parameters are an improvement over the defaults, this improvement is not considerable.

Although we highlight the limitations of the hand calibration school of thought in section 4.1.2, there are studies that claim this approach shows promise. In particular, Lyneis and Pugh (1996) claim after an experimental study that "hand calibration works, and is less of an art and is more replicable than might be expected. Moreover, it produces results which are as close to the true values as automated calibration, and are typically close enough to make no significant difference to the outcome of policy interventions". The relatively small improvement in performance that we have observed using the hand-calibrated parameters over the defaults suggests one of the following explanations; either hydraulics model 3 is simply not a sufficiently accurate representation of the behaviour of individual regional systems, or the data issues described in sections 6.2.1 and 6.4.1 (such as necessary interpolations, transformations and potential inaccuracies of series such as E_{te}) have caused significant discrepancies between hydraulics model 3 and the data.

The optimisation routine in the FIMLOF algorithm uses the log-likelihood and as such, does not prioritise any of the 5 time series. We have selected the best of our results based on the backlog RMSE as this is the most important series to BT. The

drop in performance that we have on some occasions observed for the overtime and/or shrinkage series when improving the backlog fit, not only in this section but in the previous two, could be explained as in the previous paragraph; by either hydraulics model 3 not sufficiently representing the behaviour of the regional systems, or the data issues.

6.5 Discussion

In this chapter we have attempted parameter estimation for state-space models that are based on the equations of hydraulics model 3, for 59 regional sets of historical BT time series data. Parameter estimates were initially poorer than expected, offering no improvement over the defaults. This exposed a lack of robustness to the starting values in the algorithm for the variance parameters. Compared to the simulation studies 1-3 (which added Gaussian noise) where this issue was not apparent, this issue had a major influence on results for the BT data. In section 6.4.1 we provide evidence that raises concerns regarding the assumption of Gaussian noise for the BT regional data – which is a possible explanation for why the starting values of the variance parameters were important. Issues with certain time series within the dataset and the necessary interpolation of weekly series into daily were also considered as potential additional reasons for the poor performance.

Selecting the 'best' $\underline{\sigma}^2_{R_0}$ start values for each geographical region, as in sections 6.4.2 and 6.4.3, may not be ideal, but our results demonstrate that it is useful. This provides

BT with an approach that results in significant improvements to parameter estimates. Although this is not the fully automated approach that we hoped to achieve, it does constitute methodology that can be used by BT. It is ready to be implemented immediately using our estimated parameters and can be used to recalibrate the model as the data changes in future. We argue that when necessary for certain parameters such as $\underline{\sigma}_R^2$ here, the selection of the appropriate starting values is an additional factor that requires calibration. It should also be emphasised that although we adjust the starting values for the optimisation, the parameters are still being estimated using FIMLOF. The optimisation algorithm has not been influenced in any way.

In our view this represents the most complex application of FIMLOF yet seen. Hydraulics model 3 is an extremely complex and nonlinear system dynamics model that is used by a large multinational organisation. Calibrating the model using the regional data is not an easy task. We also argue that this is one of the most thorough expositions of FIMLOF. The results exposed not only the limitations of the approach, e.g. the lack of robustness to the starting values, but also included the full effects of these. Consequently we were able to devise a solution that while not perfect, offers a practical solution to the organisation until advances can be made.

Chapter 7

Conclusions

This thesis investigates the feasibility and value of implementing recent developments in operational research and statistics, in order to enhance system dynamics models. This is undertaken in three key areas:

- Adding a discrete-event model to capture greater levels of detail, track individuals within the system and represent stochastic variation in the behaviour of these individuals (chapter 2).
- Using data-driven techniques to assess core assumptions of SD models from historical time series (chapter 3).
- Estimating the parameters of SD models using state-space models and Kalman filtering (chapters 4-6).

In chapter 2, results from the SD/DES hybrid model demonstrated some of the benefits of this approach for BT's system, compared to using a standalone SD or DES model. Incorporating the feedback from the SD model equations enabled the hybrid

to replicate the decisions of management to adjust workforce numbers. By representing individual job and engineers in the DES part of the model, the hybrid was able to directly model the important performance measure RFT. This enabled us to understand the size of the effect of increasing the standard deviation of service times on RFT, which would not have been possible without the hybrid model. Use of the hybrid also demonstrated the potential for misleading results, when relying on BT's current process of estimating RFT via regression. These results add to the literature by providing another example of the benefits of hybrid modelling over a standalone SD or DES model.

A limitation of the hybrid model was that it was based on a scaled–down, simplified version of BT's system. However, the relatively simple hybrid model can be considered as a 'proof of concept' that such models can be constructed in dedicated DES software, with the caveat that such software offers a programming language facility to enable the coding of the SD model's difference equations. However, our choice of a DES environment posed a number of additional challenges that would need to be overcome. The chapter outlines these in order to assist future researchers considering hybrid modelling within a DES environment. The successful operation of the model demonstrated that these challenges can be overcome, however alternative approaches, such as the use of separate SD and DES environments or the use of multi-discipline software, were highlighted as likely to be easier to use.

An interesting area of future work would involve the construction of a hybrid model

that represents the full–scale BT system. The SD part of such a model would be based on the full version of the hydraulics model, introduced in chapter 5, which is considerably more complex than the simplified version. The DES part would be required to represent large numbers of jobs, their queues, and all the engineers. In addition, in the full system there are multiple job types and engineer skill sets. Hence this hybrid model would be considerably more complex than the hybrid model of chapter 2. Such a model would enable BT to better understand the behaviour of the RFT measure in the full system, which could potentially result in more effective system performance.

Chapter 3 investigated core assumptions of BT's SD hydraulics model, using historical time series data. Subjective methods were used to detect 'spikes', the increases in demand, and also in the definitions of the regression variables – which were designed to track system changes around the periods of a spike to determine how the system responds. A regression model was formed at a national–level to investigate overall system behaviour, with 9 area–level regression models investigating more localised behaviour.

Although certain regression assumptions did not hold, the national model revealed that periods of increased demand were found to result in an observable response in the data as management increased engineer numbers accordingly. In terms of how a major organisation controls its backlog of jobs, this result makes sense. Since the approach of chapters 4-6 relies on the structure of the hydraulics model being an accurate representation of system behaviour, this result from the national model was

of particular importance. A number of outliers were also found in the residuals, although robust regression results provided reassurance that outliers were not unduly influencing the results. A limitation of the analysis was that subjective methods were required in the processes of spike detection and variable definitions. However, results demonstrated that the national model results were fairly robust to changes in these selections.

Area-level results suggested that the responses of each geographical area to increases in demand may not be the same. These models violated fewer assumptions than the national model, such as normality of the residuals. However the inference that can be drawn from the area–level models was limited, since these results were based on considerably less data than the national model. This highlighted a more general limitation of our approach. To obtain meaningful results from the regression model, we need a sufficient amount of spike data. This means that we cannot form conclusions separately for individual regions as the number of spikes detected in each is insufficient. Chapter 6 investigates differences in regional behaviour more formally.

In the literature there is some disagreement regarding how SD models should undergo validation tests – and whether data-driven methods are necessary. What is universally accepted is that a number of tests must be passed to build confidence in the model; the more tests that are passed increases confidence in the model. Chapter 3 demonstrated that it is possible to investigate key assumptions of a SD model from historical time series using regression methods. Such an analysis can be used to

strengthen existing structural validation tests of SD models. We also suggest that the approach followed in chapter 3 can be considered as an additional test to the group of direct structure tests. More generally, chapter 3 demonstrates the insights that can be obtained from using data-driven methods to validate the structure of a SD model.

Chapter 4 presented a novel, step by step process, for the formulation of a state-space model from a SD model. This consisted of grouping the SD variables, determining the order in which each should be updated and then performing substitutions to form Markovian difference equations; the addition of state and observational noise terms forms the state-space model. This was used to form a linear Gaussian state-space model from hydraulics model 1.

A modified FIMLOF algorithm was used to estimate the parameters of this state-space model. Simulation results demonstrated success of the method in estimating the parameters to a good degree of accuracy, especially for low amounts of added noise. The signal to noise ratio of the added noise was shown to be an important factor. Previous examples of FIMLOF in the literature present results based on using only a single simulation. Our simulation studies contribute to the literature by using Monte Carlo experiments, which enable a more thorough exposition of the performance of the FIMLOF algorithm. The simulation results can also be viewed as a first step towards estimating the parameters of the full hydraulics model from historical BT time series data in chapter 6.

The simulation studies investigated the effects of using different starting values for the structural parameters $\underline{\theta}$. However, a limitation of the studies was a lack of investigation into the effect of starting values for the variance parameters $\sigma^2_{Q_1}, \sigma^2_{Q_2}, \sigma^2_{R_1}, \sigma^2_{R_2}$. The studies revealed a key limitation of our modified FIMLOF algorithm; its sensitivity to starting values. With such a limitation, multiple sets of parameter estimates, using different starting values, are required to ensure that the modeller can have confidence in the results. In our view, a promising area of future work is to investigate optimisation techniques that are robust to the choice of starting values and hence able to consistently find the global optimum, for this type of problem. Vierhaus et al. (2014) state that this is an area they are currently working on.

The methods outlined in chapter 4 are limited to estimating parameters of linear Gaussian state-space models only. In addition, hydraulics model 1 is a considerably simplified model of BT's system – hence the state-space model formed from this is relatively simple. Chapter 5 explored extensions of the methods to more complex state-space models that are not restricted by the assumption of linearity.

In chapter 5, the novel process for the formulation of a state-space model from a SD model, was extended to nonlinear SD models of varying complexity. Hydraulics model 2 is similar in complexity to model 1. Hydraulics model 3 is considerably more complex than models 1 and 2 – this is effectively the full version used by BT. The state-space model, resulting from hydraulics model 3, is more complex and nonlinear than other models in the literature to which FIMLOF has been applied.

Two simulation studies, studies 2 and 3, presented results of applying modified FIMLOF algorithms to estimate the parameters of these nonlinear state-space models – based on hydraulics models 2 and 3 respectively. Results of study 2 demonstrated success in estimating the parameters, with the amount of added state noise σ_Q shown to be an important factor. Results of study 3, despite showing a loss in accuracy compared to studies 1 and 2, demonstrated some success, especially for low levels of added noise. Results also revealed key differences in the identifiability of each of the 6 structural parameters, resulting in some being estimated more accurately than others. The results of this study demonstrate that even for a complex, industrial–scale model, the algorithm can provide useful results, which would be of importance to an organisation such as BT. Study 3 was designed to recreate the conditions of the real system as far as possible and so was the final step prior to chapter 6 – applying FIMLOF algorithm 3 to historical BT data.

Previous uses of FIMLOF have used the extended Kalman filter for nonlinear systems. However, the literature supports the use of the UKF over the extended Kalman filter. Hence, the UKF was used within the FIMLOF algorithm for studies 2 and 3. This is the first modification of the FIMLOF method to successfully incorporate the UKF for nonlinear systems.

The modified FIMLOF algorithms of chapter 5 suffer from many of the limitations of the algorithm in chapter 4. The algorithms are again limited by their sensitivity to

the starting values used. In addition, simulation studies 2 and 3 do not investigate the effect of using different starting values for the variance parameters. However, as in study 1 of chapter 4, although the algorithms were sensitive to the structural parameter starting values, the choice of starting values did not affect the accuracy of parameter estimates for studies 2 and 3.

A key limitation of FIMLOF algorithms 1-3 is the requirement for a Gaussian noise structure. This assumption is possibly the most limiting when considering applying these techniques to complex industrial SD models. Historical time series associated with such models may be corrupted by non-Gaussian noise, but determining whether or not this is true is not always possible. In our view a promising area of future research would be to investigate approaches that enable this assumption of Gaussian noise to be dropped. Modifying FIMLOF algorithms to include particle filters, rather than the UKF, would seem to be the most promising avenue for this.

After further modifications to FIMLOF algorithm 3, chapter 6 applied this algorithm to estimate parameters of 59 regional state-space models, each based on the structure of hydraulics model 3. Historical BT time series data were used for each region. The objective was to calibrate hydraulics model 3 to enable accurate modelling at a regional level. Parameter estimates were initially poorer than expected, offering no improvement over the defaults. However during investigation a number of issues arose surrounding the historical BT data. Evidence was provided that raised serious questions over the assumption of Gaussian noise. In addition, the potential flaws of

certain time series were exposed. These included the interpolation of weekly series into daily series. Considering that the algorithms performed considerably better in the simulation studies, these data issues were considered as potential reasons for the poor performance with the BT data.

As in the three simulation studies, results exposed a lack of robustness to the starting values used in the algorithm. However, unlike the simulation studies, chapter 6 results demonstrated that the choice of different starting values for the variance parameters affected the accuracy of parameter estimates. This lead to our approach of selecting the 'best' observational variance parameters' $\underline{\sigma}^2_{R_0}$ start values for each geographical region. This was far from ideal – and not the automated parameter estimation procedure that we aimed for. However our results demonstrated that such an approach can be useful and enabled us to obtain parameter estimates that were a significant improvement over the BT defaults. This provides BT with an approach that offers a practical solution to calibrating their model, until advances can be made. We argue that the selection of the best regional starting values is an additional factor that requires calibration. In addition, due to the complexity and scale of hydraulics model 3, in our view, chapter 6 represents the most complex application of a FIMLOF algorithm yet seen.

Two of the key limitations of the FIMLOF approach highlighted earlier in this chapter were again apparent in chapter 6, and perhaps exposed more seriously. The assumption of Gaussian noise was restrictive when attempting to apply the method

to historical time series data from a complex system. The algorithm's sensitivity to starting values was problematic, since unlike the simulation studies, there were no 'true' values to compare parameter estimates against.

The concerns regarding the assumption of Gaussian noise for the BT data, further motivate the earlier suggestion of investigating the use of particle filters within a modified FIMLOF algorithm as an avenue of future work. Another interesting avenue would be to investigate use of methods such as model reference optimisation (MRO), explained in section 4.1.2. This involves the use of an objective function that attempts to minimise the prediction error (e.g. RMSE) of time series directly. This approach does not make any prior assumptions regarding Gaussian errors. MRO would also give us the freedom to incorporate priorities in the different time series - such as giving the backlog higher priority - as was the case for the BT data. MRO also does not require the system to be linear.

In this thesis we have investigated the feasibility and potential value of solution methods, in each of the three key areas outlined at the start of this chapter. Collectively for our research in the three areas, these results demonstrate the value that can be added to SD models using statistical methods. It is worth emphasising that although the focus of this thesis in terms of SD models have been the BT hydraulics models 1-3, the techniques applied in each of these three areas can be applied to a wide range of SD models, provided that appropriate historical data is available.

Bibliography

O. Abduaziz, J. K. Cheng, R. M. Tahar, and R. Varma. A hybrid simulation model for green logistics assessment in automotive industry. *Procedia Engineering*, 100: 960–969, 2015.

A. Alrabghi and A. Tiwari. A novel approach for modelling complex maintenance systems using discrete event simulation. *Reliability Engineering & System Safety*, 154:160–170, 2016.

A. Alvanchi, S. Lee, and S. AbouRizk. Modeling framework and architecture of hybrid system dynamics and discrete event simulation for construction. *Computer-Aided Civil and Infrastructure Engineering*, 26(2):77–91, 2011.

Rockwell Automation. Arena 15. *Available at: https://www.arenasimulation.com*, 2016.

Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation, Tracking and Navigation: Theory, Algorithms and Software*. Wiley, 2001.

Y. Barlas. Multiple tests for validation of system dynamics type of simulation models. *European Journal of Operational Research*, 42(1):59–87, sep 1989a.

Y. Barlas. Tests of model behavior that can detect structural flaws: Demonstration with simulation experiments. In *Computer-Based Management of Complex Systems*, pages 246–254. Springer Berlin Heidelberg, 1989b.

Y. Barlas. Model validation in system dynamics. In *Proceedings of the 1994 International System Dynamics Conference*, pages 1–10. Sterling, Scotland, 1994.

Y. Barlas. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12(3):183–210, 1996.

M. L. Baughman, P. L. Joskow, and F. S. Zerhoot. Interfuel substitution in the consumption of energy in the united states. Technical report, MIT Energy Lab, 1974.

T. Bayes. A letter from the late reverend mr. thomas bayes, frs to john canton, ma and frs. *Philosophical Transactions (1683-1775)*, 53:269–271, 1763.

A. Borshchev. Multi-method modelling: Anylogic. In S. Brailsford, L. Churilov, and B. Dangerfield, editors, *Discrete-event simulation and system dynamics for management decision making*, chapter 12, pages 248–279. John Wiley & Sons, 2014.

G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

S. Brailsford, L. Churilov, and B. Dangerfield. *Discrete-event simulation and system dynamics for management decision making*. John Wiley & Sons, 2014.

S. C. Brailsford and N. A. Hilton. A comparison of discrete event simulation and system dynamics for modelling health care systems, 2001.

S. C. Brailsford, V. A. Lattimer, P. Tarnaras, and J. C. Turnbull. Emergency and on-demand health care: modelling a large complex system. *Journal of the Operational Research Society*, 55(1):34–42, 2004.

S. C. Brailsford, S. M. Desai, and J. Viana. Towards the holy grail: combining system dynamics and discrete-event simulation in healthcare. In *Simulation Conference (WSC), Proceedings of the 2010 Winter*, pages 2293–2303. IEEE, 2010.

R. J. Brooks and S. L. Robinson. *Simulation and Inventory Control (Texts in Operational Research)*. Palgrave Macmillan, 2001.

K. Chahal and T. Eldabi. Applicability of hybrid simulation to different modes of governance in uk healthcare. In *Simulation Conference, 2008. WSC 2008. Winter*, pages 1469–1477. IEEE, 2008.

J. Chen and A. K. Gupta. *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Springer Science & Business Media, 2011.

Y. T. Chen, Y. M. Tu, and B. Jeng. A Machine Learning Approach to Policy Optimization in System Dynamics Models. *Systems Research and Behavioral Science*, 28:369, 2011.

L. Cheng and M. A. Duran. Logistics for world-wide crude oil transportation using

discrete event simulation and optimal control. *Computers & chemical engineering*, 28(6):897–911, 2004.

The AnyLogic Company. Anylogic 7 professional. *Available at: http://www.anylogic.com*, 2016.

Simul8 Corporation. Simul8 2015. *Available at: http://www.simul8.com*, 2015.

D. R. Cox and D. V. Hinkley. *Theoretical statistics*. CRC Press, 1979.

R. G. Coyle. *System dynamics modelling: a practical approach*, volume 1. CRC Press, 1996.

R. G. Coyle. Simulation by repeated optimisation. *Journal of the Operational Research Society*, pages 429–438, 1999.

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

B. Dangerfield and C. Roberts. An Overview of Strategy and Tactics in System Dynamics Optimization. *Journal of the Operational Research Society*, 47:405–423, 1996.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

R. B. Detty and J. C. Yingling. Quantifying benefits of conversion to lean manufacturing with discrete event simulation: A case study. *International Journal of Production Research*, 38(2):429–445, 2000a.

R. B. Detty and J. C. Yingling. Quantifying benefits of conversion to lean manufacturing with discrete event simulation: a case study. *International Journal of Production Research*, 38(2):429–445, 2000b.

P. J. Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press, 2013.

J. Durbin and S. J. Koopman. *Time series analysis by state space methods*, volume 24. Oxford Statistical Series, 2001.

R. L. Eberlein. Full feedback parameter estimation. In *Proceedings of the Int Systems Dynamics Conference, Sevilla, Spain. System Dynamics Society, Albany*, pages 69–83, 1986.

R. L. Eberlein and Q. Wang. Statistical estimation and system dynamics models. In *Proceedings of the Int Systems Dynamics Conference, Keystone, USA. System Dynamics Society, Albany*, pages 206–222, 1985.

D. Eddelbuettel. *Seamless R and C++ integration with Rcpp*. Springer, 2013.

A.I.E. Empresarios Agrupados. Ecosimpro 5.6.0. *Available at: http://www.ecosimpro.com*, 2016.

D. M. Ferrin, M. J. Miller, S. Wininger, and M. S. Neuendorf. Analyzing incentives and scheduling in a major metropolitan hospital operating room through simulation. In *Simulation Conference, 2004. Proceedings of the 2004 Winter*, volume 2, pages 1975–1980. IEEE, 2004.

D. M. Ferrin, M. J. Miller, and D. L. McBroom. Maximizing hospital finanacial impact and emergency department throughput with simulation. In *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*, pages 1566–1573. IEEE Press, 2007.

A. Fletcher, D. Halsall, S. Huxham, and D. Worthington. The dh accident and emergency department model: A national generic model used locally. *Journal of the Operational Research Society*, 58(12):1554–1562, 2007.

J. W. Forrester. *The impact of feedback control concepts on the management sciences.* Foundation for Instrumentation Education and Research, 1960.

J. W. Forrester. *Industrial Dynamics.* MIT Press, Cambridge, MA., 1961.

J. W. Forrester. *Market growth as influenced by capital investment.* Industrial Management Review, 1968a.

J. W. Forrester. Industrial dynamics-after the first decade. *Management Science*, 14 (7):398–415, 1968b.

J. W. Forrester. Rejoinder. *Journal of the American Statistical Association*, 75(371): 572–574, 1980a.

J. W. Forrester. Information sources for modeling the national economy. *Journal of the American Statistical Association*, 75:555–66, 1980b.

J. W. Forrester and P. M. Senge. Tests for building confidence in system dynamics models. *System dynamics, TIMS studies in management sciences*, 14:209–228, 1980.

G. E. Forsythe, C. B. Moler, and M. A. Malcolm. *Computer methods for mathematical computations*. Prentice-Hall, 1977.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

P. D. Gilbert. *Brief User's Guide: Dynamic Systems Estimation*, 2013. URL `http://cran.r-project.org/web/packages/dse/vignettes/dse-guide.pdf`.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

A. K. Graham. On Positioning System Dynamics as an Applied Science of Strategy Or: SD is Scientific. We Havent Said So Explicitly, and We Should. In *20th International Conference of the System Dynamics Society*, 2002.

M. Guo, M. Wagner, and C. West. Outpatient clinic scheduling-a simulation approach. In *Simulation Conference, 2004. Proceedings of the 2004 Winter*, volume 2, pages 1981–1987. IEEE, 2004.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics*. Wiley, New York, 1986.

J. Harrison and M. West. *Bayesian Forecasting & Dynamic Models*. Springer, 1999.

P. Harrison and N. Patel. Performance modelling of communication networks and computer architectures. addisonwesley. *Reading, Massachusetts*, 1992.

R. F. Harrod. An essay in dynamic theory. *The Economic Journal*, pages 14–33, 1939.

S. S. Haykin et al. *Kalman filtering and neural networks*. Wiley Online Library, 2001.

S. K. Heath, S. C. Brailsford, A. Buss, and C. M. Macal. Cross-paradigm simulation modeling: challenges and successes. In *Simulation Conference (WSC), Proceedings of the 2011 Winter*, pages 2783–2797. IEEE, 2011.

F. S. Hillier and S. Y. Oliver. *Queueing tables and graphs*, volume 3. North-Holland, 1981.

K. Hoad, S. Robinson, and R. Davies. Autosimoa: A framework for automated analysis of simulation output. *Journal of Simulation*, 5(1):9–24, 2011.

C. E. Holt. Forecasting trends and seasonal by exponentially weighted averages. *ONR Memorandum*, 52, 1957.

C. E. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, jan 2004.

P. J. Huber and E. M. Ronchetti. *Robust statistics, Series in probability and mathematical statistics*. John Wiley, New York, 1981.

R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer, 2008.

Ventana Systems Inc. Vensim ple. *Available at:¡ http://www. vensim. com*, 2010.

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

K. Jensen, M. Lyons, and N. Buckhurst. *Transforming Field and Service Operations: Methodologies for Successful Technology-Driven Business Transformation*, chapter System Dynamics Models of Field Force Operations, pages 47–69. Springer Berlin Heidelberg, 2013.

S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new method for the non-linear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, pages 477–482, 2000.

S. J. Julier and J. K. Uhlmann. New extension of the kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering*, 82:35–45, 1960.

W. D. Kelton and A. M. Law. *Simulation modeling and analysis*. McGraw Hill Boston, 2000.

A. Khintchine. Mathemaiische theorie der stazionaren reihe. *Matematicheskii Sbornik*, 39(4):73–84, 1932.

K. Kotiadis and J. Mingers. Combining psms with hard or methods: the philosophical and practical challenges. *Journal of the Operational Research Society*, 57(7):856–867, 2006.

A. A. Jr. Legasto and J. Macariello. System dynamics: A critical review. *TIMS Studies in Management Science*, (14):23–43, 1980.

J. D. C. Little. A proof for the queuing formula: L= λ w. *Operations research*, 9(3): 383–387, 1961.

J. D. C. Little. Or forum-little's law as viewed on its 50th anniversary. *Operations Research*, 59(3):536–549, 2011.

J. M. Lyneis and A. L. Pugh. Automated vs.hand calibration of system dynamics models: An experiment with a simple project model. In *Proceedings of the 1996 International System Dynamics Conference*, volume 42, 1996.

I. L. MacDonald. Numerical maximisation of likelihood: A neglected alternative to em? *International Statistical Review*, 82(2):296–308, 2014.

R. Mazaeda, A. Merino, C. de Prada, and F. Acebes. Hybrid modelling of batch centrifuges as part of a generic object oriented beet sugar mill library. *Simulation Modelling Practice and Theory*, 22:123–145, 2012.

D. H. Meadows. The unavoidable a priori. *Elements of the system dynamics method*, pages 23–57, 1980.

R. K. Mehra and J. S. Tyler. Case studies in aircraft parameter identification. *Identification and system parameter estimation*, pages 117–144, 1973.

R. L. Moore and F. C. Schweppe. Model identification for adaptive control of nuclear power plants. *Automatica*, 9(3):309–318, 1973.

J. Morecroft. *Strategic modelling and business dynamics: A feedback systems approach*. John Wiley & Sons, 2007.

J. D. W. Morecroft, S. Robinson, et al. Explaining puzzling dynamics: comparing the use of system dynamics and discrete-event simulation. In *Proceedings of the 23rd International Conference of the System Dynamics Society*, pages 17–21, 2005.

J. S. Morgan, S. Howick, and V. Belton. A toolkit of designs for mixing discrete event simulation and system dynamics. *European Journal of Operational Research*, 257 (3):907–918, 2017.

J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

W. D. Nordhaus. World dynamics: measurement without data. *Economic Journal*, 83:1156–1183, 1973.

R. Oliva. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151:552–568, 2003.

Y. Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.

D. W. Peterson. *Hypothesis, estimation, and validation of dynamic social models: energy demand modeling.* PhD thesis, Massachusetts Institute of Technology, 1975.

D. W. Peterson. Statistical tools for system dynamics. In *Proceedings of the 1976 International Conference on System Dynamics*, pages 224–241, 1976.

D. W. Peterson. A many-world theory of system dynamics. Address to the 2003 System Dynamics Conference. New York City, New York, July 24th., 2003.

G. Petris. An r package for dynamic linear models. *Journal of Statistical Software*, 36(12), 2010.

G. Petris, S. Petrone, and P. Campagnoli. *Dynamic linear models with R*. Springer Science & Business Media, 2009.

M. Pidd. *Computer simulation in management science*. 5th ed. Chichester: Wiley., 2004.

M. Pidd. Mixing other methods with simulation is no big deal. In *Proceedings of the Winter Simulation Conference*, page 67. Winter Simulation Conference, 2012.

M. Pidd. The ways forward: A personal view of system dynamics and discrete-event simulation. In S. Brailsford, L. Churilov, and B. Dangerfield, editors, *Discrete-event simulation and system dynamics for management decision making*, chapter 15, pages 318–336. John Wiley & Sons, 2014.

F. Pollaczek. Über eine aufgabe der wahrscheinlichkeitstheorie. i. *Mathematische Zeitschrift*, 32(1):64–100, 1930.

M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162, 1964.

M. Pruckner and R. German. A hybrid simulation model for large-scaled electricity generation systems. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, pages 1881–1892. IEEE Press, 2013.

H. Qudrat-Ullah. On the validation of system dynamics type simulation models. *Telecommunication Systems*, 51(2-3):159–166, mar 2011.

H. Qudrat-Ullah and B. S. Seong. How to do structural validity of a system dynamics type simulation model: The case of an energy policy model. *Energy Policy*, 38(5): 2216–2224, may 2010.

M. J. Radzicki, M. Kennedy, G. W. Winch, R. S. Langer, J. I. Rowe, and J. M. Yanni. Expectation Formation and Parameter Estimation in Nonergodic Systems : The System Dynamics Approach to Post Keynesian-Institutional Economics. In *Proceedings of the 22nd International Conference of the System Dynamics Society*, 2004.

G. P. Richardson. *Statistical Estimation of Parameters in a Predator-prey Model: An Exploration Using Synthetic Data.* System Dynamics Group, Sloan School of Management, Massachusetts Institute of Technology, 1981.

B. D. Ripley. *Spatial statistics*, volume 575. John Wiley & Sons, 2005.

S. Robinson. *Simulation: the practice of model development and use*. Chichester: Wiley., 2004.

A. V. Ryzhenkov. A Historical Fit of a Model of the US Long Wave. In *The 20th International Conference of The System Dynamics Society*, pages 1–30, 2002.

F. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11, 1965.

F. C. Schweppe. *Uncertain dynamic systems*. Prentice Hall, 1973.

P. M. Senge. An experimental evaluation of generalized least squares estimation. *MIT Sloan School of Management System Dynamics Group Working Paper D-1944-6*, 1974.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

V. I. Shiryaev, E. V. Shiryaev, I. Y. Golovin, V. Smolin, P. I. Davidsen, E. Mollona, V. G. Diker, R. S. Langer, and J. I. Rowe. Adaptation and Optimal Control of a Firm and its State and Parameters Estimation at the Change of a Market Situation. In *Proceedings of the 20th International Conference of the System Dynamics Society*, 2002.

R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and fore-

casting using the em algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 1982.

R. H. Shumway and D. S. Stoffer. Time series regression and exploratory data analysis. *Time Series Analysis and Its Applications: With R Examples*, pages 48–83, 2006.

J. D. Sterman. Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica*, 10(2):51–66, 1984.

J. D. Sterman. *Business dynamics: systems thinking and modeling for a complex world*, volume 19. Irwin/McGraw-Hill Boston, 2000.

ISEE systems. ithink 10. *Available at: https://www.iseesystems.com*, 2012.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

F. Tusell. Kalman filtering in r. *Journal of Statistical Software*, 39(2), 2011.

J. K. Uhlmann. Simultaneous map building and localization for real time applications. *transfer thesis, Univ. Oxford, Oxford, UK*, 1994.

J. Viana, S. C. Brailsford, V. Harindra, and P. R. Harper. Combining discrete-event simulation and system dynamics in a healthcare setting: A composite model for chlamydia infection. *European Journal of Operational Research*, 237(1):196–206, 2014.

I. Vierhaus, A. Fügenschuh, R. L. Gottwald, and S. N. Grösser. *Modern Nonlinear Optimization Techniques for an Optimal Control of System Dynamics Models*. Helmut-Schmidt-Univ., Univ. der Bundeswehr Hamburg, 2014.

A. Waller. Witness simulation software. In *Proceedings of the Winter Simulation Conference*, WSC '12, pages 436:1–436:12. Winter Simulation Conference, 2012.

E. Wan, R. Van Der Merwe, et al. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. IEEE, 2000.

M. W. Watson and R. F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

A. Wijewickrama and S. Takakuwa. Simulation analysis of appointment scheduling in an outpatient department of internal medicine. In *Simulation Conference, 2005 Proceedings of the Winter*, pages 10–pp. IEEE, 2005.

J. Windisch, K. Väätäinen, P. Anttila, M. Nivala, J. Laitila, A. Asikainen, and L. Sikanen. Discrete-event simulation of an information-based raw material allocation process for increasing the efficiency of an energy wood supply chain. *Applied energy*, 149:315–325, 2015.

P. R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342, 1960.

C. Wong, G. Geiger, Y. D. Derman, C. R. Busby, and M. W. Carter. Redesigning the medication ordering, dispensing, and administration process in an acute care academic health science centre. In *Winter Simulation Conference*, volume 2, pages 1894–1902. Citeseer, 2003.

D. C. Wynn, D. F. Wyatt, S. M. T. Nair, and P. J. Clarkson. An introduction to the cambridge advanced modeller. 2010.

D. C. Wynn, S. Cassidy, and P. J. Clarkson. Design of robust service operations using cybernetic principles and simulation. 2012.

B. P. Zeigler. Embedding dev&dess in devs: Characteristic behaviors of hybrid models. *Simulation Series*, 38(1):125, 2006.

B. P. Zeigler, H. Praehofer, and T. G. Kim. *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic press, 2000.

J. Zulkepli, T. Eldabi, and N. Mustafee. Hybrid simulation for modelling large systems: an example of integrated care model. In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, pages 1–12. IEEE, 2012.