

# Bayesian Inference for Enhanced Road Safety Analysis

NICOLA K. HEWETT



- - - A thesis submitted to Newcastle University for the degree of Doctor of  
Philosophy - - -

*School of Mathematics, Statistics and Physics,  
Newcastle University,  
Newcastle upon Tyne,  
United Kingdom*

**March 2023**



## Acknowledgements

I would first like to thank my supervisors Dr Lee Fawcett, Dr Andy Golightly, Dr Neil Thorpe, and the late Professor Richard Boys for providing me with the opportunity to pursue this PhD and for their immeasurable contributions and support throughout these last four years. I must also thank Dr Karsten Kremer from PTV group for generously giving his valuable time and insight throughout the duration of this project.

I owe a great deal of my success to Lynn Wilson, who encouraged me to pursue maths and whose belief in me has been a source of inspiration throughout my academic life. I am deeply grateful for the invaluable influence she has had on my academic pursuits.

Huge thanks must go to the PhD cohort who kept me sane throughout the years; especially those from PhD7 who (most importantly) helped me become a Mario Kart legend. I'm eternally grateful for Nick Keeper, Ryan Doran, James Murray, Kieran Peel, Holly Middleton-Spencer, Cameron Williams, Tom Flynn, Sam Hartharn-Evans, Joe Matthews and our PhD7 mascot, Gail.

I must thank my incredible friends: Caroline Sabiston, Helen Younger, Sophie Foster-Nixon, Rebecca Wilson, Sarah Horton, Faye Johnson, Bethanie Mitchinson, Steph Ross, Jess Treacy, Damian Robinson, Poonam Bagria and the WBOS gang who are an absolute joy to be around. I must give a special mention and express my heartfelt gratitude to Caroline and Helen, who have been a constant source of love, support, and encouragement throughout these past few years. They bring so much joy to my life and their unwavering faith in me has helped me to believe in myself; I cannot thank them enough for everything they do.

An immense thank you goes to Nick Keeper, who has made the last few months of my PhD far smoother than they otherwise would have been. I'm grateful for his unbounded wisdom, infinite patience, endless encouragement, culinary wizardry, and unconditional love.

Of course, none of this would've been possible without my incredible family, who have provided limitless emotional support and encouragement: Emma, Megan, John, Gill, Dad, Peach, and my wonderful Mum, who has always been my biggest cheerleader. Thank you all for believing in me. Lastly, I must extend my thanks to everyone who has been a part of shaping the past four years, I've loved (almost) every minute.



## Abstract

This thesis proposes innovative methods to analyse road traffic collision data with the aim of improving the evaluation of before/after safety schemes and predicting collision hotspots, thereby addressing significant research problems in the field of road safety. We integrate Bayesian inference, extreme value theory, and spatio-temporal modelling to create robust and flexible models for these analyses.

Specifically, the research problems include evaluating and demonstrating the limitations of frequently utilised techniques in the assessment of road safety schemes, and the need for bespoke modelling formulations for atypical before/after studies. Key findings are introduced in the form of a new model to capture treatment effect for randomised trials, and the application of extreme value theory to conduct a traffic conflict-based before/after safety scheme evaluation. To account for spatial correlation between neighbouring sites, Gaussian processes are included in the expressions for the location and scale parameters governing the generalised extreme value distribution. In terms of hotspot prediction, a Bayesian hierarchical model is proposed to segregate the seasonal and zonal effects in monthly collision data from multiple sites within fixed Traffic Administration Zones. Additionally, a spatio-temporal model for collision rates is introduced that allows for serial dependence, seasonality, and correlation between rates at nearby zones.

The key impact factors of this research are manifested in its practical applications to real-world data, including collision rate data from north Florida, USA; traffic-conflict data from Vancouver, Canada; STATS19 data from the UK; and collision count data from Tyne and Wear, UK. These applications demonstrate the effectiveness of the proposed approach in improving the evaluation of safety schemes and predicting collision hotspots, thereby offering insights that can guide stakeholders in making informed decisions on road safety interventions.



# Contents

<b>1</b>	<b>Background</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.1.1	Statistics and road safety . . . . .	2
1.1.2	Road safety schemes . . . . .	3
1.1.3	Prediction . . . . .	4
1.1.4	Aims and objectives . . . . .	5
1.2	Industrial collaborators . . . . .	6
1.2.1	PTV Group . . . . .	6
1.2.2	Gateshead Council . . . . .	7
1.2.3	Florida and New York State Departments of Transportation . . . . .	8
1.3	Datasets . . . . .	8
1.3.1	Northumbria Safety Camera Partnership data . . . . .	8
1.3.2	Florida State Department of Transport data . . . . .	10
1.3.3	Canadian leading pedestrian interval data . . . . .	10
1.3.4	UK STATS19 data . . . . .	13
1.4	Thesis structure . . . . .	15
1.5	List of publications . . . . .	18
<b>2</b>	<b>Methods</b>	<b>19</b>
2.1	Generalised linear modelling . . . . .	19
2.2	Extreme value theory . . . . .	23
2.2.1	The generalised extreme value distribution . . . . .	24
2.2.2	The generalised Pareto distribution . . . . .	26
2.2.3	Return levels . . . . .	26
2.3	Maximum likelihood estimation . . . . .	27
2.4	Bayesian inference . . . . .	31
2.4.1	Prior distributions . . . . .	32

2.4.2	Markov Chain Monte Carlo . . . . .	34
2.4.3	Bayesian inference example . . . . .	40
2.4.4	Deviance Information Criteria . . . . .	41
2.5	Dynamic linear models . . . . .	42
2.5.1	Bayesian inference for DLMs . . . . .	44
2.5.2	Example: Local level model . . . . .	47
<b>3</b>	<b>Statistical methods for traditional before/after studies</b>	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Confounding factors in traditional studies . . . . .	54
3.2.1	Regression to the mean . . . . .	54
3.2.2	Other confounders . . . . .	58
3.2.3	Road safety hotspot prediction . . . . .	59
3.3	The empirical Bayes approach for RTM . . . . .	60
3.4	Data example: The Northumbria Safety Camera Partnership data . . . . .	64
3.4.1	Empirical Bayes analysis . . . . .	64
3.4.2	Fully Bayes analysis . . . . .	65
3.4.3	Sensitivity to prior distribution . . . . .	68
3.5	Discussion . . . . .	72
<b>4</b>	<b>Bespoke modelling formulations for atypical before/after studies</b>	<b>74</b>
4.1	Introduction . . . . .	74
4.1.1	Vehicle Activated Signs . . . . .	75
4.1.2	Traffic conflict analyses . . . . .	75
4.2	Study of Vehicle Activated Signs in Tyne and Wear . . . . .	76
4.3	Study of a leading pedestrian interval intervention in Vancouver . . . . .	81
4.3.1	Including covariates . . . . .	84
4.3.2	Classifying extremes through block maxima . . . . .	85
4.3.3	Classifying extremes through threshold excesses . . . . .	87
4.4	Discussion . . . . .	97
<b>5</b>	<b>Modelling spatial extremes from the LPI intervention</b>	<b>101</b>
5.1	Background . . . . .	101
5.2	Preliminary inference . . . . .	104
5.3	Spatial extremes . . . . .	105
5.3.1	Model description . . . . .	107



5.3.2	Parameter FCDs . . . . .	108
5.3.3	Application . . . . .	111
5.4	Discussion . . . . .	119
<b>6</b>	<b>Accounting for seasonal and zonal effects in collision data</b>	<b>123</b>
6.1	Introduction . . . . .	123
6.2	Florida data analysis . . . . .	125
6.2.1	Full posterior . . . . .	126
6.2.2	Full conditional distributions . . . . .	127
6.2.3	Application . . . . .	130
6.2.4	Comparison to a fixed effects model . . . . .	131
6.3	UK data analysis . . . . .	135
6.3.1	Model description . . . . .	136
6.3.2	Application . . . . .	138
6.4	Discussion . . . . .	140
<b>7</b>	<b>Application of a sinusoidal dynamic linear model to road traffic collision data</b>	<b>143</b>
7.1	Introduction . . . . .	143
7.1.1	Data . . . . .	145
7.1.2	Zone specific model . . . . .	145
7.1.3	Joint model over zones . . . . .	147
7.2	Bayesian inference . . . . .	148
7.2.1	Simulation based inference . . . . .	149
7.2.2	Within-sample predictive density . . . . .	152
7.2.3	$k$ -step ahead prediction . . . . .	153
7.3	Application . . . . .	154
7.3.1	Single zone analysis . . . . .	154
7.3.2	Joint zone analysis . . . . .	155
7.4	Discussion and limitations . . . . .	158
<b>8</b>	<b>Conclusions and future work</b>	<b>164</b>
8.1	Conclusions . . . . .	164
8.2	Future work . . . . .	168

<b>A</b>	<b>Supplementary information</b>	<b>171</b>
A.1	Useful probability distributions . . . . .	171
A.1.1	Gamma distribution . . . . .	171
A.1.2	Weibull distribution . . . . .	172
A.1.3	Beta prime distribution . . . . .	172
A.1.4	Lognormal distribution . . . . .	173
A.1.5	Inverse-gamma distribution . . . . .	173
A.1.6	Negative binomial distribution . . . . .	174
A.1.7	Poisson distribution . . . . .	174

# Chapter 1

## Background

### 1.1 Introduction

Road safety is specifically mentioned in the United Nation's (UN) Sustainability Development Goal (SDG) 3.6 ([UN General Assembly, 2017](#)):

‘By 2020, halve the number of global deaths and injuries from road traffic accidents’,

and, UN SDG 11.2 ([UN General Assembly, 2017](#)):

‘By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons’.

Improving road safety is an important objective across the world. Approximately 1.3 million people are killed each year as a result of road traffic collisions worldwide, resulting in road traffic collisions being the 8th leading cause of death globally and the leading cause of death for children and young adults aged 5–29 years ([WHO, 2020](#)). The emotional toll of road traffic injuries is immense, affecting individuals and their families, and also causing substantial economic losses to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. In 2020, a fatal accident was calculated to cost in excess of £2 million in the UK; annually, the cost burden of road traffic

collisions in the UK is estimated to be in the billions of pounds ([Department for Transport, 2021](#)). In Europe, the ‘Vision Zero’ initiative sets the target of zero fatalities by 2050 ([Swedish Transport Administration, 2012](#)). February 2020 saw the introduction of the Stockholm Declaration at the Third Global Ministerial Conference on Road Safety: *Achieving Global Goals 2030*. Setting out a plan to achieve a 50% reduction in road deaths globally by 2030, and moving toward ‘Vision Zero’.

Clearly, it is in everyone’s best interest to reduce the number of collisions on the road. Ways to do this include education, enforcing traffic laws to reduce dangerous driving, such as drink driving or speeding, and improving infrastructure. This is reflected in road deaths being three times higher in low-income countries than in high-income countries. The number of deaths is still rising in these low to low-medium income countries, as opposed to high-income countries that are seeing a decrease. These statistics highlight the necessity of well-funded infrastructure for road safety and the importance of improving road safety in poorer countries. This comes with challenges as low-income countries don’t tend to have a vast amount of data readily available, like in richer countries, hence it is vital to find ways to improve road safety without relying on large, informative datasets. Statistics plays a crucial role in the improvement of infrastructure. By analysing collision data, experts can identify high-risk areas and patterns of incidents, which can inform the design and implementation of road safety measures. The evaluation of the effectiveness of road safety interventions, such as speed cameras, can inform the development of new strategies to improve road safety. The prediction of future road traffic collisions can help determine the most effective placement of traffic lights, speed bumps, and road signs. Overall, statistics provide valuable insights that can inform evidence-based policies and interventions to improve road safety and reduce the human and economic costs of road traffic collisions.

### 1.1.1 Statistics and road safety

In road safety, statistical methods are commonplace. As statistical methods improve, the opportunities for road safety organisations grow just as rapidly. However, in our experience, many of these organisations do not have the software or resources to use the most up-to-date methods and are using outdated techniques

in their analyses. With modern technological advances made in recent decades, the ease of data collection has led organisations to amass vast quantities of data. Coined ‘Big Data’, this resource has outpaced current, common-practice analytical tools and methods for analysing and forecasting data within reasonable time frames. Therefore, further work must be done to bridge the gap between advanced statistical methods used in academia and those used within organisations day-to-day. *Road safety GB* (RSGB) is a national road safety organisation made up of representatives from groups across the UK, including local government road safety teams. RSGB recognised the importance of being data-driven and established their annual data conference, ‘Joining the dots’, which aims to bring road safety experts – both academics and practitioners – together to showcase new developments and promote better practice. Furthermore, the *American Statistical Association’s* (ASA) *Transportation Statistics Interest Group* (TSIG) was established in 2002 with a vision to become a leading organisation of experts involved in applying statistical methods to transportation-related issues. Additionally, data science and statistics sessions at large transport conferences such as the *Transportation Research Board* (TRB) annual meeting have become more prominent.

### 1.1.2 Road safety schemes

In an attempt to reduce collisions and to make roads safer for all road users, road safety schemes can be introduced. Many schemes are developed to change driver behaviour, such as speed cameras and traffic calming measures; others work to improve the safety of the road itself, such as improving the drainage or the inclusion of crash barriers. Those tasked with implementing and analysing these schemes are often practitioners within road safety organisations. Practitioners wish to implement and evaluate the effectiveness of such schemes with the potential of reducing collisions, fatal or otherwise. Usually these schemes are introduced at carefully chosen locations, ones which have an unusually high number of collisions, considering the road type and traffic flow. Typically, roads with a higher traffic flow will generally see more collisions and so it is important to choose locations contextually, relative to collision rates at similar locations. This project will be focused on the improvement of assessing the effectiveness of safety schemes, and aiding practitioners in their evaluations. The initial issue of determining the effectiveness of a given road safety

treatment, or scheme evaluation, is choosing the optimal location for the scheme. This requires extensive road safety data for, not only the proposed location, but those similar to best inform the subsequent decision-making. Once implemented, we wish to evaluate whether the scheme has been successful in improving road safety, i.e. see a reduction in collisions/fatalities. Unfortunately, this usage of road safety data can be problematic and lead to biased conclusions if the data are not handled correctly, an effect felt most keenly where the datasets are relatively small.

Various road safety schemes are widely implemented to reduce road traffic accidents. Some examples include the use of speed cameras, leading pedestrian intervals (LPIs) at intersections, and vehicle activated signs. Speed cameras are designed to detect and capture images of vehicles that are exceeding the speed limit. The aim of these cameras is to encourage drivers to slow down and comply with speed limits. Leading pedestrian intervals (LPIs) are signals at intersections that give pedestrians more time to cross before vehicles are allowed to move forward. Vehicle activated signs are electronic road signs that are triggered by the presence of vehicles, alerting drivers to potential hazards such as sharp curves, speed limit changes, or pedestrian crossings. These signs can help to increase driver awareness.

### 1.1.3 Prediction

Within road safety, current practice for treating such ‘hotspots’ is almost always reactive: once a threshold level of collisions has been overtopped during some pre-determined observation period, treatment is applied (e.g. road safety cameras). However, more recently, methodology has been developed to predict collision counts at potential hotspots in future time periods, with a view to a more proactive treatment of road safety hotspots. Identifying road safety hotspots based on predicted (rather than observed) counts avoids the need to wait until collisions (and hence casualties/fatalities) occur before applying road safety schemes. Of course, you must have faith in the underpinning methodology, and the predictions produced by the statistical models used. The usefulness of implemented forecasting models is always questionable, as mentioned by [Box and Draper \(1987\)](#), ‘Essentially, all models are wrong, but some are useful’. Statistical methods for prediction are prominent in road safety and have become much more accessible with the advancement of

computers. More recently, practitioners have access to Bayesian analyses without the need to fully understand the underpinning statistics. In road safety literature, we refer to accident prediction models (APMs), and safety performance functions (SPFs). These are typical regression models using characteristic information in order to determine what a ‘typical’ collision count at a site displaying a given set of characteristics would be, in the case of a APM; or to evaluate the safety of a transportation system in an SPF. Time-series models are also widely used in collision count forecasting. These include use of autoregressive integrated moving average (ARIMA) models and its extensions such as SARIMA and SARIMAX, where seasonal factors and exogenous explanatory variables are included (Chen and Tjandra, 2014, Ihueze and Onwurah, 2018, Parvareh et al., 2018). ARIMA-based models are popular univariate statistical methods used for short-term forecasting purposes. Crash prediction model (CPM) studies commonly employ generalised linear models (GLMs) and are proven to be successful as they effectively model the rare, random, sporadic, and non-negative collision data. The generalised linear regression methods for CPM development mainly include Poisson regression and its various extensions, such as zero-inflated Poisson regression, negative binomial regression, and Poisson lognormal regression (El-Basyouny and Sayed, 2009a, Miaou et al., 1992, Sawalha and Sayed, 2006, Shankar et al., 2003).

#### 1.1.4 Aims and objectives

Throughout this research, we are committed to pioneering innovative analytical strategies for road traffic collision data. Our central ambition is to enhance the assessment of before/after safety schemes and to elevate the precision of predicting collision hotspots. To realise this, we will harness the intrinsic characteristics and spatial nuances of the data, thereby negating the need for additional external data, which isn’t always accessible. Critically, our approach will delve into understanding the nuances of prevalent road safety assessment techniques and introduce unique modelling methods for unconventional before/after studies.

Our work will cover two main areas: (1) enhancing road safety hotspot prediction techniques, that require minimal covariates, through Bayesian hierarchical modelling and state-space models; (2) rigorous evaluation of interventions (e.g. speed cameras)

at road safety hotspots identified for treatment through (1). To-date, most methods for hotspot prediction rely on regression methods proposed in the evaluation literature, so we will discuss the general methodology for road safety scheme evaluation first. Classically, safety scheme evaluations are reliant on tried-and-tested empirical/fully Bayesian methodology to account for confounding factors such as RTM and trend, such methods will be reviewed in Chapter 3. However, some before/after studies require more bespoke considerations, demonstrations of such will be provided in Chapter 4. In terms of prediction, we propose the use of a dynamic linear model (DLM) which is allowed to evolve over time and capture the confounding factors using linear modelling in the observation equation.

## 1.2 Industrial collaborators

### 1.2.1 PTV Group

PTV, based in Karlsruhe, Germany, are the market leader in the provision of software for transport planning and logistics. Formed as a spin-off company from Karlsruhe Institute for Technology in 1979, they have always valued academic credibility, and our collaboration with PTV through this PhD project is part of a long history of successful PTV academic collaboration (for example, with Universities in Germany, Switzerland, the US and South Africa). In our case, PTV have academic support from statisticians at Newcastle University to help develop novel statistical models in applications such as collision prediction, with the development of any such methodology being seriously scrutinised through peer-reviewed publications. At Newcastle, we benefit from access to PTV software, datasets and – crucially – translational and impact-oriented research through PTV’s software and their network of clients around the world. PTV have been assisting with the project in several significant ways, including:

- Project co-supervision — Dr. Karsten Kremer (Senior Manager Engineering).
- Hosting student industrial secondments.
- Free access to relevant traffic simulation software packages (e.g. VISIM and VISUM).
- Access to real-time traffic data.



Newcastle University has a long-standing collaborative arrangement with PTV and they have supported (both financially and in-kind) PhD research at Newcastle in the quantitative analysis of traffic data. Through their collaborative links with PTV, the supervisory team have access to vast amounts of data. The methods developed thus far for the prediction of road safety hotspots use just a fraction of this data (specifically, site-specific data pertaining to a time series of discretely-observed collision counts with associated predictor variables such as traffic volume, speed limit, average observed speed etc.). In the use cases developed by the supervisory team, collision counts for potential hotspots are usually annual, or at most monthly, allowing crude estimation of trend and seasonal effects. Through recent advances in the development of methods for analysing Big Data, PTV are confident that the methods developed will be scalable to data disaggregated to a much higher level of granularity and over a much larger geographical area. With this in mind, PTV provided training on using an SQL server and data acquisition techniques. This enabled us to gain data on every police-reported road traffic collision in the UK over an 11-year period. This dataset, outlined in Section 1.3.4, was subsequently utilised to gain valuable insights into the spatial and seasonal patterns of collisions across the UK (see Chapter 5).

PTV have software devoted to road safety in VISIM safety, within this software they allow for scheme evaluation techniques and prediction. It is, therefore, in their interest to include new and relevant statistical techniques in their software and support projects such as this one for the advancement of their software. They aided this project through providing data, industrial knowledge and training.

## 1.2.2 Gateshead Council

The Newcastle road safety team, comprising of academic staff, doctoral students and post-doctoral researchers, have a long-standing collaborative arrangement with the road safety data analysis team at Gateshead Council. Historically, members of both teams have sat on steering groups and road safety committees to ensure a co-ordinated, data-driven approach to decision-making around the deployment of road safety schemes and their subsequent evaluation. Gateshead Council have awarded the Newcastle team several rolling grants, over a period of eight years, to help develop user-friendly software tools to facilitate the deployment of sophisticated

modelling techniques in their practice. Within the remit of this PhD project, one aim was to develop a strategy for evaluating vehicle activated signs (see Chapter 4), although the pandemic meant the application of our method was paused and instead we illustrate our approach using simulated data. Gateshead have been instrumental in the development of the statistical methodology, providing expert insight and test datasets. The collaboration with Gateshead Council is ongoing.

### 1.2.3 Florida and New York State Departments of Transportation

Members of the Newcastle road safety team met representatives from Florida State Department of Transportation, and indeed other road safety practitioners in the United States, at several annual meetings of the Transportation Research Board in Washington, D.C., between 2016–2019. Collaborations have included a comparison of standard and more novel statistical models for road safety hotspot identification (e.g. [Guo et al., 2019](#)) and the provision of many datasets – including the Florida collision dataset described in Section 1.3.2 and used in Chapters 6 and 7, and a dataset for a before/after analysis of road safety cameras in New York City.

## 1.3 Datasets

In this thesis, we will utilise various real-world datasets to illustrate the discussed techniques and showcase the results that can be achieved.

### 1.3.1 Northumbria Safety Camera Partnership data

The *Northumbria Safety Camera Partnership* (NSCP) commissioned an investigation into the impact of mobile safety cameras on the demand for secondary health care at the region’s hospitals. A mobile camera is a portable unit which is operated at designated sites. The group collected data from 67 treated sites in the region from a before period (April 2001–March 2003) and an after period (April 2004–March 2006) and from 67 reference sites in the Northumbria Police force area. At each of the treated sites, the number of casualties before and after the implementation of the safety camera was observed, as well as numerous covariates:

- Average observed speed ( $x_1$  miles per hour).

- Percentage of drivers exceeding the speed limit ( $x_2$ ).
- Daily traffic flow ( $x_3$  in thousands).
- Speed limit ( $x_4$  miles per hour).
- Eighty-fifth percentile speed ( $x_5$  miles per hour).
- Percentage of drivers at least 15mph over the speed limit ( $x_6$ ).
- Road classification ( $x_7 = \text{A/B/C/Unclassified roads}$ ).
- Road type ( $x_8 = \text{single/dual/mixed carriageway}$ ).

The dataset contains a single casualty count and single observations for each covariate in the before period and then a single casualty count in the after period at each treated site. Continuous covariates were averaged over their respective 2 year period, categorical covariates remained constant throughout the data, and casualty counts were aggregated to give the total number of collisions in the before and after treatment periods, separately. In Chapter 3 we use a subset of 56 of the 67 mobile camera sites in the original study (due to missing values). The sites vary but all have an abnormally high number of casualties in the before period, which is why they were chosen for treatment. Overall there were 438 casualties before treatment; this reduced to 298 after the mobile cameras were implemented. After checking the dataset for multicollinearity between covariates and backwards stepwise regression to remove any non-significant covariates, we remove speed limit, percentage of drivers 15mph over the speed limit, road type and the 85th percentile speed. We therefore retain  $x_1$ ,  $x_2$ ,  $x_3$  and  $x_7$ . Summary statistics for this dataset (along with casualty counts  $y$ ) are provided in Table 1.1. The table presents a before/after comparison between the treated sites before and after intervention, and the comparison sites. The treated sites initially had higher average speeds, a greater percentage of drivers exceeding the speed limit, and more collisions compared to the untreated (comparison) sites; however, post-treatment, these metrics improved, suggesting that the treatment had a positive impact in fostering safer driving conditions, although they have not yet reached the safer baseline levels observed in the untreated sites.

		$x_1$	$x_2$	$x_3$	$x_{7A}$	$x_{7B}$	$x_{7C}$	$x_{7U}$	$y$
Treated (before)	Mean/Prop.	36.63	38.45	6.92	0.50	0.23	0.18	0.09	7.79
	S.D.	9.90	21.67	4.03	-	-	-	-	5.38
Treated (after)	Mean/Prop.	34.58	34.06	7.58	-	-	-	-	5.32
	S.D.	9.64	20.03	5.05	-	-	-	-	4.02
Comparison	Mean/Prop.	32.95	35.73	9.13	0.37	0.16	0.30	0.16	4.28
	S.D.	6.89	29.63	6.48	-	-	-	-	4.77

Table 1.1: Table showing the mean and standard deviations of the continuous covariates, and proportion for categorical covariates from the Northumbria before/after dataset for the treated and comparison pools of sites.

### 1.3.2 Florida State Department of Transport data

The dataset used to introduce the idea of seasonal and spatial effects in collision modelling in Chapters 6 and 7 was provided by the Florida State Department of Transport. It contains collision rates (collisions per vehicle kilometre travelled) at 49 traffic analysis zones (TAZs) across the state of Florida, U.S.A., over a period of 46 years (1960–2015) inclusive. However, a significant amount of data are missing, making any kind of longitudinal study impossible. Within each zone a number of sites were tracked and the rate of collisions per zone was calculated as the average number of collisions across those sites. The collision rates are disaggregated by month, allowing the inclusion of a seasonal component in the model, and the dataset contains the longitude and latitude of the centroid of each TAZ, enabling spatial effects to be investigated. To search for seasonal patterns in the data, we found the mean collision rate for each month individually for each zone. This is shown in the left-hand-side of Figure 1.2. The majority of zones exhibit sinusoidal patterns over a 12-month period. The right-hand-side of Figure 1.2 shows histograms of the collision rates at a select few zones, they suggest that the rates could be reasonably assumed Gaussian. In Chapter 7 we use a subset of this dataset.

### 1.3.3 Canadian leading pedestrian interval data

With the safety of pedestrians in mind, safety treatments at signalised intersections have been investigated. ‘*The leading pedestrian interval (LPI) is one treatment that has been implemented at signalised intersections to permit pedestrians to begin crossing several seconds before the release of conflicting vehicle movements*’ (Van Houten et al., 2000). We have data available from Transport Canada’s National Collision

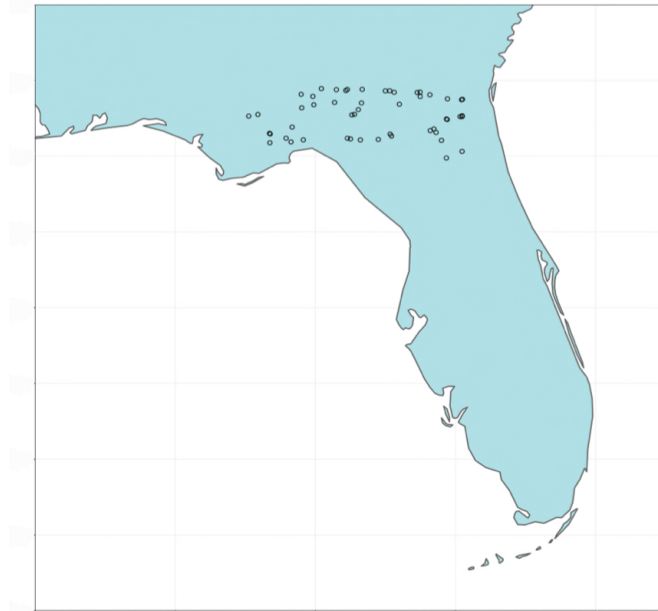


Figure 1.1: Centroids of the 49 zones overlaid onto a map of Florida.

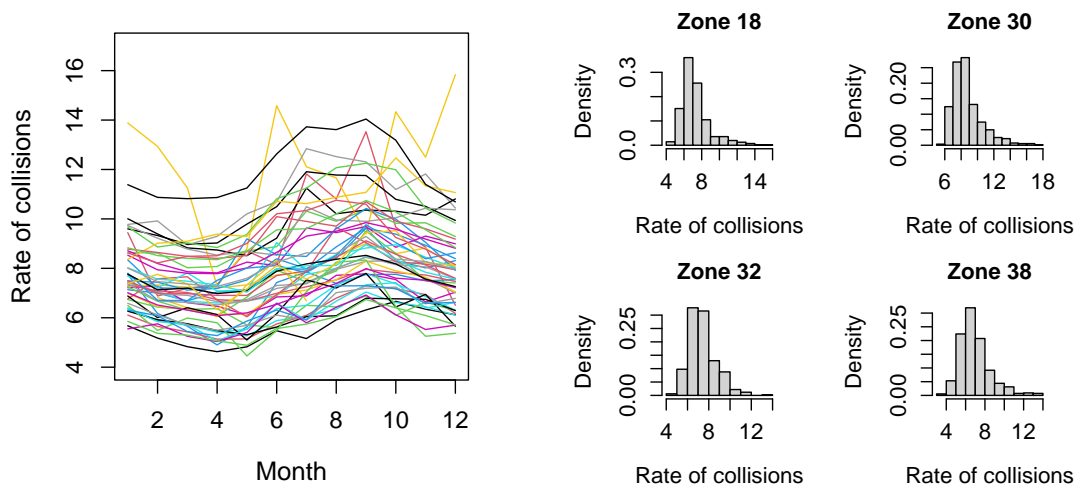


Figure 1.2: Left: Monthly mean rates of collisions for each of the 49 zones. Right: Histograms of the rates of collisions for zones 18, 30, 32, 38.

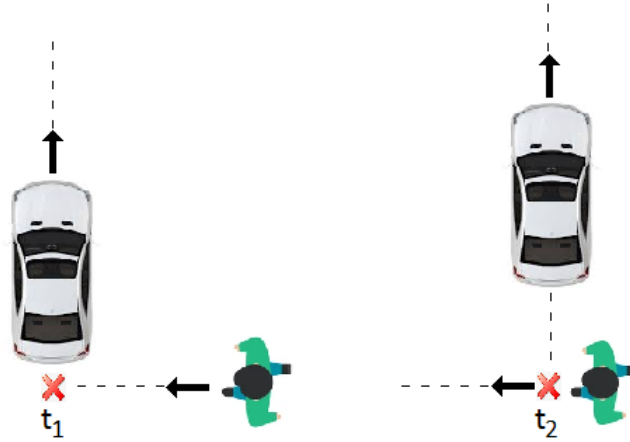


Figure 1.3: Post-encroachment times (PET),  $PET = t_1 - t_2$

Database. A number of Intersections in downtown Vancouver (see Figure 1.1) – e.g. Granville Street at Smithe Street – have had LPIs of 5s introduced to give pedestrians more time to cross before a left-turning car is released. At each intersection, one crosswalk has been treated. Data were collected over 8 months, 4 months in the before period, and 4 months in the after period in the same 12-hour period (8:00–20:00) for the before and after period to mitigate potential confounders such as seasonal variation. Data spanning exactly the same time period are also available for a further seven intersections that have not been treated with the LPI intervention, for comparison purposes. Studies employ specific thresholds of conflict indicators such as post-encroachment time (PET) to identify near-misses. PET is the time (in seconds) between the moment the first road user passes the conflicting point,  $t_1$ , and the moment the second user reaches that point,  $t_2$ . The positions of the vehicle and pedestrian are shown in Figure 1.3. We have the minimum PET in 10 minute intervals over 12 hours per day. Video data are processed with the automated traffic conflict analysis system using computer vision techniques (Saunier and Sayed, 2007, Zaki et al., 2020), with conflicts between pedestrians and left-turning vehicles at crosswalks being automatically extracted. PETs  $< 15$ s were recorded. As  $PET = t_1 - t_2$ , if  $t_1 = t_2$  then we have a collision between pedestrian and vehicle. Smaller PET values imply a near-miss, and a value close to zero would imply a dangerous situation. Figure 1.5 compares the densities of PETs from the before and after treatment time periods at sites 2 and 8. For site 8 there is a noticeable shift in the density plot towards larger values in the after period, suggesting a treatment effect. This shift in the distribution suggests that the treatment has had a positive

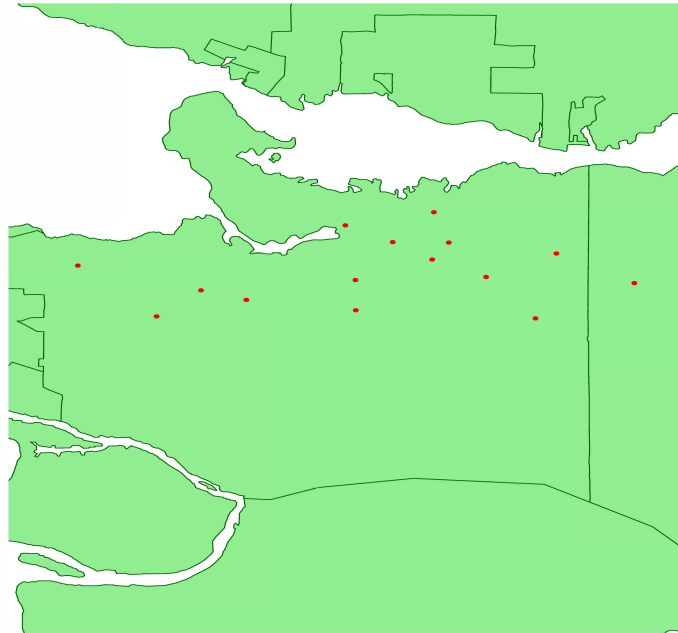


Figure 1.4: Intersection locations overlaid onto a map of Vancouver, Canada.

impact and lead to a reduction in traffic conflicts. At site 2 where no treatment has occurred, the PETs from the before and after periods are very similar with the majority of the density curves overlapping.

### 1.3.4 UK STATS19 data

The United Kingdom's Department for Transport collects data on road traffic incidents for Great Britain through STATS19 – a database comprising information on all personal injury road traffic accidents (RTAs) occurring on public highways and reported to the police within 30 days. The statistics relate only to personal injury accidents on public roads that are reported to the police, and subsequently recorded, using the STATS19 accident reporting form. STATS19 data currently provides the most robust, complete, and detailed annual statistics for road casualties across Great Britain ([Department for Transport, 2022](#)). STATS19 includes information on how many cars/pedestrians were involved, junction detail, road type, road number, longitude and latitude coordinates, urban or rural area, local authority district amongst many more contributing factors for each collision.

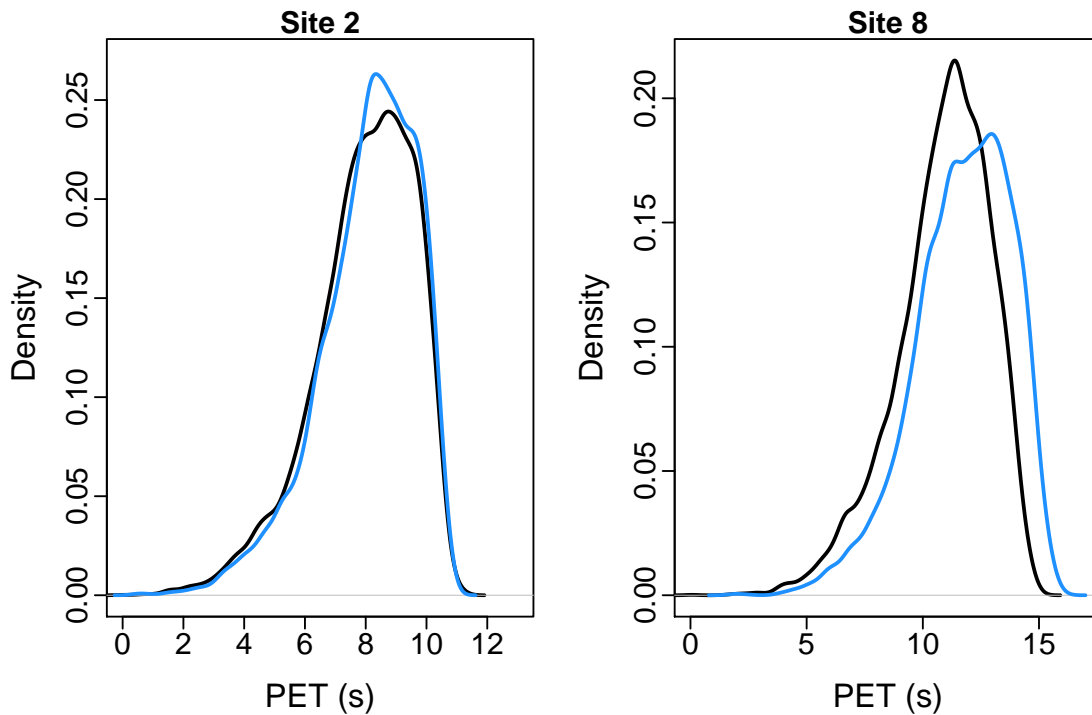
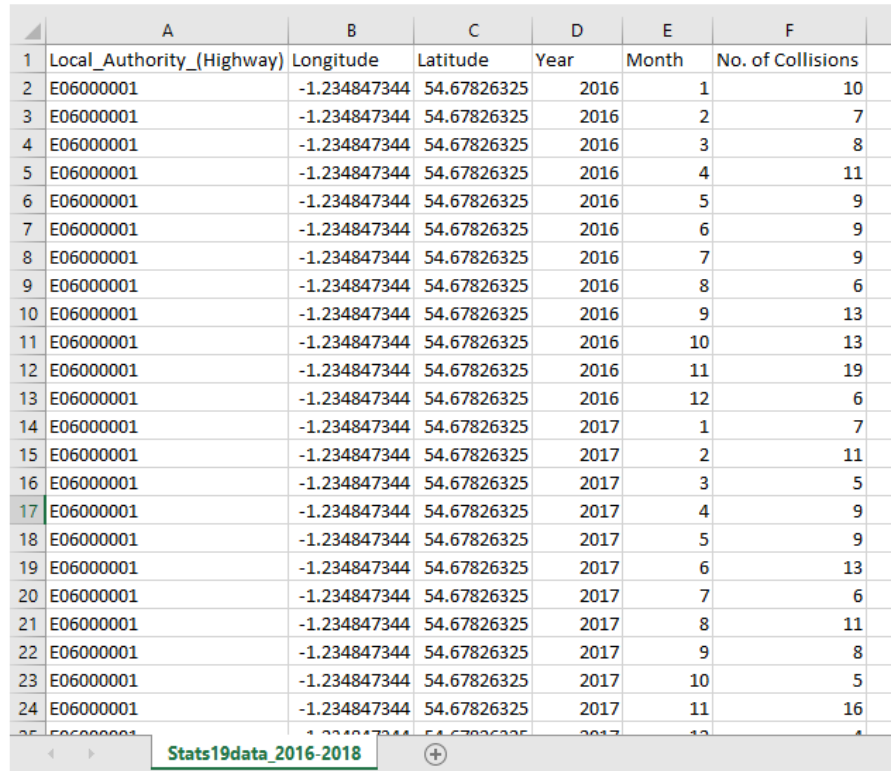


Figure 1.5: Density plots of PETs (s) at sites 2 (left) and 8 (right) in the before period (black) and after period (blue).

With help and training through multiple visits to PTV, we have been able to take downloaded STATS19 data and put it into a Microsoft SQL dataframe with counts of collisions per month per zone (as many years as necessary can be added). The SQL code creates a formatted empty table which can then take the downloaded STATS19 data, adding each year’s available data. From the STATS19 data within SQL management studio, we can group ‘zones’ through, for example, ‘Local authority highway’, ‘Local authority district’, or location ‘LSOA of accident location’. We then format date, longitude and latitude to allow us to count the number of collisions in each zone per month per year. This can then be exported to an excel spreadsheet, an example of which is shown in Figure 1.6.

STATS19 data was collected from Jan 2009–Dec 2019. Using an SQL server, we aggregated the data to give collision counts per month within a zone. Zones here are defined by their local authority highway ID, in which there are 207. There is the option to aggregate the data finer, for example, collision count per day or hour, however monthly data was chosen to omit zeros. In the dataset, we have included average longitude and latitude over all collisions in each zone to include a





	A	B	C	D	E	F
1	Local_Authority_(Highway)	Longitude	Latitude	Year	Month	No. of Collisions
2	E06000001	-1.234847344	54.67826325	2016	1	10
3	E06000001	-1.234847344	54.67826325	2016	2	7
4	E06000001	-1.234847344	54.67826325	2016	3	8
5	E06000001	-1.234847344	54.67826325	2016	4	11
6	E06000001	-1.234847344	54.67826325	2016	5	9
7	E06000001	-1.234847344	54.67826325	2016	6	9
8	E06000001	-1.234847344	54.67826325	2016	7	9
9	E06000001	-1.234847344	54.67826325	2016	8	6
10	E06000001	-1.234847344	54.67826325	2016	9	13
11	E06000001	-1.234847344	54.67826325	2016	10	13
12	E06000001	-1.234847344	54.67826325	2016	11	19
13	E06000001	-1.234847344	54.67826325	2016	12	6
14	E06000001	-1.234847344	54.67826325	2017	1	7
15	E06000001	-1.234847344	54.67826325	2017	2	11
16	E06000001	-1.234847344	54.67826325	2017	3	5
17	E06000001	-1.234847344	54.67826325	2017	4	9
18	E06000001	-1.234847344	54.67826325	2017	5	9
19	E06000001	-1.234847344	54.67826325	2017	6	13
20	E06000001	-1.234847344	54.67826325	2017	7	6
21	E06000001	-1.234847344	54.67826325	2017	8	11
22	E06000001	-1.234847344	54.67826325	2017	9	8
23	E06000001	-1.234847344	54.67826325	2017	10	5
24	E06000001	-1.234847344	54.67826325	2017	11	16
25	E06000001	-1.234847344	54.67826325	2017	12	4

Figure 1.6: Excel spreadsheet showing the aggregated data exported from the SQL server.

spatial term. Another advantage of using STATS19 data, is the numerous variables available that could be included into our modelling. Figure 1.8 shows time-series plots for four of the UK sites over the 11 year recording period.

## 1.4 Thesis structure

The focus of this thesis is to analyse two primary responsibilities performed by road safety professionals: assessing the effectiveness of road safety schemes and predicting collision hotspots. The aim is to develop methodologies that are versatile and widely applicable. The methods introduced for forecasting are focused on exploiting patterns in the observed data to achieve sensible results when there is a lack of covariate information. The modelling includes working with collision counts, collision rates and near-misses, hence, the type of data is not restricted and models are introduced for a variety of datasets.

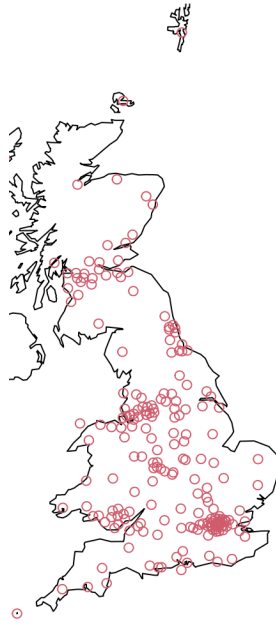


Figure 1.7: Centroids of the 207 zones overlaid onto a map of the UK.

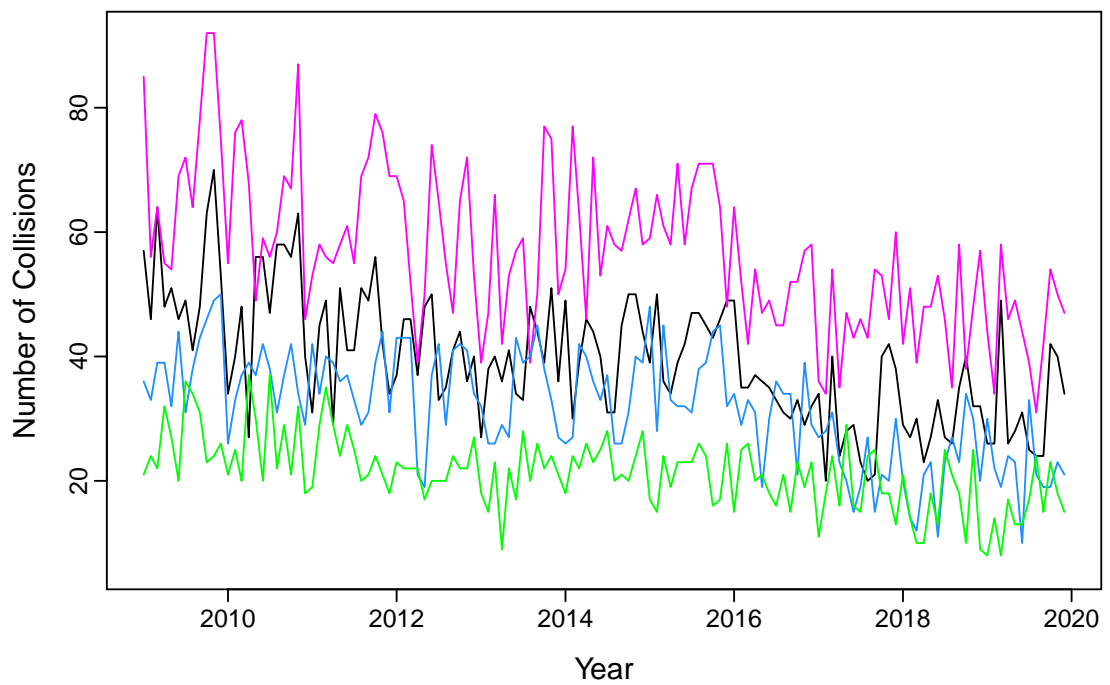


Figure 1.8: Collision counts by month from January 2009–December 2019 in Gateshead (black), Newcastle (pink), North Tyneside (blue), South Tyneside (green)

In Chapter 2, we begin with an overview of the methods used throughout the thesis and include illustrative examples where necessary. Chapter 3 discusses commonly used safety scheme evaluation methods and investigates their advantages and disadvantages. A sensitivity analysis is performed on the estimates of treatment effect, which illustrates the importance of model selection. Chapter 4 then introduces bespoke before/after safety scheme evaluation models for data that are collected in a way that varies from those modelled in Chapter 3. We firstly introduce a Poisson regression model for collision counts where the treated sites are randomly selected. Then we introduce a model for near-miss data which justifies the use of the extremal types theorem, with adjustments to the usual modelling framework to account for temporal dependence and treatment effects. Chapter 5 introduces spatial extreme value modelling methods for the near-miss data used in Chapter 4. We propose a latent variable model to introduce spatial variation in the parameters as well as capturing the treatment effect through linear modelling of these parameters. Chapter 6 turns attention to hotspot prediction and proposes a Bayesian hierarchical model which allows us to segregate the seasonal and zonal effects alongside capturing the uncertainty in the parameters sufficiently. We provide the within-sample predictive distributions for posterior predictive checks, which can be used to investigate the model fit. This random effects model is applied to two real datasets, with slight alterations. Chapter 7 introduces a joint spatio-temporal model of collision rates over multiple zones. A dynamic linear model is used at the level of a single zone, and allows for seasonality via a single harmonic with time-varying amplitude and phase parameters. We then account for spatial dependence at nearby locations by adding a spatial Gaussian process to the system equation, thereby smoothing spatial deviations from the underlying temporal process. The resulting model allows for both within- and out-of-sample forecasting for locations which are fully observed and for locations at which some data are missing. Conclusions are drawn in Chapter 8, where we also discuss potential further work.

The main research contributions of this thesis include a new model to evaluate before/after safety schemes where the treated sites are randomly assigned; using a bivariate threshold excess model to conduct a traffic conflict-based before/after safety evaluation using the extreme value theory where the treatment effects were measured through linear modelling in the scale parameter (Chapter 4). An analysis on the traffic-conflict data was performed where spatial dependence was accounted

for through the inclusion of a latent spatial process in extreme value parameters to evaluate the effectiveness of a safety scheme with linear modelling of the parameters to capture treatment effect (Chapter 5). The use of random effects methodology to capture zonal and seasonal effects within a Bayesian framework (Chapter 6); using a sinusoidal dynamic linear model with Gaussian processes to create a spatio-temporal model available for hotspot prediction (Chapter 7).

## 1.5 List of publications

- N. K. Hewett, A. Golightly, L. Fawcett and N. Thorpe, *Bayesian inference for a spatio-temporal model of road traffic collision data*. Pre-print available at: arXiv:2302.00342 (2023), <https://arxiv.org/abs/2302.00342>
- N. K. Hewett, L. Fawcett, A. Golightly, and N. Thorpe, *Using extreme value theory to evaluate the leading pedestrian interval road safety intervention*. **Writing-up in progress**

# Chapter 2

## Methods

### 2.1 Generalised linear modelling

In statistics, linear regression is an approach for modelling linear relationships between a response variable,  $Y_i$ , and one or more covariates,  $x_{1i}, x_{2i}, \dots, x_{pi}$ , where  $i$  indicates the index of each observation and  $p$  denotes the number of covariates. It assumes that the relationship between the response variable and covariates is linear. The response variable and covariates can take various forms, including but not limited to continuous, positive continuous, discrete, and count data formats. To predict the response, a typical model often includes a combination of such covariates. An example of a linear regression equation with a single covariate is,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{2.1}$$

where  $\beta_0, \beta_1$  are unknown parameters controlling the intercept and slope, respectively. The intercept is the value of the response variable at the point where the regression line crosses the  $y$ -axis. The slope term represents the rate of change of the response variable in response to a change in the covariate,  $x_i$ , and hence determines the steepness of the regression line. The error term,  $\epsilon_i$ , accounts for the uncertainty in the response variable and is typically assumed to follow a normal distribution with zero mean and a constant variance,  $\sigma^2$ . Therefore, Equation (2.1) can be written as

$$Y_i \sim N(\mu_i, \sigma^2),$$

where  $\mu_i$ , often termed the *linear predictor*, can be written as  $\mu_i = \beta_0 + \beta_1 x_i$ .

A generalised linear model (GLM) is an extension of the linear model that introduces a link function to relate the linear predictor to the response variable. There are many possibilities where this is appropriate, for example, when the relationship between the response and explanatory variables is non-linear, or the variance is not constant, or if the response is discrete and/or strictly positive. GLMs use a link function  $g(\cdot)$  to relate the linear predictor to the response variable, which allows the response variable to take on a wider range of distributions than the normal distribution assumed in linear regression. GLMs were formulated as a way of unifying various statistical models, including linear, logistic and Poisson regression (Zhao, 2012). The general form of a GLM is,

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi},$$

where  $g(\cdot)$  is the link function that relates the expected value of the response variable, denoted as  $\mu_i$ , to the linear predictor. Poisson regression is a type of generalised linear model used to analyse count data and contingency tables. It is assumed that the response variable  $Y_i$  follows a Poisson distribution, and assumes the logarithm of its expected value can be expressed as a linear combination of unknown parameters. The basic form is,

$$Y_i \sim Po(\lambda_i),$$

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip},$$

where  $\lambda_i$  is the expected number of occurrences of the event for the  $i$ th case, and  $x_{ij}$  is the value of the  $j$ th predictor variable for the  $i$ th case. The  $\beta$  values are the regression coefficients that represent the effect of each predictor variable on the response variable.

We can formulate models for all  $n$  observations  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ , with covariate vector  $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})^T$ , for the  $i$ th observation and unknown parameter vector  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ , as

$$\begin{pmatrix} g(Y_1) \\ \vdots \\ g(Y_n) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

or

$$g(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta},$$

where  $\mathbf{X}$  is known as the design matrix.

In statistics, we are trying to find the model that best fits our observed data. Usually it is relatively easy to choose a probability distribution and the model parameters such that the theoretical population mean of the model is approximately equal to the sample mean. However, this isn't as straightforward for the variance, especially for simpler models. If the observed variance is more than the model variance, we have overdispersion. Conversely, if the variance is less than the mean, underdispersion has occurred. In the case of modelling count data, as is the case for most road traffic data, we use a Poisson distribution, with rate  $\lambda$ ,

$$X \sim Po(\lambda), \quad \lambda > 0,$$

where the mean and variance are defined as

$$\mathbb{E}(X) = Var(X) = \lambda.$$

Commonly this assumption is not met and the data have a variance that is proportional to its mean, such that

$$Var(X) = \theta\lambda, \quad \theta > 0.$$

Most commonly the variance is larger than the mean and so  $\theta > 1$  which means

$$\theta = 1 + \gamma,$$

where  $\gamma$  is known as the overdispersion parameter and controls the severity of the excess variance. If  $\theta < 1$  then we have underdispersion, and  $\theta = 1$  relates to the standard Poisson distribution.

A way to acknowledge the overdispersion in count data, is within a Bayesian framework where we allow the rate parameter to be a random variable drawn from a prior distribution. Then we have the same Poisson set-up for the data however, the distribution is now conditional on the rate parameter  $\lambda$ ,

$$X|\lambda \sim Po(\lambda), \quad \lambda > 0.$$

For mathematical convenience, we set the prior distribution of  $\lambda$  to be a gamma distribution which is conjugate to the Poisson distribution and thus we gain an analytical posterior distribution for  $\lambda$ . We have,

$$\lambda \sim Ga(\alpha, \beta),$$

with PDF

$$g(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp\{\beta\lambda\}, \quad \alpha, \beta > 0.$$

We are then able to find the unconditional distribution for  $X$ ,  $f(X)$ , by integrating out  $\lambda$ ,

$$\begin{aligned} f(x) &= \int_{\Lambda} f(x|\lambda)g(\lambda) \, d\lambda, \\ &= \int_0^\infty \frac{\lambda^x}{x!} \exp\{-\lambda\} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp\{\beta\lambda\} \, d\lambda, \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)x!} \int_0^\infty \lambda^{x+\alpha-1} \exp\{-(\beta+1)\lambda\} \, d\lambda. \end{aligned}$$

We notice that the integrand has the form of a gamma distribution proportional to  $Y \sim Ga(x + \alpha, \beta + 1)$  with PDF

$$f(Y) = \frac{(\beta+1)^{(x+\alpha)}}{\Gamma(x+\alpha)} y^{x+\alpha-1} \exp\{-(\beta+1)y\}, \quad y > 0.$$

As this is a PDF we have  $\int_0^\infty f(Y) \, dy = 1$ , ergo,

$$\begin{aligned} &\int_0^\infty \lambda^{x+\alpha-1} \exp\{-(\beta+1)\lambda\} \, d\lambda \\ &= \int_0^\infty \frac{(\beta+1)^{(x+\alpha)}}{\Gamma(x+\alpha)} \frac{\Gamma(x+\alpha)}{(\beta+1)^{(x+\alpha)}} y^{x+\alpha-1} \exp\{-(\beta+1)y\} \, dy, \\ &= \frac{\Gamma(x+\alpha)}{(\beta+1)^{(x+\alpha)}} \int_0^\infty \frac{(\beta+1)^{(x+\alpha)}}{\Gamma(x+\alpha)} y^{x+\alpha-1} \exp\{-(\beta+1)y\} \, dy, \\ &= \frac{\Gamma(x+\alpha)}{(\beta+1)^{(x+\alpha)}} \times 1, \\ &= \frac{\Gamma(x+\alpha)}{(\beta+1)^{(x+\alpha)}}. \end{aligned}$$



Hence,  $f(x)$  becomes

$$\begin{aligned} f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)x!} \times \frac{\Gamma(x+\alpha)}{(\beta+1)^{(x+\alpha)}}, \\ &= \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)\Gamma(x+1)} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^x, \\ &= \binom{y+\alpha-1}{y} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^x. \end{aligned}$$

We recognise this as the probability mass function (PMF),  $f(x) = Pr(X = x)$ , of a negative binomial random variable and so

$$X \sim NB\left(\alpha, \frac{1}{\beta+1}\right)$$

is the unconditional distribution for  $X$ .

## 2.2 Extreme value theory

Let  $X_n$  denote a stationary sequence of random variables with common distribution function  $F$ , and let  $M_n = \max\{X_1, \dots, X_n\}$ . It is typically the case that, as  $n \rightarrow \infty$ ,

$$Pr(M_n \leq x) \approx F^{n\theta}(x), \quad (2.2)$$

where  $\theta \in (0, 1)$  is known as the extremal index; see, for example, [Leadbetter and Rootzén \(1988\)](#). As  $\theta \rightarrow 0$  there is increasing dependence in the extremes of the process; for an independent process,  $\theta = 1$ , as  $Pr(M_n \leq x) = Pr(X_1, X_2, \dots, X_n \leq x) = Pr(X_1 \leq x) \times Pr(X_2 \leq x) \times \dots = F^n$ . Initially concerned with the independent case (i.e.  $\theta = 1$ ), classical extreme value theory (EVT) sought families of limiting models for  $F^n$ , without reference to the marginal distribution function  $F$  as any small discrepancies in  $F$  could lead to large discrepancies in  $F^n$ .

Examining the behaviour of  $M_n$  as  $n \rightarrow \infty$  gives rise to the Extremal Types Theorem (see [Fisher and Tippett \(1928\)](#), [Gnedenko \(1943\)](#)). The limiting distribution of  $M_n$  is degenerate, that is, the distribution converges to a single point on the real line with probability 1, this single point being the upper endpoint of  $F$ . This is analogous to the sample mean  $\bar{X}$  converging to the population mean  $\mu$  with certainty in the Central Limit Theorem. The Extremal Types Theorem states that

if there exists sequences of constants  $a_n > 0$  and  $b_n$  such that, as  $n \rightarrow \infty$ ,

$$Pr \{(M_n - b_n)/a_n \leq x\} \rightarrow G(x)$$

for some non-degenerate distribution  $G$ , then  $G$  is of the same type as one of the following distributions:

$$\begin{aligned} I : G(x) &= \exp\{-\exp(-x)\}, & -\infty < x < \infty; \\ II : G(x) &= \begin{cases} 0 & x \leq 0, \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \\ III : G(x) &= \begin{cases} \exp\{-(-x)^\alpha\} & x < 0, \alpha > 0, \\ 1 & x \geq 0. \end{cases} \end{aligned}$$

Distributions  $I$ ,  $II$  and  $III$  have become known as the Gumbel, Fréchet and Weibull types (respectively), and are known collectively as the extreme value distributions. For both the Gumbel and Fréchet distributions the limiting distribution  $G$  is unbounded; that is, the upper-endpoint tends to  $\infty$ . The Weibull distribution has a finite upper bound. It should be noted that the Extremal Types Theorem does not ensure the existence of a non-degenerate limit for  $M_n$ ; nor does it specify which of types  $I$ ,  $II$  or  $III$  is applicable if a limit distribution does exist (i.e. in which *domain of attraction* the distribution of  $G$  lies (Coles, 2001)). However, when such a distribution does exist, we find that, by analogy with the Central Limit Theorem, the limiting distribution of sample maxima follows one of the distributions given by the Extremal Types Theorem, no matter what the parent distribution  $F$ .

### 2.2.1 The generalised extreme value distribution

Von Mises (1954) and Jenkinson (1955) independently derived a distribution which encompasses all three types of extreme value distribution; the generalised extreme value distribution (GEV). The GEV is the limiting model for  $F^n$ , with distribution function (d.f.)

$$\mathcal{G}(x; \mu, \sigma, \xi) = \begin{cases} \exp \left[ -(1 + \xi(x - \mu)/\sigma)^{-1/\xi} \right]_+, & \xi \neq 0 \\ \exp \left[ -\exp(-(x - \mu)/\sigma) \right], & \xi = 0 \end{cases} \quad (2.3)$$

defined on  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ , where  $-\infty < \mu < \infty$ ,  $\sigma > 0$  and  $-\infty < \xi < \infty$

are parameters of location, scale and shape, respectively. It can be shown that  $F^{n\theta}$  is also GEV with d.f.  $\mathcal{G}(x; \mu^*, \sigma^*, \xi)$ , provided long-range dependence is negligible. Let  $X_1, X_2, \dots, X_n$  be the first  $n$  observations of a stationary series satisfying Leadbetter's  $D(u_n)$  condition, which ensures that long-range dependence is sufficiently weak so as not to affect the asymptotics of an extreme value analysis (Leadbetter et al., 1983), and let  $M_n = \max\{X_1, X_2, \dots, X_n\}$ . Now let  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$  be an independent series, with  $\tilde{X}$  having the same distribution as  $X$ , and let  $\tilde{M}_n = \max\{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n\}$ . Then if  $\tilde{M}_n$  has a non-degenerate limit law given by  $\Pr\{(\tilde{M}_n - b_n)/a_n \leq x\} \rightarrow G(x)$ , it follows that

$$\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G^\theta(x). \quad (2.4)$$

The effect of dependence is simply a replacement of  $G$  in Equation (2.3) as the limit distribution with  $G^\theta$ . If  $G$  corresponds to the GEV distribution with parameters  $(\mu, \sigma, \xi)$ , then

$$\begin{aligned} G^\theta(x) &= \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}^\theta \\ &= \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu^*}{\sigma^*} \right) \right]_+^{-1/\xi} \right\} \end{aligned}$$

where  $\mu^* = \mu - \frac{\sigma}{\xi}(1 - \theta^\xi)$  and  $\sigma^* = \sigma\theta^\xi$ . Thus, if the (approximate) distribution of  $M_n$  is GEV with parameters  $(\mu, \sigma, \xi)$ , then the (approximate) distribution of  $\tilde{M}_n$  is GEV with parameters  $(\mu^*, \sigma^*, \xi)$ . Hence, for block maxima, in practical terms short-range dependence can be ignored since the distribution of block maxima falls within the same family of distributions as would be appropriate if the series were truly independent. The extreme value distribution is used to model the largest or smallest value from a group or block of data. The GEV distribution can be described as the limiting distribution of block maxima or minima and can be fit to the data using statistical methods, see Sections 2.3 and 2.4.3. In practice, the GEV is used to model maxima over some convenient calendar unit – usually years. However, choosing a suitable block size can sometimes be problematic. It is crucial to select a block size that is sufficiently large to satisfy limiting conditions, yet not too large that there are too few extreme values from which to make inferences. In some applications, in which data span only a period of months, years would not

be appropriate; blocks would need to be large enough for the limiting results to hold, yet if they're too large there will be too few maxima from which to make inferences. Furthermore, this method can lead to a significant loss of valuable data since informative observations on the extremes are often discarded.

## 2.2.2 The generalised Pareto distribution

Pickands (1975) showed that for  $\theta = 1$  and  $u$  large,  $(X - u|X > u)$  follows a Generalised Pareto Distribution (GPD) with d.f.

$$\mathcal{H}(x; \tilde{\sigma}, \xi) = \begin{cases} 1 - (1 + \xi x/\tilde{\sigma})^{1/\xi}, & \xi \neq 0 \\ 1 - \exp(-x/\tilde{\sigma}), & \xi = 0 \end{cases} \quad (2.5)$$

defined on  $x > 0$ , with scale and shape parameters  $\tilde{\sigma}$  and  $\xi$  respectively. Here,  $\tilde{\sigma}$  is related to the parameters in the corresponding GEV distribution for block maxima through  $\tilde{\sigma} = \sigma + \xi(u - \mu)$ . Threshold methods classify observations as extreme if they exceed some high threshold, usually denoted  $u$ ; then, the GPD in Equation (2.5) is fitted to the excesses over this threshold. Graphical diagnostics are available for the selection of a suitable threshold, as will be demonstrated in Section 4.3. Unlike the case of modelling block maxima with the GEV, powering  $F^n$  by  $\theta$ , as described by Equation (2.2), does not lead to another extreme value distribution whose parameters have absorbed the extremal index; thus, careful consideration of extremal dependence is required (and, unlike consecutive block maxima, extremal dependence is usually present between consecutive threshold excesses).

## 2.2.3 Return levels

Once the extreme value distribution has been fit to the data, it can be used to estimate the probability of extreme events occurring in the future. Estimates of an extreme quantile  $z_r$  can be obtained by inversion of  $\mathcal{G}(z_r)$  or  $\mathcal{H}^\theta(z_r)$  to  $1 - r^{-1}$  and then solving for  $y = z_r$ , where  $z_r$  is the  $r$ -observation return level associated with return period  $r$ . Termed return levels as they refer to *levels* that might be *returned* on average at least once every  $r$  years, and high quantiles of the fitted distribution

are used as estimates of these return levels. Specifically,

$$z_r = \begin{cases} \mu^* - \frac{\sigma^*}{\xi} \left[ 1 - (-\log(1 - r^{-1}))^{-\xi} \right], & \xi \neq 0 \\ \mu^* - \sigma^* \log(-\log(1 - r^{-1})), & \xi = 0, \end{cases}$$

when modelling block maxima. When modelling threshold excesses:

$$z_r = \begin{cases} u + \sigma \xi^{-1} [(\lambda_u^{-1} \omega_r)^\xi - 1], & \xi \neq 0 \\ u + \sigma \log(\lambda_u^{-1} \omega_r), & \xi = 0, \end{cases}$$

where  $\omega_r = 1 - [1 - (rn_y)^{-1}]^{1/\theta}$ ,  $\lambda_u$  is the rate of threshold excess and  $n_y$  is the (average) number of observations per year. We will make use of EVT in our analysis of the leading pedestrian interval data from Canada (see Section 4.3).

## 2.3 Maximum likelihood estimation

Maximum likelihood estimation is a method of estimating the parameters of a statistical model. Given a set of observations, this method seeks to find the parameter values that maximise the likelihood of obtaining the observations. To use the maximum likelihood estimation method, we first need to define a statistical model that describes the process that generated the data. This model will include one or more parameters that we want to estimate. The likelihood function then needs to be defined, which is a function that explains how well a parameter explains the observed data through the product of the individual probability density functions (PDF) at each observation to gain the joint distribution. The likelihood function is used to estimate the parameter values that maximise the likelihood of obtaining the observations; this is typically done using numerical optimisation methods. For a more in-depth tutorial-style text, see [Myung \(2003\)](#).

Suppose that there exists a parameter,  $\theta$ , that maximises the likelihood function,  $L(\theta|\mathbf{x})$ , on the set of possible parameters  $\Theta$ , i.e.

$$L(\hat{\theta}|\mathbf{x}) = \max_{\theta \in \Theta} \{L(\theta|\mathbf{x})\}.$$

Then  $\hat{\theta}$  is called the Maximum Likelihood Estimator (MLE). The MLE (usually)

satisfies the following two properties (Casella and Berger, 2021):

1. *Consistency.* We say that an estimate  $\hat{\theta}$  is consistent if  $\hat{\theta} \rightarrow \theta_0$  in probability as  $n \rightarrow \infty$ , where  $\theta_0$  is the ‘true’ unknown parameter of the distribution of the sample.
2. *Asymptotic Normality.* We say that  $\hat{\theta}$  is asymptotically normal if

$$\sqrt{n} (\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \sigma_{\theta_0}^2),$$

where  $\sigma_{\theta_0}^2$  is called the asymptotic variance of the estimate  $\hat{\theta}$ . Asymptotic normality says that the estimator not only converges to the unknown parameter, but it converges at a rate  $1/\sqrt{n}$ .

Under some regularity conditions and for  $n$  sufficiently large we have the following approximate result:

$$\sqrt{n} (\hat{\theta} - \theta_0) \approx N\left(0, \frac{1}{\mathcal{I}(\theta_0)}\right).$$

where  $\mathcal{I}(\theta)$  is called the Fisher information. The Fisher information matrix is a way of measuring the amount of information that an observable random variable  $X$  carries about an unknown parameter  $\theta$  of a distribution that models  $X$  as is given as

$$\mathcal{I}(\theta) = \mathbb{E}_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta|x) \right)^2 \right],$$

under regularity conditions this is equivalent to

$$\mathcal{I}(\theta) = -\mathbb{E}_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta|x) \right],$$

where  $\ell(\theta|x)$  denotes the log-likelihood ( $\log L(\theta|x)$ ). With independent and identically distributed (iid) data, the Fisher information can be shown to have the form

$$\mathcal{I}_n(\theta) = n\mathcal{I}(\theta).$$

Standard errors (SE) of the estimator,  $\hat{\theta}$ , are just the square roots of the diagonal terms of the variance-covariance matrix. Maximum likelihood estimation is a method from which we can derive an estimator that converges to the true parameter values as the sample size increases. Furthermore, it is relatively simple to imple-

ment and has mathematical properties that make it amenable to theoretical analysis.

We provide an example for completeness. As introduced in Section 1.3.3, we have eight months of PET data where smaller values constitute dangerous situations. We are most interested in the extreme values and therefore utilise EVT. In order to use standard methods from the EVT toolkit directly (designed for analysing ‘large’ extremes), we negate our series of PET values at each location, thus switching the focus from very small values to very large values to identify dangerous situations in our series. We block the PET data by day and model the daily maxima from site 2 – an untreated site. We firstly assume that the data follow the GEV distribution with unknown parameters,  $\mu, \sigma$  and  $\xi$ . We require the likelihood function of the GEV,  $L(\mu, \sigma, \xi|x)$ . Differentiation of Equation (2.3) gives the PDF of the GEV; this can be found to be

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi} + 1} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\},$$

for  $\xi \neq 0$ , and

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \exp \left[ - \exp \left( - \frac{x - \mu}{\sigma} \right) - \left( \frac{x - \mu}{\sigma} \right) \right],$$

when  $\xi \rightarrow 0$ . The likelihood can then be evaluated as the products of the PDFs evaluated at all maxima  $x_i$ ,

$$L(\mu, \sigma, \xi|x) = \prod_{i=1}^m g(x_i; \mu, \sigma, \xi),$$

where  $m$  is the number of block maxima. We then define the likelihood function,

$$L(\mu, \sigma, \xi|x) = \prod_{i=1}^n \left[ \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \right], \quad (2.6)$$

for  $\xi \neq 0$ , and

$$L(\mu, \sigma, \xi|x) = \prod_{i=1}^n \left[ \frac{1}{\sigma} \exp \left\{ - \frac{x_i - \mu}{\sigma} - \exp \left( - \frac{x_i - \mu}{\sigma} \right) \right\} \right], \quad (2.7)$$

for  $\xi \rightarrow 0$ . Therefore, the log-likelihood is,

$$\ell(\mu, \sigma, \xi|x) = -m \log \sigma - \left(\frac{1}{\xi} + 1\right) \log \sum_{i=1}^m \left(1 + \frac{\xi}{\sigma}(x_i - \mu)\right) - \sum_{i=1}^m \left(1 + \frac{\xi}{\sigma}(x_i - \mu)\right)^{-\frac{1}{\xi}}, \quad (2.8)$$

for  $\xi \neq 0$ , and

$$\ell(\mu, \sigma, \xi|x) = -m \log \sigma - \exp \left\{ - \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma}\right) \right\} - \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma}\right), \quad (2.9)$$

for  $\xi \rightarrow 0$ .

To find the MLEs we differentiate the log-likelihood function with respect to each parameter and set the resulting derivatives equal to zero. We then solve for the corresponding parameter to get the MLE. We therefore calculate,

$$\frac{\partial \ell}{\partial \mu} = 0, \quad \frac{\partial \ell}{\partial \sigma} = 0, \quad \frac{\partial \ell}{\partial \xi} = 0,$$

and replace  $\mu, \sigma, \xi$  with  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$  to obtain the MLEs. To verify that the estimates are maximum likelihood estimates, we compute the second derivatives of the log-likelihood function with respect to each parameter. If the second derivatives are negative this denotes the peak in the log-likelihood function about these values and hence the estimates are maximum likelihood estimates. Since the MLE equations for the GEV distribution are not solvable analytically, we use numerical approximation methods in R to maximise the likelihood function and obtain the MLEs. The MLEs are

$$\hat{\mu} = \underset{(0.0979)}{-7.9717}, \quad \hat{\sigma} = \underset{(0.0711)}{1.3541}, \quad \hat{\xi} = \underset{(0.0466)}{-0.0025}, \quad (2.10)$$

with their standard errors (SE) given in parentheses underneath. These standard errors can be used to formulate 95% confidence intervals. For location parameter  $\mu$ , the 95% confidence interval is calculated as,

$$\hat{\mu} \pm 1.96SE_{\mu},$$

where  $SE_{\mu}$  is the standard error for  $\hat{\mu}$ , similarly for the remaining parameters. The value of 1.96 is based on the fact that 95% of the area of a normal distribution is within 1.96 standard deviations of the mean. The MLEs and 95% confidence inter-



	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
MLE	-7.9717	1.3541	-0.0025
95% CI	(-8.1635, -7.7798)	(1.2147, 1.4934)	(-0.0938, 0.0888)

Table 2.1: Maximum likelihood estimates and 95% confidence intervals of the GEV parameters.

vals for the GEV parameters are shown in Table 2.1. Here, a confidence interval represents a range of values that is likely to contain the true value of a parameter, based on the sampling distribution of the estimator. A 95% confidence interval means that if the experiment were repeated many times, the true value of the parameter would be expected to fall within the interval in 95% of the experiments.

## 2.4 Bayesian inference

The majority of the statistical inference in this thesis will be done within a Bayesian framework. Bayesian statistics is an approach to data analysis based on Bayes theorem. A typical statistical analysis might formulate the likelihood function for the assumed statistical model, maximising this with respect to the parameters in that model to obtain their MLEs, see Section 2.3. In a classical sense these are sample-based estimates of fixed but unknown quantities. In the Bayesian paradigm, the likelihood is merely an ingredient in the inferential process; via Bayes Theorem, it is combined with the density of the *prior distribution* to provide a *posterior distribution* for the parameters of interest – in effect an update in our beliefs about the parameters after having observed some data, relative to our beliefs before observing these data. Crucially, the interpretation of the model parameters is different here: rather than being fixed (but unknown) constants, the parameters are now regarded as random variables. This means that in the Bayesian setting, confidence (or credible) intervals (for example) have a much more natural interpretation, with there being a probability of 0.95 that the parameter falls within the bounds of the 95% Bayesian credible interval. A Bayesian framework provides many advantages; it allows for incorporating uncertainty in models and parameters as well as allowing expert knowledge to be used to formulate the prior distributions.

Suppose we have data  $\mathbf{x} = (x_1, \dots, x_n)^T$  which we model using PDF  $f(\mathbf{x}|\theta)$ ,

which depends on a single parameter  $\theta$ . The main result in Bayesian statistics is Bayes theorem which states that the posterior distribution of  $\theta$ , denoted  $\pi(\theta|\mathbf{x})$ , can be obtained by multiplying the prior distribution of  $\theta$ ,  $\pi(\theta)$ , with the likelihood obtained by observing data  $\mathbf{x}$ ,  $L(\theta|\mathbf{x})$  (also denoted  $f(\mathbf{x}|\theta)$ ), and dividing by a normalising value constant with respect to  $\theta$ ,  $f(\mathbf{x})$ . We can write this mathematically as,

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})\pi(\theta)}{f(\mathbf{x})}.$$

where

$$f(\mathbf{x}) = \begin{cases} \int_{\Theta} \pi(\theta)L(\theta|\mathbf{x}) d\theta & \text{if } \theta \text{ is continuous,} \\ \sum_{\Theta} \pi(\theta)L(\theta|\mathbf{x}) & \text{if } \theta \text{ is discrete.} \end{cases}$$

As  $f(\mathbf{x})$  is not a function of  $\theta$ , the normalising constant is often omitted and it is sufficient to say that the posterior distribution is proportional to the likelihood multiplied by the prior distribution,

$$\pi(\theta|\mathbf{x}) \propto L(\theta|\mathbf{x})\pi(\theta). \quad (2.11)$$

### 2.4.1 Prior distributions

In Bayesian statistics, a prior distribution is a probability distribution that represents our beliefs or knowledge about the values of a parameter before observing any data. The prior is specified before observing the data and is updated to a posterior distribution using Bayes theorem. The prior can take different forms depending on the information available and the assumptions made about the parameter. For example, a uniform prior assumes that all values of the parameter are equally likely, while a normal prior assumes that the parameter follows a normal distribution with a specific mean and variance. We have substantial prior information for  $\theta$  when the prior distribution dominates the posterior, that is  $\pi(\theta|\mathbf{x}) \sim \pi(\theta)$ . If there is very little or no prior information about  $\theta$ , we choose a prior distribution which is not concentrated about any particular value, that is, one with a large variance. The majority of the information about  $\theta$  will be passed through to the posterior distribution via the data and hence,  $\pi(\theta|\mathbf{x}) \sim L(\theta|\mathbf{x})$ . When we assume no prior knowledge or prior ignorance we tend to use a Jeffreys prior.

The Jeffreys prior, named after Sir Harold Jeffreys, is a non-informative prior

distribution for a parameter space (Jeffreys, 1946). It is commonly used to depict prior ignorance. Its density function is proportional to the square root of the determinant of the Fisher information matrix. Jeffreys recommended that we represent prior ignorance by the prior distribution,

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)}.$$

The Jeffreys prior favours values of  $\theta$  for which  $\mathcal{I}(\theta)$  is large. Hence it satisfies the local uniformity property: it does not change much in the region over which the likelihood is significant and it is invariant with respect to one-to-one transformations. If  $\theta$  is a vector then Jeffreys prior is

$$\pi(\theta) \propto \sqrt{\det \mathcal{I}(\theta)},$$

which is still invariant under reparametrisation.

We provide an example of a Jeffreys prior for a random sample from an Exponential distribution, that is  $X_i|\theta \sim \text{Exp}(\theta)$  for  $i = 1, \dots, n$  (independent). We calculate the likelihood function

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^n \theta \exp\{-\theta x_i\}, \\ &= \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\}, \\ &= \theta^n \exp\{-n\bar{x}\theta\}, \end{aligned}$$

as  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . Therefore

$$\begin{aligned} \ell(\theta|\mathbf{x}) &= n \log \theta - n\bar{x}\theta, \\ \Rightarrow \frac{\partial}{\partial \theta} \ell(\theta|\mathbf{x}) &= \frac{n}{\theta} - n\bar{x}, \\ \Rightarrow \frac{\partial^2}{\partial \theta^2} \ell(\theta|\mathbf{x}) &= -\frac{n}{\theta^2}, \\ \Rightarrow \mathcal{I}(\theta) &= \mathbb{E}_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \ell(\theta|\mathbf{x}) \right] = \frac{n}{\theta^2}. \end{aligned}$$

The Jeffreys prior for this model is

$$\begin{aligned}\pi(\theta) &\propto \sqrt{\mathcal{I}(\theta)} \\ &\propto \frac{\sqrt{n}}{\theta}, \\ &\propto \frac{1}{\theta}, \quad \theta > 0.\end{aligned}$$

### 2.4.2 Markov Chain Monte Carlo

Monte Carlo methods are a broad class of computational techniques that rely on random sampling to estimate numerical solutions to problems that may be difficult or impossible to solve analytically. These methods generate random samples from the data's (assumed) underlying distribution and use these samples to estimate the desired quantity. The accuracy of the estimate improves as the number of samples increases. For example, if we consider a bivariate model,  $f_{\Theta}(\theta_1, \theta_2)$  where the marginals  $f_{\theta_1}(\theta_1)$  and  $f_{\theta_2}(\theta_2)$  are hard to compute, we can simulate realisations of both  $\theta_1$  and  $\theta_2$  and look at the two histograms for information about the marginals. We can also use these realisations to gain sample mean and variances. Monte Carlo can then be used as a way of estimating difficult integrals.

Suppose we wish to evaluate an integral of the form

$$\int_{\Theta} h(\theta) d\theta,$$

for which there is no closed analytical solution. If the integrand has the form

$$h(\theta) = \tilde{h}(\theta)f(\theta),$$

for some PDF  $f(\cdot)$ , then the integral has the form,

$$\int_{\Theta} h(\theta) d\theta = \int_{\Theta} \tilde{h}(\theta)f(\theta) d\theta = \mathbb{E}[\tilde{h}(\theta)],$$

where  $\theta$  is a random variable with PDF  $f$  with support  $\Theta$ . If we are able to simulate

iid realisations of  $\theta$ , such as  $\theta^{(1)}, \dots, \theta^{(N)}$ , we obtain an unbiased estimator:

$$\int_{\Theta} h(\theta) d\theta = \mathbb{E}[\tilde{h}(\theta)] \approx \frac{1}{N} \sum_{i=1}^N \tilde{h}(\theta^{(i)}) = \hat{I}.$$

This method for approximating integrals is called Monte Carlo integration. Provided  $\sigma_h^2 = \text{Var}[h(\theta)] < \infty$  exists,

$$\begin{aligned} \text{Var}[\hat{I}] &= \frac{1}{N^2} \text{Var} \left[ \sum_{i=1}^N \tilde{h}(\theta^{(i)}) \right], \\ &= \frac{\sigma_h^2}{N}. \end{aligned}$$

Hence, the size of error in  $\hat{I}$  is proportional to the standard deviation of the estimator.

When performing a Bayesian analysis, we are most often interested in finding the posterior distribution. However, finding an analytic form for the posterior is not always possible. Fortunately, techniques to sample from the posterior numerically have been developed. The most common approach is to use a Markov Chain Monte Carlo (MCMC) scheme, of which there are several kinds. If there is conjugacy within the parameters, where the prior and data distributions are of the same distribution family, we are able to calculate full conditional distributions (FCDs) for the parameters and draw samples directly. Conjugate and semi-conjugate models can be sampled using a Gibbs sampler (Casella and George, 1992). Through Bayes theorem we can calculate the FCDs for  $\theta$  up to proportionality and draw posterior elements successively. The general algorithm for a Gibbs sampler for  $N$  iterations is given in Algorithm 1.

The samples returned form a Markov chain whose stationary distribution equals the posterior distribution,  $\pi(\theta|\mathbf{x})$ . The chain is initialised arbitrarily, yet it is commonplace to initialise at the prior means. As the initialisation is somewhat guesswork, we wouldn't necessarily assume these values are from the posterior, hence it is common practice to discard the samples before the chain has reached convergence as 'burn-in'. Furthermore, how the chain is initialised shouldn't matter as the chain should converge at the same stationary distribution. Checking for convergence can be done numerically or by simply plotting the trace plots for each element in the

**Algorithm 1** Gibbs sampler

- 
- 1: Initialise the chain at its initial value  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$ . Set  $i = 1$
  - 2: Draw a sample for each element of  $\theta$  from its FCD,

$$\begin{aligned}\theta_1^{(i)} &\sim \pi\left(\theta_1^{(i)} \mid \theta_2^{(i-1)}, \dots, \theta_p^{(i-1)}, \mathbf{x}\right) \\ \theta_2^{(i)} &\sim \pi\left(\theta_2^{(i)} \mid \theta_1^{(i)}, \dots, \theta_p^{(i-1)}, \mathbf{x}\right) \\ &\vdots \\ \theta_p^{(i)} &\sim \pi\left(\theta_p^{(i)} \mid \theta_1^{(i)}, \dots, \theta_{p-1}^{(i)}, \mathbf{x}\right)\end{aligned}$$

- 3: If  $i = N$  stop, else set  $i = i + 1$  and go to step 2.
- 

parameter vector  $\theta$  and eyeballing that the chains converge at the same distribution. We provide an example of a Gibbs sampler for a conjugate Poisson-gamma model with unknown rate. Our conjugate model structure is,

$$\begin{aligned}X_i \mid \lambda &\sim Po(\lambda), \quad i = 1, \dots, n, \\ \lambda &\sim Ga(g_0, h_0).\end{aligned}\tag{2.12}$$

From this we can derive the FCD for  $\lambda$ ,

$$\lambda \mid \mathbf{x} \sim Ga\left(\sum_{i=1}^n x_i + g_0, n + h_0\right).$$

We demonstrate this using example data,

$$X_i \mid \lambda \sim Po(\lambda = 5), \quad i = 1, \dots, n,$$

for  $n = 200$  data points. We specify vague prior information via the hyperparameters,  $g_0 = 0.1$  and  $h_0 = 0.1$ , to give large variance. To demonstrate convergence to the posterior, we initialise at  $\lambda^{(0)} = 1, 5, 10$ . Output is shown in Figure 2.1. After the few initial iterations, all chains converge to the same distribution.

If there is no conjugacy and the posterior distribution is intractable, a Metropolis-Hastings (MH) algorithm can be used (Hastings, 1970). Here, instead of being able to draw directly from FCDs, we propose posterior values of each element of  $\theta$  and accept or reject depending on a probability given by the observed data likelihood

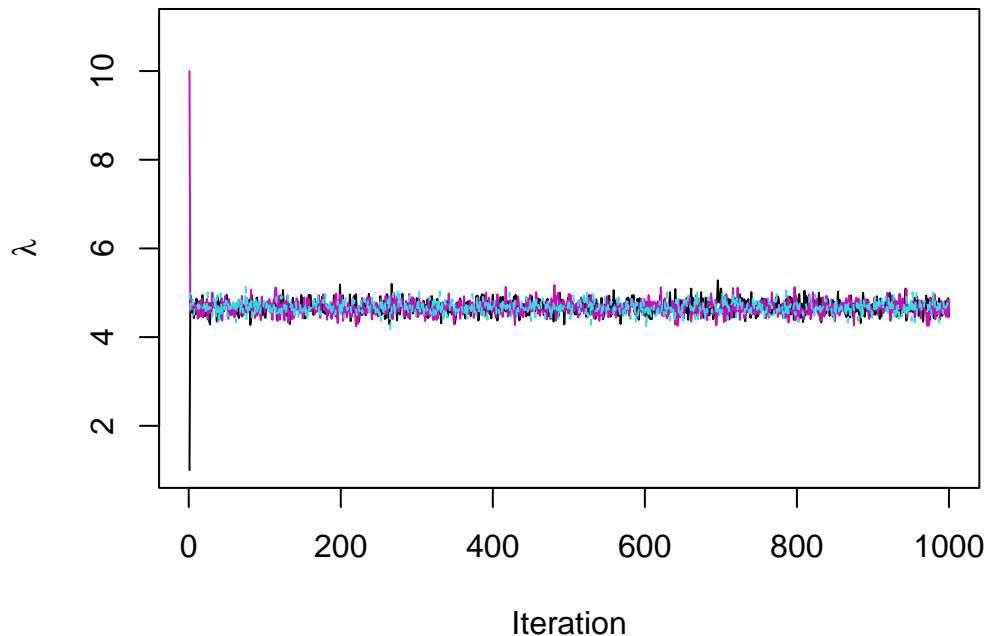


Figure 2.1: Trace plots for the posterior samples for  $\lambda$  from a Gibbs sampler for a Poisson-gamma model using initialisation points  $\lambda^{(0)} = 1, 5, 10$  as the black, blue and pink lines respectively.

and prior distributions. The MH algorithm is outlined in Algorithm 2. There are numerous options for the proposal distribution  $q(\cdot|\cdot)$  which can be used to draw our proposal value  $\theta^*$ , the most common is a normal distribution, centred on the current value of  $\theta$  with some tuning variance  $\Sigma$ ,

$$\theta^* \sim N(\theta, \Sigma).$$

An advantage of using a normal random walk is that the normal distribution is symmetric, therefore in the acceptance probability  $q(\theta^{(i-1)}|\theta^*) = q(\theta^*|\theta^{(i-1)})$  and so cancel out in the formula for  $\alpha$ . The parameter  $\Sigma$  controls the size of jumps from the current value to proposed value in the chain. If  $\Sigma$  is too large then proposed values will be far from current, accepted values and will most likely result in the majority of proposals being rejected and hence we will achieve a low acceptance rate (the number of accepted proposed values divided by the number of iterations). On

**Algorithm 2** Metropolis-Hastings

- 1: Initialise the chain at its initial value  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$ . Set  $i = 1$
- 2: Sample a proposal value  $\theta^*$  from the proposal distribution of  $\theta$ ,

$$\theta^* \sim q(\cdot|\cdot)$$

- 3: Set  $\theta^{(i)} = \theta^*$  with probability  $\alpha$ , where

$$\alpha = \min \left\{ 1, \frac{\pi(\theta^*)L(\theta^*|\mathbf{x})q(\theta^*|\theta^{(i-1)})}{\pi(\theta)L(\theta|\mathbf{x})q(\theta^{(i-1)}|\theta^*)} \right\}$$

else set  $\theta^{(i)} = \theta^{(i-1)}$

- 4: If  $i = N$  stop, else set  $i = i + 1$  and go to step 2.

the other hand, if  $\Sigma$  is too small, the majority of the proposals will be accepted and though this seems like an advantage, it will result in very slow convergence to the stationary distribution (if the initial value is far from the posterior) meaning a lot of iterations will be removed as burn-in. There will also be a high autocorrelation between successive values of the chain indicating the posterior distribution is not being explored effectively. Selecting a good value of  $\Sigma$  is referred to as tuning the chain and is important. An optimal MH scheme should aim to accept around 23.4% of the proposal values, in high dimensional settings (Roberts and Rosenthal, 2001). A formula provided to assist in achieving this acceptance rate is given using the variance of the parameters chains from a short pilot run, say 1k iterations then,  $q(\theta^*|\theta) = N(\theta^*; \theta, \Sigma)$  where the innovation matrix  $\Sigma = \gamma \widehat{Var}(\theta|x)$ , with  $\widehat{Var}(\theta|x)$  obtained from a pilot run and  $\gamma = \frac{2.38^2}{p}$  where  $p$  is the number of parameters.

For many problems of interest, FCDs may be available for sampling from for a subset of components of  $\theta$ . In this case we use a componentwise transitions algorithm as outlined in Algorithm 3. This is in fact the original form of the Metropolis algorithm. Proving that  $\pi(\theta|\mathbf{x})$  is the stationary distribution of a Markov chain defined in this way can be achieved by induction (details omitted). Note that the Metropolis-Hastings algorithm as presented in Algorithm 2 can be seen as a special case of Algorithm 3. If the FCD is available for sampling from directly, for a particular component  $\theta_i$ , it is easy to show that the resulting acceptance probability is 1. When all FCDs are available for sampling from, we obtain Algorithm 1.



---

**Algorithm 3** Componentwise transitions

---

- 1: Initialise the chain at its initial value  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$ . Set  $i = 1$
- 2: Obtain a new value  $\theta^{(i)}$  from  $\theta^{(i-1)}$  by successive generation of values,

$$\begin{aligned}\theta_1^{(i)} &\sim \pi\left(\theta_1^{(i)} \mid \theta_2^{(i-1)}, \dots, \theta_p^{(i-1)}, \mathbf{x}\right) \\ \theta_2^{(i)} &\sim \pi\left(\theta_2^{(i)} \mid \theta_1^{(i)}, \dots, \theta_p^{(i-1)}, \mathbf{x}\right) \\ &\vdots \\ \theta_p^{(i)} &\sim \pi\left(\theta_p^{(i)} \mid \theta_1^{(i)}, \dots, \theta_{p-1}^{(i)}, \mathbf{x}\right)\end{aligned}$$

using individual Metropolis-Hastings steps with proposal distributions  $q(\theta_1 \mid \theta_1^{(i-1)})$ ,  $q(\theta_2 \mid \theta_2^{(i-1)})$ ,  $\dots$ ,  $q(\theta_p \mid \theta_p^{(i-1)})$  respectively.

- 3: If  $i = N$  stop, else set  $i = i + 1$  and go to step 2.
- 

We provide an example of an MH scheme using the Poisson-gamma model from Equation (2.12) using a normal random walk MH algorithm with varying choices of innovation parameter  $\Sigma$ . Trace plots for the parameter  $\lambda$  are given in Figure 2.2. From Figure 2.2 we can deduce that when  $\Sigma = 0.005$ , the chain is too ‘hot’, it updates frequently with an acceptance rate of 92%, but the updates are very small meaning the chain takes a long time to reach its stationary distribution. Furthermore, the successive samples are highly correlated meaning the chain explores the posterior distribution inefficiently. For  $\Sigma = 0.5$  the chain is too ‘cold’, the difference between candidate values and current values of the chain are relatively large and hence results in a high rate of rejection (only 6% of proposals were accepted), vastly reducing the efficiency of the scheme. Finally, when  $\Sigma = 0.12$  we have good and quick convergence and the chain explores the posterior distribution efficiently. This results in an acceptance probability of 24%, where the updates are large enough to reduce the autocorrelation of the chain but we are not wasteful of data. This demonstrates the ‘Goldilocks Principle’ of parameter tuning, where we want an innovation parameter which is not too large or too small.

In MCMC algorithms, the log-likelihood is often used in the acceptance probability instead of the likelihood itself. This is because the log-likelihood is more numerically stable than the likelihood, especially for small probabilities. When calculating the acceptance probability, the likelihood of the proposed sample is compared to

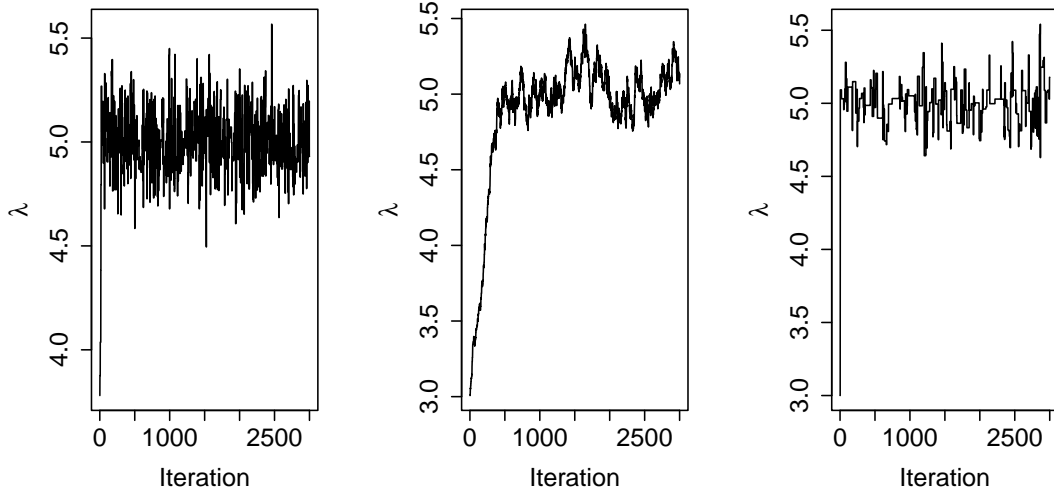


Figure 2.2: Traceplots for the Poisson rate parameter  $\lambda$  from a Metropolis-Hastings algorithm with innovation parameter  $\epsilon = 0.12, 0.005, 0.5$  respectively.

the likelihood of the current sample. If the proposed sample has a higher likelihood, it is more likely to be accepted. However, if the likelihood is very small, it can become difficult to compute accurately, leading to numerical instability. Using the log-likelihood in the acceptance probability can help to avoid this and improve the accuracy of the MCMC algorithm.

### 2.4.3 Bayesian inference example

In Section 2.3 we introduced maximum likelihood estimation for the GEV parameters. We will now perform the same inference in a Bayesian setting. Firstly, we assign independent, non-informative prior distributions to the model parameters. The joint prior distribution can be written as,

$$\pi(\mu, \sigma, \xi) = \pi(\mu)\pi(\sigma)\pi(\xi),$$

where

$$\mu \sim N(0, 100), \quad \eta \sim N(0, 100), \quad \xi \sim N(0, 100),$$

where we reparameterise  $\eta = \log \sigma$  to respect the positivity of the scale parameter. As the posterior distribution is intractable, we use MCMC with normal random walk proposals to draw realisations from the posteriors. The MCMC scheme is implemented as in Algorithm 2 using the likelihood function defined in Equations (2.6)

	$\mu$	$\sigma$	$\xi$
Mean	-7.9674	1.3683	0.0012
95% CI	(-8.1522, -7.7580)	(1.2376, 1.5207)	(-0.0837, 0.1012)

Table 2.2: Posterior means and 95% credible intervals of the GEV parameters.

and (2.7) in the acceptance probabilities. We draw realisations of  $\mu, \eta$  and  $\xi$  recursively over each iteration. The MCMC scheme was run for 10k iterations with a burn-in of 1k to ensure convergence. The posterior means and 95% credible intervals for each parameter are given in Table 2.2. The posterior means and MLEs, shown in Equation (2.10), agree to within one decimal place. However, MLEs provide a point estimate for each parameter, while Bayesian analysis generates a posterior distribution for each parameter that reflects both prior knowledge and the likelihood of the data. Bayesian analysis also produces posterior distributions that can be used to quantify uncertainty about the estimated parameters, while MLEs provide standard errors. The confidence intervals in Table 2.1 are similar to the Bayesian credible intervals in Table 2.2, but the interpretation is different. The credible intervals represent a range of values for the parameters that have a specified probability of containing the true values, given the observed data and prior knowledge, thus the 95% credible interval means that there is a 95% probability that the true value of the parameter is within the interval.

#### 2.4.4 Deviance Information Criteria

The deviance information criterion (DIC) was introduced by Spiegelhalter et al. (2002) as a metric for Bayesian hierarchical model selection. In application, DIC is analogous to the Akaike information criterion (AIC) (Akaike, 1974), it is a function of model deviance and the effective number of parameters. While AIC uses the maximum likelihood estimate, DIC's estimate is based on the posterior mean; hence, the DIC is an asymptotic approximation as the sample size becomes large, it's also only valid when the posterior distribution is approximately normal.

The DIC gives a numerical value to summarise how well the model fits the data, however as in the AIC, it penalises the model for how many parameters are used. Clearly using more parameters will give a better fit, but we want model adequacy

with the least model complexity. The statistical deviance is defined as

$$D(x, \theta) = -2 \log(L(\theta|x)).$$

where  $L(\theta|x)$  is the likelihood function, with data  $x$  and unknown parameters  $\theta$ . The deviance at the posterior mean of  $\theta$ , denoted  $\bar{\theta}$ , is

$$D_{\bar{\theta}}(x) = D(x, \bar{\theta}),$$

and the expectation of the deviance under the posterior of the un-standardised model

$$\bar{D}(x, \theta) = \mathbb{E}(D(x, \theta)|x),$$

where larger values correspond to a worse fit. The effective number of parameters,  $p_D$ , is denoted as

$$p_D = \bar{D}(x, \theta) - D(x, \bar{\theta}).$$

Furthermore, the deviance information criterion is derived as

$$DIC = \bar{D}(x, \theta) + p_D. \tag{2.13}$$

Smaller values of DIC are preferred.

## 2.5 Dynamic linear models

If we consider a time-series  $\{X_{t_i}, i > 1\}$ , then we say this is a first order Markov chain if, for  $i > 1$

$$\pi(x_{t_i}|x_{t_{1:i-1}}) = \pi(x_{t_i}|x_{t_{i-1}}),$$

that is, information about  $x_{t_i}$  carried by all observations up to time  $t_{i-1}$  is the same as that carried by  $x_{t_{i-1}}$  alone. The joint distribution of the Markov chain can be factorised as

$$\pi(x_{t_{1:n}}) = \pi(x_{t_1}) \prod_{i=2}^n \pi(x_{t_i}|x_{t_{i-1}}).$$

State-space models assume there is a Markov chain in some underlying state process ( $\theta_{t_i}$ ) and then the observations are noisy measurements of  $\theta_{t_i}$ . The observations  $X_{t_i}$  are then independent conditional on  $\theta_{t_i}$  (Petris et al., 2009).

State-space models build on the relatively simple dependence structure of a (first order) Markov chain. They are made of two main components, observed data  $(x_{t_1}, \dots, x_{t_n})$  and unobserved/latent states  $(\theta_{t_0}, \dots, \theta_{t_n})$ . Figure 2.3 shows the evolution of a simple univariate state space in which the continuous valued latent state process  $\{\theta_{t_0}, \theta_{t_1}, \dots, \theta_{t_{i-1}}, \theta_{t_i}, \dots\}$  evolves according to a first order Markov chain with transition density  $\pi(\theta_{t_i}|\theta_{t_{i-1}})$ . The continuous-valued observation process  $\{x_{t_1}, x_{t_2}, \dots, x_{t_{i-1}}, x_{t_i}, \dots\}$  is linked to the latent state process at an arbitrary time  $t_i$  via the density  $\pi(x_{t_i}|\theta_{t_i})$ ; here it is assumed that the observed data are conditionally independent given the latent states. The observable process  $(X_{t_i})$  depends on the underlying, unobservable latent state process  $(\theta_{t_i})$  and we can reasonably assume that the observation  $X_{t_i}$  only depends on the state of the system at the time the measurement is taken,  $\theta_{t_i}$ . State-space models can be used for modelling univariate or multivariate time-series in the presence of non-stationarity, structural changes and irregular patterns (see e.g. [Harvey \(1990\)](#), [West and Harrison \(2006\)](#)).

Time-series analysis typically begins with the formulation of a model that accounts for temporal dependence, for example through auto-correlation, trend or seasonality. The use of state-space models within a time-series setting allows for uncertainty quantification in both the observation process and any dynamic variables that are not observed directly. Forecasting therefore accounts for these different sources of uncertainty and, when inferences are made within the Bayesian paradigm, additional parameter uncertainty. Throughout, we focus on a particular class of state-space model within which the observation and system equations involve linear functions of the latent process. Such models are known as dynamic linear models (DLMs, see e.g. [Petris et al. \(2009\)](#), [West and Harrison \(2006\)](#)) and offer several practical benefits over their nonlinear counterparts. Notably, they admit a tractable observed data likelihood function, allowing a computationally efficient approach to inference and forecasting.

It remains that we specify the relationship between  $X_{t_i}$  and  $\theta_{t_i}$ , and between  $\theta_{t_i}$  and  $\theta_{t_{i-1}}$ . In each case, we adopt linear relationships, and further assume that the errors in the state and observed components are independent and normally

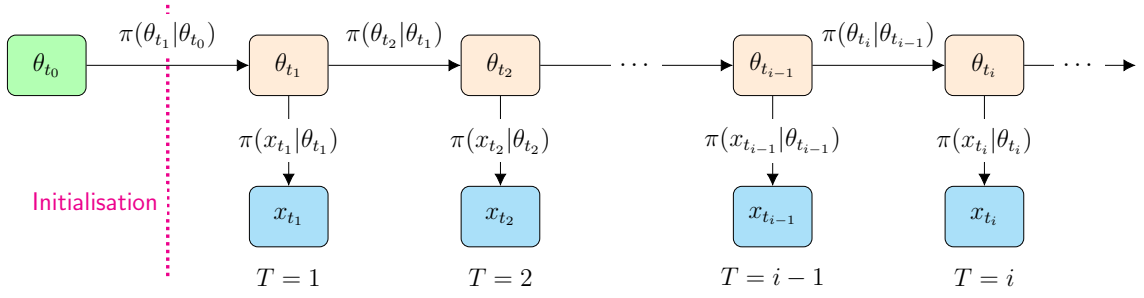


Figure 2.3: Directed acyclic graph showing the dependence structure of the state-space model.

distributed. This structure leads to a DLM, given by the following equations:

$$\text{Observation Equation : } X_{t_i} = F_{t_i} \theta_{t_i} + \nu_{t_i},$$

$$\text{System Equation : } \theta_{t_i} = G_{t_i} \theta_{t_{i-1}} + \omega_{t_i}.$$

Here,  $X_{t_i}$  is a scalar,  $\theta_{t_i}$  is a  $p \times 1$  vector,  $F_{t_i}$  is a  $1 \times p$  vector,  $G_{t_i}$  is a  $p \times p$  matrix and  $\nu_{t_i} \sim N(0, V_{t_i})$  and  $\omega_{t_i} \sim N(0, W_{t_i})$  are independent white noise processes with known variance matrices  $V_{t_i}$  and  $W_{t_i}$ , typically assumed to be constant. Assuming that the initial latent state follows a Gaussian distribution gives

$$\begin{aligned} \theta_0 &\sim N(m_0, C_0), \\ \theta_{t_i} | \theta_{t_{i-1}} &\sim N(G_{t_i} \theta_{t_{i-1}}, W_{t_i}), \\ X_{t_i} | \theta_{t_i} &\sim N(F_{t_i} \theta_{t_i}, V_{t_i}), \end{aligned}$$

for suitably chosen hyperparameters  $m_0$  and  $C_0$ .

The simplest univariate DLM is the locally constant DLM, also known as the local level model. This model has constant functions  $F_{t_i} = 1, G_{t_i} = 1$  and variance components  $V_{t_i} = V, W_{t_i} = W$ . The observation and system equations are

$$\begin{aligned} X_{t_i} &= \theta_{t_i} + \nu_{t_i}, & \nu_{t_i} &\overset{\text{indep}}{\sim} N(0, V), \\ \theta_{t_i} &= \theta_{t_{i-1}} + \omega_{t_i}, & \omega_{t_i} &\overset{\text{indep}}{\sim} N(0, W). \end{aligned} \tag{2.14}$$

### 2.5.1 Bayesian inference for DLMs

Let  $\psi = (\psi_1, \dots, \psi_{n_p})^T$  denote all fixed model parameters, let  $\theta_{t_{0:n}} = (\theta_{t_0}, \dots, \theta_{t_n})^T$  denote the vector of latent states up to time  $n$  and let  $x = x_{t_{1:n}} = (x_{t_1}, \dots, x_{t_n})$

denote the equally spaced observed data. Upon ascribing a prior density  $\pi(\cdot)$  to  $\psi$ , Bayesian inference may proceed via the joint distribution:

$$\begin{aligned} \pi(\psi, \theta_{t_{0:n}}|x) &\propto \pi(\psi)\pi(\theta_{t_0})\pi(\theta_{t_{1:n}}|\theta_{t_0}, \psi)\pi(x|\theta_{t_{1:n}}, \psi), \\ \text{where} \quad \pi(\theta_{t_{1:n}}|\theta_{t_0}, \psi) &= \prod_{i=1}^n \pi(\theta_{t_i}; G_{t_i}\theta_{t_{i-1}}, W_{t_i}), \\ \text{and} \quad \pi(x|\theta_{t_{1:n}}, \psi) &= \prod_{i=1}^n \pi(x_{t_i}; F_{t_i}\theta_{t_i}, V_{t_i}). \end{aligned}$$

If interest lies in the marginal parameter posterior then we may compute

$$\pi(\psi|x) \propto \pi(\psi)\pi(x|\psi),$$

where the marginal likelihood  $\pi(x|\psi)$  is given by

$$\pi(x|\psi) = \pi(x_{t_1}|\psi) \prod_{i=2}^n \pi(x_{t_i}|x_{t_{1:i-1}}, \psi), \quad (2.15)$$

and whose constituent terms are analytically tractable. Simulating from  $\theta_{t_{0:n}}|x, \psi$  can be achieved using a forward filtering, backward sampling algorithm (FFBS) (Carter and Kohn, 1994, Frühwirth-Schnatter, 1994). The idea is to first calculate the filtering distributions  $\pi(\theta_{t_i}|x_{t_{1:i}})$  recursively, for  $i = 1, \dots, n$ . This is achieved using a forward filter, also known as a Kalman filter (Kalman, 1960). The key steps of the forward filter for time  $t_i$ , where we have the posterior for  $\theta_{t_{i-1}}$  as  $\theta_{t_{i-1}} \sim N(m_{t_{i-1}}, C_{t_{i-1}})$ , are as follows. Note that we suppress dependence on  $\psi$  where possible.

- Using the system equation at time  $t_i$  the prior distribution is

$$\theta_{t_i}|x_{t_{i-1}} \sim N(a_{t_i}, R_{t_i}),$$

where

$$\begin{aligned} a_{t_i} &= G_{t_i}m_{t_{i-1}}, \\ R_{t_i} &= G_{t_i}C_{t_{i-1}}G_{t_i}^T + W_{t_i}. \end{aligned}$$

- From the observation equation, the one step ahead predictive distribution is

$$X_{t_i}|x_{t_{i-1}} \sim N(f_{t_i}, Q_{t_i}),$$

where

$$\begin{aligned} f_{t_i} &= E(x_{t_i}|x_{t_{i-1}}) = F_{t_i}a_{t_i}, \\ Q_{t_i} &= Var(x_{t_i}|x_{t_{i-1}}) = F_{t_i}R_{t_i}F_{t_i}^T + V_{t_i}. \end{aligned}$$

- To update the posterior density, we construct the joint density of  $(\theta_{t_i}, X_{t_i})^T$  and note that  $Cov(\theta_{t_i}, X_{t_i}) = Cov(\theta_{t_i}, F_{t_i}\theta_{t_i} + V_{t_i}) = R_{t_i}F_{t_i}^T$ , hence,

$$\begin{pmatrix} \theta_{t_i} \\ X_{t_i} \end{pmatrix} \Big| x_{t_{1:i-1}} \sim N \left\{ \begin{pmatrix} a_{t_i} \\ f_{t_i} \end{pmatrix}, \begin{pmatrix} R_{t_i} & R_{t_i}F_{t_i}^T \\ F_{t_i}R_{t_i} & Q_{t_i} \end{pmatrix} \right\}.$$

- Conditioning on  $X_{t_i} = x_{t_i}$  yields the required filtering distribution as

$$\theta_{t_i}|x_{t_i} \sim N(m_{t_i}, C_{t_i}),$$

with

$$\begin{aligned} m_{t_i} &= E(\theta_{t_i}|x_{t_i}) = a_{t_i} + R_{t_i}F_{t_i}^TQ_{t_i}^{-1}e_{t_i}, \\ C_{t_i} &= Var(\theta_{t_i}|x_{t_i}) = R_{t_i} - R_{t_i}F_{t_i}^TQ_{t_i}^{-1}F_{t_i}R_{t_i}. \end{aligned}$$

where  $e_{t_i} = X_{t_i} - f_{t_i}$  is the forecast error.

These steps can be executed for  $i = 1, \dots, n$  to give the forward filter. To obtain the terms necessary for the backward sweep, we have the joint density of  $\theta_{t_i}$  and  $\theta_{t_{i+1}}$  (given  $x_{t_{1:i}}$ ) is

$$\begin{pmatrix} \theta_{t_i} \\ \theta_{t_{i+1}} \end{pmatrix} \Big| x_{t_{1:i-1}} \sim N \left\{ \begin{pmatrix} m_{t_i} \\ G_{t_{i+1}}m_{t_i} \end{pmatrix}, \begin{pmatrix} C_{t_i} & C_{t_i}G_{t_{i+1}}^T \\ G_{t_{i+1}}C_{t_i} & G_{t_{i+1}}C_{t_i}G_{t_{i+1}}^T + W_{t_i} \end{pmatrix} \right\}.$$

Conditioning on  $\theta_{t_{i+1}}$  gives

$$\pi(\theta_{t_i}|\theta_{t_{i+1}}, x_{t_{1:i}}) = N(\theta_{t_i}; h_{t_i}, H_{t_i}),$$

where

$$\begin{aligned} h_{t_i} &= m_{t_i} + C_{t_i}G_{t_{i+1}}^T (G_{t_{i+1}}C_{t_i}G_{t_{i+1}}^T + 1)^{-1}(\theta_{t_{i+1}} - G_{t_{i+1}}m_{t_i}), \\ H_{t_i} &= C_{t_i} - C_{t_i}G_{t_{i+1}}^T (G_{t_{i+1}}C_{t_i}G_{t_{i+1}}^T + W_{t_i})^{-1}G_{t_{i+1}}C_{t_i}. \end{aligned}$$

Note that the constituent terms of the marginal likelihood in Equation (2.15) are obtained from the forward pass as

$$\pi(x_{t_i}|x_{t_{1:i-1}}, \psi) = N(x_{t_i}; f_{t_i}, Q_{t_i}).$$



The full FFBS scheme is outlined in Algorithm 4.

---

**Algorithm 4** FFBS scheme
 

---

**Forward Filtering:**

- 1: Initial distribution:  $\theta_{t_0} \sim N(m_{t_0}, C_{t_0})$ . Store the values of  $m_{t_0}$  and  $C_{t_0}$ .
- 2: For  $i = 1, \dots, n$ ,

- (a) Prior at  $t_i$ . Using the system equation, we have that

$$\theta_{t_i} | x_{t_{1:i-1}} \sim N(G_{t_i} m_{t_{i-1}}, G_{t_i} C_{t_{i-1}} G_{t_i}^T + W_{t_i}).$$

Store  $R_{t_i} = G_{t_i} C_{t_{i-1}} G_{t_i}^T + W_{t_i}$ .

- (b) One step forecast. Using the observation equation, we have that

$$X_{t_i} | x_{t_{1:i-1}} \sim N(F_{t_i} G_{t_i} m_{t_{i-1}}, F_{t_i} R_{t_i} F_{t_i}^T + V_{t_i}).$$

Store the marginal likelihood contribution

$$\pi(x_{t_i} | x_{t_{1:i-1}}) = N(x_{t_i}; F_{t_i} G_{t_i} m_{t_{i-1}}, F_{t_i} R_{t_i} F_{t_i}^T + V_{t_i}).$$

- (c) Posterior at  $t_i$ :  $\theta_{t_i} | x_{t_{1:i}} \sim N(m_{t_i}, C_{t_i})$ , where

$$\begin{aligned} m_{t_i} &= G_{t_i} m_{t_{i-1}} + R_{t_i} F_{t_i}^T (F_{t_i} R_{t_i} F_{t_i}^T + V_{t_i})^{-1} (x_{t_i} - F_{t_i} G_{t_i} m_{t_{i-1}}) \\ C_{t_i} &= R_{t_i} - R_{t_i} F_{t_i}^T (F_{t_i} R_{t_i} F_{t_i}^T + V_{t_i})^{-1} F_{t_i} R_{t_i} \end{aligned}$$

Store the values of  $m_{t_i}$  and  $C_{t_i}$ .

**Backward Sampling:**

- 3: Sample  $\theta_{t_n} | \mathbf{x}_{t_{1:n}} \sim N(m_{t_n}, C_{t_n})$ .
- 4: For  $i = n - 1, \dots, 0$ ,

- (a) Backward distribution:  $\theta_{t_i} | \theta_{t_{i+1}}, x_{t_{1:i}} \sim N(h_{t_i}, H_{t_i})$ , where

$$\begin{aligned} h_{t_i} &= m_{t_i} + C_{t_i} G_{t_{i+1}}^T \left( G_{t_{i+1}} C_{t_i} G_{t_{i+1}}^T + W_{t_i} \right)^{-1} (\theta_{t_{i+1}} - G_{t_{i+1}} m_{t_i}) \\ H_{t_i} &= C_{t_i} - C_{t_i} G_{t_{i+1}}^T \left( G_{t_{i+1}} C_{t_i} G_{t_{i+1}}^T + W_{t_i} \right)^{-1} G_{t_{i+1}} C_{t_i} \end{aligned}$$

- (b) Sample  $\theta_{t_i} | \theta_{t_{i+1}}, x_{t_{1:i}} \sim N(h_{t_i}, H_{t_i})$ .
- 

## 2.5.2 Example: Local level model

Consider the local level model given by Equation (2.14), with  $\psi = (V, W)^T$ . When using a Gibbs sampler, we can firstly simulate  $\theta_{t_{0:n}}$  from  $\pi(\theta_{t_{0:n}} | \psi, x)$  using the FFBS

scheme outlined in Algorithm 4. The next step is to simulate  $\psi$ :

1. Simulate  $\psi_1$  from  $\pi(\psi_1|\psi_2, \theta_{t_{0:n}}, x) = \pi(\psi_1|\theta_{t_{0:n}}, x)$ .
2. Simulate  $\psi_2$  from  $\pi(\psi_2|\psi_1, \theta_{t_{0:n}}, x) = \pi(\psi_2|\theta_{t_{0:n}})$ .

Where  $\psi_1 = V$  and  $\psi_2 = W$ . Specifying independent inverse gamma prior distributions for  $\psi_1$  and  $\psi_2$  results in semi-conjugacy leading to tractable FCDs. Taking  $\psi_1 \sim \text{InvGa}(\alpha_v, \beta_v)$  and  $\psi_2 \sim \text{InvGa}(\alpha_w, \beta_w)$ , the FCD of  $\psi_1$  is calculated as

$$\begin{aligned} \pi(\psi_1|\psi_2, \theta_{t_{0:n}}, x) &\propto \pi(\psi, \theta_{t_{0:n}}, x), \\ &\propto \pi(x|\theta_{t_{1:n}}, \psi_1) \pi(\theta_{t_{1:n}}|\psi_2) \pi(\theta_{t_0}) \pi(\psi), \\ &\propto \prod_{i=1}^n \pi(x_{t_i}|\theta_{t_i}, \psi_1) \prod_{i=1}^n \pi(\theta_{t_i}|\theta_{t_{i-1}}, \psi_2) \pi(\theta_{t_0}) \pi(\psi_1) \pi(\psi_2), \\ &\propto \pi(\psi_1) \prod_{i=1}^n \pi(x_{t_i}|\theta_{t_i}, \psi_1), \\ &\propto \psi_1^{-(\alpha_v + \frac{n}{2})-1} \exp\left(-\frac{1}{\psi_1} \left[\beta_v + \frac{1}{2} \sum_{i=1}^n (x_{t_i} - \theta_{t_i})^2\right]\right). \end{aligned}$$

Hence,

$$\psi_1|\theta_{t_{1:n}}, x \sim \text{InvGa}\left(\alpha_v + \frac{n}{2}, \beta_v + \frac{1}{2} \sum_{i=1}^n (x_{t_i} - \theta_{t_i})^2\right).$$

Similarly,

$$\psi_2|\theta_{t_{0:n}} \sim \text{InvGa}\left(\alpha_w + \frac{n}{2}, \beta_w + \frac{1}{2} \sum_{i=1}^n (\theta_{t_i} - \theta_{t_{i-1}})^2\right).$$

When FCDs are available to sample from and a Gibbs sampler is used to sample from the joint posterior  $\pi(\psi, \theta_{t_{0:n}}|x)$ , retaining draws of  $\psi$  gives samples from the marginal parameter posterior  $\pi(\psi|x)$ . Alternatively, although the marginal likelihood is tractable, the posterior will typically be unavailable in closed form. Hence we use MH to generate draws from  $\pi(\psi|x)$ :

1. Initialise the chain with  $\psi^{(0)}$ . Set  $j = 1$ .
2. Propose  $\psi^* \sim q(\psi^*|\psi^{(j-1)})$ .

3. Calculate the acceptance probability  $\alpha(\psi^*|\psi^{(j-1)})$  of the proposed move, where

$$\begin{aligned}\alpha(\psi^*|\psi^{(j-1)}) &= \min \{1, A(\psi^*|\psi^{(j-1)})\} \\ &= \min \left\{ 1, \frac{\pi(\psi^*|x)q(\psi^{(j-1)}|\psi^*)}{\pi(\psi^{(j-1)}|x)q(\psi^*|\psi^{(j-1)})} \right\}\end{aligned}$$

4. With probability  $\alpha(\psi^*|\psi^{(j-1)})$ , set  $\psi^{(j)} = \psi^*$ ; otherwise set  $\psi^{(j)} = \psi^{(j-1)}$ .

5. Set  $j = j + 1$ . Return to step 2.

Then, given draws from  $\pi(\psi|x)$ , the backward sampler can be used to generate draws from  $\pi(\theta|x)$ .

### 2.5.2.1 Simulated data example

To illustrate the methods, we draw simulated data from a local level DLM as in Equation (2.14), with  $V = 1$ ,  $W = 0.5$  and  $\theta_{t_0} = 0$ . We simulate data for  $n = 150$  time points as follows

- For  $i = 1, \dots, n$  :
  1. Draw  $\theta_{t_i} \sim N(\theta_{t_{i-1}}, W)$ .
  2. Draw  $x_{t_i} \sim N(\theta_{t_i}, V)$ .

We assume the variance components,  $V$  and  $W$ , follow independent Gamma distributions such that,

$$V \sim Ga(a_v, b_v); \quad W \sim Ga(a_w, b_w),$$

with  $a_v, a_w, b_v, b_w = 0.1$ . Using an MH MCMC scheme, we propose values of  $\psi = (V, W)$  at each iteration from a normal random walk where the forward filter is used to evaluate the marginal likelihood for use in the acceptance probabilities. The inference is run initially for 5k iterations with the sampled parameter values used to estimate  $Var(\psi|x)$  for use as a tuning matrix in the main monitoring run. It is then run for 30k iterations with a thin of 30 to gain sufficiently low auto-correlation. The marginal posterior distributions of  $V$  and  $W$  are summarised by plots shown in Figure 2.4. The trace plots show good mixing and convergence to the target distribution. The density plots show how the data has been informative in forming the posterior as they both now show peaks consistent with the true values,

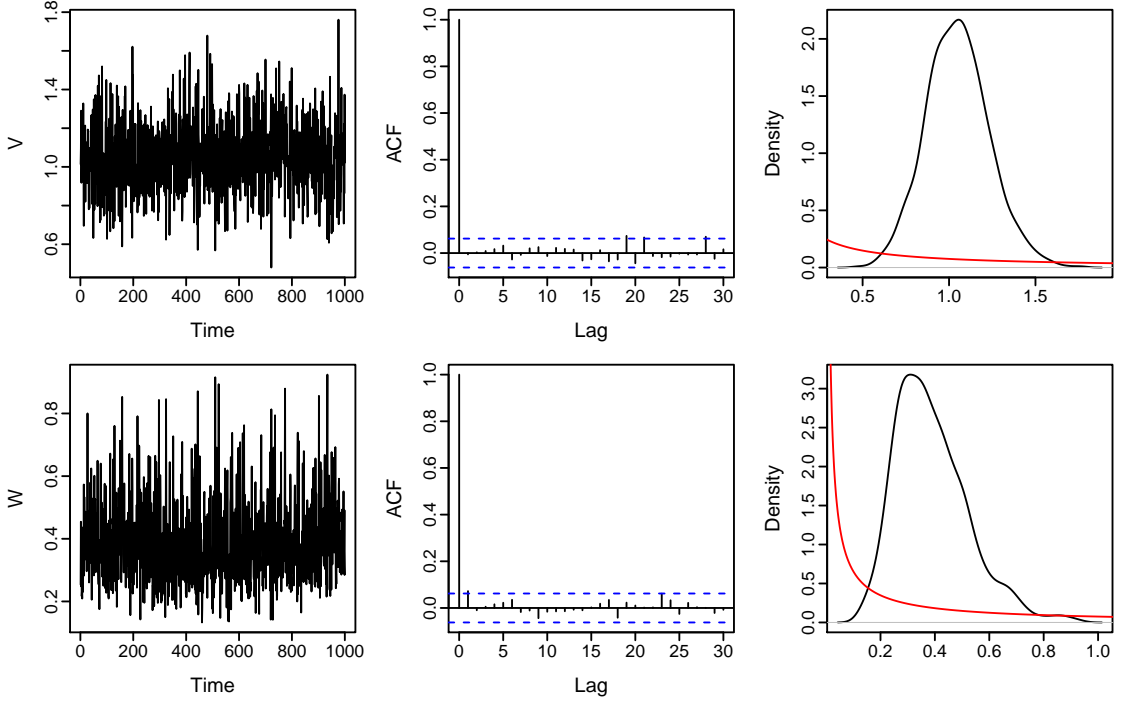


Figure 2.4: Left: trace plots, middle: ACF plots, and right: density plots of posterior estimates of  $V$  and  $W$  from the MCMC scheme, with the prior distributions overlaid in red

as opposed to the prior (shown in red). The stored values of  $m_{t_i}$  and  $C_{t_i}$  were then fed into the backward sampler, along with the posterior estimates of  $\psi$  to draw samples of  $\theta_{t_i}$ . Figure 2.5 summarises the output,  $\pi(\theta_{t_{0:n}}|\psi, x)$ , through the posterior mean in blue with 95% credible intervals in green overlaid on the real values of  $\theta_{t_i}$ . They are consistent with the true values, with the majority of the true values lying within the 95% credible intervals.

Furthermore, to check the validity of the model we can look at within-sample predictions. The within-sample predictive density is given by

$$\pi(\hat{x}_{t_{1:n}}|x_{t_{1:n}}) = \int \int \pi(\hat{x}_{t_{1:n}}|\theta_{t_{1:n}}, \psi)\pi(\theta_{t_{1:n}}, \psi|x_{t_{1:n}}) d\theta_{t_{1:n}} d\psi,$$

where

$$\pi(\theta_{t_{1:n}}, \psi|x_{t_{1:n}}) = \pi(\theta_{t_{1:n}}|\psi, x_{t_{1:n}})\pi(\psi|x_{t_{1:n}}).$$

Although the within-sample predictive density is intractable, draws from  $\pi(\theta_{t_{1:n}}, \psi|x_{t_{1:n}})$  are readily available and therefore  $\pi(\hat{x}_{t_{1:n}}|x_{t_{1:n}})$  can be sampled via Monte Carlo.

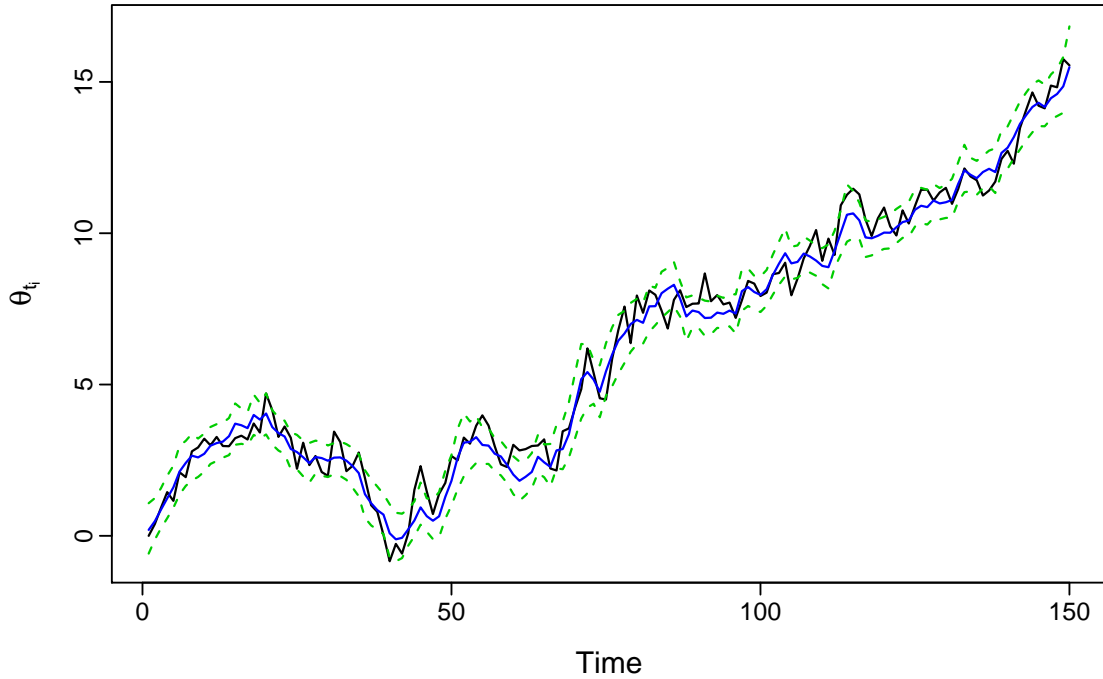


Figure 2.5: Posterior means (blue) and 95% credible intervals (green) for  $\theta_{t_i}$  from backward sampling with the true values plotted in black.

Given draws from  $(\psi^{(r)}, \theta_{t_{1:n}}^{(r)})$ ,  $r = 1, \dots, N$ , we can simulate

$$\hat{X}_{t_i}^{(r)} | \theta_{t_i}^{(r)}, \psi^{(r)} \sim N(F_{t_i} \theta_{t_i}^{(r)}, V^{(r)}),$$

for  $i = 1, \dots, n$  and where  $r$  is the number of iterations. Figure 2.6 shows the majority of the observed data falling comfortably within the 95% CIs of the within-sample predictions of  $y_{t_i}$  for all  $t_i$ , as expected for this synthetic data example.

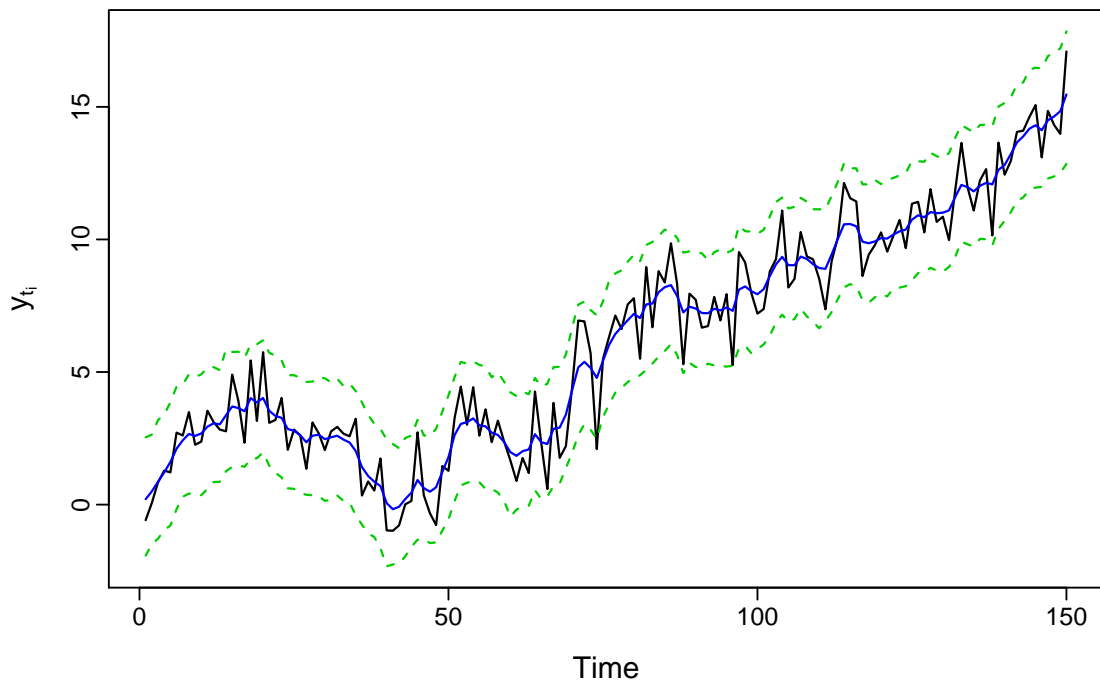


Figure 2.6: Observed data (black) with overlaid within-sample predictions – mean (blue) and 95% credible intervals (green).

# Chapter 3

## Statistical methods for traditional before/after studies

### 3.1 Introduction

Transportation-related safety schemes can include objectives such as safer vehicles, safer users, safer roads or any combination of these. Some common examples of road safety schemes include speed cameras, changes to road layout and vehicle activated signs. Most countries have a department for transport or similar who are responsible for making the road safer for all road users. The Global Network of road safety legislators was established in 2016 to address road traffic deaths and injuries globally through their corresponding parliamentarians. One of their aims is to encourage the UN to develop a set of global guidelines for standards for road safety ([The Global Network for Road Safety Legislators, 2016](#)).

Most road safety schemes (e.g. speed cameras) are evaluated retrospectively through ‘before/after studies’ ([Hauer, 1997](#)). In the UK, historically, Road Safety Partnerships (e.g. the Northumbria Safer Roads Initiative in northeast England) have been responsible for safety on the road network for which the local/regional police force has jurisdiction. These partnerships monitor their road networks continuously, identifying potential sites ‘with promise’; that is, sites that might be worthy of treatment of some sort. The Highway Safety Improvement Program (HSIP) Evaluation Guide ([Gross and Scury, 2017](#)) from the United States Department of Transportation notes that ‘evaluations are limited to sites already selected for treatment

based on other reasons including safety concerns'. Sites with promise are usually then monitored more closely during some predetermined observation ('before') period, over which vast amounts of data are collected (for example, over a three year observation period, within a predefined radius of each identified 'site', data on the number accidents, daily traffic volume and average speed might be collected). Of course, other site-specific data are available, including static data such as speed limit, road classification and surrounding land-use. The UK has access to so-called STATS19 data which has vast amounts of information associated with each individual accident (data collected by the police at the scene of the accident, mostly pertaining to casualty-level information – e.g. age of drivers involved in the accident, severity of the accident etc. ([Department for Transport, 2013](#))). At the end of the observation period, sites at which a threshold level of accident counts have been observed are selected for treatment; treatment is applied; and then data are collected during a subsequent treatment ('after') period. Until very recently, partnerships in the UK performed extremely rudimentary analyses of their before/after data, often just citing overall percentage changes in accident counts between the observation and treatment periods, and attributing the reduction entirely to the intervention.

Of course, such simple analyses are extremely over-simplistic, and treatment effects are bound to be exaggerated in this way because of selection bias: sites have been selected for treatment based on their abnormally high accident counts during the 'before' period, and counts in the subsequent 'after' period are bound to reduce anyway, even if no intervention was used. This is known as the regression to the mean (RTM) phenomenon, and is well-known and well-documented in statistics ([Fawcett and Thorpe, 2013](#), [Galton, 1889](#), [Hauer, 1980](#)).

## 3.2 Confounding factors in traditional studies

### 3.2.1 Regression to the mean

In road safety, practitioners make decisions on the implementation of safety schemes based upon data. However, issues in collecting data cause problems with the implementation or retention of safety schemes. Common issues include problems with sparsity of data, temporal trends or regression to the mean (RTM). Regression to the mean is a concept which refers to the fact that when a sample from a ran-



dom variable is extreme, eventually the following samples will return closer to the mean value. It was first documented by Sir Francis Galton in his paper ‘*Regression towards mediocrity in hereditary stature*’ (Galton, 1886), in which he noted the effect of parents who are well above the average population height having children who themselves are tall but ‘regressed towards the mean’ population height, hence shorter than their parents. The opposite was also shown true in shorter parents.

In the context of road safety, the first significant investigation into the RTM effect was by Ezra Hauer (Hauer, 1980, 1986) and can be defined in this context as *selection bias* when a treatment is applied non-randomly based on the responses on the individuals that are treated (see also, Elvik (1997), Mountain et al. (1998)). Usually, decisions on implementation of road safety schemes are based upon data collected at numerous sites. The data collected at these sites will include various traffic statistics such as traffic flow and average speed, alongside any recorded collisions. Commonly, when the number of collisions exceeds some threshold value, within a fixed time period, a safety scheme is introduced. However, the number of collisions is bound to fluctuate stochastically over time regardless of any safety scheme. This is due to collisions being caused by a variety of reasons, with one key factor being human error. This brings the question of whether the road is truly ‘unsafe’ as blame cannot be placed upon the location itself. Additionally, minor collisions are less likely to be reported and recorded formally meaning a lot of sites will have a collision count which is lower than the true value, bringing data quality issues. Recently, some countries are focusing on standardising the way collisions are reported, for example STATS19 in the UK, in a bid to remove some inconsistencies in the data. However, some inconsistencies are difficult to avoid, for example the exact location or time of a collision, hence we must conclude that it is unlikely that any dataset, especially those including minor collisions, are completely free of error.

Collision count data might be thought of as observations drawn from a Poisson distribution with rate  $\lambda$ , as the counts are discrete. From this we would expect the mean number of collisions to be  $\lambda$  and for other observations to fluctuate around this mean by, on average,  $\sqrt{\lambda}$ . If, for example, we had a site which has a mean number of 8 collisions per year ( $\lambda = 8$ ) from 2002–2018, and in 2019 we saw a surge of collisions to 12; practitioners may perceive this as a site which has become unsafe and therefore implement a safety treatment. Following this, in 2020, the treated

site observes 6 collisions. We may then conclude that the safety scheme has reduced the number of collisions by 6, however, without the safety scheme the site will have most likely seen a reduction naturally, say to 9 collisions (regressing to the mean). Therefore the safety scheme may have actually only helped avoid 3 collisions. That is why the RTM effect is important to take into account when evaluating safety schemes as we want to avoid overestimating their effectiveness.

The RTM effect varies but in some instances, the reduction in collision counts owing to RTM has been shown to be around 20–30% (Elvik, 2002, Fawcett and Thorpe, 2013, Hauer, 1997, Mountain et al., 1998). This might be viewed as the exaggerated effects of the road safety intervention, as often reported by police or in the media (Brooker and North, 2007). Hence, it is vital that this is accounted for when evaluating schemes when almost a third of the reduction could be inevitable.

If it was possible to know the number of collisions that would have taken place in the after period without the scheme, treatment effect elicitation would be trivial. We could determine that the treatment effect is the difference between the collision counts with and without the scheme in the after period,

$$\tau_i = \tilde{y}_{i,AFT} - y_{i,AFT},$$

where  $y_{i,AFT}$  is the recorded number of collisions at site  $i$  after treatment, and  $\tilde{y}_{i,AFT}$  is the unobserved collision count in the same period, at site  $i$ . Unfortunately, in reality we cannot know  $\tilde{y}_{i,AFT}$  and hence we must estimate it to allow us to deduce the treatment effect. Determining an estimate to the treatment effect at a site may seem simplistic, such as merely the difference between the collision counts in the before and after periods,

$$\hat{\tau}_i = y_{i,BEF} - y_{i,AFT}. \quad (3.1)$$

Upon setting  $m_i$  to be the underlying collision rate at site  $i$  then we have

$$\begin{aligned} \mathbb{E}[\hat{\tau}_i] &= \mathbb{E}[y_{i,BEF} - y_{i,AFT}], \\ &= \mathbb{E}[y_{i,BEF}] - \mathbb{E}[y_{i,AFT}], \\ &= m_i - (m_i - \tau_i), \\ &= \tau_i, \end{aligned}$$

and providing our assumptions of the expected values in each period are accurate, we have an unbiased estimator of the treatment effect,  $\tau_i$ . For the treatment effect over all  $n$  sites we would simply calculate

$$\hat{\tau} = \sum_{i=1}^n \hat{\tau}_i.$$

Unfortunately, this approach is over-simplistic and naive as it fails to account for RTM and other confounding factors, such as trend. As mentioned in Section 3.2.1, treated sites are most likely chosen where there is an abnormally high number of collisions in the before period. Hence, we would expect

$$\begin{aligned} y_{i,BEF} &= m_i + \rho_{i,BEF}, \\ y_{i,AFT} &= m_i + \rho_{i,AFT} + \tau_i, \end{aligned}$$

where  $m_i$  is the true, underlying collision rate, and  $\rho_{i,BEF}$  and  $\rho_{i,AFT}$  are the random deviations due to chance. Ergo, Equation (3.1) becomes

$$\begin{aligned} \hat{\tau}_i &= y_{i,AFT} - y_{i,BEF}, \\ &= (\rho_{i,AFT} - \rho_{i,BEF}) + \tau_i, \\ &= \rho_i + \tau_i, \end{aligned} \tag{3.2}$$

where  $\rho_i$  is the RTM effect at site  $i$ . If sites were chosen randomly, the RTM effect would be obsolete, and  $\rho_i = 0$ . However, as sites deemed dangerous through their high collision count are chosen, we induce a bias, such that

$$\mathbb{E}[\rho_{i,BEF}] > 0.$$

As the after period is free of selection bias and is independent to the before period we can set  $\mathbb{E}[\rho_{i,AFT}] = 0$ . Therefore, we have that the RTM effect,

$$\mathbb{E}[\rho_i] < 0.$$

Now, we no longer have an unbiased estimator of the treatment effect in Equation (3.1). Therefore, to regain unbiased estimates we must remove the RTM effect

from the estimate of the treatment effect,

$$\begin{aligned}\mathbb{E}[\hat{\tau}_i] &= \mathbb{E}[y_{i,AFT} - y_{i,BEF} - \hat{\rho}_i], \\ &= \rho_i + \tau_i - \rho_i, \\ &= \tau_i.\end{aligned}$$

Correctly quantifying the RTM effect is a challenge in before/after studies, as a decrease or increase in collision counts at a site could be attributed to this phenomenon. We estimate the RTM effect through,

$$\mathbb{E}[\rho_i] = y_{i,BEF} - \mathbb{E}[m_i],$$

where  $\mathbb{E}[\rho_i]$  represents the expected RTM effect,  $y_{i,BEF}$  is the observed collision count before implementing a safety measure, and  $\mathbb{E}[m_i]$  is the expected collision count in the absence of the safety measure.

### 3.2.2 Other confounders

Another feature not accounted for through such a simple comparison of ‘before’ with ‘after’ is trend: we might expect a change in accident counts between the observation and treatment periods owing to general temporal trends, perhaps as a result of road safety awareness campaigns. To derive Equation (3.2), we assume that the underlying collision rate,  $m_i$ , is constant and any changes to collision rates are down to RTM or treatment effects. In reality, there are a multitude of reasons as to why this rate could change over time; traffic volume, recent safety awareness schemes and vehicle safety improvements could all change the underlying rate in either direction. If we do not account for trend, our estimates for the treatment effect will be biased.

When accounted for, trend is usually factored in to any before/after analysis by applying a multiplicative factor to the change in counts between the ‘before’ and ‘after’ periods, based on historical records of accident counts across the region, or perhaps expert opinion on trend at the location (see, [Guo et al. \(2019\)](#), [Yanmaz-Tuzel and Ozbay \(2010\)](#)). Temporal trends can vary across different levels of severity, as evidenced by recent data from the UK which shows an annual rise in reported ‘slight’ accidents while accidents graded as ‘KSI’ (Killed or Seriously Injured) have steadily decreased. Therefore, taking into account the severity of accidents is an im-

portant factor to consider when building effective accident prediction models. Thus, careful consideration of accident severity might also play a part in successful accident prediction modelling. Other confounding effects include traffic volume, any other significant change to the site in the same period or simply general fluctuations in the number of accidents. [Elvik \(2002\)](#) studied seven before/after studies and found that not controlling for confounding factors lead to an overestimate of the effects of the road safety measure in five of the seven cases. Hence, such simplistic studies, which do not control for any confounding factors, should not be trusted.

Commonly, the amount of data available for before/after studies is relatively small and it's not possible to measure these effects from the treated site data alone; hence, we use data collected at comparison sites, where no scheme has been implemented, to gain estimates of the RTM and trend effects. These control sites are chosen to be similar to the treated sites in terms of characteristics, and the same amount of data are collected in the before and after periods. The differences in the before and after periods at the treated sites are then compared to control sites to account for RTM or other confounding factors ([Engel and Thomsen, 1992](#), [Yannis et al., 2014](#)). While simplistic before/after studies comparing treated and control sites can provide some insights into the safety effects of roadway treatments and account for confounding factors, a more advanced methodology is the empirical Bayes (EB) approach.

### 3.2.3 Road safety hotspot prediction

'Predictive policing' of the road network is now becoming an accepted part of practitioners' road safety plans. For example, installing a speed camera at a location with an abnormally high number of accidents during some observation period might be an unnecessary knee-jerk response, with wasted resource, given the discussion on RTM above. Hence, road safety organisations are increasingly looking towards accident prediction modelling to identify sites for treatment based on their predicted accident counts. The International Road Assessment Programme (iRAP) have developed AiRAP which uses artificial intelligence and machine learning to collect road data globally in the hope to support road safety assessment, crash risk mapping and investment prioritisation for all road users ([The International Road Assessment Programme \(iRAP\), 2022](#)). The supervisory team have also worked

closely with researchers at PTV to develop a model for predicting accidents with the following features:

- Makes use of information (observed or modelled) across a ‘global’ network (e.g. region/ municipality/ borough/ ward), obtained from within PTV software.
- Dynamically models global trend.
- Allows site-specific deviations from the global trend, should there be significant evidence at a site for such a deviation.
- Uses historical accident data (observed or modelled) at each site to estimate a site-specific underlying mean accident rate.
- Incorporates a device to allow more recent observations at a site to inform predictions at that site with more certainty, relative to observations further in the past.

The model developed with PTV is now available within PTV’s Visum software, as well as a locally-hosted prototype developed at Newcastle (essentially a suite of R-Shiny web-based applications). Currently, 19 road safety organisations globally are using this hotspot prediction model, some of which are using it to help guide investment decisions relating to road safety interventions. Many of the prediction models used, heavily rely on covariate information to provide reliable predictions; however, data availability is restricted. It is common to have collision information but not necessarily any matching covariates, such as traffic flow or average speed. Commonly, these data need to be outsourced and combined with the collision data. This matching exercise is time-consuming and not always straightforward. In Chapters 6 and 7, we present models that utilise available data to identify patterns in the response variable. Our proposed model in Chapter 7 is capable of predicting on more precise scales, providing finer-scale predictions.

### 3.3 The empirical Bayes approach for RTM

When a road safety organisation or practitioner wants to account for RTM, they usually do so through an ‘Empirical Bayes’ (EB) approach (Fawcett and Thorpe, 2013, Fawcett et al., 2017, Hauer, 1980). Here, collision counts in the before period,  $y_i$  at site  $i$ , are assumed Poisson (see Appendix A.1.7) conditional on rate  $m_i$ . As

described in Section 2.1, the unconditional distribution for  $y_i$  is negative binomial with mean  $\mu_i$  and variance  $\mu_i + \kappa\mu_i^2$ , where  $\kappa = 1/\gamma > 0$  is the negative binomial overdispersion parameter. For mathematical convenience, as the gamma distribution is the conjugate (from the same distribution family) prior for the Poisson,  $m_i$  is assumed to be gamma distributed, such that

$$\begin{aligned} y_i|m_i &\sim Po(m_i), \quad \text{independently,} \\ m_i &\sim Ga(\gamma, \gamma/\mu_i), \end{aligned}$$

where  $\gamma$  is the reciprocal of the overdispersion parameter of the unconditional negative binomial distribution of the collision counts  $y_i$ . The PDF for  $m_i$  is,

$$\pi(m_i) = \frac{1}{\Gamma(\gamma)} \left(\frac{\gamma}{\mu_i}\right)^\gamma m_i^{(\gamma-1)} \exp\left\{-\frac{\gamma}{\mu_i}m_i\right\},$$

strategically parameterised so that we have mean  $\mu_i$  and variance proportional to the overdispersion parameter. Using Bayes theorem, we can calculate the posterior up to proportionality,

$$\begin{aligned} \pi(m_i|y_i) &\propto m_i^{(\gamma-1)} \exp\left\{-\frac{\gamma}{\mu_i}m_i\right\} \times \frac{m_i^{y_i}}{y_i!} \exp\{-m_i\}, \\ &\propto m_i^{(y_i+\gamma-1)} \exp\left\{-m_i\left(\frac{\gamma}{\mu_i} + 1\right)\right\}. \end{aligned}$$

Then, the posterior distribution for  $m_i$  is found to be

$$m_i|y_i \sim Ga(\gamma + y_i, \gamma/\mu_i + 1), \tag{3.3}$$

with posterior mean

$$\begin{aligned} \mathbb{E}[m_i|y_i] &= \alpha_i\mu_i + (1 - \alpha_i)y_i, \\ \text{where } \alpha_i &= \gamma/(\gamma + \mu_i), \quad 0 < \alpha_i < 1. \end{aligned} \tag{3.4}$$

The effectiveness of a scheme is then measured through a comparison of  $y_{i,AFT}$  and  $\mathbb{E}[m_i|y_i]$ . The standard EB approach treats the posterior mean,  $\mathbb{E}[m_i|y_i]$ , as a ‘counterfactual outcome’ – in other words, it uses the underlying mean accident rate at each site ( $\mu_i$ ) and the actual observed count in the before period ( $y_i$ ) to suggest what the accident count might have been had no intervention been used. It then

compares this count (rather than the count from the ‘before’ period) to that in the ‘after’ period to account for the fact that the observed count in the before period may have been abnormally high.

Data from a reference set of sites, at which no treatments have been applied, are usually used to estimate the underlying mean accident rate  $\mu_i$  as, typically, historical records are not available. Simple log-linear regression models (so-called ‘Accident Prediction Models’ (APMs) or Safety Performance Functions (SPFs)) linking collision counts to several regressors are constructed for data at reference sites, and then applied to regressor information at the treated sites, in order to estimate  $\mu_i$ ,

$$\hat{\mu}_i = \exp \left\{ \hat{\beta}_0 + \sum_{p=1}^{n_p} \hat{\beta}_p x_{pi} \right\},$$

where  $x_{pi}$  are variables associated with attributes at site  $i$  that could have an effect on the mean number of casualties at that site (e.g. traffic flow or average vehicular speed) and  $n_p$  is the number of such variables used. The estimated regression coefficients,  $\hat{\beta}_p$ ,  $p = 0, \dots, n_p$ , are obtained from analyses on reference sites. Reference sites are chosen such that their attributes (used as explanatory variables) are as identical as possible to the treated sites, yet they observe a more ‘normal’ collision frequency. The treated sites have been chosen because of their abnormally high collision counts, but for  $\mu_i$  we need a model that will give us a more representative prediction of the mean collision count at the treated sites  $i$ . Estimation of the weight  $\alpha_i$  and thus the overdispersion parameter  $\gamma$  is required; to do this, we set the SPF to be a negative binomial GLM (see, Section 2.1), which we fit using maximum likelihood estimation and obtain our SPF coefficient vector  $\hat{\beta} = (\beta_0, \dots, \beta_{n_p})$  and an estimate of the inverse of the overdispersion parameter,  $\kappa$ . We can then gain our EB estimates of the collision frequency at each site  $i$  through Equation (3.4).

In the absence of reference sites, regional or national models are sometimes relied upon to provide an ‘off-the-shelf’ estimate of  $\mu_i$  (e.g. the ‘COBA manual’ in the UK). Further examples detailing the RTM effect, and the potential pitfalls it causes can be found in [García-Gallego et al. \(2011\)](#); further examination on the uses of empirical Bayes in before/after studies can be found in [Persaud and Lyon \(2007\)](#).

It is standard practice for frequentist techniques to be employed to estimate the



regression coefficients in the APM, which then feed through the Bayesian analysis as outlined above. Of course, in a truly Bayesian setting, these parameters' coefficients might have prior distributions themselves, and simulation-based procedures might be needed (e.g. MCMC) to estimate all parameters within a 'fully Bayesian' (FB) context. Doing so could more realistically quantify our uncertainty in estimates of RTM and treatment effects by acknowledging our uncertainty in the estimated APM.

Previous reluctance to embrace an FB methodology in industry can be explained by the enhanced level of computing and statistical ability needed to implement this method, fortunately advances of computing and software applications greatly improve the accessibility and hence, FB methods are becoming more commonplace (El-Basyouny and Sayed, 2012, Heydari et al., 2014). As an example, Bayesian methodology is used by PTV group in their software VISUM and by Gateshead council through RAPTOR. The RAPTOR suite of software applications was developed by the Newcastle road safety team to allow practitioners to implement Bayesian methods, without any technical or computational requirements (Matthews et al., 2019). The advantage stems from use of diffuse prior distributions for the coefficients in the regression equation. This allows for more realistic standard deviations and doesn't accept the estimated regression coefficients as fixed (known) values. In addition, it also allows for change of priors omitting the need for the Poisson-gamma conjugacy. Despite FB methods being accepted in the literature for a while, with Schlüter et al. (1997) providing support for a hierarchical model to replace EB in the 1990s, EB still remains in common usage currently with Høye (2015), Park and Abdel-Aty (2015), Wang et al. (2017) providing examples of studies carried out in the last several years still relying on an EB methodology for inference. For more examples in favour of a fully Bayes analyses, see Kitali and Sando (2017), Lan et al. (2009), Yanmaz-Tuzel and Ozbay (2010).

## 3.4 Data example: The Northumbria Safety Camera Partnership data

### 3.4.1 Empirical Bayes analysis

Two datasets were provided by Northumbria Safety Camera Partnership (NSCP): 56 treatment sites, where mobile safety cameras were operated, with data from a before period and an after period, and 67 reference sites to account for RTM. The data were collected from 2001–2003 in the before period and 2004–2006 in the after period. This dataset was introduced in Section 1.3.1. Generalised linear modelling and maximum likelihood estimation is used to obtain the prior mean  $\mu_i$  from the reference set,

$$\begin{aligned} \hat{\mu}_i = \exp\{ & \underset{(0.534)}{1.933} - \underset{(0.015)}{0.041x_1} - \underset{(0.004)}{0.013x_2} + \underset{(0.193)}{0.444x_3} \\ & + \underset{(0.417)}{0.674x_{4I}} + \underset{(0.422)}{0.846x_{5I}} + \underset{(0.380)}{1.060x_{6I}} \}, \end{aligned} \quad (3.5)$$

where  $x_1$ ,  $x_2$  and  $x_3$  correspond to average observed speed (mph), percentage of drivers over speed limit and traffic flow (respectively), and  $x_{4I}$ ,  $x_{5I}$  and  $x_{6I}$  are indicator variables associated with road classification (A, B and C). The regression coefficients,  $\hat{\beta}_p$ , are represented by maximum likelihood estimates with their standard errors given in parentheses underneath. The overdispersion parameter is given as  $\hat{\gamma} = 2.494$ ; maximum likelihood estimates and standard errors are from fitting the ‘glm’ function in R to the reference data. With estimates available for all regression coefficients and the overdispersion parameter, we can calculate our estimates for  $\mu_i$  at each site and in turn our weight value  $\alpha_i$ . We are able to estimate  $\mathbb{E}(m_i|y_i)$ , the number of casualties we would expect to see had no scheme been implemented via Equation (3.4). The standard deviation can be calculated through the square root of the variance of the posterior distribution,  $\pi(m_i|y_i)$  in Equation (3.3). Table 3.1 shows that after RTM we would expect to see, on average, 317 casualties over all sites, had no treatment been applied. We conclude that, through the EB method, the speed cameras have resulted in a reduction, on average, of 19 casualties across the sites.

	EB Method					$y_{i,AFT}$	Difference	
	$y_i$	$\mu_i$	$\alpha_i$	$\mathbb{E}(m_i y_i)$	$SD(m_i y_i)$		Observed	After RTM
Site $i = 2$	4	1.673	0.599	2.607	1.023	0	-4	-2.607
Site $i = 13$	12	2.063	0.547	6.561	1.723	2	-10	-4.561
Site $i = 39$	7	1.464	0.630	3.511	1.140	2	-5	-1.511
Site $i = 47$	16	4.809	0.341	12.179	2.832	5	-11	-7.179
<b>Total</b>	<b>438</b>			<b>317</b>		<b>298</b>	<b>-138</b>	<b>-19</b>

Table 3.1: Results of the EB analysis to account for RTM for four sites treated with safety cameras, as well as totals for all 56 safety camera sites.

### 3.4.2 Fully Bayes analysis

It is standard practice in road safety to use frequentist techniques (EB) for estimates of the regression coefficients and to feed these through to a Bayesian analysis, using these maximum likelihood estimates to build our prior distribution for  $m_i$ . However in a fully Bayesian (FB) setting we assign priors to the regression coefficients to quantify our uncertainty in these parameters. Therefore, moving to the FB method, we initially work with the same Poisson-gamma hierarchy, however, we now assign prior distributions to the regression coefficients  $\beta_p, p = 0, \dots, 6$ , and the negative binomial overdispersion parameter  $\kappa = 1/\gamma$ . We use diffuse, independent priors:

$$\beta_p \sim N(0, 100), \quad p = 0, \dots, 6 \quad \text{and}$$

$$\rho = \log \kappa \sim N(0, 100),$$

working with the natural logarithm to retain the positivity of  $\kappa$ , we can then gain estimates of the overdispersion parameter as  $\gamma = \exp(-\rho)$ . We use a Markov chain Monte Carlo (MCMC) algorithm with normal random walk proposals to draw realisations from the posteriors.

---

**Algorithm 5** Fully Bayes MCMC scheme

---

1: Initialise the chain  $\boldsymbol{\beta}^{(0)} = (\beta_0^{(0)}, \dots, \beta_6^{(0)})$ ,  $\rho^{(0)}$  at their MLEs. Set  $j = 1$ .

2: For each element  $\beta_p$ :

- Sample a proposal value  $\beta_p^*$  from the proposal distribution  $q(\beta_p^*|\beta_p^{(j-1)})$ ,

$$\beta_p^* \sim N(\beta_p^{(j-1)}, \Sigma_{\beta_p}),$$

where  $\Sigma_{\beta_p}$  is determined from tuning the acceptance probabilities.

---

- Set  $\beta_p^{(j)} = \beta_p^*$  with probability  $\alpha_{\beta_p}$ , where

$$\alpha_{\beta_p} = \min \left\{ 1, \frac{\pi(\beta_p^*)L(\boldsymbol{\beta}^*|\rho^{(j-1)}, \mathbf{x})q(\beta_p^*|\beta_p^{(j-1)})}{\pi(\beta_p^{(j-1)})L(\boldsymbol{\beta}^{(j-1)}|\rho^{(j-1)}, \mathbf{x})q(\beta_p^{(j-1)}|\beta_p^*)} \right\},$$

where  $\mathbf{x}$  denotes the reference data; else set  $\beta_p^{(j)} = \beta_p^{(j-1)}$ .

3: Update  $\rho$ .

- Sample a proposal value  $\rho^*$  from the proposal distribution  $q(\rho^*|\rho^{(j-1)})$ ,

$$\rho^* \sim N(\rho^{(j-1)}, \Sigma_\rho),$$

where  $\Sigma_\rho$  is determined from tuning the acceptance probabilities.

- Set  $\rho^{(j)} = \rho^*$  with probability  $\alpha_\rho$ , where

$$\alpha_\rho = \min \left\{ 1, \frac{\pi(\rho^*)L(\rho^*|\boldsymbol{\beta}^{(j)}, \mathbf{x})q(\rho^*|\rho^{(j-1)})}{\pi(\rho^{(j-1)})L(\rho^{(j-1)}|\boldsymbol{\beta}^{(j)}, \mathbf{x})q(\rho^{(j-1)}|\rho^*)} \right\},$$

else set  $\rho^{(j)} = \rho^{(j-1)}$ . Set  $\gamma^{(j)} = \exp(-\rho^{(j)})$ .

4: For  $i = 1, \dots, 56$ ,

- Calculate  $\mu_i$ .
- Sample a proposal value  $m_i^*$  from the proposal distribution  $q(m_i^*|m_i^{(j-1)})$ ,

$$m_i^* \sim N(m_i^{(j-1)}, \Sigma_{m_i}),$$

where  $\Sigma_{m_i}$  is determined from tuning the acceptance probabilities.

- Set  $m_i^{(j)} = m_i^*$  with probability  $\alpha_{m_i}$ , where

$$\alpha_{m_i} = \min \left\{ 1, \frac{\pi(m_i^*|\mu_i, \gamma^{(j)})L(m_i^*|y_i)q(m_i^*|m_i^{(j-1)})}{\pi(m_i^{(j-1)}|\mu_i, \gamma^{(j)})L(m_i^{(j-1)}|y_i)q(m_i^{(j-1)}|m_i^*)} \right\},$$

where  $y_i$  denotes the treated sites data; else set  $m_i^{(j)} = m_i^{(j-1)}$ .

5: If  $j = N$  stop, else set  $j = j + 1$  and go to step 2.

---

The MCMC scheme, outlined in Algorithm 5, was run for 400k iterations with a thin of 400 and discarding the first 1000 as burn-in. The tuning parameters for each

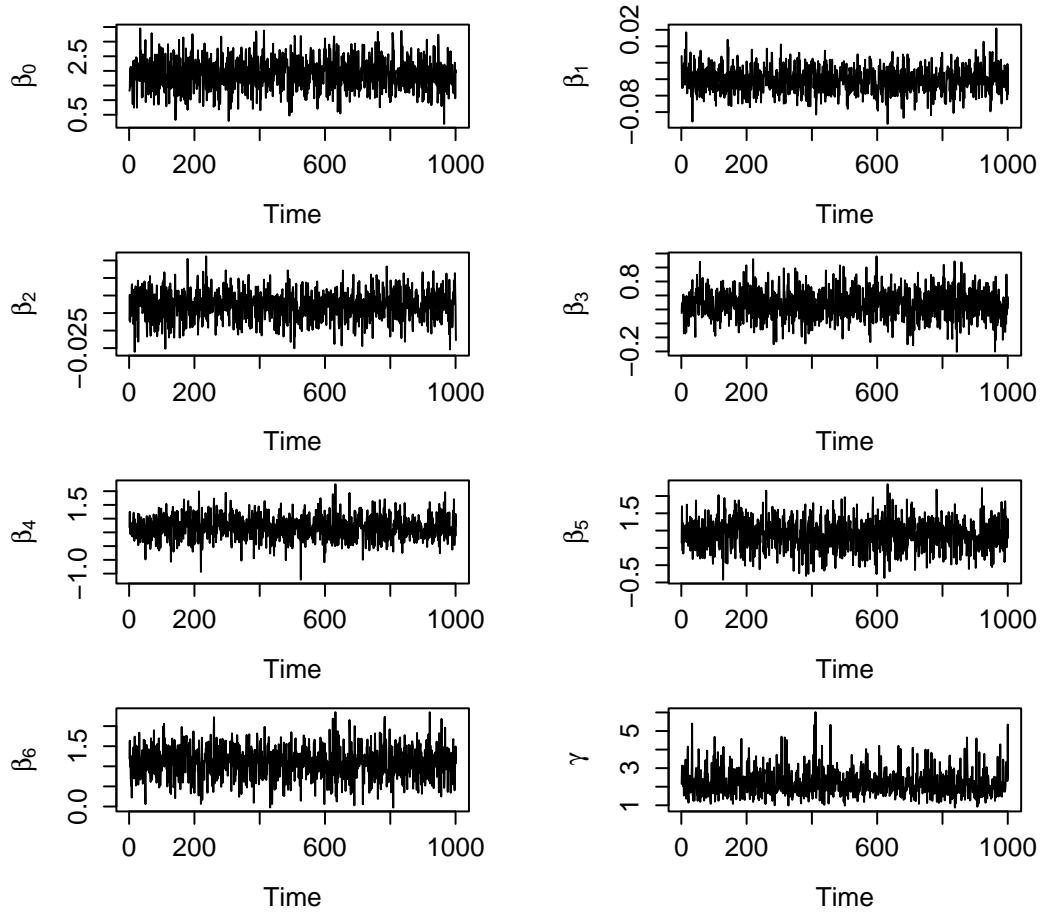


Figure 3.1: Trace plots for  $\beta_p$ ,  $p = 0, \dots, 6$  and  $\gamma$  from the MCMC scheme.

coefficient were chosen suitably to give acceptance probabilities of around 23%. The trace plots in Figure 3.1 show good convergence for all and have similar means and variances in different sections of the chains. The autocorrelation was sufficiently low after thinning and the histograms in Figure 3.2 show good estimates of marginal densities.

In Table 3.2, we show estimates for  $\mu_i$  and  $m_i$  for the same 4 sites as shown in Table 3.1 for illustrative purposes. As can be seen by comparing standard deviations for  $m_i$  from the FB analysis to those from the EB analysis in Table 3.1, posterior variability is substantially greater in the FB analysis. This is because we have acknowledged the variability of the regression parameters  $\beta_p$ ; hence, this makes our estimation for the accident prediction model (APM) more realistic as it quantifies our uncertainties in these coefficients. We see in Figure 3.3 how an FB analysis

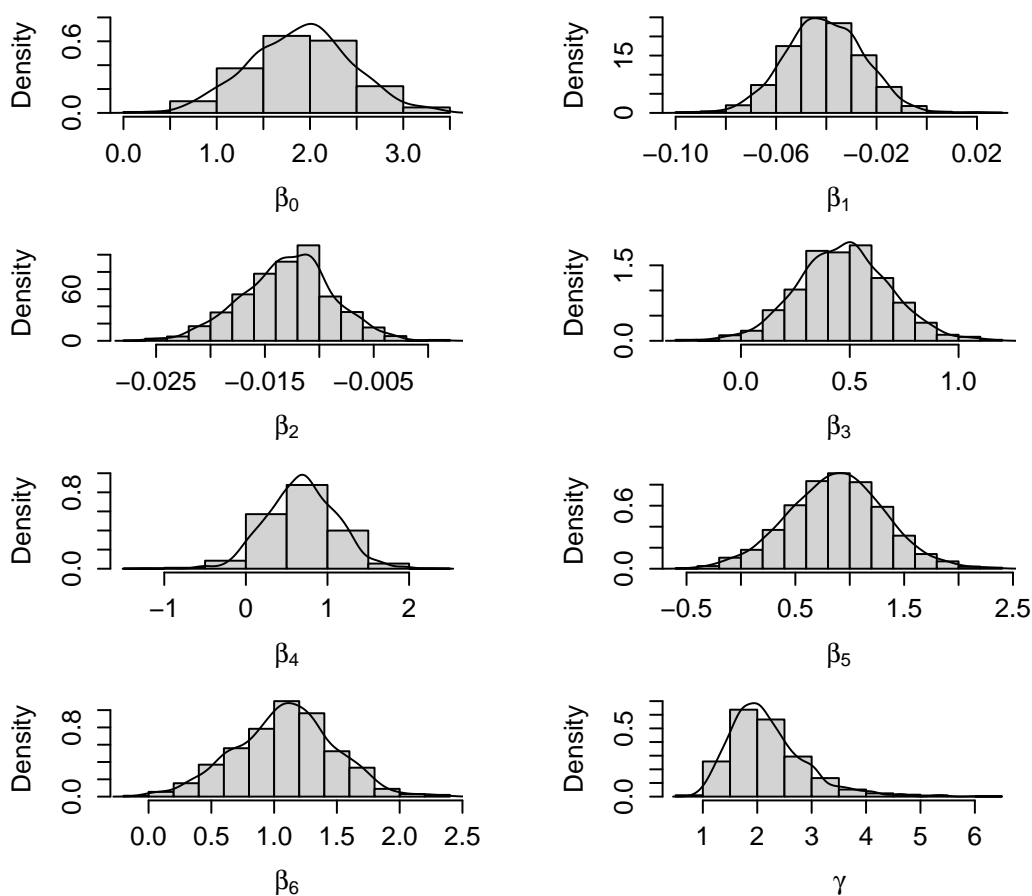


Figure 3.2: Histograms for  $\beta_p$ ,  $p = 0, \dots, 6$  and  $\gamma$  from the MCMC scheme showing marginal posterior densities.

is more realistic over the uncertainty of the parameters, with most showing larger posterior standard deviations for  $m_i$ . Additionally, by comparing the estimates of  $\mathbb{E}(m_i|y_i)$  obtained from the FB and EB analyses, an FB analysis estimates that we would expect to observe fewer casualties in the after period when accounting for RTM.

### 3.4.3 Sensitivity to prior distribution

As discussed earlier, the Poisson-gamma hierarchy was chosen out of mathematical convenience to give us a convenient form for the expectation of the posterior. However, simulation-based methods, such as MCMC, are relatively straightforward and allow us to draw realisations from the posterior without the necessity for conjugate

		Posterior			
		Mean	St. Dev.	Median	95% Credible interval
	$\beta_0$	1.962	0.539	1.962	(0.816, 2.963)
	$\beta_1$	-0.042	0.015	-0.041	(-0.073, -0.015)
	$\beta_2$	-0.013	0.004	-0.013	(-0.021, -0.005)
	$\beta_3$	0.465	0.216	0.465	(0.030, 0.848)
	$\beta_4$	0.668	0.489	0.660	(-0.268, 1.469)
	$\beta_5$	0.858	0.450	0.846	(0.014, 1.773)
	$\beta_6$	1.074	0.391	1.064	(0.296, 1.818)
	$\gamma$	2.167	0.680	2.093	(1.077, 3.488)
$\mu_i$	Site $i = 2$	1.821	0.852	1.661	(0.701, 3.983)
	Site $i = 13$	2.384	1.344	2.113	(0.809, 5.461)
	Site $i = 39$	1.587	0.722	1.445	(0.613, 3.370)
	Site $i = 47$	4.976	0.971	4.888	(3.322, 7.157)
$m_i$	Site $i = 2$	2.676	1.284	2.459	(0.860, 5.961)
	Site $i = 13$	7.070	2.642	6.728	(2.956, 12.949)
	Site $i = 39$	3.711	1.589	3.419	(1.308, 7.568)
	Site $i = 47$	12.543	3.261	12.267	(6.892, 19.292)
	<b>Total</b>	<b>311.915</b>	<b>37.286</b>	<b>311.015</b>	<b>(258.706, 366.825)</b>

Table 3.2: Posterior summaries for a fully Bayesian analysis to account for RTM for four treated sites and the total over all 56 sites.

priors. In the last Section we compared the results from a traditional EB analysis for a safety camera before/after study to those from an analogous FB analysis, noting the increase in uncertainty under the more ‘honest’ FB analysis. Here, we will compare results from five different prior specifications; all are chosen such that they have support on  $(0, \infty)$  (as we are specifying a prior for a rate parameter). The estimates of the marginal densities for each of the coefficients are not dependent on the prior for  $m_i$ , hence, these estimates vary only marginally over each MCMC output.

We explore the sensitivity of the posterior for  $m_i$  to the following prior distributions: Gamma; Weibull; lognormal; beta-prime; and inverse-gamma (see Appendix A.1). Additionally, we use a Jeffreys prior as an uninformative prior to assess the relevance of the reference data. To allow for relative comparisons between the different priors, we keep the mean and variance the same in all specifications, that is, we hold  $\mathbb{E}(m_i) = \mu_i$  and  $Var(m_i) = \mu_i^2/\gamma$ , as was the case in the Poisson-gamma specification in the last Section, with the prior mean given by the accident prediction model (APM) in Equation (3.5). For example, in the case of the lognormal

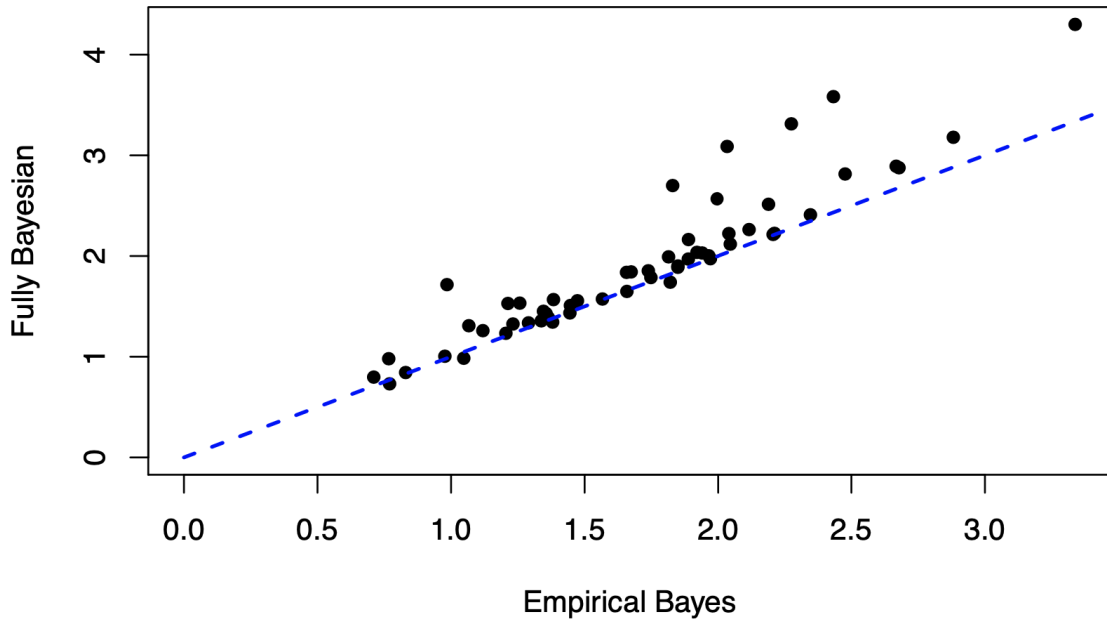


Figure 3.3: Standard deviations for  $m_i$  from EB analysis against Poisson-Gamma FB analysis

distribution, we reparameterise as

$$\begin{aligned}
 m_i &\sim \text{Lognormal}(\lambda_i, \sigma^2), \quad \text{where} \\
 \lambda_i &= \log(\mu_i) - \frac{1}{2} \log(1 + \gamma^{-1}) \quad \text{and} \\
 \sigma^2 &= \log(1 + \gamma^{-1}).
 \end{aligned}
 \tag{3.6}$$

For the inverse-gamma prior, we have

$$\begin{aligned}
 m_i &\sim \text{InvGa}(\alpha, \beta), \\
 \mu_i &= \frac{\beta}{(\alpha - 1)}, \\
 \frac{\mu_i^2}{\gamma} &= \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)},
 \end{aligned}$$

giving  $\alpha$  and  $\beta$  in terms of  $\mu_i$  and  $\gamma$ .

We use the Deviance Information Criterion (DIC; see Section 2.4.4) to compare the five model formulations outlined here. Table 3.3 shows posterior summaries for accident rates  $m_i$  totalled across all 56 sites, with summaries from the original Poisson-gamma formulation, as reported in Table 3.2, included for completeness.



	Gamma prior		Lognormal prior		Weibull prior	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
<b>Total</b>	311	(258, 366)	339	(248, 400)	317	(296, 371)
<b>RTM (%)</b>	-28.9	-(40.8, 15.6)	-22.7	-(36.7, 8.7)	-27.6	-(39.3, 15.3)
<b>DIC</b>	666.3		787.2		645.6	
	Beta-prime prior		Inv-gamma prior		Jeffreys prior	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
<b>Total</b>	333	(290, 378)	339	(298, 381)	464	(421, 506)
<b>RTM (%)</b>	-23.9	-(33.8, 13.7)	-22.9	-(31.9, 12.9)	5.9	(-3.7, +15.6)
<b>DIC</b>	754.4		773.5		805.0	

Table 3.3: Posterior means and 95% credible intervals for the total expected number of casualties and the RTM effect. Also shown are the computed values of the Deviance Information Criterion.

Also reported in Table 3.3 are posterior summaries for the estimated RTM percentage change, that is,

$$100 \left( \frac{\sum_{i=1}^{56} \mathbb{E}(m_i|y_i) - y_i}{\sum_{i=1}^{56} y_i} \right),$$

along with the DIC values for each model specification. As can be seen, there are some discrepancies in results when comparing across the five informative prior specifications. Both the lognormal and inverse-gamma priors give posterior inferences suggesting a reduction in casualties, after accounting for RTM, to around 339, compared to 333 when using a beta-prime prior or around 311 when using a gamma or 317 with a Weibull prior. This would indicate an inflated estimated treatment effect under the lognormal, inverse-gamma or beta-prime prior specifications, relative to an analysis based on a gamma or Weibull prior specification. Correspondingly, the estimated RTM effects are larger under the gamma and Weibull prior specifications, relative to the lognormal, inverse-gamma and beta-prime specifications. The DIC values suggest the Poisson-Weibull specification might be the most appropriate here, followed closely by the Poisson-gamma formulation. The other three model formulations using informed priors have returned much larger DIC values, perhaps ruling them out in this particular application. The Jeffreys prior doesn't refer to the reference data at all, and so expects the number of casualties to continue being abnormally high. Only in the very lower tail of the marginal distribution would we see a reduction in total casualties over all sites. Without the reference data, the RTM effect is somewhat ignored, as we might expect.

The results reveal some sensitivity in estimates of RTM to the choice of prior distribution for the mean casualty rate used in the NSCP study on the effectiveness

of mobile safety cameras. Of the informative priors used, the DIC suggests the least appropriate prior distribution for the mean casualty rate is the lognormal, and the most appropriate is the Weibull. Indeed, there might be some practical justification for this: taking  $\hat{\gamma} = 2.494$  and  $\hat{\mu}_{47} = 4.809$  (as reported in Table 3.1 for site 47), gives a value for  $\Pr(m_i > 40)$  under the lognormal prior that is 7 times larger than that under the gamma prior, and 77 times larger than that under the Weibull prior! Even for a relatively high casualty frequency site like site 47, it would be extremely unusual to observe a casualty rate of more than 40, and so we might trust most the prior with smallest tail probabilities. An EB analysis would, of course, miss the nuances reported here, and miss altogether the supposed best-fitting model formulation using the Weibull prior.

### 3.5 Discussion

We have introduced some commonly used methods for the evaluation of road safety schemes. Typically, in road safety before/after studies, locations with an unusually high number of collisions in the pre-determined time-period are selected. However, this approach may result in selection bias issues. These locations were chosen due to the high number of collisions over a certain period, and the number of collisions in a subsequent period is likely to decrease. This is known as regression to the mean, where the abnormally high observation tends to return to the average over time. Practitioners therefore must account for this in their evaluations of how effective a safety scheme has been. Other confounders must also be accounted for to gain a ‘clean’ estimate of the treatment effect. We introduced the commonly-used empirical Bayes (EB) and fully Bayes (FB) methods and showed the advantages of a full Bayesian framework through acknowledging the variability of the regression parameters and hence realistically quantifying the uncertainty in the coefficients. We also presented the added flexibility that comes with the FB method through the ability to stray from a Poisson-gamma formulation. The sensitivity to change in prior distributions analysis showed that the conclusions drawn on the treatment effect’s effectiveness can alter vastly with the use of different priors. For the NSCP data analysis, we showed that the best-fitting model formulation is using a Weibull prior.

In the realm of road safety research, both EB and FB methods come with their

respective advantages and limitations. EB is computationally efficient and well-suited for large datasets, especially when prior information is limited. However, it's less robust and can be sensitive to the choice of estimated hyperparameters. On the other hand, FB offers robustness and the flexibility to incorporate complex models and prior beliefs. It also provides a fuller quantification of uncertainty through its posterior distributions. Yet, FB comes with the drawback of being computationally intensive, sensitive to prior choice, and potentially less scalable for very large datasets.

# Chapter 4

## Bespoke modelling formulations for atypical before/after studies

### 4.1 Introduction

EB or FB methods are not applicable to all before/after studies. If the sites are not chosen for their abnormally high collision counts, then we no longer need to account for RTM and thus, have reference data. In recent years, with the advancement of technology and video-processing techniques, interest in near-misses has peaked. In these studies we usually model times between road users reaching a point in space, hence standard techniques would not apply. In other disciplines, randomised trials are commonplace. In medicine, randomised trials aim to avoid selection bias thus ensuring that any differences in outcome can be solely attributed to the treatment. They help to ensure that any variations between treatment groups are due to chance which reduces the potential for bias and confounding. Hence, implementing safety schemes at sites that are randomly allocated could provide reliable estimates of the treatment effect in road safety evaluation studies. Naturally, the implementation of many road safety schemes, such as installing speed bumps or changes to road layouts, can be costly – requiring significant time and resources. If these safety interventions are then proven to be ineffective and they were implemented at sites that were already deemed ‘safe’, they may be classed as a waste of money and resources. However, there are cheap and easy to install safety intervention options, such as vehicle activated signs, which could be a viable option for a randomised trial. In these instances, we must curate a model to estimate any potential treatment effect.

### 4.1.1 Vehicle Activated Signs

Vehicle activated signs (VAS) are traffic signs which activate when traffic approaches, some may activate if a speed limit is being exceeded, or if a hazard is ahead, including warning of a sharp bend in the road. They are created to look similar to usual road signs, with clear messages flashing up when activated, some may show a happy face to traffic going below a speed limit and an angry face to those exceeding. VAS are a low-cost and low-maintenance option for a traffic calming measure which have been proven to be effective in reducing speeding and accidents (Jamson et al., 2010, Winnett and Wheeler, 2002). As a result of this, VAS could be a viable option for a randomised trial. They are cheap to implement and maintain, compared to speed cameras, and are not as ‘controversial’ from a public perception or a resource point-of-view. Trials of safety schemes in which sites to be treated are chosen randomly would also not fit into the ‘standard’ model formulation. In this case we wouldn’t require reference data to account for RTM.

To avoid selection bias, Xie et al. (2011) select random segments of a highway to apply the Highway Safety Manual (HSM, AASHTO, 2010) procedures to roadway segments in Oregon. Issues of selection bias are common place, Elvik (2004) studies the extent of site selection bias in a study on Norway’s national roads. Clearly the methods outlined in Chapter 3 attempt to account for this bias, however studies can be performed omitting this bias altogether.

### 4.1.2 Traffic conflict analyses

It is not always the case that data are collected as collision counts, recently attention has turned to looking at near-misses or traffic conflicts. Traffic conflicts are generally defined as a situation in which two road users approach each other in time and space to such an extent that they will collide if their trajectories remain unchanged. Traffic conflicts are more frequent than collisions, easy to observe and are of marginal social cost (Tarko, 2018). As such, traffic conflict and near-miss before/after studies can be conducted over shorter time periods. Furthermore, new technology has been adopted, such as automated video techniques (Saunier and Sayed, 2007), traffic simulation (Wang et al., 2018), and software (for e.g. TrafXSAFE, Transoft Solutions, 2020) to allow for detecting and tracking moving objects based on their trajecto-

ries; conflict data are thus easily extracted. We are then able to analyse these traffic conflicts and near-misses through post encroachment times (PETs; outlined in Section 1.3.3) which have become popular to define conflicts in relevant literature (Navarro et al., 2022, Paul and Ghosh, 2021). When working with PETs or time to collision (TTC) it is intuitive to model using extreme value theory (EVT). Small PETs or TTCs denote a dangerous situation. Zangenehpour et al. (2016) categorised PET values in terms of severity through the following classification:

- Very dangerous,  $PET \leq 1.5s$
- Dangerous,  $1.5s < PET \leq 3s$
- Mild interaction,  $3s < PET \leq 5s$
- Safe interaction,  $PET > 5s$

Hence, we look for extreme values in the data and in before/after studies, our safety scheme has had a positive effect when we have larger values for PET or TTC in the after period. A significant reduction in potential conflicts will likely lead to a significant reduction in actual collisions. There is the added benefit of treating potential road safety hotspots proactively. Standard road safety interventions are usually analysed in a reactive manner. Collisions are rare events and so prolonged observation periods are necessary, with much waiting for collisions to happen to evaluate a treatment using a standard before/after analysis. Working with traffic conflicts means not having to wait for collisions to happen.

## 4.2 Study of Vehicle Activated Signs in Tyne and Wear

As discussed in Section 1.2.2, the road safety team at Newcastle have been working closely with Gateshead Council since 2015 on various road safety-related projects, including an evaluation of their mobile safety camera scheme (see Section 3.4 for full details). In August 2019, the council's road safety steering group gave the green light to a completely randomised trial of their newly-purchased vehicle activated signs. The aim of the trial would be an evaluation of their VAS in terms of reducing collisions at potential road safety hotspots across Tyne and Wear and Northumberland. Compared to their well-established programme of deployment for mobile

road safety cameras, VAS were a relatively new form of road safety intervention for the council, and there was an eagerness to properly assess their effectiveness before a full rollout as part of their standard road safety practices. It was hoped that a randomised trial would provide the best form of evaluation here, helping to avoid any selection bias and thus the need to account for RTM at the data analysis stage. The following model formulation was developed just before discussions around site selection were held; the evaluation was due to take place during the 2020/2021 fiscal year, but the pandemic has since put the study on hold. Thus, throughout the rest of this section we use simulated data to illustrate the model we develop for assessing the effectiveness of Gateshead Council's VAS.

Gateshead council had planned to implement the signs at five localities, and collect monthly collision data in five months before and 5 months after implementation. Therefore, the data were generated to represent the number of collisions over five localities. This was based on five records before treatment and five after. Suppose the number of collisions,  $X_{ij}$ , follow a Poisson distribution,

$$X_{ij}|\lambda_{ij} \sim Po(\lambda_{ij}), \quad \lambda_{ij} > 0,$$

where  $i = 1, \dots, 5$  denotes the specific locality and  $j = 1, \dots, 10$  denotes the time period. A Poisson regression equation was chosen to describe the rate of collisions dependent on certain factors,

$$\log \lambda_{ij} = \alpha + \beta t_{ij} + (\gamma + \delta t_{ij})(l_i - \bar{l}), \quad (4.1)$$

where  $t_{ij} = 0$  before treatment and  $t_{ij} = 1$  after treatment. The total number of collisions at the five localities are recorded ( $l_i$ ) and compared to the mean over numerous localities ( $\bar{l}$ ) in an attempt to standardise the data, making it easier to compare across sites. For experimental purposes,

$$l_i - \bar{l} = \frac{1}{5}(-2, -1, 0, 1, 2).$$

We set,

$$\alpha = 2, \beta = -2, \delta = 2, \gamma = 1,$$

in the regression equation to calculate values of  $\lambda_{ij}$ . The data were simulated in R

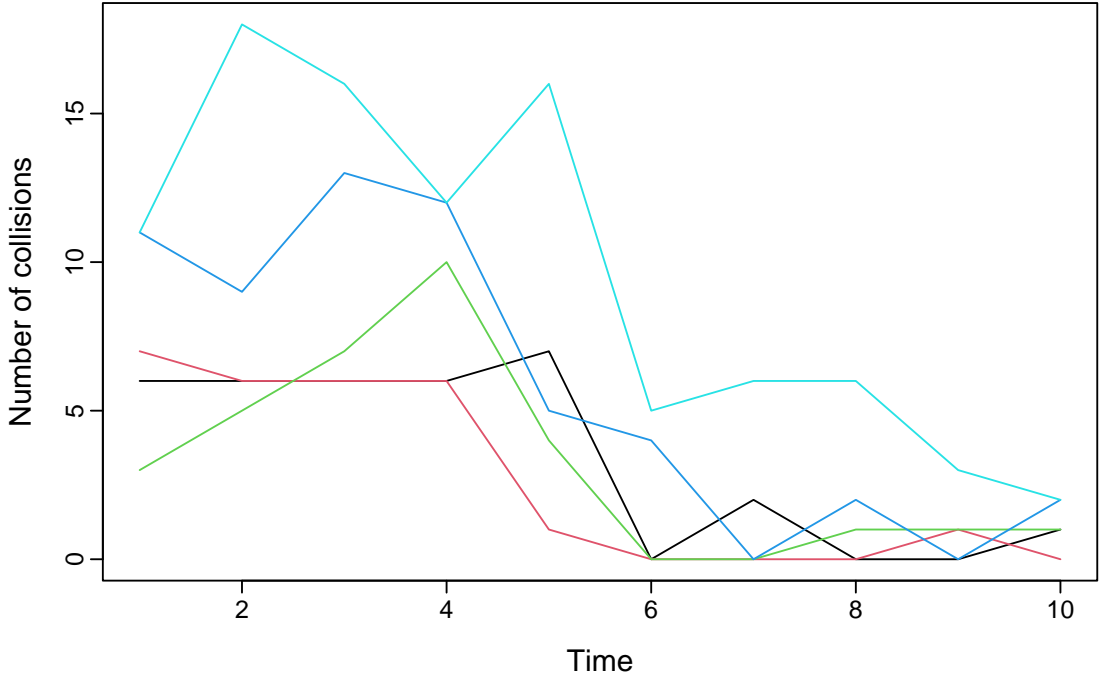


Figure 4.1: Simulated data over five localities for 5 observations before treatment and 5 after.

using a Poisson distribution (with mean  $\lambda_{ij}$ ) gaining 10 observations at each locality; five observations before treatment ( $j = 1, \dots, 5, t_{ij} = 0$ ) and five observations after treatment ( $j = 6, \dots, 10, t_{ij} = 1$ ), and are shown in Figure 4.1. Since the parameters could take any value, normal prior distributions were chosen for each parameter independently, with zero mean and large variance,

$$\alpha \sim N(0, 100), \beta \sim N(0, 100), \delta \sim N(0, 100), \gamma \sim N(0, 100),$$

such that the prior can be written as

$$\pi(\Theta) = \pi(\alpha)\pi(\beta)\pi(\delta)\pi(\gamma),$$

where  $\Theta$  denotes the parameter vector. As the posterior distribution  $\pi(\Theta|x)$  is intractable and full conditional distributions (FCDs) are not available in closed form, we draw realisations from  $\pi(\Theta|x)$  using a Metropolis-Hastings algorithm with individual normal random walk proposals on each parameter (see Section 2.4.2). The MCMC scheme is initialised at  $\alpha^{(0)} = 0, \beta^{(0)} = 0, \delta^{(0)} = 0, \gamma^{(0)} = 0$ . At each iteration  $k$ , proposals are drawn for each parameter from normal distributions such



	$\alpha$	$\beta$	$\delta$	$\gamma$
Mean	2.042	-2.185	2.415	1.347
95% Credible interval	(1.897, 2.183)	(-2.739, -1.688)	(0.833, 4.071)	(0.818, 1.840)

Table 4.1: Posterior mean and 95% credible intervals for the regression coefficients from 400k iterations and a thin of 400.

that, for example,

$$\alpha^* \sim N(\alpha^{(k-1)}, \epsilon_\alpha)$$

where  $\epsilon_\alpha = 0.3$  is the tuning parameter for the  $\alpha$  updates. For the remaining model parameters, we have  $\epsilon_\beta = 0.3$ ,  $\epsilon_\delta = 0.9$  and  $\epsilon_\gamma = 0.9$ . In the acceptance probabilities we use the log-likelihood of the Poisson distribution, such that

$$\ell(\lambda|\mathbf{x}) = \sum_{i=1}^5 \sum_{j=1}^{10} x_{ij} \log \lambda_{ij} - \sum_{i=1}^5 \sum_{j=1}^{10} \lambda_{ij} - \sum_{i=1}^5 \sum_{j=1}^{10} \log(x_{ij}!).$$

Therefore the acceptance probability for  $\alpha$  is

$$A_\alpha = \min \left\{ 0, \frac{\pi(\alpha^*)\ell(\lambda^*|\mathbf{x})q(\alpha^*|\alpha^{(k-1)})}{\pi(\alpha^{(k-1)})\ell(\lambda|\mathbf{x})q(\alpha^{(k-1)}|\alpha^*)} \right\}$$

where  $\alpha^{(k-1)}$  is the last accepted value of  $\alpha$ . The same methods apply for the remaining parameters.

The MCMC scheme is run for 400k iterations with a thin of 400 to gain sufficiently low autocorrelation between realisations. The acceptance rates for all parameters were approximately 0.21. The trace plots showed good convergence for all regression coefficients and have similar means and variances in different sections of the chains. The autocorrelation was sufficiently low after thinning and the histograms showed good estimates of marginal densities. Numerical summaries are shown in Table 4.1. Using the means of the marginal posterior distributions for each coefficient, obtained from the MCMC scheme, we can estimate the regression equation, Equation (4.1), to be

$$\log \lambda_{ij} = 2.042 - 2.185t_{ij} + (2.415 + 1.347t_{ij})(l_i - \bar{l}). \quad (4.2)$$

We can therefore see how well the model fits the data as the means of the marginal distributions are very close to the true parameter values used to simulate the data.

We noticed a high correlation between parameters  $\alpha$  and  $\beta$  ( $Cor(\alpha, \beta) = -0.91$ ) and  $\delta$  and  $\gamma$  ( $Cor(\delta, \gamma) = -0.93$ ). This suggests that we could benefit from using joint proposals combining  $\theta_1 = (\alpha, \beta)^T$  and  $\theta_2 = (\delta, \gamma)^T$  within the MCMC scheme. Hence, proposals come from multivariate normal distributions, where

$$\theta_1^* = \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix} \sim N_2 \left[ \begin{pmatrix} \alpha^{(j-1)} \\ \beta^{(j-1)} \end{pmatrix}, \begin{pmatrix} 0.17 & -0.16 \\ -0.16 & 0.17 \end{pmatrix} \right],$$

$$\theta_2^* = \begin{pmatrix} \delta^* \\ \gamma^* \end{pmatrix} \sim N_2 \left[ \begin{pmatrix} \delta^{(j-1)} \\ \gamma^{(j-1)} \end{pmatrix}, \begin{pmatrix} 2.0 & -1.88 \\ -1.88 & 2 \end{pmatrix} \right].$$

For proposals, the variance matrices for  $(\alpha, \beta)^T$  and  $(\delta, \gamma)^T$  were determined from tuning the acceptance probabilities. As discussed in Section 2.4.2, for high-dimensional target distributions formed of independent and identically distributed (iid) components, the acceptance rate optimising the efficiency of the process approaches 0.234 (23.4%) (Gelman et al., 1997), from this algorithm we have 0.256 and 0.233 for  $\theta_1$  and  $\theta_2$  respectively. From the independent proposals, the covariance matrix approximated through MCMC for the parameters can be used to generate an approximate optimal covariance matrix to use for joint normal random walks. These are given to be

$$\frac{1}{2}2.38^2 V_{\alpha, \beta} = \begin{pmatrix} 0.166 & -0.165 \\ -0.165 & 0.191 \end{pmatrix}, \quad \frac{1}{2}2.38^2 V_{\delta, \gamma} = \begin{pmatrix} 2.073 & -1.848 \\ -1.848 & 1.854 \end{pmatrix}, \quad (4.3)$$

where

$$V_{\alpha, \beta} = \begin{pmatrix} 0.059 & -0.058 \\ -0.058 & 0.067 \end{pmatrix} \quad \text{and} \quad V_{\delta, \gamma} = \begin{pmatrix} 0.732 & -0.653 \\ -0.653 & 0.655 \end{pmatrix}.$$

Updating the MCMC to use these covariance matrices gives acceptance probabilities of 0.231 and 0.242 for  $\theta_1$  and  $\theta_2$  respectively. Altering the Poisson regression to joint proposals to account for high correlation between  $\alpha$  and  $\beta$ , and  $\gamma$  and  $\delta$ , made the MCMC more efficient reducing the amount of thin needed from 400 to 100, hence a shorter run is plausible to estimate the marginals. Figures 4.2 and 4.3 show the outcome of the MCMC run for 100k iterations. The trace plots show good convergence and the histograms show good estimations to marginal densities.

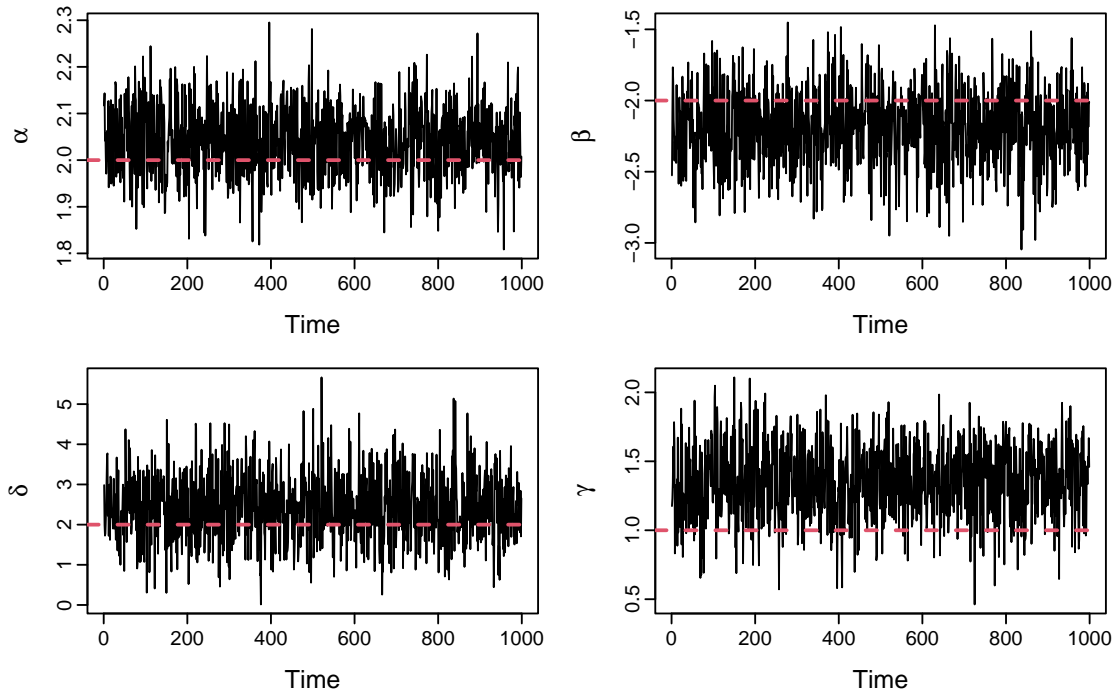


Figure 4.2: Trace plots for  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  from 100k iterations after burn-in and thin of 100 with their true values overlaid by a red dotted line.

The horizontal dotted lines on the trace plots denote the true values for the model parameters – those used to simulate the data, and hence the model has estimated them well.

### 4.3 Study of a leading pedestrian interval intervention in Vancouver

With the safety of pedestrians in mind, safety treatments at signalised intersections have been investigated in Vancouver, Canada as outlined in Section 1.3.3. The LPI was implemented at multiple sites by adjusting the signal phasing and pedestrian interval to provide a walk display of several seconds before the adjacent vehicle green display, making this an efficient and low-cost safety measure. The LPI intervention, and associated traffic conflict data on collision near-misses, justifies the use of the extremal types theorem (EVT, see Section 2.2). Small PET values imply a near-miss, a value close to zero implying a dangerous situation. Our aim is to model extremely

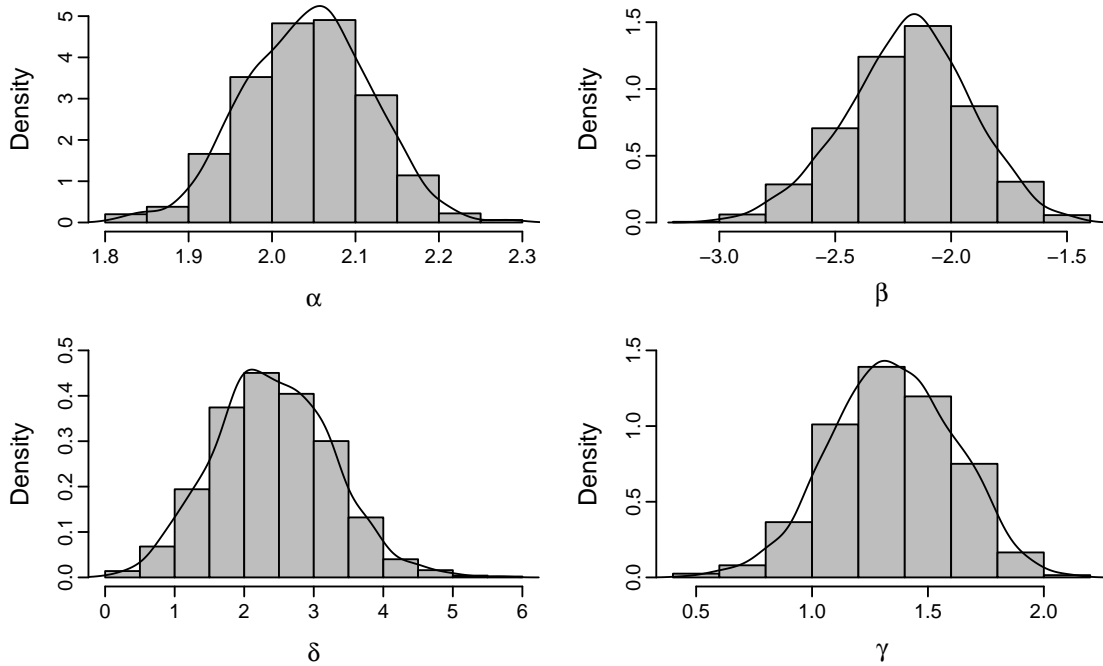


Figure 4.3: Histograms of the posterior densities of  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  from 100k iterations after burn-in and thin of 100 with their density overlaid.

small PET values and investigate differences between extremes from periods before and after the LPI treatment was introduced. Typical applications of EVT occur in the environmental sciences to model, for example, extreme precipitation events or extreme wind speeds (see, for example, [Fawcett and Walshaw, 2006](#), [Katz, 1999](#)); for interested readers, we point to the classical reference of [Gumbel \(1958\)](#) and the tutorial-style text of [Coles \(2001\)](#). Here, we focus on the lower tail of the data where small time values indicate a dangerous situation: a near-miss or collision between a pedestrian and left-turning vehicle.

EVT has become popular in traffic conflict analyses. [Zheng and Sayed \(2019a\)](#) use a peaks over threshold approach including covariates in the scale parameter for crash estimation; furthermore, they also use EVT on block maxima from a traffic conflict before/after study ([Zheng and Sayed, 2019b](#)). [Wang et al. \(2019\)](#) use bivariate EVT to predict annual crash frequencies at intersections, [Zheng et al. \(2018\)](#) use bivariate threshold excess model methods to estimate crashes relating to merging events on freeway entrance merging areas. [Fu et al. \(2021\)](#) use hierarchical EVT modelling on traffic conflict extremes for crash estimation. [Guo et al. \(2020\)](#) anal-

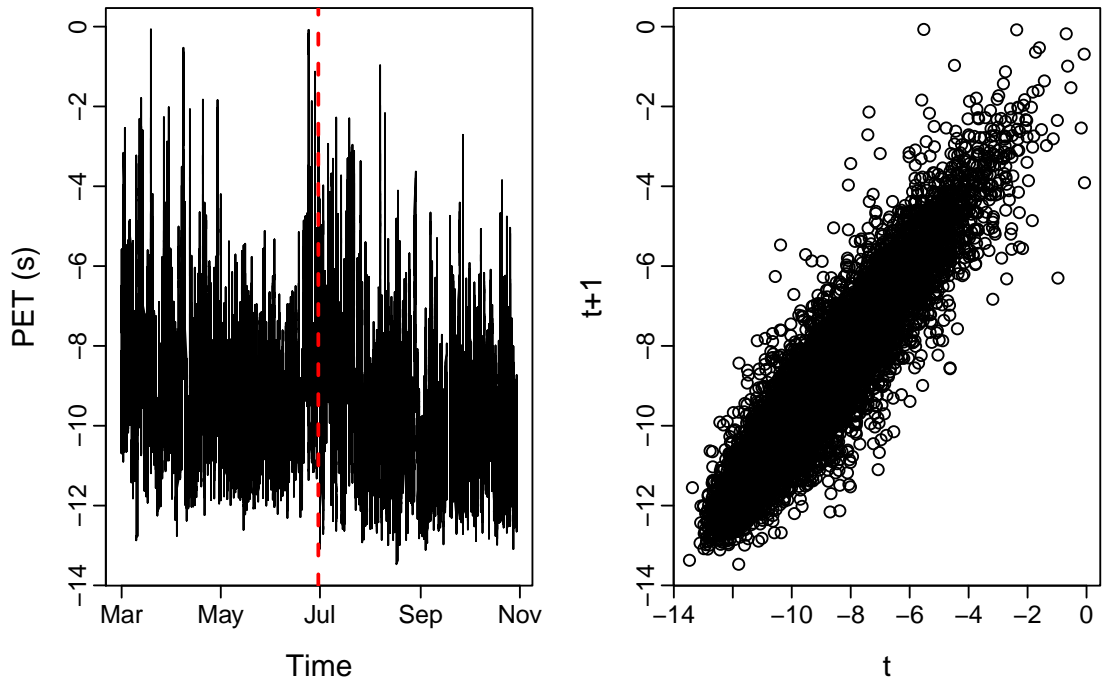


Figure 4.4: Left: Time series plot of negated PET values (seconds) before and after intervention at intersection 1, the vertical red line showing when the after period begins. Right: Temporal dependence of observations at neighbouring time points for intersection 1 ( $t$  on  $x$ -axis against  $t + 1$  on  $y$ -axis).

used the effectiveness of LPI treatment at two signalised intersections in Vancouver, Canada by modelling the scale and shape parameters of the GPD as a function of a treatment indicator, a period indicator (before/after), and an interaction of the two variables. They use a peaks over threshold approach to remove dependence between consecutive extremes. In this project we use EVT in an attempt to capture treatment effects from the LPI intervention described in Section 1.3.3. Our method aims to maximise data usage by performing a threshold-based analysis (as opposed to a block maxima analysis; see, for example, [Coles \(2001\)](#)). We include an example of the use of a block maxima approach before proposing the use of a threshold-based approach on the premise of including more data in the analysis than the block maxima approach, which can be wasteful through discarding all but the most extreme observations in each block. It is hoped that the inclusion of more data points will lend greater precision to the analysis, including reduced uncertainty in our estimated treatment effect.

In order to use standard methods from the EVT toolkit (designed for analysing maxima), we negate our series of PET values at each location thus switching the focus from very small values to very large values to identify dangerous situations in our series. Figure 4.4 shows a time series plot of these negated PET values, and a plot of observations at neighbouring time points, for one of the intersections at which the LPI was applied. Note the apparent systematic decrease in PET values after the LPI implementation; note also the clear temporal dependence between consecutive data values, persisting into the extremes of the process.

### 4.3.1 Including covariates

When the data admits non-stationarity – for example trend, or a dependence on covariates – we can attempt to incorporate this non-stationarity through linear modelling of the GEV or GPD parameters. Generally, we can write the extreme value parameters in the form  $h(X^T\beta)$ , where  $h$  is a specified function,  $\beta$  is a vector of parameters and  $X$  is the model vector. Recall from Section 2.2 that the GPD,  $\mathcal{H}(\tilde{\sigma}, \xi)$ , arises from the GEV,  $\mathcal{G}(\mu, \sigma, \xi)$ , where the GPD scale parameter is a function of the GEV location and shape parameters. Thus, attempting to model any trend in our threshold excesses is usually done through linear modelling of the location parameter when fitting a GEV, or the scale parameter  $\tilde{\sigma}$  when fitting a GPD. The PET data we are investigating have before/after time implications at each of the fifteen intersections. We attempt to capture the treatment effect through the following parameterisations. We set the location parameter of the GEV as

$$\mu_t = \beta_0 + \beta_1 t,$$

and the scale parameter in the GPD as,

$$\tilde{\sigma}_t = \exp(\beta_0 + \beta_1 t), \tag{4.4}$$

where  $t = 0$  in the before period and  $t = 1$  in the after period. We take the exponential to respect the positivity of  $\tilde{\sigma}$ . Hence, at each intersection we could use the slope parameter  $\beta_1$  as a proxy for our LPI treatment effect; an estimate of  $\beta_1$  that is deemed significantly different from zero might be indicative of a treatment effect at an intersection.

### 4.3.2 Classifying extremes through block maxima

The first step in any application of the generalised extreme value distribution is to fit Equation (2.3) to a series of block maxima,  $M_{n,i}$ . This requires choosing a block length  $n$ , and then discarding all but the maximum value in each block. However, the optimal block length isn't always apparent. Choosing an appropriate block length is crucial. Excessively short blocks may invalidate the limiting arguments that support the GEV distribution and result in maxima that are too close together to assume independence. Additionally, selecting a block length that is too large relative to the sample size  $n$  can result in insufficient data being available for analysis. Conventionally, it is common to work with convenient block lengths – e.g. annual maxima in environmental applications, however our application differs. As the data has a few reasonable block sizes available, we look at hourly, daily and weekly maxima, in addition to this we choose the maxima of a block size of approximately  $\sqrt{N}$ , where  $N$  is the total number of observations per site. Firstly, hourly maxima data showed seasonality and temporal dependence, and thus, this block length was not chosen. When using weekly maxima, we were only left with 32 maxima to model with, and although we can reasonably assume independence between maxima, we are discarding over 17k observations. Figure 4.5 shows that we can reasonably assume independence between both daily and  $\sqrt{N}$  maxima, and no seasonality was present between blocks. As a result, the limiting arguments are valid for all three block sizes, and to maximise the use of available data, we opt to model daily maxima, which provides us with 244 maxima per site.

The model is formulated as follows, we have daily maxima for each site coming from a GEV distribution with location,  $\mu_t = \beta_0 + \beta_1 t$ , scale  $\sigma > 0$  and shape  $\xi$  parameters. Deviance statistics were used to check model adequacy in terms of the additional linear trend in  $\mu$  through time. Allowing for the linear trend in time was shown to improve model fit for all treated sites.

As there are no conjugate priors for the GEV distribution, we use MCMC to draw realisations from the posterior distribution. We perform inference within a Bayesian framework, our parameter vector at each intersection can be written as,

$$\Theta = (\beta_0, \beta_1, \eta, \xi)^T,$$

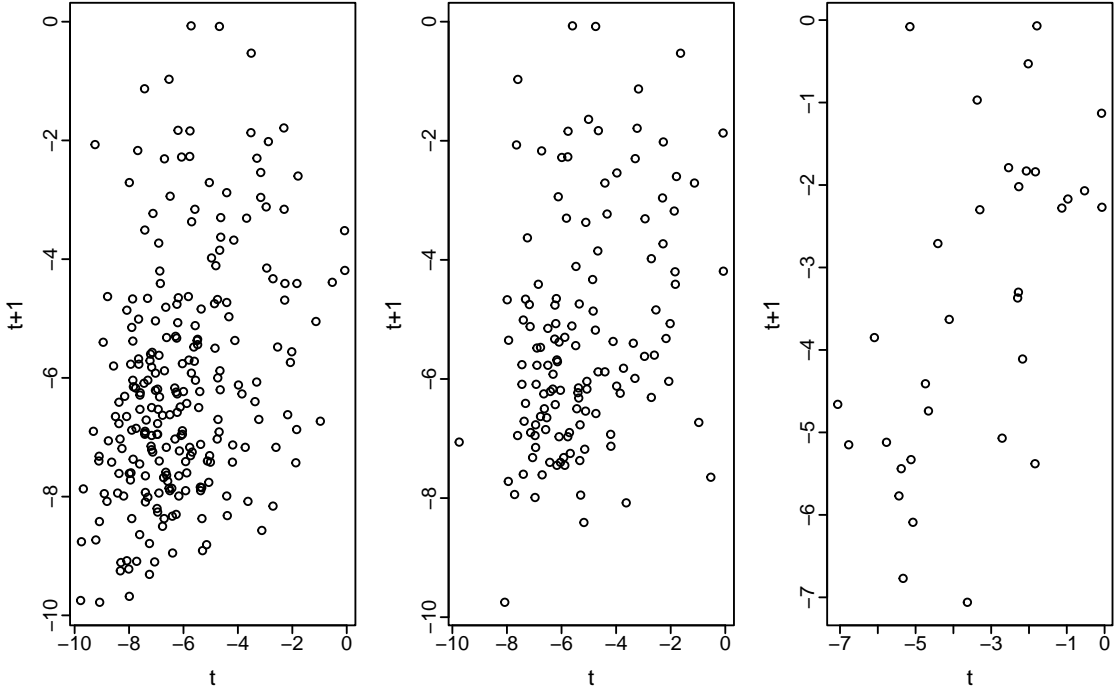


Figure 4.5: Temporal dependence of observations at neighbouring time points for intersection 1 ( $t$  on  $x$ -axis against  $t+1$  on  $y$ -axis) for blocks of daily,  $\sqrt{N}$  and weekly maxima (left to right).

where,  $\eta = \log \sigma$  to respect the positivity of  $\sigma$ . The parameters are assumed independent and hence the prior can be written as,

$$\pi(\Theta) = \pi(\beta_0)\pi(\beta_1)\pi(\eta)\pi(\xi),$$

where

$$\beta_0 \sim N(0, 10), \beta_1 \sim N(0, 10), \eta \sim N(0, 10), \xi \sim N(0, 10).$$

Our MCMC scheme evaluates the posterior through Bayes theorem, such that,

$$\pi(\Theta|x) \propto L(\Theta|x)\pi(\Theta).$$

The same methods apply for all other sites.

Bayes theorem requires the likelihood to evaluate the target posterior distribution. To avoid numerical instability and improve the accuracy of the MCMC algorithm, we will use the log-likelihood function, outlined in Equations (2.8) and



(2.9), in the acceptance probabilities. For our model, we will simply replace  $\mu$  by  $\mu_t = \beta_0 + \beta_1 t$ , where  $t$  is a vector with  $t_i = 0$  for  $i = 1, \dots, 122$  (the before period) and  $t_i = 1$  for  $i = 123, \dots, 244$  (the after period). The acceptance probability in the MH scheme then becomes,

$$\alpha_{\Theta_j} = \min \{0, A\},$$

where

$$A = \log \pi(\Theta_j^*) + \ell(\Theta_j^*|x) + \log q(\Theta_j^*|\Theta_j^{(i-1)}) - \log \pi(\Theta_j) - \ell(\Theta_j|x) - \log q(\Theta_j^{(i-1)}|\Theta_j^*).$$

The MCMC scheme with MH was run for 100k iterations with a thin of 100 and the first 20k iterations discarded as burn-in. This is run for each site individually, to gain estimates to the marginal posterior distributions for the elements of  $\Theta_j$ , where  $j$  denotes sites  $1, \dots, 15$ .

From the marginal posterior means and CIs of  $\beta_1$  in Figure 4.6, we can clearly identify the non-treated sites as the CIs include 0. Although modelling block maxima is a common approach, it does come with issues. Some of which include finding the correct block size, if the block length is too small the limiting arguments will not hold (the GEV is a limiting result, which holds approximately for large  $n$ ). Alternatively, if the block length is too large we will not have enough maxima to model. There is also evidence of possible sensitivity of GEV parameter estimates to block length (Özari et al., 2019). Additionally, it is extremely wasteful of data; we discard all but the block maxima when other observations in the block may also be extreme, just not as extreme as the maxima.

### 4.3.3 Classifying extremes through threshold excesses

#### 4.3.3.1 Temporal dependence

Our aim now is to maximise precision by including as much data in the analysis as possible; hence, we will proceed with a threshold-based approach to analysis. As discussed in Section 2.2.2, unlike the case of modelling block maxima with the GEV, powering  $F^n$  by  $\theta$ , as described by Equation (2.2), does not lead to another extreme value distribution whose parameters have absorbed the extremal index; thus, careful consideration of extremal dependence is required. In this Section, we describe two methods for handling temporal dependence present between consec-

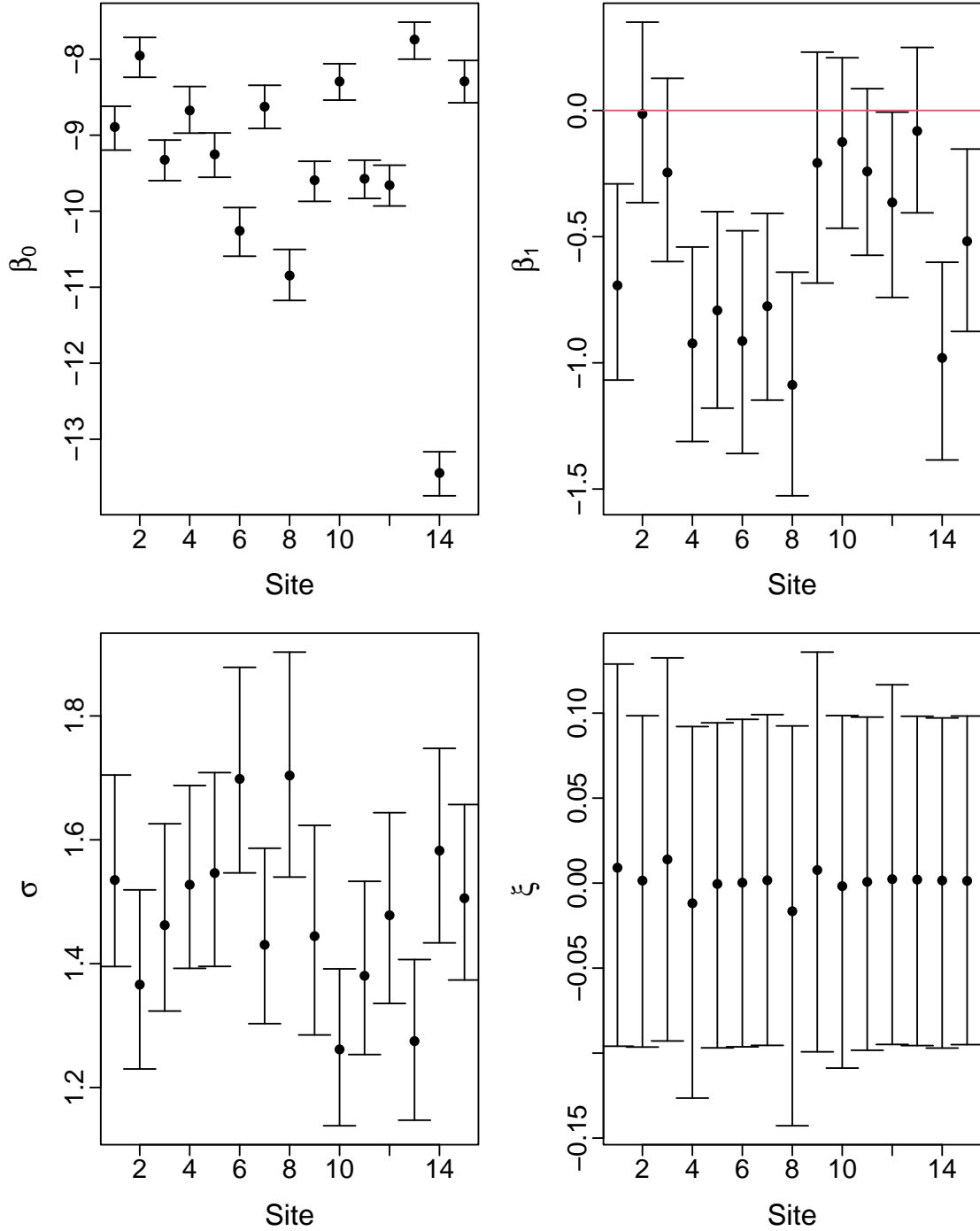


Figure 4.6: Marginal posterior means and 95% credible intervals for the GEV parameters  $(\beta_0, \beta_1, \sigma, \xi)$ .

utive threshold excesses: a declustering approach, leading to the commonly-used ‘peaks over threshold’ analysis of a filtered subset of extremes (Davison and Smith, 1990), and an approach that explicitly models the transition from one extreme to the other through a bivariate extreme value model. As the right-hand-side plot in Figure 4.4 reveals, even above a high threshold there appears to be dependence between consecutive PET values. At busier times of the day – perhaps early morning or late-afternoon – we might expect more pedestrians to be using the crosswalks at each of our intersections, and a greater number of vehicles turning into the intersections, perhaps resulting in a greater number of near-misses (with an associated clustering of small PET values) at these times. Clustering of extremes is common in many other applications of the threshold approach to extreme value modelling – for example, dependence between consecutive temperature or wind speed extremes, and serial correlation in extremes obtained from financial time series. Ignoring such dependence will likely lead to under-estimated uncertainty measures (for example, confidence intervals that are unrealistically narrow; see, Barao and Tawn, 1999, Shi et al., 1992).

A common method to handle temporal dependence is through declustering. The aim of this approach for handling extremal dependence is to extract a series of independent threshold excesses, justifying the use of  $\theta \approx 1$  in Equation (2.2). An auxiliary ‘declustering parameter’, say  $\kappa$ , is chosen and a cluster of threshold excesses is then deemed to have terminated once at least  $\kappa$  consecutive observations fall sub-threshold. This is repeated over the entire series to identify clusters of excesses. Then, the maximum (or ‘peak’) observation from each cluster is extracted, and the GPD fitted to the set of (hopefully independent) cluster peak excesses.

This approach is often referred to as the ‘peaks over threshold’ approach (POT, Davison and Smith, 1990) and is the most commonly-used approach for dealing with clustered extremes. Although this approach is quite easy to implement, there are issues surrounding the choice of  $\kappa$ ; if  $\kappa$  is too small, the cluster peaks will not be far enough apart to safely assume independence. If  $\kappa$  is too large, there will be too few cluster exceedances on which to form our inference (and of course, the approach will be wasteful of data). Furthermore, parameter estimates have been shown to be sensitive to the choice of  $\kappa$  (Fawcett and Walshaw, 2012). Given how commonplace POT analyses for threshold excesses have become, we will include results based on

this approach as a baseline for comparing our results using a first-order extreme value Markov chain for modelling dependence.

To avoid declustering, we can account for dependence between consecutive extremes by assuming a first order Markov structure. For example, we can assume the following joint density,

$$\begin{aligned} g(\mathbf{x}|\theta) &= g(x_1) \times g(x_2|x_1) \times \dots \times g(x_n|x_{n-1}), \\ &= g(x_1) \times \frac{g(x_1, x_2)}{g(x_1)} \times \dots \times \frac{g(x_{n-1}, x_n)}{g(x_{n-1})}, \\ &= \frac{\prod_{i=2}^n g(x_{i-1}, x_i)}{\prod_{i=2}^{n-1} g(x_i)}, \end{aligned} \tag{4.5}$$

and log-likelihood,

$$\log g(\mathbf{x}|\theta) = \sum_{i=2}^n \log g(x_{i-1}, x_i) - \sum_{i=2}^{n-1} \log g(x_i),$$

for our series of (negated) PET values at each intersection, where  $\theta$  is a generic parameter vector. In a threshold excess context, univariate contributions to the denominator in Equation (4.5) are given through the GPD (on differentiation of Equation (2.5)). We therefore allow consecutive extremes to follow a bivariate extreme value distribution with some dependence parameter.

### Limiting distributions for bivariate extremes

We let  $\mathbf{M}_n^* = (M_{x,n}^*, M_{y,n}^*)$  be the normalised maxima i.e.  $M_{x,n}^* = \max\{X_i\}/n$  for  $i = 1, \dots, n$ , where the  $(X_i, Y_i)$  are i.i.d. with standard Fréchet marginal distributions. Then if

$$Pr(M_{x,n}^*, M_{y,n}^*) \rightarrow G(x, y),$$

where  $G$  is non-degenerate, then  $G$  has the form

$$G(x, y) = \exp \{-V(x, y)\}; \quad x > u_x, y > u_y. \tag{4.6}$$

Appealing to bivariate EVT, transformation from GPD to unit Fréchet margins (see, for example, [Coles, 2001](#)) provides a range of models to use for contributions to

the numerator in Equation (4.5), the most commonly-used being the logistic family with,

$$V(x, y) = (x^{-1/\alpha} + y^{-1/\alpha})^\alpha, \quad (4.7)$$

in Equation (4.6). Here,  $\alpha \in (0, 1)$  controls the extent of extremal dependence in the process, with independence giving  $\alpha = 1$  and  $\alpha \rightarrow 0$  corresponding to increasing levels of extremal dependence. The logistic model is symmetric meaning the variables are exchangeable. See [Coles \(2001\)](#) for other choices for  $G$  and a more detailed discussion of bivariate EVT more generally. Where direct evaluation of Equation (2.2) is necessary – for example, when obtaining quantiles from the fitted distribution (often used as estimates of “return levels” in applications of EVT) – [Fawcett and Walshaw \(2012\)](#) provide an approximation to the extremal index  $\theta$  based on the estimated logistic dependence parameter  $\alpha$ .

We define our bivariate extremes as those observations which exceed a threshold in one or other margin. The distribution function for the exceedances of a threshold  $u_x$  by a variable  $X$ , conditional on  $X > u_x$  for large enough  $u_x$  is,

$$H(x) = 1 - \lambda_{u_x} \left\{ 1 + \frac{\xi_x(x - u_x)}{\sigma_x} \right\}^{-1/\xi_x},$$

defined on  $\{x - u_x : x - u_x > 0 \text{ and } (1 + \xi_x(x - u_x)/\sigma_x > 0)\}$ , where  $\xi_x \neq 0, \sigma_x > 0$ , and  $\lambda_{u_x} = Pr(X > u_x)$  is the empirical threshold exceedance rate. If we suppose  $X$  and  $Y$  follow a GPD with parameters  $(\sigma_x, \xi_x)$  and  $(\sigma_y, \xi_y)$  and thresholds  $u_x, u_y$  respectively, we obtain our unit Fréchet margins with the transformations

$$\begin{aligned} \tilde{x} &= - \left( \log \left\{ 1 - \lambda_{u_x} \left[ 1 + \frac{\xi_x(x - u_x)}{\sigma_x} \right]^{-1/\xi_x} \right\} \right)^{-1}, \\ \tilde{y} &= - \left( \log \left\{ 1 - \lambda_{u_y} \left[ 1 + \frac{\xi_y(y - u_y)}{\sigma_y} \right]^{-1/\xi_y} \right\} \right)^{-1}. \end{aligned}$$

For  $\xi \rightarrow 0$ , our transformations are,

$$\begin{aligned} \tilde{x} &= - \left( \log \left\{ 1 - \lambda_{u_x} \left[ \exp \left( -\frac{x - u_x}{\sigma_x} \right) \right] \right\} \right)^{-1}, \\ \tilde{y} &= - \left( \log \left\{ 1 - \lambda_{u_y} \left[ \exp \left( -\frac{y - u_y}{\sigma_y} \right) \right] \right\} \right)^{-1}. \end{aligned}$$

Differentiation of Equation (4.7), with careful censoring when one or both of  $(x, y)$  lie sub-threshold, gives pairwise contributions to the numerator in Equation (4.5). We obtain contributions to the likelihood function for a pair of observations through

$$g(x, y; \alpha) = \begin{cases} \left. \frac{\partial^2 G}{\partial x \partial y} \right|_{(x, y)}, & \text{if } x > u_x, y > u_y, \\ \left. \frac{\partial G}{\partial x} \right|_{(x, u_y)}, & \text{if } x > u_x, y < u_y, \\ \left. \frac{\partial G}{\partial y} \right|_{(u_x, y)}, & \text{if } x < u_x, y > u_y, \\ G(u_x, u_y), & \text{if } x < u_x, y < u_y. \end{cases}$$

Then, we have

$$L(\alpha|x, y) = \prod_{i=1}^n g(x_i, y_i).$$

For the logistic model,  $G(\tilde{x}, \tilde{y})$  is expressed in Equation (4.7). Therefore, for  $\xi \rightarrow 0$  our likelihood contribution for the region where  $x > u_x$  and  $y > u_y$  can be calculated by the following

$$\frac{\partial G}{\partial x} = G \times -\alpha (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^{\alpha-1} \times -\frac{1}{\alpha} \tilde{x}^{-1-\frac{1}{\alpha}} \times \frac{\partial \tilde{x}}{\partial x},$$

where

$$\frac{\partial \tilde{x}}{\partial x} = \tilde{x}^2 \left[ 1 - \lambda_{u_x} \exp\left(-\frac{x - u_x}{\sigma_x}\right) \right]^{-1} \times \frac{\lambda_{u_x}}{\sigma_x} \exp\left(-\frac{x - u_x}{\sigma_x}\right).$$

Then, with  $k = \tilde{x}^{-(1+\frac{1}{\alpha})} \frac{\partial \tilde{x}}{\partial x}$ ,

$$\frac{\partial^2 G}{\partial y \partial x} = k \frac{\partial G}{\partial y} (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^{\alpha-1} + k G \frac{\partial}{\partial y} (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^{\alpha-1},$$

giving

$$\frac{\partial^2 G}{\partial y \partial x} = G \tilde{x}^{-\frac{1}{\alpha}} \tilde{y}^{-\frac{1}{\alpha}} \frac{\partial \tilde{x}}{\partial x} \frac{\partial \tilde{y}}{\partial y} (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^{\alpha-2} \left[ (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^\alpha + \left(\frac{1}{\alpha} - 1\right) \right].$$

Furthermore, in regions below thresholds  $u_x$  and  $u_y$ ,

$$\begin{aligned} \tilde{x} &= -(\log(1 - \lambda_{u_x}))^{-1}, & x < u_x, \\ \tilde{y} &= -(\log(1 - \lambda_{u_y}))^{-1}, & y < u_y. \end{aligned}$$

as  $Pr(X < u_x) = 1 - \lambda_{u_x}$ .

In our application, we invoke bivariate extreme value theory in a serial context, in which  $X$  and  $Y$  are replaced with  $X_{i-1}$  and  $X_i$  respectively, to support contributions to the numerator in Equation (4.5).

### 4.3.3.2 Threshold selection

The threshold stability property of the GPD means that if it is a valid model for excesses over some threshold  $u_0$ , then it is valid for excesses over all thresholds  $u > u_0$ . Furthermore, for all  $u > u_0$ ,  $E[X - u | X > u]$  is a linear function of  $u$ . In practice, this expectation can be estimated empirically as the sample mean of the excesses over  $u$ . This leads to the mean residual life plot (MRL plot; see, for example, [Coles, 2001](#)): a graphical procedure for identifying a suitably high threshold for modelling extremes via the GPD in which mean excesses over  $u$  are plotted against a range of values for  $u$ , and the optimal threshold is chosen at the lowest point at which we observe linearity in the plot. The process of choosing a suitable threshold should not be taken lightly as the scale parameter,  $\sigma$ , is threshold dependent where  $\sigma^* = \sigma + \xi u$ .

For all intersections, we class extremes as those values that exceed thresholds which were checked against MRL plots to ensure linearity is observed at this value. As an example, the MRL plot for negated PET values at intersections 1 and 7 are shown in Figure 4.7, both MRL plots are mostly linear between  $(-7, -5)$ , ergo we choose a threshold of  $u_1 = -5.6930$  for intersection 1 and  $u_7 = -5.53$  for intersection 7, these values also represent the 95% quantiles at these sites.

### 4.3.3.3 Prior specification

The Canada data we are investigating has before and after implementation time periods. As outlined in Section 4.3.1, we wish to include a time covariate to see if there is a significant difference in the scale parameter,  $\sigma_t$ , from before to after the delay in releasing cars was introduced. Here, we will be most interested in the sign and magnitude of  $\beta_1$  as this will indicate any treatment effect. Now,

$$\eta_t = \log \sigma_t = \beta_0 + \beta_1 t.$$

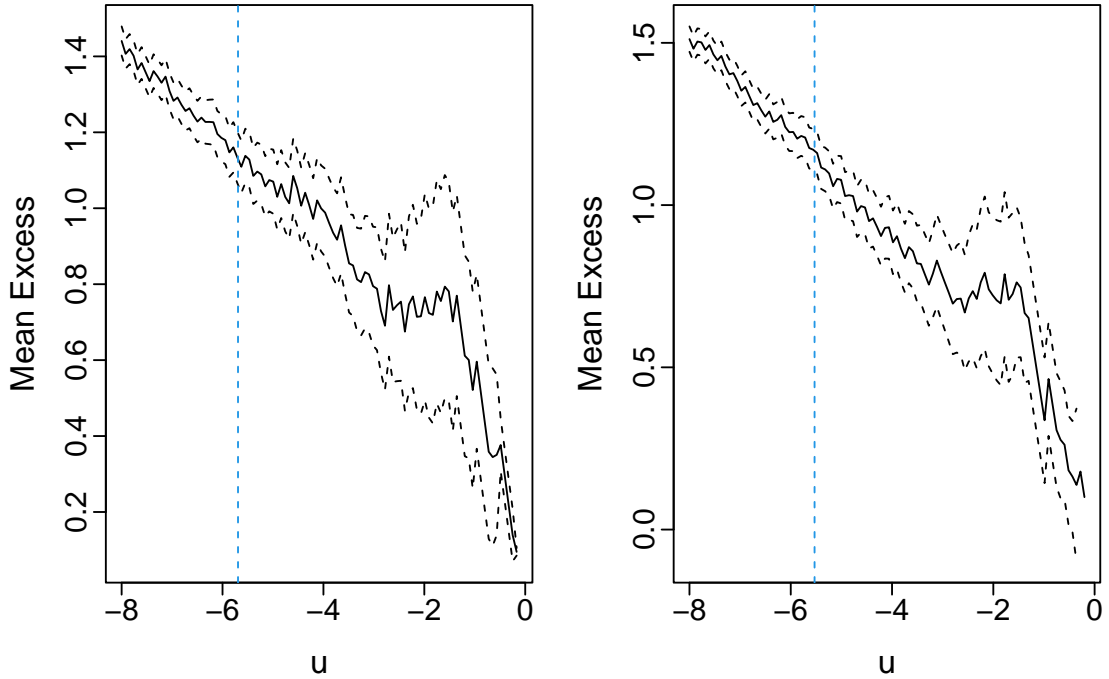


Figure 4.7: Mean residual life plot for sites 1 (left) and 7 (right). The solid line shows the empirical mean excess of values exceeding  $u$ ; the outer dashed lines show the 95% confidence band around this mean excess. The chosen threshold  $u$  is overlaid in blue,  $u_1 = -5.693$  and  $u_7 = -5.53$ .

As in the block maxima example, we are interested in finding the posterior distribution, however an analytic form for the posterior is not available. We therefore, use MCMC to estimate the posterior here. Our MCMC scheme evaluates the posterior through Bayes theorem, such that:

$$\pi(\Theta|x) \propto L(\Theta|x)\pi(\Theta),$$

where  $\Theta$  denotes all model parameters. We work with the reparameterised GPD scale parameter as in Equation (4.4).

In the absence of any expert prior information regarding our model parameters, we adopt fairly uninformative, independent prior distributions. Recall that, for each intersection, we adopt a bivariate threshold excess model for negated PET values exceeding a threshold identified by the 95% quantile and validated through use of an MRL plot. The margins are assumed GPD with a linear model enabling the scale parameter to vary between before and after periods via Equation (4.4); for the



dependence between successive threshold excesses we adopt a logistic model with a parameter quantifying the degree of serial correlation present. Thus, our parameter vector at each intersection can be written as:

$$\Theta = (\beta_0, \beta_1, \xi, \alpha)^T,$$

for which we set the following prior:

$$\pi(\Theta) = \pi(\beta_0)\pi(\beta_1)\pi(\xi)\pi(\alpha),$$

and where

$$\beta_0 \sim N(0, 10), \quad \beta_1 \sim N(0, 10), \quad \xi \sim N(0, 100) \quad \text{and} \quad \alpha \sim U(0, 1).$$

We set initial values for all parameters to their prior means, using a simple Metropolis update to give successive draws

$$(\beta_0^{(j)}, \beta_1^{(j)}, \xi^{(j)}, \alpha^{(j)}), \quad j = 1, \dots, 10^5,$$

after thinning to every tenth iteration and removing the first 2k iterations as burn-in. Within the MCMC scheme, proposals were drawn using normal random walks. We have that  $q(\Theta^*|\Theta) = N(\Theta^*; \Theta, \Sigma)$  where the innovation matrix  $\Sigma = \gamma \widehat{Var}(\Theta|x)$ , with  $\widehat{Var}(\Theta|x)$  obtained from a pilot run and  $\gamma$  is chosen to give an acceptance rate of around 25% (see Section 2.4.2).

#### 4.3.3.4 Results

Figure 4.8 shows the posterior means and 95% credible intervals for  $\beta_1$  over all sites, from analyses that (i) ignore dependence, (ii) filter out dependence through declustering, and (iii) explicitly model the dependence via our first-order extreme value Markov chain model. The treated sites are denoted with a ‘T’ on the  $x$ -axis. In our model,  $\beta_1$  captures the treatment effect as the slope term in our linear predictor for the GPD scale parameter. When we ignore dependence we use all threshold excesses and hence our credible intervals are relatively narrow owing to the maximal use of extreme data. However, we are violating the assumption that consecutive threshold excesses are independent; as such, these credible intervals are likely to be unrealistically narrow as there is less information in the data than the likelihood

requires. Declustering (here with  $\kappa = 10$ ) removes this dependence, but at a cost: reduced datasets, with site 1 (for example) now having just 107 threshold excesses post-declustering (from an original 17,467 raw excesses). We have done nothing to ‘optimise’ the declustering interval  $\kappa$  here, and it could be that our choice of  $\kappa$  is unnecessarily large resulting in a procedure that is wasteful of data (giving relatively wide credible intervals). However, even if we were to investigate an optimal choice of declustering interval  $\kappa$ , our extreme value Markov chain model for explicitly modelling dependence is probably a superior approach here, as it maximises data usage whilst also making some effort to capture the dependence in the series of threshold excesses.

As  $\beta_1$  corresponds to the treatment effect, a value of  $\beta_1 < 0$  shows a successful treatment as this indicates larger (negated) PET values in the after period (in other words, larger values on the raw PET scale, meaning a move away from a near-miss/actual collision). Of course, there is uncertainty in our inference, so we look for 95% credible intervals that are wholly below 0 to identify a successful treatment effect. In the majority of the treated sites  $\beta_1 = 0$  lies outside the range of the 95% credible intervals, hence we conclude these sites have seen an improvement post-treatment, with increased PETs.

Numerical summaries of the marginal posterior distributions for each parameter via their mean and 95% credible intervals for the three analyses – ignoring dependence, declustering and the logistic model are displayed in Table 4.2. For all sites and analyses,  $\xi$  is negative, meaning the GPD has a ‘light tail’. This implies that the distribution has a low probability of producing extreme events. As seen in Figure 4.8 for the posterior summaries of  $\beta_1$ , the posterior CIs for parameters  $\beta_0$  and  $\xi$  when using declustering are much wider than those from the logistic model. The posterior mean estimates for  $\xi$  are more negative meaning we would expect an even lower probability of extreme events. Hence, including all of the data has hugely improved the parameter estimates in terms of uncertainty.

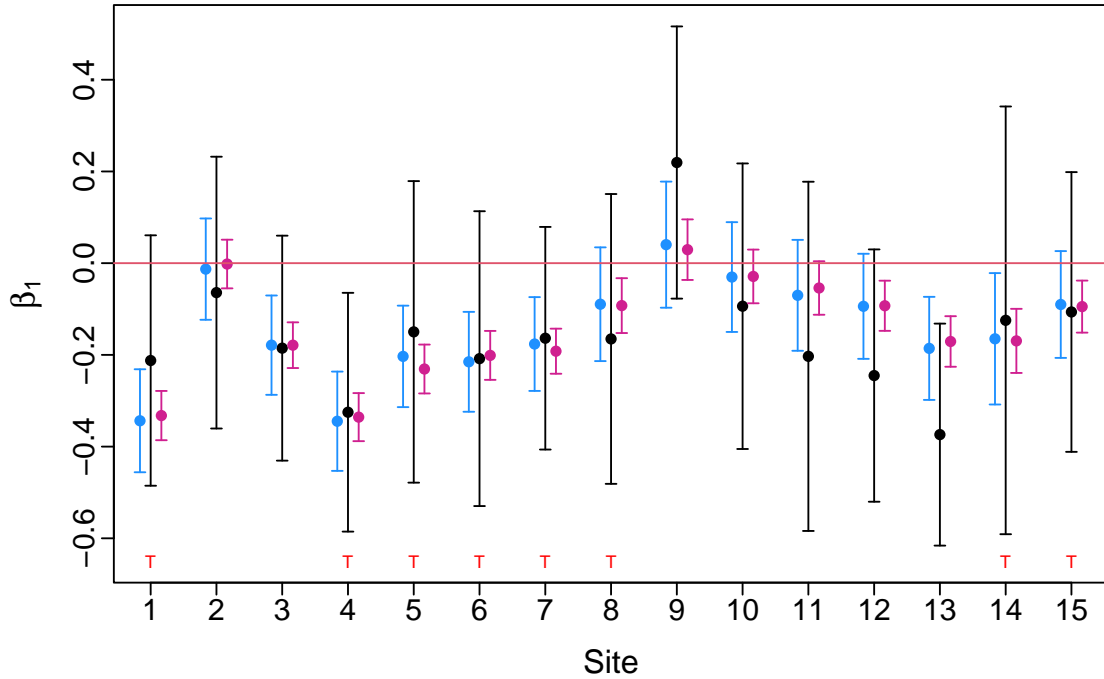


Figure 4.8: Posterior means and 95% CIs for  $\beta_1$  over all sites from ignoring dependence (pink), declustering (black) and the logistic model (blue). Treated sites are denoted with ‘T’ above the  $x$ -axis.

## 4.4 Discussion

The Poisson regression model introduced in Section 4.2 is relatively straightforward and captures both site and treatment effect through estimating the posterior mean of  $\lambda_{ij}$  in the before and after periods for each site independently. For any road traffic data that does not require accounting for RTM and reference data, the model could be used to estimate the effectiveness of a safety intervention. The model fits well with only five observations in either period and would be improved if more data were available. The data are also transformed to be in terms of the overall mean number of collisions over many sites, hence we can notice sites which do have abnormally high number of collisions relatively. Though randomised trials may seem ethically questionable as a potential waste of resources, they are a viable option as they omit selection bias. Though we may not necessarily randomly choose roads which have abnormally high collision rates, ones which ‘should’ have treatment, performing a randomised trial allows us to truly focus on whether the treatment works without concerning ourselves with issues of bias. Furthermore, VAS are an optimal choice

for randomised safety schemes as they are cheap to operate and low maintenance.

To mitigate issues of confounding treatment effects with time in Gateshead Council’s plan, a crossover or stepped wedge design could be advantageous. In a crossover design, signs could initially be placed in some localities and later rotated to others, allowing for within-site comparisons and controlling for time trends. Alternatively, a stepped wedge design could stagger the sign rollout across localities at different times, making it easier to separate the signs’ impact from other time-related factors. Both designs would yield more reliable insights into the signs’ effectiveness. The project team are hoping to implement this model with data from Gateshead’s VAS in the financial year 2023/24.

In Section 4.3, a bivariate threshold excess model was proposed to conduct a near-miss-based before/after safety evaluation using extreme value theory. The approach is able to combine near-miss incidents of different sites (treatment sites and control sites) and different periods (the before period and after period) to develop a bivariate model to capture dependence for the treatment effect estimation. The approach was applied to a before/after safety evaluation of the leading pedestrian interval (LPI). Pedestrian near-misses were collected from the treatment and control sites during the before and after periods using automated computer vision analysis techniques. The treatment effects were measured through linear modelling in the scale parameter. Bivariate EVT was used to account for dependence between consecutive extremes by assuming a first order Markov structure. Transforming the random variables to have unit Fréchet distributions and using a logistic model in the limiting distribution omits the need to remove dependence in the data which means we do not need to be wasteful of data – a common issue in EVT.

Use of the bivariate threshold excess model resulted in narrower CIs than from a declustering approach and was also more successful in estimating which sites had been treated. This was due to us being able to include all of the data in the analysis to inform the parameter estimates. The posterior distributions for the intercept term ( $\beta_1$ ) coincide for the model ignoring dependence and logistic model, with similar posterior means. This is also true for the other parameters. However, although ignoring dependence and fitting a GPD to all threshold excesses results in narrow CIs, when we ignore dependence our model is no longer valid. Using a declustering approach, to remove dependence, results in wide CIs which makes drawing conclu-

sions about the treatment effect ambiguous. This approach also results in more negative posterior estimates of the shape parameter which results in smaller tails in the fitted GPD.

The modelling approach can be extended as the data showed signs of spatial dependence and this can be accounted for through spatial extreme methods. This allows for further sharing of information across sites which is advantageous in extremes modelling as data are regularly scarce. As suggested by [Arun et al. \(2023\)](#), our approach uses stationary thresholds at the treated sites however the LPI treatment could also affect the thresholds, and holding them constant in both the before and after periods could be erroneous. Additionally, more attention could have been paid to threshold selection and the optimal declustering value.

Model	Site	Threshold, $u$		$\beta_0$	$\beta_1$	$\xi$				
Ignore Dependence	1	-5.6930	Mean	0.4108	-0.3324	-0.1809				
			95% CI	(0.3253, 0.4963)	(-0.3861, -0.2786)	(-0.2360, -0.1259)				
Declustering			Mean	0.9363	-0.2123	-0.3516				
			95% CI	(0.5911, 1.2816)	(-0.4853, 0.0607)	(-0.5916, -0.1115)				
Logistic			Mean	0.3981	-0.3436	-0.1649				
			95% CI	(0.3015, 0.4948)	(-0.4559, -0.2314)	(-0.2221, -0.1077)				
Ignore Dependence			2	-4.8800	Mean	0.1058	-0.0019	-0.1373		
					95% CI	(0.0147, 0.1968)	(-0.0549, 0.0512)	(-0.1949, -0.0797)		
Declustering					Mean	0.6136	-0.0642	-0.2552		
					95% CI	(0.2878, 0.9394)	(-0.3605, 0.2321)	(-0.4900, -0.0204)		
Logistic					Mean	0.1297	-0.0129	-0.1454		
					95% CI	(0.0251, 0.2344)	(-0.1234, 0.0975)	(-0.2082, -0.0826)		
Ignore Dependence					3	-5.9500	Mean	0.4687	-0.2308	-0.2119
							95% CI	(0.3967, 0.5408)	(-0.2841, -0.1775)	(-0.2538, -0.1700)
Declustering	Mean	0.9325					-0.1498	-0.3003		
	95% CI	(0.6369, 1.2280)					(-0.4786, 0.1790)	(-0.5286, -0.0719)		
Logistic	Mean	0.4695					-0.2033	-0.1910		
	95% CI	(0.3858, 0.5533)					(-0.3139, -0.0926)	(-0.2368, -0.1452)		
Ignore Dependence	4	-5.7400					Mean	0.4632	-0.3357	-0.2066
							95% CI	(0.3821, 0.5442)	(-0.3881, -0.2834)	(-0.2502, -0.1630)
Declustering			Mean	0.9938			-0.3250	-0.3682		
			95% CI	(0.6649, 1.3226)			(-0.5854, -0.0647)	(-0.6117, -0.1247)		
Logistic			Mean	0.4792			-0.3447	-0.2071		
			95% CI	(0.3865, 0.5719)			(-0.4529, -0.2364)	(-0.2630, -0.1513)		
Ignore Dependence			5	-5.9600			Mean	0.4687	-0.2308	-0.2119
							95% CI	(0.3967, 0.5408)	(-0.2841, -0.1775)	(-0.2538, -0.1700)
Declustering					Mean	0.9325	-0.1498	-0.3003		
					95% CI	(0.6369, 1.2280)	(-0.4786, 0.1790)	(-0.5286, -0.0719)		
Logistic					Mean	0.4695	-0.2033	-0.1910		
					95% CI	(0.3858, 0.5533)	(-0.3139, -0.0926)	(-0.2369, -0.1452)		
Ignore Dependence					6	-6.7500	Mean	0.409	-0.2011	-0.1837
							95% CI	(0.3245, 0.4935)	(-0.2544, -0.1478)	(-0.2309, -0.1364)
Declustering	Mean	0.8900					-0.2082	-0.2974		
	95% CI	(0.5393, 1.2407)					(-0.5297, 0.1133)	(-0.4987, -0.0962)		
Logistic	Mean	0.4177					-0.2151	-0.1719		
	95% CI	(0.3211, 0.5144)					(-0.3240, -0.1062)	(-0.2229, -0.1208)		
Ignore Dependence	7	-5.5300					Mean	0.3845	-0.1920	-0.1823
							95% CI	(0.3051, 0.4639)	(-0.2411, -0.1428)	(-0.2256, -0.1390)
Declustering			Mean	0.9984			-0.1637	-0.4098		
			95% CI	(0.6883, 1.3084)			(-0.4064, 0.0791)	(-0.6374, -0.1822)		
Logistic			Mean	0.3999			-0.1763	-0.2046		
			95% CI	(0.3082, 0.4915)			(-0.2786, -0.0740)	(-0.2541, -0.1551)		
Ignore Dependence			Mean	0.3183			-0.0925	-0.0834		
			95% CI	(0.2354, 0.4012)			(-0.1523, -0.0327)	(-0.1317, -0.0351)		

Table 4.2 continued from previous page

Model	Site	Threshold, $u$		$\beta_0$	$\beta_1$	$\xi$
Declustering			Mean	1.0146	-0.1652	-0.2931
			95% CI	(0.6750, 1.3542)	(-0.4812, 0.1508)	(-0.5060, -0.0801)
Logistic			Mean	0.3153	-0.0895	-0.1186
			95% CI	(0.2190, 0.4116)	(-0.2136, 0.0345)	(-0.1728, -0.0643)
Ignore Dependence			Mean	0.1779	0.0295	-0.1706
			95% CI	(0.0718, 0.2841)	(-0.0366, 0.0955)	(-0.2325, -0.1088)
Declustering	9	-6.1300	Mean	0.8241	0.2195	-0.4150
			95% CI	(0.4972, 1.1510)	(-0.0773, 0.5163)	(-0.6876, -0.1424)
Logistic			Mean	0.2068	0.0404	-0.1358
			95% CI	(0.0859, 0.3276)	(-0.0971, 0.1779)	(-0.2066, -0.0650)
Ignore Dependence			Mean	0.1200	-0.0289	-0.1340
			95% CI	(0.0279, 0.2122)	(-0.0874, 0.0296)	(-0.1901, -0.0778)
Declustering	10	-5.1600	Mean	0.6670	-0.0939	-0.2505
			95% CI	(0.3498, 0.9843)	(-0.4052, 0.2175)	(-0.4641, -0.0369)
Logistic			Mean	0.1327	-0.0303	-0.1072
			95% CI	(0.0286, 0.2368)	(-0.1500, 0.0894)	(-0.1741, -0.0402)
Ignore Dependence			Mean	0.2028	-0.0541	-0.1249
			95% CI	(0.1095, 0.2960)	(-0.1124, 0.0042)	(-0.1661, -0.0837)
Declustering	11	-6.5155	Mean	0.7097	-0.2032	-0.1831
			95% CI	(0.3112, 1.1082)	(-0.5839, 0.1776)	(-0.4316, 0.0654)
Logistic			Mean	0.2102	-0.0701	-0.1239
			95% CI	(0.1033, 0.3171)	(-0.1910, 0.0508)	(-0.1809, -0.0668)
Ignore Dependence			Mean	0.3638	-0.0929	-0.1961
			95% CI	(0.2757, 0.4518)	(-0.1476, -0.0382)	(-0.2467, -0.1455)
Declustering	12	-6.2600	Mean	1.0513	-0.2451	-0.3788
			95% CI	(0.6634, 1.4393)	(-0.5202, 0.0300)	(-0.6586, -0.0989)
Logistic			Mean	0.3778	-0.0941	-0.1688
			95% CI	(0.2774, 0.4783)	(-0.2086, 0.0205)	(-0.2255, -0.1121)
Ignore Dependence			Mean	0.2562	-0.1708	-0.1470
			95% CI	(0.1641, 0.3482)	(-0.2259, -0.1157)	(-0.1919, -0.1021)
Declustering	13	-5.0500	Mean	0.9524	-0.3738	-0.3970
			95% CI	(0.6237, 1.2811)	(-0.6159, -0.1318)	(-0.6729, -0.1212)
Logistic			Mean	0.2456	-0.1858	-0.1634
			95% CI	(0.1394, 0.3518)	(-0.2982, -0.0734)	(-0.2260, -0.1008)
Ignore Dependence			Mean	0.3635	-0.1695	-0.0788
			95% CI	(0.2895, 0.4374)	(-0.2394, -0.0996)	(-0.1207, -0.0368)
Declustering	14	-10.4100	Mean	0.7947	-0.1247	-0.0480
			95% CI	(0.4865, 1.1029)	(-0.5910, 0.3417)	(-0.2650, 0.1689)
Logistic			Mean	0.3666	-0.1649	-0.0572
			95% CI	(0.2836, 0.4496)	(-0.3082, -0.0216)	(-0.1056, -0.0087)
Ignore Dependence			Mean	0.2499	-0.0947	-0.1548
			95% CI	(0.1612, 0.3386)	(-0.1514, -0.0380)	(-0.2156, -0.0940)
Declustering	15	-5.1635	Mean	0.6906	-0.1064	-0.2655
			95% CI	(0.3828, 0.9985)	(-0.4114, 0.1986)	(-0.4958, -0.0352)
Logistic			Mean	0.2289	-0.0901	-0.1454
			95% CI	(0.1274, 0.3304)	(-0.2066, 0.0264)	(-0.2074, -0.0833)

Table 4.2: Numerical posterior summaries for model parameters at all intersections. For each parameter we show the posterior mean and 95% credible interval from the three analyses (ignoring dependence, declustering and logistic) and the chosen threshold.

# Chapter 5

## Modelling spatial extremes from the LPI intervention

In this chapter, we continue our analysis of the PET data introduced in Section 1.3.3 and analysed in Chapter 4. The data supplied longitude and latitude of all 15 sites which are potential covariates in a model that admits spatial variation. To check for spatial dependence we plotted the correlation between PET data for all pairs of sites against the distance between them in kilometres as shown in the left plot of Figure 5.1. There is a clear decay in correlation as distance between sites increases. Hence, our inference might be improved by accounting for this spatial dependence, potentially increasing the overall explanatory power of the model. By capturing such spatial correlation in the model, we aim to improve posterior parameter estimates and therefore the estimation of treatment effects and uncertainty quantification. When spatial dependence exists, usually observations at nearby locations tend to be more similar to each other than to observations at distant locations. This spatial correlation can be quantified and incorporated in the model to predict at neighbouring, unobserved sites.

### 5.1 Background

Spatial extremes modelling is used to analyse and model the occurrence of extreme events in space. Additionally, it can be used to identify areas that are particularly susceptible to extreme events, which can aid in the development of targeted adaptation and mitigation measures. In road safety analyses, it is common that

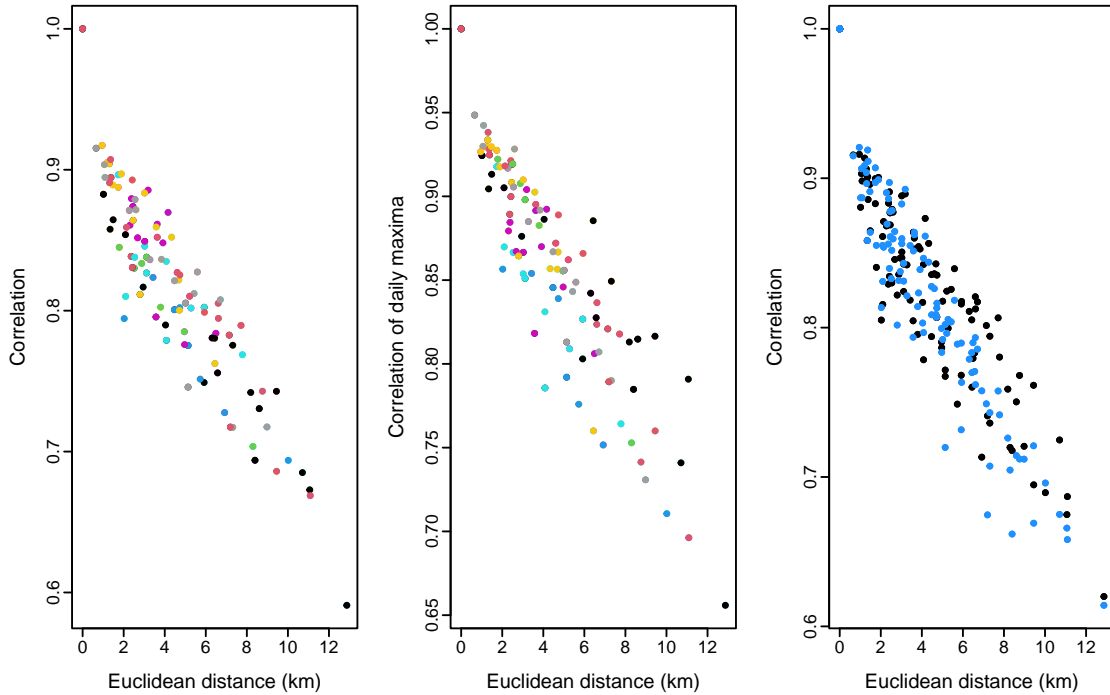


Figure 5.1: Left: Correlation of all PETs against distance between all pairs of sites. Middle: Correlation of daily maxima against distances for all pairs of sites. Right: Correlations between all PETs against distances for all pairs of sites in the before period (black) and the after period (blue).

measurements are taken at many different sites and modelling is performed assuming these sites are spatially independent. However, realistically, this is frequently false. Given their proximity, sites that are geographically close are likely to share similar characteristics such as traffic and pedestrian volumes or weather conditions. This inevitably leads to correlation. Thus, it is not unexpected to observe such spatial dependence among these sites. As mentioned, road safety practitioners are often in possession of observed data without many covariates. They are then left to attempt to pair these observations with explanatory variables over the same time period. This is very time-consuming and sometimes the extra variables do not exist. Therefore, any chance there is to exploit all possible patterns in the available data should be taken. Extreme events are sparse by definition, such as the PET data analysed in Chapter 4, and spatial modelling allows us to borrow strength across locations for better marginal estimation.

Interest lies in developing a framework in which it is possible to estimate prob-



abilities of joint events over space and of events at sites for which data are not available. The need for statistical techniques for spatial extremes data is commonplace across many disciplines. For example, it is used in climatology (Davison and Gholamrezaee, 2012, Reich and Shaby, 2019), hydrology (Huser and Davison, 2014), and public health (Vettori et al., 2019). Some important texts in this area include Banerjee et al. (2015), Cressie (2015), Stein (2012). The classical theory is outlined in Davison et al. (2012) and includes an application of the main types of spatial extremes methods on rainfall data; a more recent review of current methods is provided by Huser and Wadsworth (2022). An advantage of this approach is it allows for joint marginals for model parameters over all sites, sharing information across sites to have marginal fits at individual sites can potentially increase precision in parameter estimates. It also brings opportunity for interpolation between parameter values so for an unobserved site, meaning we can potentially estimate extreme events across zones rather than single sites.

Within road safety, researchers have found that spatial analyses are advantageous as road collisions are subject to both spatial and temporal variations (Loo and Anderson, 2015). Spatial analysis examines the relationship between road safety attributes and their impact on surrounding areas, as well as the spatial variation of explanatory parameters' influence. Agüero-Valverde (2013) uses a full Bayes hierarchical approach to estimate crash frequency using a multivariate conditional autoregressive model to describe spatial random effects. They show that the multivariate spatial model outperforms its univariate counterpart through measurement of the DIC. Zeng and Huang (2014) propose a Bayesian joint model to estimate crash frequencies of road segments and intersections in an urban road network where the spatial correlations between adjacent road segments and intersections are accounted for. Wang and Huang (2016) propose a Bayesian hierarchical joint model for road network safety evaluation including random effects to capture spatial correlation. Further examples of spatial analyses include El-Basyouny and Sayed (2009b), Noland and Quddus (2004), Quddus (2008). A recent full review of spatial analyses in road safety research can be found in Ziakopoulos and Yannis (2020). To the best of the author's knowledge, there have been no studies that apply spatial extreme value theory (EVT) methods to road safety analyses.

We propose a latent variable model to introduce spatial variation in the GEV

parameters. We have applied our methods to block maxima as it is currently the setting where extremal spatial methods are most commonly used and for which the existing theory supports. However, the extension to threshold modelling would allow for more flexible inference, and although research in this area is still in its early stages, it holds promise for future applications. Some examples of spatial models for threshold excesses are [Cooley et al. \(2007\)](#), [Roth et al. \(2012\)](#), [Sharkey and Winter \(2019\)](#); for a thorough overview of the current state of research in this field, we recommend referring to the comprehensive review by [Huser and Wadsworth \(2022\)](#). Our methodology assumes a standard extreme value model to describe the data generation process at extreme levels for each location. In addition, we make the assumption that the unobservable model parameters represent realisations of a smooth stochastic spatial process ([Casson and Coles, 1999](#)). Traditional inference methods, such as maximum likelihood, are intractable. Therefore, we use MCMC methods which are proven to be effective in estimating the parameters of latent spatial process models ([Diggle et al., 1998](#)).

## 5.2 Preliminary inference

Currently most spatial extremes research is performed within the block maxima paradigm. We therefore assume the GEV distribution for the data (see Section 2.2.1). To see how strong the spatial correlation is, we plotted the correlation between daily maxima between pairs of sites against their Euclidean distance apart (the argument behind this choice of block length is outlined in Section 4.3.2). The middle plot in Figure 5.1 shows the correlation between daily maxima decreases as sites get further apart, meaning sites closer together seem to possess similar patterns and using this in the model could be advantageous.

Firstly, the block maxima model from Section 4.3.2 is adapted to allow for data from all sites to be fed into the model at once, rather than site by site. This joint model allows for draws from the likelihood over all sites in a single MCMC algorithm. The log-likelihood is therefore calculated as

$$\ell(\mu, \sigma, \xi | x) = \sum_{j=1}^{n_s} \sum_{i=1}^N \log \{g(\mu_j, \sigma_j, \xi_j; x_{ij})\},$$

where  $n_s$  is the number of sites and  $N$  is the number of observations per site. We utilise an MCMC scheme to draw estimates of each parameter for each site and obtain posterior samples. This scheme is implemented as laid out in Section 4.3.2, however we use the joint log-likelihood in the formulation of the acceptance probabilities and draw parameter proposals from multivariate normal random walks. For example, we obtain draws for  $\beta_0$  by sampling from  $N_{n_s}(\beta_0, \Lambda_{\beta_0})$ , where  $\beta_0 = (\beta_0^1, \dots, \beta_0^{n_s})^T$  and  $\Lambda_{\beta_0}$  is the tuning matrix for the normal random walk. The same process is followed for the parameters  $\beta_1, \eta$ , and  $\xi$ , where,  $\eta = \log \sigma$  to respect the positivity of the scale parameter. All parameters were given uninformative normal prior distributions with zero mean and a variance of 100, as in the individual sites model. Upon comparison of the parameter estimates from the individual model against the joint model, the marginal posterior distributions coincide for each parameter over all sites.

As the shape parameter is estimated to be around zero for all sites, as shown in Figure 4.6, we fixed  $\xi$  over all sites, which in turn reduced the number of parameters to estimate from 60 to 46. The inference was performed again with the MCMC scheme set up as before, except we draw estimates  $\xi^* \sim N(\xi, \Lambda_\xi)$  where  $\Lambda_\xi$  is now the scalar tuning parameter for the normal random walk. Again, posterior estimates for parameters  $\beta_0, \beta_1, \eta$  coincide with the individual model estimates. The posterior mean for the shape parameter,  $\xi$  is  $-0.0026$  with 95% CI  $(-0.0965, 0.0946)$ .

### 5.3 Spatial extremes

To account for spatial correlation between neighbouring sites, we include Gaussian processes (GP) in the expressions for the location and scale parameters. As in Section 5.2, resulting from the knowledge that  $\xi$  is estimated to be approximately zero for all sites, we fix  $\xi$  across all sites and omit the need for a GP component to be included. GPs are often used to model spatial dependence because they provide a flexible and computationally efficient way to incorporate prior information about the spatial dependence of a variable of interest. One of the key advantages of GPs is that they are non-parametric, meaning that they do not rely on a fixed set of parameters to describe the underlying function. Instead, they use a covariance function, also known as a kernel function, to describe the similarity between any two points in

the spatial domain. The kernel function encodes prior information about the spatial dependence of the variable of interest, such as whether the variable is more similar between nearby locations or more distant locations. The exponential covariance kernel, also known as the radial basis function (RBF) kernel, is a popular choice for a Gaussian process because it is smooth, infinitely differentiable, and isotropic.

The model is set out as follows. As before we have

$$X_{tj} | \mu_{tj}, \sigma_j, \xi \sim GEV(\mu_{tj}, \sigma_j, \xi),$$

where  $t = 0, 1$  denotes the before and after periods respectively, and  $j = 1, \dots, n_s$  denotes the site. However, with the addition of the GPs, the location parameter becomes

$$\begin{aligned} \mu_0 &= \beta_0 + S_\mu(\alpha_\mu, \lambda_\mu), \\ \mu_1 &= \beta_0 + \beta_1 + S_\mu(\alpha_\mu, \lambda_\mu), \end{aligned}$$

where  $S_\mu$  is a zero mean, stationary Gaussian process with exponential covariance function, such that

$$S_\mu \sim N_m \left( 0, \alpha_\mu \exp \left\{ -\frac{\|h\|}{\lambda_\mu} \right\} \right),$$

with unknown sill and range parameters  $\alpha_\mu$  and  $\lambda_\mu$ . The matrix,  $\|h\|$ , is obtained by computing the pairwise distances between sites using their coordinates. Each element of  $\|h\|$  represents the Euclidean distance in kilometres between two sites  $j$  and  $j'$ . Similar formulations are used for the scale parameter.

In this instance, we are modelling a before and after period through a time varying location parameter. To avoid the time vector inclusion, we separate the location parameter to have  $\mu_0$  in the before period and  $\mu_1$  in the after period. We denote the data over all sites as  $(x_1, \dots, x_{n_s})$ , where  $n_s$  is the number of sites. We use conjugate priors where possible and let all  $\beta$  coefficients follow independent multivariate normal distributions, and take independent inverse-gamma distributions for  $\alpha_\mu$ . We use relatively uninformative gamma distributions for  $\lambda_\mu$  where no conjugate priors are available, similarly for the sill and range parameters for the scale parameter. The priors are similarly chosen for the remaining GEV parameters (Davison et al.,

2012). The full conditional distributions are defined as follows:

$$\begin{aligned}
 \pi(\mu_0|\cdots) &\propto \pi(\mu_0|\alpha_\mu, \lambda_\mu, \beta_{\mu_0})\pi(x_0|\mu_0, \sigma, \xi), \\
 \pi(\mu_1|\cdots) &\propto \pi(\mu_1|\alpha_\mu, \lambda_\mu, \beta_{\mu_1})\pi(x_1|\mu_1, \sigma, \xi), \\
 \pi(\alpha_\mu|\cdots) &\propto \pi(\alpha_\mu|\kappa_{\alpha_\mu}^*, \theta_{\alpha_\mu}^*)\pi(\mu_0|\alpha_\mu, \lambda_\mu, \beta_{\mu_0})\pi(\mu_1|\alpha_\mu, \lambda_\mu, \beta_{\mu_1}), \\
 \pi(\lambda_\mu|\cdots) &\propto \pi(\lambda_\mu|\kappa_{\lambda_\mu}^*, \theta_{\lambda_\mu}^*)\pi(\mu_0|\alpha_\mu, \lambda_\mu, \beta_{\mu_0})\pi(\mu_1|\alpha_\mu, \lambda_\mu, \beta_{\mu_1}), \\
 \pi(\beta_\mu|\cdots) &\propto \pi(\beta_\mu|\gamma_\mu^*, \Sigma_\mu^*)\pi(\mu_0|\alpha_\mu, \lambda_\mu, \beta_{\mu_0})\pi(\mu_1|\alpha_\mu, \lambda_\mu, \beta_{\mu_1}),
 \end{aligned}$$

where  $\kappa_\cdot^*$ ,  $\theta_\cdot^*$ ,  $\gamma_\cdot^*$ ,  $\Sigma_\cdot^*$  are the hyperparameters of the prior distributions. The subscript dot notation denotes all indexes related to that particular parameter. Also,  $x_0$  denotes the data in the before period and  $x_1$  in the after period, across all sites. We assume that the spatial correlation remains constant over time and therefore fix  $\alpha_\mu$  and  $\lambda_\mu$  across the before and after periods. The right-hand-side plot of Figure 5.1 compares the spatial correlations in the before and after periods and shows that they do not vary between the two time periods. The full conditional distributions for the remaining GEV parameters are similarly derived.

### 5.3.1 Model description

The reparameterised model is then given by,

$$\begin{aligned}
 \mu_0 &= \beta_0 + \alpha_\mu \exp(-\|h\|/\lambda_\mu), \\
 \mu_1 &= \beta_0 + \beta_1 + \alpha_\mu \exp(-\|h\|/\lambda_\mu), \\
 \eta &= \beta_2 + \alpha_\eta \exp(-\|h\|/\lambda_\eta), \\
 \xi &= \beta_3,
 \end{aligned}$$

where  $\eta = \log \sigma$  to respect the positivity of the scale parameter. As we have a before and after period which was previously modelled through  $t$ , we can think of this as having a location parameter  $\mu_0$  for the before period ( $t = 0$ ) and  $\mu_1$  for the after period ( $t = 1$ ). Hence in our FCDs, to draw  $\beta_0$  and  $\beta_1$  we will use their respective  $\mu_\cdot$ . Then, we can deduce that the GEV parameters follow normal distributions, and

the prior distributions are as follows:

$$\begin{aligned}
 \mu_0 &\sim N_m(\beta_0, \Sigma_\mu); & \mu_1 &\sim N_m(\beta_0 + \beta_1, \Sigma_\mu), \\
 \eta &\sim N_m(\beta_2, \Sigma_\eta); & \xi &\sim N(\beta_3, \nu_\xi), \\
 \beta_0 &\sim N_m(\gamma_{\beta_0}, \Sigma_{\beta_0}); & \beta_1 &\sim N_m(\gamma_{\beta_1}, \Sigma_{\beta_1}), \\
 \beta_2 &\sim N_m(\gamma_{\beta_2}, \Sigma_{\beta_2}); & \beta_3 &\sim N(\gamma_{\beta_3}, \nu_{\beta_3}), \\
 \alpha_\mu &\sim \text{InvGa}(\kappa_{\alpha_\mu}, \theta_{\alpha_\mu}); & \alpha_\eta &\sim \text{InvGa}(\kappa_{\alpha_\eta}, \theta_{\alpha_\eta}), \\
 \lambda_\mu &\sim \text{Ga}(\kappa_{\lambda_\mu}, \theta_{\lambda_\mu}); & \lambda_\eta &\sim \text{Ga}(\kappa_{\lambda_\eta}, \theta_{\lambda_\eta}), \\
 \nu_\xi &\sim \text{Ga}(a_0, b_0),
 \end{aligned}$$

where  $N_m(\cdot, \cdot)$  denotes a multivariate normal distribution. The variances for the normal distributions for the GEV parameters,  $\Sigma_\mu, \Sigma_\eta$ , are the corresponding Gaussian random field, e.g.  $\Sigma_\eta = \alpha_\eta \exp(-\|h\|/\lambda_\eta)$ . As we have fixed the shape parameter,  $\xi$  over all sites, and for ease of calculation,  $\nu_{\beta_3}$  and  $\nu_\xi$  are scalar precisions, where we set  $\nu_{\beta_3} = 0.01$ .

### 5.3.2 Parameter FCDs

Due to the use of conjugate priors, the full conditional distributions are available for  $\beta_0, \beta_1, \beta_2, \beta_3, \alpha_\mu, \alpha_\eta$  and  $\nu_\xi$  to sample from directly. Firstly, to derive the full conditional distribution of  $\beta_0$ , we need to use Bayes theorem and the properties of normal distributions. Bayes theorem states that:

$$\pi(\beta_0 | \mu_0, \gamma_{\beta_0}, \Sigma_\mu, \Sigma_{\beta_0}) = \frac{\pi(\mu_0 | \beta_0, \Sigma_\mu) \pi(\beta_0 | \gamma_{\beta_0}, \Sigma_{\beta_0})}{\pi(\mu_0 | \gamma_{\beta_0}, \Sigma_{\beta_0}, \Sigma_\mu)}.$$

Up to proportionality, the posterior distribution can be expressed as,

$$\pi(\beta_0 | \mu_0, \gamma_{\beta_0}, \Sigma_\mu, \Sigma_{\beta_0}) \propto \pi(\mu_0 | \beta_0, \Sigma_\mu) \pi(\beta_0 | \gamma_{\beta_0}, \Sigma_{\beta_0}).$$

Upon substituting the expressions for  $\pi(\mu_0|\beta_0, \Sigma_\mu)$  and  $\pi(\beta_0|\gamma_{\beta_0}, \Sigma_{\beta_0})$ , we have:

$$\begin{aligned} \pi(\beta_0|\cdots) &\propto \exp\left\{-\frac{1}{2}(\mu_0 - \beta_0)^T \Sigma_\mu^{-1}(\mu_0 - \beta_0)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\beta_0 - \gamma_{\beta_0})^T \Sigma_{\beta_0}^{-1}(\beta_0 - \gamma_{\beta_0})\right\}, \\ &\propto \exp\left\{\frac{1}{2}\beta_0^T (\Sigma_\mu^{-1} + \Sigma_{\beta_0}^{-1}) \beta_0 + \beta_0^T (\Sigma_\mu^{-1}\mu_0 + \Sigma_{\beta_0}^{-1}\gamma_{\beta_0})\right\}. \end{aligned}$$

We set

$$A = \Sigma_\mu^{-1} + \Sigma_{\beta_0}^{-1}, \quad B = \Sigma_\mu^{-1}\mu_0 + \Sigma_{\beta_0}^{-1}\gamma_{\beta_0},$$

then the FCD for  $\beta_0$  is expressed as,

$$\pi(\beta_0|\cdots) \propto \exp\left\{-\frac{1}{2}(\beta_0 - A^{-1}B^T A)(\beta_0 - A^{-1}B)\right\},$$

which we recognise as proportional to a multivariate normal distribution, and hence we can write

$$\beta_0|\cdots \sim N_m(A^{-1}B, A^{-1}). \quad (5.1)$$

Similarly for  $\beta_1$  we have,

$$\begin{aligned} \pi(\beta_1|\cdots) &\propto \exp\left\{-\frac{1}{2}(\mu_1 - (\beta_0 + \beta_1))^T \Sigma_\mu^{-1}(\mu_1 - (\beta_0 + \beta_1))\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\beta_1 - \gamma_{\beta_1})^T \Sigma_{\beta_1}^{-1}(\beta_1 - \gamma_{\beta_1})\right\}, \\ &\propto \exp\left\{\frac{1}{2}\beta_1^T (\Sigma_\mu^{-1} + \Sigma_{\beta_1}^{-1}) \beta_1 + \beta_1^T (\Sigma_\mu^{-1}\mu_1 - \Sigma_\mu^{-1}\beta_0 + \Sigma_{\beta_1}^{-1}\gamma_{\beta_1})\right\}. \end{aligned}$$

We set

$$C = \Sigma_\mu^{-1} + \Sigma_{\beta_1}^{-1}, \quad D = \Sigma_\mu^{-1}(\mu_1 - \beta_0) + \Sigma_{\beta_1}^{-1}\gamma_{\beta_1},$$

then the FCD is expressed as,

$$\pi(\beta_1|\cdots) \propto \exp\left\{-\frac{1}{2}(\beta_1 - C^{-1}D^T C)(\beta_1 - C^{-1}D)\right\},$$

which is proportional to a multivariate normal distribution and so we write,

$$\beta_1|\cdots \sim N_m(C^{-1}D, C^{-1}). \quad (5.2)$$

Following the same methods, the prior distribution of  $\beta_2$  is chosen to be normal with mean  $\gamma_{\beta_2}$  and variance matrix  $\Sigma_{\beta_2}$ . Therefore the mean parameters for  $\eta$  are drawn from the FCD,

$$\beta_2 | \dots \sim N_m(E^{-1}F, E^{-1}). \quad (5.3)$$

where

$$E = \Sigma_{\eta}^{-1} + \Sigma_{\beta_2}^{-1}, \quad F = \Sigma_{\eta}^{-1}\eta + \Sigma_{\beta_2}^{-1}\gamma_{\beta_2}.$$

The FCD for  $\beta_3$  is calculated as

$$\begin{aligned} \pi(\beta_3 | \dots) &\propto \exp \left\{ -\frac{1}{2} (\nu_{\xi}(\xi - \beta_3)^2 + \nu_{\beta_3}(\beta_3 - \gamma_{\beta_3})^2) \right\}, \\ &\propto \exp \left\{ -\frac{1}{2} (\beta_3^2(\nu_{\xi} + \nu_{\beta_3}) - 2\beta_3(\nu_{\xi}\xi + \nu_{\beta_3}\gamma_{\beta_3})) \right\}, \end{aligned}$$

and hence,

$$\beta_3 | \dots \sim N \left( \frac{\nu_{\xi}\xi + \nu_{\beta_3}\gamma_{\beta_3}}{\nu_{\xi} + \nu_{\beta_3}}, \nu_{\xi} + \nu_{\beta_3} \right).$$

As for the hyperparameter for the precision, with the use of conjugate priors the FCDs are calculated as:

$$\nu_{\xi} | \dots \sim Ga \left( a_0 + \frac{1}{2}, b_0 + \frac{1}{2}(\xi - \beta_3)^2 \right). \quad (5.4)$$

For the Gaussian process parameters we have the following priors:

$$\begin{aligned} \alpha_{\mu} &\sim InvGa(\kappa_{\alpha_{\mu}}, \theta_{\alpha_{\mu}}), \\ \lambda_{\mu} &\sim Ga(\kappa_{\lambda_{\mu}}, \theta_{\lambda_{\mu}}), \end{aligned}$$

similarly for those related to the scale parameter,  $\eta$ . Then, due to the use of conjugate priors, the sill parameter  $\alpha_{\mu}$  is drawn directly from an inverse-gamma distribution:

$$\begin{aligned} \pi(\alpha_{\mu} | \dots) &\propto \theta_{\alpha_{\mu}}^{\kappa_{\alpha_{\mu}}} \alpha_{\mu}^{\kappa_{\alpha_{\mu}}-1} \exp \left\{ \frac{\theta_{\alpha_{\mu}}}{\alpha_{\mu}} \right\} \\ &\times \exp \left\{ -\frac{1}{2} (\mu_0 - \beta_0)^T \Sigma_{\mu}^{-1} (\mu_0 - \beta_0) \right\} \\ &\times \exp \left\{ -\frac{1}{2} (\mu_1 - (\beta_0 + \beta_1))^T \Sigma_{\mu}^{-1} (\mu_1 - (\beta_0 + \beta_1)) \right\}, \end{aligned}$$



which equates to

$$\alpha_\mu | \dots \sim \text{InvGa}(G, H), \quad (5.5)$$

where the shape and rate are defined as

$$G = \kappa_{\alpha_\mu} + 1, \\ H = \theta_{\alpha_\mu} + \frac{1}{2} \alpha_\mu (\mu_0 - \beta_0)^T \Sigma_\mu^{-1} (\mu_0 - \beta_0) + \frac{1}{2} \alpha_\mu (\mu_1 - (\beta_0 + \beta_1))^T \Sigma_\mu^{-1} (\mu_1 - (\beta_0 + \beta_1)).$$

Similarly,  $\alpha_\eta$  is drawn from an inverse-gamma distribution with shape and rate parameters:

$$\kappa_{\alpha_\eta} + \frac{1}{2}, \\ \theta_{\alpha_\eta} + \frac{1}{2} (\eta - \beta_2)^T \Sigma_\eta^{-1} (\eta - \beta_2).$$

To update the range parameter,  $\lambda$ , where we use non-conjugate priors, we require a Metropolis step for the proposals. The full MCMC algorithm is outlined in Algorithm 6. The Jacobian term is included in the acceptance probability in the Metropolis-Hastings step for the range parameter to account for the change in volume caused by the reparameterisation in the proposal distribution. It adjusts the acceptance probability so that the Markov chain has the correct stationary distribution. It is typically calculated by taking the determinant of the Jacobian matrix of the transformation from the current state to the proposed state. Here, the Jacobian term is

$$\frac{\partial \lambda_\mu}{\partial \zeta_\mu} = \exp \zeta_\mu.$$

When applying the change of variable formula, the Jacobian is the absolute value of the derivative of the transformation function, however as the Jacobian is scalar the absolute value is not needed.

### 5.3.3 Application

In what follows, we implement the MCMC scheme from Algorithm 6. We have that  $q(\psi^* | \psi) = N(\psi^*; \psi, \Lambda)$  where the innovation matrix  $\Lambda = \gamma \widehat{\text{Var}}(\psi | x)$ , with  $\widehat{\text{Var}}(\psi | x)$  obtained from a pilot run and  $\gamma$  is chosen to give an acceptance rates of around

---

**Algorithm 6** MCMC algorithm

---

Given a current value of the Markov chain

$$\psi^{(j)} = \left( \mu_0^{(j)}, \mu_1^{(j)}, \eta^{(j)}, \xi^{(j)}, \alpha_\mu^{(j)}, \lambda_\mu^{(j)}, \alpha_\eta^{(j)}, \lambda_\eta^{(j)}, \beta_0^{(j)}, \beta_1^{(j)}, \beta_2^{(j)}, \beta_3^{(j)}, \nu_\xi^{(j)} \right),$$

the next state of the chain  $\psi^{(j+1)}$  is obtained as follows:

1: Update the GEV parameters at each site:

**for**  $t=(0,1)$  **do**

- (a) Each component of  $\mu_t^{(j)} = (\mu_{t,1}^{(j)}, \dots, \mu_{t,n_s}^{(j)})$  is updated singly through generating a proposal  $\mu_{t,s}^*$  from a symmetric random walk and compute the acceptance probability

$$A(\mu_{t,s}^* | \mu_{t,s}^{(j)}) = \min \left\{ 1, \frac{\pi(y_{t,s} | \mu_{t,s}^*, \sigma_s^{(j)}, \xi^{(j)}) \times \pi(\mu_{t,s}^* | \alpha_\mu, \lambda_\mu, \beta_\mu)}{\pi(y_{t,s} | \mu_{t,s}^{(j)}, \sigma_s^{(j)}, \xi^{(j)}) \times \pi(\mu_{t,s}^{(j)} | \alpha_\mu, \lambda_\mu, \beta_\mu)} \right\}.$$

- (b) With probability  $A(\mu_{t,s}^* | \mu_{t,s}^{(j)})$  the  $\mu_{t,s}$  component of  $\psi^{(j+1)}$  is set  $\mu_{t,s}^*$ ; otherwise,  $\mu_{t,s}^{(j+1)} = \mu_{t,s}^{(j)}$ .

**end do**

- $\sigma_s^{(j)}$  and  $\xi^{(j)}$  are updated similarly.

2: Update the regression parameters:

- Due to the use of conjugate priors,  $\beta_0$  and  $\beta_1$  are drawn directly from a multivariate normal distributions as shown in Equation (5.1) and Equation (5.2).
- The regression parameters for the GEV scale and shape are updated similarly.

3: Update the sill parameters of the covariance function:

- Due to the use of conjugate priors  $\alpha_\mu$  is drawn from an inverse-gamma distribution as shown in Equation (5.5).
- The sill parameters for the GEV scale and shape are updated similarly.

4: Update the range parameters of the covariance function:

- Here we have a non-conjugate prior and hence require a Metropolis step. To respect the positivity of  $\lambda$ , we generate proposals of  $\zeta = \log \lambda$  from a symmetric, normal random walk.
- Generate a proposal  $\zeta_\mu^* \sim N(\zeta_\mu^{(j)}, \Lambda_{\zeta_\mu})$  and compute the acceptance probability:  $A(\zeta_\mu^* | \zeta_\mu^{(j)}) = \min \{1, \Psi\}$ , where

$$\Psi = \frac{\pi(\mu_0^{(j)} | \alpha_\mu^{(j)}, \lambda_\mu^*, \beta_{\mu_0}^*)}{\pi(\mu_0^{(j)} | \alpha_\mu^{(j)}, \lambda_\mu^{(j)}, \beta_{\mu_0}^{(j)})} \times \frac{\pi(\mu_1^{(j)} | \alpha_\mu^{(j)}, \lambda_\mu^*, \beta_{\mu_1}^*)}{\pi(\mu_1^{(j)} | \alpha_\mu^{(j)}, \lambda_\mu^{(j)}, \beta_{\mu_1}^{(j)})} \\ \times \left( \frac{\lambda_\mu^*}{\lambda_\mu^{(j)}} \right)^{\kappa_{\lambda_\mu^*} - 1} \exp \left\{ -\theta_{\lambda_\mu^*}^* (\lambda_\mu^* - \lambda_\mu^{(j)}) \right\} \times \frac{\exp \zeta_\mu^*}{\exp \zeta_\mu},$$

and  $\frac{\exp \zeta_\mu^*}{\exp \zeta_\mu}$  is the Jacobian term and  $\lambda_\mu = \exp \zeta_\mu$ .

- $\lambda_\eta$  is updated similarly.

5: Draw precisions  $\nu_\xi$  from Equation (5.4).

---

25%, see Section 2.4.2. We define the prior parameters to be

$$\begin{aligned} \gamma_{\beta_0} &= -6 \cdot \mathbf{1}, & \Sigma_{\beta_0} &= \text{diag}\{10 \cdot \mathbf{1}\}, \\ \gamma_{\beta_1} &= 0 \cdot \mathbf{1}, & \Sigma_{\beta_1} &= \text{diag}\{10 \cdot \mathbf{1}\}, \\ \gamma_{\beta_2} &= 0 \cdot \mathbf{1}, & \Sigma_{\beta_2} &= \text{diag}\{10 \cdot \mathbf{1}\}, \\ \gamma_{\beta_3} &= 0, & \nu_{\beta_3} &= 0.01, \\ \kappa_{\alpha_\mu} &= 1, & \theta_{\alpha_\mu} &= 5 \\ \kappa_{\alpha_\eta} &= 1, & \theta_{\alpha_\eta} &= 5 \\ \kappa_{\lambda_\mu} &= 10, & \theta_{\lambda_\mu} &= 1 \\ \kappa_{\lambda_\eta} &= 10, & \theta_{\lambda_\eta} &= 1, \\ a_0 &= 0.1, & b_0 &= 0.1, \end{aligned}$$

with  $\mathbf{1}$  being a  $n_s \times 1$  vector of 1s. As suggested by [Banerjee et al. \(2015\)](#), informative priors should be used for the parameters  $\alpha$ , and  $\lambda$ , of the covariance functions, in order to yield non-degenerate marginal posterior distributions for them.

The MCMC scheme was run for 500k iterations with a thin of 500 for sufficiently low auto-correlation. From visual checks of all trace plots, the posterior distributions show good mixing and convergence to the target distribution. Marginal posterior distributions for parameters  $\eta$  and  $\xi$  are in agreement with those from the joint model outlined in Section 5.2, with a mean difference of 0.004 between the corresponding parameters. Like before, we are interested in the change in location parameter,  $\mu$ , from before the LPI treatment, to after. Hence here we will compare the values of  $\mu_0$  and  $\mu_1$  to measure the treatment effect. Figure 5.2 shows posterior

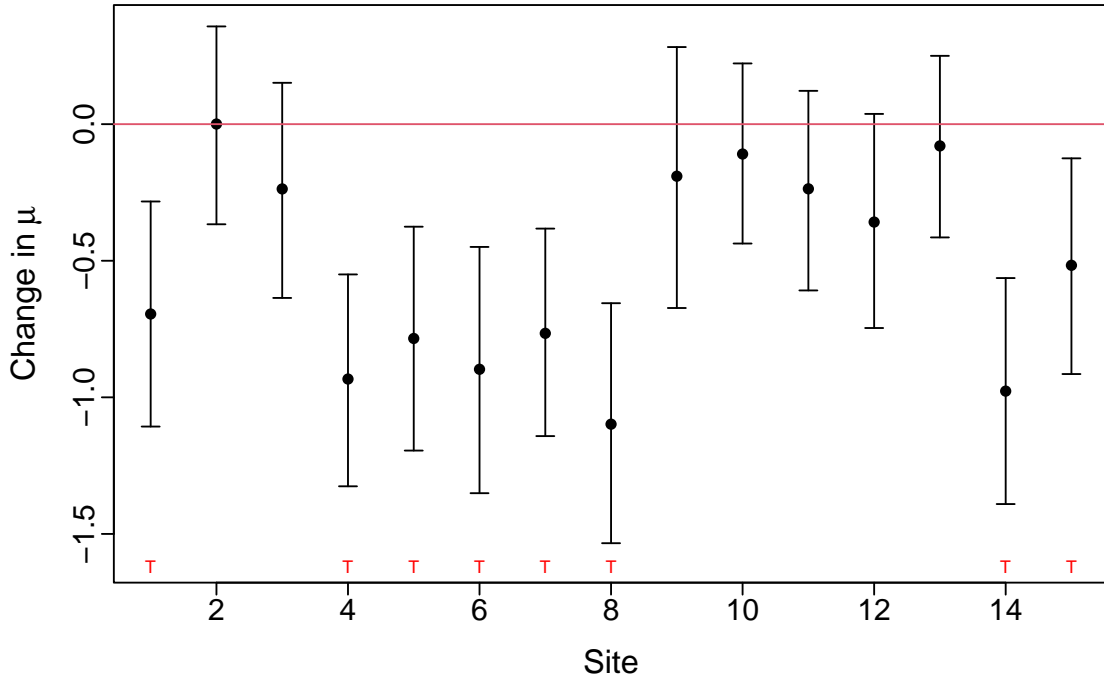


Figure 5.2: Posterior means and 95% CIs for the change in location parameter between the before and after period ( $\mu_1 - \mu_0$ ). Treated sites are denoted with a ‘T’ above the  $x$ -axis.

summaries for the treatment effect,  $\mu_1 - \mu_0$ , over all sites. The treatment effect is successfully captured in all the treated sites (the credible intervals (CI) are wholly negative), and the sites at which no treatment took place, the change in location parameter might be assumed zero (their CIs contain zero). Table 5.1 shows the posterior means and 95% CIs from the joint model and the spatial model. We see the means and CIs are very similar, however, the spatial model correctly identifies site 12 as untreated, by including zero in the CI. The posteriors for parameters  $\mu_0$ ,  $\mu_1$  and  $\eta$  are summarised through their mean and 95% CIs in Figure 5.3, alongside a density plot of the marginal posterior for  $\xi$  which is fixed across all sites. Figure 5.4 summarises the posteriors for parameters  $\mu_0$ ,  $\mu_1$  and  $\eta$  through their mean and 95% CIs however this time they’re plotted against the corresponding longitude and latitude of the site. This shows a potential decrease in both  $\mu_0$  and  $\mu_1$  from West–East and South–North, and an increase in scale parameter from South–North. A change in location parameter shifts the distribution along the  $x$ -axis and hence the North–East is potentially where we’d expect the least near-misses. The scale parameter in the GEV distribution controls the spread or variability of the distribution, therefore

Site	Treatment effect			
	Spatial model		Joint model	
	Mean	95% CI	Mean	95% CI
1	-0.6949	(-1.1069, -0.2831)	-0.7259	(-1.1217, -0.3268)
2	0.0004	(-0.3665, 0.3578)	-0.0426	(-0.4069, 0.3202)
3	-0.2371	(-0.6362, 0.1513)	-0.2820	(-0.6876, 0.1015)
4	-0.9329	(-1.3257, -0.5501)	-0.9775	(-1.3826, -0.5918)
5	-0.7845	(-1.1948, -0.3754)	-0.8399	(-1.2331, -0.4559)
6	-0.8973	(-1.3509, -0.4494)	-0.9718	(-1.4183, -0.5101)
7	-0.7657	(-1.1419, -0.3823)	-0.8035	(-1.1659, -0.4428)
8	-1.0982	(-1.5339, -0.6554)	-1.1319	(-1.5723, -0.7047)
9	-0.1906	(-0.6730, 0.2822)	-0.2425	(-0.7194, 0.2301)
10	-0.1093	(-0.4369, 0.2223)	-0.1433	(-0.4745, 0.1784)
11	-0.2366	(-0.6086, 0.1218)	-0.2681	(-0.6460, 0.0989)
12	-0.3584	(-0.7462, 0.0375)	-0.4099	(-0.8104, -0.0184)
13	-0.0796	(-0.4146, 0.2500)	-0.1046	(-0.4381, 0.2247)
14	-0.9772	(-1.3907, -0.5634)	-1.0495	(-1.4948, -0.6145)
15	-0.5167	(-0.9147, -0.1251)	-0.5478	(-0.9332, -0.1528)

Table 5.1: Treatment effect posterior mean and 95% credible intervals from the spatial model ( $\mu_1 - \mu_0$ ) and the joint model ( $\beta_1$ , as specified in Section 5.2).

	$\alpha$		$\lambda$	
	Mean	95% CI	Mean	95% CI
$\mu$	0.1276	(0.0123, 0.5133)	90.2594	(47.2135, 118.8835)
$\eta$	0.3174	(0.0250, 1.8452)	44.8333	(23.7952, 61.3935)

Table 5.2: Posterior means and 95% CIs for  $\alpha_\mu, \alpha_\eta, \lambda_\mu, \lambda_\eta$ .

we may expect to see a larger range of PETs at sites that are further North.

When using an exponential kernel in a Gaussian process to capture spatial dependence, the length scale parameter,  $\lambda$ , controls the spatial scale over which the function varies. A larger length scale means that the function can change more gradually over larger distances, whereas a smaller length scale means that the function can change more rapidly over shorter distances. The signal variance parameter,  $\alpha$ , controls the overall variability of the spatial function, with larger values indicating more variability and smaller values indicating less variability. Table 5.2 summarises the marginal posterior distributions of the hyperparameters in the GP. For the LPI data we have that the amplitude of the spatial function is determined by a relatively small value of  $\alpha$ , for both  $\mu$  and  $\eta$  indicating less variability. The spatial scale, over

which the function varies, is determined by  $1/\lambda$ . As such, the function can change rapidly over short distances. Numerical summaries for the marginal posteriors of all other model parameters are provided in Tables 5.4, 5.5 and 5.3.

When a GEV distribution is fitted to negated PETs at signalised intersections, two important quantitative measures of safety indices exist: the risk of crash (RC) and return levels. The risk of crash is calculated as,

$$RC_j = \Pr \{z_j \geq a_d\} = 1 - \mathcal{G}_j(a_d) = \begin{cases} 1 - \exp \left\{ - \left[ 1 - \xi \frac{\mu_{\cdot j}}{\sigma_j} \right]^{-\frac{1}{\xi}} \right\}, & \text{for } \xi \neq 0, \\ 1 - \exp \left\{ - \exp \left( \frac{\mu_{\cdot j}}{\sigma_j} \right) \right\}, & \text{for } \xi = 0. \end{cases}$$

If  $a_d = 0$ ,  $RC_j$  is the risk of crash at site  $j$ ,  $z_j$  is the maximum negated PET at site  $j$ , and  $\mathcal{G}_j(\cdot)$  is the fitted GEV distribution.  $RC$  is non-negative, where a  $RC$  value of zero means no risk of crash, while an  $RC$  value greater than zero indicates a risk of crash. Typically,  $a_d = 0$  which gives the probability that there are zero seconds between the pedestrian and vehicle and hence, of a crash. However we are interested in the probability of dangerous situations and severe near-misses, hence we use  $a_d = -0.5$ . This gives us the probability of a negated PET being more than  $-0.5$  seconds or the probability that a car and pedestrian will come within 0.5 seconds of each other at each site.

Using the posterior parameter estimates from the MCMC output, we draw estimates of  $RC_{tj}^{(i)}$  for each site using  $(\mu_{0j}^{(i)}, \mu_{1j}^{(i)}, \sigma_j^{(i)}, \xi^{(i)})$ , where  $i = 1, \dots, N$  and  $N$  is the number of posterior draws available from the MCMC scheme, in the before ( $t = 0$ ) and after ( $t = 1$ ) periods. The posterior distribution of  $RC_{tj}$  is summarised through the mean and 95% credible intervals at each site in Figure 5.5. For the return levels, we estimate the probability of extreme events occurring in the future. Estimates of an extreme quantile  $z_r$  can be obtained by inversion of  $\mathcal{G}(z_r)$ , where  $z_r$  is the  $r$ -observation return level associated with return period  $r$ , as outlined in Section 2.2.3. Specifically,

$$z_r = \begin{cases} \mu_{\cdot j} - \frac{\sigma_j}{\xi} \left[ 1 - (-\log(1 - r^{-1}))^{-\xi} \right], & \xi \neq 0 \\ \mu_{\cdot j} - \sigma_j \log(-\log(1 - r^{-1})), & \xi = 0. \end{cases}$$

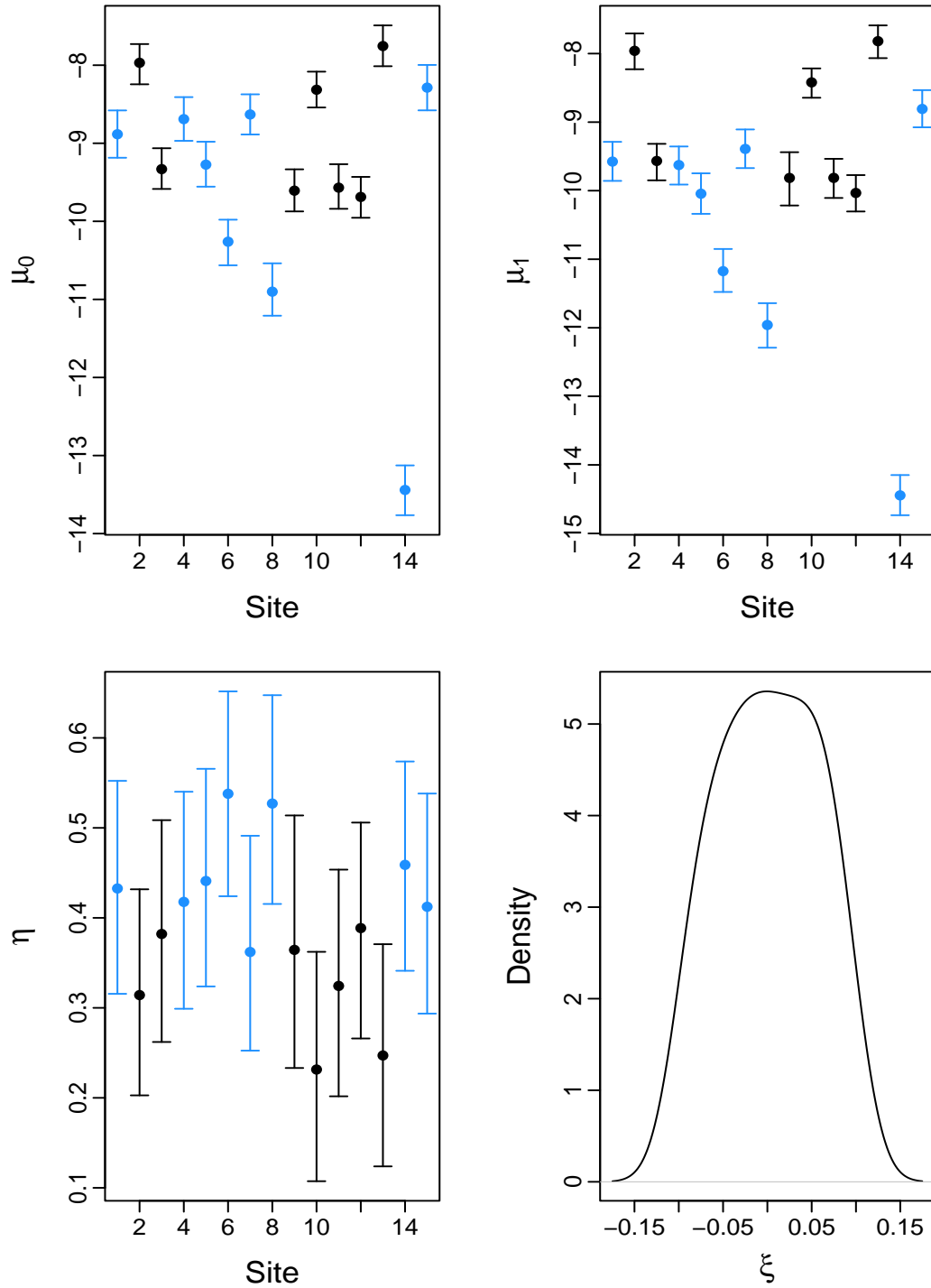


Figure 5.3: Posterior means and 95% CIs for parameters  $\mu_0, \mu_1, \eta$  for each site, with the posterior density of the shape parameter,  $\xi$ . Treated sites are shown in blue.

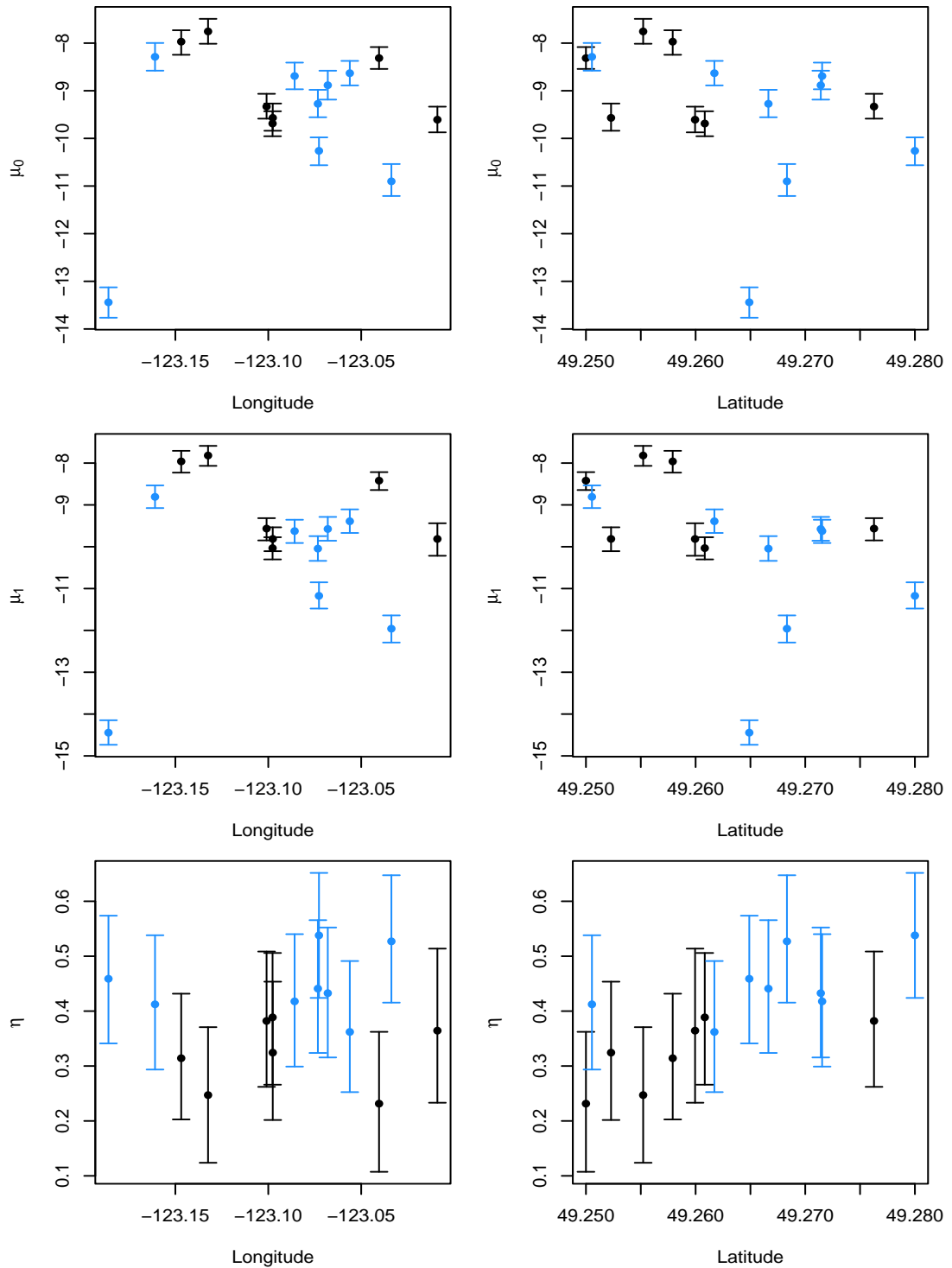


Figure 5.4: Posterior mean and 95% credible intervals for the GEV parameters ( $\mu_0$  (top),  $\mu_1$  (middle),  $\eta$  (bottom)) plotted against longitude (left) and latitude (right). Treated sites are shown in blue.



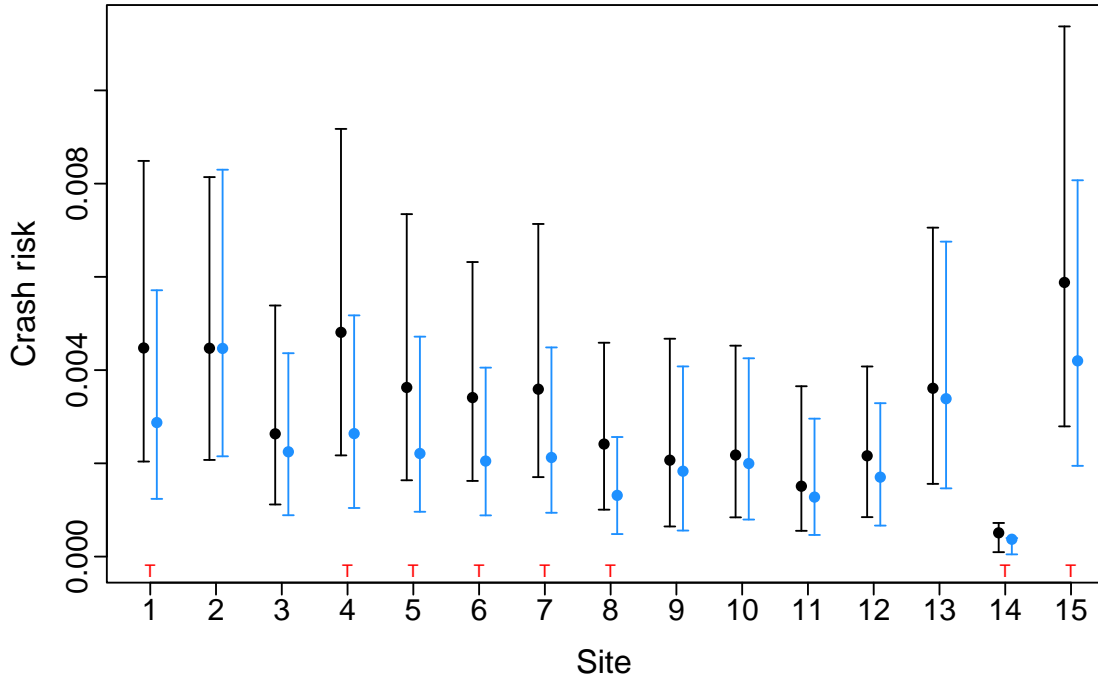


Figure 5.5: Posterior means and 95% credible intervals of  $RC$  at each site. The  $RC$  in the before period is shown in black, and the after period in blue. Treated sites are denoted with a ‘T’ above the  $x$ -axis.

Using the posterior parameter estimates, as outlined for the drawing of  $RC_{tj}$ , we find distributions for the one-year return levels for each site in the before and after periods. The means and 95% credible intervals of the distribution are shown in Figure 5.6. In both the return levels and the crash risk assessments, we see an improvement in safety in the after period for the treated sites. On average, the crash risk at the treated sites has decreased by 37%, and at the non-treated sites there has been a reduction of 11% on average. The mean one year return level has decreased by 0.83 seconds at the treated sites and by 0.17 seconds at the non-treated sites.

## 5.4 Discussion

We have developed a spatial model to capture the treatment effect at intersections treated with 5 second LPs. We use a block maxima approach as this is most common in current spatial extremes research. The model allows for spatial dependence and a varying location parameter between the before and after treatment time periods. The treatment effect is then estimated through comparison of the estimates

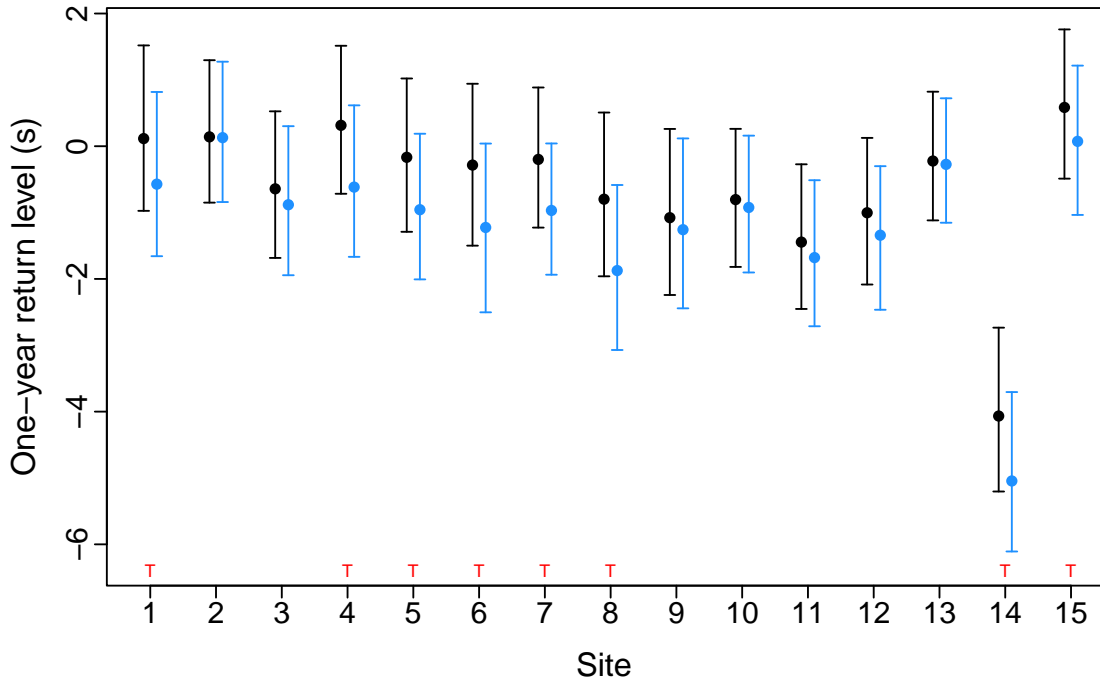


Figure 5.6: Posterior means and 95% credible intervals of the one-year return level (seconds) at each site. The return level in the before period is shown in black, and the after period in blue. Treated sites are denoted with a ‘T’ above the  $x$ -axis.

for the location parameter in the before period to those in the after period at each site. A Gaussian process is built in to the GEV location and scale parameters, using an exponential covariance function to describe the similarity between two points in the spatial domain. We fixed the shape parameter across all sites as it is estimated as approximately zero for all sites. The model can be fitted through Markov chain Monte Carlo, using Gibbs sampling where conjugacy allows and MH steps elsewhere.

We applied our approach to real data consisting of negated PETs at 15 intersections, of which 8 had been treated with the LPI safety intervention. There was clear spatial correlation shown between sites and accounting for this within the model allowed for pooling of information which is beneficial when extra covariates are not accessible. Although modelling parameter variation as a latent spatial process resulted in slightly wider credible intervals for the treatment effect for some sites compared to the joint model, it has better estimated the treatment effect in general. The untreated sites have means closer to zero and site 12 is correctly identified as untreated which contradicts what was found through the joint model, where the 95%

CI was wholly negative. This differentiation highlights the refined capability of the spatial model to discern untreated sites with means closer to zero. While both models offer valuable insights, the spatial model provides a more nuanced understanding of the underlying phenomena and represents a more precise approach for capturing the true treatment effect at specific sites. Another advantage of this model, though not pursued here, is that it can be used to estimate PETs at sites for which observations are not available. The GP allows interpolation between parameter values. Using the distances between the unobserved and the observed sites, we can construct the covariance matrix and then the conditional Gaussian for the parameters at the unobserved sites. The Gaussian conditional is then available to sample from at the unobserved sites given the parameter draws at the observed sites, see [Casson and Coles \(1999\)](#). The treatment was shown to have been advantageous as the location parameter became more negative in the after periods in all treated sites. Through return levels and risk of crash summaries, we again showed how the treatment made the intersections safer through reducing the probability of near-misses.

Our modelling approach can be improved in a number of ways. For example, the inclusion of covariate information would not only provide more insight into how these effect near-misses, but would also improve estimations and predictions. Other correlation functions could be considered and potentially found to be more beneficial ([Diggle et al., 1998](#)); more attention could be given to the specification of prior distributions; and better tuning could make the MCMC scheme more efficient.

	$\xi$	$\beta_3$	$\nu_\xi$
Mean	0.0017	-0.0087	1.0858
95% CI	(-0.0950, 0.0952)	(-2.8032, 2.3732)	(0.0707, 3.8865)

Table 5.3: Posterior means and 95% CIs for  $\xi, \beta_3$  and  $\nu_\xi$ .

Site	$\mu_0$		$\mu_1$		$\eta$	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
1	-8.8844	(-9.1857, -8.5793)	-9.5757	(-9.8571, -9.2876)	0.4327	(0.3156, 0.5522)
2	-7.9694	(-8.2448, -7.7295)	-7.9600	(-8.2287, -7.7067)	0.3142	(0.2028, 0.4317)
3	-9.3314	(-9.5851, -9.0639)	-9.5664	(-9.8491, -9.3167)	0.3821	(0.2621, 0.5086)
4	-8.6910	(-8.9690, -8.4090)	-9.6259	(-9.9108, -9.3552)	0.4177	(0.2990, 0.5401)
5	-9.2741	(-9.5562, -8.9805)	-10.0445	(-10.3378, -9.7468)	0.4410	(0.3238, 0.5656)
6	-10.2611	(-10.5633, -9.9790)	-11.1743	(-11.4775, -10.8505)	0.5379	(0.4240, 0.6517)
7	-8.6314	(-8.8878, -8.3732)	-9.3911	(-9.6705, -9.1068)	0.3621	(0.2525, 0.4912)
8	-10.9011	(-11.2090, -10.5380)	-11.9592	(-12.2907, -11.6413)	0.5270	(0.4154, 0.6474)
9	-9.6073	(-9.8730, -9.3340)	-9.8144	(-10.2163, -9.4389)	0.3644	(0.2332, 0.5138)
10	-8.3146	(-8.5418, -8.0828)	-8.4201	(-8.6442, -8.2166)	0.2315	(0.1074, 0.3622)
11	-9.5689	(-9.8394, -9.2682)	-9.8135	(-10.1061, -9.5353)	0.3244	(0.2017, 0.4536)
12	-9.6879	(-9.9544, -9.4307)	-10.0332	(-10.3034, -9.7728)	0.3886	(0.2660, 0.5059)
13	-7.7549	(-8.0135, -7.4904)	-7.8192	(-8.0664, -7.5890)	0.2471	(0.1238, 0.3707)
14	-13.4408	(-13.7645, -13.1265)	-14.4448	(-14.7344, -14.1483)	0.4589	(0.3412, 0.5738)
15	-8.2877	(-8.5784, -7.9970)	-8.8079	(-9.0756, -8.5334)	0.4123	(0.2936, 0.5381)

Table 5.4: Posterior means and 95% CIs for  $\mu_0$ ,  $\mu_1$  and  $\eta$ .

Site	$\beta_0$		$\beta_1$		$\beta_2$	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
1	-7.5217	(-8.8202, -4.5331)	-1.6373	(-3.1992, -0.6337)	0.3550	(-0.4885, 1.1920)
2	-6.6307	(-7.9185, -3.4488)	-0.9234	(-2.4593, 0.1177)	0.2368	(-0.6217, 1.0381)
3	-7.9759	(-9.2660, -4.7144)	-1.1818	(-2.7253, -0.1801)	0.3048	(-0.4756, 1.0752)
4	-7.3249	(-8.6183, -4.2132)	-1.8879	(-3.4460, -0.8402)	0.3357	(-0.4794, 1.1517)
5	-7.9070	(-9.1997, -4.7127)	-1.7228	(-3.2163, -0.6967)	0.3656	(-0.4454, 1.2069)
6	-8.9039	(-10.2233, -5.8630)	-1.8560	(-3.4739, -0.8198)	0.4549	(-0.4080, 1.3429)
7	-7.2753	(-8.5330, -4.1312)	-1.7019	(-3.2330, -0.6452)	0.2841	(-0.5866, 1.1032)
8	-9.5608	(-10.9008, -6.4892)	-1.9855	(-3.4685, -0.9293)	0.4542	(-0.3794, 1.2633)
9	-8.2911	(-9.5380, -5.1159)	-1.1189	(-2.7184, -0.1013)	0.2962	(-0.5749, 1.1600)
10	-6.9738	(-8.2271, -3.7587)	-1.0354	(-2.5959, -0.0402)	0.1531	(-0.6602, 0.9298)
11	-8.2073	(-9.5100, -5.2197)	-1.1946	(-2.7654, -0.1688)	0.2404	(-0.5902, 1.0392)
12	-8.3226	(-9.5987, -5.2641)	-1.2985	(-2.7959, -0.2785)	0.3096	(-0.4670, 1.0855)
13	-6.4074	(-7.6588, -3.2454)	-1.0038	(-2.4845, 0.0052)	0.1685	(-0.6968, 0.9547)
14	-12.1274	(-13.3797, -9.0578)	-1.9248	(-3.5600, -0.8918)	0.3898	(-0.5037, 1.1788)
15	-6.9574	(-8.2572, -3.6861)	-1.4494	(-2.9741, -0.4041)	0.3361	(-0.4937, 1.1264)

Table 5.5: Posterior means and 95% CIs for  $\beta_0$ ,  $\beta_1$  and  $\beta_2$ .

# Chapter 6

## Accounting for seasonal and zonal effects in collision data

### 6.1 Introduction

As discussed in Section 1.1, road safety practitioners strive to improve road safety and reduce the number of collisions on the road. They often do this by applying safety schemes to locations in which there has been a high number of collisions in some ‘before’ period and hope that the scheme will reduce the number of collisions in the ‘after’ period. Clearly choosing the sites in which the schemes are implemented is an important task. The sites must be chosen such that the scheme is the most effective, in that there is an expected measurable reduction in collisions after accounting for RTM and trend effects. We will refer to these sites as ‘hotspots’. Logically it would make sense for these sites to be chosen where there has been a large number of collisions and with that number following an upward trend. However, we do not want to have to wait to see this growth as people are potentially losing their lives. Therefore, we intend to be proactive over reactive by predicting the number of collisions at hotspots to inform site selection. Naturally, finding ways to predict hotspots is prominent in road safety research. [Lu et al. \(2015\)](#) use logistic regression analysis to form a prediction model for accident hotspots. [Fawcett et al. \(2017\)](#) use a Bayesian hierarchical model for predicting road safety hotspots, building upon the relatively simple EB methodology. [Deublein et al. \(2013\)](#) combine hierarchical multivariate Poisson-lognormal regression analysis and Bayesian Probabilistic Networks to predict road traffic collisions. Further studies in this area include [Abdel-Aty and](#)

Radwan (2000), El-Basyouny et al. (2014), Qu and Meng (2014), with a full literature review available in Yannis et al. (2017)

Poisson regression and negative binomial (NB) models have been utilised for years to analyse road traffic accident occurrence, however they come with limitations. The Poisson model requires the unlikely situation of the mean of the data being equivalent to the variance, then both the Poisson and NB models require the data to be uncorrelated in time. At their basic level, these regularly-used models are incapable of taking into account some unobserved heterogeneities due to spatial and temporal effects of crash data. Usually it is assumed that sites with the same attributes must have the same crash risk, and that crash counts at a site are temporally independent. However, in reality we notice that these assumptions rarely hold and some unobserved factors may exist between sites and subsequent crash counts. Hence, we must account for site-specific effects and temporal effects to avoid underestimating the uncertainty in the regression coefficients. It is common in road safety to have data which show seasonality, and in which sites differ in terms of their safety through factors that are not observed in the data. To allow us to capture such differences between seasons and sites we can use random effects models. Random effects models are an extension of the general linear model outlined in Section 2.1.

Random effects modelling is commonly used in road safety analyses to account for sources of variability that are not explicitly measured, such as weather conditions and driver behaviour. By estimating the variance components of multiple factors, practitioners can identify important predictors of road safety outcomes and provide valuable insights for improving road safety interventions and policies. Random effects modelling is a powerful tool for understanding the complex relationships between different factors that contribute to road safety, and can help decision-makers develop more effective strategies for reducing accidents and improving overall road safety. Random effects models can account for potential unobserved heterogeneity between sites and correlations between the residuals. They are also available for prediction. A random effects model is fitted to observed data with a set of predictor variables. Once the regression coefficients have been estimated, the model can be used to predict values of the response variable at each site in future time periods.

Random effect negative binomial (RENB) models have been shown to outperform regular negative binomial and Poisson models (Jiang et al., 2014, Shankar et al.,

1998). In what seems to be the first application of the RENB model to road safety, Shankar et al. (1998) indicated that the model is appropriate as geometric and traffic variables are likely to have site-specific effects. Huang and Chin (2010) use random effects to capture site-effects in their zero-inflated Poisson regression with site-specific random effects (REZIP). Chin and Quddus (2003) use a RENB model to identify the elements that affect intersection safety. We propose a Bayesian hierarchical random effects model to capture seasonal and zonal effects shown in the collision rate data provided by the Florida Department of Transport, outlined in Section 1.3.2. As we have rates, we avoid the need to model counts through Poisson or NB distributions and assume a normal distribution to describe the data. We then fit the model to UK count data, outlined in Section 1.3.4, and assume the data follow a Poisson distribution.

## 6.2 Florida data analysis

We have monthly collision data available from North Florida, USA within fixed Traffic Administration Zones (TAZ), outlined in Section 1.3.2. For each of the 49 zones there are multiple sites – the rate is the average number of collisions across those sites in each month within each TAZ. We propose a Bayesian hierarchical model which allows us to segregate the seasonal and zonal effects alongside capturing the uncertainty in the parameters sufficiently. Here we denote the rate of collisions per zone and season by  $y_{mj}$  where  $m$  denotes the season (month)  $m = 1, \dots, 12$  and  $j$  denotes the zone  $j = 1, \dots, 49$ . As to make the model more general, in this section we will denote  $n_z$  as the number of zones,  $n_s$  as the number of months and  $n_{mj}$  as the number of observations in month  $m$  in zone  $j$ .

Upon visual inspection of the raw data, it appears that the majority of the combinations of month and zone approximately follow normal distributions. This is to be expected as the rate for each zone was calculated as an average across multiple sites. Additionally, the rates across all sites are sufficiently larger than zero, with a mean of 7.8, and so the normal distributions are unlikely to pass through zero. We assume a model that has each observation  $y_{mj}^{(i)}$ , for  $i = 1, \dots, n_{mj}$  (denoting the number of observations in month  $m$  in zone  $j$ ), normally distributed with mean  $\mu_{mj}$  and precision  $\tau = 1/\sigma^2$ . The raw data over each zone and month combination has similar variances, therefore we assume constant  $\sigma^2$  and thus constant precision.

Throughout this chapter,  $N_0(A, B)$  denotes a normal distribution with mean  $A$  and precision  $B$  – precision chosen for notational convenience. The model takes the form,

$$\begin{aligned} y_{mj}^{(i)} | \mu_{mj}, \tau &\sim N_0(\mu_{mj}, \tau), \quad \text{with} \\ \mu_{mj} &= \mu + \epsilon^{(j)} + \gamma^{(m)} \quad \text{and} \\ \tau = 1/\sigma^2 &\sim \text{Ga}(\alpha, \beta), \end{aligned}$$

for  $i = 1, \dots, n_{mj}$ ,  $m = 1, \dots, n_s$  and  $j = 1, \dots, n_z$ . We let  $\alpha = 1, \beta = 0.01$  to have an uninformative prior on  $\tau$ . All effects for  $\mu_{mj}$  are assumed to be normally and independently distributed:

$$\begin{aligned} \mu &\sim N_0(f, g), \\ \epsilon^{(j)} &\sim N_0(0, \nu_\epsilon), \\ \gamma^{(m)} &\sim N_0(0, \nu_\gamma), \end{aligned} \tag{6.1}$$

where  $f = 0$  and  $g = 0.01$ . The overall mean  $\mu$  is separate to the random effects to capture any similarities between the zones and months, leaving  $\epsilon^{(j)}$  and  $\gamma^{(m)}$  to capture the variability. We set the random effects priors to be normal distributions with zero mean and precisions  $\nu_\epsilon$  and  $\nu_\gamma$ , respectively. We choose non-informative (large variance) priors for the hyperparameters, giving

$$\begin{aligned} \nu_\epsilon &\sim \text{Ga}(a, b), \\ \nu_\gamma &\sim \text{Ga}(c, d), \end{aligned}$$

where  $a = 1, b = 0.01, c = 1$  and  $d = 0.01$ .

### 6.2.1 Full posterior

The model is complex with a large number of parameters. Hence, calculating the full posterior is problematic. Using Bayes' theorem, we have

$$\begin{aligned} \pi(\epsilon^{(1:n_z)}, \gamma^{(1:n_s)}, \mu, \tau, \nu_\epsilon, \nu_\gamma | y) &\propto \\ \pi(\mu) \pi(\tau) \pi(\nu_\epsilon) \pi(\nu_\gamma) \pi(\epsilon^{(1:n_z)} | \nu_\epsilon) &\pi(\gamma^{(1:n_s)} | \nu_\gamma) \pi(y | \mu, \epsilon^{(1:n_z)}, \gamma^{(1:n_s)}, \tau), \end{aligned} \tag{6.2}$$



where  $y = \{y_{mj}^{(i)}\}, i = 1, \dots, n_{mj}$  and

$$\begin{aligned}\pi(\epsilon^{(1:n_z)}|\nu_\epsilon) &= \prod_{j=1}^{n_z} N_0(\epsilon^{(j)}; 0, \nu_\epsilon), \\ \pi(\gamma^{(1:n_s)}|\nu_\gamma) &= \prod_{m=1}^{n_s} N_0(\gamma^{(m)}; 0, \nu_\gamma), \\ \pi(y|\mu, \epsilon^{(1:n_z)}, \gamma^{(1:n_s)}, \tau) &= \prod_{m=1}^{n_s} \prod_{j=1}^{n_z} \prod_{i=1}^{n_{mj}} N_0(y_{mj}^{(i)}; \mu + \epsilon^{(j)} + \gamma^{(m)}, \tau).\end{aligned}$$

However, Equation (6.2) is intractable; therefore, we will use an MCMC scheme to draw posterior realisations of the parameters. As we have semi-conjugate priors for the parameters governing the random effects, and full conjugacy elsewhere, we are able to use Gibbs sampling to perform inference. We alternate between draws of

1.  $\epsilon^{(j)}|\cdot \sim \pi(\epsilon^{(j)}|y_{\cdot j}, \gamma^{(1:n_s)}, \nu_\epsilon), \quad j = 1, \dots, n_z,$
2.  $\gamma^{(m)}|\cdot \sim \pi(\gamma^{(m)}|y_{m\cdot}, \epsilon^{(1:n_z)}, \nu_\gamma), \quad m = 1, \dots, n_s,$
3.  $\mu|\cdot \sim \pi(\mu|y, \epsilon^{(1:n_z)}, \gamma^{(1:n_s)}, \tau),$
4.  $\tau|\cdot \sim \pi(\tau|y, \mu, \epsilon^{(1:n_z)}, \gamma^{(1:n_s)}),$
5.  $\nu_\epsilon|\cdot \sim \pi(\nu_\epsilon|\epsilon^{(1:n_z)}),$
6.  $\nu_\gamma|\cdot \sim \pi(\nu_\gamma|\gamma^{(1:n_s)}).$

Where, for example,  $y_{\cdot j} = \{y_{mj}^{(i)}\}, m = 1, \dots, 12, i = 1, \dots, n_{mj}.$

### 6.2.2 Full conditional distributions

Having  $y_{mj}$  follow a normal distribution permits Gibbsian updates of all unknowns. The full conditional distributions (FCDs) are available for all parameters to sample from directly. Consider seasonal effect  $\gamma^{(m)}$  for  $\mu_{mj}$  with distribution given in Equation (6.1), then the density of each month effect  $\gamma^{(m)}, m = 1, \dots, n_s$  is given by

$$\pi(\gamma^{(m)}|\nu_\gamma) = \sqrt{\frac{\nu_\gamma}{2\pi}} \exp\left\{-\frac{\nu_\gamma}{2} (\gamma^{(m)})^2\right\}.$$

Thus, the joint density is

$$\pi(\gamma^{(1:n_s)}|\nu_\gamma) = \left(\frac{\nu_\gamma}{2\pi}\right)^{\frac{n_s}{2}} \exp\left\{-\frac{\nu_\gamma}{2} \sum_{m=1}^{n_s} (\gamma^{(m)})^2\right\}.$$

The prior is

$$\pi(\nu_\gamma) \propto \nu_\gamma^{c-1} \exp\{-d\nu_\gamma\}.$$

We can now write the posterior distribution as

$$\pi(\nu_\gamma|\gamma^{(1:n_s)}) \propto \nu_\gamma^{(c+\frac{n_s}{2})-1} \exp\left\{-\nu_\gamma \left[d + \frac{1}{2} \sum_{m=1}^{n_s} (\gamma^{(m)})^2\right]\right\},$$

and then, the FCD for  $\nu_\gamma$  is

$$\nu_\gamma|\cdot \sim Ga\left(c + \frac{n_s}{2}, d + \frac{1}{2} \sum_{m=1}^{n_s} (\gamma^{(m)})^2\right).$$

By similar arguments, the FCD for  $\nu_\epsilon$  is

$$\nu_\epsilon|\cdot \sim Ga\left(a + \frac{n_z}{2}, b + \frac{1}{2} \sum_{j=1}^{n_z} (\epsilon^{(j)})^2\right).$$

The full conditional for  $\tau$  is available for Gibbs sampling. Each  $y_{mj}^{(i)}$  has density given by

$$\pi\left(y_{mj}^{(i)}|\tau, \epsilon^{(j)}, \gamma^{(m)}, \mu\right) = \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2} \left(y_{mj}^{(i)} - (\mu + \epsilon^{(j)} + \gamma^{(m)})\right)^2\right\}. \quad (6.3)$$

Thus, the likelihood is

$$\pi(y|\epsilon^{(j)}, \gamma^{(m)}, \mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\tau}{2} \sum_{j=1}^{n_z} \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} \left(y_{mj}^{(i)} - (\mu + \epsilon^{(j)} + \gamma^{(m)})\right)^2\right\},$$

where  $N = \sum_{j=1}^{n_z} \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} 1$ . The prior density is

$$\pi(\tau) \propto \tau^{\alpha-1} \exp\{-\beta\tau\}, \quad \alpha, \beta > 0.$$

Therefore the full conditional distribution (FCD) for  $\tau$  is

$$\tau | \epsilon^{(j)}, \gamma^{(m)}, \mu, y \sim Ga \left( \alpha + \frac{N}{2}, \beta + \frac{1}{2} \sum_{j=1}^{n_z} \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - (\mu + \epsilon^{(j)} + \gamma^{(m)}) \right)^2 \right).$$

We then have the prior for  $\gamma^{(m)}$  as

$$\pi(\gamma^{(m)}) = \sqrt{\frac{\nu_\gamma}{2\pi}} \exp \left\{ -\frac{\nu_\gamma}{2} (\gamma^{(m)})^2 \right\}.$$

Using Equation (6.3), the full conditional density for  $\gamma^{(m)}$  is

$$\begin{aligned} \pi(\gamma^{(m)} | \cdot) &\propto \exp \left\{ -\frac{\tau}{2} \sum_{j=1}^{n_z} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - (\mu + \epsilon^{(j)} + \gamma^{(m)}) \right)^2 - \frac{\nu_\gamma}{2} (\gamma^{(m)})^2 \right\} \\ &\propto \exp \left\{ -\frac{\nu_\gamma + n_m \tau}{2} \left[ \gamma^{(m)} - \frac{\tau \sum_{j=1}^{n_z} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - \epsilon^{(j)} - \mu \right)}{\nu_\gamma + n_m \tau} \right]^2 \right\}, \end{aligned}$$

where  $n_m = \sum_{j=1}^{n_z} n_{mj}$ . Hence, the FCD for  $\gamma^{(m)}$  is

$$\gamma^{(m)} | \cdot \sim N \left( \frac{\tau \sum_{j=1}^{n_z} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - \epsilon^{(j)} - \mu \right)}{\nu_\gamma + n_m \tau}, \nu_\gamma + n_m \tau \right).$$

Similarly, the FCD for  $\epsilon^{(j)}$  is

$$\epsilon^{(j)} | \cdot \sim N_0 \left( \frac{\tau \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - \gamma^{(m)} - \mu \right)}{\nu_\epsilon + n_j \tau}, \nu_\epsilon + n_j \tau \right),$$

where  $n_j = \sum_{m=1}^{n_s} n_{mj}$ . To find the FCD for  $\mu$  we have, again, Equation (6.3) with prior distribution

$$\pi(\mu) = \sqrt{\frac{g}{2\pi}} \exp \left\{ -\frac{g}{2} (\mu - f)^2 \right\}.$$

Following similar arguments as above, we find the FCD for  $\mu$  to be

$$\mu | \cdot \sim N_0 \left( \frac{fg + \tau \sum_{j=1}^{n_z} \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} \left\{ y_{mj}^{(i)} - (\gamma^{(m)} + \epsilon^{(j)}) \right\}}{g + N\tau}, g + N\tau \right).$$

### 6.2.3 Application

The model is applied to the Florida collision rate data defined in Section 1.3.2. After a couple of initial runs, we simplified the model to have the seasonal effect as fixed, the argument being that once we had accounted for zonal effects, we would assume that the accident rate would not change by more than  $\pm 2$ . Therefore, we set  $\nu_\gamma = 1$  and hence  $\gamma^{(m)} \sim N(0, 1)$ . This removed a step from the sampler making it run faster and improved mixing. The Gibbs sampler draws realisations for each parameter from their FCD over each iteration. We run the Gibbs sampler for 300k iterations and thinned by 300 which gave reasonable mixing and sufficiently low autocorrelation. Our results are summarised by Figures 6.1, 6.2, 6.3. There is a clear sinusoidal pattern for the rate of collisions seasonally shown in Figure 6.2, suggesting that we'd expect a higher collision rate over the autumn months and a lower rate over spring time. This is expected to be down to weather patterns; Florida's wet season is from May to October and hurricanes pose a regular threat from June to November. There also seems to be a slight increase in rate from West to East shown when the zonal effect is plotted against longitude in Figure 6.3.

For validation we predict within the sample for month  $m$  and zone  $j$  to check that the observed data are consistent with the predictive distributions. We have posterior samples of parameters, that is draws of  $\psi = (\mu, \epsilon^{(j)}, \gamma^{(m)}, \tau)$  from  $\pi(\psi|y)$ . The within-sample predictive density at month  $m$  and site  $j$  is given by

$$\pi(\tilde{y}_{mj}|y) = \int_{\psi} \pi(\tilde{y}_{mj}|\psi) \pi(\psi|y) d\psi.$$

Although the within-sample predictive density is intractable, we can get draws from  $\pi(\tilde{y}_{mj}|y)$  via Monte Carlo:

For  $k = 1, \dots, n_{iters}$ :

1. Draw

$$\tilde{y}_{mj}^{(k)} \sim N\left(\left\{\mu_{(k)} + \epsilon_{(k)}^{(j)} + \gamma_{(k)}^{(m)}\right\}, \tau_{(k)}\right), \quad (6.4)$$

where  $\mu_{(k)}$ ,  $\epsilon_{(k)}^{(j)}$ ,  $\gamma_{(k)}^{(m)}$  and  $\tau_{(k)}$  are the  $k^{th}$  posterior draws of  $\mu$ ,  $\epsilon^{(j)}$ ,  $\gamma^{(m)}$  and  $\tau$ ; the sampled values can then be used to obtain histograms or summaries, as desired.

2. Overlay observed data points.

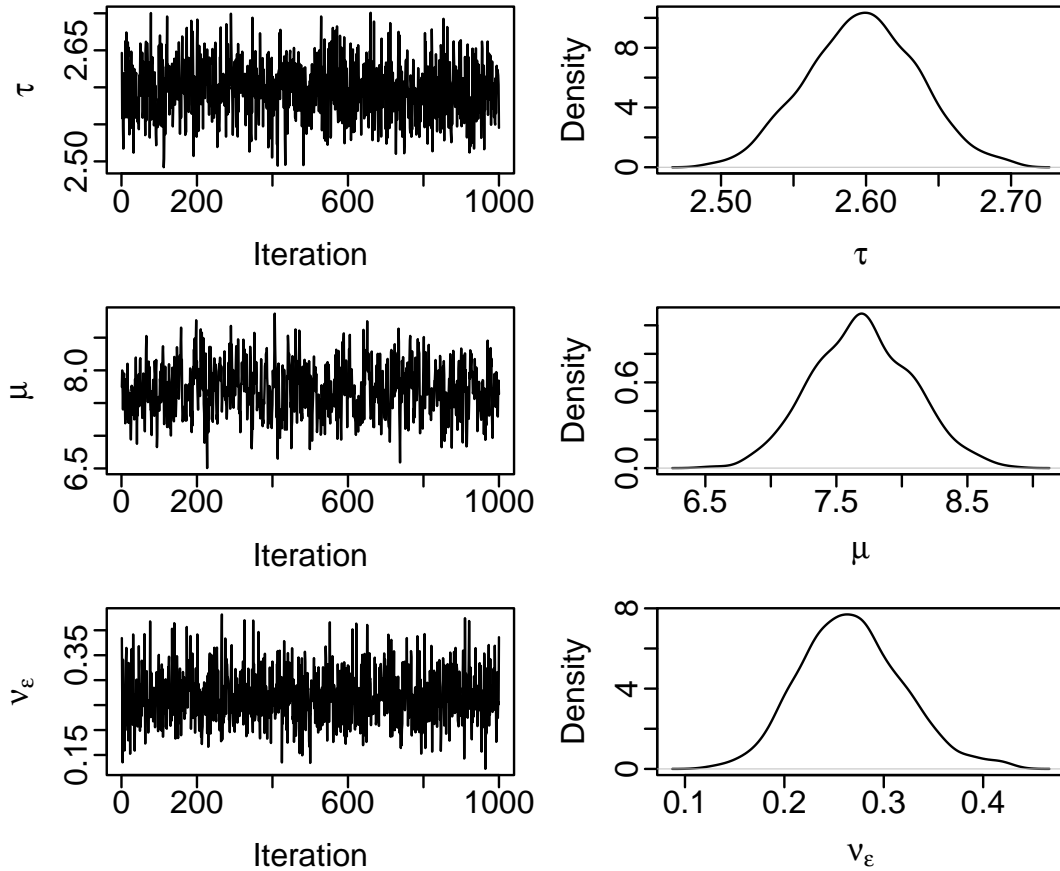


Figure 6.1: Trace plots and density plots for  $\tau$ ,  $\mu$  and  $\nu_\epsilon$  from 300k iterations of the Gibbs sampler and a thin of 300.

Results of the within-sample predictions are shown in Figure 6.4. The marginal posterior densities include all observed values, showing the model fits well to the data. Note that the model discussed thus far ignores any temporal dependence exhibited by the data, and without additional covariate information, out of sample forecasts are likely to be poor.

#### 6.2.4 Comparison to a fixed effects model

To justify the use of random effects to capture the seasonal and zonal effects, we compare this to a ‘simpler’ fixed effect model where we have data in month  $m$  and

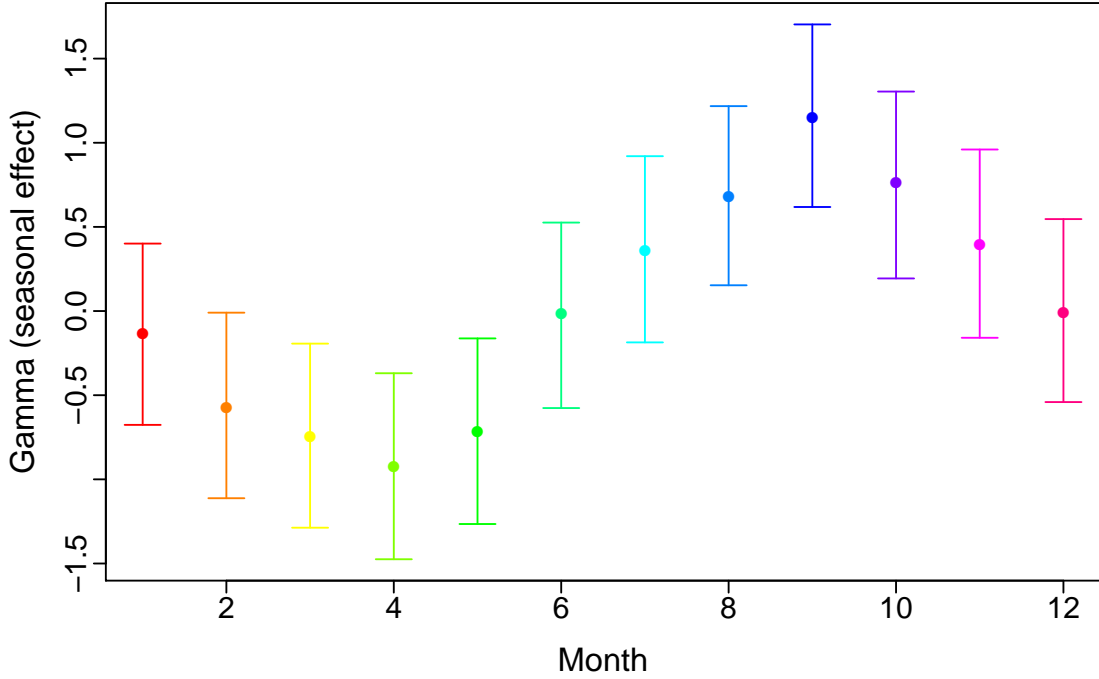


Figure 6.2: Posterior mean and 95% credible intervals for seasonal effects ( $\gamma^{(m)}$ ) from the analysis of the Florida data.

zone  $j$  following independent normal distributions, with mean  $\mu_{mj}$  and precision  $\tau$ ,

$$\begin{aligned} y_{mj} | \mu_{mj}^*, \tau &\sim N_0(\mu_{mj}^*, \tau), \\ \mu_{mj}^* &\overset{\text{indep}}{\sim} N_0(a, b), \\ \tau &\sim Ga(c, d). \end{aligned}$$

As we have conjugate priors, we are able to draw realisations via a Gibbs sampler using FCDs for  $\mu_{mj}^*$  and  $\tau$ . The FCDs are calculated as

$$\begin{aligned} \mu_{mj}^* | \cdot &\sim N \left( \frac{ba + \tau \sum_{i=1}^{n_{mj}} y_{mj}^{(i)}}{b + n_{mj}\tau}, b + n_{mj}\tau \right), \\ \tau | \cdot &\sim Ga \left( c + \frac{N}{2}, d + \frac{1}{2} \sum_{j=1}^{n_z} \sum_{m=1}^{n_s} \sum_{i=1}^{n_{mj}} \left( y_{mj}^{(i)} - \mu_{mj}^* \right)^2 \right). \end{aligned}$$

Therefore, the MCMC scheme draws realisations from the FCDs for each iteration

1.  $\mu_{mj}^* | \cdot \sim \pi(\mu | y, \tau)$  for  $j = 1, \dots, n_z, m = 1, \dots, n_s$ ,

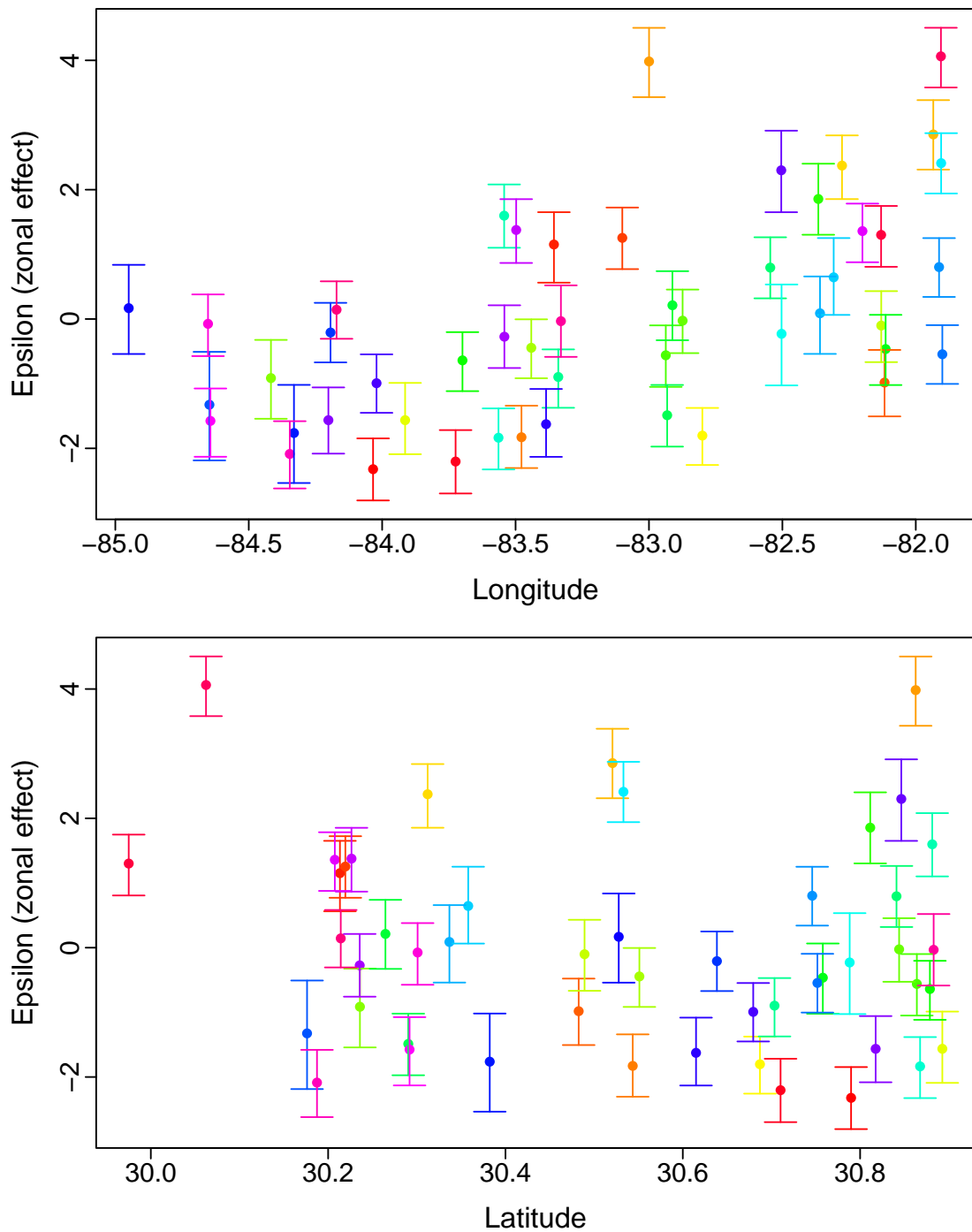


Figure 6.3: Posterior mean and 95% credible intervals for zonal effects ( $\epsilon^{(j)}$ ) against the longitude (top) and latitude (bottom) of the corresponding Florida TAZ centroid.

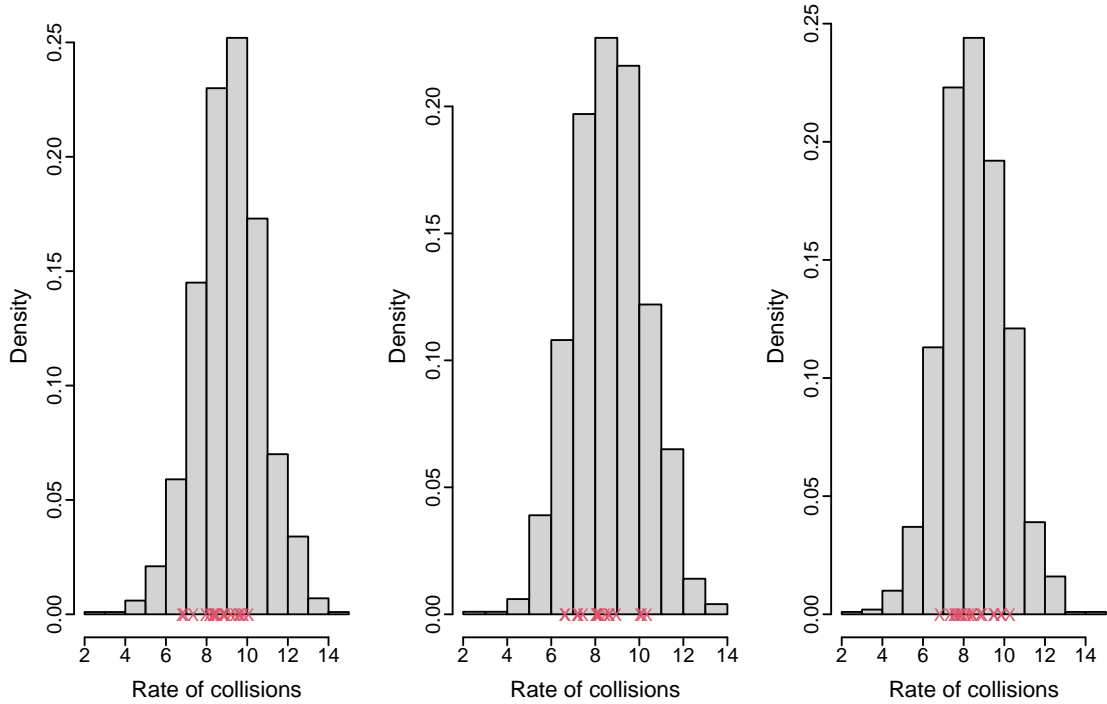


Figure 6.4: Within-sample predictions: Histogram of predictive posterior draws of  $\tilde{y}_{mj}$  for  $j = 48$  and  $m = 1, 2, 3$  with true observations overlaid on the  $x$ -axis.

$$2. \tau | \cdot \sim \pi(\tau | y, \mu_{mj}^*),$$

where,  $y = \{y_{mj}^{(i)}\}, i = 1, \dots, n_{mj}$ . As in Section 6.2, we use uninformative priors for all parameters such that  $a = 0, b = 0.01, c = 0.1, d = 0.01$ . As the model has fewer parameters than the random effects model, and all can be drawn using Gibbs sampling, the MCMC can be run for fewer iterations to gain convergence and good mixing. Therefore, the MCMC is run for 30k iterations with a burn-in of 20k. The results were then compared to the random effects model. The mean values for  $\tau$  from the two models were within 0.01 of each other, agreeing on a posterior mean value of approximately 0.39. Figure 6.5 shows that the posterior mean estimates of each  $\mu_{mj}$  and  $\mu_{mj}^*$  coincide with them being linear and lying on or close to the overlaid line of  $y = x$ . However, the advantages of modelling using random effects is evident when we see the improvement in precision in all  $\mu_{mj}$  over  $\mu_{mj}^*$  from the fixed effect model. The standard deviations are on average three times larger in the fixed effect model. Furthermore, Figure 6.6 shows posterior mean estimates of the within-sample predictions ( $\tilde{y}_{mj}$ ) drawn from Equation (6.4) for the random effects, against those drawn from the fixed effects model. We see both models estimate the value of  $\tilde{y}_{mj}$  similarly, however the majority of the standard deviations from the



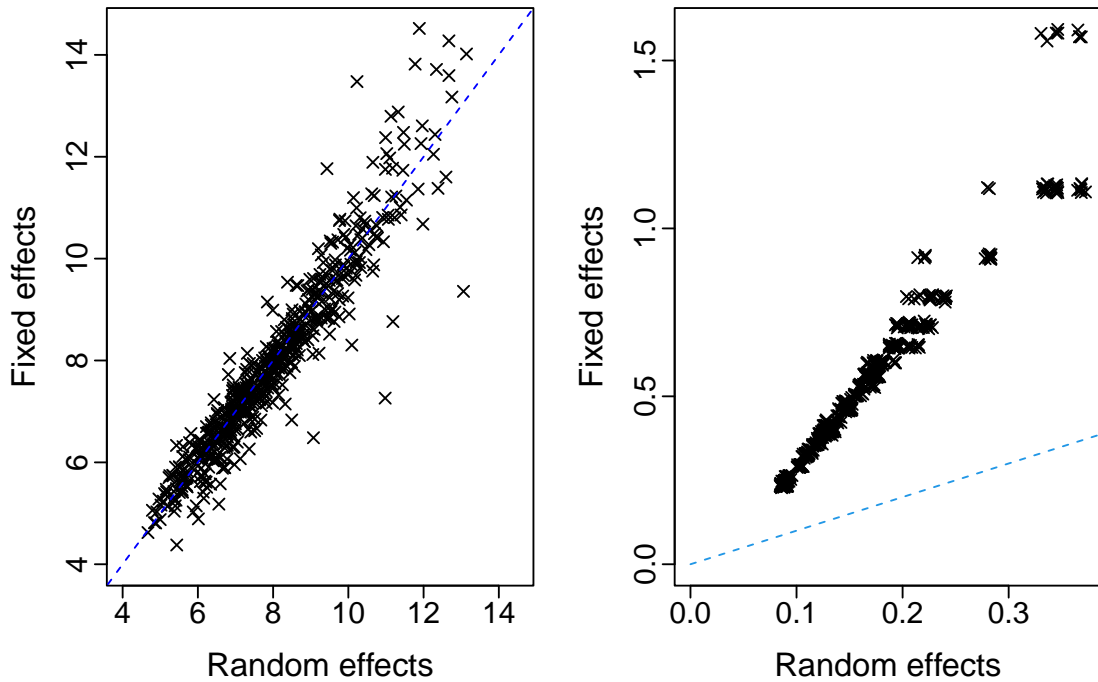


Figure 6.5: Left: Posterior mean values of  $\mu_{mj}^*$  from the fixed effects analysis, against  $\mu_{mj}$  from the random effects. Right: Posterior standard deviations of  $\mu_{mj}^*$  from a fixed effects model against  $\mu_{mj}$  from a random effects model. Both include the line  $y = x$  overlaid in blue.

fixed effect model are larger than their random effects equivalents.

### 6.3 UK data analysis

For further impact, we wish to be able to share our research with local practitioners that we currently work with, e.g. Highways England. Therefore, we fit a similar random effects model to STATS19 data for the UK, outlined in Section 1.3.4. Using a SQL server we collected 11 years of collision data (2009–2019 inclusive) over the UK and aggregated it to give the number of collisions per zones. The 207 zones are grouped by their local authority highway ID from the STATS19 data and we have the number of collisions for each month in each year.

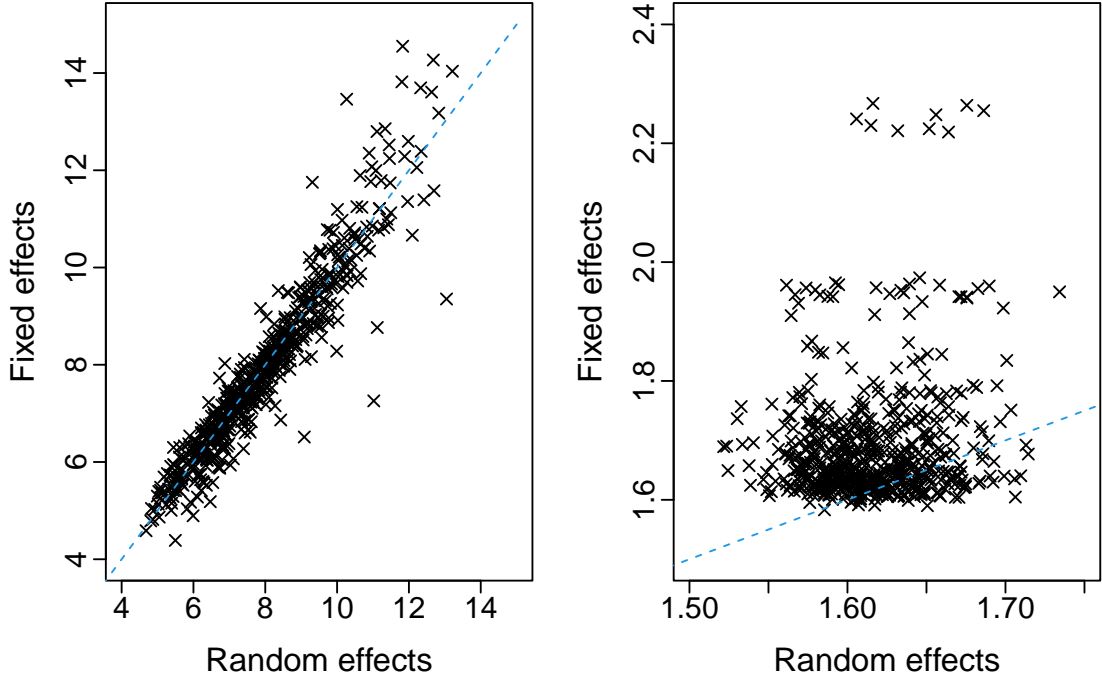


Figure 6.6: Left: Posterior mean values of within-sample predictions  $\tilde{y}_{mj}$  from the fixed effects analysis, against those from the random effects. Right: Posterior standard deviations of within-sample predictions  $\tilde{y}_{mj}$  from a fixed effects model against random effects. Both include the line  $y = x$  overlaid in blue.

### 6.3.1 Model description

Here, as this is count data, the number of collisions is assumed to follow a Poisson distribution

$$y_{mj} | \lambda_{mj} \sim Po(\lambda_{mj}).$$

We define the rate of the Poisson distribution as

$$\lambda_{mj} = \exp\{\epsilon_{\lambda}^{(j)} + \gamma_{\lambda}^{(m)}\},$$

to respect the positivity of  $\lambda$ . Similarly to Section 6.2, all effects for  $\lambda_{mj}$  are assumed normal and independently distributed,

$$\begin{aligned} \epsilon_{\lambda}^{(j)} &\sim N_0(\mu_{\epsilon_{\lambda}}, \nu_{\epsilon_{\lambda}}), \\ \gamma_{\lambda}^{(m)} &\sim N_0(0, \nu_{\gamma_{\lambda}}), \end{aligned}$$

with normal distributions parameterised in terms of mean and precision. We choose non-informative hyperparameters giving,

$$\begin{aligned}\mu_{\epsilon_\lambda} &\sim N_0(d^*, e^*), \\ \nu_{\epsilon_\lambda} &\sim Ga(f^*, g^*), \\ \nu_{\gamma_\lambda} &\sim Ga(a^*, b^*),\end{aligned}$$

with  $e^*, f^*, g^*, a^*, b^* = 0.01$  and  $d^* = 0$ . Furthermore, in the model, the data has had to be reduced as to avoid the likelihood becoming too large, and hence we have used  $y_{mj}/10$  to estimate  $\epsilon_\lambda$  and  $\gamma_\lambda$ .

As for the Florida analysis, the full posterior is intractable. Hence, we will use an MCMC scheme to perform inference. We alternate between draws of

1.  $\epsilon_\lambda^{(j)} | \cdot \sim \pi \left( \epsilon_\lambda^{(j)} | y_{\cdot j}, \gamma_\lambda^{(1:n_s)}, \mu_{\epsilon_\lambda}, \nu_{\epsilon_\lambda} \right), \quad j = 1, \dots, n_{uk},$
2.  $\gamma_\lambda^{(m)} | \cdot \sim \pi \left( \gamma_\lambda^{(m)} | y_{m\cdot}, \epsilon_\lambda^{(1:n_{uk})}, \nu_{\gamma_\lambda} \right), \quad m = 1, \dots, n_s,$
3.  $\mu_{\epsilon_\lambda} | \cdot \sim \pi \left( \mu_{\epsilon_\lambda} | \epsilon_\lambda^{(1:n_{uk})} \right),$
4.  $\nu_{\epsilon_\lambda} | \cdot \sim \pi \left( \nu_{\epsilon_\lambda} | \epsilon_\lambda^{(1:n_{uk})} \right),$
5.  $\nu_{\gamma_\lambda} | \cdot \sim \pi \left( \nu_{\gamma_\lambda} | \gamma_\lambda^{(1:n_s)} \right).$

Where, for example,  $y_{\cdot j} = \{y_{mj}^{(i)}\}$ ,  $m = 1, \dots, n_s$ ,  $i = 1, \dots, n_{mj}$ . Steps 1. and 2. are MH steps using normal random walks and 3.–5. are Gibbs steps. The same techniques are used to find the FCDs for  $\mu_{\epsilon_\lambda}, \nu_{\epsilon_\lambda}$  and  $\nu_{\gamma_\lambda}$  as those outlined in Section 6.2.2. Hence, steps 3.–5. are performed by drawing from

$$\begin{aligned}\mu_{\epsilon_\lambda} | \cdot &\sim N \left( \frac{d^* e^* + \nu_{\epsilon_\lambda} \sum_{m=1}^{n_s} \sum_{j=1}^{n_{uk}} y_{mj}}{e^* + n_s \nu_{\epsilon_\lambda}}, e^* + n_s \nu_{\epsilon_\lambda} \right), \\ \nu_{\epsilon_\lambda} | \cdot &\sim Ga \left( f^* + \frac{n_{uk}}{2}, g^* + \frac{1}{2} \sum_{j=1}^{n_{uk}} \left( \epsilon_\lambda^{(j)} - \mu_{\epsilon_\lambda} \right)^2 \right), \\ \nu_{\gamma_\lambda} | \cdot &\sim Ga \left( a^* + \frac{n_s}{2}, b^* + \frac{1}{2} \sum_{m=1}^{n_s} \left( \gamma_\lambda^{(m)} \right)^2 \right),\end{aligned}$$

recursively, for each iteration.

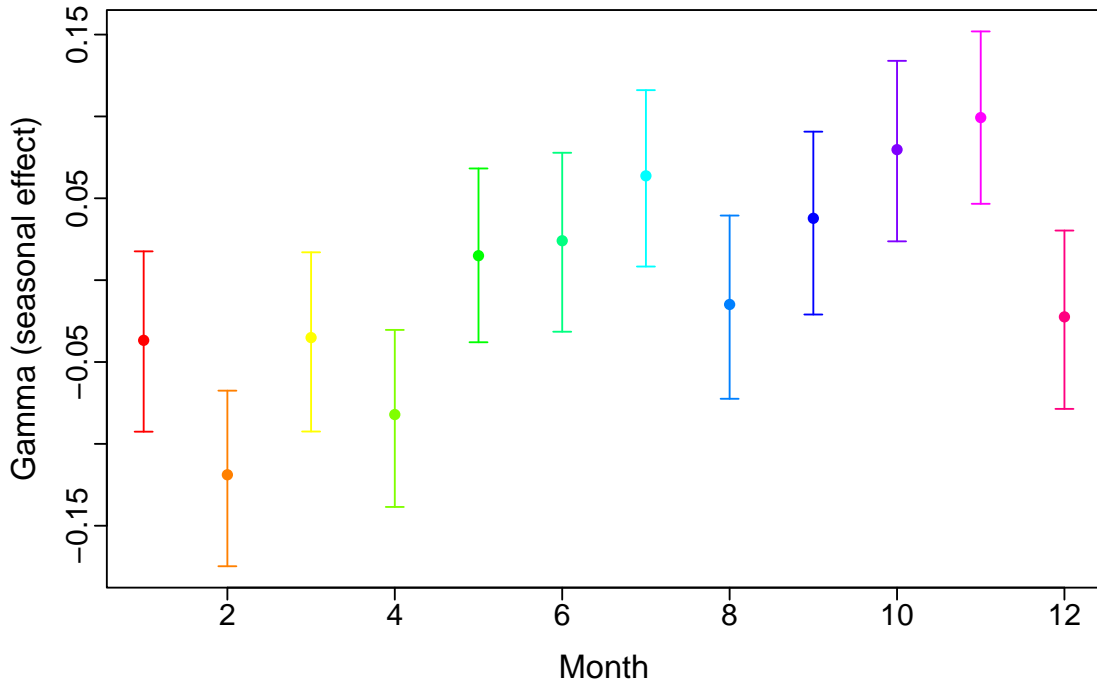


Figure 6.7: Posterior mean and 95% credible intervals for seasonal effects from the UK data,  $\gamma_{\lambda}^{(1:n_s)}$ , from a run of 500k, the first 400k iterations discarded as burn-in and a thin of 100.

### 6.3.2 Application

We run the MCMC scheme for 500k iterations with a thin of 400 after the first 100k iterations were discarded as burn-in. From visual checks of the trace plots, the chains had converged, there was good mixing and sufficiently low auto-correlation. The MCMC scheme draws realisations from the FCDs for  $\mu_{\epsilon_{\lambda}}$ ,  $\nu_{\epsilon_{\lambda}}$  and  $\nu_{\gamma_{\lambda}}$ ; our posterior estimates for  $\epsilon_{\lambda}$  and  $\gamma_{\lambda}$  are drawn using proposal distributions which are tuned to give reasonable acceptance probabilities. Figure 6.7 shows estimated seasonal effects, from this we would expect fewer collisions in February–April than over the summer months May–July, with peaks in October–November. Figure 6.8 shows how  $\epsilon_{\lambda}$  varies with longitude and latitude, respectively. There seems to be an increase in the estimated number of collisions across from West to East. There also seems to be a decrease in the number of collisions moving from South to North. Results of the within-sample predictions are shown in Figure 6.9 for site 2 in January, February and March. The marginal posterior densities include all observed values, showing the model fits well to the data.

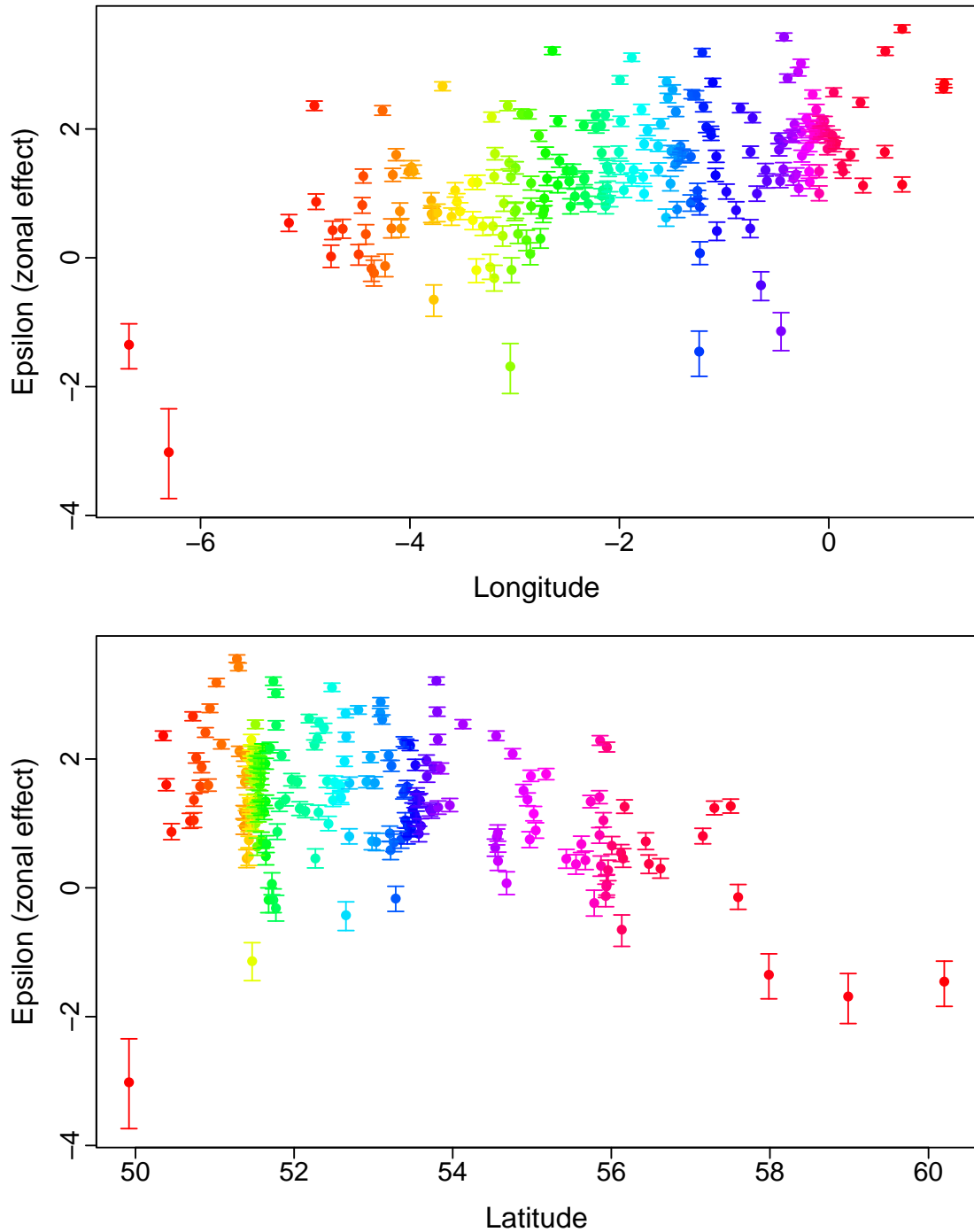


Figure 6.8: Posterior mean and 95% credible intervals for zonal effects,  $\epsilon_{\lambda}^{(1:n_{uk})}$ , plotted against the longitude (top) and latitude (bottom) of the corresponding UK zone centroid.

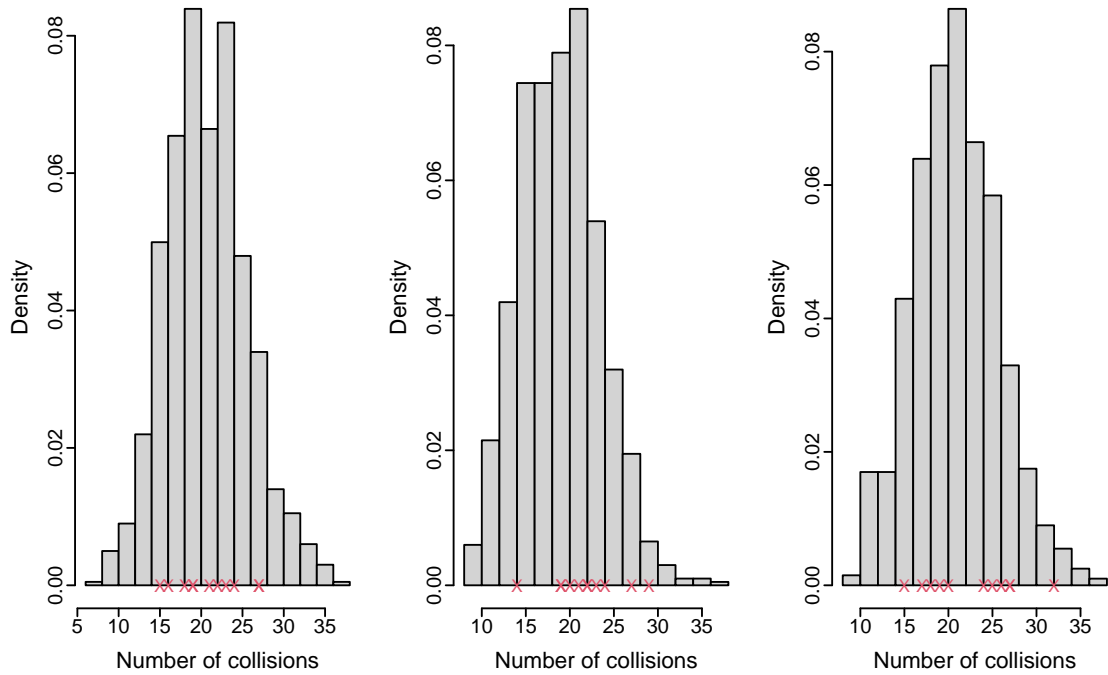


Figure 6.9: Within-sample predictions: Histogram of posterior predictive draws of  $\tilde{y}_{mj}$  for  $j = 2$  and  $m = 1, 2, 3$  with true observations overlaid on the  $x$ -axis.

## 6.4 Discussion

Random effects modelling is a powerful statistical technique that can be used to analyse data and identify patterns that are not immediately obvious. One of the key benefits of this approach is that it allows us to exploit patterns in the response variable, making us less reliant on covariate information. This can be particularly useful in road safety analyses, where the available data may be limited or incomplete. By using random effects modelling, practitioners can gain valuable insights into monthly trends and other patterns in the data that can be used to make predictions for future time periods. Analyses may reveal spatial or seasonal patterns; this information can be used to develop more accurate predictions for future time periods, even in zones where no data are available. By exploiting patterns in the data, meaningful models can be developed that allow for quality predictions without the need for additional covariates. This is important in road safety analyses, where accurate predictions can help practitioners and other decision-makers develop effective interventions and policies to reduce accidents and improve overall road safety.

From this analysis on the Florida data, we have gained insight into monthly trends and how the rate of collisions at zones differ from each other. The estimates for seasonal effects suggested a sinusoidal pattern within the rate of collisions. The estimates suggest that the rates would be lowest over the spring time and with peaks over autumn. This is likely down to weather patterns in Florida, where the rainy season is defined as occurring between May and October, during which thunderstorms bring heavy rainfall. This could explain the steady increase in rate from April–September. After October, a dry season sets in across most of Florida, this dry season typically lasts until late April in most areas which again could explain the decrease in rate from October–April.

The zonal effects estimates suggest an increase in collision rate from West–East. This could be down to the terrain or landscape in these areas differing. Another possibility is that population density or traffic volume may vary across these regions, with more drivers on the road in certain areas leading to an increased risk of collisions. Weather conditions may also be a contributing factor. Finally, it is possible that cultural or demographic differences between the eastern and western regions of Florida may play a role, with differences in driving habits, attitudes towards road safety, or other factors contributing to a higher rate of collisions in certain areas. Further investigation and analysis would be necessary to determine the specific factors contributing to the observed increase in road collisions from West to East in Florida.

By comparing the random effects model to a fixed effects model, we found that the random effects model provided more precise marginal posterior parameter estimates for  $\mu_{mj}$ . Specifically, the level of uncertainty surrounding these estimates was lower in the random effects model compared to the fixed effects model. Less uncertainty is advantageous as our predictions from the random effects model would also be more precise, this is shown in Figure 6.6 where the majority of the standard deviations for within-sample predictions of  $\tilde{y}_{mj}$  are smaller from the random effects model than from the fixed effects. Note that the model discussed ignores any temporal dependence exhibited by the data, and without additional covariate information, out of sample forecasts are likely to be poor.

The model was then altered to allow analysis of count data. The monthly collision count data were chosen to follow a Poisson distribution which altered the top

layer in the hierarchical model from Section 6.2. Again the data were exploited and seasonal and zone effects were extrapolated to show when we would expect more collisions in a year period, and how collision counts over zones differ. The estimates for seasonal effects suggested that there was an overall increase in collisions counts from February to November, where we expect to see the peak in November across sites and the lowest in February. This could be explained by weather conditions. The winter months in the UK can be particularly challenging for drivers due to adverse weather conditions such as snow, ice, and fog, increasing the risk of collisions. There is also evidence of increased traffic over the summer months in the UK due to tourism and events. More people on the roads can increase the risk of collisions, especially in popular holiday destinations and major cities.

There is evidence of a increase in the rate from North to South and from West to East. We were able to conclude from this that we would expect higher collision counts in the South-East of the UK. This could be down to variables such as population density, urbanisation and infrastructure. This information is invaluable when we aim to predict collision counts at zones where no data are available. As for the Florida analysis, further investigation and analysis would be necessary to determine the specific factors contributing to the observed increase in rate of road collisions geographically in the UK.



# Chapter 7

## Application of a sinusoidal dynamic linear model to road traffic collision data

### 7.1 Introduction

Working with collision counts can introduce issues of zero-inflation, especially over short time-frames. By working with rates over zones, we have the advantage of fewer zeros in the dataset and upon removing these, we may treat the data as continuous, which can be mathematically convenient in terms developing a tractable model. Most road traffic data are recorded sequentially over time and it is common for there to be dependencies between each observation. Hence, it is necessary to account for these dependencies in the model via a time-series model, such as a state-space model. The use of state-space models in road safety analysis is relatively new and uncommon, though they provide advantages for prediction.

State-space models can be used for modelling univariate or multivariate time-series in the presence of non-stationarity, structural changes and irregular patterns (see e.g. [Harvey, 1990](#), [West and Harrison, 2006](#)). Time-series analysis begins with the formulation of a model that accounts for temporal dependence, for example through auto-correlation, trend or seasonality. The use of state-space models within a time-series setting allows for uncertainty quantification in both the observation process and any dynamic variables that are not observed directly. Forecasting there-

fore accounts for these different sources of uncertainty and, when inferences are made within the Bayesian paradigm, additional parameter uncertainty. Throughout, we focus on a particular class of state-space model within which the observation and system equations involve linear functions of the latent process. Such models are known as dynamic linear models (DLMs, see e.g. [Petris et al., 2009](#), [West and Harrison, 2006](#)) and offer several practical benefits over their nonlinear counterparts. Notably, they admit a tractable observed data likelihood function, allowing a computationally efficient approach to inference and forecasting.

[Gamerman and Migon \(1993\)](#) give a list of hierarchical dynamic linear models (DLMs) used for the state evolution, smoothing and filtering through the stages of the hierarchy. Although state-space models and DLMs in particular, have been to date rarely exploited in the road safety context (see e.g. [Buddhavarapu, 2015](#), [Fei et al., 2011](#)), they have been ubiquitously applied in environmental settings. For example, [Lai et al. \(2020\)](#) use a spatio-temporal model to forecast sensor output consisting of temperature and humidity measurements at five locations in North East England. The signal is described using coupled dynamic linear models, with spatial effects specified by a Gaussian process (GP). A related approach in the context of emissions data can be found in [Shaddick and Wakefield \(2002\)](#).

In this thesis we propose a joint spatio-temporal model of collision rates over multiple zones. A DLM is used at the level of a single zone, and allows for seasonality via a single harmonic with time varying amplitude and phase parameters. We then account for spatial dependence at nearby locations by adding a spatial Gaussian process to the system equation, thereby smoothing spatial deviations from the underlying temporal process. The resulting model allows for both within- and out-of-sample forecasting for locations which are fully observed and for locations for which some data are missing. A Bayesian approach is used to infer both dynamic and static model components and leverages the tractability of the observed data likelihood, which can be efficiently computed via a forward filter (see e.g. [Carter and Kohn, 1994](#), [Frühwirth-Schnatter, 1994](#)). We apply the inference scheme to a real data application. We assess the assumption of time-varying parameters governing the seasonal component to each zone separately before considering a joint model of all zones.

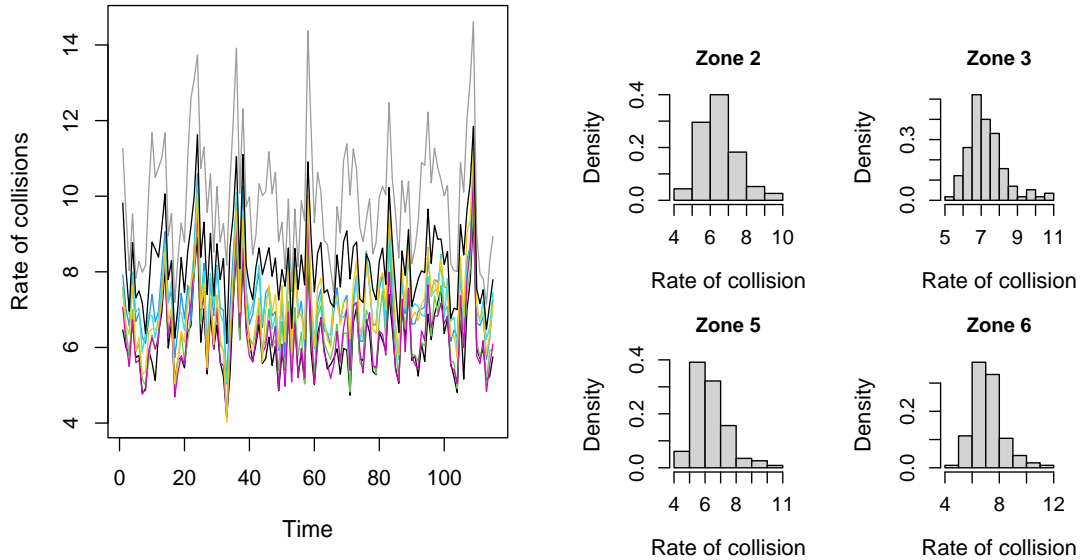


Figure 7.1: Left: Time series plots of monthly collision rate in each of the 8 zones. Right: Histograms of collision rates for zones 2,3,5,6.

### 7.1.1 Data

We consider a subset of the Florida monthly collision rate data outlined in Section 1.3.2. We focus on 8 zones, where for each zone we have 115 months of observations where the most recent observations are from April 2014. Figure 7.1 shows the multiple data streams over time for the different zones. For all zones, the monthly collision rates exhibit sinusoidal patterns over a 12 month period. Histograms of the monthly collision rates suggest that a Gaussian observation model may adequately describe the observation process. Through scatter plots, we determined that there was clear temporal dependence between certain months in year  $t$  to year  $t + 1$ , precluding the use of a simpler model with “month” as a fixed effect. Furthermore, zones geographically closer are more strongly correlated (see Figure 7.2).

### 7.1.2 Zone specific model

The dataset described in Section 1.3.2 showed seasonality in that, over all zones there was a clear sinusoidal pattern about the rate of collisions over a year. Therefore, to account for this within the DLM we include a single harmonic. Note that it

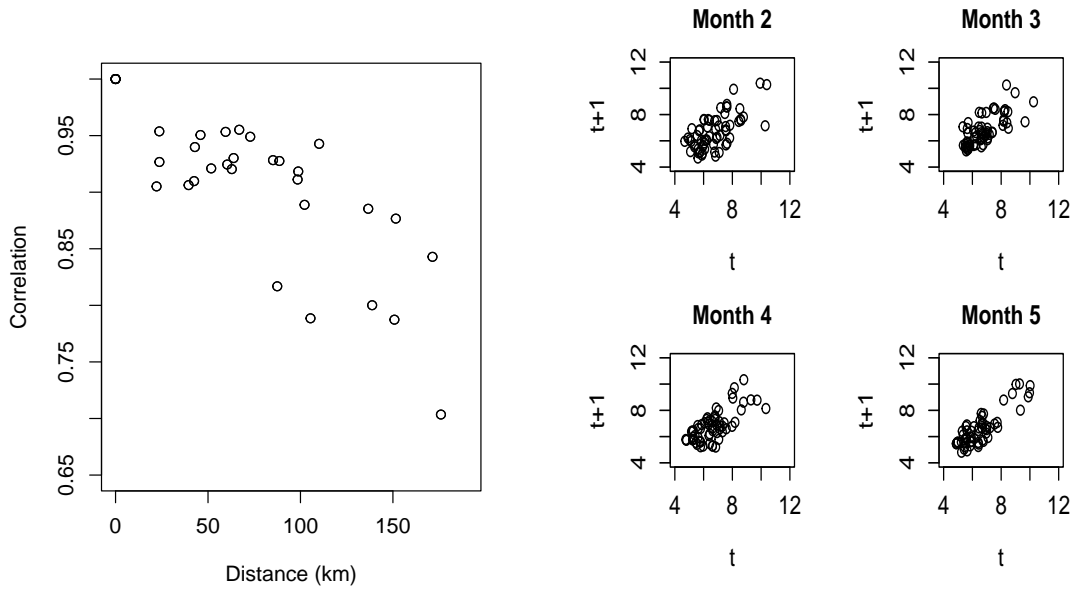


Figure 7.2: Left: Correlation between the 8 zones against distance between zones (km). Right: The temporal dependence between observations in months 2,3,4,5 in consecutive years across all zones.

is possible to account for seasonality through the inclusion of multiple harmonics in the system equation (see e.g. [Petris et al., 2009](#)), however, we find that using a single harmonic and allowing the amplitude and phase to vary over time, provides a parsimonious modelling approach.

Consider first a single location. We assume constant variance matrices  $V$  and  $W$  and data at irregularly spaced times  $t_1, t_2, \dots, t_n$ . The observation equation becomes

$$X_{t_i} = F_{t_i} \theta_{t_i} + \nu_{t_i}, \quad \nu_{t_i} \overset{indep}{\sim} N(0, V), \quad (7.1)$$

where  $\theta_{t_i} = (\theta_{1,t_i}, \theta_{2,t_i}, \theta_{3,t_i})^T$  and the observation matrix is given by

$$F_{t_i} = \left( \sin \left( \frac{2\pi t_i}{P_x} \right), \cos \left( \frac{2\pi t_i}{P_x} \right), 1 \right),$$

where  $P_x$  is the time corresponding to one complete period ( $P_x = 12$  for seasonal data). Note that the observation equation can be written as

$$X_{t_i} = \tilde{\theta}_{1,t_i} \cos \left( \frac{2\pi t_i}{P_x} - \tilde{\theta}_{2,t_i} \right) + \theta_{3,t_i} + \nu_{t_i} \quad (7.2)$$

where the dynamic parameters in Equation (7.1) and (7.2) are related using

$$\tilde{\theta}_{1,t_i} = \sqrt{\theta_{1,t_i}^2 + \theta_{2,t_i}^2}, \quad \tilde{\theta}_{2,t_i} = \tan^{-1} \left( \frac{\theta_{1,t_i}}{\theta_{2,t_i}} \right). \quad (7.3)$$

We impose some smoothness in these dynamic parameters by taking the system equation to be of the form

$$\theta_{t_i} = \theta_{t_{i-1}} + k_{t_i} \omega_{t_i}, \quad \omega_{t_i} \stackrel{indep}{\sim} N(0, W).$$

Which has been further altered to allow for measurements that are irregularly spaced on a temporal grid. That is, we include a coefficient,  $k_{t_i}$ , on the variance in the state equation such that  $k_{t_i}^2 = t_i - t_{i-1}$ . Hence the sinusoidal form DLM captures seasonality via a single harmonic whose amplitude  $\tilde{\theta}_{1,t_i}$  and phase  $\tilde{\theta}_{2,t_i}$  vary according to two transformed independent random walk processes.

### 7.1.3 Joint model over zones

We now consider a model of monthly collision rates that captures both the seasonality, and additionally, the correlation between nearby zones. Let  $X_{t_i} = (X_{t_i}^1, \dots, X_{t_i}^{n_z})^T$  denote the collection of monthly collisions rates at time  $t_i$  with  $X_{t_i}^j$  corresponding to zone  $j$ , and  $j = 1, \dots, n_z$ . In Section 7.3.1 we find that amplitude and phase are plausibly constant for each zone. Therefore, for ease of notation, in what follows we fix  $\theta_{1,t_i} = \theta_1$  and  $\theta_{2,t_i} = \theta_2$ . The model at zone  $j$  is

$$\begin{aligned} X_{t_i}^j &= \theta_1^j \sin \frac{\pi t_i}{6} + \theta_2^j \cos \frac{\pi t_i}{6} + \theta_{3,t_i}^j + \nu_{t_i}^j, \quad \nu_{t_i}^j \stackrel{indep}{\sim} N(0, V^j), \\ \theta_{3,t_i}^j &= \theta_{3,t_{i-1}}^j + k_{t_i} \omega_{t_i}^j + p_{t_i}^j, \quad \omega_{t_i}^j \stackrel{indep}{\sim} N(0, W^j). \end{aligned}$$

To induce correlation between nearby zones, we include the term  $p_{t_i}^j$ , as a component of a spatially smooth error process  $p_{t_i} = (p_{t_i}^1, \dots, p_{t_i}^{n_z})^T$ . We model  $\{p_{t_i}, t_i = 1, \dots, n\}$  using independent (over  $i$ ) zero mean Gaussian processes so that

$$p_{t_i} \stackrel{indep}{\sim} GP\{0, f_3(\cdot; \eta_3)\}.$$

We impose smoothness by taking a squared exponential kernel for the covariance

function. Hence, the covariance between spatial errors at locations  $j$  and  $j'$  is

$$f_3(d_{jj'}; \eta_3) = \text{Cov}(\theta_{3,t_i}^j, \theta_{3,t_i}^{j'}) = \sigma_3^2 \exp(-\phi_3 d_{jj'}), \quad (7.4)$$

with  $\eta_3 = (\sigma_3, \phi_3)^T$  parameterising the kernel; note that  $\phi_3$  determines the decay ratio of the correlation as the distance between zones  $j$  and  $j'$  ( $d_{jj'}$ ) increases (Banerjee et al., 2014). Similarly, we adopt GP priors for  $\theta_1 = (\theta_1^1, \dots, \theta_1^{n_z})^T$  and  $\theta_2 = (\theta_2^1, \dots, \theta_2^{n_z})^T$  so that  $\theta_1 \sim GP(m_1(\cdot), f_1(\cdot; \eta_1))$  and  $\theta_2 \sim GP(m_2(\cdot), f_2(\cdot; \eta_2))$  with  $f_1$  and  $f_2$  defined analogously to Equation (7.4) with the addition of  $m_1(\cdot)$  and  $m_2(\cdot)$  as appropriate mean functions. Hence, the full spatial DLM model (over all locations) is

$$\begin{aligned} X_{t_i} &= F_{t_i} \theta_{t_i} + \nu_{t_i}, & \nu_{t_i} &\overset{\text{indep}}{\sim} \text{N}(0, \text{diag}\{V^1, \dots, V^{n_z}\}), \\ \theta_{3,t_i} &= \theta_{3,t_{i-1}} + k_{t_i} \omega_{t_i}, & \omega_{t_i} &\overset{\text{indep}}{\sim} \text{N}(0, \text{diag}\{W^1, \dots, W^{n_z}\} + K_3), \end{aligned}$$

where  $F_{t_i} = \text{diag}(F_{t_i}^1, \dots, F_{t_i}^{n_z})$ ,  $\theta_{3,t_i} = (\theta_{3,t_i}^1, \dots, \theta_{3,t_i}^{n_z})^T$ ,  $\theta_{t_i} = (\theta_1^1, \theta_2^1, \theta_{3,t_i}^1, \dots, \theta_1^{n_z}, \theta_2^{n_z}, \theta_{3,t_i}^{n_z})^T$  and  $K_3$  is an  $n_z \times n_z$  matrix with  $(i, j)$ th element  $f_3(d_{ij}, \eta_3)$ .

## 7.2 Bayesian inference

For simplicity, suppose we have  $n_z$  zones with  $n$  observations in each zone. Let  $V = (V^1, \dots, V^{n_z})^T$  and  $W = (W^1, W^2, \dots, W^{n_z})^T$ . Furthermore, let  $\eta_3 = (\sigma_3, \phi_3)^T$  denote the hyperparameters governing  $f_3(\cdot)$ , with  $\eta_1 = (\sigma_1, \phi_1)^T$  and  $\eta_2 = (\sigma_2, \phi_2)^T$  denoting the hyperparameters governing  $f_1(\cdot)$  and  $f_2(\cdot)$  respectively. Let  $x^j = (x_{t_1}^j, \dots, x_{t_n}^j)^T$  denote the vector of collision rates at zone  $j$  so that  $x = (x^1, \dots, x^{n_z})$  denotes the complete dataset over all zones. The joint posterior over all dynamic and static parameters is proportional to the marginal static parameter posterior multiplied by the conditional posterior of the dynamic process  $\theta_3 = (\theta_{3,t_0}, \dots, \theta_{3,t_n})$  such that

$$\begin{aligned} \pi(\theta_1, \theta_2, V, W, \eta_1, \eta_2, \eta_3, \theta_3 | x) &\propto \\ &\pi(\theta_1, \theta_2, V, W, \eta_1, \eta_2, \eta_3 | x) \times \pi(\theta_3 | \theta_1, \theta_2, V, W, \eta_1, \eta_2, \eta_3, x). \end{aligned}$$

Let  $\psi$  denote all fixed model parameters. To simulate realisations from the joint posterior we use a two step approach:

1. Simulate from the marginal posterior  $\psi \sim \pi(\psi | x)$ .

2. Simulate from the conditional posterior  $\theta_3 \sim \pi(\theta_3|\psi, x)$ .

For step 1, as the marginal static parameter posterior is intractable, we use Markov chain Monte Carlo (see e.g. [Gilks et al., 1995](#)). For step 2 we use a forward filter backward sampling algorithm (see e.g. [West and Harrison, 2006](#)) to directly draw from  $\pi(\theta_3|\psi, x)$ . We provide details as follows.

### 7.2.1 Simulation based inference

Let  $\theta_{3,t_0:n} = (\theta_{3,t_0}, \theta_{3,t_1}, \dots, \theta_{3,t_n})$  denote the collection of latent states up to time  $t_n$  and let  $x = x_{t_1:n} = (x_{t_1}, \dots, x_{t_n})$  denote the observed data. Note that  $\theta_{3,t_i} = (\theta_{3,t_i}^1, \dots, \theta_{3,t_i}^{n_z})^T$  and  $x_{t_i} = (x_{t_i}^1, \dots, x_{t_i}^{n_z})^T$ . Upon assuming an independent prior specification for the constituent terms of  $\psi$ , Bayesian inference may proceed as follows. Integrating out the dynamic parameters, gives us the marginal posterior:

$$\begin{aligned} \pi(\theta_1, \theta_2, V, W, \eta_1, \eta_2, \eta_3|x) &\propto \\ &\pi(\theta_1|\eta_1)\pi(\theta_2|\eta_2) \left[ \prod_{j=1}^{n_z} \pi(V^j)\pi(W^j) \right] \times \\ &\pi(\eta_1)\pi(\eta_2)\pi(\eta_3) \times \pi(x|\theta_1, \theta_2, \eta_3, V, W), \end{aligned}$$

where the marginal likelihood  $\pi(x|\theta_1, \theta_2, \eta_3, V, W)$  is given by

$$\pi(x|\theta_1, \theta_2, \eta_3, V, W) = \pi(x_{t_1}|\theta_1, \theta_2, \eta_3, V, W) \prod_{i=1}^{n-1} \pi(x_{t_{i+1}}|x_{t_{1:i}}, \theta_1, \theta_2, \eta_3, V, W), \quad (7.5)$$

and whose constituent terms are analytically tractable. Moreover,  $\pi(\theta_1|\eta_1) = N(\theta_1; m_1, K_1)$  and  $\pi(\theta_2|\eta_2) = N(\theta_2; m_2, K_2)$  are multivariate normal densities,  $\pi(V^j)$  and  $\pi(W^j)$  are the prior densities ascribed to  $V^j$  and  $W^j$ ,  $\pi(\eta_1)$ ,  $\pi(\eta_2)$  and  $\pi(\eta_3)$  are the prior densities ascribed to  $\eta_1$ ,  $\eta_2$  and  $\eta_3$ .

The marginal likelihood can be efficiently evaluated using a forward filter. It will be helpful here to define

$$\tilde{X}_{t_i} \equiv X_{t_i} - \theta_1 \sin \frac{\pi t_i}{6} - \theta_2 \cos \frac{\pi t_i}{6} = \tilde{F}_{t_i} \theta_{3,t_i} + \nu_{t_i}, \quad \nu_{t_i} \stackrel{indep}{\sim} N(0, \text{diag}\{V^1, \dots, V^{n_z}\}),$$

so that

$$\tilde{X}_{t_i}|\theta_{3,t_i} \sim N(\tilde{F}_{t_i} \theta_{3,t_i}, \text{diag}\{V\}),$$

where  $\tilde{F}_{t_i}$  is the  $n_z \times n_z$  identity matrix and will be omitted for ease of notation in

what follows. We also write

$$\theta_{3,t_i} | \theta_{3,t_{i-1}} \sim N(\theta_{3,t_{i-1}}, \tilde{W}_{t_i}),$$

where  $\tilde{W}_{t_i} = k_{t_i}^2 (\text{diag}\{W^1, \dots, W^{n_z}\} + K_3)$ .

---

**Algorithm 7** Forward filter

---

1. Initial distribution:  $\theta_{3,t_0} \sim N(m_0, C_0)$ . Store the values of  $m_0$  and  $C_0$ .
2. For  $t_i, i = 1, \dots, n$ ,
  - (a) Prior at  $t_i$ . Using the system equation, we have that

$$\theta_3 | \tilde{x}_{t_{1:i-1}} \sim N(m_{t_{i-1}}, C_{t_{i-1}} + \tilde{W}_{t_i}).$$

Store  $R_{t_i} = C_{t_{i-1}} + \tilde{W}_{t_i}$ .

- (b) One step forecast. Using the observation equation, we have that

$$\tilde{X}_{t_i} | \tilde{x}_{t_{1:i-1}} \sim N(m_{t_{i-1}}, R_{t_i} + V).$$

Store the marginal likelihood contribution

$$\pi(\tilde{x}_{t_i} | \tilde{x}_{t_{1:i-1}}) = N(\tilde{x}_{t_i}; m_{t_{i-1}}, R_{t_i} + V).$$

- (c) Posterior at  $t_i$ :  $\theta_{3,t_i} | \tilde{x}_{t_{1:i}} \sim N(m_{t_i}, C_{t_i})$  where

$$\begin{aligned} m_{t_i} &= m_{t_{i-1}} + R_{t_i} (R_{t_i} + V)^{-1} (\tilde{x}_{t_i} - m_{t_{i-1}}), \\ C_{t_i} &= R_{t_i} - A_{t_i} Q_{t_i} A_{t_i}^T, \end{aligned}$$

where  $A_{t_i} = R_{t_i} Q_{t_i}^{-1}$  and  $Q_{t_i} = R_{t_i} + V$ . Store the values of  $m_{t_i}$  and  $C_{t_i}$ .

---

Algorithm 7 gives the steps of the forward filter. This is akin to the forward filter in Algorithm 4 from Section 2.5, where  $\theta_{3,t_i}$  is scalar and  $\tilde{F}_{t_i}$  and  $G_{t_i}$  are set to unity. We see that the constituent terms in Equation (7.5) are obtained from the forward pass as

$$\pi(\tilde{x}_{t_i} | \tilde{x}_{t_{1:i-1}}, \theta_1, \theta_2, \eta_3, V, W) = N(\tilde{x}_{t_i}; m_{t_{i-1}}, R_{t_i} + V),$$



where  $R_{t_i} = C_{t_{i-1}} + \tilde{W}_{t_i}$  and  $m_{t_{i-1}}, C_{t_{i-1}}$  are updated recursively; we refer the reader to [Petris et al. \(2009\)](#) (see also [Carter and Kohn, 1994](#), [Frühwirth-Schnatter, 1994](#), [West and Harrison, 2006](#), for further details).

Although the marginal likelihood is tractable, the posterior will typically be unavailable in closed form. Hence we use Metropolis-Hastings to generate draws from  $\pi(\psi|\tilde{x})$ ; see Algorithm 8.

---

**Algorithm 8** MCMC scheme

---

- 1 Initialise the chain with  $\psi^{(0)}$ . Set  $r = 1$ .
- 2 Propose  $\psi^* \sim q(\psi^*|\psi^{(r-1)})$ .
- 3 Calculate the acceptance probability  $\alpha(\psi^*|\psi^{(r-1)})$  of the proposed move, where

$$\begin{aligned} \alpha(\psi^*|\psi^{(r-1)}) &= \min \{1, A(\psi^*|\psi^{(r-1)})\} \\ &= \min \left\{ 1, \frac{\pi(\psi^*|\tilde{x}_{1:n})q(\psi^{(r-1)}|\psi^*)}{\pi(\psi^{(r-1)}|\tilde{x}_{1:n})q(\psi^*|\psi^{(r-1)})} \right\}. \end{aligned}$$

- 4 With probability  $\alpha(\psi^*|\psi^{(r-1)})$ , set  $\psi^{(r)} = \psi^*$ ; otherwise set  $\psi^{(r)} = \psi^{(r-1)}$ .
  - 5 Set  $r := r + 1$ . Return to step 2.
- 

It remains that, given draws of  $\psi^{(1)}, \dots, \psi^{(N)}$  we can sample  $\theta_3^{(r)} \sim \pi(\theta_3|\psi, x)$ ,  $r = 1, \dots, N$ . This can be achieved by noting the factorisation

$$\pi(\theta_3|\psi, x) = \pi(\theta_{3,t_n}|\psi, x_{t_{1:n}}) \prod_{i=0}^{n-1} \pi(\theta_{3,t_i}|\theta_{3,t_{i+1}}, \psi, x_{t_{1:i}}),$$

where the constituent densities are tractable and can be sampled recursively via a backward sampling algorithm. The key steps are given in Algorithm 9.

### Missing data

Missing observations are commonplace. That is, only observations on a subset of components of  $X_t$  may be available at time  $t_i$ . To account for this in the model we let  $\tilde{X}_{t_i}^o$  denote the observed rates at time  $t_i$ . The observation model is then written as

$$\tilde{X}_{t_i}^o = P_{t_i} \tilde{X}_{t_i}, \tag{7.6}$$

where the  $n_{obs} \times n_z$  incidence matrix  $P_{t_i}$  determines which components are observed at time  $t_i$  ([Lai et al., 2020](#)). For example, if we have data from 5 zones and data are

---

**Algorithm 9** Backward sampler

---

**Backward Sampling**

3. Sample  $\theta_{3,n} | \tilde{x}_{t_{1:n}} \sim N(m_n, C_n)$ .

4. For  $t_i, i = n, \dots, 1$ ,

(a) Backward distribution:  $\theta_{3,t_i} | \theta_{3,t_{i+1}}, \tilde{x}_{t_{1:i}} \sim N(h_{t_i}, H_{t_i})$ , where

$$\begin{aligned} h_{t_i} &= m_{t_i} + C_{t_i} (C_{t_i} + \tilde{W}_{t_{i+1}})^{-1} (\theta_{3,t_{i+1}} - m_{t_i}), \\ H_{t_i} &= C_{t_i} - C_{t_i}^T (C_{t_i} + \tilde{W}_{t_{i+1}})^{-1} C_{t_i}. \end{aligned}$$

(b) Sample  $\theta_{3,t_i} | \theta_{3,t_{i+1}}, \tilde{x}_{t_{1:i}} \sim N(h_{t_i}, H_{t_i})$ .

---

missing at the second and third zone at time  $t_i$ , then the incidence matrix is

$$P_{t_i} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The forward filter and backward sampler can be modified straightforwardly to allow for this scenario. In brief, each occurrence of  $\tilde{F}_{t_i}$  is replaced by  $P_{t_i} \tilde{F}_{t_i}$  and each occurrence of  $V$  is replaced by  $P_{t_i} V$  in Algorithm 7.

### 7.2.2 Within-sample predictive density

In order to assess model fit, we consider the within-sample predictive density. The within-sample predictive density is given by

$$\pi(\hat{x}_{t_{1:n}} | x_{t_{1:n}}) = \int \int \pi(\hat{x}_{t_{1:n}} | \theta_{3,t_{1:n}}, \psi) \pi(\theta_{3,t_{1:n}}, \psi | x_{t_{1:n}}) d\theta_{3,t_{1:n}} d\psi,$$

where

$$\pi(\theta_{3,t_{1:n}}, \psi | x_{t_{1:n}}) = \pi(\theta_{3,t_{1:n}} | \psi, x_{t_{1:n}}) \pi(\psi | x_{t_{1:n}}).$$

Although the within-sample predictive density is intractable, draws from  $\pi(\theta_{3,t_{1:n}}, \psi | x_{t_{1:n}})$  are readily available and therefore  $\pi(\hat{x}_{t_{1:n}} | x_{t_{1:n}})$  can be sampled via Monte Carlo. Given draws  $(\psi^{(r)}, \theta_{3,t_{1:n}}^{(r)})$ ,  $r = 1, \dots, N$ , we can simulate

$$\hat{X}_{t_i}^{(r):j} | \theta_{t_i}^{(r):j}, \psi^{(r):j} \sim N(F_{t_i} \theta_{t_i}^{(r):j}, V^{(r):j}), \quad r = 1, \dots, N, \quad i = 1, \dots, n, \quad j = 1, \dots, n_z, \quad (7.7)$$

where  $\theta_{t_i}^{(r),j} = (\theta_1^{(r),j}, \theta_2^{(r),j}, \theta_{3,t_i}^{(r),j})$  denotes the  $r$ th sample of  $\theta_{t_i}^j$ , with  $\hat{X}_{t_i}^{(r),j}$  defined similarly. Draws obtained from (7.7) can be summarised (e.g. via the mean, upper and lower quantiles) and bench marked against the observed data.

### 7.2.3 $k$ -step ahead prediction

The system and observation forecast distributions can be obtained by exploiting the linear Gaussian structure of the DLM. The one-step ahead system forecast density is given by

$$\begin{aligned}\pi(\theta_{3,t_{n+1}}|x_{t_{1:n}}) &= \int \int \pi(\theta_{3,t_{n+1}}|\theta_{3,t_n}, \psi, x_{t_{1:n}})\pi(\theta_{3,t_n}|\psi, x_{t_{1:n}})\pi(\psi|x_{t_{1:n}}) d\theta_{3,t_n} d\psi, \\ &= \int \pi(\theta_{3,t_{n+1}}|\psi, x_{t_{1:n}})\pi(\psi|x_{t_{1:n}}) d\psi,\end{aligned}$$

where

$$\pi(\theta_{3,t_{n+1}}|\psi, x_{t_{1:n}}) = N(\theta_{3,t_{n+1}}; m_{t_n}, C_{t_n} + \tilde{W}_{t_{n+1}}).$$

Similarly, the one-step ahead observation forecast density is given by

$$\pi(x_{t_{n+1}}|x_{t_{1:n}}) = \int \pi(x_{t_{n+1}}|\psi, x_{t_{1:n}})\pi(\psi|x_{t_{1:n}}) d\psi,$$

where

$$\pi(x_{t_{n+1}}|\psi, x_{t_{1:n}}) = N(x_{t_{n+1}}; m_{t_n}, C_{t_n} + \tilde{W}_{t_{n+1}} + V).$$

Hence, given  $N$  posterior summaries  $(m_{t_n}^{(r)}, C_{t_n}^{(r)})$ ,  $r = 1, \dots, N$  from  $\pi(\theta_{3,t_n}|\psi, x_{t_{1:n}})$  and  $\psi^{(r)}$  from  $\pi(\psi|x_{t_{1:n}})$ , the one-step ahead state and observation forecast distributions can be sampled via Monte Carlo, by drawing

$$\begin{aligned}\theta_{3,t_{n+1}}^{(r)}|\psi^{(r)}, x_{1:n} &\sim N(m_{t_n}^{(r)}, C_{t_n}^{(r)} + \tilde{W}_{t_{n+1}}^{(r)}), \\ \tilde{X}_{t_{n+1}}^{(r)}|\psi^{(r)}, x_{t_{1:n}} &\sim N(m_{t_n}^{(r)}, C_{t_n}^{(r)} + \tilde{W}_{t_{n+1}}^{(r)} + V^{(r)}).\end{aligned}$$

Then,  $X_{t_{n+1}}^{(r)}$  can be obtained from  $\tilde{X}_{t_{n+1}}^{(r)}$  by adding the term  $\theta_1^{(r)} \sin(\frac{\pi t_{n+1}}{6}) + \theta_2^{(r)} \cos(\frac{\pi t_{n+1}}{6})$  to the latter. For the general  $k$ -step ahead forecast, the above draws are replaced by

$$\begin{aligned}\theta_{n+k}^{(r)}|\psi, x_{1:n} &\sim N\left\{m_{t_n}^{(r)}, R_{t_{n+k}}^{(r)}\right\}, \\ \tilde{X}_{t_{n+k}}^{(r)}|\psi^{(r)}, x_{t_{1:n}} &\sim N\left\{m_{t_n}^{(r)}, R_{t_{n+k}}^{(r)} + V^{(r)}\right\},\end{aligned}$$

where

$$R_{t_{n+k}}^{(r)} = C_{t_n}^{(r)} + \sum_{i=1}^k \tilde{W}_{t_{n+i}}^{(r)}.$$

## 7.3 Application

In what follows, and where required, we implement the MCMC scheme from Section 7.2 by taking a random walk proposal with Gaussian innovations. We have that  $q(\psi^*|\psi) = N(\psi^*; \psi, \Sigma)$  where the innovation matrix  $\Sigma = \gamma \widehat{Var}(\psi|x)$ , with  $\widehat{Var}(\psi|x)$  obtained from a pilot run and  $\gamma$  is chosen to give an acceptance rate of around 25% (Roberts and Rosenthal, 2001). Within the MCMC scheme, for mathematical convenience, we will work with precisions such that  $\tau_V = 1/V, \tau_W = 1/W$ . Moreover, for parameter vectors whose components that must be strictly positive (i.e.  $V, W, \eta$ ) we implement the proposal on the log scale.

### 7.3.1 Single zone analysis

In this section we assess the assumption that amplitude and phase vary with time. We present results for zone 4 and note similar findings (namely that amplitude and phase are plausibly constant) for the remaining zones.

For the single zone model,  $\psi = (\tau_V, \tau_{W_1}, \tau_{W_2}, \tau_{W_3})^T$  is the vector of precision parameters. We set the mean and variance of  $\theta_{t_0}$  to be  $m_0 = (1.5, 1.5, 6)$  and  $C_0 = \text{diag}\{1.5, 1.5, 20\}$  respectively. We take an uninformative and independent prior specification for the components of  $\psi$ , via  $\tau_V, \tau_{W_1}, \tau_{W_2}, \tau_{W_3} \sim Ga(0.1, 0.1)$ . The MCMC scheme was run for 22k iterations with a thin of 20 and the first 2k discarded as burn-in, leaving 20k iterations on which to base posterior summaries.

The marginal MH scheme gives the estimated marginal posterior densities for the components of  $\psi$  shown in Figure 7.3 with their prior densities overlaid. The  $\psi$  samples were thinned to obtain 1k (near uncorrelated) draws from the marginal parameter posterior, denoted  $\{\psi^{(r)}\}_{r=1}^{1000}$ . The FFBS algorithm was then executed for each  $\psi^{(r)}$ , to obtain samples of the dynamic parameter vector,  $\{\theta_{t_i}^{(r)}\}_{r=1}^{1000}, i = 1, \dots, n$ , from the within-sample predictive. Samples of the dynamic components  $\theta_{1,t_i}$  and  $\theta_{2,t_i}$  can be transformed via (7.3) to obtain phase and amplitude draws from their

respective within-sample predictive densities; see Section 7.2.2 for further details regarding the method for obtaining samples from these predictive distributions. These distributions are summarised in Figure 7.4 via their means and 95% credible intervals. We can conclude that, upon allowing for the uncertainty in amplitude and phase, they are plausibly constant over time for this zone. Performing the analysis on the remaining zones shows that the same conclusions can be drawn. This suggests that the dynamic parameters  $\theta_{1,t_i}$  and  $\theta_{2,t_i}$ ,  $i = 1, \dots, n$ , can reasonably be replaced with static parameters  $\theta_1$  and  $\theta_2$ .

We assess the validity of the proposed model for a single zone by comparing observed data with their model-based within-sample posterior predictive distributions and with model-based out-of-sample forecast distributions. For the latter, we withheld the last 10 observations when fitting the model. Figure 7.5 shows the within-sample predictive distribution for the observation process, summarised by the mean and 95% credible interval calculated for each time point. This suggests that the model is able to reasonably account for the observation process. Similarly, the 10-step ahead forecast distribution is summarised by the mean and 95% credible interval at each time point. We see that the forecast distribution is able to capture the general trend exhibited by the observations.

### 7.3.2 Joint zone analysis

We now consider the joint model over all zones detailed in Section 7.1.3. Our prior specification takes the following form.

We expect that amplitude and phase should be similar at nearby zones. Recall that  $\theta_1 \sim GP(m_1(\cdot), f_1(\cdot; \eta_1))$ ,  $\theta_2 \sim GP(m_2(\cdot), f_2(\cdot; \eta_2))$  and the Gaussian process components in the dynamic mean process are  $p_{t_i} \stackrel{indep}{\sim} GP\{0, f_3(\cdot; \eta_3)\}$ . We take the mean functions to be constant so that  $m_1(\cdot) = m_2(\cdot) = 1.5 \cdot \mathbf{1}$ , with  $\mathbf{1}$  defined as an  $n_z \times 1$  vector of 1s. We have that  $f_k(d_{jj'}; \eta_k) = \sigma_k^2 \exp(-\phi_k d_{jj'})$ ,  $k = 1, 2, 3$ . We take  $\log \sigma_k \stackrel{indep}{\sim} N(\log(0.1), 0.1)$  representing fairly strong prior beliefs about the amplitude variance and phase within a zone. For the logarithm of the inverse length scales, we take  $\log \phi_k \sim N(\log(0.1), 0.1)$ , giving typical length scales of around 10km and reflecting typical distances between zones. The precisions of the observation equations governing each zone are  $\tau_{V_j} \stackrel{indep}{\sim} Ga(0.1, 0.1)$  and similarly for the system variances,  $\tau_{W_j} \stackrel{indep}{\sim} Ga(0.1, 0.1)$ ,  $j = 1, \dots, n_z$ . Finally, the initial values  $\theta_{3,t_0}^j$

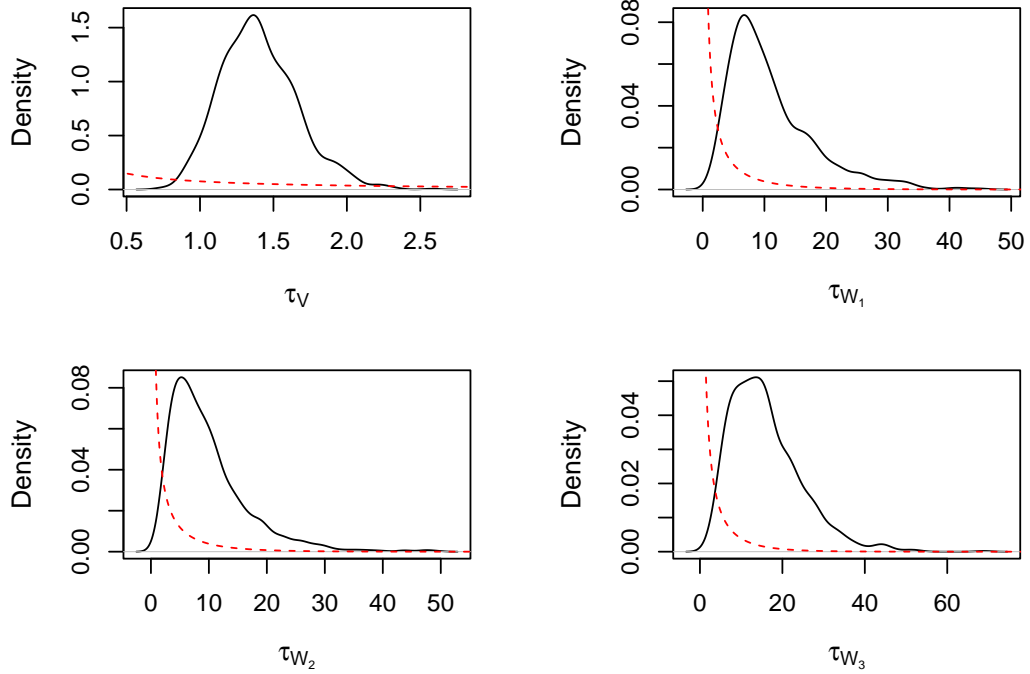


Figure 7.3: Density plots of  $\tau_V, \tau_{W_1}, \tau_{W_2}, \tau_{W_3}$  respectively, from 20k iterations and a thin of 20 with prior densities overlaid in red.

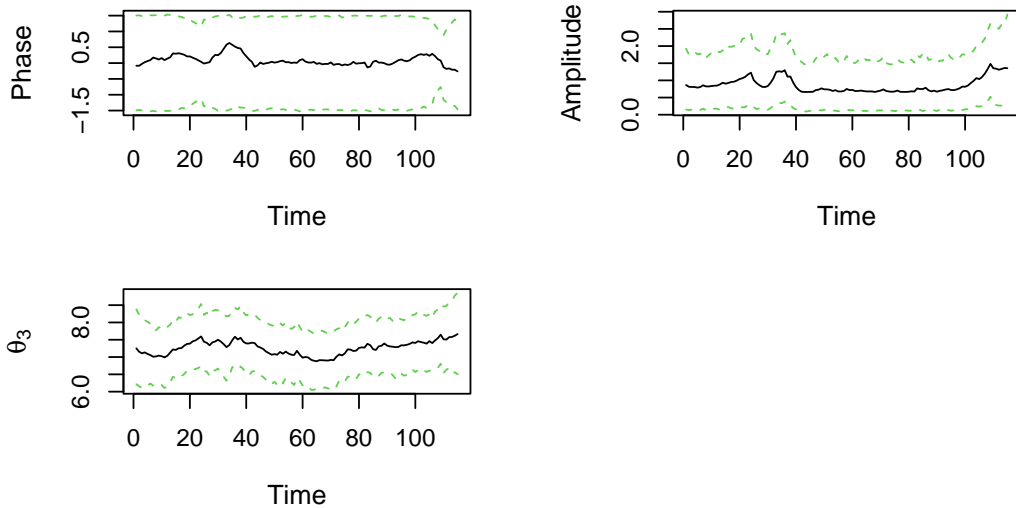


Figure 7.4: Phase, amplitude and  $\theta_{3,t_i}$  mean and 95% CI at zone 4 from time  $t_1$  to  $t_{115}$ .

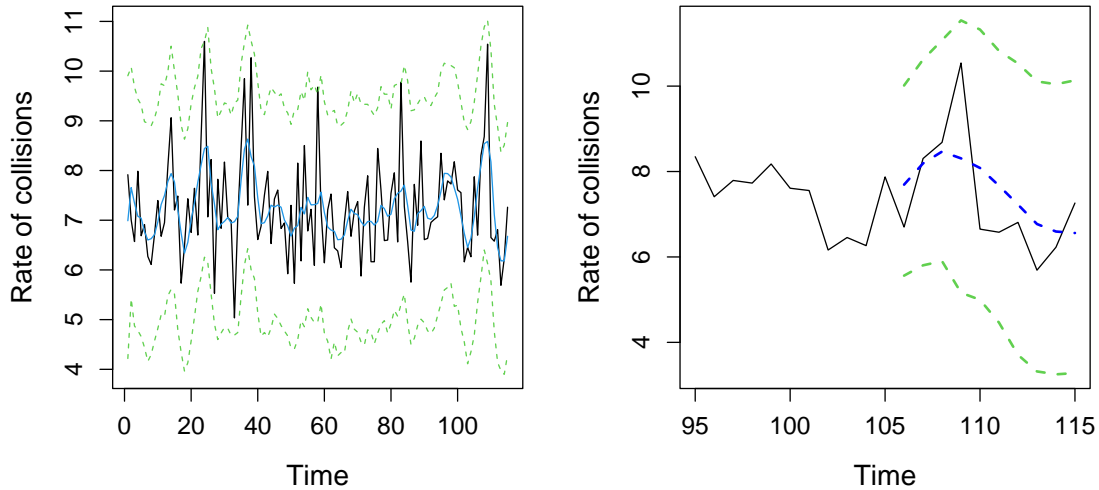


Figure 7.5: Left: Zone 4 observed data (black) with overlaid within-sample predictions – mean (blue) and 95% credible intervals (green). Right: Observed data (black) with overlaid 10 step ahead predictions – mean (blue) with 95% credible interval (green).

for each site were assumed to follow  $N(6, 20)$  distributions.

The MCMC scheme was run for  $1 \times 10^6$  iterations; the output is summarised by Table 7.1. Figure 7.6 shows the mean value and 95% credible interval of the posterior densities for amplitude and phase at each zone against longitude. There are signs of spatial dependence as the phase seems to decrease and amplitude increases in zones further to the east. Figure 7.7 shows a single period of the sine curve, averaged over draws of amplitude and phase for the most eastern versus most western zone with 95% credible intervals. From this we would expect to see more pronounced fluctuations in the rate of collisions across the year for eastern zones. Furthermore, we would expect the highest rate of collisions to be a month sooner (August) in eastern zones than that in western zones (September).

Figure 7.8 shows summaries (mean and 95% credible interval) of the difference between observations and the within-sample predicted observation process for zones 2, 4 and 8. The left-hand-side plots show the differences from the single zone analysis and the right-hand-side from the joint analysis. It is clear that the mean difference

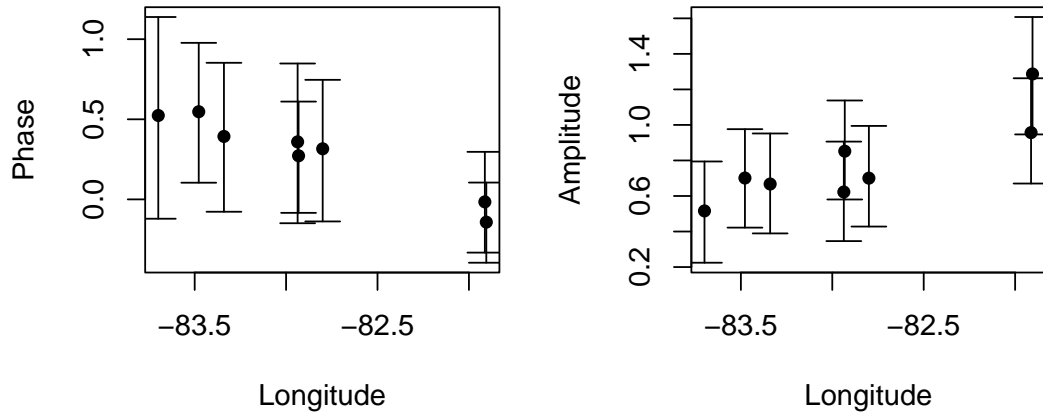


Figure 7.6: Mean amplitude and phase with 95% credible intervals against longitude for each zone.

at each time-zone combination is small and that a mean difference of zero is plausible (the 95% credible intervals include zero). Comparing left to right, shows the improvement in the within-sample predictions from a single zone analysis to a joint model; that is, the spatial information included through the GP has increased prediction precision. We additionally calculated the root mean square error (RMSE) at each time-point (observation vs prediction) and averaged this measure over all time points for each zone; the results are shown in Table 7.2. We see that the mean RMSEs are approximately 5 times larger for the single zone analysis, giving further evidence of an improvement in fit when considering a joint model over all zones.

Figure 7.9 shows 10-step ahead predictions for zones 2, 4, 6 and 8, following application of the method in Section 7.2.3. Note that the last 10 observations were removed from each zone before running the inference scheme. The figure shows that the forecast distributions are consistent with the data as they lie within the forecast intervals for all 7 zones. As we would expect, uncertainty grows as we move away from the last recorded observation.

## 7.4 Discussion and limitations

We have developed a spatio-temporal model for collision rates that allows for serial dependence, seasonality and correlation between rates at nearby zones. We



$\psi$	Mean	95% CI	$\psi$	Mean	95% CI
$V^1$	0.034	(0.021, 0.052)	$\theta_1^4$	0.251	(-0.045, 0.532)
$V^2$	0.025	(0.015, 0.039)	$\theta_1^5$	0.226	(-0.065, 0.514)
$V^3$	0.059	(0.039, 0.084)	$\theta_1^6$	0.249	(-0.039, 0.526)
$V^4$	0.037	(0.022, 0.058)	$\theta_1^7$	-0.181	(-0.501, 0.133)
$V^5$	0.031	(0.018, 0.048)	$\theta_1^8$	-0.014	(-0.308, 0.271)
$V^6$	0.041	(0.023, 0.066)	$\theta_2^1$	0.585	(0.301, 0.877)
$V^7$	0.119	(0.059, 0.196)	$\theta_2^2$	0.651	(0.367, 0.944)
$V^8$	0.045	(0.026, 0.071)	$\theta_2^3$	0.566	(0.285, 0.856)
$W^1$	0.021	(0.011, 0.037)	$\theta_2^4$	0.424	(0.144, 0.722)
$W^2$	0.024	(0.012, 0.041)	$\theta_2^5$	0.809	(0.530, 1.098)
$W^3$	0.023	(0.011, 0.044)	$\theta_2^6$	0.601	(0.311, 0.896)
$W^4$	0.025	(0.012, 0.044)	$\theta_2^7$	1.264	(0.931, 1.587)
$W^5$	0.024	(0.013, 0.043)	$\theta_2^8$	0.945	(0.660, 1.249)
$W^6$	0.034	(0.016, 0.061)	$\sigma_1$	1.688	(1.315, 2.309)
$W^7$	0.099	(0.031, 0.213)	$\sigma_2$	1.545	(1.201, 2.253)
$W^8$	0.029	(0.014, 0.055)	$\sigma_3$	1.352	(1.349, 1.355)
$\theta_1^1$	0.357	(0.066, 0.642)	$\phi_1$	1.527	(1.278, 1.917)
$\theta_1^2$	0.213	(-0.077, 0.494)	$\phi_2$	1.603	(1.387, 1.903)
$\theta_1^3$	0.213	(-0.084, 0.499)	$\phi_3$	1.103	(1.098, 1.107)

Table 7.1: Marginal parameter posterior means and quantile-based 95% credible intervals obtained from the MCMC scheme.

Zone	Mean RMSE	
	Single zone	Joint zone
1	1.384	0.198
2	1.066	0.167
3	1.214	0.258
4	1.083	0.206
5	1.114	0.185
6	1.191	0.217
7	1.339	0.378
8	1.107	0.227

Table 7.2: The mean RMSE over all time-points for each zone, from the single zone and joint zone analyses.

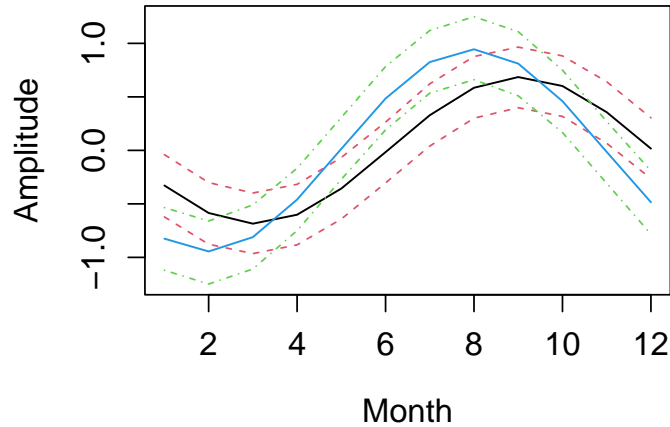


Figure 7.7: Mean and 95% credible intervals for the seasonal component for the most western zone (black) against the most eastern (blue).

considered a dynamic linear model (DLM) whose observation equation takes the form of a single harmonic with a smoothly time-varying amplitude and phase, thus accounting for seasonality and potential long term changes. Spatial consistency is accounted for at nearby zones by adding a Gaussian process (GP) component in the system equation. The model can be fitted in a Bayesian paradigm using an efficient two-stage Markov chain Monte Carlo procedure, targeting the joint posterior over the parameters, the latent time-varying harmonic coefficients (amplitude and phase) and dynamic mean. At the first stage, parameter samples are generated from the marginal parameter posterior using a random walk Metropolis algorithm with the likelihood evaluated via a forward filter. At the second stage, samples of the dynamic parameters are generated conditionally on the static parameter draws from stage one using a backward sampler. Further details of this forward filter, backward sampling (FFBS) approach can be found in [Petris et al. \(2009\)](#) (see also [Carter and Kohn, 1994](#), [Frühwirth-Schnatter, 1994](#)).

We applied our approach to a real dataset consisting of 115 months of collision rates over eight traffic administration zones in Florida, USA. An exploratory analysis that considered separate models for each zone found that the phase and amplitude were plausibly constant. We were therefore able to simplify the joint model over all zones by treating the harmonic components as static, with a GP prior allowing correlation between these parameters at nearby zones. The validity of both the

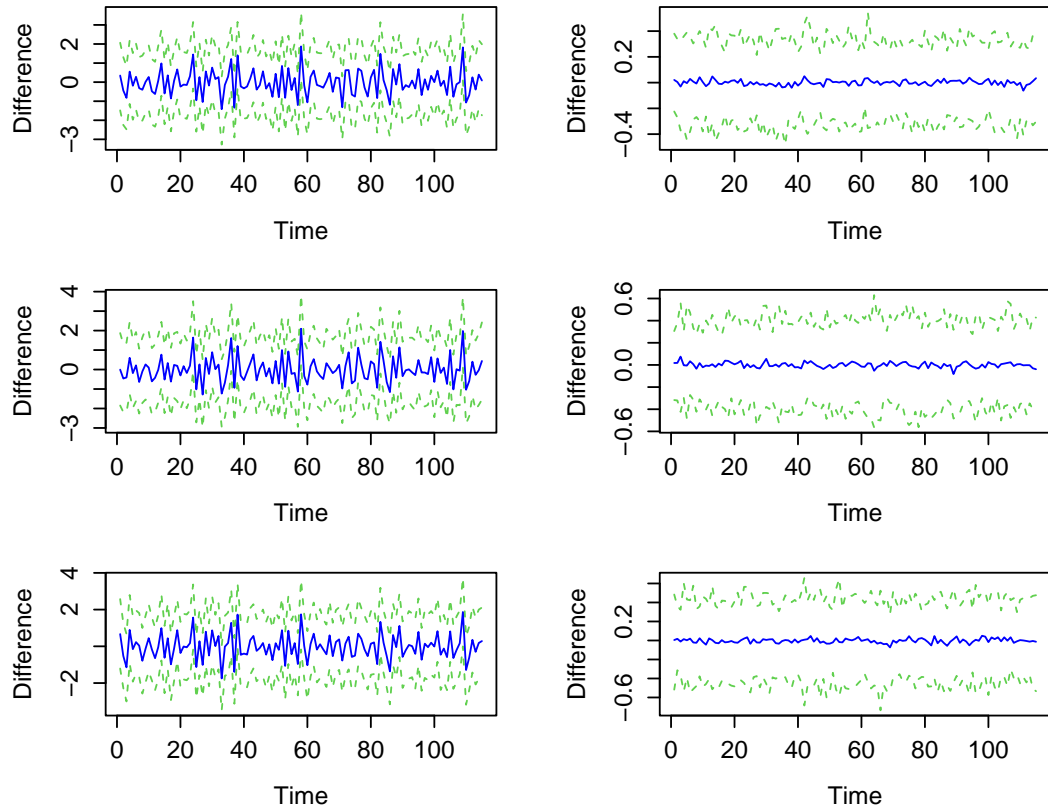


Figure 7.8: Mean (blue) and 95% credible intervals (green) for the difference between the within-sample predictive and the observations over time. Each row shows the differences from the single zone analysis (left) and the joint zone analysis (right) for zones 2 (top), 4 (middle) and 8 (bottom).

single zone and joint models was assessed using within-sample posterior predictive distributions, which suggested a satisfactory fit in both cases. Moreover, the within-sample predictions were improved substantially when using the joint model, with the credible intervals of our predictions narrowing almost tenfold, and a reduction in root mean squared error (RMSE) between the observations and predictions of around a factor of 5.

Our analysis suggests clear spatial patterns between phase and longitude as well as amplitude and longitude. For all zones we found that the lowest rates of collisions would fall earlier in the year. The model also suggests that for western zones, the lowest rates would be in March, and in February for eastern zones. It appears that peak collision rates are in September in the East and August in the West. We would

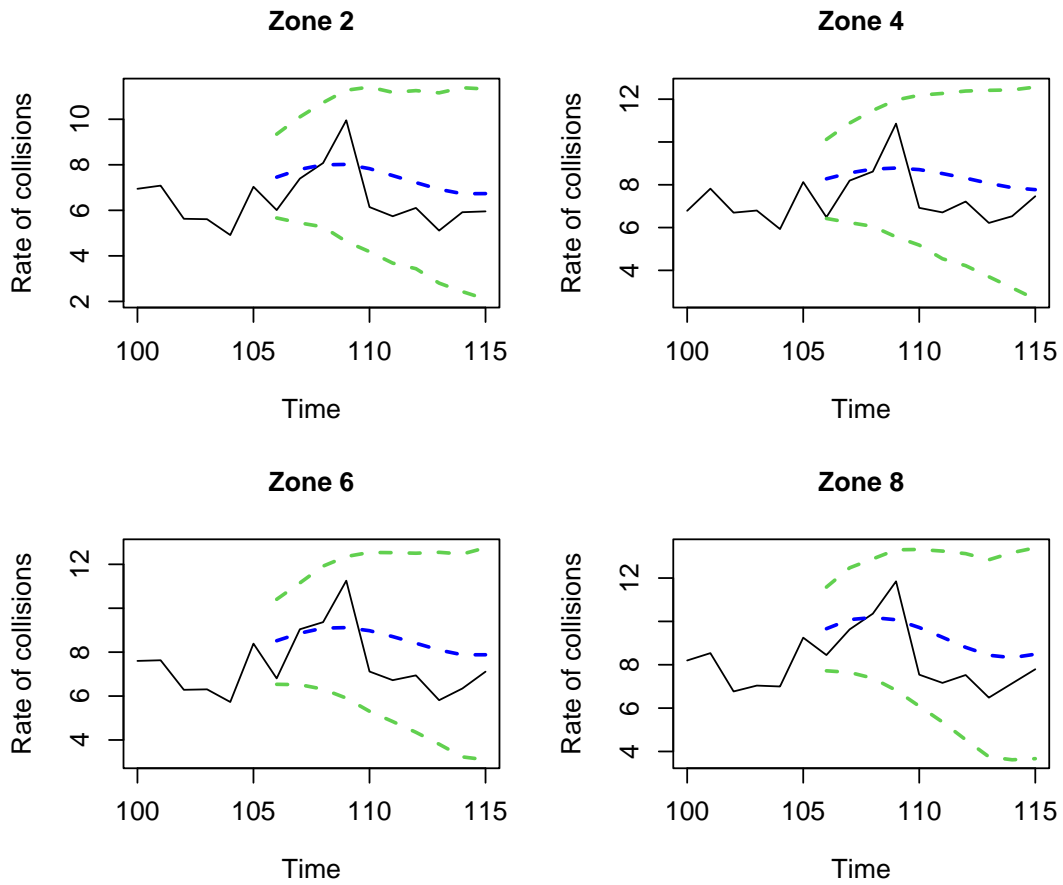


Figure 7.9: Rates of collisions in Florida zones 2, 4, 6 and 8 with overlaid out-of-sample 10-step ahead predictions – mean (blue) and 95% credible intervals (green).

also expect to see a larger fluctuation in the rate of collisions in an eastern zone. Our interest also lies in the ability to forecast collision rates in future months. Model-based out-of-sample forecast distributions suggest that our model is able to capture observed trend and seasonality in monthly collision rates up to around a year ahead.

In the context of road safety, the time-variant parameter  $\theta_{3,t_i}$  plays a pivotal role in capturing the overall mean rate of collisions at each site. This parameter is particularly useful for identifying both temporal trends and location-specific risk factors. It allows us to dissect the data to uncover how safety conditions evolve over time at individual locations, providing critical insights for targeted safety interventions. Furthermore, given its capability to model site-specific and temporal dynamics, this DLM with GP component could be extended to address the appar-

ent correlation between multivariate time-series data, such as rates of fatal, serious, and slight injury collisions. By modelling these categories simultaneously, we can gain a comprehensive understanding of how different severity levels of incidents are interrelated, thereby enabling more nuanced road safety evaluations and strategies.

In our pursuit of refining the model, various avenues were explored, including data transformation to address apparent positive skewness. However, this transformation turned out to be inconsequential, as it neither improved nor significantly altered the inference results. This suggests that the existing model already provides a satisfactory representation of the underlying data distribution. Beyond this, there are several other avenues for potential improvement in our modelling approach. For example, it is common to have covariate information such as traffic flow or average speed associated with a particular location at which a collision has occurred. However, pooling such data over zones is time-consuming and not always straightforward. Nevertheless, incorporation of covariates into the DLM framework is straightforward in principle, via the observation equation, and we anticipate improved prediction in this scenario. Although not pursued here, our model can also be used to predict collision rates at zones for which observations are not available. Interpolation of the fitted GP component in the system equation governing the dynamic mean and GP prior over the static parameters governing the harmonic, can be performed for unobserved zones of interest; see e.g. [Rasmussen and Williams \(2005\)](#) for further details.

# Chapter 8

## Conclusions and future work

### 8.1 Conclusions

In this thesis we have investigated challenges faced by road safety practitioners and introduced statistical techniques that enable them to use their data efficiently in safety scheme evaluation studies and in the prediction of hotspots. Chapter 1 emphasised the significance of enhancing road safety and the toll, in terms of both human lives and financial burden, inflicted by road collisions in the UK and worldwide. Chapter 2 outlined the concept of Bayesian inference and explained its usefulness in statistical analysis. We introduced background information on methods used throughout this thesis, including Bayesian inferential techniques such as Markov chain Monte Carlo, which are indispensable in the implementation of many statistical models.

Chapter 3 outlined the predicaments confronted by road safety practitioners who rely on data to make informed decisions, specifically the challenge posed by regression to the mean, which can yield erroneous conclusions if not addressed appropriately. We then outlined methods to overcome the issues of RTM by utilising external data from other sites within the network to detect any anomalous data at treated sites. We focused on road safety scheme evaluation and the conventional statistical approaches employed to perform an evaluation analysis. We discussed the possible influence of RTM and other confounders in misleading the scheme evaluation analysis. As a result, we highlighted the reason why a simplistic before/after comparison could result in an inflated estimate of the treatment effect. The inference was then shifted to take place completely within a Bayesian paradigm. This

*full Bayes* model allowed us to filter effects of regression to the mean from genuine treatment effects; hence, we were able to gain a ‘cleaner’ estimate of the treatment effect. The full Bayes approach was shown to include benefits such as capturing the uncertainty in the parameters more realistically and adding flexibility to our choice of prior distribution. We demonstrated the fitting of multiple priors and compared the resulting models in terms of their goodness-of-fit to the data using the Deviance Information Criterion.

Chapter 4 emphasised the necessity for bespoke modelling formulations for atypical before/after studies. We focused on two instances where the data were not of the ‘standard’ before/after collision count form. Firstly, we introduced a new model created for a study involving randomly allocated vehicle activated signs (VAS). The novel model was designed to capture both site and treatment effects and without choosing sites due to abnormally high collision counts, the model isn’t required to account for RTM. The model fit well to synthetic data and is ready for application to real data as and when the VAS study progresses. Secondly, we proposed a bivariate threshold excess model to conduct a traffic conflict-based before/after safety scheme evaluation of the leading pedestrian interval using extreme value theory. We introduced a block maxima model where treatment effects were measured through linear modelling of the location parameter. The generalised extreme value (GEV) distribution was fit to daily maxima and correctly identified the treated sites, where the coefficient of the time indicator forming the intercept term in the location parameter was wholly negative. To permit the use of more data in our analysis, we moved from modelling block maxima and modelled threshold excesses. We introduced a peaks over threshold (POT) approach, capturing treatment effects via linear modelling of the scale parameter, and removing temporal dependence through declustering. We then introduced an alternative which avoids having to remove temporal dependence altogether by using bivariate EVT. We accounted for dependence between consecutive extremes by assuming a first-order Markov structure, meaning we were able to include all of the data in the analysis to inform the parameter estimates. The bivariate threshold excess model resulted in narrower credible intervals than the POT approach and was also more successful in identifying which sites had been treated.

Chapter 5 continued the analysis of the LPI data and used the longitude and latitude of the sites to determine the presence of any spatial correlation. The in-

clusion of spatial extremes methods is commonly done so within the block maxima paradigm. Therefore, we expanded upon the daily maxima model which drew estimates of the GEV parameters of each site separately and altered this to become a joint model. The joint model allowed for draws from the likelihood over all sites in one MCMC algorithm. To account for spatial correlation between neighbouring sites, Gaussian processes were included in the expressions for the location and scale parameters, and the shape parameter was fixed across all sites. Although the model is computationally expensive, many of the parameters can be drawn using Gibbs sampling via their full conditional distributions. As before, treatment effects were measured within the location parameter, we modelled separate location parameters for the before and after periods and compared their posterior estimates to draw conclusions of the treatment effect. Including spatial dependence seemed to improve the estimation of which sites were treated and which were not. We also discussed two measures of safety indices - return levels and risk of crash. Both showed that at treated sites, there was evidence to suggest that the sites are safer after the introduction of LPI.

Chapter 6 turned attention away from safety scheme evaluation and onto ‘hotspot’ prediction. We proposed a Bayesian hierarchical model which allows the segregation of seasonal and zonal effects. The model was firstly applied to monthly collision rate data from zones in Florida. The data followed a normal distribution with mean varying with zone and month. The mean was modelled linearly where an underlying mean captured similarities between zones and two random effects parameters capture the zonal and seasonal variability. Owing to the semi-conjugate priors, we were able to use Gibbs sampling to perform inference. There was a clear sinusoidal pattern for the rate of collisions seasonally, which suggests there are higher collision rates in autumn months and lower rates in spring months. Zonal effects suggested rates would be higher in more eastern zones. Within-sample posterior predictive distributions showed good fits to the observed data. The advantages of modelling using random effects were discussed through a comparison with an equivalent fixed effects model. An improvement in precision of the mean of the normal distribution governing the data was shown, where posterior standard deviations from the fixed effects model were 5 times larger than those from the random effects model. The model was then altered to fit to UK collision count data which had been sourced from STATS19 data and aggregated via an SQL server. At the top-level, the data



now followed a Poisson distribution. Again, patterns in the data were exploited and seasonal and zonal effects were extrapolated which aided in conclusions that we would expect higher collision counts in the South-East of the UK. The ability to extrapolate this information from the data is invaluable when wanting to predict collision counts/rates at unobserved zones.

Chapter 7 introduced a spatio-temporal model for collision rates that allowed for serial dependence, seasonality and correlation between rates at nearby zones. To account for seasonality and potential long term changes, we proposed a sinusoidal dynamic linear model for a single zone with a smoothly time-varying amplitude and phase. From this, the phase and amplitude were found to be plausibly constant for all zones. We extended this to become a joint model and included a Gaussian process component in the system equation which accounted for spatial consistency at nearby zones. We allowed for spatial correlation between the static harmonic components via a GP prior. In both cases the validity of the model was checked using within-sample posterior predictive distributions, which suggested satisfactory fits in both cases. The inclusion of spatial correlation in the model improved the within-sample predictions reducing the root mean squared error between the observations and predictions by around a factor of 5. We were able to draw conclusions on collision rates at zones spatially, in that we have evidence to suggest that the lowest rates would come in March for western zones and in February in eastern zones. There also seemed to be larger fluctuations in the rate of collisions in eastern zones than in western. Model-based out-of-sample forecast distributions suggested the model was able to capture observed trend and seasonality in monthly collision rates up to around a year ahead.

In this thesis, a variety of statistical methodologies—including Empirical Bayes (EB), Fully Bayes (FB), Extreme Value Theory (EVT), random effects models, and Dynamic Linear Models (DLMs)—have been applied to road safety analysis. Each of these approaches comes with its own advantages and limitations, offering different perspectives on the complex issue of road safety. It's important to note that beyond the methods discussed in this thesis, other methodologies exist for both before/after studies as well as hotspot prediction, such as the propensity score method (Rosenbaum and Rubin, 1983), for example. It is noteworthy that causal inference has increasingly gained prominence in the road safety literature. Originating from

seminal works by Fisher (1925), Rubin (1974, 1978), Splawa-Neyman et al. (1923), the Rubin Causal Model (RCM), named by Holland (1986), offers an alternative framework that explicitly considers the mechanism of site selection for treatment and introduces the concept of counterfactual outcomes for causal effect estimation. This approach has been robustly employed in other scientific disciplines like epidemiology, economics, and sociology (Imbens and Wooldridge, 2009, Morgan and Winship, 2007, Robins et al., 2000). For its application in the context of road safety, readers may refer to the work by Davis (2000). Despite the broad range of methodologies considered in this work, causal inference represents an important avenue for future research. Its growing relevance in road safety studies suggests that incorporating this framework could potentially provide a more nuanced understanding of treatment effects, complementing the methods already employed in this thesis.

Throughout this research, we have methodically achieved our set objectives. We confronted and addressed challenges in road safety using innovative statistical methodologies. Our in-depth exploration of safety schemes and our advancements in hotspot prediction highlight the effectiveness of our approach. Leveraging a Bayesian framework, we managed to distinguish genuine treatment effects from confounding variables. Furthermore, our emphasis on bespoke modelling and the integration of spatial correlations underscores our commitment to precision and innovation. In sum, this thesis fulfills the aims and objectives we set out with, marking a significant stride in road safety analytics.

## 8.2 Future work

There are several promising avenues for future work in the areas explored in this thesis. One direction is to apply the treatment effects model, in Section 4.2, to real-world data once Gateshead Council proceeds with the random allocation of vehicle-activated signs. This would aid in the validation of the model's effectiveness and allow for the evaluation of the effectiveness of this type of road safety scheme using a robust statistical framework.

Additionally, there is potential to use the spatial extremes model developed in Chapter 5 for forecasting purposes, especially at unobserved sites. Forecasting collision rates in advance can aid in the planning of road safety interventions and

resource allocation. Forecasting at unobserved sites is also important because it can help prioritise limited resources towards the locations with the highest risk of road collisions. By leveraging information from nearby locations with available data, models can provide estimates of collision rates at unobserved locations, allowing decision-makers to make informed decisions about where to allocate resources such as traffic calming measures, increased enforcement, or other interventions to reduce the likelihood of collisions. This can be especially useful in countries where data collection may be more difficult or costly, as it can help focus limited resources on areas where they are most needed.

Another area for future exploration is the application of threshold techniques to the spatial extremes model. Such models are available but not as common as block maxima models, however benefits include the ability to be less wasteful of data. It may also be beneficial to incorporate covariate information into the models developed in this thesis, as this could improve the accuracy of the predictions and enable a more nuanced understanding of the factors that contribute to road safety. There is potential to expand upon the analysis conducted on UK data. This could include examining collision counts by day or time of day, incorporating covariate information, and exploring zero-inflated models to account for the excess zeros in the data.

The DLM with GP model in Chapter 7 can be improved by including covariate information such as traffic flow or average speed associated with a particular location at which a collision has occurred. The incorporation of covariates into the DLM framework is straightforward in principle, via the observation equation, and we anticipate improved prediction in this scenario. Additionally, our model can also be used to predict collision rates at zones for which observations are not available. Interpolation of the fitted GP component in the system equation governing the dynamic mean and GP prior over the static parameters governing the harmonic, can be performed for unobserved zones of interest. There is also scope to develop a similar model for application to count data. Times series consisting of counts will require a different approach. Ideally, a state space model with a Poisson observation process and a random walk (or related) state process (to capture the log of the mean of the observed process), such as a dynamic generalised linear model (Gamerman, 1998). Unfortunately, inference for this model will be complicated by the intractability of

the marginal likelihood. We will therefore appeal to recently developed pseudo-marginal MH approaches (see e.g., [Andrieu et al., 2010](#)). This approach is likely to be practically infeasible for long time series. We will therefore also consider a tractable approximation as a DLM, that may be appropriate for example, when the data are aggregated.

# Appendix A

## Supplementary information

### A.1 Useful probability distributions

We take this opportunity to introduce some probability distributions used throughout this thesis, see [Forbes et al. \(2011\)](#) for further information on these distributions.

#### A.1.1 Gamma distribution

The gamma distribution is a continuous probability distribution. A random variable  $X$  which is gamma-distributed with shape  $\alpha > 0$  and rate  $\beta > 0$ , is denoted

$$X \sim Ga(\alpha, \beta),$$

with probability density function (PDF)

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \quad x > 0, \alpha > 0, \beta > 0. \quad (\text{A.1})$$

Where  $\Gamma(\cdot)$  denotes the gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp\{-x\} dx, \quad \text{for } \alpha > 0.$$

The mean and variance are calculated as

$$\mathbb{E}(x) = \frac{\alpha}{\beta}; \quad \text{Var}(x) = \frac{\alpha}{\beta^2}.$$

From Equation (A.1), if we set  $\alpha = 1$ ,

$$f(x; \alpha = 1, \beta) = \beta \exp\{-\beta x\}, \quad x > 0,$$

we gain the PDF of an Exponential distribution. Hence we can conclude that a  $Ga(1, \beta)$  is equivalent to  $Exp(\beta)$ . More generally, the sum of  $n$  independent  $Exp(\beta)$  random variables is equivalent to a  $Ga(n, \beta)$  random variable.

### A.1.2 Weibull distribution

The Weibull is a continuous probability distribution. A random variable  $X$  which is Weibull-distributed with shape  $\kappa > 0$  and scale  $\lambda > 0$ , is denoted

$$X \sim Weibull(\kappa, \lambda),$$

with PDF

$$f(x; \kappa, \lambda) = \frac{\kappa}{\lambda} \left(\frac{x}{\lambda}\right)^{\kappa-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^\kappa\right\}, \quad x \geq 0, \kappa > 0, \lambda > 0.$$

The mean and variance are calculated as

$$\begin{aligned} \mathbb{E}(x) &= \lambda \Gamma(1 + 1/\kappa), \\ Var(x) &= \lambda^2 [\Gamma(1 + 2/\kappa) + (\Gamma(1 + 1/\kappa))^2]. \end{aligned}$$

The Weibull distribution is also equivalent to an Exponential distribution when  $\kappa = 1$ .

### A.1.3 Beta prime distribution

The beta prime distribution (or inverted beta distribution) is a continuous probability distribution. A random variable  $X$  which is beta prime-distributed with shape  $\alpha > 0$  and scale  $\beta > 0$  is denoted

$$X \sim \beta'(\alpha, \beta),$$

with PDF

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha, \beta)}, \quad x \geq 0, \alpha > 0, \beta > 0.$$

Where  $B(\cdot, \cdot)$  denotes the beta function, defined as

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

The mean and variance are calculated as

$$\begin{aligned} \mathbb{E}(x) &= \frac{\alpha}{\beta - 1}, \quad \text{if } \beta > 1, \\ \text{Var}(x) &= \frac{\alpha(\alpha + \beta - 1)}{(\beta - 2)(\beta - 1)^2}, \quad \text{if } \beta > 2. \end{aligned}$$

### A.1.4 Lognormal distribution

A lognormal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable  $X$  is lognormally distributed, then  $Y = \log(X)$  has a normal distribution. This would be denoted as

$$X \sim \text{Lognormal}(\mu, \sigma^2)$$

with PDF

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0, -\infty < \mu < \infty, \sigma > 0.$$

The mean and variance are calculated as

$$\begin{aligned} \mathbb{E}(x) &= \exp\left(\mu + \frac{\sigma^2}{2}\right), \\ \text{Var}(x) &= [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2). \end{aligned}$$

### A.1.5 Inverse-gamma distribution

An inverse-gamma distribution is a continuous probability distribution. If a random variable  $Y$  is gamma distributed, then  $1/Y$  is inverse-gamma distributed. A random variable  $X$  which is inverse-gamma distributed with shape  $\alpha > 0$  and scale  $\beta > 0$  is denoted as

$$X \sim \text{InvGa}(\alpha, \beta),$$

with PDF

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right), \quad x > 0, \alpha > 0, \beta > 0.$$

The mean and variance are calculated as

$$\begin{aligned} \mathbb{E}(x) &= \frac{\beta}{\alpha - 1} \text{ for } \alpha > 1, \\ \text{Var}(x) &= \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ for } \alpha > 2. \end{aligned}$$

### A.1.6 Negative binomial distribution

A negative binomial distribution is a probability distribution that describes the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified number of failures occur. The negative binomial distribution is characterised by two parameters: the probability of success in each trial,  $p$ , and the number of failures until the experiment is stopped,  $r$ . A random variable,  $X$  which is negative binomial distributed is denoted as

$$X \sim NB(r, p),$$

with probability mass function (PMF),

$$Pr(x = k) = \binom{k + r - 1}{k} p^r (1 - p)^k$$

where  $k$  is the number of successes. The mean and variance are calculated as

$$\begin{aligned} \mathbb{E}(x) &= \frac{r(1 - p)}{p}, \\ \text{Var}(x) &= \frac{r(1 - p)}{p^2}. \end{aligned}$$

### A.1.7 Poisson distribution

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event (Haight, 1967). A discrete random variable  $X$  which is



Poisson distributed with rate parameter  $\lambda > 0$  is denoted as

$$X \sim Po(\lambda),$$

with PMF

$$Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

The mean and variance are calculated as

$$\mathbb{E}(x) = Var(x) = \lambda.$$

# Bibliography

- AASHTO (2010). *Highway Safety Manual*. Number v. 1 in Highway Safety Manual. American Association of State Highway and Transportation Officials.
- Abdel-Aty, M. A. and Radwan, A. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5):633–642.
- Aguero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of costa rica. *Accident Analysis & Prevention*, 59:365–373.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.
- Arun, A., Lyon, C., Sayed, T., Washington, S., Loewenherz, F., Akers, D., Ananthanarayanan, G., Shu, Y., Bandy, M., and Haque, M. M. (2023). Leading pedestrian intervals – yay or nay? a before-after evaluation of multiple conflict types using an enhanced non-stationary framework integrating quantile regression into bayesian hierarchical extreme value analysis. *Accident Analysis & Prevention*, 181:106929.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015). *Hierarchical modeling and analysis for apatial data, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

- 
- Barao, M. I. and Tawn, J. A. (1999). Extremal analysis of short series with outliers: Sea-levels and athletics records. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 48(4):469–487.
- Box, G. E. and Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.
- Brooker, C. and North, R. (2007). Speed cameras: the twisted truth. <https://www.telegraph.co.uk/motoring/road-safety/2749419/Speed-cameras-the-twisted-truth.html>. [Online; accessed 24-May-2021].
- Buddhavarapu, P. N. V. S. R. (2015). *On Bayesian estimation of spatial and dynamic count models using data augmentation techniques: application to road safety management*. PhD thesis, The University of Texas at Austin.
- Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.
- Casella, G. and Berger, R. (2021). *Statistical Inference*. Cengage Learning.
- Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.
- Casson, E. and Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1:449–468.
- Chen, Y. and Tjandra, S. (2014). Daily collision prediction with sarimax and generalized linear models on the basis of temporal and weather variables. *Transportation Research Record*, 2432(1):26–36.
- Chin, H. C. and Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*, 35(2):253–259.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London.
- Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

- Cressie, N. (2015). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley.
- Davis, G. A. (2000). Accident reduction factors and causal inference in traffic safety studies: a review. *Accident Analysis & Prevention*, 32(1):95–109.
- Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2138):581–608.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical science*, 27(2):161–186.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.
- Department for Transport (2013). Statement of administrative sources, stats19 – personal injury road traffic accidents. [Online; accessed 24-May-2021].
- Department for Transport (2021). Ras60: Accident and casualty costs. <https://www.gov.uk/government/statistical-data-sets/reported-road-accidents-vehicles-and-casualties-tables-for-great-britain#accident-and-casualty-costs-ras60>. [Online; accessed 24-Sept-2022].
- Department for Transport (2022). Linking stats19 and tarn: an initial feasibility study. <https://www.gov.uk/government/statistics/linking-stats19-and-tarn-an-initial-feasibility-study/linking-stats19-and-tarn-an-initial-feasibility-study>. [Online; accessed 4-Oct-2021].
- Deublein, M., Schubert, M., Adey, B. T., Köhler, J., and Faber, M. H. (2013). Prediction of road accidents: A bayesian hierarchical approach. *Accident Analysis & Prevention*, 51:274–291.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 47(3):299–350.
- El-Basyouny, K., Barua, S., and Islam, M. T. (2014). Investigation of time and weather effects on crash types using full bayesian multivariate poisson lognormal models. *Accident Analysis & Prevention*, 73:91–99.

- El-Basyouny, K. and Sayed, T. (2009a). Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, 41(5):1118–1123.
- El-Basyouny, K. and Sayed, T. (2009b). Urban arterial accident prediction models with spatial effects. *Transportation research record*, 2102(1):27–33.
- El-Basyouny, K. and Sayed, T. (2012). Measuring direct and indirect treatment effects using safety performance intervention functions. *Safety science*, 50(4):1125–1132.
- Elvik, R. (1997). Evaluations of road accident blackspot treatment: A case of the iron law of evaluation studies? *Accident Analysis & Prevention*, 29(2):191–199.
- Elvik, R. (2002). The importance of confounding in observational before-and-after studies of road safety measures. *Accident Analysis & Prevention*, 34(5):631–635.
- Elvik, R. (2004). To what extent is there bias by selection?: Selection for road safety treatment in norway. *Transportation Research Record*, 1897(1):200–205.
- Engel, U. and Thomsen, L. K. (1992). Safety effects of speed reducing measures in danish residential areas. *Accident Analysis & Prevention*, 24(1):17–28. Special Issue Speed Management Through Traffic Engineering.
- Fawcett, L. and Thorpe, N. (2013). Mobile safety cameras: estimating casualty reductions and the demand for secondary healthcare. *Journal of Applied Statistics*, 40(11):2385–2406.
- Fawcett, L., Thorpe, N., Matthews, J., and Kremer, K. (2017). A novel bayesian hierarchical model for road safety hotspot prediction. *Accident Analysis & Prevention*, 99:262–271.
- Fawcett, L. and Walshaw, D. (2006). A hierarchical model for extreme wind speeds. *Applied statistics.*, 55(5):631–646.
- Fawcett, L. and Walshaw, D. (2012). Estimating return levels from serially dependent extremes. *Environmetrics*, 23(3):272–283.
- Fei, X., Lu, C.-C., and Liu, K. (2011). A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*, 19(6):1306–1318.

- Fisher, R. (1925). Statistical methods for research workers (oliver and boyd, edinburgh, uk).
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.
- Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011). *Statistical distributions*. John Wiley & Sons.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202.
- Fu, C., Sayed, T., and Zheng, L. (2021). Multi-type bayesian hierarchical modeling of traffic conflict extremes for crash estimation. *Accident Analysis & Prevention*, page 106309.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263.
- Galton, F. (1889). *Natural Inheritance*, volume 42. Macmillan.
- Gamerman, D. (1998). Markov chain monte carlo for dynamic generalised linear models. *Biometrika*, 85(1):215–227.
- Gamerman, D. and Migon, H. S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(3):629–642.
- García-Gallego, A., Georgantzís, N., Navarro-Martínez, D., and Sabater-Grande, G. (2011). The stochastic component in choice and regression to the mean. *Theory and decision*, 71(2):251–267.
- Gelman, A., Gilks, W. R., and Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, pages 423–453.

- Gross, F. and Scury, K. (2017). Hsip resources - safety: Federal highway administration. <https://safety.fhwa.dot.gov/hsip/docs/fhwas17039.pdf>. [Online; accessed 1-Nov-2022].
- Gumbel, E. J. (1958). *Statistics of extremes*. Columbia university press.
- Guo, X., Wu, L., Zou, Y., and Fawcett, L. (2019). Comparative analysis of empirical bayes and bayesian hierarchical models in hotspot identification. *Transportation Research Record*, 2673(7):111–121.
- Guo, Y., Sayed, T., and Zheng, L. (2020). A hierarchical bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals. *Accident Analysis & Prevention*, 147:105772.
- Haight, F. (1967). *Handbook of the Poisson Distribution*. Operations Research Society of America. Publications in operations research. Wiley.
- Harvey, A. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hauer, E. (1980). Bias-by-selection: Overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. *Accident Analysis & Prevention*, 12(2):113–117.
- Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis & Prevention*, 18(1):1–12.
- Hauer, E. (1997). *Observational Before/After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Emerald Group Publishing Limited.
- Heydari, S., Miranda-Moreno, L. F., and Liping, F. (2014). Speed limit reduction in urban areas: A before–after study using bayesian generalized mixed linear models. *Accident Analysis & Prevention*, 73:252–261.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

- 
- Høyve, A. (2015). Safety effects of section control—an empirical bayes evaluation. *Accident analysis & prevention*, 74:169–178.
- Huang, H. and Chin, H. C. (2010). Modeling road traffic crashes with zero-inflation and site-specific random effects. *Statistical Methods & Applications*, 19(3):445–462.
- Huser, R. and Davison, A. C. (2014). Space—time modelling of extreme events. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(2):439–461.
- Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *WIREs Computational Statistics*, 14(1):e1537.
- Ihueze, C. C. and Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of anambra state, nigeria. *Accident Analysis & Prevention*, 112:21–29.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Jamson, S., Lai, F., and Jamson, H. (2010). Driving simulators for robust comparisons: A case study evaluating road safety engineering treatments. *Accident Analysis & Prevention*, 42(3):961–971. Assessing Safety with Driving Simulators.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.
- Jiang, X., Abdel-Aty, M., and Alamili, S. (2014). Application of poisson random effect models for highway network screening. *Accident Analysis & Prevention*, 63:74–82.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.



- Katz, R. (1999). Extreme value theory for precipitation: sensitivity analysis for climate change. *Advances in Water Resources*, 23(2):133–139.
- Kitali, A. E. and Sando, P. T. (2017). A full bayesian approach to appraise the safety effects of pedestrian countdown signals to drivers. *Accident Analysis & Prevention*, 106:327–335.
- Lai, Y., Golightly, A., and Boys, R. J. (2020). Sequential bayesian inference for spatio-temporal models of temperature and humidity data. *Journal of Computational Science*, 43:101125.
- Lan, B., Persaud, B., Lyon, C., and Bhim, R. (2009). Validation of a full bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. *Accident Analysis & Prevention*, 41(3):574–580.
- Leadbetter, M., Lindgren, G., Rootzén, H., and Nosko, V. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer New York.
- Leadbetter, M. R. and Rootzén, H. (1988). Extremal theory for stochastic processes. *The Annals of Probability*, 16(2):431 – 478.
- Loo, B. and Anderson, T. (2015). *Spatial Analysis Methods of Road Traffic Collisions*. CRC Press.
- Lu, T., Dunyao, Z., Lixin, Y., and Pan, Z. (2015). The traffic accident hotspot prediction: Based on the logistic regression method. In *2015 International Conference on Transportation Information and Safety (ICTIS)*, pages 107–110.
- Matthews, J., Newman, K., Green, A., Fawcett, L., Thorpe, N., and Kremer, K. (2019). A decision support toolkit to inform road safety investment decisions. *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, 172(1):53–67.
- Miaou, S.-P., Hu, P. S., Wright, T., Rathi, A. K., and Davis, S. C. (1992). Relationship between truck accidents and highway geometric design: a poisson regression approach. *Transportation Research Record*.

- Morgan, S. L. and Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.
- Mountain, L., Maher, M., and Fawaz, B. (1998). Improved estimates of the safety effects of accident remedial schemes. *Traffic Engineering & Control*, 39(10).
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.
- Navarro, B., Miranda-Moreno, L., Saunier, N., Labbe, A., and Fu, T. (2022). Do stop-signs improve the safety for all road users? a before-after study of stop-controlled intersections using video-based trajectories and surrogate measures of safety. *Accident Analysis & Prevention*, 167:106563.
- Noland, R. B. and Quddus, M. A. (2004). A spatially disaggregate analysis of road casualties in england. *Accident Analysis & Prevention*, 36(6):973–984.
- Özari, Ç., Eren, Ö., and Saygin, H. (2019). A new methodology for the block maxima approach in selecting the optimal block size. *Tehnički vjesnik*, 26(5):1292–1296.
- Park, J. and Abdel-Aty, M. (2015). Development of adjustment functions to assess combined safety effects of multiple treatments on rural two-lane roadways. *Accident Analysis & Prevention*, 75:310–319.
- Parvareh, M., Karimi, A., Rezaei, S., Woldemichael, A., Nili, S., Nouri, B., and Nasab, N. E. (2018). Assessment and prediction of road accident injuries trend using time-series models in kurdistan. *Burns & Trauma*, 6.
- Paul, M. and Ghosh, I. (2021). Development of conflict severity index for safety evaluation of severe crash types at unsignalized intersections under mixed traffic. *Safety Science*, 144:105432.
- Persaud, B. and Lyon, C. (2007). Empirical bayes before–after safety studies: Lessons learned from two decades of experience and future directions. *Accident Analysis & Prevention*, 39(3):546–555.
- Petris, G., Petrone, S., and Campagnoli, P. (2009). Dynamic linear models. In *Dynamic Linear Models with R*, pages 31–84. Springer.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119 – 131.

- Qu, X. and Meng, Q. (2014). A note on hotspot identification for urban expressways. *Safety Science*, 66:87–91.
- Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of london crash data. *Accident Analysis & Prevention*, 40(4):1486–1497.
- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press.
- Reich, B. J. and Shaby, B. A. (2019). A spatial markov model for climate extremes. *Journal of Computational and Graphical Statistics*, 28(1):117–126.
- Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367.
- Robins, J. M., Ángel Hernán, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Roth, M., Buishand, T. A., Jongbloed, G., Klein Tank, A. M. G., and van Zanten, J. H. (2012). A regional peaks-over-threshold model in a nonstationary climate. *Water Resources Research*, 48(11).
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58.
- Saunier, N. and Sayed, T. (2007). Automated analysis of road safety with video data. *Transportation Research Record*, 2019(1):57–64.
- Sawalha, Z. and Sayed, T. (2006). Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 33(9):1115–1124.
- Schlüter, P., Deely, J., and Nicholson, A. (1997). Ranking and selecting motor vehicle accident sites by using a hierarchical bayesian model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(3):293–316.

- 
- Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(3):351–372.
- Shankar, V. N., Albin, R. B., Milton, J. C., and Mannering, F. L. (1998). Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model. *Transportation Research Record*, 1635(1):44–48.
- Shankar, V. N., Ulfarsson, G. F., Pendyala, R. M., and Nebergall, M. B. (2003). Modeling crashes involving pedestrians and motorized traffic. *Safety Science*, 41(7):627–640.
- Sharkey, P. and Winter, H. C. (2019). A bayesian spatial hierarchical model for extreme precipitation in great britain. *Environmetrics*, 30(1):e2529. e2529 env.2529.
- Shi, D., Smith, R. L., and Coles, S. G. (1992). Joint versus marginal estimation for bivariate extremes. Technical report, North Carolina State University. Dept. of Statistics.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.
- Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472.
- Stein, M. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York.
- Swedish Transport Administration (2012). Vision zero on the move. [https://ec.europa.eu/transport/road\\_safety/sites/roadsafety/files/pdf/20151210\\_1\\_sweden.pdf](https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/20151210_1_sweden.pdf). [Online; accessed 24-May-2021].
- Tarko, A. P. (2018). Surrogate measures of safety. In *Safe mobility: challenges, methodology and solutions*. Emerald Publishing Limited.

- The Global Network for Road Safety Legislators (2016). Manifesto #4road-safety - priorities for road safety policy and legislation 2020 & beyond. [https://d1c2gz5q23tkk0.cloudfront.net/assets/uploads/3008576/asset/Manifesto-EN\\_%281%29.pdf?1605172127](https://d1c2gz5q23tkk0.cloudfront.net/assets/uploads/3008576/asset/Manifesto-EN_%281%29.pdf?1605172127). [Online; accessed 1-Nov-2022].
- The International Road Assessment Programme (iRAP) (2022). Ai-rap.
- Transoft Solutions (2020). Trafxsafe. <https://safety.transoftsolutions.com/trafxsafe/>. [Online; accessed 6-Dec-2022].
- UN General Assembly (2017). Resolution 71/313. work of the statistical commission pertaining to the 2030 agenda for sustainable development. adopted 6 july 2017. Technical report, A/RES/71/313.
- Van Houten, R., Retting, R. A., Farmer, C. M., and Van Houten, J. (2000). Field evaluation of a leading pedestrian interval signal phase at three urban intersections. *Transportation Research Record*, 1734(1):86–92.
- Vettori, S., Huser, R., and Genton, M. G. (2019). Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3):831–841.
- Von Mises, R. (1954). La distribution de la plus grande de n valeurs. in (ed.), selected papers (vol. ii, pp. 271-294). providence, ri. *American Mathematical Society*.
- Wang, C., Xu, C., and Dai, Y. (2019). A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accident Analysis & Prevention*, 123:365–373.
- Wang, C., Xu, C., Xia, J., Qian, Z., and Lu, L. (2018). A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. *Transportation Research Part C: Emerging Technologies*, 90:281–291.
- Wang, J. and Huang, H. (2016). Road network safety evaluation using bayesian hierarchical joint model. *Accident Analysis & Prevention*, 90:152–158.
- Wang, J.-H., Abdel-Aty, M., and Wang, L. (2017). Examination of the reliability of the crash modification factors using empirical bayes method with resampling technique. *Accident Analysis & Prevention*, 104:96–105.

- West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
- WHO (2020). Road traffic injuries. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>. "[Online; accessed 24-May-2021]".
- Winnett, M. and Wheeler, A. (2002). Vehicle-activated signs: a large scale evaluation.
- Xie, F., Gladhill, K., Dixon, K. K., and Monsere, C. M. (2011). Calibration of highway safety manual predictive models for oregon state highways. *Transportation Research Record*, 2241(1):19–28.
- Yanmaz-Tuzel, O. and Ozbay, K. (2010). A comparative full bayesian before-and-after analysis and application to urban road safety countermeasures in new jersey. *Accident Analysis & Prevention*, 42(6):2099–2107.
- Yannis, G., Dragomanovits, A., Laiou, A., La Torre, F., Domenichini, L., Richter, T., Ruhl, S., Graham, D., and Karathodorou, N. (2017). Road traffic accident prediction modelling: a literature review. *Proceedings of the Institution of Civil Engineers - Transport*, 170(5):245–254.
- Yannis, G., Kondyli, A., and Georgopoulou, X. (2014). Investigation of the impact of low cost traffic engineering measures on road safety in urban areas. *International Journal of Injury Control and Safety Promotion*, 21(2):181–189. PMID: 23651448.
- Zaki, M. H., Sayed, T., and Billeh, M. (2020). Enhancing unsupervised video-based vehicle tracking and modeling for traffic data collection. *Canadian Journal of Civil Engineering*, 47(8):982–997.
- Zangenehpour, S., Strauss, J., Miranda-Moreno, L. F., and Saunier, N. (2016). Are signalized intersections with cycle tracks safer? a case-control study based on automated surrogate safety analysis using video data. *Accident Analysis & Prevention*, 86:161–172.
- Zeng, Q. and Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis & Prevention*, 67:105–112.
- Zhao, Y. (2012). *R and Data Mining: Examples and Case Studies*. Elsevier Science.

- Zheng, L., Ismail, K., Sayed, T., and Fatema, T. (2018). Bivariate extreme value modeling for road safety estimation. *Accident Analysis & Prevention*, 120:83–91.
- Zheng, L. and Sayed, T. (2019a). Bayesian hierarchical modeling of traffic conflict extremes for crash estimation: a non-stationary peak over threshold approach. *Analytic methods in accident research*, 24:100106.
- Zheng, L. and Sayed, T. (2019b). A full bayes approach for traffic conflict-based before–after safety evaluation using extreme value theory. *Accident Analysis & Prevention*, 131:308–315.
- Ziakopoulos, A. and Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis & Prevention*, 135:105323.