

Emergency Departments: Predicting the Unpredictable

Michael John Ucci

October 26, 2023

*Thesis submitted for the degree of
Master of Philosophy
in
Applied Mathematics & Statistics
at The University of Adelaide
Faculty of Science, Engineering and Technology
School of Computer and Mathematical Sciences*



THE UNIVERSITY
of ADELAIDE

Contents

Acknowledgements	vii
Signed Statement	viii
Abstract	ix
1 Introduction	1
2 Emergency department overcrowding	3
2.1 Causes and effects	3
2.2 Transformative interventions	4
3 Modelling approaches	6
3.1 Queueing theoretic models	6
3.2 Simulation	7
3.3 Statistical models	9
3.4 Generalisable ED models	10
3.4.1 Project goals in context	11
4 The data set	12
4.1 The data set	12
4.1.1 Data aggregation	13
4.1.2 Variable descriptions	13
4.2 Missing and erroneous data	17
4.2.1 Inclusion criteria	21
4.3 Exploratory insights	24
4.3.1 Distributional observations	24
4.3.2 Pattern based analyses	30
5 Temporal modelling methods	34
5.1 Problem Specification	34
5.2 Time series modelling theory	35

5.2.1	Time series processes	35
5.2.2	Time series model estimation	36
5.3	Model assessment protocols	38
6	Predictive modelling	44
6.1	Baseline establishment	44
6.2	Model development	48
6.2.1	Identifying deterministic trends	48
6.2.2	Modelling stochastic behaviour	52
6.3	Results	55
6.4	Prediction platform	58
7	Deep learning for time series	62
7.1	Autoregressive neural networks	62
7.2	Modelling building	64
7.3	Implementation: results and performance	65
8	Conclusion	67
A	Detailed results	70
B	ARIMA models	74
	Bibliography	81

List of Tables

4.1	Data scraping output	16
4.2	Hospital information	16
4.3	Summary statistics : Capacity Index	24
4.4	Summary statistics : Expected Arrivals	30
6.1	Prediction benchmarks	46
6.2	Linear model comparison	49
6.3	Linear model results	50
6.4	Initial ARIMA results	52
6.5	Time series results : LMH	56
6.6	Time series results : All hospitals	56
7.1	Deep learning results: LMH	66
A.1	Time series results : LMH	70
A.2	Time series results : RAH	71
A.3	Time series results : TQEH	71
A.4	Time series results : NHS	72
A.5	Time series results : MH	72
A.6	Time series results : WCHP	73

List of Figures

4.1	Emergency department dashboard example	14
4.2	Data collection model	15
4.3	Missing data model: Dashboard failure	18
4.4	Missing data model: Web scrape failure	18
4.5	Missing data model: Partially missing data	19
4.6	Missing data model : Outdated data	20
4.7	Missing data model: Contradicting records	20
4.8	Missing data model: Distinct observations	21
4.9	Missing data plot	23
4.10	Capacity Index density plot	25
4.11	Time continuously over capacity	27
4.12	Average Waiting Time box plot	28
4.13	Expected Arrivals box plot	29
4.14	Daily profile: Capacity Index	31
4.15	Weekly profile: Capacity Index	32
5.1	Standard cross validation schematic	40
5.2	Rolling origin cross validation schematic	41
5.3	Rolling cross validation parameter estimate errors	42
5.4	Adapted rolling origin cross validation schematic	43
6.1	Prediction benchmark	47
6.2	Residuals analysis	51
6.3	Training set size sensitivity	54
6.4	Time series results	57
6.5	Prototype Application: Visualisation	60
6.6	Prototype Application: Prediction	61
7.1	ANN diagram	63
7.2	Time series ANN diagram	64
B.1	Time series example: AR(2)	77

B.2	Time series example: MA(2)	78
B.3	Time series example: ARMA(2,2)	79

Acknowledgements

I am overwhelmed with gratitude for all those who have helped, supported and educated me throughout this degree and all of my previous education. Among the definitions and theorems I've learnt, are lessons of the true value to be found in people, experiences and learning. Above all else, this journey has taught me the power and beauty of both curiosity and passion.

To my supervisors, it goes without saying that this work would not exist without your support and guidance. I cannot thank each of you enough for your time, care and commitment. Throughout this experience your interest and passion has naturally fed my own.

To my peers, colleagues and mates; you've taught me the power of friendship and the beauty of all the in-between moments. My fondest memories of this journey are all thanks to you, and the love you have shown.

Finally to my family; Mum, Dad and Sara. Your ever present support, love and care is inspiring to say the least. I cannot thank you enough for all that you have done for me.

Ucci

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed:

Date: October 25, 2023
.....

Abstract

On a global scale, hospital Emergency Departments (EDs) are experiencing increasing rates of overcrowding. This work details the development of a framework intended for practical use by hospital practitioners to forecast ED demand.

The increasing amount of overcrowded EDs, due to factors including increases in population sizes as well as complexity and severity of health issues, have been shown to correlate with increased patient mortality as well as broader effects on health care systems, not to mention financial losses for providers. Mathematical methods and analyses have proven to be powerful tools against ED overcrowding, the most popular including simulation models, queueing theory, and statistical analyses. That said, limited work has investigated the use of generalised models for the purposes of multi-hour demand prediction.

Here a hierarchical framework is developed for the purpose of forecasting short-term ED demand and validating its use on six South Australian EDs. Through the collection of public data, and subsequent application of the framework, a stepwise predictive model was shown to consistently outperform a forecasting baseline in all six EDs. The predictive model is then compared against readily implemented deep learning approaches which produce comparable results, but at great computational cost. A methodology for time series cross-validation is also developed and applied for the purposes of model selection and comparison.

Alongside the collection and aggregation of South Australian ED data for future research, the results of this work provide an improved prediction baseline, and a proof of concept for generalised short-term statistical modelling of emergency departments.

Chapter 1

Introduction

Emergency Department (ED) overcrowding has been a topic of focus in areas of public health, operations research and statistics for a number of decades. Here, a novel predictive modelling framework is developed that utilises openly accessible, generic data for the purposes of supporting tactical decisions such as staff allocation, patient placement or resource usage. This work develops a widely applicable tool which produces short-term (one to five hours) forecasts of emergency department demand through a proxy variable, known as the *Capacity Index* (a ratio of patients to capacity). In pursuit of this goal, three key contributions are made including:

- the aggregation and collation of an openly available data set,
- the development of a cross-validation process for time series data, and
- the development of a predictive framework and model which improves upon established predictive baselines.

This thesis is initiated with an investigation of the problem space, reviewing the issue of overcrowding and previously applied modelling approaches. Following this, a discussion of the data set being analysed is presented, detailing its collection as well as descriptive analyses. Using the results from the data analysis, the model development process is detailed, concluding with the presentation of the predictive modelling results.

A review of research into the application of quantitative methods to the overcrowding problem categorises methods into three broad approaches being queueing theoretic, simulation based, and statistical models. It was also recognised that studies regularly consider strategic-scale time horizons and analyses are purpose built for specific hospitals, inhibiting the broader application of their results. The intent of this work therefore was to build a model for more operational insights, which are more transportable across different hospitals and health networks. More concretely, this intent can be restated as developing a generalisable modelling framework, providing predictions up to five hours

into the future, for the purpose of supporting decisions of workforce and resource usage along with patient management.

To form a data set regarding seven public hospital EDs in South Australia, the *SA Health Emergency Department Dashboard* was accessed at regular intervals over a number of months in 2021. Data cleaning and descriptive analysis were performed on the data providing insights into the similarities and differences of behaviours observed across the multiple hospitals.

Informed by this data analysis, discussions of the predictive modelling approaches lead to the application of a hierarchical modelling framework consisting of a systematic temporal dependency step, as well as a step with more specific allowances for the autodependencies found in the data. The latter of these steps involved the application of time-series modelling approaches, for which a novel model assessment technique was developed.

The application of the hierarchical framework using the model assessment technique generated a set of models which outperformed prediction baselines across all hospitals. For predictions five hours into the future (a five-hour prediction horizon), the minimum improvement of Mean RMSE was found to be approximately 17% (*Royal Adelaide* and *Queen Elizabeth* hospitals), with the most significant being a 37% improvement (*Women's and Children's Hospital*). Practically, the implication of these results is an improved prediction capability, enabling tactical decision support for purposes such as resourcing or staff allocation.

Finally, for the purposes of validation and exploration of what may be considered more contemporary models, the same framework was applied using Neural Network models in place of ARIMA models. Neural network based predictions provided a 1% reduction in mean prediction RMSE at a five-hour horizon, however, this approach was found to be computationally prohibitive in the current state, taking on the order of two weeks for models to be fit, which compared with minutes for the ARIMA models.

This work entails the collection, and cleaning of a data set regarding multiple hospitals which has been used to aid the development and assessment of a hierarchical modelling framework to produce predictions of emergency department demand up to five hours into the future.

Before moving into the predictive model development and analysis process, the following chapter will first provide an understanding of ED overcrowding

Chapter 2

Emergency department overcrowding

In this chapter, a discussion of Emergency Department (ED) overcrowding is approached through two lenses being the causes and effects of overcrowding on health care, as well as some transformative interventions which have been published. Not only does this contextualise this work in the broader setting, but it allows for an understanding of the driving factors which motivate the development of a data-driven decision support tool.

2.1 Causes and effects

Data provided by the *Australian Institute for Health and Welfare* (2019-2020) stated that over 30% of individuals presenting to Australian Emergency Departments remain there for longer than the previously mandated four-hour target; for South Australia, that number is 39% [1].

The causes of overcrowding in emergency departments and the consequential effects have been thoroughly researched and documented, demonstrating an undeniable demand for improved systems and policies [2–7]. Since the 1960’s with authors such as Hyman [8], research has been detailing and quantifying overcrowding in healthcare systems, emergency departments in particular. Hyman cited causes of overcrowding to be increases in population size and mobility, as well as a trend of patients preferring to attend a hospital emergency department rather than the more expensive alternative of inpatient care systems. More recent work such as Derlet and Richards [2], Trzeciak and Rivers [3], claim the key drivers of overcrowding are increased complexity and severity of conditions as a result of changing behaviours and ageing populations, as well as inefficiencies within emergency departments and health care systems more broadly. Consultation of the literature allows one to deduce one thing with certainty; that the causes of ED overcrowding are nothing if not complicated and interrelated.

It is to some degree self-evident that an overcrowded ED has a negative impact on emergency care and the health system more broadly. With this said, there is a vast

amount of literature detailing the exact modes and degrees to which the adverse outcomes are manifested. A review paper by Hoot and Aronsky [4] classified the effects of overcrowding in emergency departments into four broad categories; adverse outcomes, reduced quality, impaired access, and provider losses. Regarding the reduced quality of care category, a series of papers have been published establishing significant relationships between the overcrowding of emergency departments and increased patient mortality [5–7]. This establishes a clear motivation for increasing the efficiency and efficacy of emergency department care.

2.2 Transformative interventions

While it is clear that the current state of ED overcrowding is sub-standard, there have been significant interventions applied locally and globally which have shown to be beneficial for health care outcomes. Throughout a patient’s experience of an emergency department, there are a number of potential bottlenecks which contribute to overcrowding. Some of these potential issues across the patient’s journey have been studied and addressed as detailed below.

Pre-arrival

The first of these issues occurs prior to arriving at the ED. Early work in 1979 indicated that improved ambulance response times often reduced mortality rate and complexity of care required [9]. Since then, a large amount of work has been undertaken to advise on ambulance locations and deployment strategies as well as investigating ambulance diversion; a process by which ambulances may avoid busier EDs as a destination, in favour of an ED which may be further away but more capable of providing the necessary care. As detailed in the review by Goldberg [10], tools from operations research have been used to effectively model dispatch systems. In particular, Peleg and Pliskin [11] simulated the use of a decision support tool for ambulance location in an Israeli district, exhibiting that the proportions of ambulance response within a predetermined eight-minute time threshold could be increased from 34% to 94%. An alternate decision support tool was validated in Kentucky by Repede and Bernardo [12], achieving a 36% reduction in response time, without increasing the amount of available resources.

In care

Assessment and improvement of the processes within EDs have been the focus of the large majority of the efforts regarding overcrowding. Upon arrival at an ED, one of the most widely used tools is applied to all patients in a triage assessment. Initially developed during World War I, the triage system is designed to prioritise patients as they present to an ED. As detailed by Saghaian et al. [13], the triage process has been extensively studied

in order to better understand the finer details of the triage process, including the most effective classification system, required information for effective triaging and the optimal qualifications of triaging staff.

The effective use of resources, namely staffing and bed numbers has also been the object of considerable work. A common approach to many resource focused problems is to get an estimate of the required ED capacity to achieve a predetermined goal. However, the results of such efforts tend to be undesirable not only due to financial and economic forces but as matters of efficacy as well. It was noted by Richardson [14] that objectives should be focused on effective allocation more so than an increased amount of resources. It was shown in fact by Han et al. [15] that in some cases expanding an ED provides no tangible improvement on lengths of stay or ambulance diversions.

The scheduling of nurse and physician staff has been the focus of a significant amount of work, in particular from Green and collaborators [16–20]. A number of studies have combined the use of stochastic models and simulation tools in order to build proof of concept staffing policies designed to find optimal staffing schedules for a given set of circumstances [16, 17, 20–22]. Sinreich and Jabali [21] applied optimisation techniques to characterise the utility of particular resources in an ED, allowing for reduced lengths of stay even with a reduced number of staff. Later work by Sinreich et al. [22] involved the development of algorithmic scheduling which significantly reduced patient waiting time and lengths of stay.

Discharge

The final step in a patient’s ED experience is of course their departure, either through discharge or admission as an inpatient into the hospital. The latter of these options is the final stage at which ED overcrowding issues can originate. It goes without saying, that a patient can only be admitted to a hospital if there is a bed for them. Hence, if particular sections of the hospital are at or near capacity, it is common that a build up of patients will start to occur in the ED. This phenomenon is commonly referred to in the literature as *bed blocking* or *boarding*. A significant amount of work has been published evidencing the fact that the effect of boarding is material to ED overcrowding, with some papers claiming it is one of the chief causes of overcrowding [23, 24]. A limited amount of work has considered developing and testing interventions to reduce the impact of boarding, including the work of Shi et al. [25] which considered alternate inpatient discharge schedules, and that of Kverndokk and Melberg [26] which considered a game theoretic approach introducing financial disincentives for hospitals to allow boarding.

As can be seen from this contextual summary of the literature, an emphasised focus on effective planning and decision making tends to provide positive outcomes. It should be noted that longer term strategic planning is the focus of the majority of these interventions. This is in contrast to the focus of this thesis, which looks to develop decision support tools to enable shorter term tactical planning and decision making.

Chapter 3

Modelling approaches

The previous chapter detailed research regarding the causes and effects of Emergency Department (ED) overcrowding. In this chapter, the focus is shifted to the research into modelling approaches which have previously been investigated. Key results from the literature are discussed, highlighting particular research gaps. In particular, it is concluded that the gap which is to be addressed by this work is the development of a generalisable decision support tool, which generates predictions of performance on a shorter (multi-hour) time scale.

Given the prevalence of ED overcrowding on an international scale as well as the magnitude of its consequences, the amount of research conducted to date on the issue is vast. The structure of EDs and the nature of the overcrowding problem naturally lends itself to formulation in mathematical contexts, allowing for the issues around overcrowding to be analysed in a quantifiable fashion.

Broadly speaking, most if not all approaches to health care modelling can be classified into one of three broad categories; *queueing theoretic models*, *simulation based models* and *statistical models*. The breadth and depth of research in this area makes an entirely comprehensive review of the literature out of the scope of this work, with this said, the following sections detail key work from the three areas.

3.1 Queueing theoretic models

Queueing theory is an area of mathematics involving the use of stochastic processes to construct systems resembling waiting lines or queues. Careful and effective construction of these systems can allow for insights into the emergent behaviours in the relevant applications. For example, queueing theory has been used extensively in telecommunications networks, traffic flow problems and of course to study physical queues in industries ranging from retail to manufacturing. As ED overcrowding is by definition an issue centred on queues, queueing theory has naturally been a common tool for researchers to approach

the problem with.

Largely, work in this area considers $M/M/s$, $M/G/s$, and $G/G/s$ queues (in Kendall's Notation¹), or variations thereof. These models have been applied in a variety of ways, allowing for different characteristic behaviours to be investigated as discussed below.

In 2006, Green et al. [17] provided examples of the ways in which these standard queueing models can be used to estimate probabilities of delay in a theoretical obstetrics unit, as well as aid in determining required staffing levels in a New York ED [17]. Singer and Donoso [28] used an $M/G/s$ model to capture the behaviours of ambulance services in order to aid the design of effective key performance indicators. Financial impacts of patient renegeing (individuals leaving the queue prior to receiving service) were investigated using an $M/M/1/K$ queue, where the additional K is the maximum length of the queue [29].

It is noted by Green et al. [17] that while queueing models are useful, the necessary data required to build effective models can be difficult to access. It was also noted that standard queueing models are unable to capture the effects of finer details, even when models are built for specific departments. These details tend to involve the impact of temporal dependencies, patient acuity and bed blocking among others.

A number of papers have considered more sophisticated queueing constructions in order to represent healthcare systems more realistically. A crucial assumption of standard queueing models is that the arrival and service distributions are stationary and time-independent. Armony and Mandelbaum [30] make the observation that this assumption is not satisfied in reality, and instead test alternate models. Both time-dependent, and state-dependent queueing models are considered with results indicating that the state-dependent model most accurately aligns with observed data [30]. Other previously noted issues such as patient acuity and time-dependent arrivals were considered in a queueing network model developed by Cochran and Roche [31]. McNamara et al. [32] applied network analysis in a novel approach which used openly accessible data to develop a framework for analysis of more holistic system-wide models, allowing investigation into more strategic and planning-oriented questions.

3.2 Simulation

The capability of simulation models to replicate complex dynamical systems means they are suited to the task of detailed modelling of patient flows in emergency departments. An emergent feature of the literature is that simulation studies are regularly carried out in

¹*Kendall's notation* is the standard language for describing a queue with mathematical formalism. An $A/S/s$ queue is one in which the random processes reflecting the arrival and service distributions are of type A and S respectively. In the cases described above; M stands for a Markovian (memoryless) process (Poisson distributed arrivals and exponentially distributed service times), and G stands for a general distribution [27]. Lastly, the s is the number of servers in the queue being considered.

direct partnership with an individual hospital, to analyse and improve patient flow in the hospital's ED. As a result of this, researchers often have access to unit record data from particular hospitals and are given sufficient auxiliary information (processes, plans etc.). This allows simulation models of the specific EDs to be built, to analyse performance under different strategies and policies.

Simulation studies can be classified into three sets of techniques including Discrete Event Simulation (DES) models, System Dynamics (SD) models, and Agent-Based simulation (ABS) models. The depth of research applying simulation techniques to EDs is vast, to the degree that confining a search to exclusively review articles results in a comprehensive set of papers [13, 33–41].

The majority of simulation studies use DES models, which replicate patients progressing through a queueing system (an ED for example) according to a set of rules and procedures. The objective of these papers is to run *in silico* experiments to test how different operational policies affect various key performance indicators. Commonly, the policies being considered regard resources, staffing distributions and ambulance diversion, with the indicators including patient waiting times, lengths of stay, and the number of patients leaving without being seen.

Ruohonen et al. [42] use simulations of a Finnish ED to exhibit how altered triage assessment systems could be used to reduce average length of stay (LOS) by 26%. Simulations of altered triage systems were also shown to provide significant improvements in LOS by Wang et al. [43], along with the introduction of a simulated 'floating nurse' (a surplus nurse to share workload), and a mandated physician consultation within 30 minutes of arrival. Hoot et al. [44] took a novel approach, applying a DES model in order to forecast ED crowding.

A valuable contribution was made by Lane et al. [45] in using an SD simulation of a hospital (as opposed to just an ED). Results showed that while some interventions may be considered beneficial in the context of an ED, the same intervention can result in adverse effects in other departments of the hospital.

The key limitation of DES, and SD models is the limited ability to model the impact of particular entities (agents) involved in the treatment process. ABS models however are built with the intention of doing exactly this [46–49]. Jones and Evans [47] show the utility of ABS models by exhibiting the similarity of simulated and empirical data. Laskowski and Mukhi [48] develop an ABS model with the goal of modelling the impact of staff scheduling on queue lengths. It is noted however that the utility of their model is limited without associating it with more detailed empirical data. Wang et al. [43] construct an ABS model to replicate an ED in Virginia, considering five different agents (patients, triage nurses, bed nurses, physicians and radiologists). The simulated data demonstrated that improved efficiency in radiology services results in a four fold reduction in patient length of stay.

It is important to note that in the literature regarding simulations health systems,

conflicting findings are not uncommon. For example, Kirtland et al. [50] and McGuire [51] come to opposite conclusions about intermediate waiting areas, while the findings Blasak et al. [52] and those of González et al. [53] conflict on the matter of increased nursing staff. It can be deduced from this that; while simulation models can provide useful insights for particular hospitals, the necessary level of detail, and ability to tailor models to particular EDs mean that great care should be taken when extrapolating findings globally across different health systems.

3.3 Statistical models

Statistical models, in contrast to the previously mentioned approaches, look at the overcrowding from a *data-centric* perspective as opposed to a *model-first* perspective. In this manner, papers will analyse a data set with the intent of understanding its patterns and structures. Evidently, such research requires access to data sets which often come as part of an agreement exclusive to particular hospitals, meaning that most studies are carried out on a hospital-by-hospital basis. Work using statistical methods is almost exclusively oriented toward creating forecasting models, largely for the purpose of resource allocation and planning. The two differentiating factors that should be noted in forecasting models are the scale of the forecasting horizon and the model type.

The scales of forecasting horizons are largely dependent on the observation frequency of the data in use. In the literature, these frequencies vary from hourly observations through to monthly and annual observations. Evidently, the observation frequency of a particular data set naturally influences the time scales on which effective forecasting can be performed (prediction horizons can not be smaller than observation intervals). The majority of studies consider forecasting models built on daily data, allowing forecasts of behaviour for one to multiple days into the future.

The nature of data used in these studies also guides the type of model that is used and the intended outcomes. Discussed below are the main types of models in use, spanning time series models (both univariate and multivariate) and regression style models.

A study of a French hospital ED by Kadri et al. [54] displayed effective use of time series models to generate forecasts of daily attendance. The data set contained patient count data stratified by GEMSA (a retrospective classification system²) classes, allowing the authors to build specific ARIMA models for each of the GEMSA classes. Results indicated that the additional class information allowed for improved forecasts. A similar approach was taken in a different French ED by Afilal et al. [55], who also considered the CCMU classification (A second retrospective classification system³). The study had

²Groupes d'Etude Multicentrique des Services d'Accueil, or Multicentric Emergency Department Study Group in English, is one of two retrospective emergency patient classification systems used in France. The GEMSA classification is based on the patient's outcome upon leaving the ED

³The second of the two classification systems used in France, the CCMU classification system regards

two-fold goals of building both long-term forecasts using exogenous data as well as daily forecasts with time series models, again with a model for each classification. The shorter-term forecasts were shown to be robust against short periods of epidemic illness (Flu and Acute diarrhoea). The use of exogenous information (namely weather data) was also used for daily attendance forecasts in a Singaporean ED by Sun et al. [56], demonstrating a positive correlation of attendance with average ambient temperature over the study period. Another observation by Sun et al. [56] noted that Sundays and Mondays tend to be the most active days of the week, with the increased activity attributed mostly to low-acuity patients. Further to this, hourly census and imaging data were used by Jones et al. [57] to build multivariate autoregressive models, which were shown to outperform exponential smoothing methods.

Models outside of traditional time series constructions have also been implemented for ED prediction. A suite of regression models including Generalised Additive Models (GAMs) were used to build daily forecasts of ED presentation numbers, with results indicating that GAMs were able to outperform traditional time series models [58]. Another non-standard prediction method was used by Sun et al. [59], retrospectively implementing a quantile regression method in order to predict patient waiting times based on acuity information and the current state of the ED. After classifying patients by their acuity levels, the resultant models achieved median absolute prediction errors of 11.9 minutes and 15.7 minutes for the medium and low acuity categories respectively.

3.4 Generalisable ED models

The models described above are intentionally developed to solve very specific problems, in particular settings, namely individual EDs. In this fashion, researchers are able to build models that can solve specific and narrow problems effectively. The drawback of this approach however is that the results can rarely be generalised to apply to problems outside the scope of the research, in particular solving analogous problems in other EDs.

A subset of work in the area of simulation health care modelling considers generalisable models that can be applied more broadly. In this fashion, baseline best practice procedures can be more easily established and attained. Fletcher and Worthington [60] provide a useful description of the variety of ways in, and degrees to which a model can be *generic*. Fletcher and Worthington [60] separate models into four categories; (i) *Generic principle models* that are not industry specific (generic queueing models for example), (ii) *Generic framework models* which can provide a toolkit to develop models when combined with local understanding and data, (iii) *Setting specific generic models* which describes transportable models for which inputs can be altered for different settings and finally (iv) *Setting specific models* which are custom built for a particular setting with limited transferability. It is key to note that *generic* as an adjective for a model, can be with regard to model abstraction, the necessary resources used to treat the patient.

model transportability or implementation software.

More recently, papers by Furian et al. [61] and Boyle et al. [62] have addressed the requirement of a framework for generalisable ED modelling. Furian et al. [61] carried out a comprehensive review of ED modelling literature and proposed a conceptual framework for *Generic* ED models (in contrast to *Specific* ED models). Boyle et al. [62] refines the dichotomy of models to include *Generalisable* models, which sit between generic and specific models both for their level of abstraction and requirement for setting specific detail. The authors then go on to define a framework for generalisable DES models, which is applied and validated against an Australian ED.

A consensus on the ideal model and level of generality is yet to be met, but a number of models and frameworks have been developed with these ideas in mind. Initial work by Sinreich and Marmor [63] worked to validate and develop a foundation for generic simulations by analysing the patient flows from five separate EDs. The use of generic simulation tools was also validated by exhibiting statistical similarities in patient paths across the hospitals. Yang et al. [64] take a slightly different approach by developing a simulation model in a purely theoretical context as a tool to test triaging processes and resource allocation strategies. A generic ED simulation model was developed by Gunal and Pidd [65], which was subsequently embedded within a broader hospital simulation tool allowing generic studies of *bed blocking* [66].

A number of studies have applied generic tools to specific EDs to varying degrees of success. A financially driven study by Ferrin et al. [67] develops a generic simulation tool, and applies it to an ED in Arizona, with results exhibiting a 7% reduction in average length of stay alongside a 20% in average daily attendances. It is important to note that Fletcher et al. [68], cited a number of implementation difficulties of generic models including data quality and organisational dysfunction.

3.4.1 Project goals in context

With this understanding of the literature, the goals of this work become more defined. The extremely unique nature of the *SA Health Emergency Department Dashboard* lends itself to taking a statistical modelling approach. The generic nature of the data from the dashboard means that rather than building a prediction method for specific EDs, this project will, unlike the literature discussed develop a universal statistical modelling framework which can be applied to any ED with access to data with either the same or analogous measurements of behaviour. Clearly, the temporal nature of the data will imply that time series models will play a key role in the model building process, in particular, the half-hourly frequency of observations will allow for shorter-term forecasting (0.5-5 hour prediction horizons) which is in contrast to the vast majority of previous work. The research also has the objective of being end-user focused, in that the final product should not require a high level of technical skill to yield useful information for decision making purposes.

Chapter 4

The data set

In the previous chapter, it was noted that the existing literature has exposed a requirement for models which provide short-term decision support, whilst also being generalisable and widely applicable. The data set on which this work is focused provides a novel opportunity to carry out research distinct from the existing literature. Relative to data previously used, the data set considered in this work is extremely unique in that it is reported at a highly generic level of detail, and at a higher frequency, lending itself to the creation of a generalisable modelling framework that can be used for short-term predictions. The collection and collation of this data provides a foundational testing bed on which generalisable models can be investigated, built and baselined. This chapter contains a discussion of the data and its characteristics.

The following sections detail the collection process for the data and describe both the features and challenges it poses. In the final part of this chapter, exploratory insights which inform the model building approach are detailed.

4.1 The data set

This work is centred on an accumulation of data published on the *SA Health Emergency Department Dashboard* [69]. The online dashboard is part of an initiative developed by *SA Health*, with the intent of aggregating and displaying information regarding the status of seven main public EDs in South Australia. The goal of the dashboard is to aid prospective ED patients by providing up-to-date information about their potential destinations for treatment.

Each half-hour, the dashboard presents an updated snapshot of the emergency departments in the following seven hospitals:

- Flinders Medical Centre
- Lyell McEwin Hospital

- Royal Adelaide Hospital
- Modbury Hospital
- Noarlunga Hospital
- The Queen Elizabeth Hospital
- Women’s and Children’s Hospital

As displayed in Figure 4.1 for each hospital, a range of tables help build an overview of the ED’s current state. The main *Status* table in the top left of the dashboard describes all seven hospitals at a high level, while the remaining tables provide a more detailed description of a given hospital

4.1.1 Data aggregation

Data published on the dashboard was accessed and aggregated (*scraped*) by an **R** program utilising the `jsonlite` package, along with a `cron` infrastructure. A schematic of the data collection model is contained in Figure 4.2. The scraping program was set to run every fifteen minutes, five minutes after the dashboard was updated. In this way, each update of the dashboard would be attained twice allowing for infrastructure breakdowns of either the scraping program or the dashboard itself. Despite these efforts, several issues persisted, the details of which will be contained in the following sections.

When data is attained every fifteen minutes, it is collected as ten individual `.json` files. The information on the dashboard is partitioned into the ten files as described in Table 4.1. Given that such a storage format is not conducive to analysis, for each scrape, the collection of files was unpacked and collated into a tabular long form more fit for analysis and model fitting.

The large majority of analyses completed in this work concern data from the *status* table which can be found in the top-left of the dashboard.

4.1.2 Variable descriptions

The data used in this work is predominantly composed of observations of the following variables from the *status* table. Variables are listed in the order they appear on the table.

Hospitals

The seven hospitals that form the subjects of this work are detailed in Table 4.2.



Figure 4.1: An example of the Emergency department dashboard from July 2020. In this screenshot it can be seen that FMC (Flinders Medical Center) is selected in the top left status table, indicating that the remaining tables all describe the FMC emergency department in further detail.

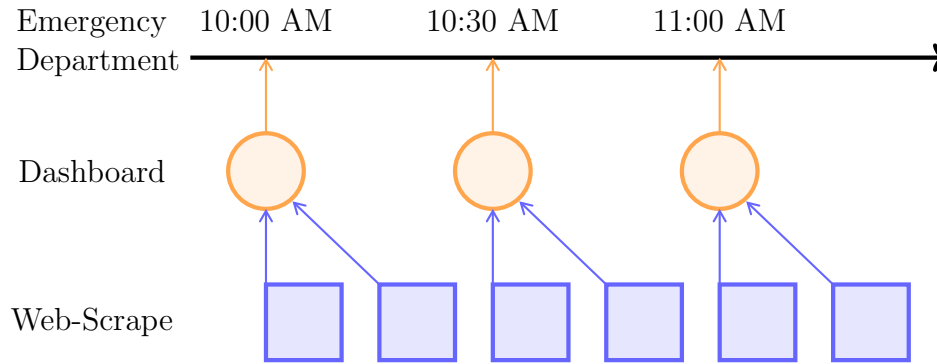


Figure 4.2: A schematic of the data collection model. The continuous black line resembles the continuously evolving situation in a given emergency department, the orange circles indicate the discrete observations that the dashboard is making of the emergency department. Finally, the blue squares indicate the duplicated data collected from web scrapes occurring every 15 minutes, 5 minutes offset from the dashboard update.

Expected Arrivals

These observations (labelled *EA* in Figure 4.1) resemble the information emergency departments have regarding anticipated emergency arrivals, namely from ambulance services or rarely from other sources.

Patients Waiting for Assessment

The number of patients waiting for assessment (*WTBS* in Figure 4.1) reflects the number of people who have presented to an Emergency Department and have been assessed by a triage nurse, but are yet to be seen by a clinician.

Patients in Treatment

Referred to as *Com Treat* in the status table, this is the number of patients whose treatment has *commenced*, that is; they have been seen by a clinician. It is important to note, however, that while they may have been seen by a clinician, patients in this category may still be waiting for a bed.

For example, it can be seen in Figure 4.1 that the *Flinders Medical Centre* states this number as 59, however the *Total* row in the *FMC Streams* table indicates that 31 of the 59 patients are in treatment, while 24 of them are still waiting for a bed (the remaining 4 patients are accounted for in the penultimate column which resembles the extended emergency care unit, a ward in which patients are monitored for 24 hours before they are allowed to leave).

File prefix	Information
ED001	Status table
ED002	Activity Last 24 hours table
ED003	Streams table
ED004	Waiting Times table
ED005	Information for bottom 3 plots
ED006	Triage Categories table
ED007	Expected Arrivals table
ED009	Departures table
ED_Etl_Cntrl	Date and time record
ED_Message_Board	Messaging and information

Table 4.1: *Allocation of the dashboard data into the ten .json files attained in each scrape with reference to information displayed in Figure 4.1.*

Hospital	Abbreviation	Location	Capacity
Flinders Medical Centre	FMC	South	65
Lyell McEwin Hospital	LMH	North	51
Modbury Hospital	MH	Inner North	35
Noarlunga Hospital	NHS	South	26
Royal Adelaide Hospital	RAH	CBD	69
The Queen Elizabeth Hospital	TQEH	Inner North West	29
Women’s and Children’s Hospital	WCHP	Inner City	26

Table 4.2: *Hospitals being considered for the analysis. The stated capacity is that according to the dashboard, and location is relative to Adelaide.*

Capacity

Labelled *CAP* on the dashboard, this resembles the quoted capacity of each of the emergency departments.

Average Waiting Time

The average waiting time observations (*Avg Wait* in Figure 4.1) are stated to be the average waiting time (difference between the time of triage and assessment from a clinician) over the past 2 hours, excluding high priority cases. It should be noted that no information is provided regarding what is classed as a *high priority* case, nor the way in which the average is calculated (the inclusion criteria for the calculation).

Capacity Index

The final variable detailed here does not appear on the dashboard but rather is calculated using variables on the dashboard. The capacity index is inspired by the *traffic light style* colouring system in the penultimate column of the *Status* table which reflects the ratio of the Patients in Treatment to the Capacity.

Rather than using a statistic describing the number of patients in treatment, it was decided instead to consider a statistic describing the number of patients *needing* treatment. Hence, the Capacity Index for a hospital H was defined to be:

$$I^{(H)} = \frac{N_T^{(H)} + N_W^{(H)}}{C^{(H)}}, \quad (4.1)$$

where $N_T^{(H)}$, $N_W^{(H)}$ are the numbers of patients in treatment, and waiting for treatment respectively, and $C^{(H)}$ is the stated capacity in hospital H . In this way, an entirely full ED will have a capacity index equal to one, while a half-full ED will register a value of 0.5. Note that in the interest of legibility, the superscript (H) will be omitted in further references to these variables. It will be treated as implicit that variables will be referring to observations from individual hospitals, and that the seven hospitals will be considered in parallel.

A key benefit to the use of the capacity index is its normalised nature lends itself to the building of generic models applicable to any hospital with the ability to capture elementary data.

4.2 Missing and erroneous data

Initial analyses of the data set were focused on identifying and understanding complications that result from the data collection and attainment process. Infrastructure failures on the part of either the dashboard or the data collection process can result in either missing or corrupted observations. Building models with potentially missing or corrupt data can result in models that are biased or do not accurately describe a system. As such, care was taken to properly understand and detect these issues in the data, in the following pages the types and prevalence of failures will be detailed.

Dashboard failure

In cases where the infrastructure on which the dashboard is hosted is either not functional or undergoing maintenance, no data can be recorded. As indicated in Figure 4.3, one or multiple entire snapshots will be missing from the final data set, meaning the information will go unobserved. While the scraping strategy allows for short outages (fifteen minutes or less), there have been prolonged outages which meant data went unobserved.

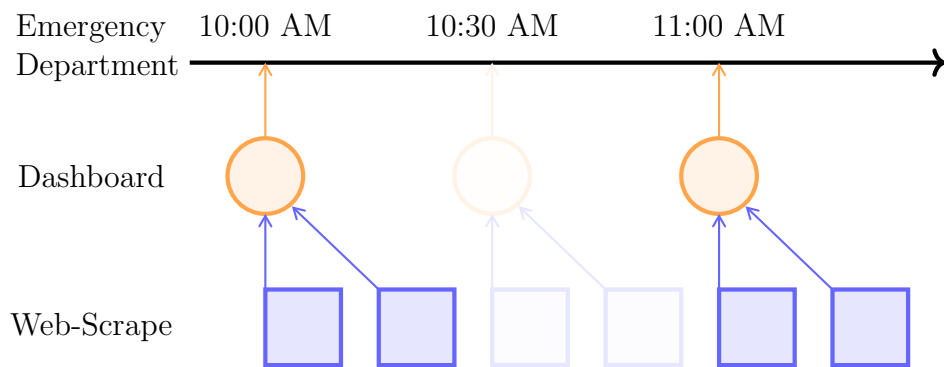


Figure 4.3: *Dashboard Failure:* In a similar fashion to Figure 4.2, this schematic depicts the collection of data for the case in which the dashboard is unable to provide data. This is commonly the case when the dashboard is non-functional resulting in a failed scrape.

Web scrape failure

With results analogous to dashboard failure, there have also been cases in which data was unable to be collected due to issues with the web scraping infrastructure. As the scraping program was run locally on a computer, power and internet failures resulted in the program being unable to run and attain data. This failure mode is represented by Figure 4.4.

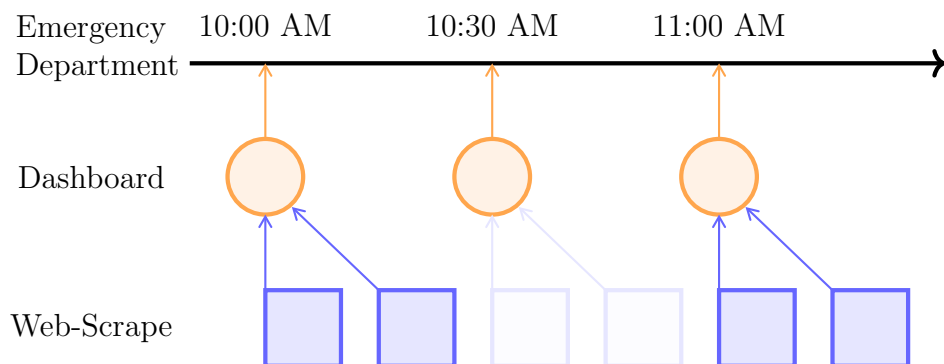


Figure 4.4: *Web scrape failure:* This figure describes the case in which the web scraping infrastructure breaks down due to power failures or other technical issues. This failure manifested in a nearly identical way to that in Figure 4.3.

Partially missing data

An instance that is not uncommon, is when a subset of data is missing from a particular scrape or set of scrapes. This can be due to breakdowns at a number of stages in the

process and for a number of reasons. There are cases where hospitals have undergone a system reconfiguration or migration, resulting in the hospital being unable to report any observations for an extended period of time. There are also cases in which individual measures are missing from the data set, which may be due to a lack of observation or reporting errors. These errors are depicted in Figure 4.5

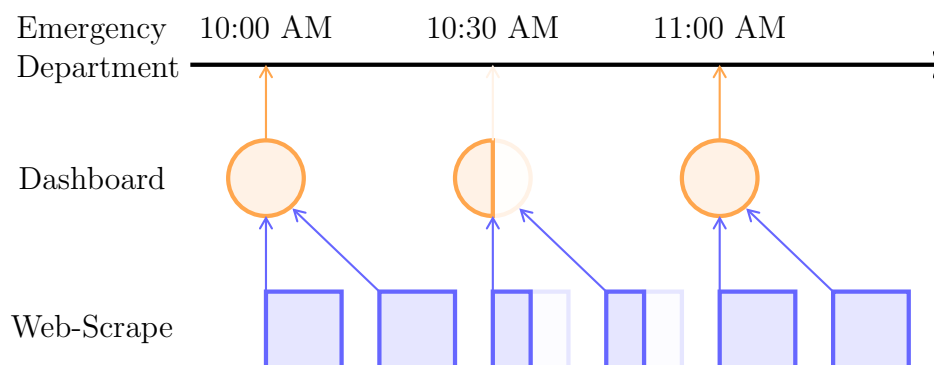


Figure 4.5: *Partially missing data: The schematic depicts a situation in which some proportion of the data set has gone unobserved or unrecorded. This most commonly occurs on a hospital-level basis, but also regularly occurs on a measurement-wise basis.*

Outdated data

There are times in which while the dashboard is still functional, a breakdown has occurred in such a way that the data has not been updated, therefore causing the scraped data to be outdated. This results in the previous data set being scraped twice extra, yielding four duplicated scrapes as indicated in Figure 4.6. This type of event is easily detected using timestamps in the respective tables, which allows for the removal of the duplicated data. It does however mean that a time point will have missing data.

Contradicting records

In the eight .json files that are collected in each web scrape, there exist two distinct records of the time at which the data is current. Largely, these times are identical, however, there are events in which the two-time records are contradictory. Most often, this occurs as a one-time record indicating it is the most recent observation, and the other indicating it is for the prior observation as depicted in Figure 4.7.

Distinct observations

An unexpected issue was the identification of distinct observations for the same time stamp. Figure 4.8 displays a schematic of this occurrence, showing two scrapes of what

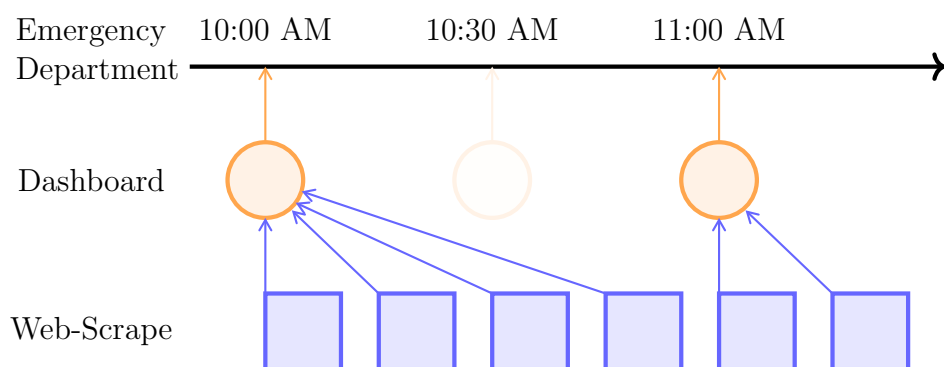


Figure 4.6: *Outdated Data:* This schematic depicts the collection of data for the case in which the dashboard is unable to provide updated data. As indicated by the diagram, in the absence of any updated data, the web-scraping program naturally accumulates two extra duplicated reads of data from the previous outdated time step.

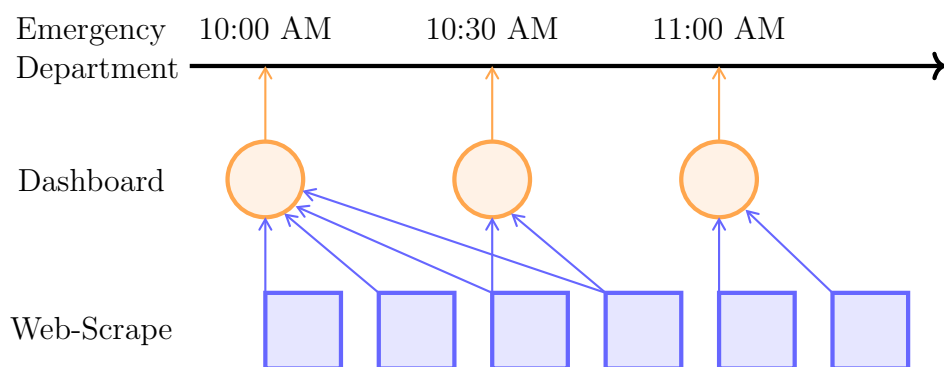


Figure 4.7: *Contradicting records:* The multiple records of observation time do not always agree in single scrapes. As shown in the figure above, the scrapes which should reflect data current at 10:30 a.m., simultaneously indicate they correspond to the 10:30 a.m. as well as the 10:00 a.m. data.

should be duplicated data being recorded as distinct observations.

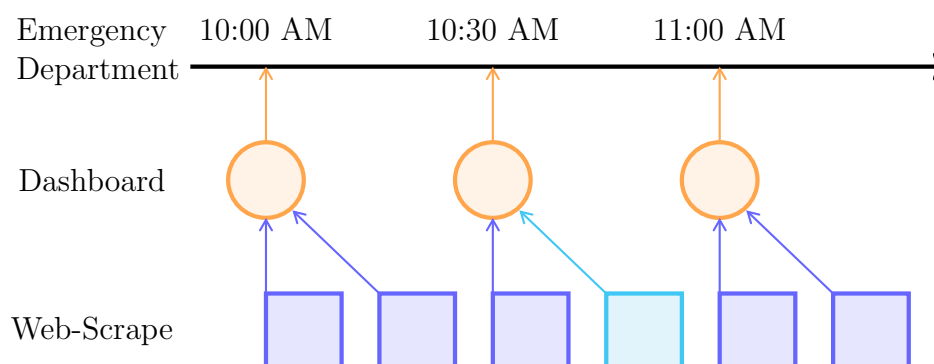


Figure 4.8: *Distinct Observations:* Whilst uncommon, the data set contains cases where observations which should be duplicates of each other, in fact disagree on their records symbolised by the differing colours of the web scraped data.

4.2.1 Inclusion criteria

The prevalence of the errors described above poses potential risks for the predictive capability, and robustness of models to be built. As such, inclusion criteria are imposed on observations to limit the impact of erroneous observations on model results. The criteria regard factors that are indicative of the reliability of individual observations and the data set as a whole. The three factors considered are the missingness, self-consistency, and reliability of the data.

When errors such as those in Figures 4.3 to 4.6 occur, the result is missing observations either entirely or partially. The set of missing observations is visualised in Figure 4.9. In Figure 4.9, there appears to be three discrete hospital-wise tiers of missingness, namely the first regards the LMH, MH and WCHP hospitals with the least missingness, followed by the NHS, RAH and TQEH, finally followed by the FMC which distinctly has the most missing data of the seven hospitals.

Noting the large blocks of missing data in the FMC data set, the supervision panel agreed that the exclusion of FMC data from the work would be advisable. As such, future analysis of the data will consider the remaining six hospitals. In future work, the implementation of sophisticated imputation methods could be applied in order to carry out research into the dynamics of the FMC ED.

In the interest of carrying out analyses that provide consistent and robust insights, it was decided that limiting the original two-year interval of observations to the interval spanning January 2021 to mid-September 2021. This decision was made due to data attainment issues prior to January 2021 causing corresponding observations to be regularly incomplete and at irregular intervals. A dashboard infrastructure change in September 2021 caused reporting intervals to be increased to one hour, causing significant issues for the time series modelling approach applied (discussed in Chapter 5). It is acknowledged

that using this interval of data will limit the model's ability to capture seasonal, and annual variations. It is recommended that any work extending that of this thesis considers a larger interval to inform this investigation.

Regarding self-consistency and reliability of data, it was decided that given the size of the data set in use, a conservative approach would be taken in order to maintain model robustness. As a result, data displaying the following features was removed from the data set prior to analysis:

- Outdated or conflicting records of time (as in fig. 4.6,fig. 4.7),
- Conflicting observations (as in fig. 4.8) and
- Evidently erroneous observations.

Imputation process

The presence of missing data, due to both organic missingness and active removal of observations poses challenges in a model building process, particularly in a time series modelling framework. In order to combat these challenges, missing values from the remaining six hospitals were imputed via a linear model. Discussion of the linear model, its building process and assumptions can be found in Chapter 6.

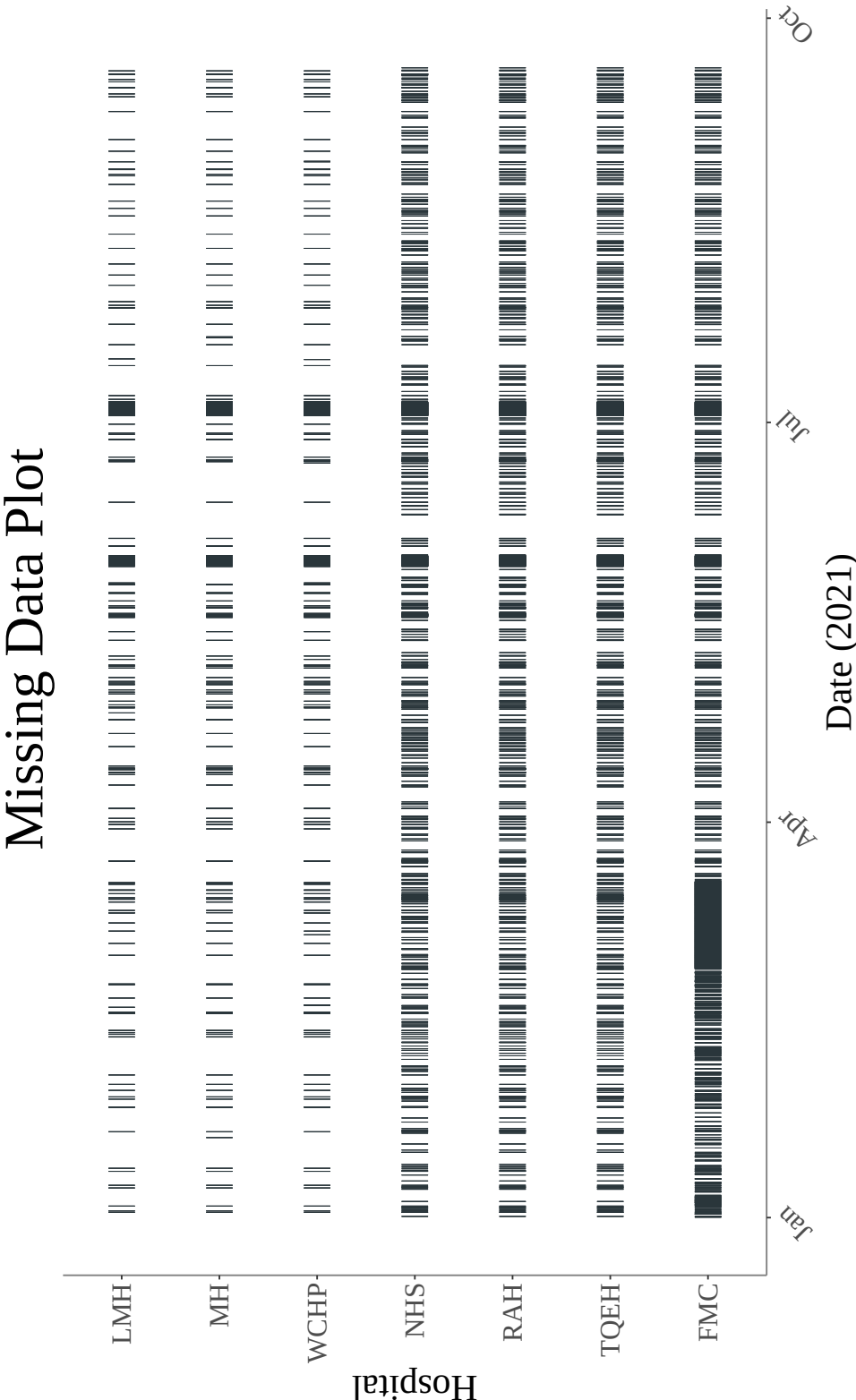


Figure 4.9: Across the data being considered for this analysis, all missing observations from the data set are indicated by vertical lines for each hospital. The three tiers of missingness are observable in this plot, with the top three hospitals (LMH, MH, WCHP) having the least missing observations, followed by the second tier consisting of (NHS, RAH, TQEH), with the FMC clearly having the most missing data.

4.3 Exploratory insights

With the nature of the data set and its collection process detailed above, the focus is shifted to results derived from the data exploration process. The following pages provide insights that provide context to the data being considered, whilst also guiding the model building process detailed in Chapter 6.

4.3.1 Distributional observations

Capacity Index

Recall that the capacity index, as defined in (4.1) is a reflection of the number of patients requiring treatment compared to the capacity and resources to provide that treatment in the given emergency department. Univariate statistics contained in Table 4.3, and visualised in Figure 4.10 provide a stark exhibition of the overcrowding issues in South Australian Emergency Departments.

Hospital	Mean	Standard Deviation	Median	Interquartile Range
LMH	1.18	0.24	1.18	0.31
MH	0.91	0.24	0.91	0.34
NHS	1.00	0.36	0.96	0.54
RAH	1.02	0.16	1.03	0.22
TQEH	1.07	0.26	1.07	0.34
WCHP	0.92	0.43	0.88	0.62

Table 4.3: *Summary statistics of the Capacity Index for all hospitals.*

It can be seen in Table 4.3 that while the Women’s and Children’s and Modbury Hospitals are on average below their capacity, the remaining four hospitals all have sample average capacities above one. Clearly, in some commercial contexts these results may be satisfactory or even desirable however, noting that the purpose of an ED is to have the capability to provide aid in emergency cases these results suggest that South Australian EDs on average do not have a significant amount of spare capacity for new emergency cases.

In Figure 4.10 it is noted that the distributions observed in the six hospitals are certainly distinct. While four hospitals, namely the LMH, MH, RAH and TQEH have roughly symmetric distributions, NHS and WCHP exhibit some severe multimodality and skewness. It can also be observed in both Table 4.3 and Figure 4.10 that NHS and WCHP hospitals have a much broader distribution when compared to the remaining four hospitals, the narrowest of which is the RAH.

An interesting line of enquiry regards the types and modes of behaviours observed in EDs. For example Figure 4.11 contains a histogram of times spent continuously over-

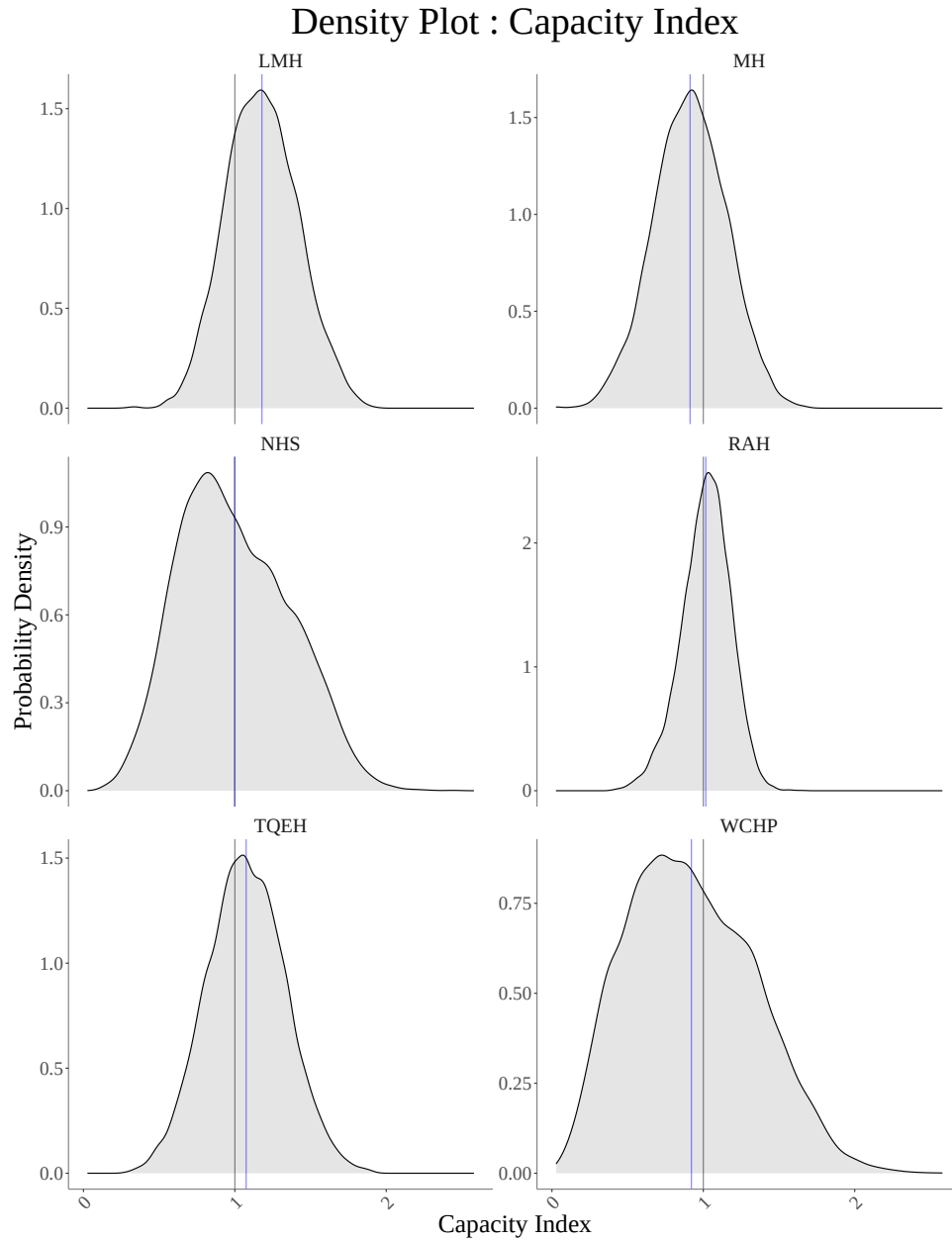


Figure 4.10: *Plots of the empirical probability densities of the Capacity index measure in each of the hospitals. Grey vertical lines in each of the plots indicate a capacity index of one, that is, when the number of patients requiring treatment is equal to the capacity of the ED. Blue vertical lines indicate the sample mean of the Capacity Index for the respective hospitals. Note that vertical axes are not uniform across the six plots, allowing the reader to compare the distinct nature of the distribution shapes.*

capacity. That is; once an ED exceeds its capacity, how long will it be before the ED is under capacity again?

The distributions of times exceeding capacity, across all hospitals exhibit a positively skewed distribution, with mode at 1 hour. Interestingly, hospitals LMH, NHS and WCHP prove to have somewhat heavier tails and multi-modality. It is noted that the latter two of these hospitals are also the hospitals that had the broader and multi-modal empirical densities in Figure 4.10. This suggests that these hospitals experience distinct regimes of demand in which the requirements of, and ability to provide care differ significantly.

Average Waiting Time

The average waiting times in all hospitals can be seen in Figure 4.12 to have somewhat similar distributions in that all display a heavy positive skew, with a mode at around 100 minutes. The hospitals with the three lowest medians are MH, RAH and TQEH with medians of just over an hour, while the greatest median average waiting time is that of LMH with a value of 106 minutes. The key differentiating factor in the distributions is the variation in widths that are observed. Each of the heavily skewed distributions can also be seen to contain numerous outliers in the positive tails.

Expected Arrivals

Figure 4.13 contains a side-by-side box plot for the observed values of expected arrivals at each of the hospitals. While all hospitals tend to exhibit a positively skewed distribution each with a number of outliers, the scales and locations of the empirical distributions differ significantly. Figure 4.13 shows that the RAH has a significantly greater median and interquartile range in comparison to other EDs. This result is not unexpected given that the RAH is the most centrally located hospital in the data set, making it the destination for the greatest proportion of ambulance journeys. Another expected result is NHS, the hospital furthest from the city in the data set having the least median and interquartile range. In a similar fashion to the average waiting times, Table 4.4 indicates that the variances of the distribution tend to scale with the mean. This is a feature also of the Poisson distribution which is commonly used as an arrival distribution in queueing analyses.

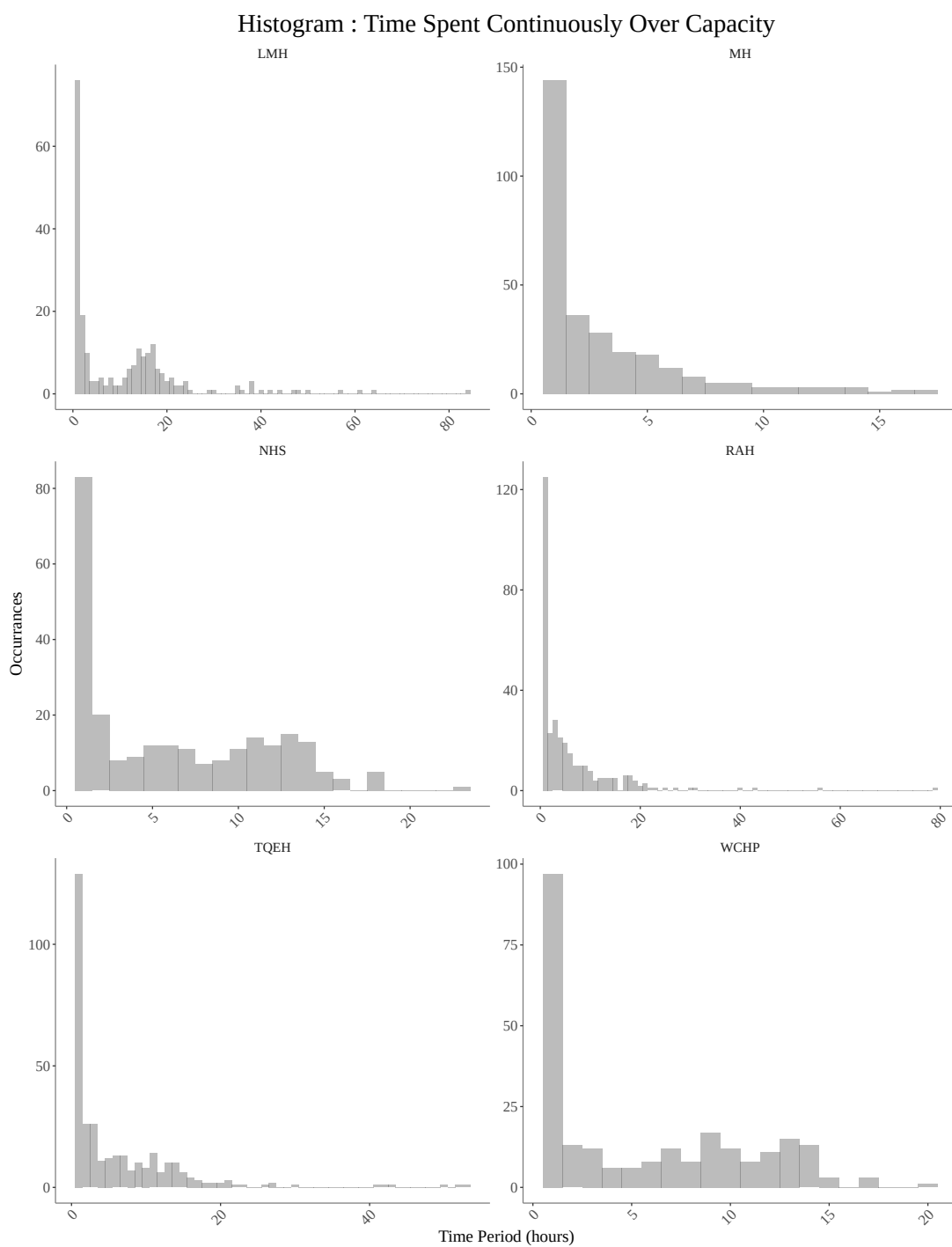


Figure 4.11: Histograms of the times spent continuously over capacity. For each hospital, the number of occurrences each hospital has continuously been over capacity for a particular number of hours is shown.

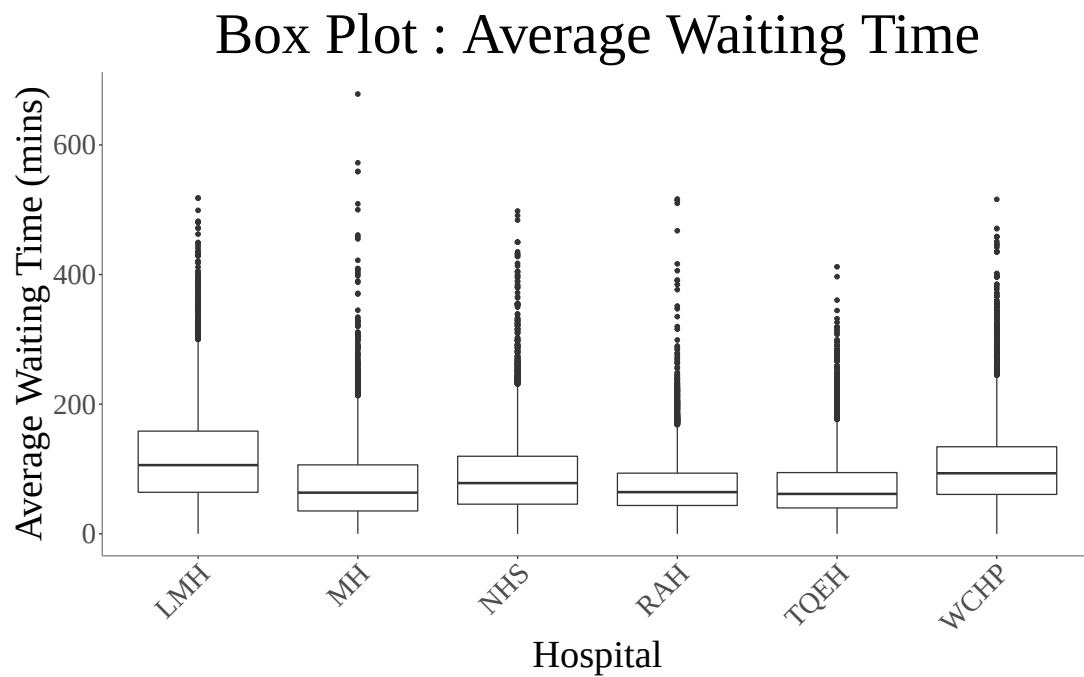


Figure 4.12: *Box plot of the empirical probability densities for the average waiting time in each of the hospitals. Each of the hospitals exhibit a clear positive skew, each with significant numbers of positive outliers.*

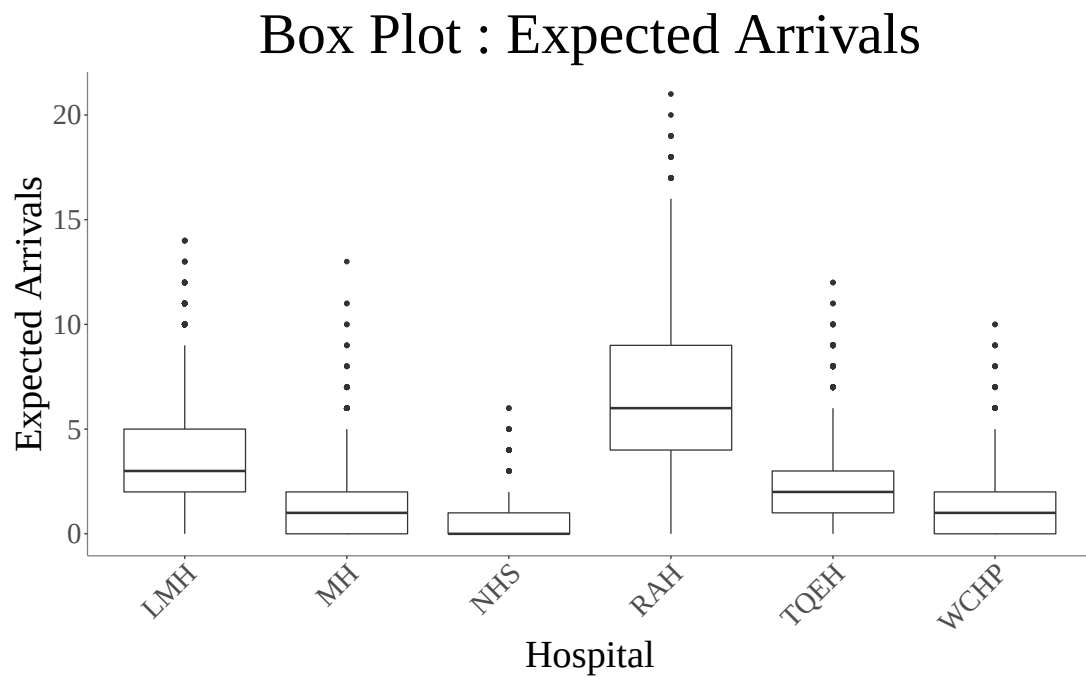


Figure 4.13: Side-by-side box plot of the expected arrivals in each of the hospitals. The notable takeaway from this plot is the inflated values corresponding to the RAH, consistent with its central location in the Adelaide CBD.

Hospital	Mean	Standard Deviation	Median	Interquartile Range
FMC	4.97	3.15	5.00	4.00
LMH	3.62	2.40	3.00	3.00
MH	1.53	1.47	1.00	2.00
NHS	0.60	0.84	0.00	1.00
RAH	6.08	3.54	6.00	5.00
TQEH	2.17	1.79	2.00	2.00
WCHP	1.06	1.25	1.00	2.00

Table 4.4: *Summary statistics of the number of Expected Arrivals for all hospitals. It can be noticed that across the hospitals, the variability of the distributions tends to scale with the mean.*

4.3.2 Pattern based analyses

In an effort to develop an understanding of structures in the data, the possible bivariate relationships in the data were explored. In this section, these patterns are exposed in order to better motivate and inform the model building in the following sections.

Temporal patterns

A natural first place to look for structured patterns was by understanding the temporal scales on which behaviours tend to cycle. As such, investigations into temporal patterns throughout the day and over a week, became extremely informative.

Figure 4.14 exhibits the daily patterns existent in the data, namely for the Capacity Index as a proxy for how *busy* each of the hospitals are. Clearly, each of the hospitals appear to follow quite a similar trend in which a trough is often observed early morning between six and eight o'clock in the morning, before each hospital becomes busier, peaking between early afternoon and evening. This may be due to a lower level of engagement in potentially dangerous activities in the mornings as opposed to the remaining parts of the day. Another distinct point from the plot is the depth of the trough for WCHP. Recalling that the *Women's and Children's* hospital in the inner city, this observation may be explained by an accentuation of the aforementioned difference in morning behaviour, as the vast majority of those eligible to attend the WCHP would be attending or preparing to attend school.

As may be expected, the widths of empirical densities in Figure 4.10 are correlated with the range of mean behaviours observed in the daily profiles. In particular, it is noted that the RAH and MH EDs have the tightest bounds on their mean behaviour, while the WCHP and NHS exhibit varied values in their mean behaviour.

Expanding the temporal scale, weekly patterns present in the data are also considered. Figure 4.15 exhibits a number of interesting features. While one might expect hospitals to

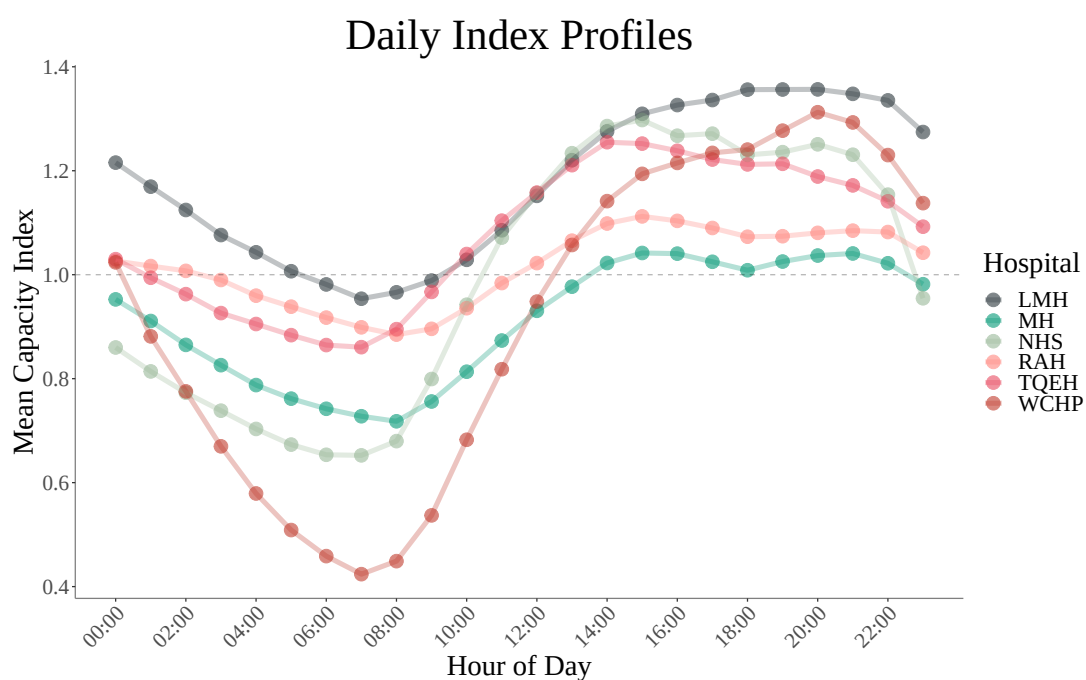


Figure 4.14: Plot of the daily profile of the Capacity Index in each of the hospitals. Each point resembles the mean value of the Capacity Index at the given hour of day. The horizontal dashed line represents a capacity index of one, that is, a full emergency department.

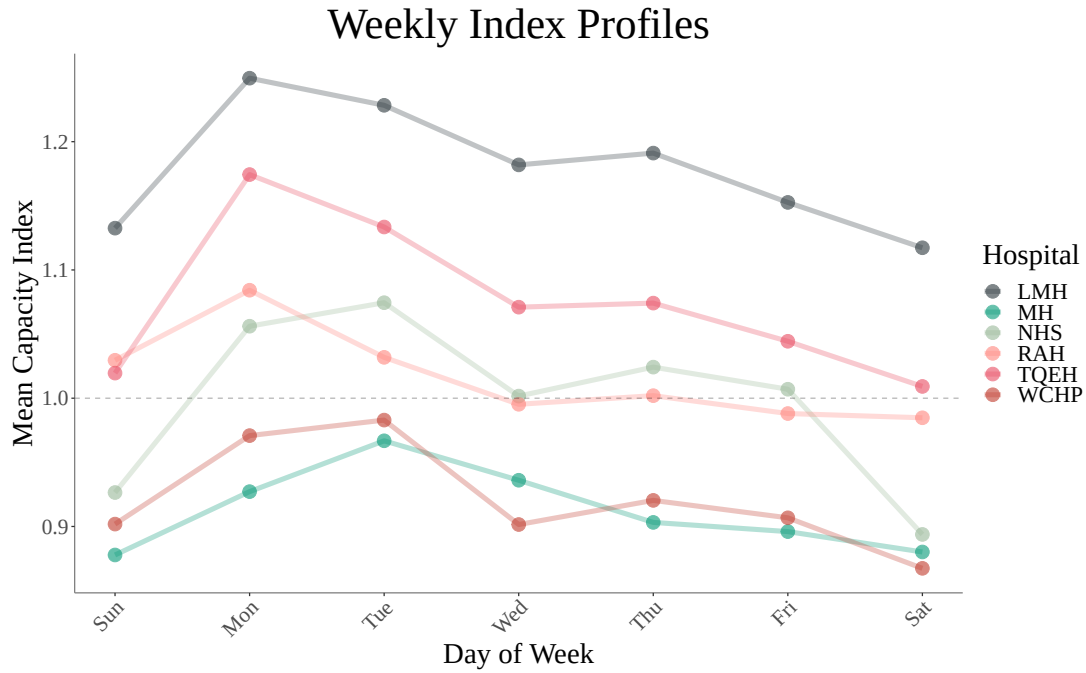


Figure 4.15: Plot of the mean values of the Capacity Index in each of the hospitals throughout the week. Each point in the plot resembles the mean value of the Capacity Index for a particular hospital, on a particular day of the week. The horizontal dashed line represents a capacity index of one, that is, a full emergency department.

reach peak business on weekend days, this is not the case in the observed data. Rather, all of the six hospitals are, on average observed to be busiest on either Mondays or Tuesdays. This result may be attributable to a number of possible explanations. Firstly, while staffing information is not available it is plausible that extra staffing is provided on weekends to increase ED throughput. Secondly, there may be a build-up of hospital admissions during a weekend, resulting in a lack of free in-hospital beds for admission from ED cases hence causing a backlog of patients to develop. Finally, a behavioural reason may be that patients are reluctant to attend an ED on a weekend until their symptoms reach a threshold at which point they decide to take action on a Monday or Tuesday.

From these initial analyses, a few key ideas have been concluded which should be kept in mind in the sections that follow. In particular, the nature in which the six emergency departments exhibit both similar and different behaviours. It is observed that even when using a normalised variable in the Capacity Index, the individual hospitals vary on different scales, and exhibit different behaviours based on location and the nature of the care they provide. Contrastingly however, there is also a striking similarity in the ways the EDs operate throughout the day and the week. Understanding these similarities and differences aids the process of building a robust prediction model.

In this chapter, the source, attainment process and nature of the data set in use for this work have been described and investigated. The challenges posed by the data set, as well as the characteristic properties and behaviours it exhibits, have been detailed and will form the basis for the model building process discussed in Chapter 6. Next however, Chapter 5 develops the theoretical fundamentals required to model the temporal dependencies observed in this chapter.

Chapter 5

Temporal modelling methods

In Chapter 4, characteristic behaviours in the data were investigated to guide the process of building predictive models. These results, along with the time-series nature of the data imply that the use of temporal modelling methods is crucial to achieve the goals of short-term prediction. In this chapter, the underlying theory of temporal modelling methods and model assessment techniques are discussed and developed. Application of these techniques then follows in Chapter 6.

5.1 Problem Specification

As has been previously noted, the utility of a predictive model lies in the accuracy of the prediction, and the quality of information you can use to predict. As such, the choice of outcome or response variable for the predictive model is crucial to the utility of this work. For this reason, it was discerned that the capacity index Equation (4.1), which for a given hospital was previously defined as:

$$I = \frac{N_T + N_W}{C},$$

is the most suitable statistic for prediction. Recall that in the definition of the capacity index, N_W and N_T are the numbers of patients waiting and in treatment respectively, and C is the stated capacity of the relevant emergency department. The capacity index I has been chosen as it provides a useful summary of the pressure under which an ED is operating, and importantly, its normalised nature allows for translation of the metric across hospitals.

It was noted in Chapter 3 that the objective of this work is to create models that produce short-term forecasts of ED behaviour, using statistical methods. More concretely, the goal is to effectively predict I from thirty minutes up to five hours into the future, for each of the six hospitals being considered.

Central to this work is the process of using historical data to build forecasts of the future, namely through statistical models. That is to say, using observed data up until a time t_0 , in the form of the previous capacity index values $I_{0:t_0}$, and other observed variables $\mathbf{V}_{0:t_0}$, the objective is to produce a forecast of the value $\hat{I}_{t_0+h}$, the capacity index h time steps into the future. This is written as:

$$I_{t_0+h} = \hat{I}_{t_0+h} + \varepsilon \quad (5.1)$$

$$= f(I_{0:t_0}, \mathbf{V}_{0:t_0+h}) + \varepsilon, \quad (5.2)$$

where h is the desired forecast horizon (the time between the most recent observation and the forecast), and ε is a random variable indicative of the forecast error.

The process of identification, estimation and evaluation of the model f , is the crux of this work along with the parameterisation of ε . The following chapters are focussed on building, comparing and assessing the functional forms f can take and their corresponding utility.

5.2 Time series modelling theory

In the process of identifying the most appropriate model f , an answer is being found for the question, how best can historic data (from times 0 up to t_0), be used to predict the immediate future (at $t_0 + h$)? This question naturally leads research in the direction of time series models which explicitly define the ways in which data is *autocorrelated*, that is, how correlated the current behaviour of a temporal random variable is with its history. In the following section, the details of time series theory are examined, providing context for their use in later in this work.

5.2.1 Time series processes

The vast majority of statistical methods are concerned with the use of data containing *independent* observations to better understand relationships between variables for the purpose of either discerning the significance of the relationships or extrapolating predictions beyond the original data set. In contrast however, the problem at hand involves a data set containing observations that are intrinsically dependent on each other. The temporal evolution of EDs makes the assumption of independent observations (and hence any methods reliant upon it) invalid. As such, attention is turned to a methodology that leverages this dependence between observations for the purposes of inference and prediction.

The suite of tools with which these temporal dependencies can be examined are time series models. Predominantly, this work will be considering *Autoregressive Integrated*

Moving Average (ARIMA) processes, which are a broad class of stochastic processes that can be customised in order to apply to a variety of contexts. ARIMA processes combine white noise processes (sequences of normal random variables with zero mean and specified variance) with three types of dependency structures namely; autoregressive (AR), general trends (I¹, and moving average (MA). This combination allows for the specification of an ARIMA(p, d, q) process Y_t as :

$$\tilde{Y}_t^{(d)} = \left(\sum_{m=1}^p \varphi_m \mathbf{B}^m \right) \tilde{Y}^{(d)} + \left(1 + \sum_{m=1}^q \theta_m \mathbf{B}^m \right) \varepsilon_t, \quad (5.3)$$

where:

- p is the order of autoregressive dependency,
- d is the order of differencing to allow for general trends,
- q is the order of the moving average dependency,
- φ_m and θ_m are the parameter set of the model,
- ε_t is a white noise process, and
- $\mathbf{B}$ is an operator defined by: $\mathbf{B}Y_t := Y_{t-1}$.

A detailed discussion of ARIMA models, their constituent components and the process to arrive at this expression is contained in Appendix B. In the following sections, the tools and processes used to apply this model and identify its parameters are detailed.

5.2.2 Time series model estimation

With this framework of time series processes, focus is shifted to the the theory underpinning the methods required to apply them to data sets in order to understand and exploit the dependencies in the observations. The model estimation process consists of two key components, first is the identification of the model (*i.e.* choosing p , d and q), followed by parameter estimation (*i.e.* finding φ_m and θ_m) which is carried out via maximum likelihood estimation.

¹The I in ARIMA stands for integrated, referring to the summing required for back-transformation of differenced time series. For a more in-depth discussion of the differencing process, see Appendix B

Model identification

Time series model identification consists of discerning the most appropriate values for p , d and q in the ARIMA model. Approaches detailed in the literature (namely from Box and Jenkins [70] and Hyndman and Athanasopoulos [71]) are now discussed to motivate the modelling approaches applied to the ED data.

Differencing Order: The first step in model identification is the assessment of the data's stationarity and subsequent determination of d . The assessment of stationarity can be carried out in a number of ways, two of which will be used in conjunction. Box and Jenkins [70] describe a more qualitative assessment of the sample Autocorrelation Functions (ACFs, see definition in Appendix B) in which one discerns stationarity or lack thereof. It is noted that a stationary time series will exhibit an ACF, and hence sample ACF which rapidly tails to zero, if such behaviour is not observed, authors advise that an order of differencing is applied to the data and the process is then repeated. In contrast, Hyndman and Athanasopoulos [71] details a process in which the *Kwiatkowski–Phillips–Schmidt–Shin test* is applied, providing an approach with a notion of statistical significance (for the interested reader, details of the KPSS test can be found in Kwiatkowski et al. [72]).

AR, MA Order: In a similar way to above, Box and Jenkins [70] and Hyndman and Athanasopoulos [71] provide contrasting approaches to identifying p and q . Box and Jenkins [70] detail a process using ACF and PACF plots (for definitions and discussions of characteristic features, see Appendix B) to inform a useful model, before reassessing the result by considering the ACF and PACF plots of the model's residuals. The process would then be repeated until a final model is settled.

Hyndman and Athanasopoulos [71] describe a more algorithmic approach to the identification of p and q , with the goal of finding the combination that produces the model with the minimal Corrected Akaike Information Criterium (AICc).

The AIC [73], and its modification *AICc* [74] are measures developed in information theory as estimates of the *information lost* when a model (such as f from Equation (5.2)) is used for estimation (details can be found at [73]). In this way, a comparison of models' AIC values provides an effective comparison of model fit. Specifically, the original measure, AIC is defined as:

$$AIC = 2k - \ln(\hat{L}),$$

where k is the number of parameters in the given model, and $\hat{L}$ is the maximum of the likelihood function for the given model and data. As one may anticipate, in any comparison between two models the model producing the smaller AIC is desirable, as this indicates less information is lost by the model. In this way, the use of AIC favours both a better fitting model and a more parsimonious model.

The modification (or *correction*) made to the AIC measure by [74], produces the following formula:

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1},$$

where n is the number of data points in the training set. The modification, produces an estimate of information loss that is more robust against smaller data sets, with the modification becoming insignificant for larger data sets [74].

The algorithmic approach detailed by Hyndman and Athanasopoulos [71], initially developed by Hyndman and Khandakar [75] is described below. Once an appropriate value of d has been identified, the following steps are followed:

Algorithm 1: Hyndman-Khandakar algorithm

1. Fit the four initial models:
 - ARIMA(0,d,0),
 - ARIMA(1,d,0),
 - ARIMA(0,d,1),
 - ARIMA(2,d,2)
 2. Set the *current model* to be that with the smallest AICc
 3. Variations to the *current model* are made by:
 - Alter p and/or q by ± 1
 - Include/remove a constant c from the model
 4. Repeat steps 2. and 3. until the AICc value cannot be improved.
-

This algorithm allows for exploration of the model space by making stepwise simplifications and complications, using the AICc measure as a heuristic for model quality. As will be seen in the following sections, the ideas laid out in [70] and [71] guide the modelling procedure carried out on the ED data.

5.3 Model assessment protocols

As the key objective of this work is to build an effective predictive model, the model analysis methodology forms a crucial component of the modelling process. In particular, the temporal nature of the problem requires careful attention to how the time-dependent

structures in the data are identified and understood. In this section, details of the strategies with which the models are tested are discussed. The forecasting nature of this work has particular consequences on the ways in which models can be built and analysed.

With the model from Equation (5.2) in mind, as t_0 increases, the response variables previously predicted become predictor variables for future prediction efforts. Consequently, one cannot cleanly partition the data into two sets of *predictor* and *response* variables. As a result of this from a model assessment perspective, one also cannot simply separate the data into distinct *training*, *testing* and *validation*. Instead, to assess model performance for the purposes of model selection or parameter tuning, a method that evolves with t_0 is required.

Standard Cross Validation

In the processes of model development and tuning of standard predictor-response style statistical models, a standard approach is to apply a k -fold cross-validation method, depicted in Figure 5.1. To apply this technique, observations in the data set are first partitioned into k folds at which point, a model can be trained on k subsets of the data which are each composed of $k - 1$ of the partitions. Each of the k resultant models can then be tested on the partition that was omitted from the respective training sets. This manner of producing k pairs (splits) of training and testing data means that all data are used as both training and testing data, leading to more efficient use of the observations.

For the problem at hand however, this style of model assessment cannot be implemented for a number of reasons. Chief among these reasons regards the fundamental assumption of cross-validation; that each of the observations are both *independent* and *identically distributed*. It is clear that the data being considered for this project does not satisfy these assumptions of independence, in fact, the models being built (5.2), explicitly rely on the dependence between adjacent observations.

Second, it is noted that the behaviours of EDs and the way populations interact with them evolve in time. This fact combined with the structure of time series models indicates that the models being built can only be tested on data that immediately follows the training data (in a temporal sense). This allows for seamless prediction in time series models and ensures that the model is tested on data that is representative of the training set.

Rolling Origin Cross Validation

As such, the methods implemented for this work are an adaption of the methods described by Hyndman and Athanasopoulos [71]. Authors describe *Rolling Origin* cross-validation (ROCV) method for time series, whereby a series of models are built and tested by sequentially rolling forward the *origin* (the point of forecast) by one observation at a time, as depicted in Figure 5.2. In this fashion, the training set increases in size by one

Standard cross validation splits

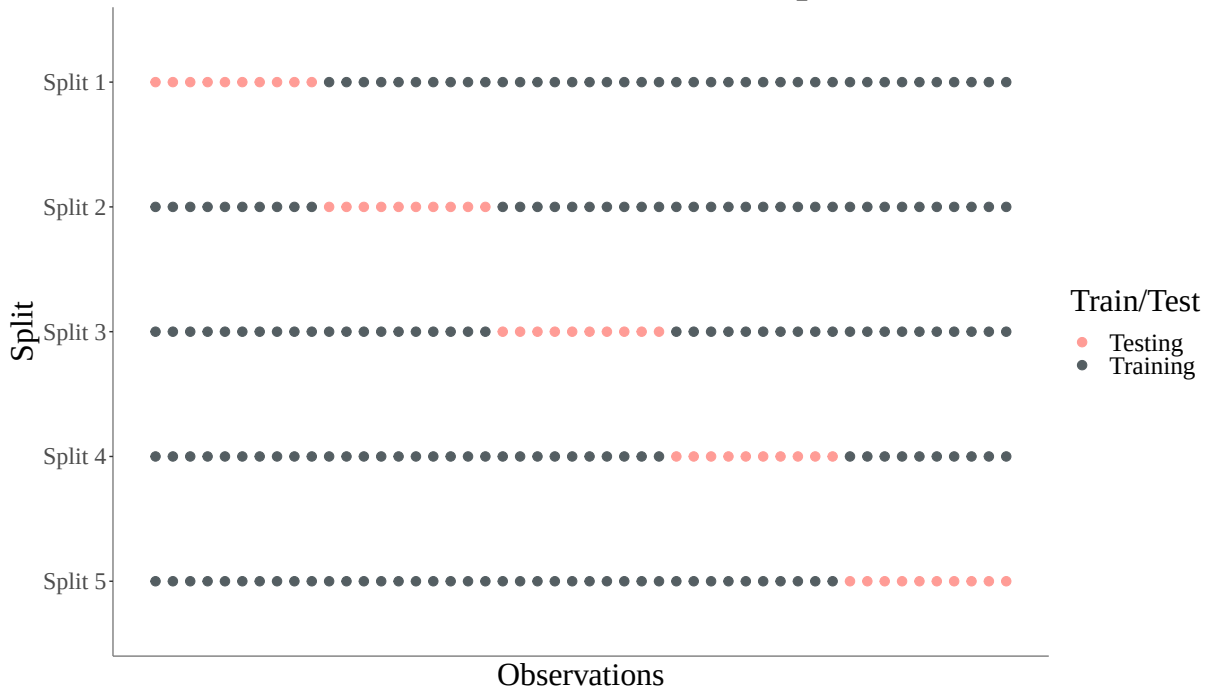


Figure 5.1: A schematic of the k -fold cross validation methodology on a data set with 50 observations with $k = 5$. Each point represents a specific data point, where in each of the splits the grey and pink points are training and testing data respectively.

observation for each sequential model, and the testing set shifts in line with the rolling origin.

This method also looks to use all observations (except the first training set, and final test set) for both model training and testing, but has the advantage of allowing models to be built and then tested on data immediately following the training set.

Drawbacks of this method however centre on the increasing size of the training set for each subsequent training/testing split. The first of these is the emphasis the procedure applies to early observations. As the earlier observations are in the training set for more models than later observations, the models and hence error estimates become biased toward data appearing earlier in the data set.

The second, and arguably more material drawback is the non-uniformity of the error distributions that result. It is claimed that directly comparing the errors in the consecutive test sets does not necessarily make sense, as the distributions of the forecasts change as a function of the training set size.

In order to illustrate this difference in the fits and forecast distributions, an experiment

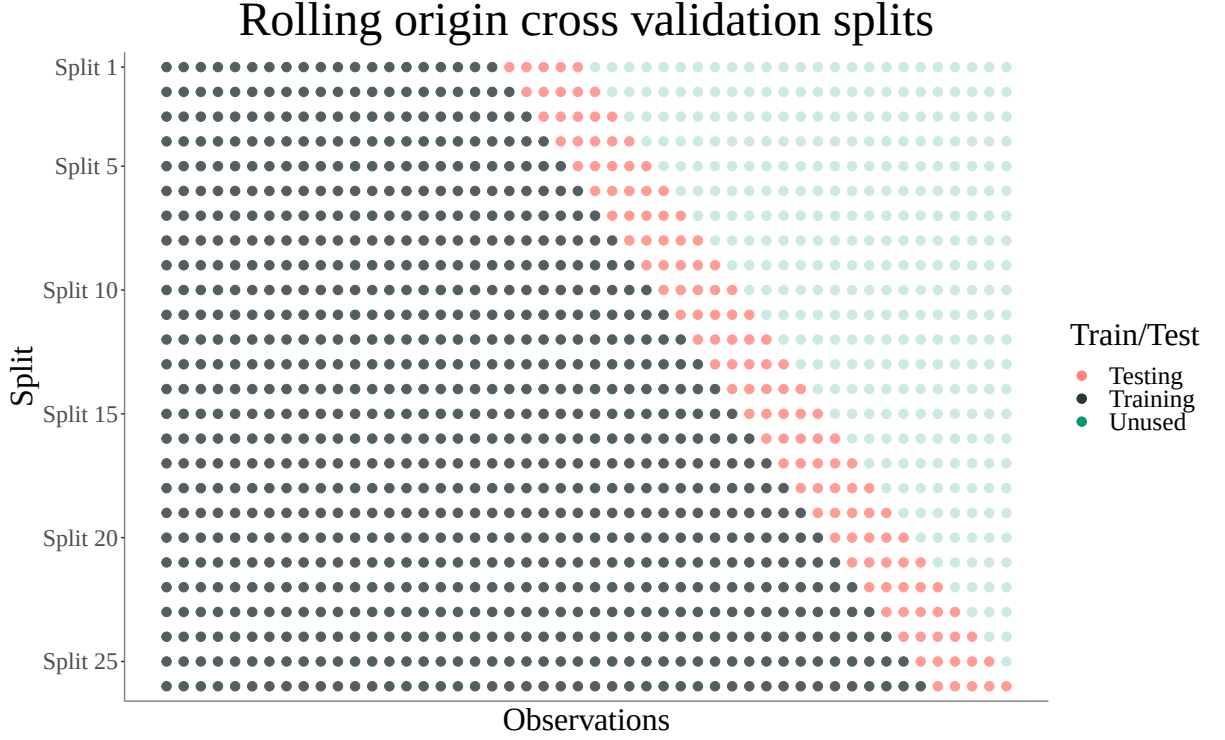


Figure 5.2: A schematic of the rolling origin cross-validation methodology on a data set with 100 observations. Each point represents a specific data point, wherein each of the splits the grey and pink points are training and testing data respectively. The fainter teal points are unused in the particular split

is carried out using a set of time series simulations.

First, a particular time series model is chosen, and simulations are generated from this model. Defining a particular observation as the *origin* (data point to be forecasted), we can fit a series of models using increasing amounts of historical data. Repeating this process a number of times to gauge average behaviour, we can then assess the validity of comparing models fit with different amounts of data.

For the purposes of illustration an AR(2) process is simulated, in particular, the process described by (B.2). Recall that in such a process, its value at each time point explicitly depends only on the values of the two previous observations. A collection of 1000 simulated sample paths are generated, each of length 500.

For each of the 1000 simulations; AR(2) models are fit, using different size training sets ($N = 20, 50, 100, 200, 500$), indexed backwards from the 500th observation. The resultant model fits are then evaluated, allowing for comparison across training set sizes. To assess the impact of training set size on model fit, the absolute errors of parameter estimates are plotted in Figure 5.3.

AR(2) Model accuracy vs training set size

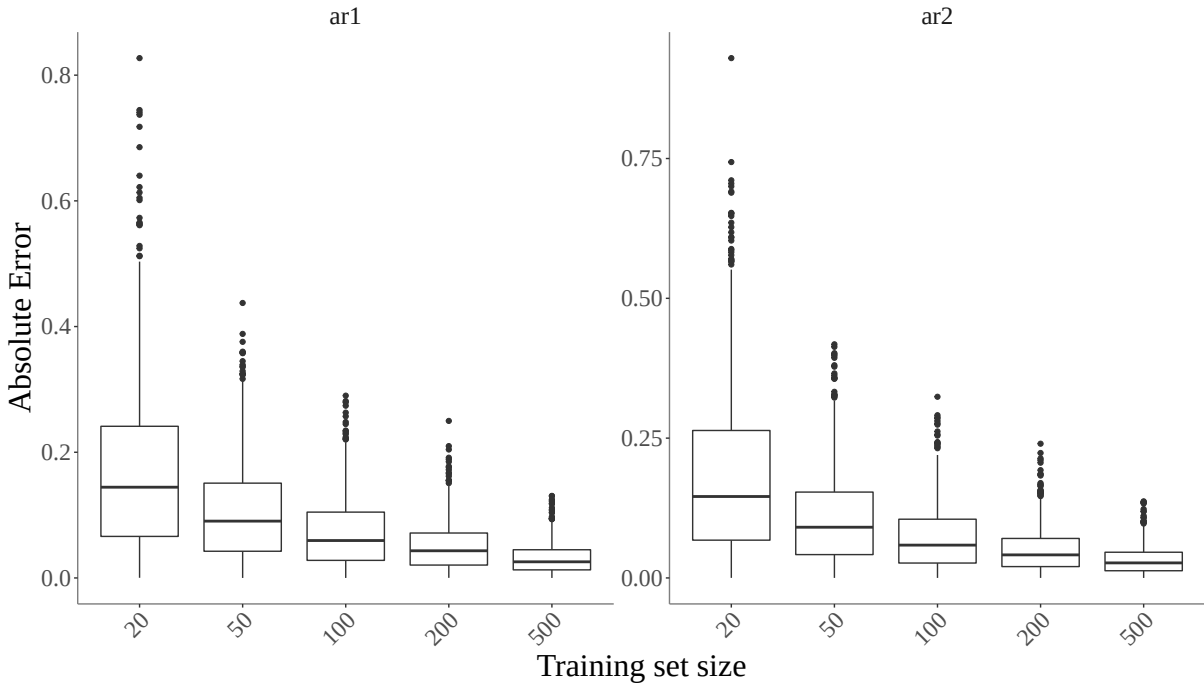


Figure 5.3: *Side-by-side box plots of parameter estimate errors. For each training set size, a box plot of all absolute parameter estimate errors has been generated. The left and right plots display the errors for the first and second autoregressive coefficients respectively. It can be seen in both cases that the fit is significantly improved for larger training set sizes.*

It can be seen in Figure 5.3, that an increase in training set size significantly reduces the error in estimates of both autoregressive coefficients. As a result of this, it is deduced that direct comparisons of models fit with different size training sets can potentially cause inconsistencies in the resultant conclusions.

Adapted ROCV

In order to protect against these possible inconsistencies, a new adaption is introduced to the ROCV method producing *Adapted Rolling Origin Cross Validation* (AROCV). To preserve training set size, rather than rolling forward only the origin (point of forecast) the initial point in the the training set is also rolled forward in a corresponding fashion as depicted in Figure 5.4. In this way, cross-validation can be carried out in a way that allows for comparison of results, and ensures that testing data is sufficiently representative of the training data.

This new protocol for model building and testing provides a more robust methodology

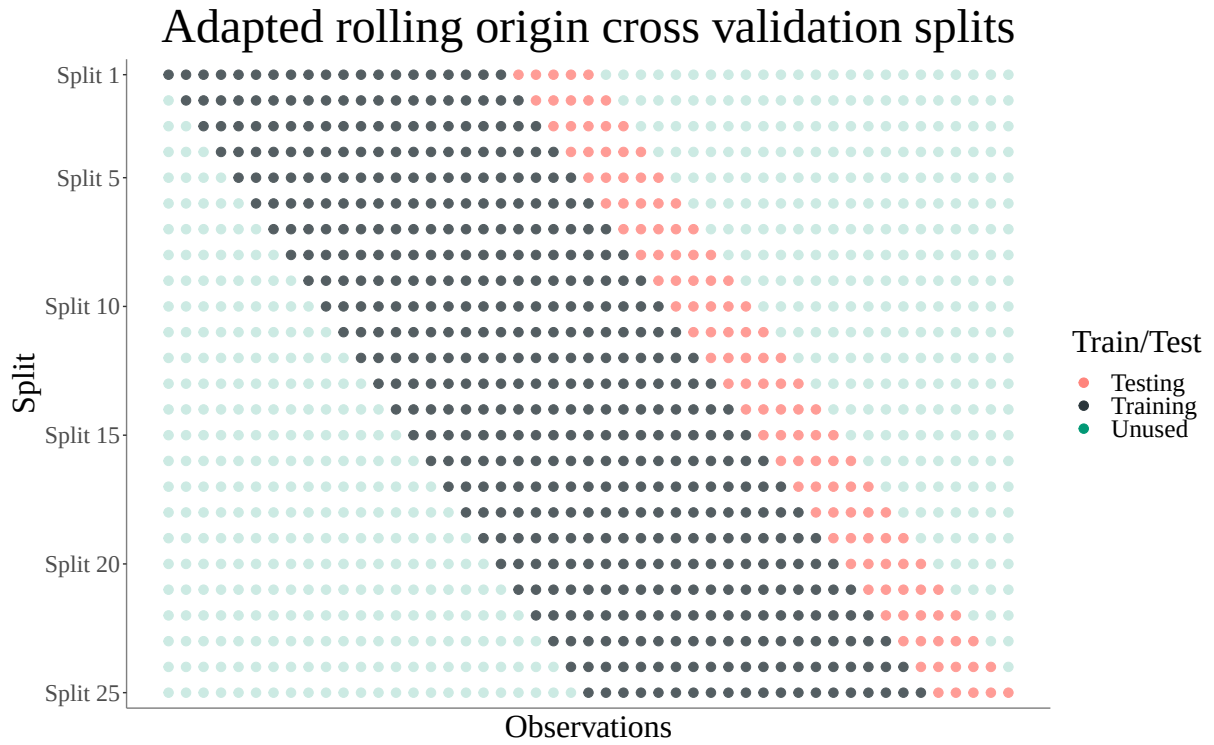


Figure 5.4: A schematic of the adapted rolling origin cross-validation methodology on a data set with 100 observations, depicted in the same way as Figure 5.2.

for model assessment in the context of time series models. In the sections that follow, the AROCV method will be used as the critical tool with which models will be assessed and compared for quality.

In this chapter, the foundational concepts of time series processes and models have been developed, introducing the family of ARIMA processes which will be used in the following sections to model ED observations. The characteristic properties of these models have been examined, leading to a discussion of an applicable model building process in the Hyndman-Kandakar Algorithm [75]. Finally, the methods with which time series models are assessed and compared for predictive ability have been examined for their suitability in context. This resulted in the introduction of the AROCV method for the comparison of model predictions. The combination of the Hyndman-Kandakar Algorithm and AROCV will be applied in the following chapter for the purposes of validating, identifying and assessing ARIMA models for the purposes of prediction with the data at hand.

Chapter 6

Predictive modelling

In this section, temporal modelling techniques developed in Chapter 5 are applied to the ED data and evaluated for their predictive ability. Initially, a set of prediction baselines are established for the data set, followed by the development of the predictive model framework as well as its application to the ED data. Finally, the results are compared to the established baselines, and the prototype platform built as part of this work is discussed.

6.1 Baseline establishment

The first step is to establish what will be referred to as capacity index *prediction baselines*, a set of benchmarks against which the efficacy of predictive models can be measured. To develop these benchmarks, a series of models which can be considered *naive* to some level will be applied to the data and analysed for their predictive utility, providing a baseline from which can be improved upon.

To analyse model quality, the *root-mean-square error* of a model across a range of prediction horizons will be considered. As has been established, the primary goal of this work is to build a model for short-term prediction of a hospital's capacity index Equation (4.1), specifically from one to five hours into the future. For this data set, which contains half hourly observations, this corresponds to prediction horizons of up to ten time steps into the future. As such, baselines will be developed that correspond to such prediction horizons.

The naive models being considered are detailed below. For notational purposes, the models are specified below assuming data has been observed up to and including time index t_0 , with a prediction horizon h half hour steps into the future. Noting that each of the hospitals are being modelled independently, indices to specify hospital have been dropped.

- Mean prediction model

As the name suggests, this model produces predictions equal to the sample mean of the capacity index for the hospital. In a dataset containing N observations, this is defined as:

$$\hat{I}_{t_0+h} = \frac{1}{N} \sum_{0 < t < t_0} I_t \quad \text{for all horizons } h.$$

- Last Observation Carried Forward (LOCF)

The LOCF (often referred to as a persistence model) method, is also self-describing. In a given hospital, for all prediction horizons h , this model forecasts that the capacity index will remain unchanged from the most recent observation:

$$\hat{I}_{t_0+h} = I_{t_0} \quad \text{for } h = 1, 2, \dots, 10.$$

This model is analogous to an AR(1) model, in which the autoregressive coefficient is equal to one. Importantly however, such a model would never result from fitting a model to data as the result would be considered non-stationary¹.

- Rolling four-hour mean

Including slightly more information, the rolling mean model generates predictions equal to the mean of the previous four hours (eight-time steps) of observed values:

$$\hat{I}_{t_0+h} = \frac{1}{8} \sum_{i=0}^7 I_{t_0-i} \quad \text{for } h = 1, 2, \dots, 10.$$

Compared to the LOCF model, the rolling mean model has increased robustness against anomalous observations but neglects to take into account any potential trend in the recent observations.

- Rolling four-hour weighted mean

Only subtly different from the rolling mean model, the rolling weighted mean biases more recent observations. This biased model is produced by weighting the coefficients in such a way that they decrease in magnitude in a linear fashion:

$$\hat{I}_{t_0+h} = \sum_{i=0}^7 \omega_i \times I_{t_0-i} \quad \text{for } h = 1, 2, \dots, 10.$$

¹The sole root of the characteristic polynomial (eq. (B.1)) in this case is $b = 1$. As the root lies on the boundary of the unit disc, the model would be non-stationary.

Where ω_i are the weighted coefficients. Assuming that the capacities I_t are identically distributed, and then producing an unbiased estimate of the mean capacity index results in weights given by $\omega_i = (i + 1)/36$.

To establish the prediction benchmarks, the four naive models were applied to the data collected from the six hospitals across the calendar year of 2021. The resultant forecasts of the naive models were calculated for prediction lags $h = 1, 2, \dots, 10$. Calculating the RMSE for each forecast, for each model in each hospital, their performance can be assessed with results displayed in Figure 6.1.

Across the hospitals, similar behaviour is exhibited from the naive models. Of the three models which predict dynamically (LOCF, and the two rolling mean models), it is clear that the LOCF model outperforms the rolling mean approaches regardless of the prediction horizon. As one may expect, each of these three models decrease in performance as the prediction horizon increases, to the point where the stagnant mean prediction (which is clearly the worst model for small lags) outperforms the dynamic models at larger lags.

In order to consolidate these results, the minimum RMSE across all naive methods at each lag for each hospital will be taken as the benchmark. These results ordered *most predictable* hospital to least are contained in Table 6.1.

Hospital	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$	$h = 9$	$h = 10$
RAH	0.06	0.07	0.09	0.10	0.12	0.13	0.14	0.15	0.16	0.16
MH	0.07	0.10	0.13	0.15	0.17	0.18	0.20	0.21	0.23	0.24
LMH	0.06	0.09	0.12	0.14	0.16	0.18	0.20	0.22	0.23	0.24
TQEH	0.09	0.13	0.15	0.18	0.20	0.22	0.24	0.26	0.26	0.26
NHS	0.11	0.16	0.20	0.24	0.27	0.30	0.33	0.36	0.36	0.36
WCHP	0.11	0.16	0.21	0.26	0.30	0.34	0.38	0.42	0.43	0.43

Table 6.1: *Benchmark RMSE values for each prediction horizon h , at each hospital. These are the values against which will be used to compare the predictive ability of models in future sections.*

The prediction benchmarks can be seen to follow a similar trend to the intrinsic variability observed in each hospital. Understanding that at large lags, the most effective naive model is the mean model; the RMSE of this model is simply the sample standard deviation of the capacity index from each hospital. Nevertheless, it is observed that the RAH appears to have the highest benchmark to improve upon (i.e. it is the most predictable), while the NHS and WCHP appear to be the least predictable.

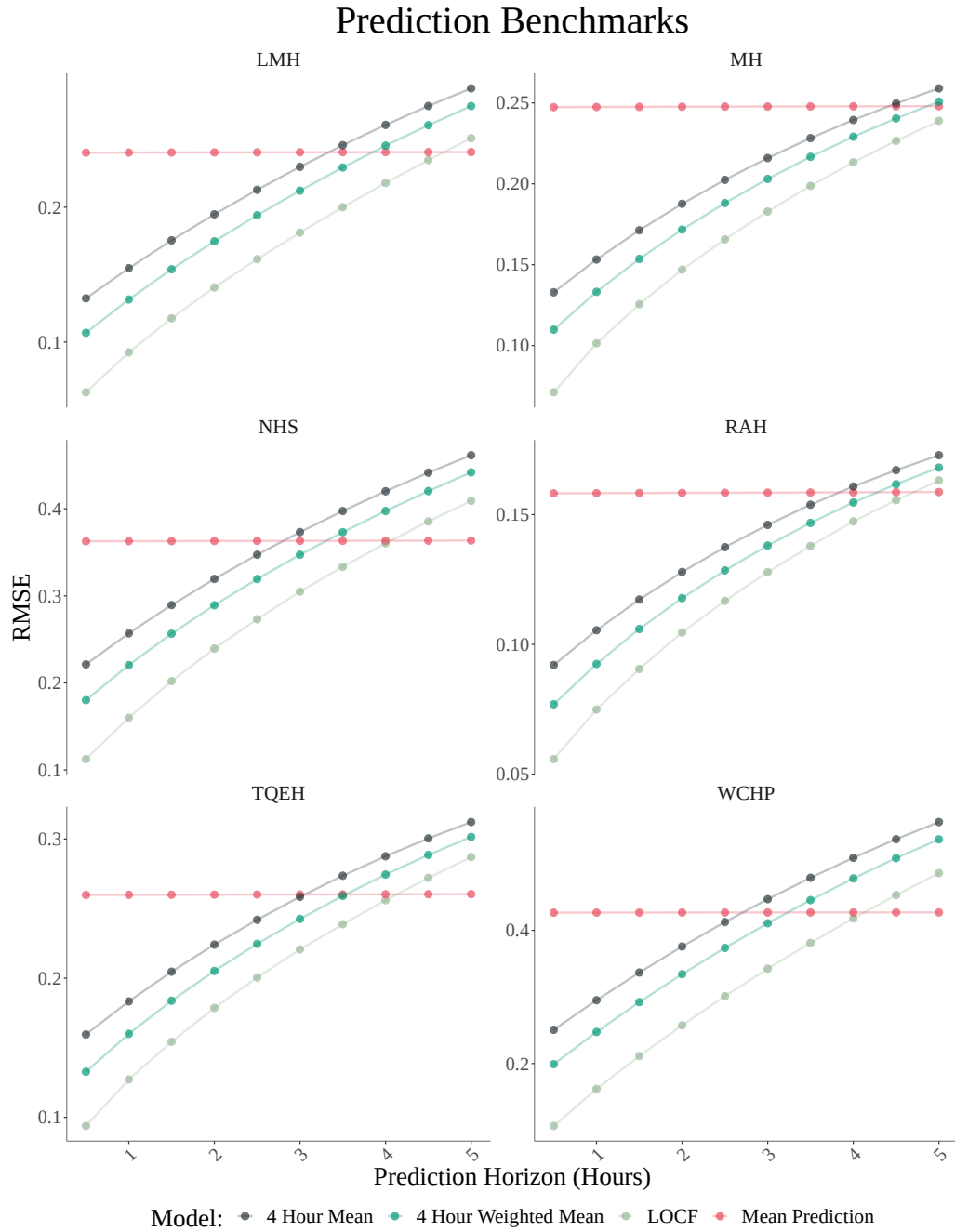


Figure 6.1: For each of the six hospitals, the RMSE prediction error from each naive model is plotted for the respective prediction horizons. The results from each of the hospitals are being considered independently, as such the scales of the vertical axes across the six plots vary.

6.2 Model development

With the tools that have been developed, attention is turned to the application of time series models in order to determine whether they can generate forecasts that improve upon the benchmarks established. The modelling approach followed is guided by the exploratory results in Section 4.3 and grounded in the fundamentals detailed in Section 5.2.2, by first establishing stationarity and then exploiting autodependencies. Recalling that the principle guiding heuristic is to minimise prediction error on test observations (orange observations in Figure 5.4), in this way, the final model will provide maximum utility in practice.

Crucial conclusions reached in Section 4.3, regarded the similarities and differences amongst the six EDs. Univariate analyses indicated that the shapes, locations and spreads, particularly of the *Capacity Index* varied significantly across the hospitals. For this reason, each hospital ED is modelled individually, producing six predictive models to suit the scales and variances of the six EDs. Contrastingly however, the bivariate analyses indicated that all hospitals display strikingly similar dynamics, particularly in their temporal patterns throughout each day and week. Noting this, and recalling the priority of developing a generic modelling framework; it should be noted that while the models will be for individual EDs, the model framework will be consistent such that the methodology is transportable across EDs within and beyond the data set.

6.2.1 Identifying deterministic trends

It was shown in Section 4.3, that there are clear temporal patterns in the variation of the capacity index on daily and weekly time scales. Such patterns suggest that the quantities being predicted have time-dependent characteristics and are naturally non-stationary.

Investigation of these temporal patterns allows for the removal of deterministic trends in the data and hence reduction of non-stationary behaviours, as well as providing a model that can become the next baseline model that would be considered deterministic to some degree. To do so, a model consisting of a deterministic, and a stochastic part can be developed:

$$I_t = \left\{ \begin{array}{c} \textit{Deterministic} \\ \textit{Portion} \end{array} \right\} + \left\{ \begin{array}{c} \textit{Stochastic} \\ \textit{Portion} \end{array} \right\} \quad (6.1)$$

To model the deterministic portion, a linear model is constructed for the capacity index in which the predictor variables are exclusively the :

- Hour of the day,
- Day of the week and
- A first-order interaction between the day and hour.

F Statistics				
Hospital	Included Predictors			
	Day and Hour	Hour only	Day only	
LMH	3.28	9.09	58.77	
MH	2.74	4.62	30.07	
NHS	5.02	10.18	82.80	
RAH	2.10	7.19	24.82	
TQEH	4.36	10.97	41.43	
WCHP	2.19	3.70	102.50	

Table 6.2: Resultant statistics from F -tests comparing the model Equation (6.2) to simpler counterparts for each hospital. The first column of results regards comparisons with models containing both day and hour (no interaction), while the second and third regard comparisons with one factor models containing only hour and day respectively.

As this model uniquely identifies each of the 168 hours in the week it is equivalent to a one-factor model which has the 168 necessary levels. For notational purposes, define a map $\mathcal{H}$ which simply transforms any time index t to the integer reflecting the necessary hour in the week, allowing the definition of the model as:

$$I_t = \mu_{\mathcal{H}(t)} + X_t, \quad (6.2)$$

where X_t represents the error of the model, and is hence the stochastic part which will be analysed in the following sections.

From linear model theory, it is known that the estimates of this model will be the sample mean of all training observations corresponding to the level, hence the nomenclature $\mu_{\mathcal{H}(t)}$. In this way, the removal of deterministic trends via this model allows for the investigation of autodependencies in the resultant values of x_t .

This model was fitted to all available data, using the `stats` package [76] in order to assess its suitability. Comparisons were made with simpler models, in which the interaction as well as each of the *day of week* and *hour of day* variables were removed. These comparisons, in the form F -tests, validated the additional complexity, with results contained in Table 6.2. It is clear from the results that the additional complexity provides improved explanatory power.

Assumption Checking

In order to ensure the validity of this model, the four common linear model assumptions must be validated. Firstly, it is noted that the assumption of independence between observations is invalid for the data being considered, in fact the key goal of this chapter is to further investigate this dependence. Having said this, studying the assumptions

Hospital	Mean Model	Linear Model
RAH	0.16	0.14
LMH	0.24	0.18
MH	0.24	0.21
TQEH	0.26	0.21
NHS	0.36	0.25
WCHP	0.42	0.27

Table 6.3: *A table containing the RMSE errors of both the original mean model, and the linear hour of week model. Results indicate the linear model can be used to describe the deterministic behaviours on a weekly time scale.*

regarding the structure and distribution of the model's errors x_t ensures that further investigation is well grounded.

For purposes of illustration, Figure 6.2 contains three diagnostic plots to assess the assumptions of the linear model when applied to the LMH data. All other EDs exhibited corresponding behaviour, and hence for the sake of brevity have been excluded. The assumption of linearity for a one-factor model is, by construction inherently satisfied as can be seen in the top panel of Figure 6.2. The normality of the residuals is largely satisfied, as can be seen in the second panel the small number of residuals that have contrasting quantiles appear in the tails. The first and last panels both evidence the homoscedasticity present in the residuals. Across the range of fitted values and predictor levels, the variance remains approximately constant.

Results

To assess the utility of modelling deterministic behaviour through a linear model, the resulting RMSEs are considered. The natural benchmarks to compare against are the RMSEs from the naive mean model (equivalent to the sample standard deviation of the original data). These results are contained in Table 6.2, which shows that for all hospitals, a significant amount of the variability is removed by applying the deterministic linear model, allowing the remaining stochastic variability to be modelled in the following sections. It is acknowledged that this approach could be further investigated in future work, potentially with more complex and dynamical models that build in longer scale, multiple-seasonality dependencies which cannot be elicited with the current data set. Such models are detailed in Kolassa et al. [77].

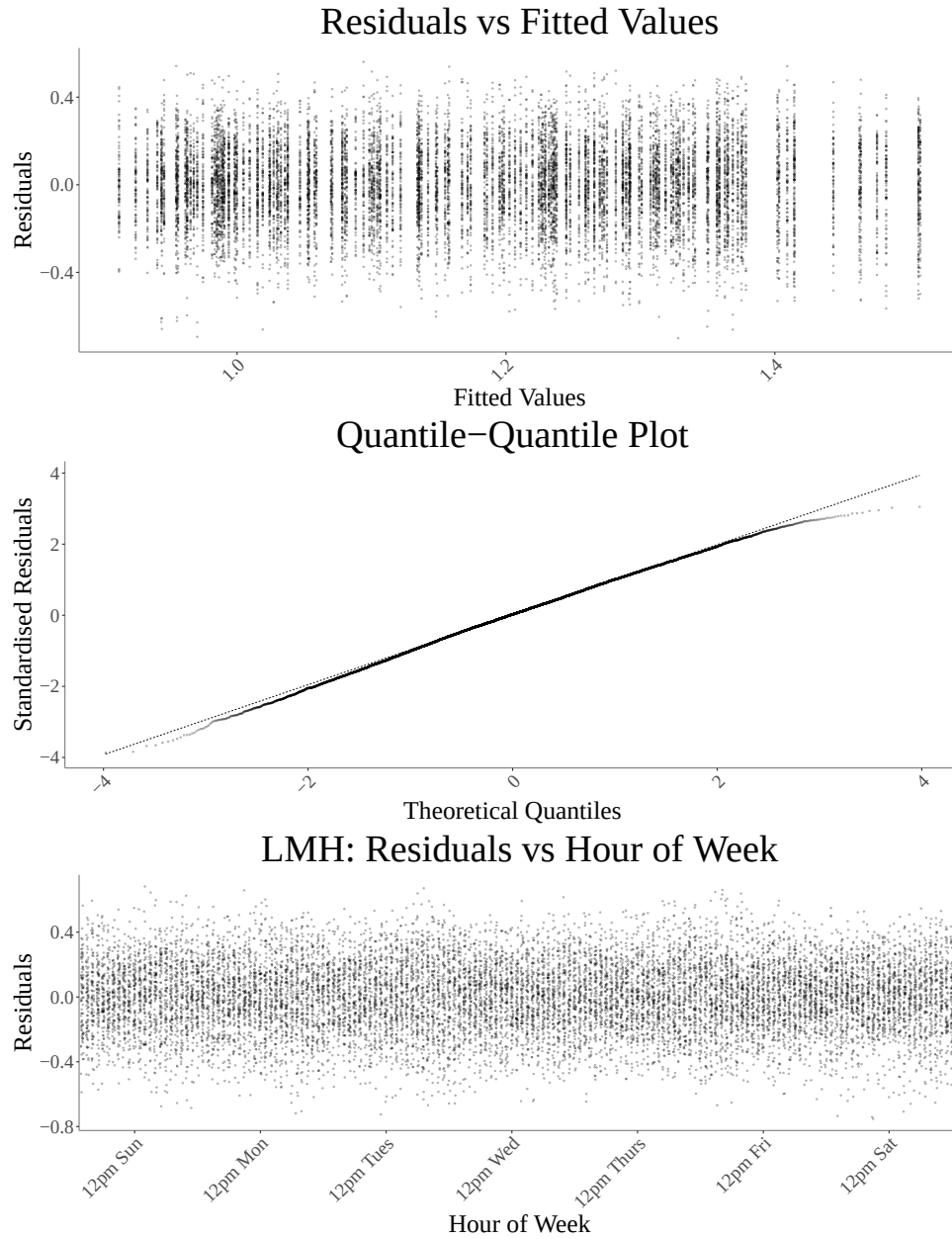


Figure 6.2: Diagnostic plots of the residual errors (realisations of X_t in Equation (6.2)), resulting from fitting to the LMH data. The top frame, containing the residuals versus fitted values plot indicates that roughly constant variance is observed across the range of values. The second frame shows that the body of the distribution follows a normal distribution almost exactly with some deviations in both tails. In the final frame very similar distributions of errors for each level of the predictor are observed, indicating that there is no weekly variation in the residuals.

6.2.2 Modelling stochastic behaviour

Having used linear models to extract the deterministic behaviour from data, the stochastic behaviour can now be modelled. In particular, the time series tools developed earlier will be applied to the observations X_t (6.2) in order to understand and exploit the autocovariance structures for the purpose of prediction.

The process to develop the stochastic models largely consists of two phases. First, is an *Exploratory* phase in which a subset of data is used to gauge the types of time series models that are appropriate to model the autocovariance observed. This is followed by a *Build and Test* phase in which the exploratory results are used to develop an array of models, which can be *built and tested* and to examine their predictive ability.

Exploratory phase

To understand the class of time series models that can effectively model the stochastic behaviour, a subset of the data was used to carry out some exploratory model building, namely from January of 2021 (1448 observations per hospital). The initial step in this process was to remove deterministic trends via the linear methods detailed above, leaving the observations X_t (6.2) for each hospital. The process described by Algorithm 1 was carried out on these observations using methods contained in the `fable` package [78]. In this way, a set of models that accurately describe the behaviours in the observed data can be validated and assessed. The most successful models and the information criteria resulting from their fit to the January data can be found in Table 6.4.

The models contained in the table are indicative of models that can successfully describe the dynamics observed. As such, this set of models can be used to guide the potentially applicable models in the next stage. Combining the methodology of Hyndman and Khandakar [75] with Box and Jenkins [70], the ACF and PACF of the residuals from these fits were also considered.

Hospital	Model	AICc	BIC
LMH	ARMA(2,2)	-4114.40	-4082.62
MH	ARMA(2,2)	-3929.70	-3892.64
NHS	AR(2)	-2393.88	-2377.98
RAH	ARMA(3,1)	-4627.11	-4590.05
TQEH	ARMA(2,1)	-3240.58	-3219.39
WCHP	ARIMA(3,1,1)	-3035.80	-3004.03

Table 6.4: Results from initial ARIMA modelling, this table contains the resultant AICc, and BIC scores from the most successful models resulting from the Hyndman-Khandakar Algorithm 1.

the subsequent trials doubling the training set size to a maximum of 16 weeks (5376 observations).

To understand the impact of training set size on prediction performance, the RMSE across all models was calculated at each lag, and each for each Hospital by applying the AROCV method with each of the four training set sizes (two, four, eight, sixteen weeks). The results observed across the six hospitals exhibit very similar behaviours, as such results from the LMH will be used to illustrate the observations. Figure 6.3 depicts the changes in prediction performance as the size of the training set increases. It is clear that as the training set size increases, the prediction performance is improved, and the magnitude of improvements remain reasonably consistent across different lags. However, recalling that the size of training sets doubled between the samples depicted, it is clearly a case of diminishing returns beyond a training set size of eight weeks, with an increase to sixteen weeks providing only a minor improvement. With this in mind, AROCV results presented from this point forward will be only from models built with eight-week training sets.

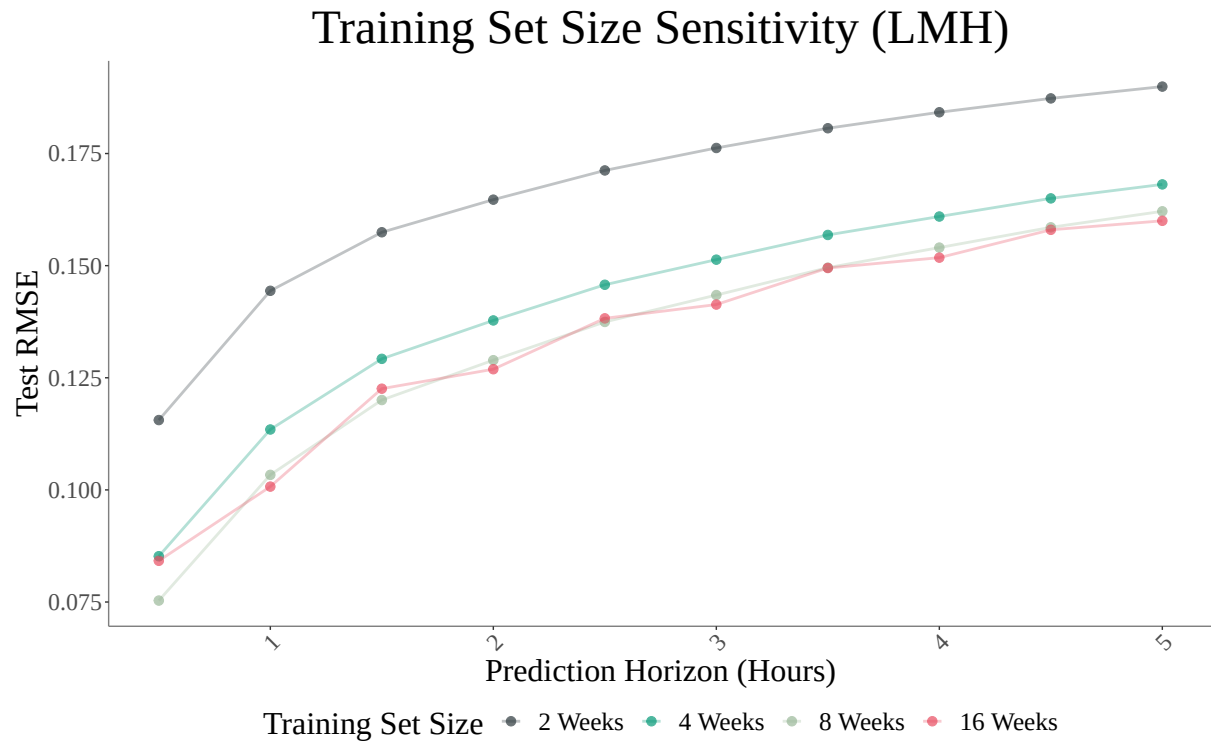


Figure 6.3: For all LMH observations outside of the exploratory set, the test RMSE across the models and splits was calculated for each training set size at each lag, producing the plot above.

6.3 Results

The assessment of results for this chapter is comprised of two parts, first the discernment of the most successful models for each hospital, followed by the assessment of their utility against the benchmarks previously set (Table 6.1).

Model assessment

In order to identify the most effective predictive model for each hospital, each of the candidate models were applied using the AROCV approach, with training sets corresponding to eight weeks worth of observations. The metrics which the results would be compared were initially centred on predictive power at a five-hour horizon. As will be observed in the results that follow, several of the candidate models have effectively equivalent predictive power at the stated horizon. As such, the mean AICc and BIC indices were decided to be relevant tools to compare models that have equivalent predictive power, allowing models achieving better *goodness-of-fit* to be discerned.

Noting that similar model performance behaviours were observed across all hospitals, the process and details of model comparison will be discussed in more detail for the LMH results, while conclusions will be stated for the remaining hospitals, with supporting evidence contained in Appendix A.

To assess predictive power at the five-hour horizon, the RMSE of the five-hour predictions were calculated for each model applied using the AROCV process. The results are contained in Table 6.5. With eleven of the twelve models yielding *five-hour-test-RMSE* results lying in the interval (0.1655, 0.1688), it is clear that there is limited utility in using the test RMSE alone to choose a superior model. Consulting the mean AICc (effectively equivalent to AIC for training sets spanning 8 weeks of data) and BIC scores provide some extra insight, showing that in the case of the LMH, the *ARMA*(2, 1) model achieves by far the best mean BIC score, as well as a competitive mean AICc score and test RMSE.

Across the six hospitals, multiple candidate models were the most successful. The results of the most successful models for the six hospitals are contained in Table 6.6. As with the LMH results, models were chosen first, if their predictive power outperformed the other models; failing that, the AICc and BIC scores were used. As indicated in the table, the model chosen most often was the simplest of the successful models, the *ARMA*(1, 1) model, and the most complex was the *ARIMA*(3, 1, 2).

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
ARMA(2,1)	0.1042	0.1302	0.1466	0.1579	0.1663	-7480.7762	-7457.2049
ARMA(2,2)	0.1043	0.1302	0.1465	0.1579	0.1663	-7480.7927	-7451.3288
ARMA(3,1)	0.1042	0.1302	0.1465	0.1579	0.1663	-7480.7340	-7451.2722
ARMA(3,2)	0.1044	0.1303	0.1466	0.1580	0.1664	-7481.8442	-7446.4963
ARMA(2,3)	0.1043	0.1302	0.1466	0.1580	0.1664	-7480.1528	-7444.8041
ARIMA(3,1,2)	0.1034	0.1291	0.1455	0.1569	0.1655	-7474.0358	-7438.6900
ARMA(1,3)	0.1049	0.1313	0.1472	0.1584	0.1667	-7458.8213	-7429.3609
ARMA(1,2)	0.1053	0.1312	0.1477	0.1592	0.1677	-7449.7757	-7426.2044
AR(2)	0.1048	0.1302	0.1463	0.1575	0.1658	-7440.5983	-7422.9176
ARMA(1,1)	0.1068	0.1309	0.1465	0.1576	0.1658	-7423.7395	-7406.0588
ARIMA(3,1,1)	0.1025	0.1288	0.1462	0.1589	0.1688	-7433.1535	-7403.6950
AR(1)	0.1072	0.1351	0.1522	0.1638	0.1721	-7346.0530	-7334.2643

Table 6.5: *LMH time series results: The Test Root Mean Square Error, at each of the prediction horizons resulting from the application of the named ARIMA models using the build and test process with the AROCV methodology. Rows have been sorted by ascending BIC scores.*

Hospital	Model	5 Hr Test RMSE	Mean AICc	Mean BIC
LMH	ARMA(2,1)	0.1663	-7480.7	-7457.2
MH	ARIMA(3,1,1)	0.1903	-6698.1	-6668.6
NHS	ARMA(1,1)	0.2535	-4386.6	-4368.9
RAH	ARMA(1,1)	0.1327	-8051.8	-8034.1
TQEH	ARMA(1,1)	0.2155	-5255.5	-5237.9
WCHP	ARIMA(3,1,2)	0.2689	-4816.1	-4780.8

Table 6.6: *Successful time series models: This table details the most successful model across each of the six hospitals, along with the metrics upon which they were assessed.*

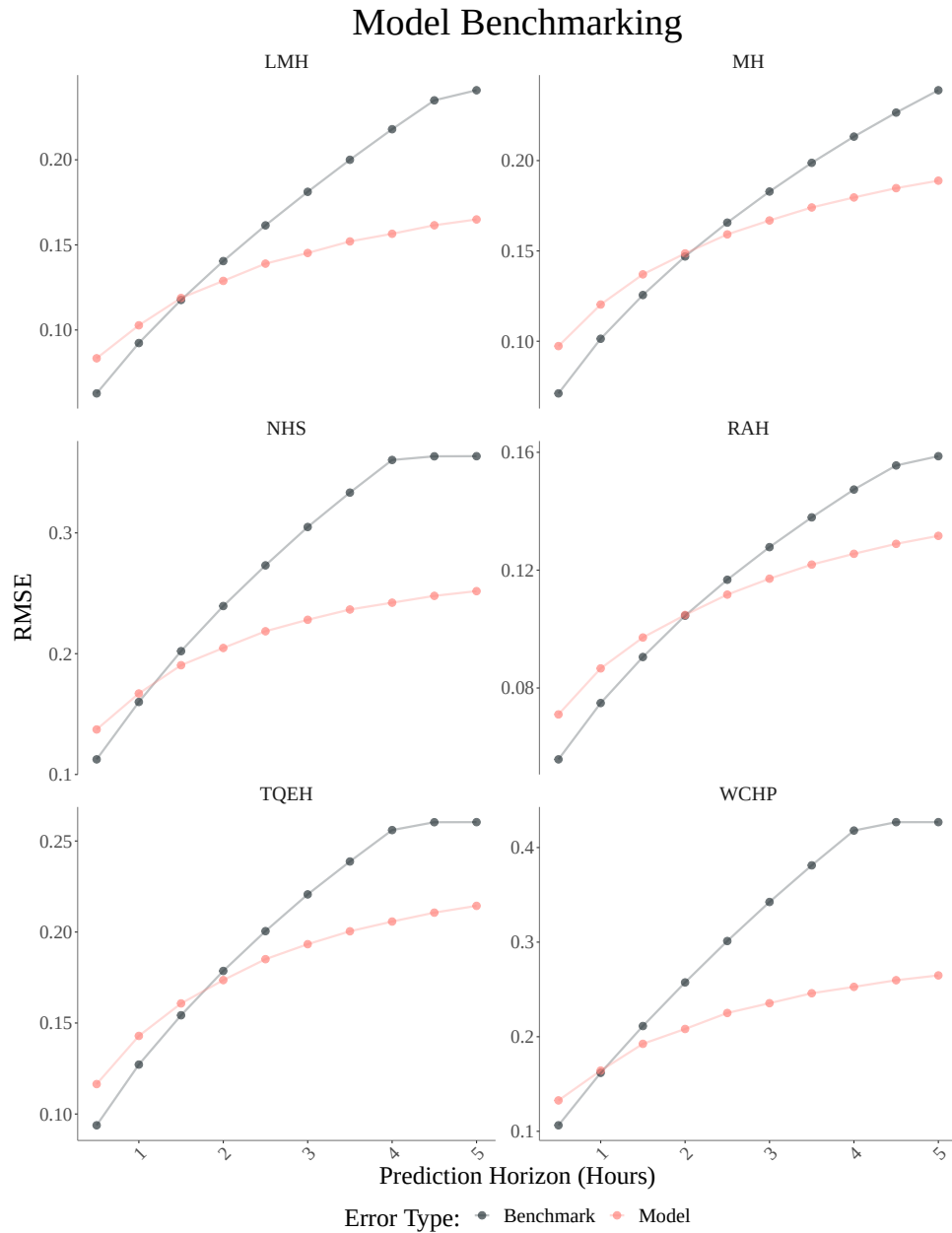


Figure 6.4: For each hospital and at each prediction horizon, the test RMSE was calculated for the model deemed optimal in each of the hospitals. The model test RMSE values are plotted in orange, in contrast to the prediction benchmarks plotted in grey. Note the vertical axes are independently scaled, allowing the different behaviours across hospitals to be exhibited.

Predictive power

Turning attention to the utility of these models, the predictive power of these models can be compared to the benchmarked models detailed earlier. In particular, direct comparison of the models using the test RMSEs at the various prediction horizons between 0.5 and 5 hours. These test RMSE values for both the benchmarks and the chosen models are plotted in Figure 6.4 for each of the six hospitals. One can make a number of clear conclusions from the plots;

- A) As expected, *predictability* drops off as the horizon increases,
- B) For greater prediction horizons (> 2.5 hours) time series models consistently outperform benchmarked models,
- C) While outperformance is consistently observed at the five-hour horizon, the level of improvement varies across hospitals. As a rule of thumb, the level of improvement is correlated to the initial variability,
- D) Particularly for short prediction horizons, the naive models consistently outperform time series models.

6.4 Prediction platform

Carrying on from the results of the predictive models, attention is turned to a proof of concept for model implementation. The goal of this work has been to develop a predictive modelling framework to enable tactical decision support. Following the development of the modelling framework, it was acknowledged that the framework and models in isolation provide limited ability to make any concrete impact on emergency department overcrowding. In fact, it was noted that if an impact were to be made, a tool would have to be in the hands of individuals with limited to no requisite understanding of statistics or modelling. These factors encouraged the development of a web-based application that provides a platform for users to interact with the results of this work. The application can be accessed via this [link](#)²

The prototypical application was developed with the goal of allowing a user to interact and experiment with empirical data, as well as collect and make predictions from real-time data. The resultant application consists of the following four pages with corresponding capabilities:

- **Static Visualisations:** Enabling exploratory data analysis of data spanning a six-month window (limited due to the lightweight nature of the application). Users can investigate distributional plots, as well as plots of notable daily and weekly trends.

²The platform is accessible via the address:

<https://datascience-adelaideuniversity.shinyapps.io/EmergencyDepartmentPrediction/>

The variables considered include the Capacity index, the number of patients in and waiting for treatment, and the number of expected arrivals. An example is shown in Figure 6.5.

- **Dynamic Visualisations:** Allowing users to visualise the distribution of observations in a sliding window, exhibiting the dynamic nature of the distributions in time.
- **Live Data:** Providing users the ability to scrape data live, and observe the current state of the seven public hospitals for which data is available.
- **Prediction:** The most crucial of the pages, users are able to generate short-term predictions to inform decision making. At any given time, the back end of the platform holds the two most recent weeks of data for each hospital, allowing models to be rapidly built, assessed and then used for prediction. This page is shown in Figure 6.6.

In its current state, this platform is a functional proof of concept. With further investment into the concept, a more thoughtful and robust architecture is possible allowing more intuitive and robust insights to be generated. It is recommended that any efforts to progress the findings of this work are motivated by user-centred design, and prioritise further development of a tool for practical implementation.

Throughout this chapter, after establishing a foundation of time series modelling techniques, the AROCV method for developing and testing time series models was motivated and introduced. Following the development of this technique, a series of benchmark models were developed, allowing a baseline of ‘predictability’ to be established in each of the hospital’s data. Finally, a stepwise modelling process was carried out to identify and extract deterministic and stochastic behaviour patterns, establishing models which were conclusion provide improved predictive power over established baselines at larger prediction horizons.

Static Visualisation

The tool below allows you to investigate characteristic features of historic demand data for one South Australian hospital across the first six months of 2021.

Patients Waiting for Treatment

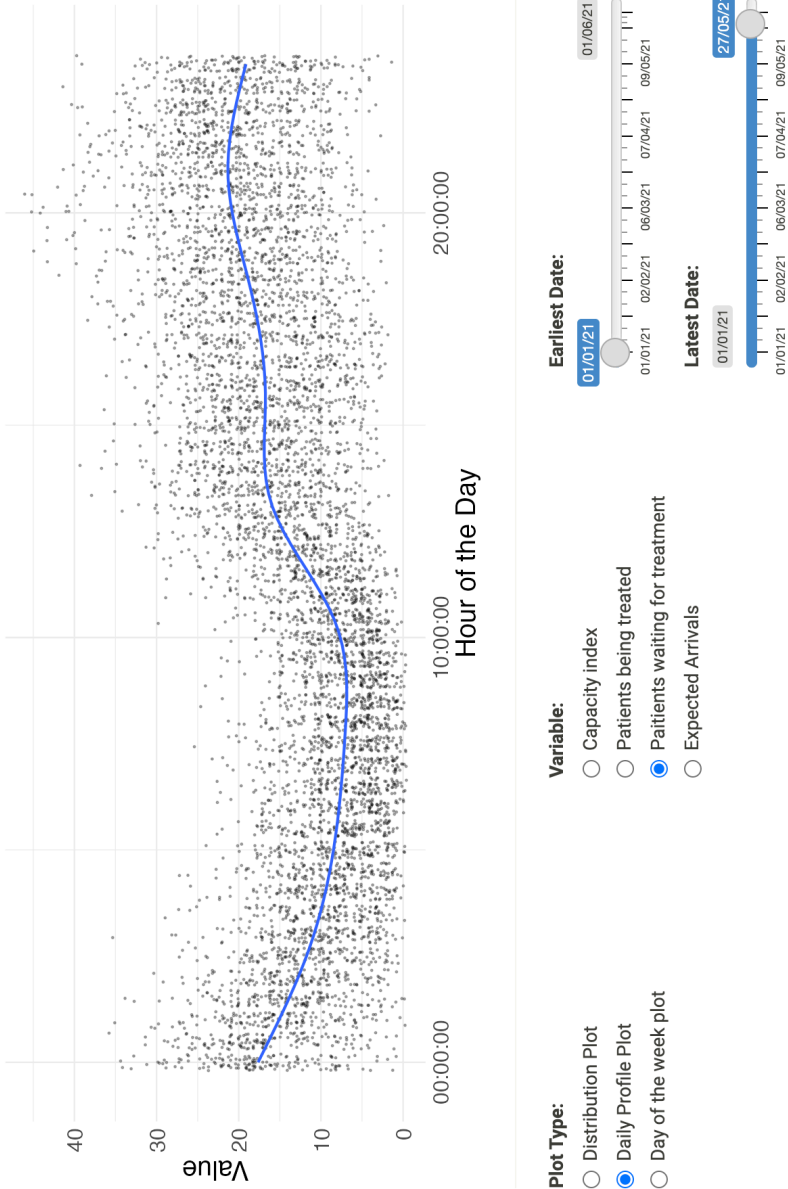


Figure 6.5: A screenshot of the prototype application, showing one of the data visualisation pages. As specified by the selection of radio buttons and sliders, the plot depicts the daily trend in the number of patients waiting for treatment.

Build and Assess models:

By clicking the button below, a set of models will be built on a live data set consisting of the most recent two weeks of observations. Results of the five models will be displayed in a table below.

ASSESS MODELS

Show 10 entries

Search:

		RMS Error	Log Likelihood	AIC	BIC
1	Deterministic Model	0.0129	245.47	-488.93	-485.15
2	Deterministic & AR(1)	0.0032	470.84	-936.67	-929.11
3	Deterministic & AR(2)	0.0032	470.53	-935.06	-923.72
4	Deterministic & MA(1)	0.0066	354.54	-705.09	-697.52
5	Deterministic & MA(2)	0.0044	420.71	-835.42	-824.08
6	Deterministic & ARMA(1,1)	0.0032	470.5	-935	-923.66
7	Deterministic & ARMA(2,1)	0.0032	472.77	-937.53	-922.41

Showing 1 to 7 of 7 entries

Previous

1

Next

Prediction

In order to produce predictions, first select a model by clicking a row in the results table above, and then click the *Generate Prediction* button below:

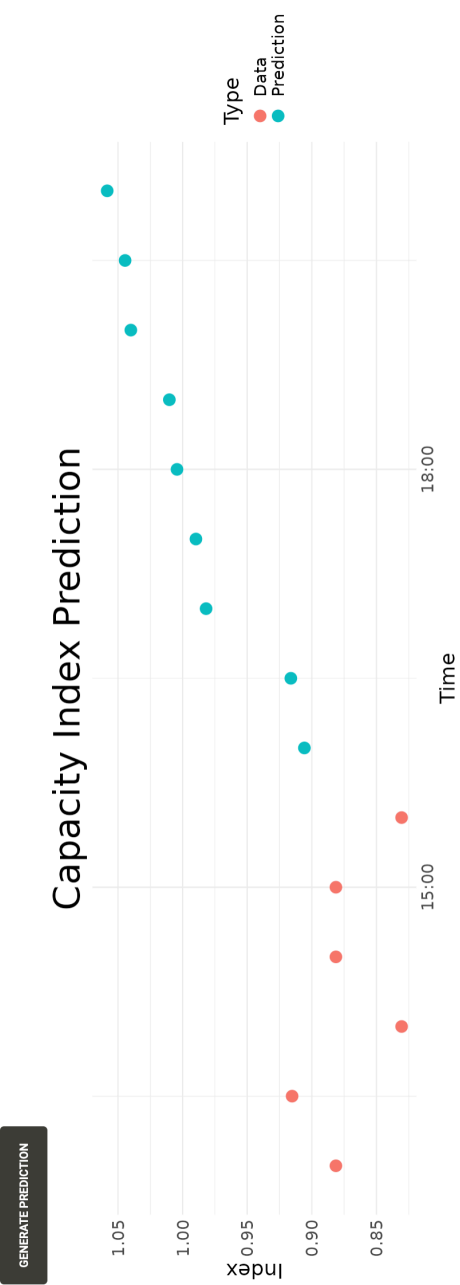


Figure 6.6: A screenshot of the prototype application, indicating its ability to fit and assess models, along with the ability to generate predictions.

Chapter 7

Deep learning for time series

In the previous chapter, the *ARIMA* framework was used to produce forecasting models. In order to further benchmark the results of the *ARIMA* models and validate the use of this framework, this chapter will investigate the use of a neural network technique applied to the same data.

7.1 Autoregressive neural networks

Neural Networks

As evidenced by Schmidhuber [79] Neural networks have been the focus of a vast amount of research both for the purpose of further theoretical development and for use in prediction problems. A technical discussion of Neural Networks is not considered to be in the scope of this work, but rather the application of a prediction scheme through an existing package implementation. For the interested reader, a detailed discussion and theoretical construction of neural networks can be found in Aggarwal [80].

In a general sense, neural networks can be considered as a construction that allows for the approximation of functions through the repeated composition of linear combinations with scalar modifications. In a multilayer feed-forward model, this process of composition is mediated via the *layers* of nodes as will be shown below. The scope of this work is limited to networks with a single hidden layer, a diagram of which is shown in Figure 7.1. The visual shows the layered nature of the network via colour, with data flowing from left to right taking a vector of n input observations $\mathbf{x}$, and transforming them to a single scalar output y , via a hidden layer of m nodes denoted $\mathbf{h}$. The operations that transform the data between each layer consist of both a linear combination, where weights and biases are contained in matrices W_l and vectors $\mathbf{b}_l$ for each layer l , and a non-linear modification function denoted Φ as per the following equations :

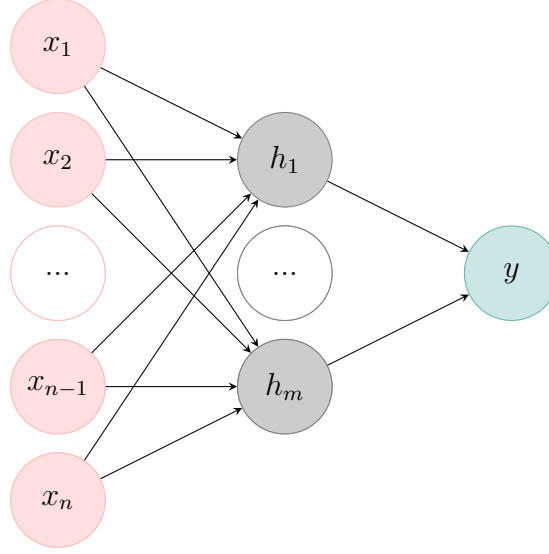


Figure 7.1: A diagram of a multilayer neural network with one hidden layer, transforming an input x_i ($i = 1, \dots, n$) to an output y . The network consists of input nodes (coloured pink), m nodes in a single hidden layer (coloured grey), and a single output node coloured green. Black edges joining nodes in consecutive layers represent the computations occurring in the feed-forward structure.

$$\mathbf{x} \rightarrow \mathbf{h} :$$

$$\mathbf{h} = \Phi(\mathbf{b}_1 + W_1^T \mathbf{x})$$

$$\mathbf{h} \rightarrow y :$$

$$y = \Phi(b_2 + W_2^T \mathbf{h}),$$

where Φ applies an element-wise scalar modification. In this work, the modification function will be the *sigmoid* function :

$$\Phi(s) = \frac{1}{1 + e^{-s}},$$

which has the property of reducing the impact of more extreme outliers by giving greater weight to more central observations. Therefore for a network consisting of n input nodes, m hidden nodes and a single output node, the matrices W_1 and W_2 will be of dimensions $n \times m$ and $m \times 1$ respectively. The elements of the parameters W_l and b_l form the parameter set which needs to be *learned* algorithmically.

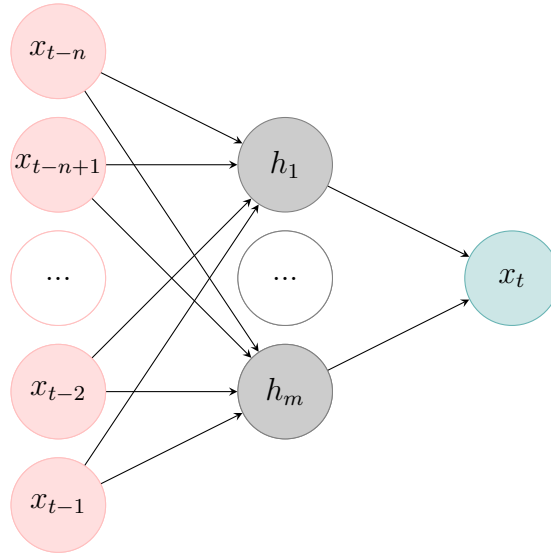


Figure 7.2: A diagram of a neural network constructed for time series prediction. The distinction to be made with Figure 7.1 is that the input layer of the model consists of n lagged values of the response variable.

Time series construction

In the context of time series the construction of the model is very similar, apart from the distinction between input and output observations becoming less clear. As with the ARIMA models previously detailed, the inputs to the model are lagged values of the time series observations rather than a distinct set of covariates. In this way, a neural network constructed to predict an observation at time t , denoted x_t would have the structure shown in Figure 7.2, where the input layer is now made up of the lagged values of the response variable x .

7.2 Modelling building

Comparison against ARIMA models was a key objective of this work, hence the implementation of the neural network structure sought to leverage the analysis framework previously developed, and existing methods contained in the **fable** package. As such, the model building process mirrored that of the analysis in Chapter 6, swapping the ARIMA model for a neural network.

That is to say, neural networks were built under the AROCV framework (section 5.3) using the following process:

1. Isolate training/testing observations for a given split,

2. Apply the deterministic model (6.2):

$$I_t = \mu_{\mathcal{H}(t)} + X_t,$$

to the training data:

- Calculating fitted estimates $\hat{\mu}_{\mathcal{H}(t)}$
 - Calculating errors x_t
3. Using **fable** methods, fit Neural Network models to errors x_t
 4. Generate test predictions: $\hat{I}_t = \hat{\mu}_{\mathcal{H}(t)} + \hat{x}_t$.

The **fable** methods referred to above are implemented in a similar fashion to the ARIMA models previously detailed, instead using the **NNETAR** model fitting routine. The **NNETAR** function contains methods to determine the network architecture; that is, the number of input and hidden layer nodes (n , and m from fig. 7.2 respectively). The model's AIC is used as a heuristic to optimise the number of input nodes, and the number of hidden nodes is calculated as half the number of input nodes plus one.

Optimisation of the parameters contained in the matrices W_1 and W_2 is carried out using the MSE as the cost function. As with the ARIMA models, the out-of-the-box implementation is built to optimise one-step predictions, with multi-step forecasts calculated recursively. The potential risk of arriving at local extrema is avoided by initialising each model with 20 independent neural networks, initialised with randomly sampled parameters. Further details of the **NNETAR** model can be found in resources from Hyndman and Athanasopoulos [71] and O'Hara-Wild et al. [78].

7.3 Implementation: results and performance

It was decided that implementation of the neural network model via the AROCV method would be carried out on one hospital for experimental purposes, with the intent of validating its application and gauging its performance on the ED data. As such, given that the LMH has been used for illustrative purposes throughout this work, the model was applied to data from the LMH for these purposes.

Results of the implementation are contained in table 7.1, with a comparison against the most successful model developed in Chapter 6. At each of the five horizons, the neural network is seen to have a marginally improved test RMSE over the $ARMA(2,1)$, with the improvements reducing as the horizon increases.

While minor prediction improvements were observed, the model was found to have significant drawbacks in fitting efficiency. Use of the **NNETAR** model for only the LMH data took on the order of two weeks to generate a predictive model, compared with ARIMA models (trained on the same data) took a number of minutes to generate.

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs
ARMA(2,1)	0.104	0.130	0.147	0.158	0.166
NNAR	0.094	0.124	0.143	0.155	0.164

Table 7.1: *LMH time series results: The Test Root Mean Square Error, at each of the first five hour prediction horizons from both the previously detailed ARMA(2,1) model and from the application of the NNAR model under the AROCV methodology.*

Marginal improvements are observed using the neural network model, hence validating the potential for the use of neural network techniques, however it was decided that the significant computational requirement of the model is prohibitive to further investigation. This work does however provide a proof of concept for future research further investigating optimised network architectures for the problem, not to mention more efficient implementation schemes under the AROCV framework.

Chapter 8

Conclusion

This work developed a generalisable framework that produces short-term forecasts of emergency department demand. This provides tactical decision support which contributes to the effort addressing ED overcrowding. In the process of achieving this objective, a comprehensive data set containing observations from seven South Australian public hospital EDs was collected and collated; allowing an investigation of dependency structures and patterns found within and across the hospitals. This data analysis formed the basis for the development of a hierarchical modelling framework which was used to demonstrate a predictive capability surpassing that of standardised benchmarks.

Result Summary

The first of the key contributions of this work is the collection and collation of a comprehensive and generic data set for multiple hospitals. The data set spans seven hospitals, over 2 years which not only provides a unique window into the public emergency departments in Adelaide, but allows for the development of a more generic modelling approach. As the data contains measures that are common across all hospitals (average waiting time, number of patients in care, etc.), and are non-invasive to collect, this data set provides an opportunity to experiment and prototype models which can then be transported and tested across different health systems and networks. The descriptive analysis of the data set provided insights including but not limited to consistent daily and weekly behaviour patterns, and that approximately 50% of the time, the emergency departments had demand exceeding their capacity.

Following the theme of prototyping generalisable models, the results of the initial analysis guided the development of a hierarchical modelling framework that exploited the observed dependency structures whilst placing limited liability on input complexity. It was discerned that a normalised *capacity index* (the ratio of patients requiring treatment to the stated capacity) is suitably generic feature to have as the point of focus for a predictive model, and model-free prediction benchmarks were developed for each hospital

at each prediction horizon. This focus on the capacity index was foundational for the development a generalisable model, as it allows any emergency department to leverage the framework purely using historical attendance data.

Development of the hierarchical framework resulted in a two-stage model to be applied independently to each hospital. The first of the stages accounts for the systematic behaviours occurring in daily and weekly cycles, providing an intuitive baseline estimate which is then nuanced in the second stage by accounting for autocovariance structures via ARIMA models.

In order to effectively develop these models and carry out robust model comparison which prioritised predictive accuracy, the AROCV process was designed. AROCV allows the application of a cross-validation technique to time series data while mitigating data leakage and maintaining comparable predictions.

The models resulting from applying the hierarchical framework under the AROCV process improved upon the model-free benchmarks, with the minimum improvement at a five-hour horizon being a 17% reduction in mean prediction RMSE (*Royal Adelaide* and *Queen Elizabeth* hospitals), and the maximum providing a 37% reduction (*Women's and Children's Hospital*).

Following the finding of these results, time series models were implemented in a web-based prototype application¹, a proof of concept tool to provide practitioners with simple and practical tactical decision support.

Finally, in an effort to validate the success of the second stage ARIMA models and explore possible extension models, the same hierarchical framework was applied with autoregressive neural networks in place of ARIMA models. This experiment was carried out on the LMH data, providing a 1% improvement in mean predictive RMSE. This improvement however came at an extreme computational cost extending analysis time from a number of minutes to over two weeks. The outcome of this experiment were twofold, in validation of the ARIMA model results as well as a proof of concept for the application of autoregressive neural networks for this purpose.

Future scope

A natural next step for this work would be the implementation of the prediction methodology into a more mature tool for use by practitioners, in order to assess its utility in an emergency department. For this work to commence, significant and extensive engagement with hospital stakeholders would be necessary to ensure effective design and functionality as well as effective adoption.

This work also highlighted a number of avenues for further research in mathematics and statistics. The opportunities for future work can be categorised into two broad areas, namely increasing model sophistication and increasing process efficiency.

¹The platform is accessible via the address:
<https://datascience-adelaideuniversity.shinyapps.io/EmergencyDepartmentPrediction/>

Increasing the sophistication of a model can fine tune predictions to increase accuracy even further. There is great potential in increasing model complexity by increasing the number of input parameters. Supplementary parameters may be intrinsic, by using more of the observations contained in the collected data to, for example, effectively model the multiple-seasonality which is likely present, as well as external variables such as weather or social activity data. Further complexity could also be investigated by considering inter-dependencies between the hospitals, not to mention targeted investigation deeper into the application of neural networks. A third evident direction of increased complexity would be to investigate the use of other hierarchical, temporal modelling structures which exploit further auto and covariations, as well as multiple seasonality patterns.

Increases in process and modelling efficiency would be viewed as an enabling improvement, whereby more computationally efficient model fitting algorithms would enable easier investigation of the model space. A more computationally efficient and robust implementation of the AROCV methodology would open up the opportunity for a broader range of models to be tested on larger data sets. Following a similar theme, a more targeted investigation into neural network applications may provide a more efficient model fitting process which would provide an opportunity for further improvements in predictive accuracy.

In conclusion, this work developed a generalisable tool that produces short-term forecasts of emergency department demand. Although further work is required to validate its use, this work has provided the foundation for the development of generalisable tactical decision support tools to address Emergency Department overcrowding.

Appendix A

Detailed results

Detailed below are tables containing the resultant model metrics across the six hospitals, from the application of the time series analysis. The metrics detailed are the mean RMSE at the first five prediction horizons, along with the mean AICc and BIC.

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
ARMA(2,1)	0.1042	0.1302	0.1466	0.1579	0.1663	-7480.7762	-7457.2049
ARMA(2,2)	0.1043	0.1302	0.1465	0.1579	0.1663	-7480.7927	-7451.3288
ARMA(3,1)	0.1042	0.1302	0.1465	0.1579	0.1663	-7480.7340	-7451.2722
ARMA(3,2)	0.1044	0.1303	0.1466	0.1580	0.1664	-7481.8442	-7446.4963
ARMA(2,3)	0.1043	0.1302	0.1466	0.1580	0.1664	-7480.1528	-7444.8041
ARIMA(3,1,2)	0.1034	0.1291	0.1455	0.1569	0.1655	-7474.0358	-7438.6900
ARMA(1,3)	0.1049	0.1313	0.1472	0.1584	0.1667	-7458.8213	-7429.3609
ARMA(1,2)	0.1053	0.1312	0.1477	0.1592	0.1677	-7449.7757	-7426.2044
AR(2)	0.1048	0.1302	0.1463	0.1575	0.1658	-7440.5983	-7422.9176
ARMA(1,1)	0.1068	0.1309	0.1465	0.1576	0.1658	-7423.7395	-7406.0588
ARIMA(3,1,1)	0.1025	0.1288	0.1462	0.1589	0.1688	-7433.1535	-7403.6950
AR(1)	0.1072	0.1351	0.1522	0.1638	0.1721	-7346.0530	-7334.2643

Table A.1: *LMH time series results*

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
AR(2)	0.0890	0.1076	0.1195	0.1277	0.1336	-8053.8512	-8036.1705
ARMA(1,1)	0.0902	0.1074	0.1189	0.1269	0.1327	-8051.8230	-8034.1423
ARMA(1,2)	0.0895	0.1074	0.1192	0.1273	0.1332	-8057.6644	-8034.0931
ARMA(2,1)	0.0893	0.1073	0.1191	0.1271	0.1330	-8057.4164	-8033.8443
ARMA(1,3)	0.0894	0.1075	0.1192	0.1273	0.1332	-8056.9528	-8027.4924
ARMA(2,2)	0.0893	0.1073	0.1192	0.1274	0.1333	-8056.7840	-8027.2468
ARMA(3,1)	0.0892	0.1073	0.1191	0.1272	0.1331	-8056.3287	-8026.8062
ARMA(3,2)	0.0894	0.1074	0.1193	0.1274	0.1334	-8056.3026	-8020.9546
ARMA(2,3)	0.0891	0.1074	0.1191	0.1273	0.1332	-8055.9212	-8020.5048
ARIMA(3,1,2)	0.0882	0.1061	0.1182	0.1265	0.1329	-8043.9447	-8008.5990
ARIMA(3,1,1)	0.0875	0.1055	0.1178	0.1265	0.1332	-8031.6858	-8002.2272
AR(1)	0.0924	0.1131	0.1257	0.1340	0.1396	-7889.0461	-7877.2574

Table A.2: *RAH time series results*

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
AR(2)	0.1458	0.1766	0.1959	0.2081	0.2163	-5258.2808	-5240.6001
ARMA(1,1)	0.1471	0.1762	0.1950	0.2072	0.2155	-5255.5847	-5237.9040
ARMA(1,2)	0.1462	0.1763	0.1955	0.2078	0.2161	-5258.5697	-5234.9984
ARMA(2,1)	0.1460	0.1762	0.1953	0.2075	0.2158	-5258.1344	-5234.5631
ARMA(1,3)	0.1463	0.1765	0.1957	0.2080	0.2163	-5257.7911	-5228.3307
ARMA(2,2)	0.1462	0.1764	0.1959	0.2082	0.2166	-5257.7287	-5228.1605
ARMA(3,1)	0.1458	0.1760	0.1952	0.2073	0.2157	-5252.6650	-5223.1087
ARMA(2,3)	0.1462	0.1765	0.1960	0.2084	0.2168	-5257.1771	-5221.7070
ARMA(3,2)	0.1462	0.1764	0.1958	0.2081	0.2165	-5256.4977	-5221.1498
ARIMA(3,1,2)	0.1440	0.1741	0.1939	0.2066	0.2155	-5246.6886	-5211.3428
ARIMA(3,1,1)	0.1434	0.1739	0.1943	0.2078	0.2177	-5224.0165	-5194.5579
AR(1)	0.1511	0.1844	0.2039	0.2154	0.2224	-5137.0484	-5125.2597

Table A.3: *TQEH time series results*

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
AR(2)	0.1697	0.2076	0.2302	0.2444	0.2543	-4409.9851	-4392.3043
ARMA(2,1)	0.1698	0.2077	0.2305	0.2448	0.2547	-4415.0095	-4391.4382
ARMA(3,1)	0.1696	0.2073	0.2300	0.2444	0.2546	-4415.2839	-4385.8137
ARMA(2,2)	0.1699	0.2074	0.2302	0.2445	0.2544	-4415.1994	-4385.7369
ARMA(1,2)	0.1706	0.2081	0.2310	0.2453	0.2552	-4406.8970	-4383.3257
ARMA(1,3)	0.1702	0.2081	0.2303	0.2444	0.2542	-4412.2489	-4382.7885
ARMA(2,3)	0.1696	0.2072	0.2299	0.2442	0.2542	-4415.1618	-4379.8039
ARMA(3,2)	0.1699	0.2076	0.2303	0.2447	0.2546	-4413.9718	-4378.6238
ARIMA(3,1,2)	0.1684	0.2062	0.2293	0.2439	0.2543	-4407.7530	-4372.4073
ARMA(1,1)	0.1726	0.2077	0.2296	0.2436	0.2535	-4386.6741	-4368.9934
ARIMA(3,1,1)	0.1678	0.2057	0.2293	0.2444	0.2554	-4394.8545	-4365.3960
AR(1)	0.1772	0.2186	0.2419	0.2556	0.2643	-4237.4828	-4225.6942

Table A.4: *NHS time series results*

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
AR(2)	0.1246	0.1534	0.1721	0.1852	0.1950	-6696.0943	-6678.4136
ARMA(2,1)	0.1245	0.1535	0.1724	0.1856	0.1955	-6699.7506	-6676.1793
ARMA(1,2)	0.1250	0.1540	0.1729	0.1862	0.1960	-6697.4073	-6673.8360
ARMA(1,1)	0.1262	0.1537	0.1720	0.1849	0.1946	-6690.6441	-6672.9634
ARMA(2,2)	0.1242	0.1533	0.1723	0.1855	0.1954	-6699.4579	-6669.9608
ARMA(3,1)	0.1238	0.1530	0.1722	0.1855	0.1955	-6699.4020	-6669.9275
ARIMA(3,1,1)	0.1218	0.1499	0.1681	0.1808	0.1903	-6698.1444	-6668.6859
ARMA(1,3)	0.1248	0.1541	0.1728	0.1859	0.1957	-6697.9023	-6668.4419
ARMA(2,3)	0.1236	0.1528	0.1716	0.1850	0.1950	-6700.2079	-6664.8196
ARMA(3,2)	0.1245	0.1535	0.1723	0.1855	0.1952	-6700.1388	-6664.7908
ARIMA(3,1,2)	0.1221	0.1500	0.1681	0.1807	0.1900	-6700.0657	-6664.7200
AR(1)	0.1269	0.1581	0.1780	0.1915	0.2014	-6638.2559	-6626.4672

Table A.5: *MH time series results*

Model	1 Hr	2 Hrs	3 Hrs	4 Hrs	5 Hrs	Mean AICc	Mean BIC
ARMA(3,2)	0.1708	0.2147	0.2435	0.2619	0.2749	-4862.7256	-4827.3776
ARMA(2,1)	0.1705	0.2147	0.2434	0.2618	0.2747	-4811.8893	-4788.3180
ARMA(2,2)	0.1707	0.2146	0.2435	0.2619	0.2750	-4814.6775	-4785.2171
ARMA(3,1)	0.1703	0.2145	0.2434	0.2619	0.2750	-4814.5815	-4785.1211
ARIMA(3,1,2)	0.1683	0.2110	0.2388	0.2565	0.2689	-4816.1995	-4780.8538
ARMA(2,3)	0.1704	0.2143	0.2432	0.2617	0.2748	-4813.2527	-4777.9047
ARMA(1,3)	0.1714	0.2160	0.2440	0.2621	0.2750	-4759.6320	-4730.1716
ARMA(1,2)	0.1716	0.2158	0.2448	0.2632	0.2762	-4744.8451	-4721.2738
AR(2)	0.1718	0.2148	0.2429	0.2611	0.2741	-4719.8930	-4702.2123
ARIMA(3,1,1)	0.1662	0.2125	0.2465	0.2708	0.2903	-4725.2753	-4695.8167
ARMA(1,1)	0.1745	0.2156	0.2432	0.2612	0.2741	-4685.4280	-4667.7473
AR(1)	0.1771	0.2234	0.2520	0.2696	0.2813	-4583.0602	-4571.2716

Table A.6: *WCHP time series results*

Appendix B

ARIMA models

In Chapter 5, it is stated that an $\text{ARIMA}(p, d, q)$ process Y can be specified as:

$$\tilde{Y}_t^{(d)} = \left(\sum_{m=1}^p \varphi_m \mathbf{B}^m \right) \tilde{Y}^{(d)} + \left(1 + \sum_{m=1}^q \theta_m \mathbf{B}^m \right) \varepsilon_t.$$

Here, a more detailed discussion of ARIMA models is included. After introducing some initial notation and analysis tools, the constituent processes are defined and investigated for their characteristic behaviours.

The backwards shift operator

Consider a stochastic process Y_t indexed by time $t \in \mathbb{N}$, which evolves with some as yet unspecified dynamics. To this specification, the operator $\mathbf{B}$ referred to as the *backward shift* or *lag* operator is defined as :

$$\mathbf{B}Y_t := Y_{t-1}.$$

In this fashion, repeated application of the operator corresponds to lagged observation $\mathbf{B}^n Y_t = Y_{t-n}$.

ACF and PACF

Key tools routinely used for the analysis of time series processes and models are the *Autocorrelation* and *Partial Autocorrelation Functions* (ACF and PACF). These two functions are constructed to reflect a measure of similarity (correlation) a stochastic process has with itself through time. For the purpose of definition, consider the same stochastic process Y_t , and denote its mean as μ and variance as σ^2 . The Autocorrelation Function at lag k is simply defined to be the correlation of Y_t with Y_{t+k} :

$$\begin{aligned}
\rho_k &:= \text{Cor}(Y_t, Y_{t+k}) \\
&= \frac{\text{Cov}(Y_t, Y_{t+k})}{\sigma^2} \\
&= \frac{\mathbb{E}[(Y_t - \mu)(Y_{t+k} - \mu)]}{\sigma^2}.
\end{aligned}$$

In a similar fashion, the PACF is defined as:

$$\begin{aligned}
\alpha_k &:= \text{Cor}(Y_t, Y_{t+k} | Y_{t+1}, \dots, Y_{t+k}) \\
&= \frac{\text{Cov}(Y_t, Y_{t+k} | Y_{t+1}, \dots, Y_{t+k})}{\sqrt{\text{Var}(Y_t | Y_{t+1}, \dots, Y_{t+k}) \text{Var}(Y_{t+k} | Y_{t+1}, \dots, Y_{t+k})}},
\end{aligned}$$

that is; a measure of the similarity given information about all interim values of the process. As will be seen examples below, the sample ACFs and PACFs are often calculated and analysed for characteristic behaviours, allowing for consistency in simulation settings and assessment of model appropriateness in the model building process.

Next, the types of dependencies contained in ARIMA processes will be detailed, allowing for an informed discussion of the model building process in the sections that follow.

Autoregressive processes

Motivating the simplest of time series models, autoregressive (AR) processes are constructed by aggregating a linear combination of previous values of the process along with a perturbation. In this work, the perturbations will be a sequence of normal random variables with mean 0 and a specified variance σ_ε^2 . In this way, an autoregressive process of order p is written as:

$$Y_t = \left(\sum_{m=1}^p \varphi_m \mathbf{B}^m \right) Y_t + \varepsilon_t,$$

where ε_t are the normal random variables, and φ_m are the m^{th} order coefficients. The coefficients φ_m can be interpreted as the level of influence associated with the m -lagged value of the process. These can be thought of as being analogous to the coefficients in a simple linear regression model, hence the name *autoregressive* process. An AR process of order p , is one in which p is the greatest lag for which φ_m is non-zero, this is denoted as an $AR(p)$ process.

In practice, the use of autoregressive processes is restricted to stationary processes, that is, processes that do not have a time-dependent mean (see ARIMA processes below for

non-stationary models). It is important to note however, that in order for an AR process to be stationary the model must satisfy particular requirements. As detailed by Metcalfe and Cowpertwait [81], in order for the process Y_t described above to be stationary, it is required that the following polynomial equation in b :

$$\phi_p(b) := \sum_{m=1}^p \varphi_m b^m = 0, \quad (\text{B.1})$$

referred to as the *characteristic polynomial* has roots that sit exclusively outside the unit circle ($|b| > 1$).

To examine emergent features of an autocorrelation process, consider an example in the form of a simulated AR(2) process. Using the `arma.sim` method from the `stats` package [76] a sample path of 300 observations can be generated from the process:

$$Y_t = 0.5Y_{t-2} + 0.45Y_{t-1} + \varepsilon_t, \quad (\text{B.2})$$

for $t = 1, 2, \dots, 300$, where $\varepsilon_t \sim N(0, 1)$. The resultant sample path is displayed in Figure B.1, along with the ACF and PACF evaluated for the first twenty lags. The signature features of an autoregressive process are exhibited in the ACF and PACF plots. In particular, an AR(p) will have a PACF that *cuts off*¹ after p lags, while the magnitude of the ACF will tail off.

Moving Average processes

Moving Average processes (MA), in a similar way to AR processes are simply constructed through linear combinations. In contrast however, the linear combination in an MA process, is one of the previous normal random variables ε_i rather than the observations. Commonly, MA processes are constructed to be centred on a specified mean μ , allowing an MA process of order q to be written as:

$$Y_t = \mu + \left(1 + \sum_{m=1}^q \theta_m \mathbf{B}^m \right) \varepsilon_t.$$

In a similar way to the AR construction, the coefficients θ_m can be interpreted as the level of influence of the m^{th} lag, and for a MA process to be of order- q the q is the greatest value for which θ_q is non-zero. By construction, the distribution of the ε_t has a zero mean for all t , meaning all MA models have a time-independent mean and are hence stationary processes.

¹By “cuts off”, it is meant that for a particular lag, and any lag thereafter the autocorrelation function is statistically indistinguishable from zero. The value below which the cut-off occurs is resembled by the blue dashed line in ACF and PACF plots.

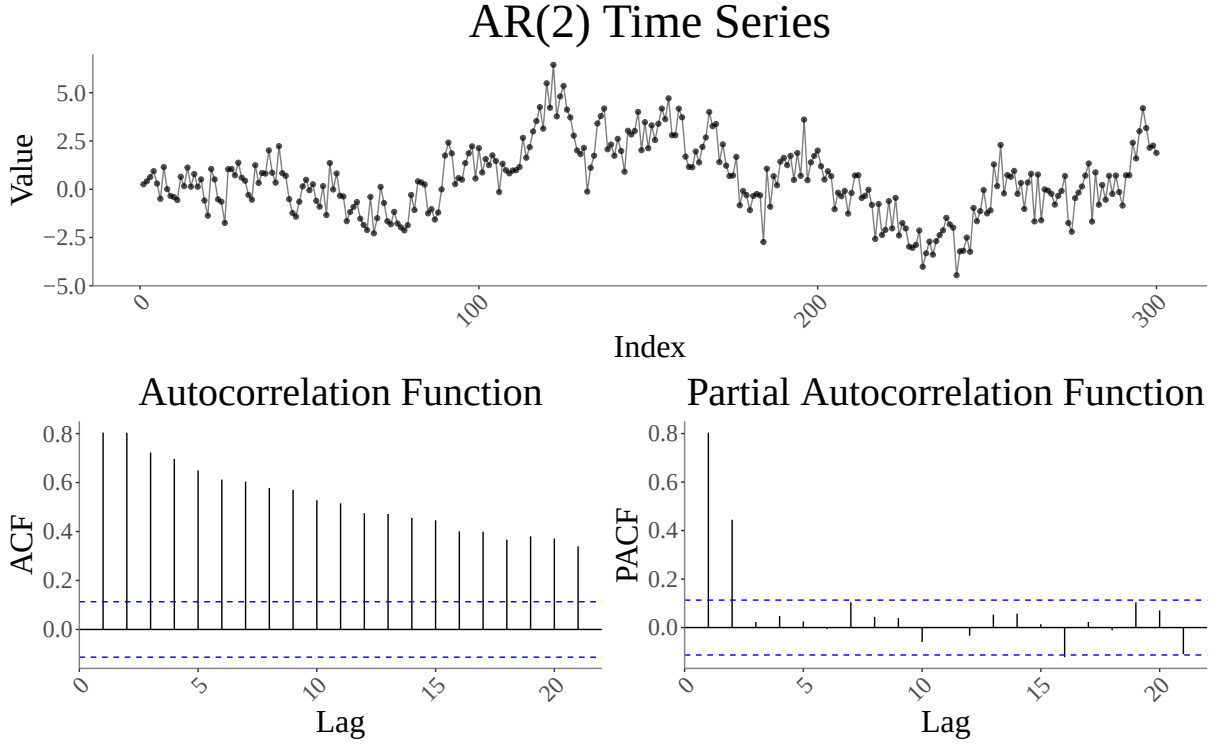


Figure B.1: Visualisation set of a sample bath from the $AR(2)$ process described by eq. (B.2). The top plot exhibits displays the evolution of the process as the index increases. The bottom left exhibits the sample ACF which is seen to slowly decay as the lag increases. The bottom right plot shows the PACF which cuts off after two lags. Each of these features are characteristic of an $AR(2)$ process.

A simulated example of an $MA(2)$ process is exhibited in Figure B.2. For illustrative purposes, consider the following MA model:

$$Y_t = \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t, \quad (\text{B.3})$$

where the $\varepsilon_i \sim N(0, 1)$, and generate a sample path consisting of 200 observations. As previously with the AR model, the ACF and PACF exhibit the key characteristics expected to be seen in an MA process. Typically, behaviour that is opposite to that of an AR model is observed; an MA process of order q will have an ACF that cuts off after q lags and a PACF that decays in magnitude as the lag increases.

Mixed processes:

From these two fundamental processes, a range of models can be constructed to model more detailed dynamics using combinations thereof. The resultant mixed *Autoregressive*

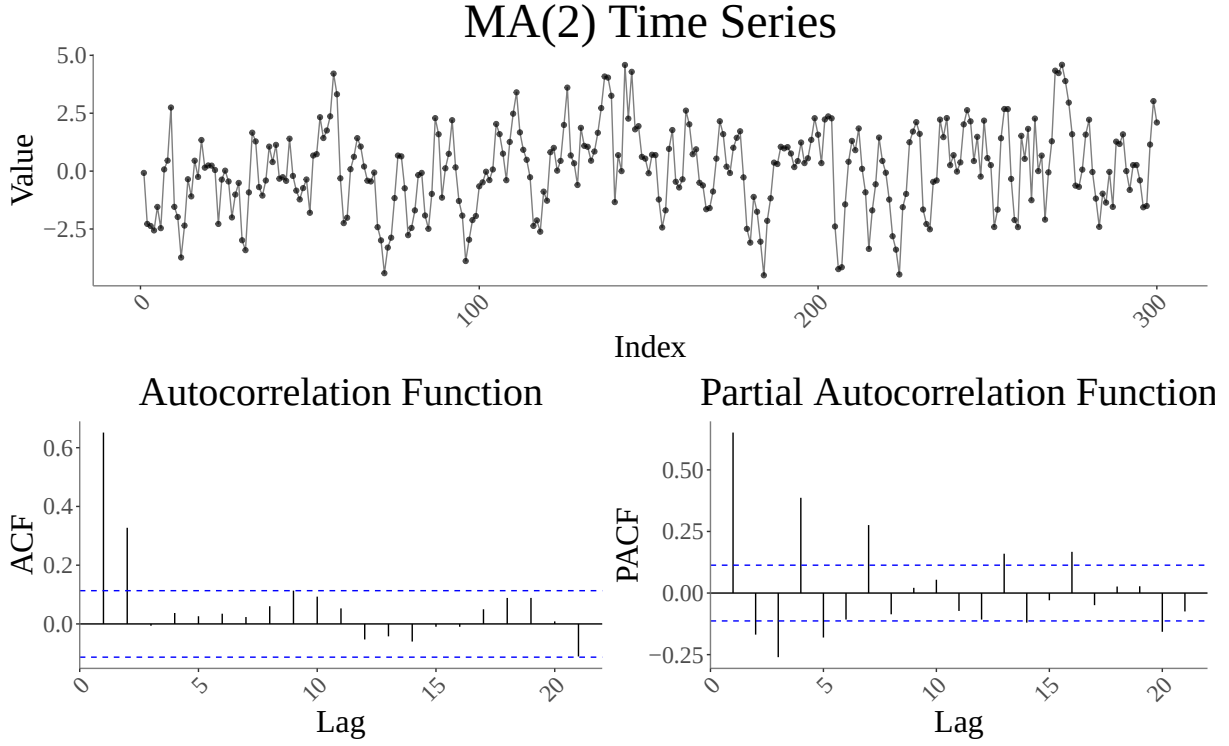


Figure B.2: A visualisation set of a sample path from the $MA(2)$ process described by eq. (B.3). The top plot displays the evolution of the process, while the bottom left plot displays ACF cutting off after two lags, and the bottom right plot shows the PACF decays in magnitude as the lag increases.

Moving Average (ARMA) model is constructed through linear combinations of both the prior observations and normal perturbations:

$$Y_t = \left(\sum_{m=1}^p \varphi_m \mathbf{B}^m \right) Y_t + \left(1 + \sum_{m=1}^q \theta_m \mathbf{B}^m \right) \varepsilon_t,$$

resulting in what is referred to as an $ARMA(p,q)$ model, where p and q reflect the order of the AR and MA terms respectively. As one may expect, the condition of stationarity is identical to that of AR processes as detailed by (B.1).

As an example, Figure B.3 displays a sample path from an $ARMA(2,2)$ process. The particular process simulated is chosen to be a combination of the previous two examples:

$$Y_t = 0.5Y_{t-2} + 0.45Y_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t, \quad (\text{B.4})$$

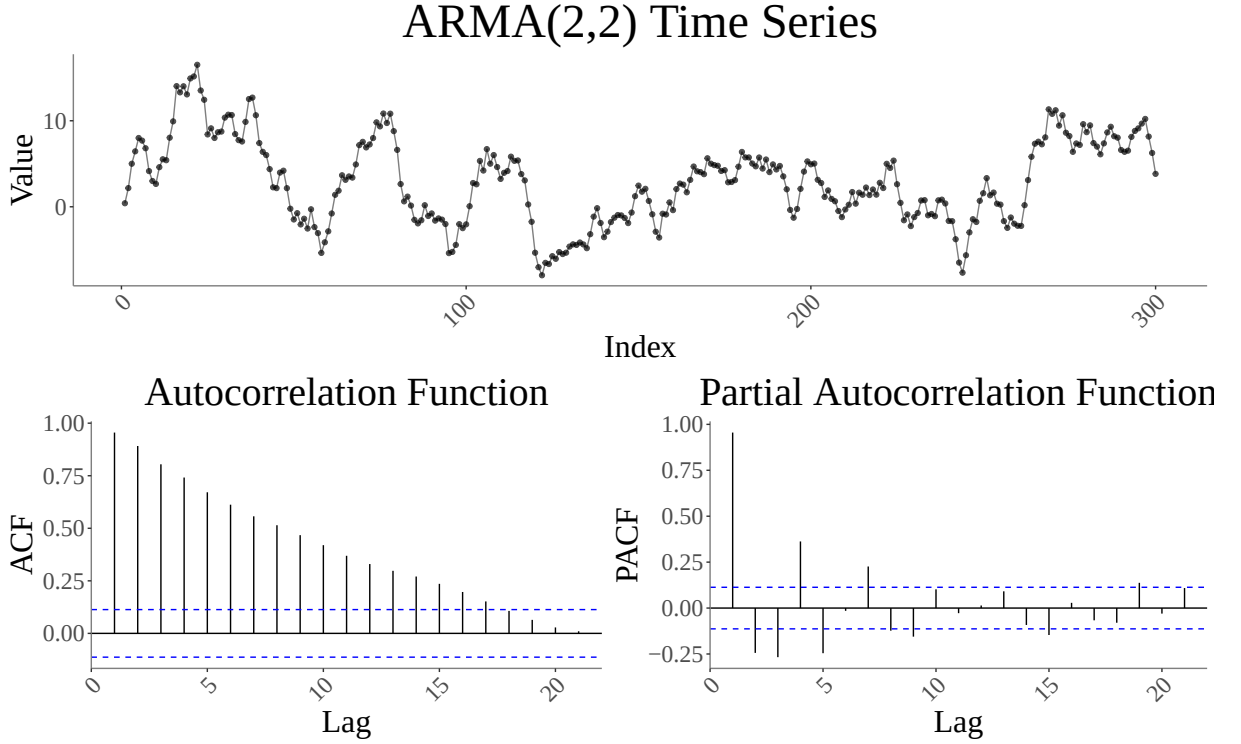


Figure B.3: A visualisation set of a sample path from the $ARMA(2,2)$ process described by eq. (B.4). The evolution of the process in the top plot is observed to sample a greater portion of the space, and does so with more structure, while the ACF and PACF plots are both observed to decay in magnitude as the lag increases.

The behaviour of the process can be seen to have a greater range and structure than the AR and the MA process previously. The characteristic properties of an ARMA process are also observed, with the ACF *and* PACF decaying in magnitude as the lag increases. This is in contrast to the simpler models where only one of the functions tends to decay while the other cuts off.

Non-stationary processes: the ARIMA model

It was noted previously that non-stationary processes are best constructed using different tools. Box and Jenkins [70] detail that while there are an infinite number of ways a process can be non-stationary, only a slight modification of the ARMA process is required to replicate the dynamics of systems in focus here. The slight adaption made is, rather than considering the true values of the process Y_t , the focus is instead on the relative differences:

$$\tilde{Y}_t := Y_t - Y_{t-1} = (1 - B)Y_t.$$

In this way, the dynamics of non-stationarity are not modelled, focusing efforts instead on the variance of the process explainable from the differenced data. It is noted that this process, referred to as *differencing* can be applied multiple times as:

$$\tilde{Y}_t^{(d)} := (1 - \mathbf{B})^d Y_t,$$

allowing for the removal of higher-order non-stationarity. In this work however, only cases in which $d = 1$ are considered. Applying this modification to the ARMA model produces the *Autoregressive Integrated Moving Average Model* (ARIMA) where the *integrated* refers to the summing required in the back-transformation from $\tilde{Y}_t$ to Y . This concept of differencing along with the construction of the ARMA process from above, produces what is denoted as the ARIMA(p, d, q) process:

$$\tilde{Y}_t^{(d)} = \left(\sum_{m=1}^p \varphi_m \mathbf{B}^m \right) \tilde{Y}^{(d)} + \left(1 + \sum_{m=1}^q \theta_m \mathbf{B}^m \right) \varepsilon_t.$$

Note that for specific values of p , d , and q in the ARIMA process, the simpler models can be recovered. For example, setting:

- $q = 0, d = 0$ produces the AR(p) process,
- $d = 0, p = 0$, produces the MA(p) process, and
- $d = 0$ produces the ARMA(p, q) process.

Bibliography

- [1] Australian Institute of Health and Welfare. Emergency department care - australian institute of health and welfare, 2021. URL <https://www.aihw.gov.au/reports-data/myhospitals/sectors/emergency-department-care>.
- [2] Robert W. Derlet and John R. Richards. Overcrowding in the nation's emergency departments: Complex causes and disturbing effects. *Annals of Emergency Medicine*, 35:63–68, 1 2000. ISSN 01960644. doi: 10.1016/S0196-0644(00)70105-3.
- [3] Stephen Trzeciak and E. P. Rivers. Emergency department overcrowding in the united states: An emerging threat to patient safety and public health, 9 2003. ISSN 13510622.
- [4] Nathan R. Hoot and Dominik Aronsky. Systematic review of emergency department crowding: Causes, effects, and solutions, 2008. ISSN 01960644.
- [5] Oscar Miró, Maite Antonio, Santos Jiménez, Antoni. De Dios, Miquel Sánchez, Antoni Borrás, and Jose Millá. Decreased health care quality associated with emergency department overcrowding. *European journal of emergency medicine : official journal of the European Society for Emergency Medicine*, 6:105–107, 1999. ISSN 09699546. doi: 10.1097/00063110-199906000-00003.
- [6] Drew B. Richardson. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Medical Journal of Australia*, 184: 213–216, 3 2006. ISSN 0025729X. doi: 10.5694/j.1326-5377.2006.tb00204.x. URL <https://onlinelibrary.wiley.com/doi/full/10.5694/j.1326-5377.2006.tb00204.x><https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.2006.tb00204.x><https://onlinelibrary.wiley.com/doi/10.5694/j.1326-5377.2006.tb00204.x>.
- [7] Peter C. Sprivulis, Julie Ann Da Silva, Ian G. Jacobs, Amanda R.L. Frazer, and George A. Jelinek. The association between hospital overcrowding and mortality among patients admitted via western australian emergency departments. *Medical Journal of Australia*, 184:208–212, 3 2006. ISSN 0025729X. doi: 10.5694/j.1326-5377.2006.tb00203.x.

- [8] Samuel Hyman. The cook county hospital emergency department. new methods to meet changing needs. *Postgraduate medicine*, 40:375–377, 1966. ISSN 00325481. doi: 10.1080/00325481.1966.11695972. URL <https://www.tandfonline.com/doi/abs/10.1080/00325481.1966.11695972>.
- [9] Jonathan D. Mayer. Emergency medical service: Delays, response time and survival. *Medical Care*, 17:818–827, 1979. ISSN 15371948. doi: 10.1097/00005650-197908000-00004. URL <https://oce-ovid-com.proxy.library.adelaide.edu.au/article/00005650-197908000-00004/HTML>.
- [10] Jeffrey Goldberg. Operations research models for the deployment of emergency services vehicles. *EMS Management Journal*, 1:20–39, 2004.
- [11] Kobi Peleg and Joseph S. Pliskin. A geographic information system simulation model of ems: Reducing ambulance response time. *American Journal of Emergency Medicine*, 22:164–170, 5 2004. ISSN 07356757. doi: 10.1016/j.ajem.2004.02.003.
- [12] John F. Repede and John J. Bernardo. Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky. *European Journal of Operational Research*, 75:567–581, 6 1994. ISSN 03772217. doi: 10.1016/0377-2217(94)90297-6.
- [13] Soroush Saghaian, Garrett Austin, and Stephen J. Traub. Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Transactions on Healthcare Systems Engineering*, 5:101–123, 4 2015. ISSN 1948-8300. doi: 10.1080/19488300.2015.1017676. URL <http://www.tandfonline.com/doi/full/10.1080/19488300.2015.1017676>.
- [14] Drew B Richardson. Reducing patient time in the emergency department. *Medical Journal of Australia*, 179:516–517, 11 2003. ISSN 0025-729X. doi: 10.5694/j.1326-5377.2003.tb05672.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.2003.tb05672.x>.
- [15] Jin H. Han, Chuan Zhou, Daniel J. France, Sheng Zhong, Ian Jones, Alan B. Storrow, and Dominik Aronsky. The effect of emergency department expansion on emergency department overcrowding. *Academic Emergency Medicine*, 14:338–343, 4 2007. ISSN 10696563. doi: 10.1197/j.aem.2006.12.005. URL <http://doi.wiley.com/10.1197/j.aem.2006.12.005>.
- [16] Linda V. Green, Peter J. Kolesar, and João Soares. Improving the sipp approach for staffing service systems that have cyclic demands. *Operations Research*, 49:549–564, 2001. ISSN 0030364X. doi: 10.1287/opre.49.4.549.11228.

- [17] Linda V. Green, João Soares, James F. Giglio, and Robert A. Green. Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine*, 13:61–68, 1 2006. ISSN 10696563. doi: 10.1197/j.aem.2005.07.034.
- [18] Linda V. Green, Peter J. Kolesar, and Ward Whitt. Coping with time-varying demand when setting staffing requirements for a service system, 2007. ISSN 10591478.
- [19] Linda V. Green. Using operations research to reduce delays for healthcare, 9 2008. URL <http://pubsonline.informs.org>.
- [20] Natalia Yankovic and Linda V. Green. Identifying good nursing levels: A queueing approach. *Operations Research*, 59:942–955, 7 2011. ISSN 0030364X. doi: 10.1287/opre.1110.0943.
- [21] David Sinreich and Ola Jabali. Staggered work shifts: A way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science*, 10:293–308, 9 2007. ISSN 13869620. doi: 10.1007/s10729-007-9021-z. URL <https://pubmed-ncbi-nlm-nih-gov.proxy.library.adelaide.edu.au/17695139/>.
- [22] David Sinreich, Ola Jabali, and Nico P. Dellaert. Reducing emergency department waiting times by adjusting work shifts considering patient visits to multiple care providers. *IIE Transactions (Institute of Industrial Engineers)*, 44:163–180, 3 2012. ISSN 0740817X. doi: 10.1080/0740817X.2011.609875.
- [23] Rahul K. Khare, Emilie S. Powell, Gilles Reinhardt, and Martin Lucenti. Adding more beds to the emergency department or reducing admitted patient boarding times: Which has a more significant influence on emergency department congestion? *Annals of Emergency Medicine*, 53, 2009. ISSN 10976760. doi: 10.1016/j.annemergmed.2008.07.009.
- [24] Sandra Schneider, Frank Zwemer, Andrew Doniger, Robbin Dick, Tim Czapranski, and Eric Davis. Rochester, new york: A decade of emergency department overcrowding. *Academic Emergency Medicine*, 8:1044–1050, 2001. ISSN 10696563. doi: 10.1111/j.1553-2712.2001.tb01113.x.
- [25] Pengyi Shi, Mabel C. Chou, J. G. Dai, Ding Ding, and Joe Sim. Models and insights for hospital inpatient operations: Time-dependent ed boarding time. *Management Science*, 62:1–28, 1 2016. ISSN 15265501. doi: 10.1287/mnsc.2014.2112.
- [26] Snorre Kverndokk and Hans Olav Melberg. Using fees to reduce bed-blocking: a game between hospitals and long-term care providers. *European Journal of Health*

Economics, 22:931–949, 8 2021. ISSN 16187601. doi: 10.1007/s10198-021-01299-9. URL <https://doi.org/10.1007/s10198-021-01299-9>.

- [27] David G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics*, 24:338–354, 1953. ISSN 0003-4851. doi: 10.1214/aoms/1177728975. URL https://www-jstor-org.proxy.library.adelaide.edu.au/stable/2236285?seq=1#metadata_info_tab_contents.
- [28] Marcos Singer and Patricio Donoso. Assessing an ambulance service with queuing theory. *Computers and Operations Research*, 35:2549–2560, 8 2008. ISSN 03050548. doi: 10.1016/j.cor.2006.12.005.
- [29] Jeffery K. Cochran and James R. Broyles. Developing nonlinear queuing regressions to increase emergency department patient safety: Approximating reneging with balking. *Computers and Industrial Engineering*, 59:378–386, 10 2010. ISSN 03608352. doi: 10.1016/j.cie.2010.05.010.
- [30] Mor Armony and Avishai Mandelbaum. Routing and staffing in large-scale service systems: The case of homogeneous impatient customers and heterogeneous servers. *Operations Research*, 59:50–65, 1 2011. ISSN 0030364X. doi: 10.1287/opre.1100.0878.
- [31] Jeffery K. Cochran and Kevin T. Roche. A multi-class queuing network analysis methodology for improving hospital emergency department performance. *Computers and Operations Research*, 36:1497–1512, 5 2009. ISSN 03050548. doi: 10.1016/j.cor.2008.02.004.
- [32] Lauren McNamara, Peter T. Vanberkel, David Petrie, and Alix J.E. Carter. An application and framework for evaluating emergency department networks using location analysis and geographic information systems. *Computers & Industrial Engineering*, 149, 11 2020. ISSN 0360-8352. doi: 10.1016/J.CIE.2020.106766.
- [33] Kazi Badrul Ahsan, M. R. Alam, Doug Gordon Morel, and M. A. Karim. Emergency department resource optimisation for improved performance: a review. *Journal of Industrial Engineering International*, 15:253–266, 12 2019. ISSN 2251712X. doi: 10.1007/s40092-019-00335-x.
- [34] Sally Brailsford, Paul Harper, Bijendra Patel, and Martin Pitt. An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3:130–140, 9 2009. ISSN 1747-7778. doi: 10.1057/jos.2009.10. URL <https://www.tandfonline.com/doi/full/10.1057/jos.2009.10>.
- [35] Mahsa Dehghani, Nazila Moftian, Peyman Rezaei-Hachesu, and Taha Samad-Soltani. A step-by-step framework on discrete events simulation in emergency department;

- a systematic review. *Bulletin of emergency and trauma*, 5:79–89, 4 2017. ISSN 2322-2522. URL <http://www.ncbi.nlm.nih.gov/pubmed/28507994><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5406177>.
- [36] David Fone, Sandra Hollinghurst, Mark Temple, Alison Round, Nathan Lester, Alison Weightman, Katherine Roberts, Edward Coyle, Gwyn Bevan, and Stephen Palmer. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health*, 25:325–335, 12 2003. ISSN 1741-3842. doi: 10.1093/pubmed/fdg075. URL <https://academic.oup.com/jpubhealth/article-lookup/doi/10.1093/pubmed/fdg075>.
- [37] Muhammet Gul and Ali Fuat Guneri. A comprehensive review of emergency department simulation applications for normal and disaster conditions. *Computers and Industrial Engineering*, 83:327–344, 5 2015. ISSN 0360-8352. doi: 10.1016/J.CIE.2015.02.018. URL <https://www.sciencedirect.com/science/article/pii/S0360835215000972>.
- [38] Syed Mohiuddin, John Busby, Jelena Savović, Alison Richards, Kate Northstone, William Hollingworth, Jenny L. Donovan, and Christos Vasilakis. Patient flow within uk emergency departments: A systematic review of the use of computer simulation modelling methods. *BMJ Open*, 7:e015007, 5 2017. ISSN 20446055. doi: 10.1136/bmjopen-2016-015007. URL <http://bmjopen.bmj.com/>.
- [39] Sharoda A. Paul, Madhu C. Reddy, and Christopher J. Deflitch. A systematic review of simulation studies investigating emergency department overcrowding, 8 2010. ISSN 00375497.
- [40] Andrew Salmon, Sebastian Rachuba, Simon Briscoe, and Martin Pitt. A structured literature review of simulation modelling applied to emergency departments: Current patterns and emerging trends. *Operations Research for Health Care*, 19:1–13, 12 2018. ISSN 22116923. doi: 10.1016/j.orhc.2018.01.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S2211692317301042>.
- [41] Lien Vanbrabant, Kris Braekers, Katrien Ramaekers, and Inneke Van Nieuwenhuysse. Simulation of emergency department operations: A comprehensive review of kpis and operational improvements. *Computers and Industrial Engineering*, 131:356–381, 5 2019. ISSN 03608352. doi: 10.1016/j.cie.2019.03.025.
- [42] Toni Ruohonen, Pekka Neittaanmäki, and Jorma Teittinen. Simulation model for improving the operation of the emergency department of special health care. pages 453–458, 2006. ISBN 1424405017. doi: 10.1109/WSC.2006.323115.

- [43] Junwen Wang, Jingshan Li, Kathy Tussey, and Kay Ross. Reducing length of stay in emergency department: A simulation study at a community hospital. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 42:1314–1322, 2012. ISSN 10834427. doi: 10.1109/TSMCA.2012.2210204.
- [44] Nathan R. Hoot, Larry J. LeBlanc, Ian Jones, Scott R. Levin, Chuan Zhou, Cynthia S. Gadd, and Dominik Aronsky. Forecasting emergency department crowding: A discrete event simulation. *Annals of Emergency Medicine*, 52:116–125, 8 2008. ISSN 01960644. doi: 10.1016/j.annemergmed.2007.12.011.
- [45] David C. Lane, Camilla Monefeldt, and Jonathan Rosenhead. Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the Operational Research Society*, 51:518–531, 2000. ISSN 14769360. doi: 10.1057/palgrave.jors.2600892.
- [46] Wang Lu. An agent-based simulation for workflow in emergency department. pages 19–23, 2009. ISBN 9781424445325. doi: 10.1109/SIEDS.2009.5166148.
- [47] Spencer S. Jones and R. Scott Evans. An agent based simulation tool for scheduling emergency department physicians. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2008:338–342, 2008. ISSN 1942597X. URL /pmc/articles/PMC2656074//pmc/articles/PMC2656074/?report=abstract<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656074/>.
- [48] Marek Laskowski and Shamir Mukhi. Agent-based simulation of emergency departments with patient diversion. volume 1 LNICST, pages 25–37. Springer, Berlin, Heidelberg, 2009. ISBN 3642004121. doi: 10.1007/978-3-642-00413-1_4. URL https://link-springer-com.proxy.library.adelaide.edu.au/chapter/10.1007/978-3-642-00413-1_4.
- [49] Electron efficiency measurements with the atlas detector using the 2015 lhc proton-proton collision data, 2016. URL <https://cds.cern.ch/record/2157687/files/ATLAS-CONF-2016-024.pdf>.
- [50] Anthony Kirtland, Jacqueline Lockwood, Karen Poisker, Linda Stamp, and Patricia Wolfe. Simulating an emergency department ’is as much fun as...’. pages 1039–1042. IEEE, 1995. doi: 10.1145/224401.224769.
- [51] Frank McGuire. Using simulation to reduce length of stay in emergency departments. *Journal of the Society for Health Systems*, 5:81–90, 1997. ISSN 10431721. doi: 10.1109/wsc.1994.717446.
- [52] Ruby E. Blasak, Darrell W. Starks, Wendy S. Armel, and Mary C. Hayduk. The use of simulation to evaluate hospital operations between the emergency department

- and a medical telemetry unit. volume 2, pages 1887–1893, 2003. doi: 10.1109/wsc.2003.1261649.
- [53] Carlos J. González, Merbil González, and Nilda M. Ríos. Improving the quality of service in an emergency room using simulation-animation and total quality management. *Computers and Industrial Engineering*, 33:97–100, 10 1997. ISSN 03608352. doi: 10.1016/s0360-8352(97)00049-1.
 - [54] Farid Kadri, Fouzi Harrou, Sondès Chaabane, and Christian Tahon. Time series modelling and forecasting of emergency department overcrowding. *Journal of Medical Systems*, 38, 2014. ISSN 1573689X. doi: 10.1007/s10916-014-0107-0.
 - [55] Mohamed Afilal, Farouk Yalaoui, Frédéric Dugardin, Lionel Amodeo, David Laplanche, and Philippe Blua. Forecasting the emergency department patients flow. *Journal of Medical Systems*, 40, 7 2016. ISSN 1573689X. doi: 10.1007/s10916-016-0527-0.
 - [56] Yan Sun, Bee Hoon Heng, Yian Tay Seow, and Eillyne Seow. Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine*, 9:1–9, 1 2009. ISSN 1471227X. doi: 10.1186/1471-227X-9-1/FIGURES/5. URL <https://bmccemergmed.biomedcentral.com/articles/10.1186/1471-227X-9-1>.
 - [57] Spencer S. Jones, R. Scott Evans, Todd L. Allen, Alun Thomas, Peter J. Haug, Shari J. Welch, and Gregory L. Snow. A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics*, 42:123–139, 2 2009. ISSN 15320464. doi: 10.1016/j.jbi.2008.05.003.
 - [58] Kalpani I. Duwalage, Ellen Burkett, Gentry White, Andy Wong, and Mery H. Thompson. Forecasting daily counts of patient presentations in australian emergency departments using statistical models with time-varying predictors. *Emergency Medicine Australasia*, 32:618–625, 8 2020. ISSN 1742-6723. doi: 10.1111/1742-6723.13481. URL <https://onlinelibrary-wiley-com.proxy.library.adelaide.edu.au/doi/full/10.1111/1742-6723.13481><https://onlinelibrary-wiley-com.proxy.library.adelaide.edu.au/doi/abs/10.1111/1742-6723.13481><https://onlinelibrary-wiley-com.proxy.library.adelaide.edu.au/doi/>.
 - [59] Yan Sun, Kiok Liang Teow, Bee Hoon Heng, Chee Kheong Ooi, and Seow Yian Tay. Real-time prediction of waiting time in the emergency department, using quantile regression. *Annals of Emergency Medicine*, 60:299–308, 9 2012. ISSN 01960644. doi: 10.1016/j.annemergmed.2012.03.011.
 - [60] Adrian Fletcher and Dave Worthington. What is a 'generic' hospital model?-a comparison of 'generic' and 'specific' hospital models of emergency patient flows.

Health Care Management Science, 12:374–391, 11 2009. ISSN 13869620. doi: 10.1007/s10729-009-9108-9.

- [61] Nikolaus Furian, Dietmar Neubacher, Michael O’Sullivan, Cameron G Walker, and Christian Pizzera. Gedmod – towards a generic toolkit for emergency department modeling. *Simulation Modelling Practice and Theory*, 87:239–273, 9 2018. ISSN 1569-190X. doi: 10.1016/J.SIMPAT.2018.07.010.
- [62] Laura M. Boyle, Adele H. Marshall, and Mark Mackay. A framework for developing generalisable discrete event simulation models of hospital emergency departments. *European Journal of Operational Research*, 302:337–347, 10 2022. ISSN 0377-2217. doi: 10.1016/J.EJOR.2021.12.033.
- [63] David Sinreich and Yariv N. Marmor. A simple and intuitive simulation tool for analyzing emergency department operations. volume 2, pages 1994–2002, 2004.
- [64] Kum Khiong Yang, Sean Shao Wei Lam, Joyce M.W. Low, and Marcus Eng Hock Ong. Managing emergency department crowding through improved triaging and resource allocation. *Operations Research for Health Care*, 10:13–22, 9 2016. ISSN 22116923. doi: 10.1016/j.orhc.2016.05.001.
- [65] Murat M. Gunal and Michael Pidd. Understanding accident and emergency department performance using simulation. pages 446–452, 2006. ISBN 1424405017. doi: 10.1109/WSC.2006.323114.
- [66] Murat M. Gunal and Michael Pidd. Interconnected des models of emergency, out-patient, and inpatient departments of a hospital. pages 1461–1466, 2007. ISBN 1424413060. doi: 10.1109/WSC.2007.4419757.
- [67] David M. Ferrin, Marty J. Miller, and Diana L. McBroom. Maximizing hospital finanacial impact and emergency department throughput with simulation. pages 1566–1573, 2007. ISBN 1424413060. doi: 10.1109/WSC.2007.4419774.
- [68] 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*, 58:1554–1562, 12 2007. ISSN 0160-5682. doi: 10.1057/palgrave.jors.2602344. URL <https://www.tandfonline.com/doi/full/10.1057/palgrave.jors.2602344>.
- [69] Government of South Australia. Emergency department dashboard — sa health, 2021. URL <https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/about+us/our+performance/our+hospital+dashboards/about+the+ted+dashboard/emergency+department+dashboard>.

- [70] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*, volume 14. HOLDEN-DAY, Inc., revised ed edition, 1977. ISBN 0-8162-1104-3.
- [71] Robert J Hyndman and George Athanasopoulos. *Forecasting: Principles and practice* (3rd ed), 2021. URL <https://otexts.com/fpp3/>.
- [72] Denis Kwiatkowski, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54: 159–178, 10 1992. ISSN 03044076. doi: 10.1016/0304-4076(92)90104-Y.
- [73] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974. ISSN 15582523. doi: 10.1109/TAC.1974.1100705.
- [74] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297, 6 1989. ISSN 00063444. doi: 10.2307/2336663.
- [75] Rob J. Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27:1–22, 7 2008. ISSN 15487660. doi: 10.18637/jss.v027.i03. URL <https://www.jstatsoft.org/index.php/jss/article/view/v027i03>.
- [76] R Core Team. *R: A language and environment for statistical computing*, 2020. URL <https://www.r-project.org/>.
- [77] Stephan Kolassa, Bahman Rostami-Tabar, and Enno Siemsen. *Demand Forecasting for Executives and Professionals*. CRC Press, 2023. URL <https://dfep.netlify.app/>.
- [78] Mitchell O’Hara-Wild, Rob Hyndman, Earo Wang, Gabriel Caceres, Tim-Gunnar Hensel, and Timothy Hyndman. *fable: Forecasting models for tidy time series*, 5 2021. URL <https://cran.r-project.org/package=fable>.
- [79] Juergen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 4 2014. doi: 10.1016/j.neunet.2014.09.003. URL <http://arxiv.org/abs/1404.7828><http://dx.doi.org/10.1016/j.neunet.2014.09.003>.
- [80] Charu C. Aggarwal. An introduction to neural networks. *Neural Networks and Deep Learning*, pages 1–52, 2018. doi: 10.1007/978-3-319-94463-0_1. URL https://link.springer.com/chapter/10.1007/978-3-319-94463-0_1.
- [81] Andrew V. Metcalfe and Paul S.P. Cowpertwait. *Introductory Time Series with R*. Springer New York, 2009. doi: 10.1007/978-0-387-88698-5.