Imperial College London
Department of Epidemiology and Biostatistics

Bayesian spatio-temporal hierarchical models for disease mapping and detection

Areti Boulieri

Submitted for the degree of Doctor of Philosophy

December 2017

Declaration of Originality

I hereby declare that this thesis contains my own original work, unless explicitly acknowledged. Any contributions made by co-authors are detailed in the text, and any information derived from published or unpublished sources is appropriately referenced as such.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

In recent years, emerging computational algorithms have revolusionised the application of sophisticated Bayesian hierarchical models in numerous fields, including public health. Such models have become widespread in disease mapping and detection studies, where data of complex spatio-temporal structures arise, due to their flexibility to accommodate associated statistical challenges.

This thesis contributes to the scientific literature by building upon existing disease mapping and detection spatio-temporal models within the Bayesian hierarchical framework in order to answer important epidemiological questions. Motivated by chronic respiratory diseases and road traffic accidents, two major public health concerns characterised by inherent geographical structures, three main objectives are established.

First, a Bayesian model is proposed for the joint analysis of road traffic accidents by taking into account dependences across space, time and also between severity levels, in order to perform disease mapping and highlight increased risk of accident.

Second, the chronic respiratory disease is investigated by using a recently proposed Bayesian detection model. To achieve this, three data sources are analysed: mortality, hospital admissions, and drug prescriptions, each capturing a different level of disease severity. Along with providing spatio-temporal patterns of the chronic respiratory disease, the detection of unusual areas in terms of temporal behaviour is enabled, thus providing insights for hypothesis generation.

Following on from the previous work, a more flexible detection model is finally developed and evaluated through an extensive simulation study. Comparisons against the benchmark model are performed, suggesting great improvements of the proposed model. To illustrate the model, two case studies are carried out, on road traffic accidents and chronic obstructive pulmonary disease.

Acknowledgments

My thanks go firstly to my main supervisor Marta Blangiardo for her continuous guidance, encouragement, and genuine trust throughout the past four years. She has been the best mentor I could have hoped for and contributed greatly to make this PhD an enjoyable learning experience. I am also grateful to my second supervisor Anna Hansell for her advice on epidemiological aspects of this PhD. Her insightful and critical comments helped me to put this research into perspective. Furthermore, I would like to thank the Medical Research Council (MRC) Doctoral Training Partnership (DTP) for funding this research.

Thanks to Silvia Liverani and James Bennett for their scientific inputs, and the SAHSU team for their assistance with the data extraction. Thanks to my officemates, past and present: Faridah, Francesca, Gianluca, Gijs, Javiera, Jelena, Karin, Kawthar, Li, Olivia, Pauline, Saredo, Todd, and Ville for creating a fun environment in the VC room.

On a more personal note, I would like to thank my dear friends for the great times we had outside of work. Special thanks to Lida for sharing the highs and lows of PhD life. Most of all thanks to my parents Panagiotis and Dimitra, and my brother Yiannis, for supporting me spiritually along the way.

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**

1.1 Descriptive epidemiology 2

1.1.1 Public health outcomes 4

1.1.2 Challenges in small area studies 5

1.2 Bayesian inference 6

1.2.1 Bayesian computation 8

1.2.2 Convergence diagnostics 11

1.3 Statistical modelling 12

1.3.1 Models for disease mapping 12

1.3.2 Models for surveillance and detection 19

1.4 Structure of thesis 27

1.4.1 Provision of data 29

1.4.2 Contributions of co-authors 30

2 A multivariate Bayesian model for road traffic accidents data in England **31**

2.1 Introduction 31

2.2 Data 36

2.2.1 Road traffic accidents data 36
2.2.2 Traffic volume data 38
2.3 Statistical framework 39
2.3.1 Independent space effects models 39
2.3.2 Common space effects models 42
2.3.3 Correlated space effects models 43
2.3.4 Model comparison and checking 45
2.3.5 Posterior distribution of accident rates 46
2.4 Results 47
2.5 Discussion 58

3 Investigating spatio-temporal trends in asthma and COPD using a Bayesian detection model 62
3.1 Introduction 62
3.2 Data sources 65
3.2.1 GP drug prescription data 66
3.2.2 HES and mortality data 67
3.3 Statistical Analysis 69
3.4 Results 71
3.4.1 Spatial Patterns 71
3.4.2 Temporal patterns 75
3.4.3 Detection of unusual areas 77
3.5 Discussion 81

4 A more flexible Bayesian detection model for public health surveillance 85
4.1 Introduction 85
4.2 Proposed model 87

4.3 Simulation study 90
4.3.1 Data generation and study design 90
4.3.2 Results 91
4.4 Case study 1: Road traffic accidents 95
4.4.1 Data 95
4.4.2 Results 96
4.5 Case study 2: COPD hospital admissions 101
4.5.1 Data 101
4.5.2 Results 101
4.6 Discussion 104

5 Conclusions **109**
5.1 Summary of findings 109
5.2 Contributions to the state-of-the-art 111
5.3 Future directions 112

Bibliography **116**

A Supplementary material for Chapter 2 **133**

B Supplementary material for Chapter 3 **146**

C Supplementary material for Chapter 4 **150**

List of Figures

2.1 Temporal trend of the accident rate in England over 2005-2013 under Model E; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values. 51

2.2 Crude and posterior accident rates across wards in England for low severity in 2005; the insets show enlarged views of London. 53

2.3 Crude and posterior accident rates across wards in England for high severity in 2005; the insets show enlarged views of London. 54

2.4 Maps of posterior probability that residual accident rates belong to the top 800 ranked areas for low and high severity in England; the insets show enlarged views of London. 57

2.5 Area ranks based on crude rates versus area ranks based on residual accident risk for the top 100 areas. 58

3.1 Posterior means of the spatial relative risk $\exp(\eta_i)$ of CRD across CCGs in England for prescriptions (a), admissions (b), and deaths (c); the insets show enlarged views of London. 73

3.2 Heatmap showing correspondence across three data sources on CRD: admissions, deaths and prescriptions. 74

3.3 Temporal trends of CRD across admission, mortality, and prescription data; The solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values. 76

3.4 Temporal trends of CRD across across prescription, mortality, and admission data for the Isle of Wight CCG; the solid lines show the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values. 78

3.5 Temporal trends of CRD for the unusual CCGs in England plotted against the national trend under admission data; the solid lines show the posterior mean of the temporal effects $(\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines represent the corresponding 95% CI values. 80

4.1 Posterior means of the spatial accident rate $\exp(\eta_i)$ across districts in England; the insets show enlarged views of London. 97

4.2 Temporal trend of accident rates in England over 2005-2015; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% values. 98

4.3 Temporal trends of accident rates for three selected unusual districts plotted against the national trend in England; the solid lines show the posterior mean of the temporal effects $(\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines show the corresponding 95% CI values. $*$ represents the time of the aberration. 100

4.4 Posterior means of the spatial relative risk $\exp(\eta_i)$ of COPD admissions across CCGs in England; the insets show enlarged views of London. 103

4.5 Temporal trend of COPD admissions in England; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values. 104

4.6 Temporal trends of COPD admissions for three selected unusual CCGs plotted against the national trend in England; the solid lines show the posterior mean of the temporal effects $(\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines show the corresponding 95% CI values. $*$ represents the time of the aberration. 105

A.1 Plots showing observed versus predicted (posterior means) data for low severity across all years under each of the models considered in the study. 134

A.2 Plots showing observed versus predicted (posterior means) data for high severity across all years under each of the models considered in the study. 135

A.3 Maps showing areas with extreme p_{Bayes} values under Model E for low severity for years 2005 and 2013; the insets are enlarged views of London. 136

A.4 Maps showing areas with extreme p_{Bayes} values under Model E for high severity for years 2005 and 2013; the insets are enlarged views of London. 137

A.5 Posterior spatial effects $\exp(\theta_i + \phi_i)$ under three alternative Wishart priors for a sample of 800 areas for low severity. 139

A.6 Posterior spatial effects $\exp(\theta_i + \phi_i)$ under three alternative Wishart priors for a sample of 800 areas for high severity. 140

A.7 Posterior accident rates under two alternative Wishart priors for low and high severity level. 141
A.8 Posterior temporal effects $\exp(\delta_t)$ under three alternative priors for low and high severity; the solid lines show the means and the dashed lines represent the corresponding 95% CI values. 142
A.9 Map of posterior accident rates for high severity linked with the shapefile of the motorways; the insets are enlarged views of London. 143

B.1 Correlation among admission, mortality, and prescription data on CRD. 146
B.2 Maps of crude rates for GP drug prescription data on CRD across months August 2010 to March 2011; the insets are enlarged views of London. 147
B.3 Maps of crude rates for HES admission data on asthma and COPD across months August 2010 to March 2011; the insets are enlarged views of London. 148
B.4 Maps of crude rates for mortality data on asthma and COPD across months August 2010 to March 2011; the insets are enlarged views of London. 149

C.1 Temporal scenarios for the main simulation study (15 and 30 time points). 152
C.2 Temporal scenarios for the simulation study with short time series (8 time points). 155
C.3 Map of England with common and unusual districts in terms of road traffic accident risk; the insets are enlarged views of London. 158

C.4 Area-specific time trends plotted against the national trend for road traffic accidents in England; the first half of the total 44 unusual areas are plotted. 159
C.5 Area-specific time trends plotted against the national trend for road traffic accidents in England; the second half of the total 44 unusual areas are plotted. 160
C.6 Map of England with common and unusual CCGs in terms of COPD admissions relative risk; the insets are enlarged views of London. . . 162
C.7 Area-specific time trends plotted against the national trend for COPD admissions in England. 163

List of Tables

2.1 Summary statistics for counts of road traffic accidents at ward level by severity and year in England. 37

2.2 Models under different assumptions for spatial and temporal dependences between low and high accident severities. 40

2.3 Hierarchy of the spatial and temporal random effects for the developed models. 41

2.4 DIC values for the competing models; the best model is shown in **bold**. 48

2.5 Mean posterior estimates for the parameters of all the models developed in the study; standard deviations are in parentheses. 49

2.6 Proportion of areas with extreme p_{Bayes} values and the mean p_{Bayes} values for low and high severity under Model E; standard deviations are in parentheses. 52

3.1 Summary statistics for counts of GP drug prescriptions 67

3.2 Summary statistics for counts of HES admissions and deaths 68

4.1 Scenarios for the simulated data 92

4.2 Posterior means of performance measures for scenarios S1-S9; credible intervals are in parentheses. 94

4.3 Performance for scenario S10 95

C.1 Summary statistics for counts of asthma hospitalisations used as expected counts in the simulation study. 153
C.2 Posterior means of performance measures for the simulation study with short time series; credible intervals are in parentheses. 156
C.3 Summary statistics for counts of road traffic accidents at district level by year. 157
C.4 Summary statistics for counts of HES admissions on COPD at CCG level by month. 161

1 Introduction

Spatial epidemiology dates back to 1854 when John Snow built a map to investigate a cholera outbreak in London and to identify its sources (Snow, 1855). Over the last few decades, this area of research has seen an enormous growth, partly because of the increasing availability of data with geographical information, and partly because of advances in computing technology that allow the implementation of algorithms that were not possible in the past (Elliott et al., 2001).

The overarching goal of this PhD is to develop, apply and evaluate Bayesian spatio-temporal models for disease mapping and surveillance. I build upon the existing methodology within the Bayesian hierarchical framework to answer questions related to non-communicable diseases and road safety. Road traffic accidents have similarities to non-communicable diseases due to their intrinsic geographical patterns, and the fact that they cause a major burden in public health. In particular, the following research objectives are established:

1. To develop a multivariate Bayesian spatio-temporal hierarchical model for the analysis of road traffic accidents by severity level and investigate into the dependence structures across space and time and between severities.

2. To apply a recently developed Bayesian model for non-communicable diseases

to help understanding patterns of chronic respiratory diseases (CRDs), with a focus on the detection of unusual temporal patterns.

3. To develop a novel modelling framework for disease mapping and detection by using the aforementioned model as a foundation, and apply this on two case studies, i) road traffic accidents, and ii) chronic obstructive pulmonary disease (COPD).

The rest of this chapter introduces the theoretical concepts applied to this thesis. Section 1.1 defines the objectives of descriptive epidemiology in the context of georeferenced data, along with the public health outcomes considered in this work and challenges commonly encountered in small area studies. In section 1.2 the fundamental knowledge of Bayesian inference is provided, followed by the state-of-the-art review of statistical methods that will serve as the basis for this thesis in section 1.3. Finally, section 1.4 presents a detailed overview of the subsequent chapters.

1.1 Descriptive epidemiology

Countless studies have been conducted within the area of spatial epidemiology, providing outcomes useful in public health as they can inform on decisions for policy making. The spectrum varies from purely descriptive to formalised aetiological objectives.

Descriptive epidemiology is concerned with the description of the behaviour of public health conditions across space and/or time and is a fundamental tool for hypothesis generation. It thus provides a mandatory complement to aetiological studies which aim at formally assessing certain relationships between outcomes and explanatory variables. In this perspective, two closely related subgroups are

identified by disease mapping and surveillance, both playing a substantial role in public health practice.

Disease mapping illustrates geographical variations in public health conditions, often accounting for explanatory variables such as demographic, environmental, and socioeconomic factors. The construction of maps allows the visualisation of spatial patterns of the condition thus communicating complex information rapidly. By comparing these against maps of potential risk factors one can derive hypotheses as to the aetiology of the condition. When temporal information is incorporated, a more complete picture of the public health problem is provided.

Disease mapping may be considered to be embedded into the broader research area of public health surveillance. The aims of the latter however are thought to be more policy making oriented. Public health surveillance lies at the interface of clustering and temporal surveillance, and focuses on the detection of anomalies across space and/or time in order to implement interventions, rather than just provide descriptive measurements. Thacker and Stroup (1994) define it as the ongoing systematic data collection, analysis, interpretation and dissemination of information in order for action to be taken. It is an important task in numerous fields such as infectious and chronic diseases, healthcare malpractice, bioterrorism and road safety. The Centers for Disease Control and Prevention (CDC) in the USA, and the Health Protection Agency (HPA) and Public Health England (PHE) in England and Wales have routinely carried out surveillance data analysis with aiming at aiding policy formation and resource allocation (Lawson and Kleinman, 2005).

1.1.1 Public health outcomes

Numerous disease mapping studies have been conducted within public health; early examples include lip cancer (Clayton and Kaldor, 1987), asthma (Weiss and Wagener, 1990), adult leukemia (Olsen et al., 1996), suicides (Congdon, 1996) and later road traffic accidents (Miaou et al., 2003) and veterinary parasitology (Cringoli et al., 2005).

Public health surveillance on the other hand has traditionally focused on communicable diseases, which are characterised by the presence of occasional outbreaks. Nevertheless the public health interest is shifting from communicable to non-communicable diseases with the latter now causing a larger burden of disease than the former (World Health Organization, 2009a). The detection of anomalies in patterns of non-communicable diseases may indicate the presence of an emerged localised risk factor, an impact of an intervention, or differences in the quality of data, such as misdiagnosis of a disease, and underreporting of cases (Best and Wakefield, 1999). Similarly to disease mapping, such findings along with other epidemiological evidence and laboratory studies may suggest a possible aetiology (Elliott and Wartenberg, 2004).

In addition to non-communicable diseases, road traffic accidents have also become an issue of great concern, with World Health Organization (2009b) reporting that these are predicted to develop into the seventh leading cause of death globally by 2030. In order to improve road safety, better capacity is required for road traffic accidents analysis. For instance, highlighting dangerous zones and detecting anomalous behaviour in their spatio-temporal patterns may contribute to more efficient program planning.

In disease mapping and surveillance studies, the counts of cases are commonly

aggregated at small administrative regions with well defined boundaries, and derive from administrative sources. The recent digitalisation of National Health Service (NHS) general practitioners (GPs) services has resulted in the availability of a vast amount of information on population health, such as GP visits, hospital admissions and deaths, thus providing a great opportunity for epidemiological research. These are routinely available from national registries containing information on demographic characteristics and public health outcomes, and become available at the small-area level due to confidentiality, particularly when it comes to rare conditions.

The Department for Transport equally offers a reliable source of data on road traffic deaths and injuries that can be used for road safety research.

1.1.2 Challenges in small area studies

The analysis of small-area data, with complex structures containing geographical and temporal information, gives rise to numerous statistical challenges when it comes to performing disease mapping and detection.

The main one is the so-called small area problem, which arises when the sample size within a small area is too small to provide accurate estimates from the data. This is quite common when rare conditions are modelled, in which cases significant results may be due to pure chance producing false findings.

In addition, spatial correlation plays an important role in the analysis of geographical data, as locations that are close to each other in space often have similar values of outcomes while locations far apart are often different. Likewise, temporal correlation is evident across consecutive observed outcomes. It is crucial that these spatial and temporal dependences are accounted for in the statistical analysis so that reliable estimates are provided (Olsen et al., 1996).

As in all data analyses, the quality and representativeness of data are essential

in order for the models to be useful. Particularly with aggregated data, drawbacks include misdiagnosis, double counting and misclassification, and these need to be taken into account when assessing the quality and interpreting results.

Finally, in small-area studies the spatial and temporal resolution also play a vital role; the choice of an appropriate spatial and temporal scale is a key decision that will influence the outcomes of the study. This may be determined in relation to policy translation and implementation, the variation in population size or areal size, however the options are often destricted by practical issues such as data availability (Arsenault et al., 2013).

1.2 Bayesian inference

The majority of the statistical challenges encountered in spatial analyses can be accommodated by the Bayesian approach, which treats the model parameters as unknown random variables, unlike frequentist approaches which assume that these are fixed. By combining observed data with prior knowledge, Bayesian inference is provided on the parameters of interest through distributions rather than point and interval estimates. From a philosophical point of view, the representation of uncertainty about parameters provide more intuitive probability interpretations.

The work by Berger (1985) and Bernardo and Smith (1994) have been fundamental in establishing the Bayesian statistical approach. At the root of Bayesian theory however is the famous paper by Reverend Thomas Bayes (1701-1761) (Bayes, 1763), where he introduces the so-called Bayes' theorem, an uncontroversial formula which expresses the updated probability about the model parameters after seeing the data.

Given a parameter of interest θ and data y, the posterior distribution of θ is

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{1.1}$$

where $p(\theta)$ is the prior distribution of θ expressing the uncertainty before seeing the data, $p(y|\theta)$ is the likelihood function of the data, and $p(\theta|y)$ is the posterior distribution of θ which is used to make inference. The normalising constant $p(y)$ ensures that $p(\theta|y)$ is a valid probability distribution

$$p(y) = \int p(y|\theta)p(\theta)d(\theta) \tag{1.2}$$

Equation (1.1) can be thus written as

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{1.3}$$

which says that the posterior distribution of θ depends on the corresponding prior distribution and the likelihood. Prior distributions are used to either provide additional data for the problem at hand, i.e. external knowledge from previous studies or expert opinion, or to allow for certain assumptions to be considered in the modelling framework. Their parameters can have fixed values or be regarded as stochastic with corresponding distributions, known as hyperparameters. The posterior distribution contains all the information needed for Bayesian inference and is used to calculate quantities of interest such as the mean, variance or exceedance probabilities. In small area studies, where the problem of spatial correlation arises, Bayesian modelling allows for the inclusion of the spatial correlation at a hierarchical level, thus providing a more flexible framework for combining complex data structures and assumptions. In this context, the parameter of interest is usually the risk of the

disease, and its posterior distribution would describe the behaviour of the disease risk after the data is seen and assumptions are made.

Prior distributions and likelihood provide two sources of information. The relative influence of the prior and the data depend on the sample size, and how informative the prior is. When external knowledge is not available, it is common practice to use a prior distribution that does not give preference on certain values of the parameter. This has a minimal impact on the posterior, and can be described as vague, flat, or non-informative prior (Gelman et al., 2003). Nevertheless the choice of priors is an essential part of Bayesian analysis, as these may affect the posterior considerably, particularly when the sample size is small. In general, sensitivity analysis is required to test the impact of the prior on the posterior estimates (Bernardinelli et al., 1995).

Non-informative priors can often be improper, meaning that the normalising constant in eq. (1.2) is not finite. This is not always an important problem since improper priors do not necessarily lead to improper posterior distributions, however several limitations arise such as identifiability, i.e. the ability to distinguish across model parameters.

In order to obtain posterior distributions, Bayesian inference relies on approximative methods because of the intractable numerical integrations that are often involved. Such methods have become available only recently thanks to the development of computing technology and they have empowered Bayesian data analysis.

1.2.1 Bayesian computation

Posterior sampling is central to Bayesian inference as it allows to obtain samples from the posterior distribution which are then summarised to yield estimates of interest. The most popular methods for sampling from complex Bayesian posterior

distributions are the Markov chain Monte Carlo (MCMC) methods, originally developed in the area of physics and image processing, and later further developed for applications in Bayesian modelling (Gilks et al., 1995).

MCMC make use of a Markov chain in order to obtain a stationary distribution for the parameters of interest. A Markov chain is a sequence of random variables $X^{(0)}, X^{(1)}, \ldots$ if for all t, the distribution of $X^{(t+1)}$, conditional on the value of $X^{(t)}$ is independent of all other preceding values $X^{(t-1)}, X^{(t-2)}, \ldots, X^{(0)}$

$$X^{(t+1)} \sim p_{tr}(x|X^{(t)}) \tag{1.4}$$

where the term $p_{tr}(x|X^{(t)})$ is the transition distribution of the Markov chain.

Suppose that we wish to generate a sample from a distribution $p(\theta|y)$; the key idea is to construct a Markov chain with arbitrary starting values and perform iterative simulation of parameter values, until a stationary distribution $\pi(x)$ is reached. This distribution will be independent of the initial values and will correspond to the posterior distribution, given that it is sufficiently close to stationarity (if $X^{(t)} \sim \pi$ then $X^{(t+1)} \sim \pi$). Hence, by selecting a suitable transition probability density, a sequence of realisations can be drawn that are approximately from the posterior distribution of interest.

The standard MCMC algorithms are Gibbs (Geman and Geman, 1984; Gelfand and Smith, 1990) and Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) also implemented in the statistical software BUGS (Bayesian inference using Gibbs sampling), used for all data analyses in this thesis.

Gibbs sampling

Gibbs sampling makes use of the full conditional distributions for each parameter which are the transition probability distributions of a Markov chain. Say that

we want to sample from the posterior distribution $p(\boldsymbol{\theta}|y)$. Given a set of values $\{\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_k^{(t)}\}$ at iteration t:

1. Sample $\theta_1^{(t+1)}$ from $p(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \ldots, \theta_k^{(t)}, y)$
2. Sample $\theta_2^{(t+1)}$ from $p(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_k^{(t)}, y)$
3. ...
4. Sample $\theta_k^{(t+1)}$ from $p(\theta_k|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \ldots, \theta_{k-1}^{(t+1)}, y)$

By the end of these steps, a new realisation of the vector $\boldsymbol{\theta}^{(t+1)}$ is completed. This process is then repeated several times until a sample from the 'true' posterior is drawn.

The Metropolis-Hastings algorithm

When the samples cannot be generated from each of the full conditional distributions $p(\theta_i|\theta_{j\neq i}, y)$, for $i = 1, \ldots, k$, the Metropolis-Hastings algorithm can be used, as it only requires a function that is proportional to the distribution to be sampled. Say that we want to sample from the posterior distribution $p(\boldsymbol{\theta}|y) \propto \pi(\boldsymbol{\theta})$. Given a valid density function $q(\theta^*|\theta^{(t)})$, and a set of values $\{\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_k^{(t)}\}$ at iteration t, the algorithm proceeds as follows:

1. Sample θ^* from the proposal distribution $q(\theta^*|\theta^{(t)})$
2. Compute acceptance ratio

$$r = \frac{\pi(\theta^*)q(\theta^{(t)}|\theta^*)}{\pi(\theta^{(t)})q(\theta^*|\theta^{(t)})}$$

3. Accept θ^* as $\theta^{(t+1)}$ with probability min(r,1). If θ^* is rejected then $\theta^{(t+1)} = \theta^{(t)}$.

As in Gibbs sampling, the above steps are repeated until a sufficient sample is drawn from $p(\boldsymbol{\theta}|y)$. Gibbs sampling is a special case of the Metropolis-Hasting algorithm, where proposal distributions are the posterior conditionals and the full conditional distribution always has an acceptance ratio of 1.

1.2.2 Convergence diagnostics

When performing MCMC algorithms, diagnostics are required to assess whether the iterative simulations have reached the stationary distribution of the Markov chain. The iterations need to run for long enough until they can be assumed to provide approximately correct samples from the posterior distribution of interest. The initial period until convergence is reached, known as burn-in, is discarded from the sample, and this can vary considerably for different analyses.

Several criteria have been proposed to help with judging whether convergence has been reached, however this is in general subjective. Informally, convergence can be assessed through time series plots of the parameters, which should look like a random scatter around a mean value. Formal diagnostic criteria include those proposed by Cowles and Carlin (1996) and Gelman and Rubin (1992), later improved by Brooks and Roberts (1998). The latter one, widely known as the Brooks-Gelman-Rubin (BGR) diagnostic, is based on M chains of different initial values. The authors propose to split the iteration range T into Q batches of equal size and for each of these batches compare the average width of the credible intervals of the parameters across M chains, denoted by $W(q)$, against the corresponding width of the pooled sample of length $M \times T$, denoted by $B(q)$, where $q = 1, \ldots, Q$. By plotting the ratio $\hat{R}(q) = B(q)/W(q)$ against q it helps to determine the number of iterations required before convergence is reached, as this should approach 1.

Finally, the autocorrelation of the chains needs to be evaluated as when this

is high, the efficiency of the posterior sample decreases substantially. A solution to this is thinning of the chain, where only every k^{th} realisation from the sampler is retained and the rest are discarded. The parameter k is selected such that an independent sample is ensured (Lunn et al., 2012).

1.3 Statistical modelling

A range of statistical models have been developed over the last decades for small area studies and this is a very active research area that continues to grow. This section provides an overview of the literature in the area of disease mapping and detection, with a particular focus on the Bayesian methodology that will serve as the basis for this thesis. The methods are distinguished into those focusing on disease mapping and those on detection, although as previously mentioned, the classification is not always straightforward due to overlapping.

1.3.1 Models for disease mapping

Consider disease counts Y_i observed for a set of areas $i = 1, 2, \ldots, N$ which partition the study region. The counts follow a binomial distribution, however in cases where the binomial probability is small (i.e rare diseases and/or small areas) a Poisson model is used as an approximation to a binomial model.

In disease mapping studies, the standard statistical model considered to describe the distribution of the data is as follows

$$Y_i \sim \text{Poisson}(E_i \lambda_i) \tag{1.5}$$

where E_i is a fixed quantity representing the expected count of disease in area i, after adjusting for population characteristics through indirect standardisation. This

is calculated as

$$E_i = \sum_j N_{ij} p_j \tag{1.6}$$

where N_{ij} is the population count in area i and in stratum j, and p_j is the stratum-specific rate in area i. The parameter λ_i is the disease risk in area i relative to the expected one and is the parameter of interest in most disease mapping studies.

The simplest model would make use of the maximum likelihood to provide point estimates and associated variability for the parameters. The maximum likelihood estimate for λ_i is the standard mortality ratio (SMR) or standard incidence ratio (SIR)

$$\hat{\lambda}_i = \frac{Y_i}{E_i} \tag{1.7}$$

The corresponding variance of the estimator then equals

$$Var(\hat{\lambda}_i) = \frac{Y_i}{E_i^2} \tag{1.8}$$

which increases importantly when the size of the population at risk decreases, i.e. the denominator E_i in eq. (1.8), leading to inaccurate estimation of the disease risk.

Random effects can be incorporated into the modelling framework to allow for similarities in the disease across areas. The Bayesian framework is naturally suited for the implementation of random effects, providing extreme flexibility in the modelling. In addition, within this framework it is easy to specify external information (where available) on any parameter through informative prior distributions (Best et al., 2005).

The standard Bayesian hierarchical model has the following specification

$$\log(\lambda_i) = \alpha_0 + \nu_i \tag{1.9}$$

including an intercept α_0 representing the overall level of the disease risk and a random effect ν_i for the ith area. A non-informative flat prior is usually selected for the intercept α_0, such as a normal $N(0, 1000)$ or a flat prior. The parameter ν_i can be assumed to be independent and identically distributed, following an exchangeable normal prior distribution $N(0, \sigma^2)$, thus accounting for the extra Poisson variation in the data. The model in eq. (1.9) however does not account for spatial dependence which can be expected for areas that are in close proximity.

Various approaches have been proposed to account for spatial dependence including those by Stern and Cressie (2000), Leroux et al. (2000), MacNab and Dean (2000) and more recently Rodrigues and Assunção (2012); yet the most popular choice by far in small area studies is the BYM (Besag York Mollié) specification (Besag et al., 1991).

The BYM model defines the spatial dependency structure based on neighbourhood, i.e areas i and j are assumed to be neighbours if they share a common boundary, and decomposes the spatial random effects into a spatially unstructured and a spatially structured component

$$\log(\lambda_i) = \alpha_0 + \eta_i \tag{1.10}$$

$$\eta_i = \theta_i + \phi_i \tag{1.11}$$

The term η_i is called a convolution prior. The spatially unstructured component θ_i follows an exchangeable prior $N(0, \sigma^2)$, while the spatially structured component ϕ_i follows a conditional autoregressive (CAR) prior (Besag, 1974), which induces spatial autocorrelation based on the neighbourhood structure through the adjacency structure of the areal units

$$\phi_i|\phi_{(-i)} \sim \text{Normal}\Big(\bar{\phi}_i, \frac{\sigma_\phi^2}{n_i}\Big) \tag{1.12}$$

$$\bar{\phi}_i = \frac{\sum\limits_{k\in D_i} \phi_k}{n_i} \tag{1.13}$$

Here, $\phi_{(-i)}$ denotes all the elements of the vector ϕ except for area i. D_i represents the set of areas that are adjacent to area i (neighbours) and n_i is the total number of those areas. The parameter σ_ϕ^2 is the variance for the spatially structured random effects. The CAR prior thus assumes that the conditional expectation is the average of the random effects in neighbouring areas, while the conditional variance is inversely proportional to the number of neighbours. By 'borrowing strength' across areal units, i.e. using information from the neighbours to estimate area-specific random effects, the CAR prior improves local estimates, resulting in the smoothing of extreme rates, known as 'local smoothing'. Since the CAR prior is improper, a constraint $\sum_i \phi_i = 0$ is imposed for identifiability purposes, i.e. to distinguish among model parameters (Thomas et al., 2004).

In the BYM models it is often of interest to compute the relative contribution of structured and unstructured effects to the overall variability. This is quantified through the fraction of the marginal variability of the structured random effects σ_ϕ^2, over the total marginal variability $\sigma_\theta^2 + \sigma_\phi^2$. The parameter σ_ϕ^2 is not directly available, since by definition, the spatial effects ϕ_i are conditional on the neighbouring effects. The conditional variance is used instead, in order to estimate the empirical marginal variance

$$\hat{\sigma}_\phi^2 = \frac{\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2}{n-1} \tag{1.14}$$

The spatial fraction is then given by

$$\text{frac}_{\phi} = \frac{\hat{\sigma}_{\phi}^2}{\sigma_{\theta}^2 + \hat{\sigma}_{\phi}^2} \tag{1.15}$$

When this is close to 1, the structured effects explain most of the variability of the model, while when this is close to 0 the unstructured random effects dominate.

The modelling framework described so far provides static risk estimates. Extensions to space-time modelling have been proposed in order to account for the temporal evolution of spatial patterns over time. Given time points $t = 1, \ldots, T$, the model in eq. (1.5) becomes

$$Y_{it} \sim \text{Poisson}(E_{it}\lambda_{it}) \tag{1.16}$$

Bernardinelli et al. (1995) presented an approach in which an area specific intercept and a time trend are modelled as random effects, allowing for some form of space-time interactions. To address its restrictive assumption that the time trend is linear, Waller et al. (1997) proposed a model in which the spatially structured and unstructured effects are nested within time thus allowing for the spatial patterns at each time point to be completely different. Knorr-Held and Besag (1998) proposed to combine spatial and temporal effects additively, accounting for information shared both in space and time

$$\log(\lambda_{it}) = \alpha_0 + \eta_i + \delta_t \tag{1.17}$$

$$\eta_i = \theta_i + \phi_i \tag{1.18}$$

The term δ_t represents the temporally structured component, modelled hierarchically using a prior distribution. A common choice is the standard random

walk prior of order 1 (RW1), which assumes that the effects are dependent on the previous time point. It is also possible to define a version of the RW1 model where dependence is assumed on both the previous and the following time points. This corresponds to the 1-dimensional CAR prior where the neighbouring points of t are $t-1$ and $t+1$. Similar to the spatial CAR prior, the underlying assumption is a structure where the neighbouring time points are assumed to be similar (Fahrmeir and Lang, 2001). The RW1 prior is specified as

$$\delta_t|\delta_{(-t)} \sim \begin{cases} \text{Normal}\left(\delta_{t+1}, \sigma_\delta^2\right) & \text{for } t = 1 \\ \text{Normal}\left(\dfrac{\delta_{t-1}+\delta_{t+1}}{2}, \dfrac{\sigma_\delta^2}{2}\right) & \text{for } t = 2, 3, \ldots, T-1, \\ \text{Normal}\left(\delta_{t-1}, \sigma_\delta^2\right) & \text{for } t = T \end{cases} \tag{1.19}$$

where $\delta_{(-t)}$ denotes all elements of δ except for time point t and σ_δ^2 denotes the variance in the temporal effects. Similarly to the spatially structured effects, a constraint $\sum_t \delta_t = 0$ is imposed for identifiability.

An alternative choice to the first-order random walk (RW1) is the second-order random walk (RW2), which as the name suggests, assumes dependence of t on the two preceding time points $t-1$ and $t-2$ and possibly the two following time points $t+1$ and $t+2$. Various other temporal models have been proposed including the changepoint model (Carlin et al., 1992), spline models (MacNab and Dean, 2001) and the family of the moving average models that are typically used in time series analyses. Such models include the so-called autoregressive integrated moving average (ARIMA) model of Box and Jenkins (1976), the integer-valued autoregressive (INAR) model, introduced by McKenzie (1985) and further studied by Al-Osh and Alzaid (1987) and the exponentially weighted moving average (EWMA) (Hunter,

1986) and have been employed by many authors within the spatio-temporal setting, where the objective is usually forecasting (Zhou and Lawson, 2008; Aldor-Noiman et al., 2013; Briët et al., 2013).

An additional term γ_t can be added in eq. (1.17) which is specified by means of an exchangeable prior $N(0, \sigma_\gamma)$ to account for temporal heterogeneity in the data. In addition, the inclusion of a term x_{it} in eq. (1.17) may allow for space-time interactions as recommended by Knorr-Held (2000) who proposed four different models for x_{it}, respectively assuming independence, purely spatial correlation, purely temporal correlation and spatio-temporal correlation.

The parameters θ_i and ϕ_i can be interpreted as surrogates for unmeasured covariates that do or do not display spatial structure respectively. The same stands for the temporal parameters δ_t and γ_t. If information on relevant covariates is available, then this can be easily included in the model specification through explanatory variables.

As discussed in section 1.2, a hierarchical model requires hyperparameters which are themselves assigned prior distribution and these may have a considerable impact on the results. Common choices for the variances are the weakly-informative inverse-gamma priors $IG(0.5, 0.0005)$ (Kelsall and Wakefield, 1999; Wakefield et al., 2000) and $IG(\epsilon, \epsilon)$ where ϵ is a small value e.g. 1, 0.01, or 0.001 used by Lunn et al. (2012). The latter has been criticised by Gelman (2006) for producing inference that is sensitive to the choice of ϵ. The author suggested for a non-informative choice the half-normal $N(0, \sigma^2)$ on the natural scale of standard deviation, with σ set to a high value, e.g. 100, and for a slightly more informative choice the half-Cauchy with mean zero and scale set to a high value, e.g. 25.

Bayesian hierarchical disease mapping models have been extended to several directions aiming at providing estimates for the incidence of a disease. For instance,

MacNab and Dean (2001) and later MacNab and Gustafson (2007) developed space-time models with a combination of a global trend and a local trend, which were estimated using splines. Other variations have been proposed by Sun et al. (2000), Best et al. (2005), and Rushworth et al. (2014).

To jointly model two diseases, Knorr-Held and Best (2001) proposed the shared-component model which includes one common component for two diseases, and disease-specific components. This was applied to male and female lung cancer by Richardson et al. (2006), and later extended to jointly model multiple diseases (Held et al., 2005). Tzala and Best (2008) extended the CAR model to analyse different cancers in Greece, while several other authors have contributed to multivariate disease mapping (Gelfand and Vounatsou, 2003; Jin et al., 2005; MacNab, 2011; Martinez-Beneito, 2013). Prospective modelling of spatio-temporal data has also become increasingly popular in recent years. As an example, Bennett et al. (2015) developed a Bayesian hierarchical model to forecast mortality and life expectancy in England and Wales.

Disease mapping models are mostly developed for non-communicable diseases, however they have also been applied to infectious diseases where they are used as purely descriptive models. Adjustments are usually made in this case, in order to take into account the cases that are close to each other due to the transmission characteristics of this type of diseases (Cressie and Mugglin, 2000; Knorr-Held and Richardson, 2003; Zhou and Lawson, 2008; Lawson and Song, 2010).

1.3.2 Models for surveillance and detection

A range of statistical methods have been proposed for surveillance of counts of disease, and these can be broadly distinguished into test statistics and model-based methods (Lawson and Kleinman, 2005; Robertson et al., 2010).

Test-based methods

Methods based on test statistics seek to determine whether a defined subset (temporal, spatial, or spatio-temporal) is unusual compared to the incidence in the study region as a whole. Test statistics have a long history, see for instance the ones proposed by Knox and Bartlett (1964) and Mantel (1967) which determine spatio-temporal interactions, by comparing the number of disease counts that are related in space and time against the expected count under a null hypothesis of no interaction using likelihood ratio tests. The temporal historical limits method by Stroup et al. (1989) compares the current number of reported cases with a baseline value that is calculated based on comparable historical data, thus accounting for seasonality by design. The method has been routinely used for outbreak detection by the HPA in England and Wales (Levin-Rector et al., 2015) and the CDC in USA (Lawson and Kleinman, 2005).

Process control charts are a class of methods that involve the construction of a test statistic based on historical data. The test statistic accumulates all observed counts that exceed the expected count over time, and when this is outside predetermined lower and upper limits an alarm is signaled. Examples include the Shewhart chart (Shewhart, 1931) and the cumulative sum (CUSUM) method, introduced by Page (1954) and later extended to incorporate spatial trends (Rogerson, 1997) and multiple diseases (Woodall and Ncube, 1985). The exponentially weighted moving average (EWMA) charts (Hunter, 1986) were developed to give less weight to older historical data.

These were followed by scan statistics, originally developed by Naus (1965a,b). In their simplest form they incorporate only temporal information; given a time period of length T, a scanning window of fixed length $w < T$ passes over the entire time period and the maximum number of cases in the window is used as the definition for

the test statistic. Significance is determined by comparison to the expected disease count generated from a Poisson process. Kulldorff and Nagarwalla (1995) used this idea to develop a spatial version of the scan statistic, which was later further improved (Kulldorff, 1997) and extended to the spatio-temporal setting (Kulldorff, 2001). The popular SatScan (Kulldorff et al., 2005) is available through a user-friendly software (`http://www.satscan.org/`) and has been extensively used for health care applications (Coleman et al., 2009; Hughes and Gorton, 2013; Sherman et al., 2014).

Model-based methods

Compared with the test statistics, model-based approaches provide a flexible framework for the inclusion of random effects and explanatory variables to adjust the disease risk. The simplest model for disease surveillance is the log linear regression model by Farrington et al. (1996), routinely used by the US CDC for outbreaks detection (Lawson, 2013). The model computes a threshold using recent observations to compare against the current count of disease and declares an alarm when this is exceeded. The class of Generalized Linear Mixed Models (GLMM's) have also been extensively used for the purpose of disease surveillance. Zeger (1988) used a GLMM for the detection of polio counts across space and time in the USA. Kleinman et al. (2004) applied a binomial model to detect bioterrorist attacks, while later this was extended to a Poisson version (Kleinman, 2005) and was applied by Johnson (2008) on west nile virus data.

Within the Bayesian hierarchical setting, a large literature exists for surveillance of communicable diseases, where transmissions of the disease among cases need to be considered. Knorr-Held and Richardson (2003) proposed a model to analyse space-time surveillance data on meningococcal diseases retrospectively. The relative risks in the modelling formulation were decomposed on the log scale into four terms

modelling temporal, seasonal, and spatial trends, describing endemic patterns, and a fourth term describing the potential additional risks in the presence of an epidemic through a hidden Markov model. Zou et al. (2012) proposed a model to analyse the number of cases through a hierarchical model prospectively. The authors include two components in the model; the first represents the baseline distribution of the cases in the absence of an epidemic, following a Poisson distribution, and the second is added to the baseline when an epidemic occurs. Banks et al. (2012) adopted the aforementioned model to analyse symptom data on drug abuse from a range of US hospitals. Similar model-based approaches have been considered in Zhou and Lawson (2008), Corberán-Vallet and Lawson (2014) and Napier et al. (2016).

While some test-based methods may have applications in non-communicable diseases because of their simplicity in the assumptions, the majority of model-based methods were designed specifically for communicable diseases. The models that are typically used for disease surveillance of non-communicable diseases are the ones that were originally developed for disease mapping. Key examples include the works by Abellan et al. (2008) and Li et al. (2012).

Abellan et al. (2008) used a standard binomial logit model to account for spatial and temporal effects through a hierarchical specification. The authors integrated an outbreak detection mechanism by inclusion of a space-time interaction term v_{it} in the linear predictor, where v_{it} follows a mixture of two normal distributions, one with a larger variance than the other, in order to show evidence of unusual

variability in the time pattern of the risk. The model for v_{it} is given by

$$\begin{aligned} v_{it} &\sim \rho \,\text{Normal}(0, \tau_1^2) + (1-\rho)\,\text{Normal}(0, \tau_2^2) && (1.20)\\ \tau_1 &\sim \text{Normal}(0, 0.01)\text{I}(0, \infty) && (1.21)\\ \tau_2 &= \tau_1 + k && (1.22)\\ k &\sim \text{Normal}(0, 100)\text{I}(0, \infty) && (1.23) \end{aligned}$$

where the prior for ρ is uniform on [0, 1].

If z_{it} are latent allocation variables taking values 0 or 1 depending on whether v_{it} comes from Normal(0, τ_1) or Normal(0, τ_2) respectively, then the posterior probabilities $p_{it} = \text{P}(z_{it} = 1|Y_{it})$ are used to classify areas as stable or unstable. Two different classification rules were used to classify an area as unstable, with a different degree of conservativeness. A disadvantage of this modelling approach is that it does not incorporate specific time patterns, for example step changes that could signal the occurence of a new emerging risk factor.

This issue is accommodated by BaySTDetect (Li et al., 2012), a powerful model that showed better performance when compared to the widely used SaTScan (Kulldorff et al., 2005), and has been applied to several studies since then (Li et al., 2014; Duncan et al., 2016). The model is described below in detail as it provides fundamental methodological knowledge related to this thesis.

BaySTDetect model

Following the epidemiological structure on disease mapping, the authors considered a Poisson log-normal model for the disease counts as described in eq. (1.16). The relative risk λ_{it} is modelled as a mixture of two alternative models, a *Common model* and an *Area-Specific model*. The *Common model* assumes one global time trend

for all areas, while the *Area-Specific model* estimates a time trend for each area independently. This aims to distinguish between areas that follow the expected time trend and those that exhibit unusual behaviour.

More specifically, the second level of the hierarchy models λ_{it} as follows

$$\log(\lambda_{it}) = z_i \log(\lambda_{it}^{(C)}) + (1 - z_i) \log(\lambda_{it}^{(AS)}) \quad (1.24)$$

$$\log(\lambda_{it}^{(C)}) = \alpha_0 + \eta_i + \delta_t \quad \textit{(Common model)} \quad (1.25)$$

$$\log(\lambda_{it}^{(AS)}) = \nu_i + \kappa_{it} \quad \textit{(Area-Specific model)} \quad (1.26)$$

where z_i is the mixing parameter that selects between estimates $\lambda_{it}^{(C)}$ from the *Common model* and estimates $\lambda_{it}^{(AS)}$ from the *Area-Specific model*.

The *Common model* follows a standard disease mapping approach where the relative risk λ_{it} consists of a global spatial component η_i and a global temporal component δ_t. The spatial component η_i is assigned a convolution prior, implemented as

$$\eta_i \sim \text{Normal}(v_i, \sigma_\eta^2) \quad (1.27)$$

where v_i is a spatially structured term following a conditional autoregressive prior,

$$v_{1:N} \sim \text{CAR}(\mathbf{W}, \sigma_v^2) \quad (1.28)$$

$\mathbf{W}$ is the adjacency matrix of size $N \times N$ specifying the spatial neighbourhood structure, such that the entries $w_{ij} = 1$ if areas i and j share borders and $w_{ij} = 0$ otherwise. Similarly, the temporal component δ_t follows a Gaussian random walk

prior of order 1 (RW1), which is implemented as

$$\delta_{1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_\delta^2) \tag{1.29}$$

Here, $\mathbf{Q}$ specifies the temporal neighbourhood structure such that $q_{th} = 1$ if $|t - h| = 1$ and $q_{th} = 0$ otherwise, with t and h representing indexes of time. The overall intercept follows a flat prior

$$\alpha_0 \sim \text{Uniform}(-\infty, +\infty) \tag{1.30}$$

The *Area-Specific model* consists of an area-specific intercept that follows a weakly informative prior

$$v_i \sim \text{Normal}(0, 1000) \tag{1.31}$$

and an area-specific temporal component

$$\kappa_{i1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_{i\kappa}^2) \tag{1.32}$$

where $\mathbf{Q}$ is specified as above.

The mixing parameter z_i follows a Bernoulli prior assuming that around 5% of the areas are unusual

$$z_i \sim \text{Bernoulli}(0.95) \tag{1.33}$$

The model is implemented using MCMC sampling, and, at each iteration, the mixing parameter z_i takes values 1 or 0 taking into account which model (*Common* or *Area-specific*) is more likely under the data. At the end of the MCMC sampling, the posterior mean of z_i, namely $f_i = \text{P}(z_i = 1|Y_{it})$, reflects the posterior probability that area i is common or unusual over the whole time period.

A half-normal prior $N(0,1)$ is used for the hyperparameters σ_η^2, σ_v^2, and σ_γ^2, while $\sigma_{i\kappa}^2$ is modelled hierarchically as

$$\log(\sigma_{i\kappa}^2) \sim \text{Normal}(a, b^2) \tag{1.34}$$

where $b \sim N(0, 2.5^2)$, bounded below by 0, and $a \sim N(0, 1000)$. The authors discuss that this reflects the belief that a priori around 10% of areas would have a temporal variability 10 times greater than the expected value, i.e. $\text{P}\big(\sigma_{i\kappa}^2 > 10 \cdot \exp(a)\big) \approx 0.1$.

Multiple testing problem

The multiple testing problem is crucial in detection methods; this arises when multiple tests are performed, and false detections are likely to occur due to pure chance. Benjamini and Hochberg (1995) was the first to introduce controlling of false positives (FDR) as a way to address the multiple testing issue using a frequentist approach. Within the Bayesian setting, FDR rules were later suggested by many authors (Storey, 2003; Newton et al., 2004; Muller et al., 2006; Scott and Berger, 2006; Whittemore, 2007).

BaySTDetect uses the specification suggested by Newton et al. (2004) and later used by Ventrucci et al. (2010) in order to account for multiple testing. The authors base the FDR statistic on the posterior model probability, which represents how likely it is to follow the common trend model, i.e. to exhibit a usual pattern in the risks. Specifically, if $f_i = P(z_i = 1|Y_{it})$ is the posterior frequency that area i follows the common trend, a threshold $C = f(k)$ is chosen where k is the maximum integer such that

$$1/k \sum_{j=1}^{k} f_{(j)} < \alpha \tag{1.35}$$

where $f_{(j)}$ is the jth ordered posterior probability, and α a preset level usually taken

equal to 0.05. A version where the sum of the ordered posterior probabilities is computed rather than the mean can be also considered, leading to more conservative results. The areas with corresponding f_i below threshold C are then classed as unusual. This specification ensures that no more than $(k \times \alpha)$ of the detected areas are false positives.

While the importance of controlling for multiple testing is clear in classical inference, the relevant problem in the Bayesian setting is not always considered to be valid. Gelman et al. (2012) argues that one should not worry about excess false detections but rather not appropriately modelling the data. The authors state that when the data are modelled through a Bayesian hierarchical framework accounting for all possible different sources of uncertainty, the interpretation of the posterior probability is straightforward and there is no need to be adjusted. It is stressed though that in order for the model to provide valid estimates, all known sources of variability need to be accounted for in the modelling framework. For instance, the BaySTDetect model assumes that the area-specific probabilities z_i are modelled independently (eq. (1.33)), not accounting for dependences across the unusual areas. The multiple testing problem therefore arises and the authors selected the rule described in eq. (1.35) to adjust for this.

1.4 Structure of thesis

Chapters 2 to 4 present three studies in the from of self-contained scientific papers, each concerned with one of the three research objectives of this thesis:

Chapter 2

Boulieri, A., S. Liverani, K. d. Hoogh, and M. Blangiardo (2017). A space-time multivariate Bayesian model to analyse road traffic accidents by severity.

Journal of the Royal Statistical Society. Series A: Statistics in Society 180(**1**), 119-139.

Chapter 3

Boulieri, A., A. L. Hansell, and M. Blangiardo (2016). Investigating trends in asthma and COPD through multiple data sources: A small area study. *Spatial and spatio-temporal epidemiology 19*, 28-36.

Chapter 4

Boulieri, A., J. Bennett, and M. Blangiardo. A mixture modelling approach for chronic disease surveillance. *Biostatistics* (under revision)

Chapter 2 is concerned with the public health issue of road safety. With the aim to provide disease mapping, I develop a statistical model within the Bayesian hierarchical setting to analyse traffic accidents at the ward level in England over the years 2005 to 2013. The novel part of this work is the joint modelling of road traffic accidents of low and high severity, which allows borrowing of strength between the two severity levels thus leading to more robust estimations. To investigate the underlying dependency structure, I develop a number of alternative models and carry out an extensive model comparison. Aiming at also identifying potential areas or increased risk, I make use of posterior probability maps of extreme rates to suggest the presence of hot spots in a public health perspective.

In chapter 3, I employ BaySTDetect, described in eqs. (1.24) to (1.34), to explore the potential of the model in epidemiological studies for formal detection of unusual patterns. BaySTDetect is used to investigate CRDs by analysing multiple data sources on asthma and COPD to help understanding the relationships between disease prevalence, morbidity and mortality. In specific, GP drug prescriptions, hospital admissions, and deaths are analysed at the clinical commissioning group (CCG) level in England from August 2010 to March 2011 (8 months), thus providing

a more complete picture of the disease.

The aforementioned study leads to the identification of several limitations of BaySTDetect which forms the motivation for the work presented in chapter 4. Here, I develop an improved version of BaySTDetect for public health surveillance. The proposed model is able to provide estimates of the disease risk in space and time, and also to detect areas with unusual behaviour, while it is specifically designed to deal with a range of spatial and temporal patterns in the data, and with time series of different lengths. An extensive simulation study is performed to assess its performance under different scenarios, and comparisons against BaySTDetect are presented. The proposed model is also applied to i) a set of road traffic accidents data at the district level in England over the years 2005 to 2015, and ii) COPD hospital admissions at the CCG level in England from April 2010 to March 2011.

Finally, chapter 5 presents overall conclusions and future directions for this very active area of research.

1.4.1 Provision of data

Road traffic accidents data used in chapters 2 and 4 are publicly available and were accessed from the Department for Transport (Road Traffic Statistics Branch). Likewise, GP drug prescriptions, or Prescription Cost Analysis (PACT) data used in chapter 3, were accessed from the NHS Business Services Authority. Population data used in chapters 3 and 4 were obtained from the Office for National Statistics (ONS).

Mortality and Hospital Episode Statistics (HES) admissions data used in chapters 3 and 4 were obtained from the UK Small Area Health Statistics Unit (SAHSU) at Imperial College London. Mortality data were supplied by ONS and HES admission data by the Health and Social Care Information Centre. The work of SAHSU

is funded by PHE as part of the MRC-PHE Centre for Environment and Health, funded also by the UK Medical Research Council. SAHSU holds approvals from the National Research Ethics Service - reference 12/LO/0566 and 12/LO/0567 - and from the Health Research Authority Confidentially Advisory Group (HRA-CAG) for Section 251 support (HRA - 14/CAG/1039). Hospital Episode Statistics data are copyright © 2014, re-used with the permission of the Health and Social Care Information Centre. All rights reserved.

1.4.2 Contributions of co-authors

I was the lead author of all three papers and performed all required analyses. Marta Blangiardo provided guidance and oversaw research for all papers. Silvia Liverani provided advice on statistical methodology and aided by reviewing and editing the final document for the first paper. Kees de Hoogh supplied the offset data used in the first paper. Anna Hansell offered epidemiological advice, facilitated the provision of health data and oversaw work for the second paper. James Bennett contributed to technical discussions and reviewed the final document for the third paper.

2 A multivariate Bayesian model for road traffic accidents data in England

2.1 Introduction

Road traffic accidents are considered to be a major public health issue, with consequences similar to those of cancer, cardiovascular and other non-communicable diseases. According to the annual reports of casualties in Great Britain by the Department for Transport for years 2005 to 2013, on average, 2,400 people die on Britain's roads every year, making it the leading cause of mortality for ages 15-34. At the same time, 222,000 non-fatal accidents occur every year causing immediate and later physical, social and psychological consequences to those involved.

Road traffic accidents have an intrinsic spatial structure which is important to take into account properly in order to be able to identify areas of particularly high risk. In addition, a temporal trend could also be identified when data on multiple time points (e.g. years) are considered. The aim of this work is to develop a statistical framework to take into account the spatio-temporal structure of road traffic accidents and to highlight areas characterised by excess risk, to inform future health policy strategies. Policy makers can then implement measures for accidents prevention and prioritise expenditure by focusing on 'hot spots', i.e. areas which are shown to be more dangerous.

Both classical and Bayesian methods have been used to deal with road traffic accidents data analysis. In the context of classical statistics, Poisson regression techniques, which are suitable for count data, have been used by many authors in the past. For instance, Miaou and Lum (1993) compared this technique to the conventional linear regression to assess accidents and highway geometric design relationships, while Jovanis and Chang (1986) used Poisson regression to assess the effects of travel mileage on accident occurrence. Other examples of previous work include Kim and Yamashita (2002) and Graham and Glaister (2003) who focused on the association between road traffic accidents and potential risk factors such as driver behaviour and environmental or roadway features.

In order to relax the overdispersion constraint imposed by the Poisson model, several authors have used negative binomial regression models (Miaou, 1994; Shankar et al., 1995; Noland and Quddus, 2004). Other alternatives exist, such as zero-inflated models which are appropriate for data that exhibit excess zeros (Miaou, 1994). Yet, all the above methods have serious limitations, mainly because they give unstable estimates due to the large variability from one area to another, especially when the population size and/or the geographical scale of the analysis is small.

The Bayesian hierarchical framework is a potential valid alternative, being more flexible and able to handle data with low counts, and also to easily account for spatial correlation. As discussed in section 1.2 on page 6, Bayesian hierarchical methods facilitate smoothing by borrowing information from neighbouring units, an essential point in case of low counts, since it leads to more stable estimates (Ghosh et al., 1998; Maiti, 1998). Bayesian hierarchical models have been applied to road traffic accidents by many authors (Miaou et al., 2003; MacNab, 2004; Qin et al., 2004; Torre et al., 2007). The Poisson log-normal model, including random effects specified through the 'Besag-York-Mollie' (BYM) structure (Besag et al., 1991) (described in

section 1.3.1 on page 12) has been shown to be the most appropriate for road traffic accidents analysis (Aguero-Valverde and Jovanis, 2006; Quddus, 2008a).

A key issue with the Poisson log-normal model is the choice of the offset. Typically in disease mapping the expected number of cases is used, but this is not feasible in the context of road traffic accidents. Several authors have focused on alternative ways of obtaining a proxy of the population at risk. Various examples are based on total population, numbers of registered vehicles, or numbers of licensed drivers, but these are generally recognized as poor surrogates for the actual amount of accident risk, and as mentioned by Wolfe (1982), the most easily obtained offset measures are rarely the most desirable ones.

Traffic volume seems to be the most appropriate offset for road traffic accidents analysis (Aguero-Valverde, 2011). To approximate this to areal-level, Fridstrom et al. (1995) used petrol sales, while Eksler (2008) used information of traffic counts for each region multiplied by the corresponding road length. Alternatively, the levels of population and employment density were used to represent travel volume by Graham and Glaister (2003) and also used by Noland and Quddus (2004). Quddus (2008a), on the other hand, used a function of the number of registered cars as an offset, while the authors accounted for traffic volume by including this as an explanatory variable (covariate) in the Poisson model. Jones et al. (2008) and Ackaah and Salifu (2011) also included indicators of traffic volume as explanatory variables in their models.

Extended research has focused on the modelling of road traffic accidents using time series models which are able to overcome dependency issues in time-related data. The 'autoregressive integrated moving average' (ARIMA) model, introduced by Box and Jenkins (1976) has been used to model time series count data in many applications over the last few decades (Van den Bossche et al., 2006; Goh, 2005),

including road traffic accidents data (Houston and Richardson, 2002; Friedman et al., 2007). The 'integer-valued autoregressive' (INAR) model initially introduced by McKenzie (1985) and further studied by Al-Osh and Alzaid (1987) was applied to road traffic accidents in Great Britain by Quddus (2008b) showing important improvements over the ARIMA specification. The 'demand for road use, accidents and their gravity' (DRAG) approach is a three step approach, that considers risk exposure, accident rate and its severity, and has also been extensively used (Gaudry et al., 1993; Tegnér and Loncar-Lucassi, 1997). In state space models, also known as structural time series models or unobserved components models, the three important road safety components, i.e. exposure, accidents, and accident severity, can be modelled simultaneously. Examples of multivariate state space models in the area of road safety can be found in Bijleveld et al. (2010), and Durbin and Koopman (2012).

In the scientific literature, a separate univariate analysis of road traffic accidents is usually carried out per severity level (slight, severe and fatal) identifying different patterns and estimates for each category (Aguero-Valverde and Jovanis, 2006; Quddus, 2008a; Jones et al., 2008). However, it is reasonable to assume that accident severities are not independent but instead there is some degree of correlation among them and therefore, a potential statistical problem arises when this is not taken into account in the modelling framework. This issue was studied by Bijleveld (2005) who showed that there is a need for multivariate modelling in the road traffic accidents analysis. Joint approaches were attempted by Ma and Kockelman in a series of papers via multivariate Poisson (MVP) and multivariate Poisson-lognormal (MVPLN) models (Ma and Kockelman, 2006; Ma et al., 2008). The latter method was also used by Park and Lord (2007) and Aguero-Valverde and Jovanis (2009). All these studies suggested interactions among accident severities, thus showing

advantages of the joint specification, however none of them accounted for spatial and temporal correlation.

Very limited research has combined ideas from a multivariate setting and spatial analysis for road traffic accidents data. Song et al. (2006) proposed a number of priors for Bayesian multivariate spatial models, while Wang and Kockelman (2013) analysed spatial pedestrian crash data through a Poisson log-normal multivariate model. Miaou and Song (2005) and Wang et al. (2011) considered both spatial and temporal dependences in road traffic accidents data, the former through a generalised linear mixed model (GLMM), and the latter through a two-stage mixed model.

In this chapter, I develop a Bayesian multivariate statistical framework to jointly model road traffic accidents by severity level, while accounting for spatial and temporal correlation in the data. The spatio-temporal approach from Knorr-Held and Besag (1998) is used as a baseline, while the multivariate conditional autoregressive (MCAR) formulation initially proposed by Mardia (1988) is incorporated to capture dependences across space and between accident severities. The temporal correlation is modelled through a random walk independently for each severity level. A series of alternative models are presented for comparisons in order to assess the performance of the proposed approach. The data consist of road traffic accidents at electoral ward level in England for the years 2005 to 2013. Informative maps of England are also produced, allowing to visualise the pattern of the accident rates and identify areas with elevated risk.

This chapter is structured as follows. In section 2.2 I describe the data used for the analysis, and in section 2.3 the statistical methodology. Section 2.4 presents the results of the analysis and conclusions are discussed in section 2.5.

2.2 Data

2.2.1 Road traffic accidents data

In Great Britain, every road accident on the public highway, which includes human injury or death, is recorded on a STATS19 report form by police officers (Department for Transport, 2010). This form collects a range of information such as time and location, road condition, behaviour of the driver as well as the vehicles that were involved and the severity level of the accident. The STATS19 form is completed at either the scene of the accident, or when the accident is reported to the police. Although this is not a complete record of all casualties, as there might be a small proportion of minor accidents not reported, especially when no human injury was incurred, STATS19 data provide the most detailed and reliable available source on accidents. The Department for Transport has overall responsibility for the design and the collection system of the STATS19 data, which are publicly available.

For each accident the location and its severity are available, which can take one of three values: *slight*, *severe* or *fatal*. An accident is classed as fatal when a death occurred within a month of the collision and severe when a hospital treatment is required. Otherwise it is classed as slight. Road traffic accidents data for England over the years 2005 to 2013 were used for the analysis. I aggregated the accidents at the electoral ward level (7,932 in England) by severity for each year, considering fatal and severe in the same category, as fatalities account for an average of 0.014% of the total number of accidents for each year. In the rest of the chapter, *low severity* refers to slight accidents, while *high severity* refers to severe or fatal ones.

A number of around 199,000 accidents occurred in England in 2005 with a decrease down to 139,000 in 2013. The majority of them were of low severity -

around 85% of the total number of accidents for all years. Table 2.1 presents the summary statistics of the accidents data at ward level.

Year	Min	1st Qu.	Median	Mean	3rd Qu.	Max	%0s
Low severity							
2005	0	7.00	14	19.19	25.00	376	0.73
2006	0	7.00	13	18.06	23.00	330	0.77
2007	0	7.00	12	17.37	22.00	291	0.82
2008	0	6.00	11	16.24	21.00	307	1.16
2009	0	6.00	11	15.63	20.00	269	0.92
2010	0	5.00	10	14.87	19.00	263	1.27
2011	0	5.00	10	14.53	18.00	313	1.60
2012	0	5.00	9	13.89	18.00	301	1.70
2013	0	5.00	9	13.26	17.00	302	2.18
High severity							
2005	0	1.00	2	3.07	4.00	53	16.81
2006	0	1.00	2	3.05	4.00	60	16.91
2007	0	1.00	2	2.98	4.00	64	17.26
2008	0	1.00	2	2.75	4.00	59	19.16
2009	0	1.00	2	2.62	4.00	59	20.01
2010	0	1.00	2	2.44	3.00	45	22.29
2011	0	1.00	2	2.50	3.00	45	20.88
2012	0	1.00	2	2.50	3.00	57	21.18
2013	0	1.00	2	2.35	3.00	50	21.84

Table 2.1: Summary statistics for counts of road traffic accidents at ward level by severity and year in England.

2.2.2 Traffic volume data

The annual average daily flow (AADF) data for the major road network (motorways and A-roads) in England were obtained from the Department for Transport. These are the traffic counts for each junction-to-junction link on the major roads (980 and 16,941 for respectively motorways and A-roads) containing spatial information in the form of a postcode. In order to estimate the ward-level traffic volume, used as an offset for the analysis, the following protocol was followed: The AADF dataset was joined to the major road network shapefile, also obtained from the Department for Transport. Each AADF point was associated to a road segment based on unique road names, or, if not available, based on distance. For the road segments with no corresponding AADF count, this was estimated as the average AADF of the bordering road segments, and in the few occasions when this was not possible the middle value of AADF was used (Eeftens et al., 2012; de Hoogh et al., 2013). The resulting shapefile, containing information on all road segments in England with their corresponding AADF values was intersected with the wards 2001 shapefile. The traffic volume of each intersected road segment was then calculated by multiplying the length of the road segment by its AADF value. Finally, the ward level traffic volume was estimated by summing the traffic volumes of all road segments lying within the ward:

$$\mathrm{TV_W} = \sum_{\mathrm{RS}\in\mathrm{W}} \mathrm{TV_{RS}} \tag{2.1}$$

$$\mathrm{TV_{RS}} = \mathrm{length(RS)} \times \mathrm{AADF_{RS}} \tag{2.2}$$

where TV is the traffic volume, W represents the ward level, and RS represents the road segment.

The AADF counts are very stable across all years considered in the study (2005

to 2013), and hence the traffic volume dataset was calculated based on AADF counts of the middle year 2009. The traffic volume dataset was created using Geographic Information Systems (GIS).

2.3 Statistical framework

The analysis is carried out within a Bayesian hierarchical framework. Accidents of low and high severity are modelled jointly using as a baseline the formulation by Knorr-Held and Besag (1998) (eqs. (1.16) to (1.17) on page 16). I develop a series of models under various assumptions regarding the spatial and temporal dependences between low and high accident severities. The models are classified into three main groups according to the type of spatial dependence between the two severities: *Independent* space effects assume independence between the spatial structure of low and high severities, *Common* space effects assume the same spatial structure for the two severities, while *Correlated* space effects assume a certain degree of dependence in the spatial random effects. In all groups the temporal dependence is modelled via a random walk either independent or common for the severity levels. Table 2.2 presents the specification of the developed models, and table 2.3 shows how the spatial and temporal random effects are introduced, to help understand the hierarchy. All models are described in detail in the rest of this section.

2.3.1 Independent space effects models

I will first present the general modelling approach assuming a complete independence between the two levels of accident severities (Model A in table 2.2), as each parameter in the model is severity-specific (low and high).

In the first level of the hierarchy, the observed counts of accidents $Y_{it}^{(j)}$ are

Model	Model specification	Spatial component	Temporal component
Independent space effects			
A	$\alpha^{(j)} + \text{BYM} + \delta_t^{(j)}$	$\text{BYM} = \theta_i^{(j)} + \phi_i^{(j)}$	$\delta_t^{(j)} \sim \text{RW1}$
B	$\alpha^{(j)} + \text{BYM} + \delta_t$	$\theta_i^{(j)} \sim \text{Normal}$ $\phi_i^{(j)} \sim \text{CAR}$	$\delta_t \sim \text{RW1}$
Common space effects			
C	$\alpha^{(j)} + \text{BYM} + \delta_t^{(j)}$	$\text{BYM} = \theta_i + \phi_i$	$\delta_t^{(j)} \sim \text{RW1}$
D	$\alpha^{(j)} + \text{BYM} + \delta_t$	$\theta_i \sim \text{Normal}$ $\phi_i \sim \text{CAR}$	$\delta_t \sim \text{RW1}$
Correlated space effects			
E	$\alpha^{(j)} + \text{MBYM} + \delta_t^{(j)}$	$\text{MBYM} = \Theta_i + \Phi_i$	$\delta_t^{(j)} \sim \text{RW1}$
F	$\alpha^{(j)} + \text{MBYM} + \delta_t$	$\Theta_i \sim \text{MNormal}$ $\Phi_i \sim \text{MCAR}$	$\delta_t \sim \text{RW1}$

Table 2.2: Models under different assumptions for spatial and temporal dependences between low and high accident severities.

modelled as

$$Y_{it}^{(j)} \sim \text{Poisson}(\lambda_{it}^{(j)} \, \text{Off}_i) \tag{2.3}$$

for ward i, time point t and severity j. There are $N = 7,932$ wards and 9 time points $t = 1, 2, \ldots, 9$ which correspond to the years $2005, 2006, \ldots, 2013$ respectively. The severity level is low or high if $j = 1$ or $j = 2$ respectively. The accident rate is denoted by $\lambda_{it}^{(j)}$ and the offset variable by Off_i, which here is taken as the traffic volume described in section 2.2.

The second level of the hierarchy models the parameter $\lambda_{it}^{(j)}$. In the disease mapping context this is usually referred to as a relative risk, however in this example, where the offset is the traffic volume and not an expected count of accidents, it is interpreted as an accident risk. It is a function of a spatially unstructured component

	Spatially structured effects			Spatially unstructured effects			Temporal effects	
	$\theta_i^{(j)}$	θ_i	Θ_i	$\phi_i^{(j)}$	ϕ_i	Φ_i	$\delta_t^{(j)}$	δ_t
Model A	✓			✓			✓	
Model B	✓			✓				✓
Model C		✓			✓		✓	
Model D		✓			✓			✓
Model E			✓			✓	✓	
Model F			✓			✓		✓

Table 2.3: Hierarchy of the spatial and temporal random effects for the developed models.

$\theta_i^{(j)}$, a spatially structured component $\phi_i^{(j)}$, and a temporally structured component $\delta_t^{(j)}$. A severity-specific intercept $\alpha^{(j)}$ is also included in the model.

$$\log(\lambda_{it}^{(j)}) = \alpha^{(j)} + \underbrace{\theta_i^{(j)} + \phi_i^{(j)}}_{\text{BYM}^{(j)}} + \delta_t^{(j)} \tag{2.4}$$

In the third level, an exchangeable prior is assigned to the spatially unstructured random effects

$$\theta_i^{(j)} \sim \text{Normal}(0, \sigma_\theta^{2(j)}), \tag{2.5}$$

while the spatially structured random effects are assigned a conditional autoregressive prior

$$\phi_i^{(j)} \sim \text{CAR}(\mathbf{W}, \sigma_\phi^{2(j)}) \tag{2.6}$$

The full specification of the CAR prior is described in eq. (1.12) on page 15. The

parameters $\sigma_\theta^{2(j)}$ and $\sigma_\phi^{2(j)}$ represent the corresponding variances for the spatially unstructured and spatially structured random effects. For the temporal component a standard random walk prior of order 1 is used

$$\delta_t^{(j)} \sim \text{RW1} \tag{2.7}$$

Similarly, the full specification of the RW1 prior is described in eq. (1.19) on page 17. A gamma prior $G(0.5, 0.0005)$ is assigned to the precisions $\tau_\theta^{(j)} = 1/\sigma_\theta^{2(j)}$, $\tau_\phi^{(j)} = 1/\sigma_\phi^{2(j)}$ and $\tau_\delta = 1/\sigma_\delta^{2(j)}$, controlling for variations conditional on the neighbouring spatial and temporal points. A flat prior is finally assumed for the intercept α.

In the group of *Independent* space effects, Model B is also included, differing from Model A only in the temporal effects $\delta_t \sim \text{RW1}$, which in this case are taken to be common between the two severities.

2.3.2 Common space effects models

Following the Poisson log-normal specification described in eqs. (2.3) to (2.7), I present two additional models that assume a common spatial effect for accidents of low and high severity level. This is implemented in the model through a common distribution (not severity-specific). The spatially unstructured effects θ_i are assigned a common normal distribution and the spatially structured effects ϕ_i are assigned a common CAR distribution. This means dropping the j superscript in eqs. (2.5) and (2.6). The temporal random effects are specified either as independent (Model C) or common (Model D). The prior specification of the parameters is the same as in section 2.3.1.

2.3.3 Correlated space effects models

Finally, I extend the BYM specification to the multivariate BYM (MBYM) which considers a multivariate setting across severities for both spatially structured ϕ_i and unstructured θ_i effects quantifying their correlation.

The spatially structured effects ϕ_i follow an MCAR prior. Consider a two-dimensional vector of spatially structured random effects $\Phi_i = (\Phi_{i1}, \Phi_{i2})$ for each area $i = 1, \ldots, N$, where the subscripts 1 and 2 represent low and high severity respectively. The CAR specification in eq. (1.12) on page 15 is then extended as follows

$$\Phi_i \mid \Phi_{(-i)} \sim \text{Normal}\big(\bar{\Phi}_i, \frac{\Sigma_\Phi}{n_i}\big) \tag{2.8}$$

$$\bar{\Phi}_i = (\bar{\Phi}_{i1}, \bar{\Phi}_{i2}) \tag{2.9}$$

$$\bar{\Phi}_{ip} = \frac{\sum_{k \in D_i} \Phi_{kp}}{n_i}, \quad p = 1 \text{ or } 2 \tag{2.10}$$

where $\Phi_{(-i)} = (\Phi_{1(-i)}, \Phi_{2(-i)})$ denotes the elements of the matrix Φ excluding area i. As in the univariate CAR case, D_i and n_i are the set of areas adjacent to area i and the number of those areas, respectively. Σ_Φ is the 2×2 covariance matrix with diagonal elements $\sigma_\phi^{2^{(1)}}$ and $\sigma_\phi^{2^{(2)}}$ representing the conditional variances for each severity level. The within-area conditional correlation of random effects of low and high severity ρ_ϕ is modelled via the off-diagonal terms

$$\Sigma_\Phi = \begin{pmatrix} \sigma_\phi^{2^{(1)}} & \rho_\phi \sigma_\phi^{(1)} \sigma_\phi^{(2)} \\ \rho_\phi \sigma_\phi^{(1)} \sigma_\phi^{(2)} & \sigma_\phi^{2^{(2)}} \end{pmatrix} \tag{2.11}$$

Hence, MCAR not only takes into account the spatial correlation among areas for each severity, but also allows for correlation between severities in each spatial unit.

For each area $i = 1, \ldots, N$ a two-dimensional vector of spatially unstructured random effects $\Theta_i = (\Theta_{i1}, \Theta_{i2})$ is also included in the model, where again the subscripts 1 and 2 represent low and high severity respectively. These are assigned a multivariate normal distribution

$$\Theta_i \sim \text{MVNormal}(\mathbf{0}, \Sigma_\Theta) \tag{2.12}$$

where $\mathbf{0}$ is a 1×2 vector and Σ_Θ is a 2×2 covariance matrix

$$\Sigma_\Theta = \begin{pmatrix} \sigma_\theta^{2(1)} & \rho_\theta \sigma_\theta^{(1)} \sigma_\theta^{(2)} \\ \rho_\theta \sigma_\theta^{(1)} \sigma_\theta^{(2)} & \sigma_\theta^{2(2)} \end{pmatrix} \tag{2.13}$$

with diagonal elements $\sigma_\theta^{2(1)}$ and $\sigma_\theta^{2(2)}$ representing the conditional variances for each severity, while the off-diagonal elements ρ_θ model the within-area conditional correlation of random effects of low and high severity. In addition to ρ_ϕ and ρ_θ, the parameter ρ_{tot} which is the within-area conditional correlation for the total random effect (i.e. sum of spatially structured and unstructured components).

The precision matrix $\Omega_\Phi = \Sigma_\Phi^{-1}$ is assigned a Wishart(A, k) prior where A and k denote the inverse scale matrix and the degrees of freedom respectively. The parameter k is set to 2 for a weakly informative specification. The entries of the inverse scale matrix A are typically assigned reasonable values in advance, as recommended by Gelman (2006). Following Kass and Natarajan (2006), the observed data are used to obtain empirical estimates of matrix A, thus setting the entries on the diagonals to 500 and the off-diagonal entries to 0.0005. Although such external information does not usually bias the main estimates, it may have some influence on the precision of the estimates, and it is important to explore this through sensitivity analysis. Similarly to the spatially structured case, a Wishart(A, k) prior is assigned

to the precision matrix $R_\Theta = \Sigma_\Theta^{-1}$, with the same parameters k and A. To evaluate the robustness of the posterior estimates using different priors, the diagonals of the Wishart distribution on the MCAR and MVNormal priors are changed to 1 and 1000, while the off-diagonals remain 0.0005. The prior specification for the remaining components are as defined in section 2.3.1.

The MBYM specification described in this section, where MBYM= Φ_i + Θ_i, is then coupled with independent time effects $\delta_t^{(j)}$ in Model E and common time effects δ_t in Model F.

2.3.4 Model comparison and checking

The Deviance Information Criterion (DIC) (Spiegelhalter et al., 1998) has been extensively used for comparisons among hierarchical models in which the number of parameters is not clearly defined. This is a generalisation of Akaike's Information Criterion (AIC), as it trades off model fit against a measure of complexity. The model fit is measured by the posterior mean of deviance $\bar{\mathrm{D}}$, which has often been used alone to compare models informally (Dempster 1997; Richardson and Green 1997), and the penalty for complexity is given by the number of effective parameters p_D. Similarly to AIC, lower values indicate better performance. The formula of DIC is given by:

$$\mathrm{DIC} = \bar{\mathrm{D}} + \mathrm{p_D} = \mathrm{D}(\bar{\theta}) + 2\mathrm{p_D} \tag{2.14}$$

where $\mathrm{D}(\bar{\theta})$ is the deviance evaluated at the posterior means of the parameters $\bar{\theta}$. Similarly to AIC, lower values indicate better performance.

Although DIC is a useful tool for comparisons among different models, it does not give any information as to whether a specific model is adequate. To assess this, I make use of posterior predictive checks and Bayesian p-values, as suggested by

Lunn et al. (2012). A predictive distribution of accident counts Y^{rep} is generated for each severity level, year, and area under each model specified in table 2.2. These predictions are then compared to the observed data. If the model fits the data adequately, the replicated data should look similar to the observed data. In addition to this graphical check, a Bayesian p-value is also estimated, interpreted as the predictive probability of getting an extreme result:

$$p_{Bayes} = \Pr\left(\mathrm{T}(Y^{rep}, \lambda) > \mathrm{T}(Y, \lambda) \mid Y\right) \tag{2.15}$$

$\mathrm{T}(Y, \lambda)$ is a test statistic, and here the common choice is used: $\mathrm{T}(Y, \lambda) = Y$ and $\mathrm{T}(Y^{rep}, \lambda) = Y^{rep}$ to check for individual outliers. A p_{Bayes} value close to 0.5 suggests that the generated data are compatible to the model, whereas values close to 0 and 1 are considered extreme, suggesting poor fit.

2.3.5 Posterior distribution of accident rates

The posterior accident rates $\lambda_{it}^{(j)}$ are displayed in the form of maps in order to visualise their spatial patterns, and also to assess the degree of smoothing by comparing those against maps of crude rates. With the aim to highlight dangerous areas, I also make use of posterior probability maps. By borrowing ideas from the concept of ranking, extensively used in road traffic accidents analysis (Christiansen et al., 1992; Schlüter et al., 1997; Miranda-Moreno et al., 2005; Aguero-Valverde and Jovanis, 2007), I calculate the posterior probability that the spatial component of the accident rates $\left(\exp(\theta_i + \phi_i)\right)$ is ranked among the top 800, accounting for roughly 10% of the total number of areas (7,932), for each severity level j. This is given by

$$\Pr\Big(\mathrm{rank}\big(\exp(\theta_i + \phi_i)^{(j)}\big) < 800\Big). \tag{2.16}$$

In addition, the rank order of areas based on crude rates (averaged across years) is compared against the rank order of areas under the model in order to highlight potential differences and evaluate the impact of the modelling approach. To estimate the latter one, the posterior mean of the rank (Miaou and Song, 2005; Tunaru, 2002) of the spatial component $\big(\exp(\theta_i + \phi_i)\big)$ is used.

2.4 Results

R was used for the data manipulation and the plots, and models were implemented in OpenBUGS. Two chains were run for each parameter for each model with different initial values for around 50,000 iterations depending on the complexity of the model, from which 5,000-10,000 are discarded as a burn-in, and estimates were based on the remaining samples. The simulations took between 20 to 27 hours per model on an Intel Core processor 3.40GHz with 8GB RAM. The convergence diagnostics used include visual check of trace plots, BGR statistic, autocorrelation plots and Monte Carlo error which should be less than 5% of the posterior standard deviation.

Analysing the DIC values of the models in table 2.4 allows us to make several observations regarding their fit to the road traffic accidents data.

Regarding the spatial effects, which are of primary interest in this paper, it is shown that *Common* space effects provide the worst fit, with high relative differences in the DIC than *Independent* space effects that follow.

The benefit of including a multivariate structure in the spatial effects (MBYM specification) can be seen in the *Correlated* space effects, where the DIC decreased greatly. This indicates that there is a non-negligible correlation between low and high severities, in both spatially structured and unstructured effects which needs to be considered in the model. I have also developed a model with a multivariate

structure either in ϕ_i or in θ_i alone, followed by an independent structure for the other parameter, in order to further investigate the complex structure of the data. However, these models did not provide any important improvement, indicating that it is the synergy of the multivariate structure on ϕ_i and θ_i that best supports the data. In addition, the fact that Model E is favoured over Model F suggests that there are differences in the temporal trends between low and high severities over the period 2005 to 2013.

Independent		Common		Correlated	
A	B	C	D	E	F
527,400	540,100	542,300	543,200	**392,900**	413,300

Table 2.4: DIC values for the competing models; the best model is shown in **bold**.

Table 2.5 reports the parameter estimates of all the models that are considered in this study. In general, it can be observed that most of the estimates are consistent among the different models. The overall mean accident rate is around 0.19 for low severity and 0.03 for high severity. As expected, some differences in the variances for the spatially structured and unstructured effects exist between the models of different groups. More specifically, the *Independent* space effects give an average of 0.85 and 0.63 for the variability in ϕ_i for low and high severity respectively, while this is somewhere between the two values under the *Common* space effects models. Similarly, the average of the corresponding variability in θ_i is 0.09 for low severity and 0.16 for high severity under the *Independent* space effects, and somewhere in between under the *Common* space effects models.

Under the *Correlated* space effects models, the variability in ϕ_i decreases by around 0.2 for low severity and 0.1 for high severity, while at the same time, the variability in θ_i increases by around 0.3 for both severities, which is important

Parameter	Independent space effects models				Common space effects models				Correlated space effects models			
	A		B		C		D		E		F	
	Low severity	High severity	Low severity	High severity	Low severity	High severity	Low severity	High severity	Low severity	High severity	Low severity	High severity
Intercept												
$\exp(\alpha)$	0.192 (<.001)	0.032 (<.001)	0.192 (<.001)	0.032 (<.001)	0.196 (<.001)	0.033 (<.001)	0.196 (<.001)	0.033 (<.001)	0.192 (<.001)	0.033 (<.001)	0.192 (<.001)	0.033 (<.001)
Variances												
$\tilde{\sigma}^2_\phi$	0.854 (0.014)	0.630 (0.017)	0.850 (0.014)	0.630 (0.016)	0.790 (0.013)		0.795 (0.013)		0.641 (0.013)	0.534 (0.013)	0.647 (0.014)	0.537 (0.014)
σ^2_θ	0.089 (0.011)	0.159 (0.014)	0.095 (0.012)	0.157 (0.013)	0.105 (0.012)		0.095 (0.012)		0.414 (0.011)	0.415 (0.011)	0.411 (0.011)	0.413 (0.011)
σ^2_δ	0.0028 (0.002)	0.0028 (0.002)	0.0026 (0.002)		0.0027 (0.002)	0.0028 (0.002)	0.0027 (0.002)		0.0028 (0.002)	0.0028 (0.002)	0.0027 (0.002)	
Correlations												
ρ_ϕ	–		–		–		–		0.769 (0.009)		0.722 (0.009)	
ρ_θ	–		–		–		–		0.643 (0.010)		0.641 (0.010)	
ρ_{tot}	–		–		–		–		0.740 (0.007)		0.741 (0.007)	
Spatial fraction												
frac_ϕ	0.906 (0.012)	0.797 (0.017)	0.900 (0.013)	0.801 (0.017)	0.883 (0.013)		0.716 (0.014)		0.646 (0.009)	0.602 (0.010)	0.650 (0.009)	0.606 (0.010)

Table 2.5: Mean posterior estimates for the parameters of all the models developed in the study; standard deviations are in parentheses.

considering the small standard deviations associated with these estimates. Potentially, this is because part of the unstructured variability σ_θ^2 in the data can only be captured under a multivariate specification. The impact of including a multivariate setting for both ϕ_i and θ_i is also notable in the reduction in uncertainty associated with the variance estimates due to the information that is borrowed between the two severities. The larger uncertainty in the *Independent* case is due to the correlation between severities which is not taken into account in the model. The spatial fraction changes across groups of models as this is calculated based on the variances, with the estimates becoming slightly more precise under *Correlated* space effects.

Within the *Correlated* group of models, and by looking at the temporal effects σ_δ, it can be observed that their magnitude seems to be the same for low and high severities, however, as Model E is favoured over Model F, different patterns in the temporal trend between the two severities are suggested.

Focusing on the preferred model (Model E) in table 2.5, several conclusions can be made regarding the sources of variability in road traffic accidents data. A comparison of $\hat{\sigma}_\phi^2$ and σ_θ^2 indicates that the spatially structured effects are somewhat stronger than the spatially unstructured for both low and high severities. This can also be seen by the values of the spatial fraction (frac_ϕ), which is distributed around 0.646 for accidents of low severity, and 0.602 for accidents of high severity, indicating that although the impact of spatially structured effects is stronger in both categories, the unstructured effects are also non-negligible. The conditional within-area correlations for ϕ_i and θ_i are 0.77 and 0.64 respectively, and the corresponding correlation due to the total random effects is 0.74. Figure 2.1 plots the temporal effects on the exponential scale $(\exp(\delta_t^{(j)})$ to show the pattern of the accident risk across time. For low severity an almost linear downward pattern is observed, while for high severity the trend becomes flatter after 2010. The findings overall suggest

that the present data have a complex structure as the variation is due to several sources of variability; spatially structured, spatially unstructured, and temporal random effects, as well as the correlation between severities are needed to explain this.

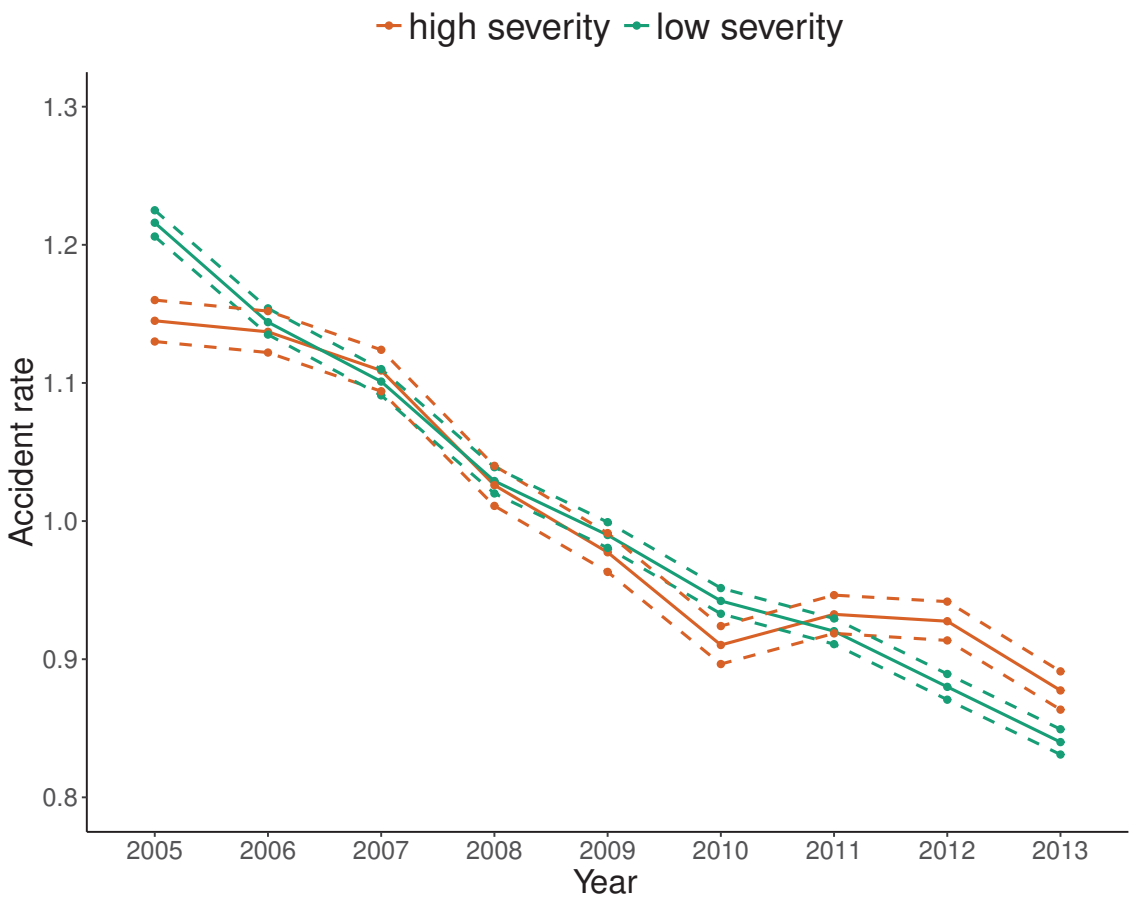


Figure 2.1: Temporal trend of the accident rate in England over 2005-2013 under Model E; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values.

Predicted data are plotted against observed data in figs. A.1 and A.2 in appendix A, suggesting an adequate model fit overall, with the corresponding variability being clearly larger for high severity than for low severity. Table 2.6 shows the proportion of areas with extreme p_{Bayes} values by year and severity level, and the

corresponding mean values under Model E. An adequate model fit is suggested, as only 13% to 17% of areas are assigned extreme p_{Bayes} values, while these are close to 0.5, varying from 0.55 to 0.63. This is confirmed by maps of p_{Bayes} values which can be found in figs. A.3 and A.4 in appendix A.

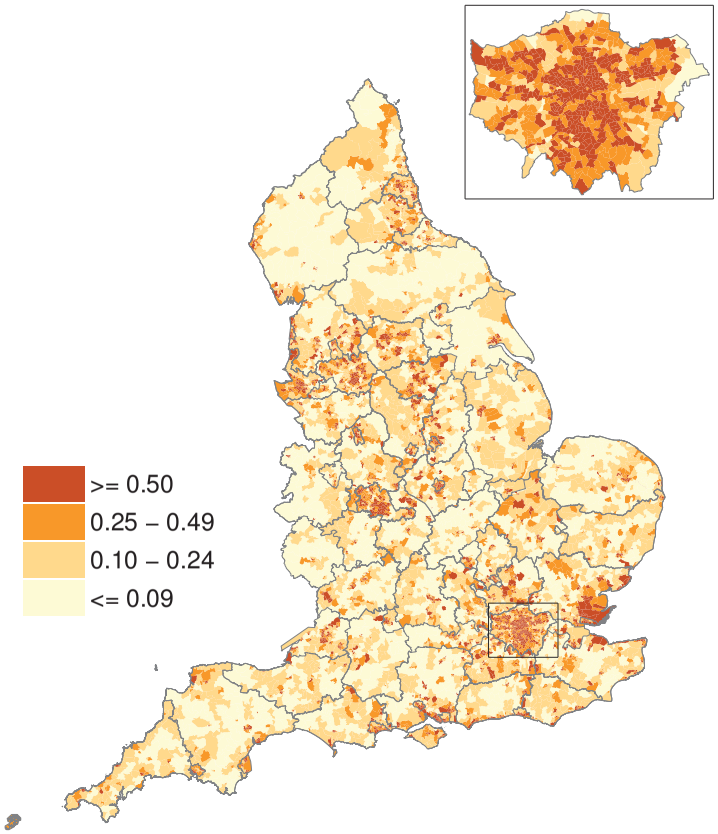
Year	% areas with extreme p_{Bayes} values		Mean p_{Bayes} values	
	Low severity	High severity	Low severity	High severity
2005	12.7%	15.5%	0.549 (0.285)	0.618 (0.254)
2006	12.3%	14.7%	0.551 (0.281)	0.618 (0.251)
2007	13.0%	14.7%	0.556 (0.284)	0.622 (0.249)
2008	11.0%	15.3%	0.552 (0.273)	0.622 (0.251)
2009	11.7%	15.6%	0.552 (0.275)	0.628 (0.249)
2010	11.3%	15.4%	0.545 (0.275)	0.629 (0.247)
2011	12.2%	15.6%	0.543 (0.282)	0.630 (0.246)
2012	12.4%	15.6%	0.547 (0.285)	0.628 (0.247)
2013	13.9%	17.3%	0.540 (0.2695)	0.634 (0.251)

Table 2.6: Proportion of areas with extreme p_{Bayes} values and the mean p_{Bayes} values for low and high severity under Model E; standard deviations are in parentheses.

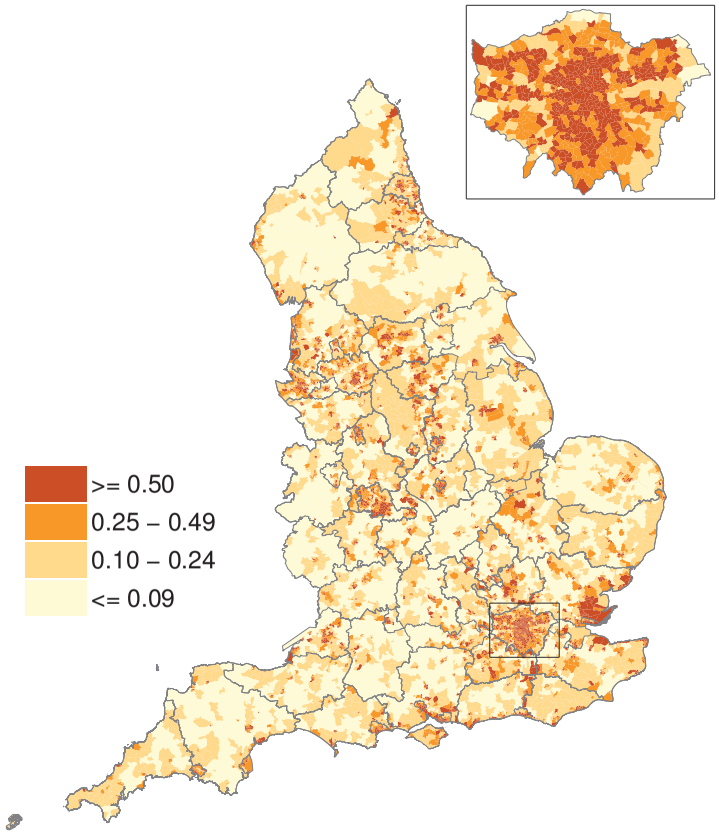
Sensitivity analyses did not highlight noticeable differences on the spatial component ($\theta_i + \phi_i$) as well as on the temporal component, as seen in figs. A.5 to A.8 in appendix A, suggesting the robustness of the results to the model specification.

Maps are produced to visualise how the risk of road traffic accidents is distributed across space. Figure 2.2 shows maps of crude accident rates and posterior accident rates respectively, calculated for a specific year (2005) for low severity. Figure 2.3 shows the respective maps for high severity level.

First, comparing crude and posterior mean rates for high severity, it can be seen

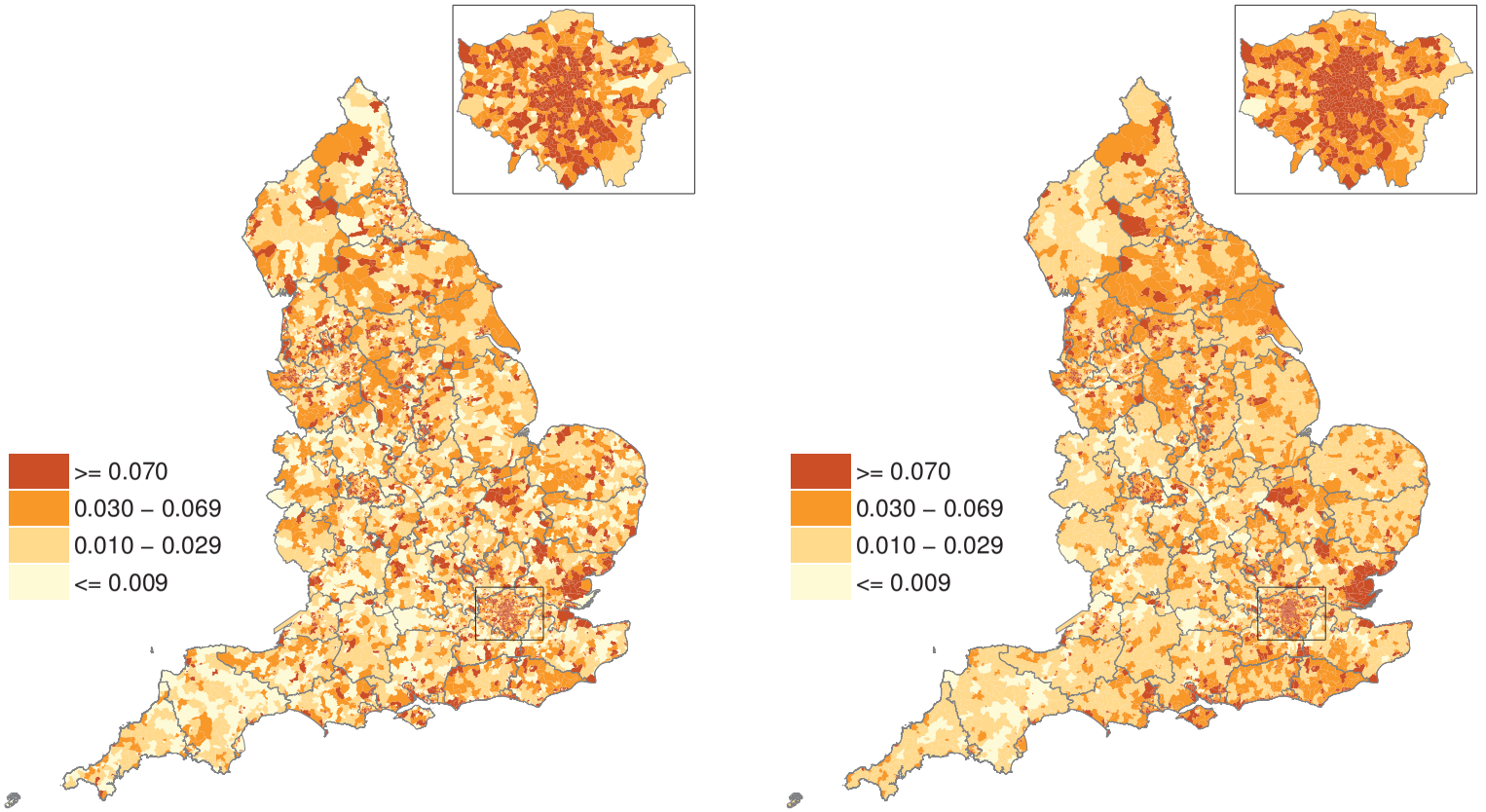


(a) Crude accident rates



(b) Posterior accident rates

Figure 2.2: Crude and posterior accident rates across wards in England for low severity in 2005; the insets show enlarged views of London.



(a) Crude accident rates

(b) Posterior accident rates

Figure 2.3: Crude and posterior accident rates across wards in England for high severity in 2005; the insets show enlarged views of London.

that these are greatly smoothed out under the model. The map of crude rates in fig. 2.3a does not indicate a particular pattern of risk, showing fairly similar effects across the southwest, central, and central east of England. The corresponding map of posterior rates in fig. 2.3b uncovers a clearer pattern; the effects in the southwest are eliminated, while these become stronger towards the eastern part of England and even stronger in the north, mainly in the area that includes Liverpool, Manchester, Leeds and Sheffield, as well as along the southern coast. Clusters of dangerous areas are now detected in the region of Southend-on-Sea, and around Hull and Bridlington. This is in line with common knowledge, as the road connecting those (A165) is believed to be the most dangerous in East Riding. Interestingly, motorways across England can also be clearly distinguished in fig. 2.3b. To illustrate this, I have linked the map with the motorway road network and the resulting map can be seen in fig. A.9 in appendix A. It appears that the wards including motorways, e.g M24, M1, M4, M11, show low posterior rates compared to the surrounding regions. This is obvious even in central London, where the inner part suffers most and while moving outside, the wards including motorways become less risky. The high accident rates within London and other big cities can be explained by the high numbers of pedestrians and cyclists involved in those, while the high rates in areas between big cities can be explained by the excess speed of vehicles and road conditions that lead to severe and fatal accidents. Other areas that are shown to be prone to accidents of high severity are countryside areas, such as Northumberland, North York, Peak District, and Yorkshire Dales National Parks, where road conditions and characteristics (narrow roads, bends, slopes, etc.) might be an explanation for this.

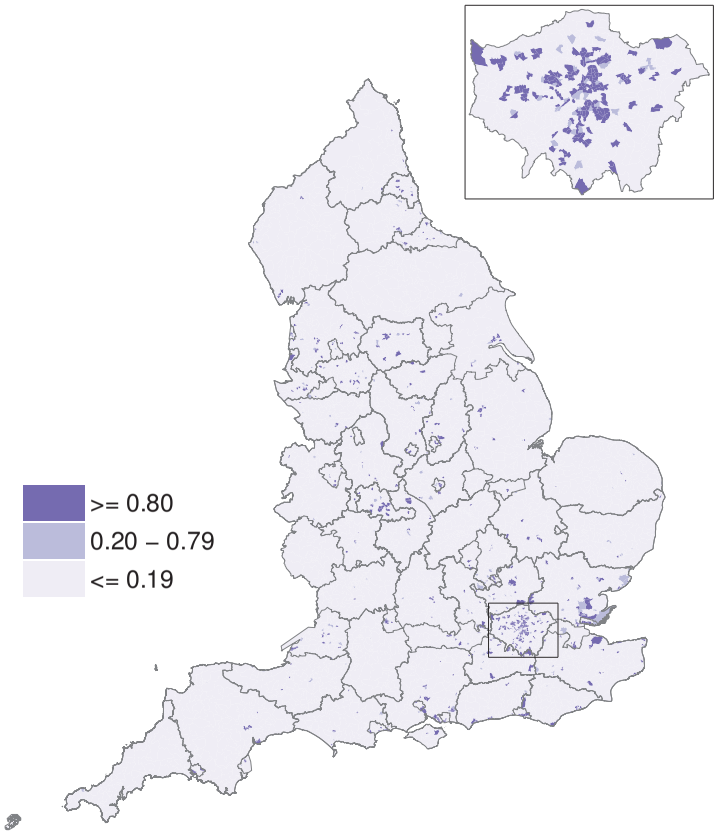
Second, looking at the posterior accident rates for low severity in fig. 2.2b, it can be seen that urban areas in England appear to be more dangerous. The risk for

accident incidence is focused mainly within big cities, with central London showing a significantly increased risk. Other big cities, such as Newcastle, Birmingham, Liverpool, Leeds, Sheffield and Manchester are also shown to be dangerous. This could be again due to the high numbers of pedestrians and cyclists, and presence of bus and cycle lanes.

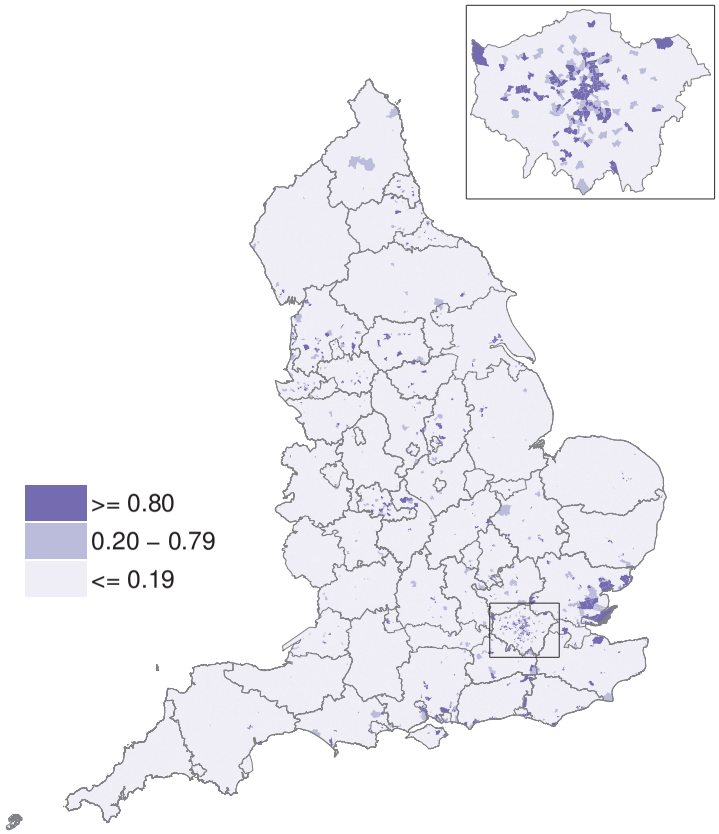
In contrast to high severity, there are no obvious differences between maps of crude and posterior rates when comparing figs. 2.2a and 2.2b. The fact that the model does not affect the rates for low severity importantly is reasonable considering that these consist of high counts in general, and the model is designed to provide smoothing to low counts that suffer from great variability. The information of low severity in the model yet contributes to stable estimates for high severity, and this is one of the main aspects of this work; strength is borrowed not only between areas, but also between severity levels.

Moreover, probability maps in fig. 2.4 help discriminate specific areas of excess risk in England, and focus on those which show strong evidence of increased risk, after accounting for the uncertainty in the posterior estimates. The west, east and northest parts of England are no longer shown as dangerous, and clusters of areas that consistently belong to the top 800 ranked areas are observed in central London, and in the area around Southend-on-Sea. This gives a clearer picture of the most dangerous areas at a national level, and therefore provides evidence for prioritising interventions to reduce rates.

Finally, by comparing ranks generated under the model against ranks under crude rates for the top 100 areas in fig. 2.5, it is indicated that for low severity these are somewhat consistent (fig. 2.5a), whereas for high severity (fig. 2.5b), these differ importantly. Among the top 100 ranked areas, the proportion of those that have a difference of more than 15 places between the two ranking criteria is 0.02% for low

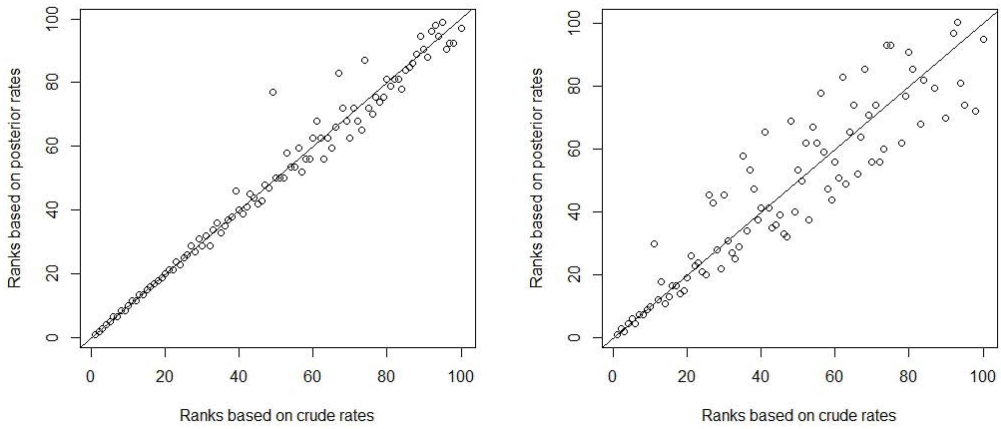


(a) low severity



(b) high severity

Figure 2.4: Maps of posterior probability that residual accident rates belong to the top 800 ranked areas for low and high severity in England; the insets show enlarged views of London.



(a) low severity (b) high severity

Figure 2.5: Area ranks based on crude rates versus area ranks based on residual accident risk for the top 100 areas.

severity and 0.26% for high severity; when the top 800 ranked areas are considered these numbers increase to 0.27% and 0.75% respectively, suggesting that the features of the model have an important impact on the results.

2.5 Discussion

In this chapter, I have developed models that combine ideas from a Bayesian hierarchical framework with space-time effects and multivariate analysis to produce a flexible framework for the analysis of road traffic accidents by severity level. I have compared models under various assumptions on the spatial and temporal dependences between levels of severity for road traffic accidents data.

The MBYM model has been identified as the most appropriate in this context. This model consists of both multivariate spatially structured and unstructured effects that assume a degree of correlation in the respective components between severities, allowing at the same time the quantification of this correlation. The

results suggest a high correlation between severities in both spatially structured and unstructured effects. The temporal effects are better modelled via completely separate components rather than a common one, indicating fairly different trends for the two severities. A multivariate specification on the temporal effects could be also considered in a future study, in order to explore potential correlations between severities in the temporal effects.

Probability maps and rankings are used to aid the detection of areas characterised by excess risk. One of the strengths of this model is that it does not only provide top ranked areas based on the mean rates of accidents, but in addition it takes into account the uncertainty associated. The results have shown that this could greatly influence high severity accidents, due to their small numbers.

From the point of view of policy making, the results of this study provide a high level of information to draw priority plans for actions. These could be potentially used centrally to decide where to allocate funding to decrease the number of injuries and fatalities due to road accidents. Among the policies that can be implemented to prevent road traffic accidents in the most high-risk areas, there are environmental changes (such as the introduction of speed cameras, or marked pathways for cyclists) and safety education and skills training (such as road safety media campaigns or providing free safety equipment). Different policies can be used for high-risk areas for high and low severity accidents.

Potential risk factors can also be formally assessed by the model. For instance socio-economic and demographic factors and adverse weather conditions could be good candidates to explain the pattern seen for slight and severe accidents. It would be of interest to also consider the type of road (i.e. major, minor), and other characteristics, and particularly whether a ward includes a motorway or not, in order to formally investigate the findings that were suggested by the current

analysis. In addition, information on the classification of an area as rural, suburban, or urban can be incorporated in the model.

It is important to note at this point that the MBYM specification has been shown to be the most suitable for the road traffic accidents data at hand and the results should not be generalised. Nevertheless the modelling approach that was followed in this chapter applies to any multivariate data with similar characteristics to road traffic accidents of low and high severity level, that would be reasonable to analyse jointly. Such data are spatio-temporal non-communicable diseases that are known to share characteristics; for instance lung and oral cancer, or asthma and chronic obstructive pulmonary disease (COPD). A similar framework could be therefore followed, where alternative models under different spatial and temporal dependency assumptions are compared, and the best is selected for inference, while probability maps can also be used to highlight excess risk.

Finally, as it is usually the case with disease mapping models, this study should be treated as an explorative analysis, aiming at investigating the spatial and temporal patterns of accidents and at identifying areas characterised by particularly high risks, and not be placed in the context of hot spot analysis (HSA). HSA involves a deep investigation from identifying the dangerous locations, ranking these locations according to various criteria, and providing explanations of why certain locations are hot spots, and which usually consider information on the cost of the accidents (Miaou and Song, 2005; Brijs et al., 2007). Instead, the modelling approach proposed here can serve as a first step for policy making, which should be followed by further investigation.

Aiming at detecting clusters of areas with an unusual road accidents pattern (e.g. increasing in time while the general temporal trend shows a decrease) more complex models including an interaction term could be considered, which would

provide information on changes across space and time jointly. The rest of this thesis focuses on such framework, and explores its potential in public health analyses.

3 Investigating spatio-temporal trends in asthma and COPD using a Bayesian detection model

3.1 Introduction

Asthma and chronic obstructive pulmonary disease (COPD) are the most common chronic respiratory conditions worldwide, contributing to heavy social and economic burden (World Health Organization, 2012).

The number of people suffering from asthma in 2014 was estimated to be 334 million around the world (World Health Organization, 2014) and this number is projected to rise to 400 million by 2025. Around 250,000 deaths per year are caused by the disease, with the majority of them considered to be preventable (Masoli et al., 2004). COPD has a lower prevalence of 64 million people but much higher mortality, with 3 million deaths annually, an estimated 6% of all deaths worldwide. COPD is predicted to become the third leading cause of death by 2030 (World Health Organization, 2014). In the UK, asthma affects 1 in 5 households, and COPD is the fifth leading cause of death after cancer and cardiovascular disease (Masoli et al., 2004).

Asthma and COPD have similarities in symptoms and treatment and there may be considerable overlap between these conditions making them difficult to distinguish clinically (Drazen et al., 2015). Asthma commonly starts in childhood

and is often allergic in origin, while a large proportion of COPD is caused by smoking and the condition starts in mid to later life. Common symptoms in both conditions are shortness of breath and wheeze, with worsening of symptoms with respiratory infections, with similarities in treatments of bronchodilators, steroids and antibiotics for infections. Triggers for exacerbations, the major determinants of admissions and possibly deaths are likely to be influenced by infectious disease trends and by common environmental factors with a spatio-temporal structure, such as air pollution (Eeftens et al., 2012).

A study of trends of chronic respiratory diseases (CRDs) is important, as it allows a better understanding of the characteristics of the disease, to determine whether health policies or preventive measures are effective, and to identify high-risk populations that might require additional care and treatments. A challenge, however, is the choice of data to use for asthma and COPD analyses. Studies on mortality (Sin et al., 2006) investigate only the highest degree of severity and findings may differ from those for hospital admissions particularly for asthma (Hansell et al., 2003), which is a more heterogeneous condition. Most studies have used data from secondary and tertiary care, such as hospital admissions and emergency care data, but these will not capture milder cases seen in primary care. It is estimated that only 20% of asthmatic patients and less than half of COPD patients suffer from severe symptoms (Lindebeg et al., 2006).

A useful addition to asthma and COPD research is the use of General Practice (GP) drug prescription data which consist in the numbers of items that are prescribed in England by GPs and are dispensed anywhere in UK or Europe. GP drug prescriptions can be very relevant for asthma and COPD as these are long-term conditions that are controlled by regular medication. These capture patients of any severity of the disease, from mild to severe, and hence they can provide a

general picture of the respiratory health of the population at small area level. Only a few authors have used GP drug prescriptions to investigate asthma and COPD trends in the literature (Hansell et al., 2003; Vegni et al., 2005; Laurent et al., 2009; Sofianopoulou et al., 2013).

In addition, the geographical trends of asthma and COPD have been barely studied. Hansell et al. (2003) found that COPD mortality, hospital admissions and GP prescriptions for COPD were higher in urban areas and northern regions of England, but less clear patterns were seen for asthma in comparisons, using age-sex standardised event ratios. Other examples include Joo et al. (2007) who used a negative Binomial model to explore variations in COPD patients across a health system in various geographic regions in the USA, and Sofianopoulou et al. (2013) who analysed GP drug prescriptions on asthma and COPD in the Newcastle and North Tyneside area in the UK through a mixed effects regression model. Bayesian hierarchical approaches were employed by Holt et al. (2011) and Nandram et al. (2000) who analysed COPD hospitalisations and mortality rates respectively across the USA.

Studies (Hacking et al., 2011; Wells and Gordon, 2008; Hansell et al., 2003) suggest that there is a significant difference of morbidity and mortality within and between regions of the UK over the last 40 years. This needs to be taken into account in order to both provide reliable statistical estimates, as well as to help public health policy makers more clearly identify target areas with great needs and improve disease prevention and treatment.

The objective of this study is to investigate trends in CRDs at the population level across England by using the BaySTDetect method, a mixture hierarchical model which is appropriate for disease mapping and detection. Multiple data sources are analysed to help understand the relationships between disease prevalence, morbidity

and mortality. Spatial and temporal patterns of hospital admissions, deaths, and GP drug prescriptions will be explored, in order to evaluate if different behaviours can be seen for different data sources underlining different condition severity. A focus of this study is the detection of unusual areas, i.e characterised by a temporal trend which deviates from the general one, suggesting the presence of a policy or an emerged localised factor. In this analysis information on asthma and COPD is combined, given the similarities in these conditions, issues distinguishing between them (Drazen et al., 2015) and the fact that the available GP prescription dataset does not provide information on diagnosis.

The remainder of the chapter is structured as follows. Section 3.2 describes the data sources used for the analysis, and section 3.3 describes the statistical modelling framework. In section 3.4 the results of the study are presented and discussed, and finally section 3.5 summarises the main findings of the study.

3.2 Data sources

To gain a better understanding of asthma and COPD, three different data sources are used: *i*) GP drug prescription data of treatments used for these conditions, which capture patients with mild to severe symptoms and will give a general picture of the disease prevalence across the study region; *ii*) Hospital Episode Statistics (HES) admissions with primary diagnosis of asthma or COPD; *iii*) mortality data with asthma and COPD disease as cause of death. The latter two data sources will inform on cases characterised by higher severity. Each data source is described in the rest of this section.

3.2.1 GP drug prescription data

The Prescription Cost Analysis (PACT) data include the prescriptions of all drugs from 8003 GPs across England from August 2010 onwards at a monthly temporal resolution. In this study the prescriptions on Salbutamol, Ventolin and Clenil Modulite are used, with corresponding British National Formulary (BNF) codes 1011R0AAAPAP, 0301011R0BEAIAP and 0302000C0BPABBF respectively. These account for more than 90% of the total prescription of short acting beta2-agonist (SABA), a class of drugs that relieves patients from bronchospasm which characteristically occurs in acute symptoms (Drazen et al., 2015). Every GP is part of a local clinical commissioning group (CCG), which is the authority responsible for local healthcare services including local hospitals and NHS services, according to the 2012 Health and Social Care Act. GPs of the same CCG collaborate to evaluate local needs, monitor services, set priorities and make area-specific decisions to promote healthcare services for local residents. This suggests that GPs within the same CCG should share similarities. Therefore, the available GP data are aggregated at CCG level (211 in England) to be used for the analysis. The time period considered for the study covers months from August 2010 to March 2011. The PACT data also contain the number of patients registered within each GP, with information on age group and sex. Appropriate covariates are included in the modelling framework in order to adjust for age and sex. As an indicator of age, the percentage of active population (age group 15 to 64) is calculated, while as an indicator of sex, the male to female ratio. Summary statistics for the GP drug prescription data can be seen in table 3.1, and maps of crude rates across the 8-month period can be seen in fig. B.2 in appendix B.

Time	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Prescriptions (in '000s)						
Aug 10	1	4.00	5	6.09	7.00	21
Sep10	2	4.00	6	7.06	8.00	25
Oct 10	2	4.00	6	7.04	8.00	25
Nov 10	2	5.00	6	7.26	9.00	26
Dec 10	2	5.00	7	8.36	10.00	30
Jan 11	2	4.00	6	6.94	8.00	25
Feb 11	2	4.00	5	6.51	8.00	23
Mar 11	2	5.00	6	7.44	9.00	26
Covariates						
Male to female ratio	0.87	0.97	0.99	0.99	1.00	1.16
% Active population	0.59	0.65	0.66	0.67	0.68	0.78

Table 3.1: Summary statistics for counts of GP drug prescriptions

3.2.2 HES and mortality data

Health data for England from August 2010 to March 2011 were obtained from the Small Area Health Statistics Unit (SAHSU) at Imperial College London. HES admission data contain the number of admissions with a primary diagnosis of asthma or COPD, and may also include readmissions of the same patient. Deaths with underlying cause of death (UCD) of asthma or COPD was collected. International Classification of Disease coding version 10 (ICD-10) was used for admission and mortality coding throughout this time period and asthma and COPD were defined as ICD-10 codes 490–496. Linkage between HES and mortality data to identify individual patients was not possible. Population data were obtained from the Office for National Statistics (ONS), with individual-level information on age and sex and

these were used for the calculation of the expected number of cases. Similarly to the GP drug prescription data, HES, mortality, and population data were all aggregated to CCG level. Summary statistics for HES and mortality data can be seen in table 3.2, and the corresponding maps of crude rates across the 8-month period in figs. B.3 and B.4 in appendix B.

Time	Min	1st Qu.	Median	Mean	3rd Qu.	Max
HES admission data						
Aug 10	9	46.50	66	78.27	99.50	310
Sep10	18	70.00	106	116.10	145.00	535
Oct 10	21	73.00	110	124.50	155.50	536
Nov 10	21	66.00	95	109.30	134.00	485
Dec 10	22	86.50	122	139.30	172.00	492
Jan 11	21	75.00	103	122.40	152.00	460
Feb 11	13	61.00	88	100.50	125.00	361
Mar 11	19	60.50	89	103.10	131.00	394
Mortality data						
Aug 10	0	4.00	7	7.81	10.00	27
Sep10	0	4.00	7	8.17	11.00	33
Oct 10	1	6.00	8	9.67	13.00	32
Nov 10	0	5.00	8	9.80	12.00	31
Dec 10	0	9.00	13	15.06	19.00	50
Jan 11	0	9.00	14	16.10	20.50	61
Feb 11	1	8.00	8	10.18	13.00	29
Mar 11	1	6.00	9	10.84	14.00	37

Table 3.2: Summary statistics for counts of HES admissions and deaths

3.3 Statistical Analysis

The analysis is conducted within a Bayesian hierarchical framework that takes into account the complex dependence patterns of asthma and COPD over space and time. I employ the BaySTDetect method and explore how it can be used to help understand patterns of CRDs in space and time, along with detecting unusual behaviour of the disease over time.

The original BaySTDetect model is described in detail in section 1.3.2 on page 23, and is reproduced here adapted to the data at hand. Following the epidemiological structure on disease mapping, let Y_{it} denote the observed counts in CCG $i = 1, \ldots, N = 211$ at time points $t = 1, \ldots, T = 8$, corresponding to months August 2010, September 2010,..., March 2011. Similarly E_i represents the expected cases in CCG $i = 1, \ldots, N = 211$. A Poisson log-normal model is assumed for the disease counts

$$Y_{it} \sim \text{Poisson}(\lambda_{it} E_i), \tag{3.1}$$

where the rate λ_{it} follows a mixture of two components

$$\begin{aligned} \log(\lambda_{it}) &= z_i \log(\lambda_{it}^{(C)}) + (1 - z_i) \log(\lambda_{it}^{(AS)}) && (3.2) \\ \log(\lambda_{it}^{(C)}) &= \alpha_0 + \eta_i + \delta_t \quad \textit{(Common model)} && (3.3) \\ \log(\lambda_{it}^{(AS)}) &= \nu_i + \kappa_{it} \quad \textit{(Area-Specific model)} && (3.4) \end{aligned}$$

The mixing parameter z_i selects between estimates $\lambda_{it}^{(C)}$ from the *Common model* and estimates $\lambda_{it}^{(AS)}$ from the *Area-Specific model*.

Priors are specified for all model parameters following Li et al. (2012). For the *Common model*, the spatial component η_i is assigned a convolution prior, $\eta_i \sim N(v_i, \sigma_\eta^2)$ where v_i follows a conditional autoregressive prior, $v_{1:N} \sim \text{CAR}(\mathbf{W}, \sigma_v^2)$,

with an adjacency matrix $\mathbf{W}$ of size $N \times N$, specifying the first-order spatial neighbourhood structure. Likewise, the temporal component δ_t follows a random walk prior of order 1, implemented as $\delta_{1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_\gamma^2)$. Here, $\mathbf{Q}$ is of size $T \times T$ specifying the first-order temporal neighbourhood structure. The overall intercept follows a flat prior $\alpha_0 \sim U(-\infty, +\infty)$. For the *Area-Specific model*, the area-specific intercept follows a weakly informative prior $\nu_i \sim N(0, 1000)$, and the area-specific temporal component follows a conditional autoregressive prior $\kappa_{i1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_{i\kappa}^2)$, where $\mathbf{Q}$ is defined as previously.

A weakly informative half-normal prior $N(0, 1)$ is assigned to each of the parameters σ_h, σ_v and σ_γ while a $N(\alpha, \beta^2)$ prior is assigned to the $\log(\sigma_{ik}^2)$ with parameters α and β following priors $N(0, 1000)$ and $N(0, 2.5^2)$, based on the specification in Li et al. (2012).

The model indicator z_i follows a Bernoulli prior with parameter 0.95

$$z_i \sim \text{Bernoulli}(0.95), \tag{3.5}$$

and the posterior estimates of the indicator z_i are used for the classification of areas as common or unusual. In this study, the areas that satisfy the following standard rule are classified as unusual

$$r_j \leq 0.05, \tag{3.6}$$

where r_j is the jth ordered posterior z_i.

The same BaySTDetect model is fitted to all three datasets described in section 3.2 for the sake of comparison. In order to justify the choice of a Poisson distribution (eq. (3.1)) for GP drug prescriptions, the original large counts are divided by 1,000 as can be seen in table 3.1.

In the prescription model, Y_{it} is the number of GP drug prescriptions in CCG i in

month t, and E_i are the corresponding expected counts averaged across months.In the admission model, Y_{it} are the observed cases of asthma and COPD admissions in CCG i in month t and E_i are the expected ones based on age and sex indirect standardisation using the whole of England as standard population, averaged across months. Similarly, in the mortality model, Y_{it} is the observed number of deaths due to asthma or COPD in CCG i at month t and E_i is the corresponding expected number based on age and sex indirect standardisation using the whole of England as standard population, averaged across months.

3.4 Results

The analysis was carried out in R and the models are implemented in OpenBUGS. Three chains are run for each parameter per model with different initial values for 80,000 iterations, from which 20,000 are discarded as burn-in, and estimates are based on the remaining samples using only every 5th iteration to limit auto-correlation. The simulations took around 17 hours per model on an Intel Xeon processor 2.50GHz with 23.4 RAM. Convergence was assessed through trace plots, Brooks-Gelman-Rubin (BGR) statistic, autocorrelation plots, and Monte Carlo error.

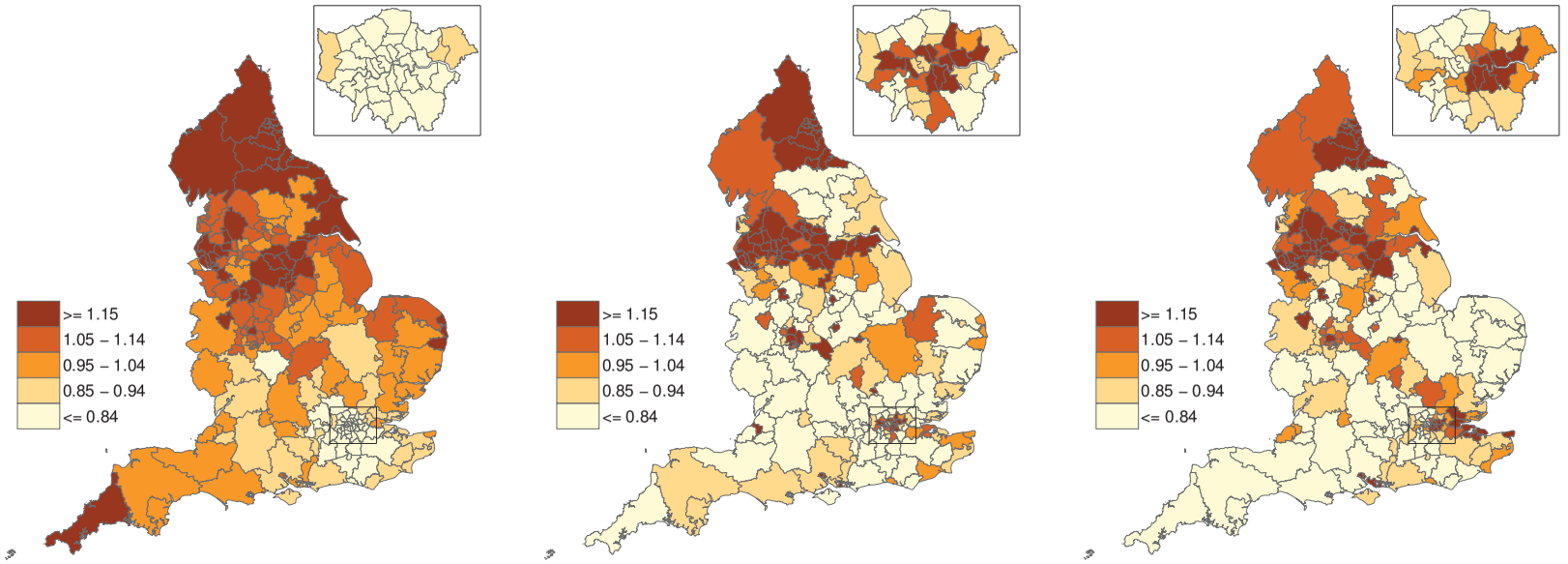
3.4.1 Spatial Patterns

Maps of posterior rates are produced to investigate the geographical patterns of asthma and COPD across the three different data sources and evaluate whether these share any similarities or differences. Figure 3.1a shows the residual relative risk of the spatial component $\exp(h_i)$ of the *Common model* for the GP drug prescriptions across England during the period August 2010 to March 2011, while figs. 3.1b and 3.1c show the corresponding risk for HES admissions and deaths respectively.

A clear pattern is observed across all datasets with a strong increasing effect from south to north. High risk is focused mainly on the region that includes Liverpool, Manchester, Leeds, and Sheffield, as well as on the region Newcastle and Durham, with the highest risks being observed in the north east of England. Potential shared risk factors may include deprivation, which is highest in urban areas of the north-east and north-west, environmental factors such as air pollution and smoking prevalence (Prescott and Vestbo, 1999; Lopez et al., 2006; Andersen et al., 2011). The north west is the areas with the highest prevalence of smokers (HSCIC, 2015).

By comparing the risk distribution across different data sources, it can be observed that for admissions and deaths, these are somewhat consistent (figs. 3.1b and 3.1c) while for prescriptions important differences are highlighted (fig. 3.1a). This observation is confirmed by the heatmap in fig. 3.2 which shows the correspondence across the three data sources using the spatial relative risk; the area in light yellow represents the CCGs in England that share a low relative risk while the area in brown represents the ones with a high relative risk. It is clear that the proportion of areas with a low/high relative risk across all three data sources is relatively small compared to the proportion of areas with a low/high relative risk among HES admissions and mortality data. The correlation of the latter two was estimated to be 0.77, while the corresponding correlations of GP drugs and mortality, and GP drugs and HES admissions were close to 0.5 (see additional plot in fig. B.1 in appendix B).

Low risk is observed for admissions and deaths (figs. 3.1b and 3.1c) in the south of England, and in the coastal areas. At the same time these areas exhibit high prescription rates (fig. 3.1a). This suggests a number of hypotheses for further investigation e.g. that higher prescription rates are a causal factor in lower morbidity and mortality, that there are more patients with milder disease in the south-east



(a) GP drug prescription data (b) HES admission data (c) Mortality data

Figure 3.1: Posterior means of the spatial relative risk $\exp(\eta_i)$ of CRD across CCGs in England for prescriptions (a), admissions (b), and deaths (c); the insets show enlarged views of London.

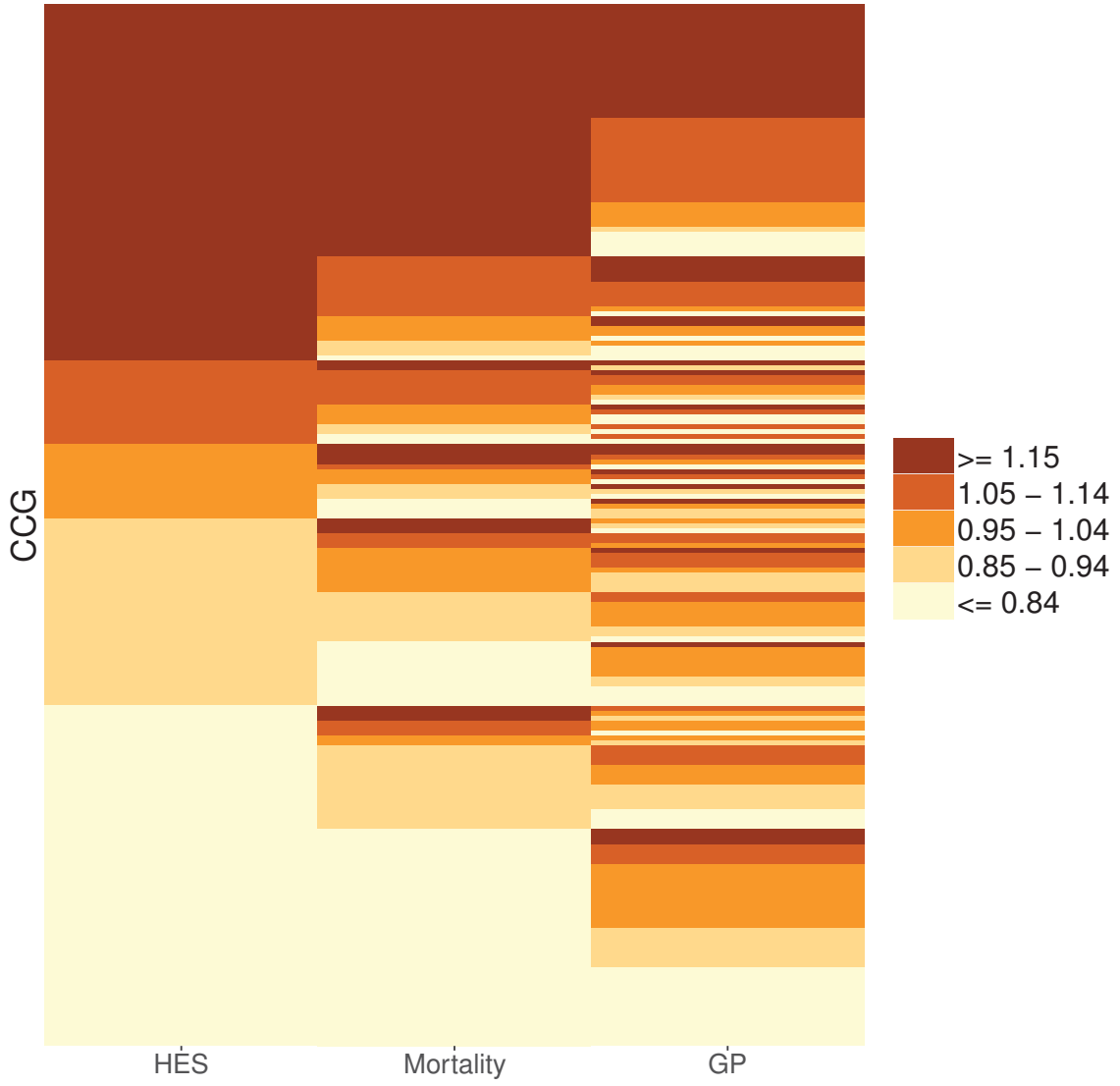


Figure 3.2: Heatmap showing correspondence across three data sources on CRD: admissions, deaths and prescriptions.

potentially related to lifestyle factors (lower smoking) and environmental factors (lower air pollution), that deprivation results in higher morbidity and mortality but lower use of primary care. Age-sex effects could also explain these differences; these are accounted for at an aggregate level for drugs, through the inclusion

of appropriate covariates at the GP level in the model, since information was not available at an individual level, while indirect standardisation was used for admissions and mortality.

Only a few CCGs in the south appear to have high admission and mortality rates (figs. 3.1b and 3.1c), including Portsmouth and Southampton CCGs. Increased smoking prevalence and high deprivation can be related to this. Unlike the rest of big cities in England, London appears to have low prescription rates (fig. 3.1a) potentially related to the low number of smokers in the area (HSCIC, 2015). Comparing between mortality and admission spatial patterns, a stronger south to north effect can be seen for mortality (fig. 3.1c), showing the south eastern and central west parts of England as the least risky for disease death. Another region which potentially represents a group of patients with mild symptoms is around the North Yorkshire area, where national parks are, and this area clearly stands out in the admissions map (fig. 3.1b).

3.4.2 Temporal patterns

The general temporal patterns in England from August 2010 to March 2011 under the different datasets considered in the study can be seen in fig. 3.3, which plots the component $\exp(\delta_t)$ of the *Common model*.

Generally, there was a seasonal pattern in the monthly risk for asthma and COPD across all data sources over the period August 2010 to March 2011 however the trends show important discrepancies. The highest risk was recorded around Christmas across all data sources. This is likely to represent peaks in respiratory infections, especially influenza and respiratory syncytial virus (RSV) that peak around this year. These result in increased respiratory symptoms from mild to severe and cause exacerbations of disease that are occasionally fatal (Fleming et al., 2015). The lowest peak was recorded in August for all three data sources, a period

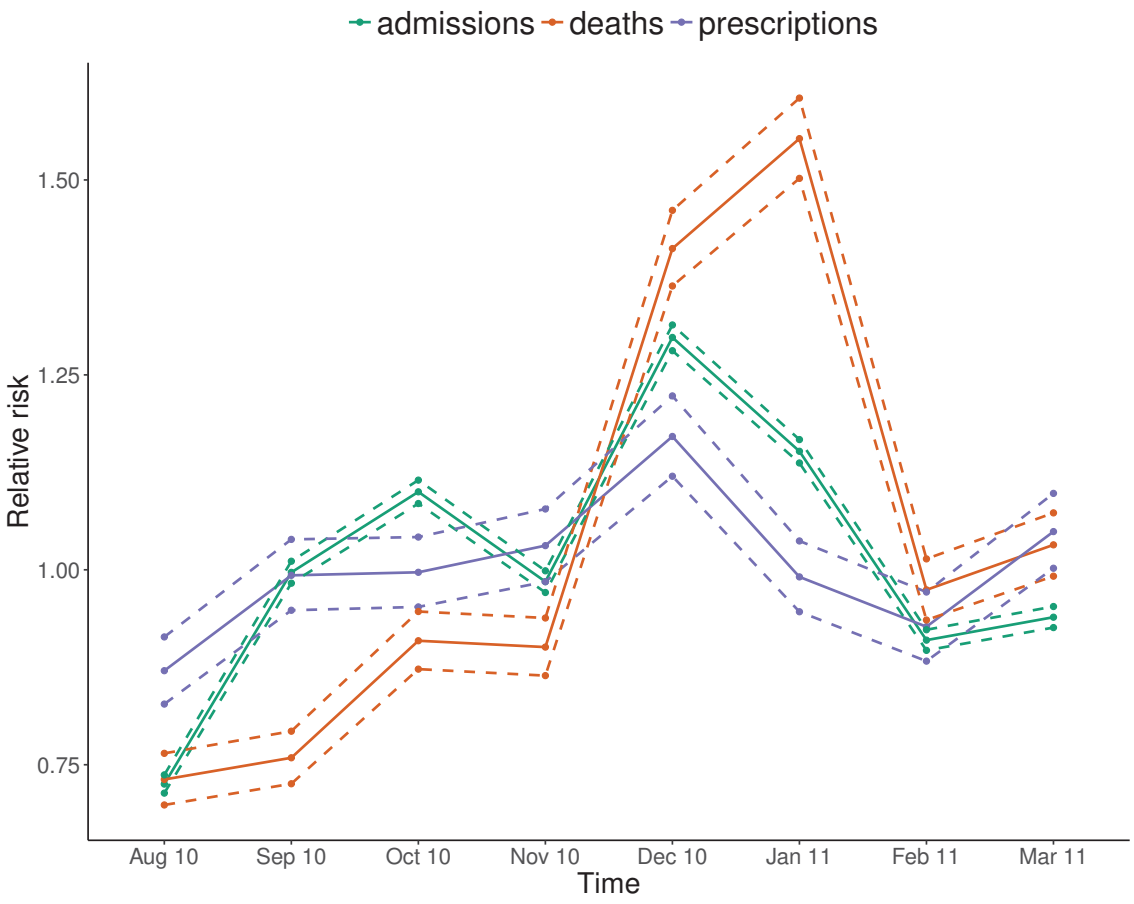


Figure 3.3: Temporal trends of CRD across admission, mortality, and prescription data; The solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values.

when people are on holidays, away from their homes. The raise in the rates across all datasets around September coincides with the start of term at schools.

By comparing temporal trends across different data sources, it can be seen that there is a high temporal variation for mortality, whereas this is lower for admissions and prescriptions. This might be due to the fact that hospitals are either closed or understaffed during holidays, and thus not well functioning. Interestingly, a time lag is apparent between admissions and mortality. The highest peak for admissions is December, whereas for mortality is January and this might reflect the group of

admissions which were then followed by death.

3.4.3 Detection of unusual areas

Finally, I obtain the posterior estimates of the indicator z_i for each model in order to classify areas as unusual based on the rule described in eq. (3.6).

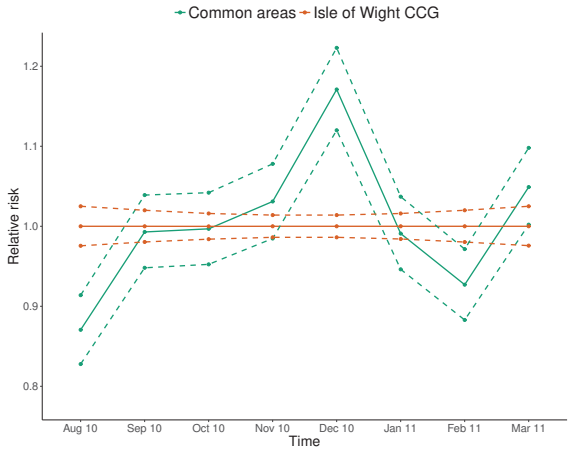
For GP drug prescription and mortality data, no areas were detected as unusual, meaning that all posterior estimates of the parameter z_i were above the *a priori* threshold value of 0.05. When ranking areas by probability, the Isle of Wight was the area with the smallest probability of following either the common prescriptions temporal pattern (with a probability of 0.13), or the common mortality temporal pattern (with a probability of 0.11).

Looking at the temporal patterns for this area in fig. 3.4, a flat pattern is apparent for prescriptions, indicating low rates with no peaks throughout the whole time period (fig. 3.4a), while for mortality, a near exponential increase is observed, which after December 2010 exceeds the national death rates importantly (fig. 3.4b). The corresponding plot for admissions shows a relatively stable pattern (fig. 3.4c).

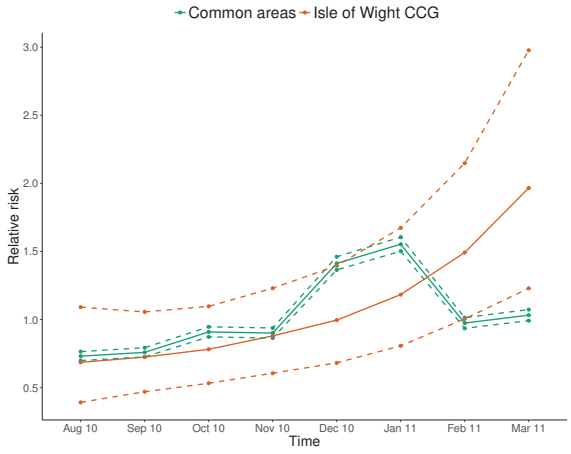
Despite the differences observed between the common and area specific-trend for the Isle of Wight under all three data sources, the model was not able to pick up the signal, potentially because of the probability specification in eq. (3.5) which spans over the entire time period and might not be very sensitive if only a few time points are deviating.

An additional observation that can be made based on these plots is the large uncertainty that characterises the area-specific time trends compared to the one for the common time trends, which is reasonable considering the small numbers of counts that are modelled through the spatially independent *Area-Specific model*.

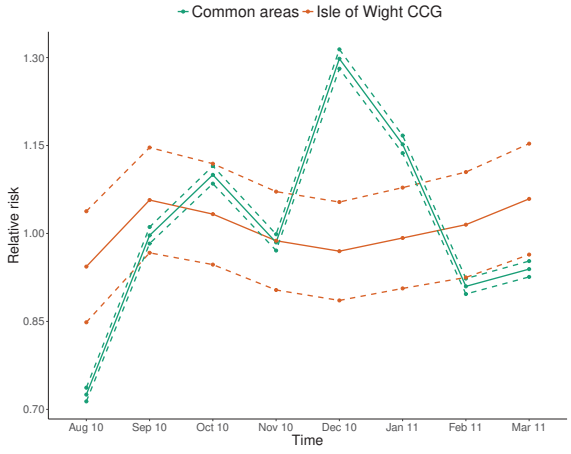
After investigating the CCG of Isle of Wight, I found that an intervention was



(a) GP drug prescription data



(b) Mortality data



(c) HES admission data

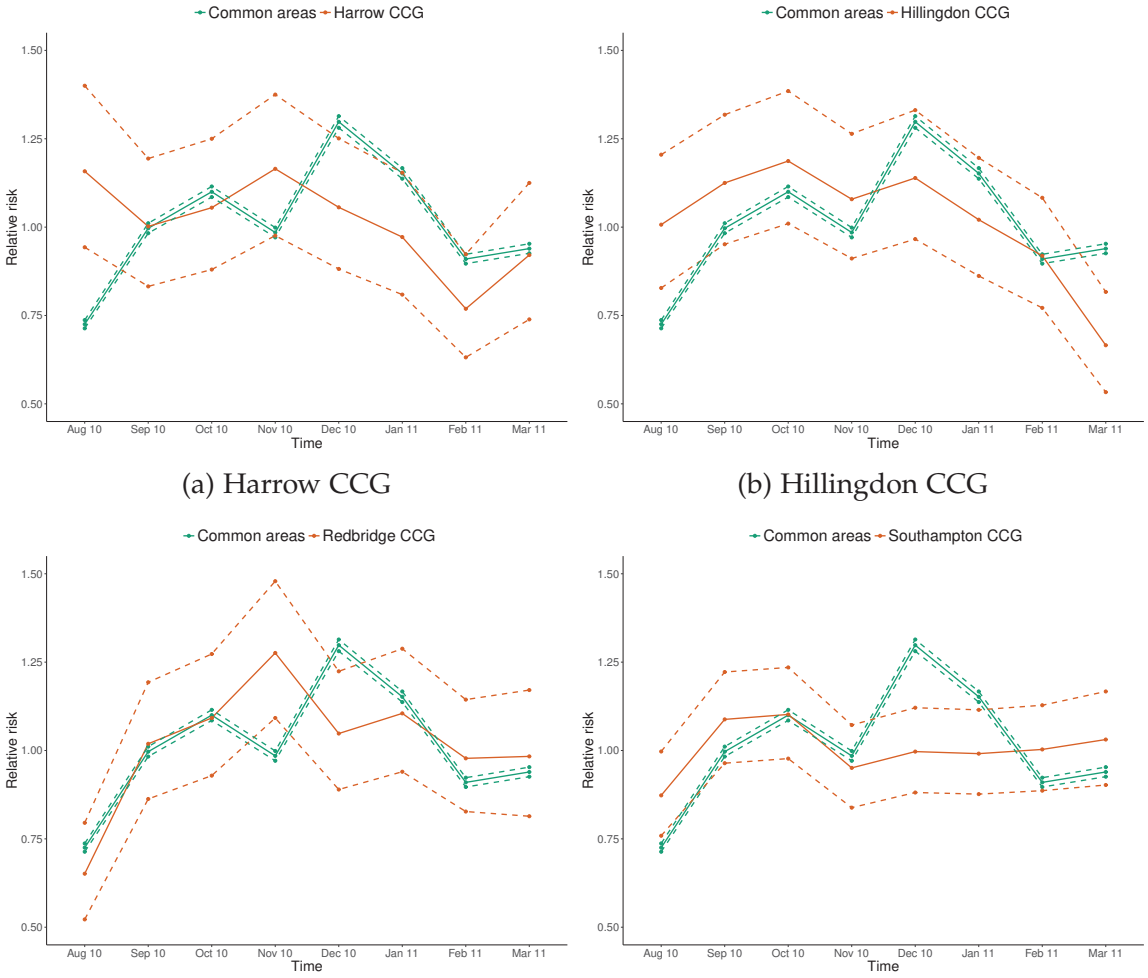
Figure 3.4: Temporal trends of CRD across across prescription, mortality, and admission data for the Isle of Wight CCG; the solid lines show the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values.

implemented prior to 2008/2009 in order to reduce the high prevalence of respiratory diseases in the area, which was combined with an excess expenditure on respiratory medication. The project entitled 'Isle of Wight Respiratory Inhaler Project', led by the National Institute for health and Clinical Excellence (NICE), involved training of healthcare professionals in the use of the inhaler, patient training, and assessment of the inhaler technique. According to the report, which became available by NICE in February 2011 and was accessed online in October 2015 (NICE, 2015), at the end of the project in 2009, the costs on selective beta-agonists fell by 22.7%, the number of prescriptions fell by 25.2%, and within 12 months, the emergency admissions due to asthma were reduced by 50%, and associated deaths by 75%.

Given the above, the flat pattern in prescriptions seen in fig. 3.4a is well supported, as well as the one for admissions (fig. 3.4c). However, the increasing temporal trend in deaths in fig. 3.4b which peaks in the last month of the study period generates questions. Since the findings reported by NICE consider only asthmatic patients, meaning that COPD patients that use the same medication are not represented, further analysis is required for the CCG of Isle of Wight separately on asthma and COPD in order to understand the underlying causes of the mortality trend. Additional data need to be also analysed in order to see the progression of the rates long after the intervention was implemented. Moreover, COPD is often misdiagnosed as asthma due to the similar symptoms that these two diseases share, and also depending on the availability of spirometry (Walker et al., 2006). This calls for further investigation into the misdiagnosed cases of COPD as asthma in the CCG of Isle of Wight which might be an adverse effect of the project.

On the other hand, four areas were detected as unusual using a 0.05 threshold level for hospital admissions model: Harrow, Hillingdon, Redbridge and Southampton. The time plots in fig. 3.5 show the temporal trends for each unusual area,

compared to the general temporal trend.



(a) Harrow CCG

(b) Hillingdon CCG

(c) Redbridge CCG

(d) Southampton CCG

Figure 3.5: Temporal trends of CRD for the unusual CCGs in England plotted against the national trend under admission data; the solid lines show the posterior mean of the temporal effects $(\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines represent the corresponding 95% CI values.

As can be seen, the corresponding areas do not follow particularly extreme patterns, but there are discrepancies from the common one that the model identifies as important. For instance, although Harrow in general exhibits lower than the average national admission rates due to asthma or COPD, an unusual high peak is observed in August, as well as one in November, which months are considered safe

for the rest of England. Hillingdon, on the other hand appears to have an extreme decrease in March 2011. Redbridge shows a different pattern in admissions from the other areas in that admissions increase in early November. London is a city known for its diversity, and for the population movements within it, hence it is not easy to make conclusions as to what has happened in those CCGs. Factors such as differences in local population characteristics, or differences in the quality of hospital care and the support people receive to manage their condition across time could be responsible.

The fourth area that was detected as unusual in terms of admission trends is Southampton. This is one of the few areas on the South Coast that show an increased risk for severe symptoms and deaths due to asthma and COPD (fig. 3.1). Interestingly, it seems that the low risk observed in fig. 3.5d from December onwards is the outcome of an intervention that was implemented in order to address the excess number of admissions in the CCG of Southampton. Wilkinson et al. (2014) found that 34 patients were responsible for 22% of the total COPD admissions over a 3-year period. The authors present an admission avoidance strategy that was constructed for this group of patients, and as a result the readmission rate fell from 13.4 to 1.9%.

3.5 Discussion

In this chapter I have investigated the spatial and temporal patterns of asthma and COPD, with a special focus on the detection of areas that follow an unusual temporal pattern compared to the general one, by employing the promising BaySTDetect method. Effective high quality datasets were used, aggregated by CCG which is meaningful for healthcare practices. There are two main conclusions that can be

drawn from this study.

First, the detection mechanism that is provided by the Bayesian mixture model, together with inference on the spatial, and temporal variation, can aid health care professionals and public health practitioners identify target areas, assess a policy impact or the quality provided by hospitals, and hence develop effective prevention programs to improve population health.

A strong aspect of the modelling approach is that it identifies unusual behaviour not only in terms of increased risk, but also in terms of any risk pattern that deviates from the expected one. In the original paper the model was applied on annual data by Li et al. (2012), while in this study monthly data were used as the seasonal pattern of asthma and COPD was an important aspect of the analysis. The expected cases were therefore averaged across the study time period for this model (E_i), contrary to the original one (E_{it}).

An additional advantage of BaySTDetect is that it only detects areas when a certain criterion is met, and the specification of this depends on how conservative we want to be. Li et al. (2012) chose a Bayesian False Discovery Rate (FDR) to adjust for multiple testing. Using this rule in the present study, the number of positives increased considerably compared to using a standard classification rule, and therefore it was not employed in order to avoid possible false positives.

The second main conclusion of this study is that the analysis of multiple data sources representing different degrees of disease severity gives a more comprehensive picture of the respiratory health problem with potential implications for healthcare. Some similarities across the different datasets both in spatial and temporal variation are indicated, however also important discrepancies are apparent reflecting the different groups of patients that are represented.

Asthma and COPD are jointly considered in the analysis, as the indication of the

disease was not possible for the prescriptions, and one of the objectives of the study was to compare between the different data sets. It would worth exploring separately asthma and COPD trends for all ages and for children, adults and elderly to see if the findings of this study hold for these conditions individually, and to better understand the temporal patterns of the areas found to be unusual.

Public health researchers may also be interested in formally investigating possible confounders, effect modifiers and causal factors such as deprivation, smoking, seasonal respiratory virus activity and service access, and to also adjust for socio-economic status, as this is likely to affect the spatial patterns. It is relatively easy to introduce covariates into the BaySTDetect model to start to explore some of the hypotheses as to why some CCGs differ from national trends. As an example I have introduced a variable to represent deprivation - the Index of Multiple Deprivation (IMD) score, which was initially available at Lower Layer Super Output Area (LSOA) level for England. This was aggregated at CCG level, and then converted into a categorical variable of 5 levels (quintiles) which was introduced into the model through dummy variables. The results showed no important effect; the exponential of the posterior estimates of the coefficients of the dummy variables varied from 0.99 to 1.02 with all credible intervals including 1. The inclusion of deprivation in the model did not affect the posterior estimates of the other parameters, nor the areas that were detected as unusual.

Despite the great advantages of BaySTDetect over standard disease mapping models, its applicability remains fairly limited in the field of public health surveillance. For instance, the model does not inform on the exact temporal point at which an unusual observation occurs. In addition, it has only been tested on a temporal window of 8 time points and it is questionable whether it can handle longer time series data given the fixed probability model over time (eq. (3.5)). This is the subject

of the next chapter, where I further build on BaySTDetect and develop a detection mixture model that is more flexible to use for disease mapping and detection.

4 A more flexible Bayesian detection model for public health surveillance

4.1 Introduction

The importance to perform surveillance on public health data has been extensively discussed in this thesis (see section 1.1.1 on page 4). A better understanding of spatial and temporal patterns can help policy makers to wisely allocate funding, take preventive measures and eventually improve population health. A crucial point in public health surveillance research is the development of powerful tools to properly analyse spatio-temporal data and extract useful information. A detailed literature on such methods has been presented in section 1.3.2 on page 19.

Among the methods discussed, careful attention was paid to BaySTDetect (section 1.3.2 on page 23), a Bayesian hierarchical mixture model which is able to not only describe spatio-temporal trends of a disease, but to also detect unusual behaviour. This is of great interest from a public health point of view, as it can lead to important findings, e.g. the identification of localised risk factors. It can also be a useful tool for practitioners who wish to assess a certain hypothesis, such as the impact of an intervention. The BaySTDetect model was applied to the chronic respiratory disease (CRD) in chapter 3, showing the amount of information it can provide in epidemiological studies.

Despite its great improvements over standard disease mapping models within the small area context, it suffers from drawbacks that limit its applicability. More specifically:

1. The mixing parameter z_i in the BaySTDetect model (eq. (1.24) on page 24) assumes that the probability that area i follows the common or unusual time trend remains the same across the whole time period. This assumption is perhaps too restrictive and makes the model appropriate only for a small number of time points, i.e. Li et al. (2012) tested the model on simulated data with 8 time points (years) and recommended to be used for a maximum length of 10. The implications are different depending on the time length covered and the type of data. For instance, when monthly data are considered rather than annual for the same outcome, the temporal variability usually increases due to seasonality thus making this assumption even more restrictive.

2. The posterior estimates of z_i give information regarding the location of the unusual behaviour but not the corresponding time point/points when this occurs.

3. The prior on z_i in eq. (1.33) on page 25 assumes independence across unusual areas, however in epidemiological studies these can be spatially correlated, and not accounting for this can result in unreliable estimates.

4. In addition, z_i assumes that a priori 5% of areas are unusual (eq. (1.33) on page 25) which might induce bias when this does not agree with the data, for instance when no outbreaks exist or a much larger proportion of areas are unusual. This is particularly likely to happen when very sparse data are considered, i.e. short time series and/or low counts, where the prior can have a strong effect on the posterior. The impact of this prior specification on the

detection power of the model has been described in Li et al. (2011).

5. Because of the nature of z_i, multiple independent tests are performed and hence the multiple testing problem arises. The Bayesian false discovery rate (FDR) specification is incorporated into the baseline model to address this, however as can be seen in the original paper, the proportion of false positives is not adequately low for 2 out of 3 scenarios considered in the simulation study (see patterns 2 and 3 in Figure 3 in Li et al. (2012)).

In this chapter, I describe a more flexible modelling approach for public health surveillance. By using BaySTDetect as the baseline model, I adopt its mixture specification and further extend it to address its limitations. The proposed model is able to detect temporal changes in the spatial setting for scenarios where the baseline model lacks performance.

The remainder of this chapter is structured as follows: In section 4.2, I describe the proposed modelling approach to overcome the above limitations. In section 4.3, I present a simulation study to assess the performance of the model by comparison against the baseline model. Two case studies are carried out to illustrate the method using data on the main public health concerns of this thesis, i) road traffic accidents in England from 2005 to 2015 (section 4.4), and ii) chronic obstructive pulmonary disease (COPD) admissions in England from April 2010 to March 2011 (section 4.5). Discussion points are raised in section 4.6.

4.2 Proposed model

The proposed model follows the Poisson lognormal specification of BaySTDetect (section 1.3.2 on page 23) with eqs. (1.25) and (1.26) on page 24 remaining the same. The limitations of the baseline model are then addressed as follows:

To make the approach able to deal with long time series, the mixing parameter z_i in eq. (1.24) on page 24 becomes z_{it}; the probability of whether area i is common or unusual therefore changes across time (limitation 1). In addition, the posterior mean z_{it} provides information on both where and when unusual behaviour is observed (limitation 2).

The parameter z_{it} follows a hierarchical model:

$$z_{it} \sim \text{Bernoulli}(\phi_{it}) \tag{4.1}$$

$$\text{logit}(\phi_{it}) = \pi_i + \delta_t + \text{logit}(\tau) \tag{4.2}$$

$$\tau \sim \text{Uniform}(0.9, 1) \tag{4.3}$$

The spatial component π_i is assigned a conditional autoregressive prior $\pi_i \sim \text{CAR}(\mathbf{W}, \sigma_\pi^2)$, and the temporal component δ_t is assigned the temporal analogue of CAR, $\delta_t \sim \text{CAR}(\mathbf{Q}, \sigma_\delta^2)$, where $\mathbf{W}$ and $\mathbf{Q}$ are the spatial and temporal adjacency matrices respectively as specified in section 1.3.2 on page 23. Through this specification, the model accounts for dependences of the unusual observations in space and time (limitation 3); it is realistic to assume that the probability that area i is unusual at time point t depends on the corresponding probability of a neighbouring area at time point t, as well as the corresponding probability of area i at time points $t-1$ and $t+1$.

The hyperparameters σ_π^2 and σ_δ^2 are assigned a weakly informative half Normal prior $N(0,1)$. All parameters and hyperparameters of eqs. (1.25) and (1.26) on page 24 are specified as in BaySTDetect, except for the area-specific variance $\sigma_{i\kappa}^2$ which now also follows a less restrictive half Normal prior $N(0,1)$. Given the hierarchical structure of z_{it}, and the fact that the proposed model is appropriate for longer time series and therefore more information is available, it is no longer needed to

specify an informative prior on the area-specific variances, but rather allow for more flexibility in the detection mechanism of the model.

The term logit(τ) is also included in the model, shifting the expectation of ϕ_{it} from 0.5 to the parameter τ which is modelled hierarchically. The rationale behind the choice for the prior for parameter τ is that it needs to be relatively high so that a small proportion of observations, roughly 5–10%, follows the *Area-Specific model* (eq. (1.26) on page 24), thus characterised as unusual, while at the same time it allows for flexibility to deal with scenarios with no unusual observations. A uniform prior with a range between 0.9 and 1 is therefore specified, assuming that the proportion of unusual observations varies from 0 (when no unusual behaviour is evident) to 10% (limitation 4).

An additional positive impact of modelling z_{it} hierarchically is that low proportions of false positives are achieved (limitation 5). As stated in section 1.3.2 on page 26, within the Bayesian modelling framework the interpretation of the posterior z_{it} is straightforward; as long as the model is properly specified, accounting for known sources of variability, f_{it} = P(z_{it}=1| Y_{it}) represents the posterior probability that area i follows the common trend at time point t (Gelman et al., 2012) and therefore should not be subject to producing false detections. Since the probability model now accounts for spatial and temporal dependences through a hierarchical structure, contrary to the original BaySTDetect, the multiple testing problem is no longer evident. Areas with probability f_{it} less than the standard statistical significance level of 0.05 are selected and classed as unusual at the corresponding time point/points t.

4.3 Simulation study

A simulation study has been performed to evaluate the performance of the proposed model under various realistic scenarios, as well as to compare it against the baseline model. For comparison purposes, I followed closely the simulation design used in Li et al. (2012).

4.3.1 Data generation and study design

Count data for 211 Clinical Commissioning Groups (CCGs) in England over 15 time points were generated under the *Common model* (eq. (1.25) on page 24).

I used data on asthma hospital admissions in the same areas to (i) calculate the expected number of cases after adjusting for age and sex characteristics using indirect standardisation and (ii) obtain posterior parameter estimates from the *Common model*, which were then used to simulate the outcome variable. Summary statistics of the asthma hospitalisation dataset can be found in table C.1 in appendix C.

The simulated counts represent the behaviour of the disease under normal conditions. In order to evaluate whether the model is able to pick up unusual behaviour, I selected a number of areas to follow a temporal trend that differs from the national one. For the selection of areas I considered two different spatial scenarios: i) isolated, where unusual areas are not located close together, i.e. they do not share geographical borders, and ii) clustered, where unusual areas are close to each other, i.e. they share geographical borders.

Accounting for the fact that the spatial risk and the magnitude of the expected cases will affect the ability of the model to detect a particular area, 15 unusual areas were selected in total (for details see appendix C). For each of the two spatial scenarios, three temporal scenarios were considered: i) one where unusual time

points are isolated, ii) one where they are consecutive and characterised by a high degree of variability, and iii) one where they are consecutive and show a stable pattern, i.e. they do not show high variability (figs. C.1a to C.1c in appendix C). This will allow to evaluate whether the model is able to detect unusual behaviour when spatial and temporal dependences are evident.

In addition to the above 6 scenarios, the following are also considered: i) one scenario with 30 time points in order to evaluate the performance of the model on much longer time series (fig. C.1d in appendix C), ii) one scenario with reduced expected cases and iii) one scenario with increased expected cases, in order to assess the impact of the magnitude of the expected counts. Summary statistics for the expected counts used in the latter two scenarios can be seen in table C.1 in appendix C. Finally, I consider iv) one scenario with no aberrations in order to assess whether the model produces positives when there is no signal. All 10 scenarios are summarised in table 4.1. In order to ensure that the proposed model performs well also in the case of short time series, where the baseline model has shown to be appropriate, we carried out a simulation study with 8 time points, results of which are shown in table C.2 in appendix C. 50 simulated data were generated in total for each scenario.

4.3.2 Results

For scenarios $S1$ to $S9$ (table 4.1), the objective is to detect as many aberrations as possible while keeping a low proportion of false positives. In the case of the proposed model, an aberration is a combination of area i and time point t for which f_{it} is below 0.05. In the case of the baseline model, an aberration is an area i for which f_i is below threshold $C = f(k)$, which satisfies the rule described in eq. (1.35) on page 26 with the preset level α equal to 0.05. Model performance was assessed

Scenario	Spatial pattern	Temporal pattern	Expected cases
S1	Isolated	Isolated	Normal
S2	Isolated	Consecutive-variable	Normal
S3	Isolated	Consecutive-stable	Normal
S4	Clustered	Isolated	Normal
S5	Clustered	Consecutive-variable	Normal
S6	Clustered	Consecutive-stable	Normal
S7	Isolated	Long time series	Normal
S8	Isolated	Consecutive-variable	Increased
S9	Isolated	Consecutive-variable	Reduced
S10		no aberrations	

Table 4.1: Scenarios for the simulated data

through sensitivity $\big(\text{TP}/(\text{TP+FN})\big)$, specificity $\big(\text{TN}/(\text{TN+FP})\big)$, proportion of false positives $\big(\text{FP}/(\text{TP+FP})\big)$ and global error $\big((\text{FP+FN})/(\text{TN+TP+FP+FN})\big)$, where TP, FP, TN, FN are the numbers of true positives, false positives, true negatives, and false negatives respectively. For the proportion of false positives, when there are no declared unusual observations ($\text{TP} + \text{FP} = 0$), this is set to 0. In order to assess performance of the scenario with no aberrations ($S9$), I take the proportion of simulations (of the total 50) with at least one aberration declared, and the corresponding average number of these.

Table 4.2 presents the averaged values of the performance criteria across 50 simulations for scenarios $S1$ to $S9$. As can be seen, the mean proportion of false positives from the baseline model is above 0.2 for all scenarios, which says that at least 20% of the declared unusual observations are not truly unusual. The range of values is detailed by the scenario characteristics: the minimum, equal to 0.2, corresponds to the case where reduced expected cases are assumed, while the

highest values, 0.32 and 0.33, correspond to the scenarios with a longer time series ($S7$) and increased expected cases ($S8$) respectively. From this it can be concluded that when more information is added, i.e the data are less sparse, the performance of the baseline model in terms of false positives gets worse. While sensitivity is very high for the same model, and specificity and global error have adequate values, the high proportion of false positives makes the model inappropriate to use for scenarios with long time series, i.e. 11 time points and above. In table C.2 in appendix C, it can be seen that for short time series (8 time points), the baseline model is powerful and gives relatively low proportions of false positives, as was shown in Li et al. (2012).

The proposed method, on the other hand, gives adequately high sensitivity values and low proportion of false positives for most scenarios (table 4.2). More specifically, for scenarios $S1$, $S2$, $S4$, $S5$, $S7$, $S8$, the proposed model outperforms the baseline model: sensitivity ranges from 0.71 to 0.97, while the proportion of false positives is always below 0.05. When the temporal pattern is stable (scenarios $S3$ and $S6$), the proportion of false positives increases to 0.15 and 0.23 respectively, still lower than the corresponding values of the baseline model, i.e. 0.26 and 0.29. In the case of reduced expected cases ($S9$), the proposed model has a mean proportion of false positives equal to only 0.01, however sensitivity is 0.5, indicating that it is not able to pick up unusual behaviour sufficiently. In the case of short time series, it seems that the baseline model is slightly better than the proposed model overall, however the associated mean proportion of false positives is higher than 0.05 for all scenarios.

In the extreme case of no aberrations ($S10$), a proportion of 8% simulations (of the total 50) produces positives with an average number of declared aberrations equal to 1.25, contrary to a corresponding 26% with an average 1.31 produced by

Scenario	%FP	Sensitivity	Specificity	Global error
Baseline model				
S1	0.250 (0.211, 0.286)	0.965 (0.933, 1.000)	0.976 (0.969, 0.980)	0.025 (0.019, 0.028)
S2	0.264 (0.211, 0.296)	1.000 (0.985, 1.000)	0.972 (0.969, 0.980)	0.027 (0.019, 0.032)
S3	0.271 (0.222, 0.318)	0.967 (0.933, 1.000)	0.972 (0.964, 0.980)	0.029 (0.024, 0.033)
S4	0.268 (0.250, 0.318)	1.000 (0.999, 1.000)	0.971 (0.964, 0.974)	0.027 (0.024, 0.033)
S5	0.286 (0.286, 0.318)	1.000 (1.000, 1.000)	0.969 (0.964, 0.969)	0.029 (0.028, 0.033)
S6	0.295 (0.250, 0.318)	1.000 (1.000, 1.000)	0.968 (0.964, 0.974)	0.030 (0.024, 0.033)
S7	0.322 (0.286, 0.306)	1.000 (0.933, 1.000)	0.963 (0.959, 0.969)	0.035 (0.028, 0.041)
S8	0.327 (0.286, 0.348)	1.000 (1.000, 1.000)	0.962 (0.959, 0.969)	0.035 (0.028, 0.038)
S9	0.200 (0.143, 0.222)	0.857 (0.800, 0.933)	0.984 (0.980, 0.990)	0.025 (0.019, 0.032)
Proposed model				
S1	0.022 (0.000, 0.036)	0.710 (0.671, 0.750)	1.000 (0.999, 1.000)	0.006 (0.005, 0.007)
S2	0.019 (0.015, 0.031)	0.796 (0.763, 0.827)	1.000 (0.999, 1.000)	0.005 (0.004, 0.006)
S3	0.150 (0.118, 0.187)	0.799 (0.762, 0.814)	0.998 (0.997, 0.999)	0.005 (0.004, 0.006)
S4	0.030 (0.000, 0.052)	0.881 (0.838, 0.917)	0.999 (0.999, 1.000)	0.003 (0.002, 0.003)
S5	0.029 (0.014, 0.041)	0.928 (0.907, 0.947)	0.999 (0.999, 1.000)	0.002 (0.002, 0.003)
S6	0.225 (0.190, 0.266)	0.932 (0.911, 0.956)	0.996 (0.995, 0.997)	0.005 (0.004, 0.006)
S7	0.027 (0.022, 0.034)	0.750 (0.710, 0.770)	0.973 (0.966, 0.978)	0.008 (0.007, 0.008)
S8	0.057 (0.040, 0.075)	0.969 (0.960, 0.987)	0.999 (0.998, 0.999)	0.002 (0.002, 0.003)
S9	0.012 (0.000, 0.027)	0.480 (0.427, 0.507)	1.000 (1.000, 1.000)	0.013 (0.012, 0.014)

Table 4.2: Posterior means of performance measures for scenarios S1-S9; credible intervals are in parentheses.

the baseline model (table 4.3).

	Baseline model	Proposed model
%sims that detect at least 1 aberration	26%	8%
Mean number of aberrations detected	1.31	1.25

Table 4.3: Performance for scenario S10

4.4 Case study 1: Road traffic accidents

For illustration of the method, a case study is carried out on a set of road traffic accidents in England. This study can be considered as an extension of the analysis in chapter 2, where accidents of low and high severity were modelled jointly through a multivariate disease mapping model, and dangerous areas were highlighted via probability ranking maps. Here, I analyse accidents of high severity using the proposed model in order to describe the overall accident risk across space and time, and to identify areas with unusual temporal trends via the detection mechanism that is integrated into the model.

4.4.1 Data

The road traffic accidents data are held by the Department for Transport. Details as to how the data are collected and manipulated can be found in section 2.2.1 on page 36. Accidents of high severity level (severe and fatal) were aggregated for this study at the district level in England (326 in total). As more data became available since the first study on road traffic accidents, the temporal window considered here increased by 2 years (2005 to 2015). Summary statistics for the accidents data are presented in table C.3 in appendix C.

It has been previously discussed that the choice of offset in traffic accidents

analysis is an important aspect and needs to be considered carefully (see section 2.1 on page 33). For the offset data, I followed the same protocol that is described in section 2.2.2 on page 38 and estimated the traffic volume dataset at the district level in England using R. Given annual average daily flow (AADF) data for the major road network (motorways and A-roads) in England for the years 2005 to 2015, also obtained from the Department for Transport, and the districts shapefile, I calculated district-level traffic volume counts based on the equations

$$\mathrm{TV_D} = \sum_{\mathrm{RS}\in\mathrm{D}} \mathrm{TV_{RS}} \tag{4.4}$$

$$\mathrm{TV_{RS}} = \mathrm{length(RS)} \times \mathrm{AADF_{RS}} \tag{4.5}$$

where TV is the traffic volume, D represents the district level, and RS represents the road segment.

The traffic volume data were calculated for all years considered in the study, i.e. 2005 to 2015, and the correlation was found to be 0.99. The middle year 2010 was therefore selected for the analysis.

4.4.2 Results

The model was implemented in R using the R2OpenBUGS package. Two chains were run for 80000 MCMC iterations storing every 2nd iteration, from which 20000 were discarded as a burn-in. The model took 5 hours on an Intel Core processor at 3.40 GHz with 16 Gbytes of RAM. I carried out convergence diagnostics through visual check of the trace plots, Brooks-Gelman-Rubin (BGR) statistic, autocorrelation plots, along with assessing the Monte Carlo error.

Figure 4.1 plots the posterior estimates of the spatial component of the *Common* model (eq. (1.25) on page 24) showing the overall pattern of accident risk across

England.

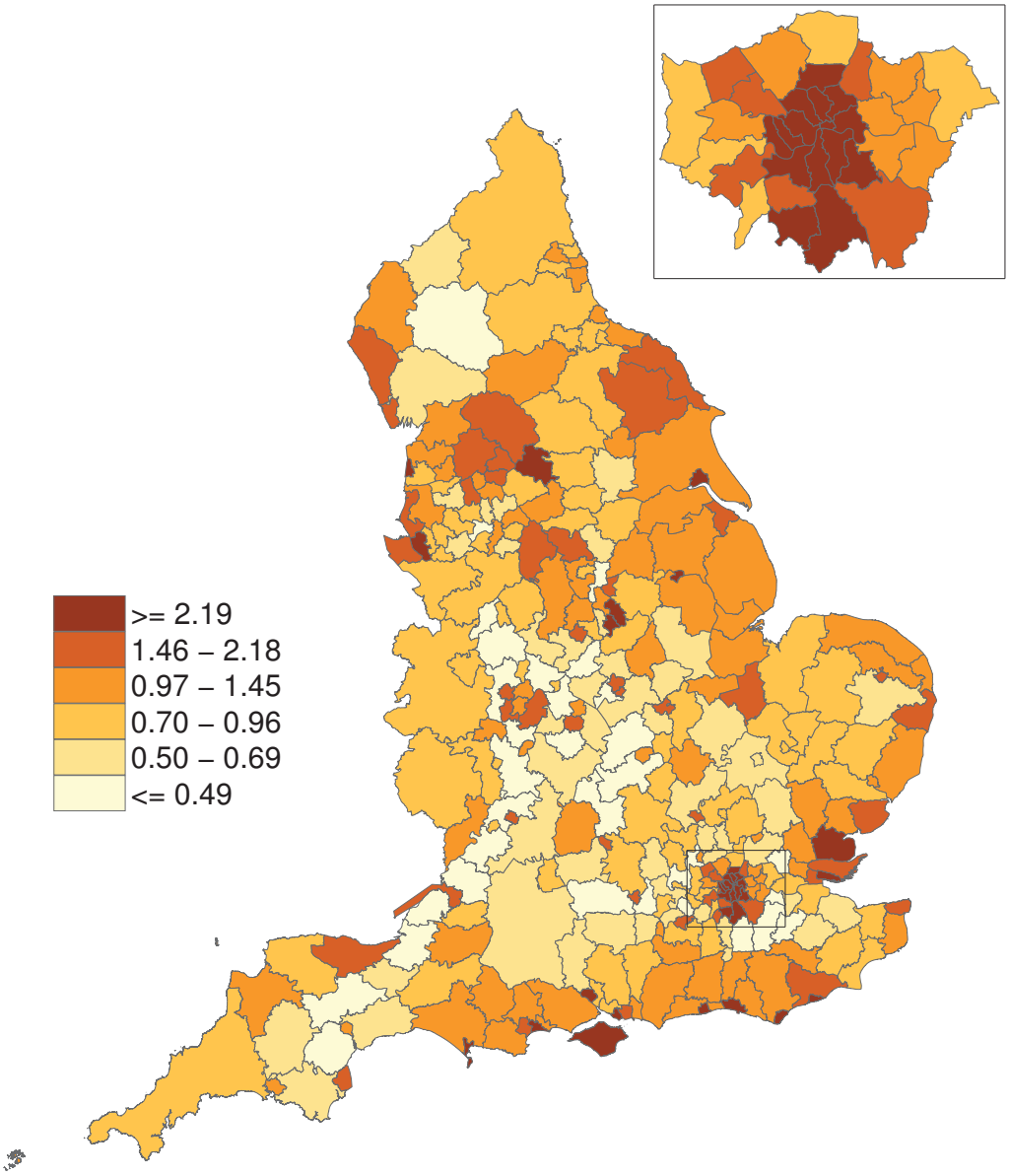


Figure 4.1: Posterior means of the spatial accident rate $\exp(\eta_i)$ across districts in England; the insets show enlarged views of London.

The map suggests higher accident risk in London and other big cities such as

Manchester and Leeds, while low accident risk is observed along certain motorways such as M1 and M5. The posterior estimate of the temporal component of the *Common model* in fig. 4.2, shows the pattern of accident risk over the period 2005 to 2015, which can be thought of as the national trend. There is a nearly linear downward trend over time up until 2010, while afterwards the trend becomes flatter. Overall results agree with those found in chapter 2, where a reduced road traffic accidents dataset was analysed (years 2005 to 2011) at a smaller spatial resolution using a disease mapping model.

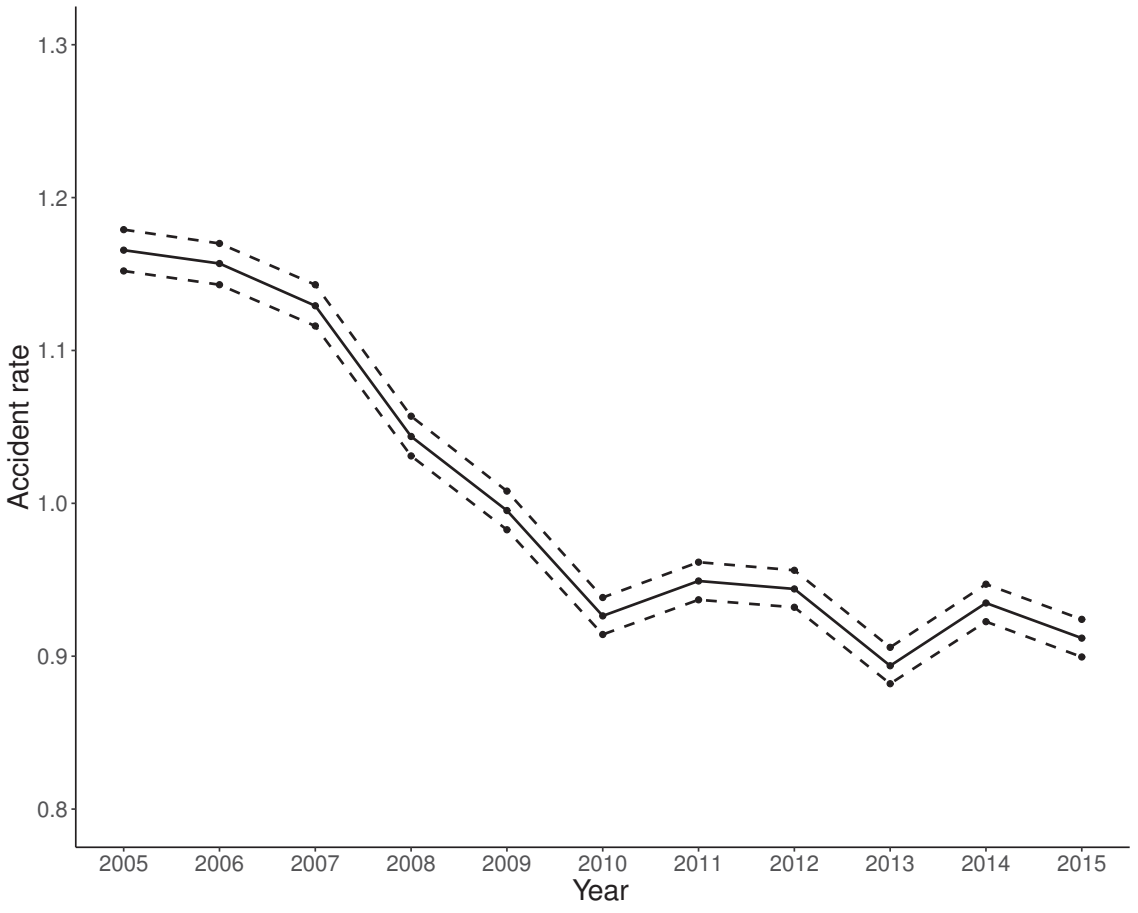


Figure 4.2: Temporal trend of accident rates in England over 2005-2015; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% values.

The temporal trend of accident risk in England over the years 2005 to 2015 is plotted in fig. 4.2. Under the model, 96 observations were detected as unusual out of which 44 correspond to unique areas (see map in fig. C.3, appendix C). These are in the form of isolated areas as well as clustered. The unusual time trends of all 44 areas can be found in figs. C.4 and C.5. Here I present the area-specific temporal trends of three selected areas plotted against the national trend (fig. 4.3). These correspond to the parameter κ_{it} of the *Area-specific model* (eq. (1.26) on page 24), where i is the selected unusual area.

The district of Eden in Cumbria shows a temporal trend with a much higher accident rate than the national average at the beginning of the time period, with an aberration detected by the model in year 2005. According to published reports, the rates of deaths and serious injuries in Cumbria are significantly higher than the national average, with the district of Eden having the highest rate in the county (Critchley and Whitfield, 2015; Brown, 2015).

The district of Southampton follows a nearly opposite temporal pattern to the national one, with aberrations detected at several time points over the study period and the highest accident risk observed in 2011. This might be related to the increased popularity of cycling in Southampton potentially leading to high accident rates. In fact, the time trend of the accident risk in Southampton somewhat resembles the trend corresponding to cyclist numbers in the area (Southampton City Council, 2016).

Similarly, the district of Elmbridge in Surrey exhibits an overall upward trend with an aberration detected in year 2015, a finding which is in line with recent reports ranking Surrey among the areas with the sharpest increase in the numbers of people killed or seriously injured on the roads (Surrey Police, 2015). In specific, road accidents involving vans and lorries increased by a third in Surrey in 2015

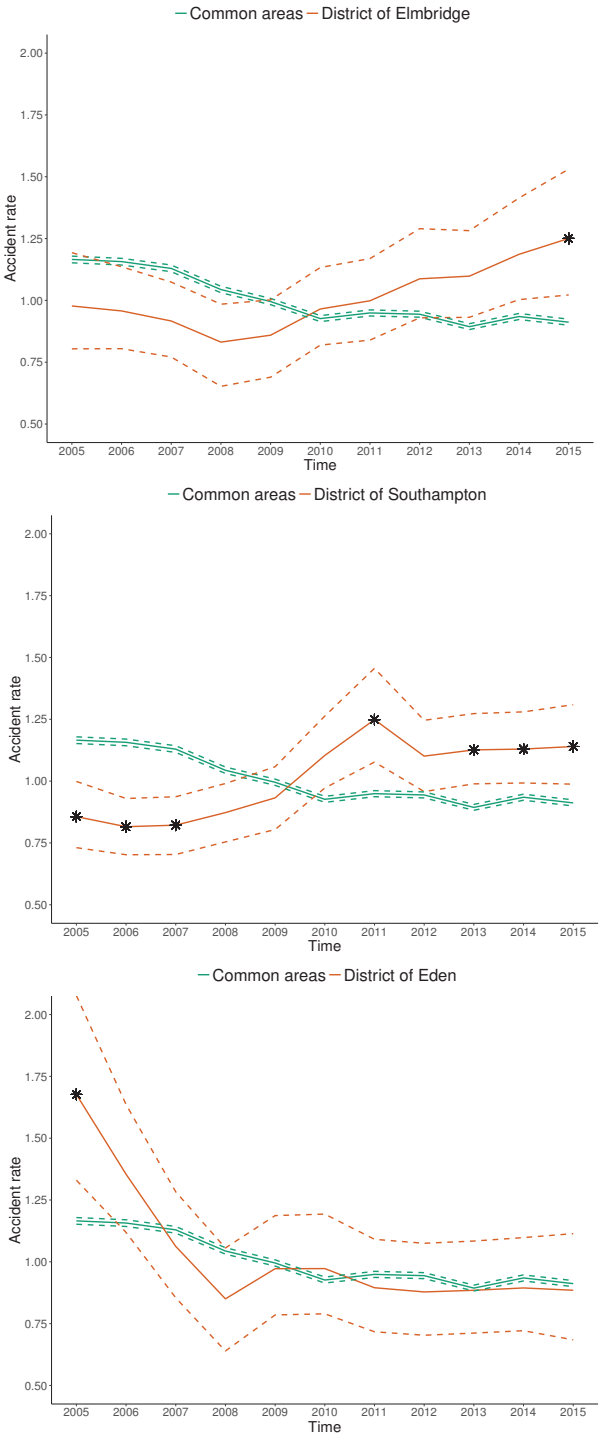


Figure 4.3: Temporal trends of accident rates for three selected unusual districts plotted against the national trend in England; the solid lines show the posterior mean of the temporal effects ($\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines show the corresponding 95% CI values. $*$ represents the time of the aberration.

compared to the previous year, while the number of pedestrians who were injured on the roads of Elmbridge rose by 54% for the same year (Scott E, 2016a,b).

4.5 Case study 2: COPD hospital admissions

An additional case study is performed on COPD hospital admissions. Following the investigation of asthma and COPD in England over a period of 8 months (August 2010 to March 2011) by using the BaySTDetect model in chapter 3, I further explore COPD hospital admissions in England over 12 months (April 2010 to March 2011) using the proposed modelling framework.

4.5.1 Data

Hospital Episode Statistics (HES) admission data, were obtained from the Small Area Health Statistics Unit (SAHSU) at Imperial College London for England from April 2010 to March 2011. The data contain the number of admissions with a primary diagnosis of COPD, and may also include readmissions of the same patient. Population data were also obtained from the Office for National Statistics (ONS), with individual-level information on age and sex which were used for the calculation of the expected number of cases. COPD admissions and population data were aggregated at CCG level. Summary statistics can be found in table C.4 in appendix C.

4.5.2 Results

Similarly to the case study on road traffic accidents data, the model was implemented in R using the R2OpenBUGS package, and two chains were run for 80000 MCMC iterations storing every 2nd iteration, from which 20000 were discarded as a burn-in.

The model for this analysis took 7 hours on an Intel Core processor at 3.40 GHz with 16 Gbytes of RAM. Convergence diagnostics were carried out to assess convergence of the model.

Figure 4.4 plots the posterior estimates of the spatial component of the *Common model* (eq. (1.25) on page 24) showing the overall pattern of the relative risk across England. The map shows a very similar pattern to the one observed in fig. 3.1b on page 73, which corresponds to asthma and COPD admissions over a subset time period of 8 months. This may suggest that COPD dominates over asthma, which is also evident from the respective counts in table 3.2 on page 68, but also that the two diseases share very similar spatial patterns.

The posterior estimates of the temporal component of the *Common model* are plotted in fig. 4.5. This shows the pattern of COPD admissions over the 12-month period considered in the study which seems to follow closely the corresponding pattern observed under the case study in chapter 3 (fig. 3.3 on page 76) for the shared time period (August 2010 to March 2011).

Under the model, 24 observations were detected as unusual out of which 20 correspond to unique areas (see map in fig. C.6 in appendix C). The unusual time trends of all 20 areas can be found in fig. C.7. It can be seen that 3 out of the 4 areas that were detected as unusual for the asthma and COPD admissions using the original BaySTDetect model (fig. 3.5 on page 80) are also detected as unusual in this study under the proposed model. The corresponding temporal trends are plotted in fig. 4.6 against the national trend, and they also show quite similar patterns to the ones observed using the original BaySTDetect model. The fact that a larger number of areas are detected here is likely related to the flexible specification of the probability model in eq. (4.1).

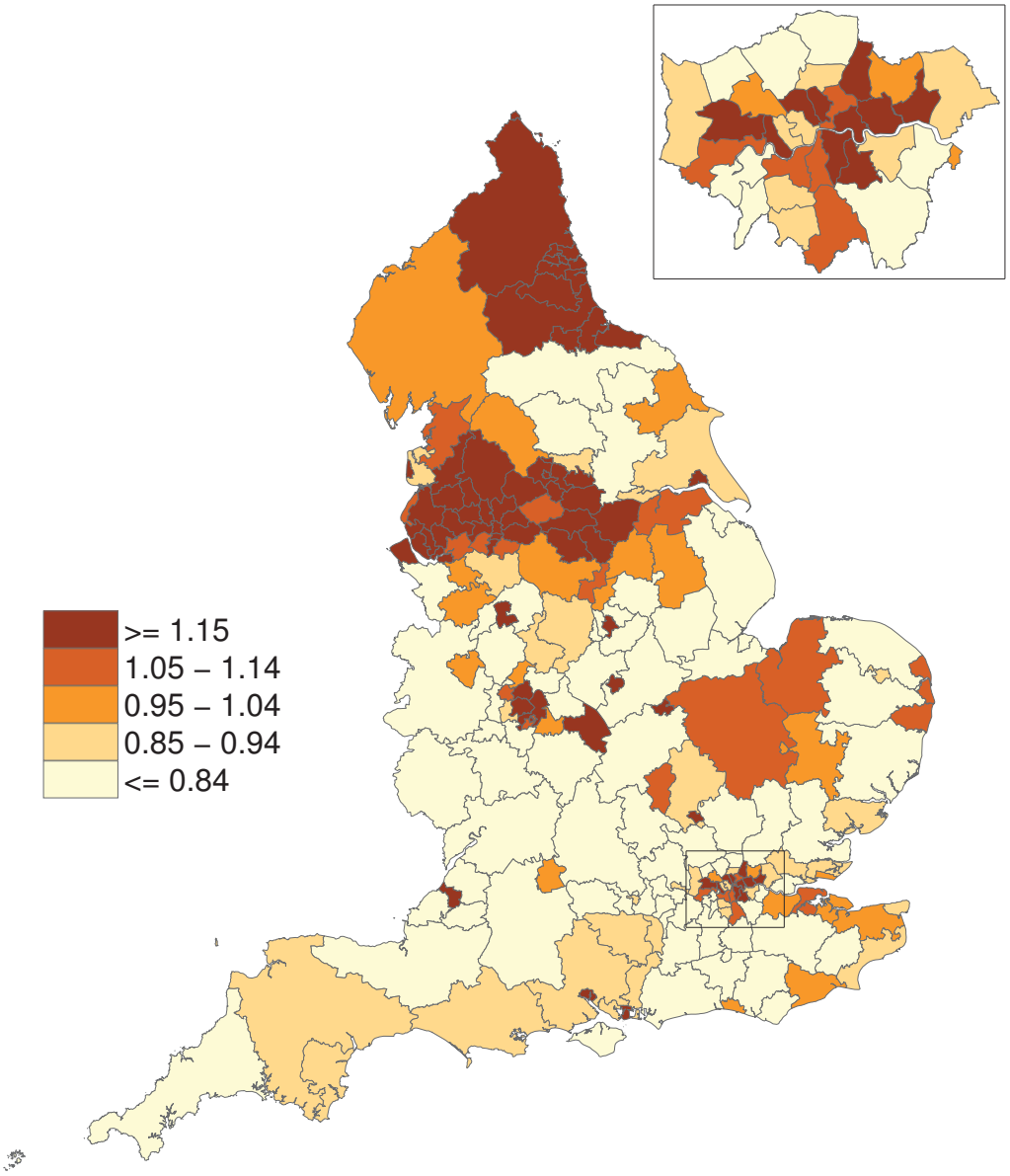


Figure 4.4: Posterior means of the spatial relative risk $\exp(\eta_i)$ of COPD admissions across CCGs in England; the insets show enlarged views of London.

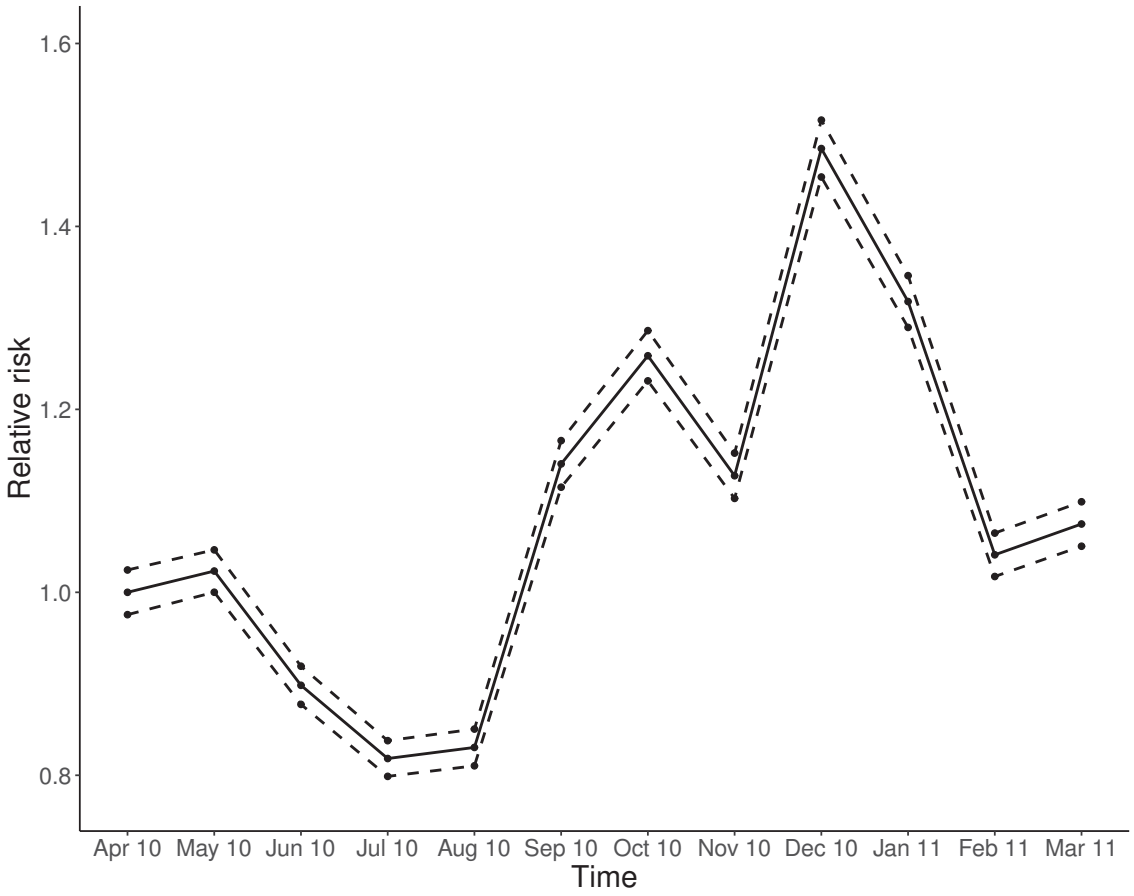


Figure 4.5: Temporal trend of COPD admissions in England; the solid line shows the posterior mean of the temporal effects $\exp(\delta_t)$ and the dashed lines represent the corresponding 95% CI values.

4.6 Discussion

In this chapter, I have described a Bayesian spatio-temporal model for public health surveillance that can be used for the detection of areas with unusual behaviour over time. I used the BaySTDetect method as a baseline, and overcame its limitations extending it to a more generic framework. Allowing the mixing parameter to change across time results in a model which is appropriate for longer time series, a case where the original BaySTDetect model showed lack of performance. Information

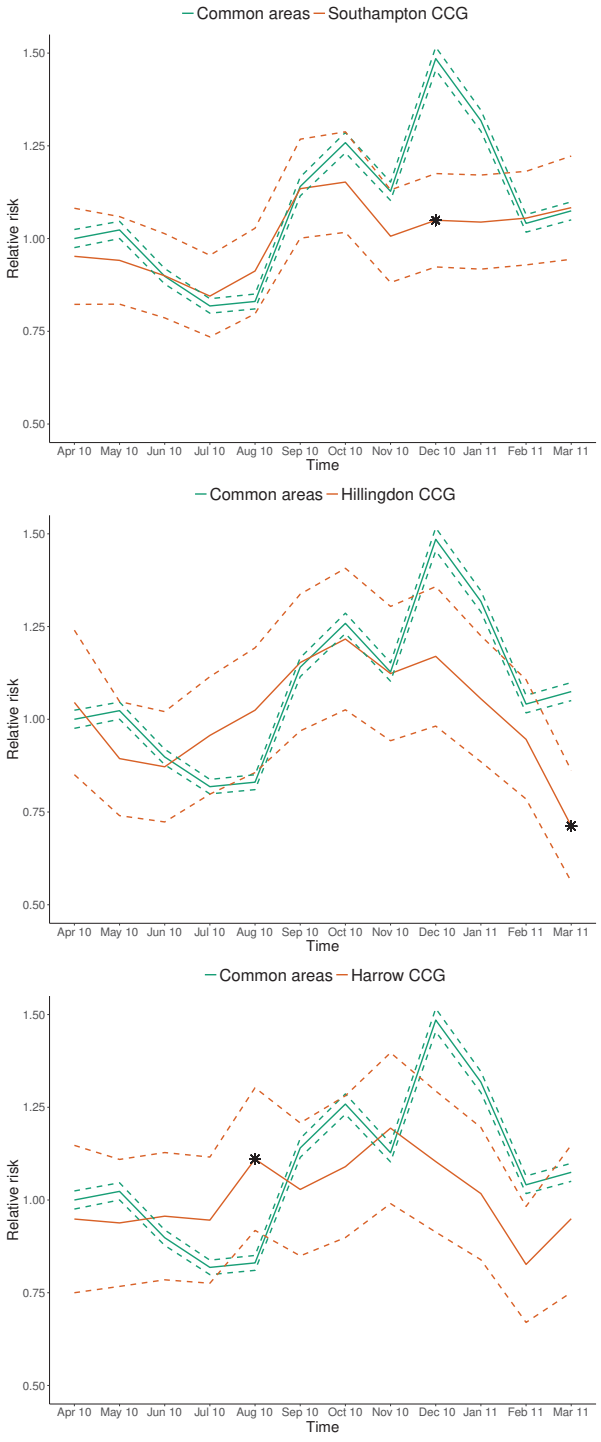


Figure 4.6: Temporal trends of COPD admissions for three selected unusual CCGs plotted against the national trend in England; the solid lines show the posterior mean of the temporal effects ($\exp(\kappa_{it})$ for the unusual trends and $\exp(\delta_t)$ for the national trend) and the dashed lines show the corresponding 95% CI values. $*$ represents the time of the aberration.

on the time points when unusual behaviour occurs is also provided to facilitate interpretation and aetiological investigation. The prior of the mixing parameter changes to a less informative hierarchical prior that accounts for spatial and temporal dependences making the model more flexible for detection of several realistic unusual patterns.

Surveillance methods are typically optimised for one sole criterion depending on the problem at hand. In epidemiological surveillance, where the financial resources are usually limited and the priority is to spend expenditure wisely, it is very important to control the proportion of false positives. The baseline model, for instance, which is able to correctly detect nearly all unusual observations as such, at the cost of excess numbers of false detections, is not appropriate in this context. The proposed model, on the other hand, has the great advantage of keeping the proportion of false positives below 5% for most scenarios. This suggests that the user can be confident that the observations that are detected as unusual under the model are indeed unusual. In addition, unusual behaviour is detected only when this is evident, as shown from the scenario with no aberrations ($S10$), thus confirming the hypothesis that the strong prior specification on z_i of the baseline model induces bias when this does not agree with the data. At the same time, the corresponding sensitivity and specificity values are satisfactory.

An additional strength of the proposed model is that there is no need to adjust for the multiple testing problem as in the case of the baseline model. While this is important particularly in classical inference, the relevant problem in the Bayesian setting is not multiple testing but rather not appropriate modelling (Gelman et al., 2012). The hierarchical specification of the mixing parameter in the proposed model in eq. (4.1) allows to define naturally the rule for classifying observations as unusual, simply by selecting those with a probability of following the common trend less

than 0.05.

The model is designed specifically for detection of a small proportion of unusual observations when these exist, ranging a priori between 0 and 10%. The upper bound could be chosen based on each specific case; however a much larger proportion of unusual trends, say 20% and more, wouldn't be seen as unusual but rather as a different common trend. The objective would then be clustering of multiple time trends, and a model following nonparametric Bayesian approaches would be more appropriate (see for instance Dass et al., 2015).

From a policy making perspective, the proposed model could be especially valuable for healthcare professionals who can use it for spatial monitoring at the population level, in order to detect areas with unusual behaviour over time. Detecting areas with increased demand is essential for implementation of targeted intervention strategies, however further investigation would be normally required to better understand unusual patterns, and generate hypotheses regarding associated risk factors.

While the model has shown its potential in detecting unusual patterns under various temporal and spatial scenarios, its performance was less satisfactory when the unusual time trend was stable (i.e. consecutive-stable patterns), as it gave a relatively high proportion of false positives. On the other hand, when the unusual trend is more variable (i.e. isolated and consecutive-variable patterns), the model is more conservative as the proportion of false positives remains very low, but also the power slightly decreases. This is perhaps because of the space-time separable structure of the probability model in eq. (4.1) and a solution to this could be the inclusion of an additional parameter in order to allow for space-time interactions.

In addition, the model was not able to adequately pick up signal when expected cases were reduced to half. This might be due to the specification of the area-specific

variances, which are defined on a specific scale of data. Potentially a way to address this is to assume that these are relative to the variance of the *Common model*. It was also shown that for short time series (8 time points) there is no additional benefit of using the proposed model rather than the baseline model. Nevertheless, in the context of public health surveillance, the availability of data has recently shifted the interest from yearly to monthly or weekly time resolution, hence resulting in longer time series where the proposed approach would be beneficial.

5 Conclusions

Motivated by various public health problems, from road traffic accidents to non-communicable diseases, this thesis contributes to the scientific literature in the field by providing tools to properly analyse complex spatio-temporal data so as to extract information that will help improve population health.

Three main research objectives were investigated in this thesis, namely i) to develop a model for the joint analysis of road traffic accidents by severity level, ii) to employ a recently proposed detection model for the investigation of chronic respiratory diseases (CRDs), and iii) to address the limitations of the aforementioned model by further extending it to a flexible framework for public health surveillance.

5.1 Summary of findings

The first research objective was addressed in chapter 2, where a number of Bayesian hierarchical models were developed to jointly analyse spatio-temporal road traffic accidents data of low and high severity level. An extensive comparison among the competing models differing in the dependence assumptions of severities was carried out in order to identify the best model for the data at hand. The model with a multivariate specification in the spatial effects and independent specification

in the temporal effects was found to be the best, indicating similarities in the underlying spatial structures of slight and severe traffic accidents, and differences in the corresponding temporal structures. The model was evaluated based on several goodness of fitness criteria, including Bayesian p-values and predictive probabilities. Overall results showed improvement in estimating severe accidents, which were characterised by particularly small counts, as these were strengthened by slight traffic accidents, characterised by large counts and therefore more power. Probability ranking maps were also constructed in order to provide detection, suggesting the most dangerous areas in terms of accident risk.

Chapter 3 addressed the second research objective, by employing BaySTDetect to investigate the CRD at the small area level in England over an 8 month period. Mortality and hospital admission data on asthma and chronic obstructive pulmonary disease (COPD) were analysed alongside a set of drugs prescribed to asthma and COPD patients. The conclusions of this study were twofold. On the one hand a clearer picture of CRD patterns were provided through modelling different data sources representing patients of different severity levels, while on the other hand examination into the strengths and limitations of the detection model was carried out.

Chapter 4 follows from the results of chapter 3; the limitations of BaySTDetect were first described, and then an improved version of the model was proposed. The model is able to deal with longer time series, to account for different possible sources of probability in the data, and to provide information on the time when unusual behaviour occurs. Focusing more on the methodological aspects of disease detection, chapter 4 provided a simulation study considering different spatial and temporal patterns, and different time lengths, in order to assess the performance of the proposed model. By comparison against the original BaySTDetect, results

showed great improvements in terms of detection power as well as in terms of proportions of false positives.

5.2 Contributions to the state-of-the-art

The outputs of this thesis extend beyond the scope of each chapter and are expected to add value to the state-of-the-art both from an epidemiological and a statistical point of view.

The model presented in chapter 2 may be useful in other public health applications which would benefit from a joint modelling framework. While the proposed multivariate specification cannot be generally assumed to be the most appropriate for any multivariate dataset, but it is rather specific to the road traffic accidents data analysed in the respective chapter, the overall modelling process can be generally adopted for joint data analyses. For example, to analyse non-communicable diseases that are known to share characteristics, one may develop alternative models as to the dependence structures and perform comparisons in order to find the most appropriate. In addition, the exceedance probability maps can be particularly useful in highlighting areas characterised by excess risk, and the concept of area ranking, although so far mainly used in road traffic accidents analyses, may also benefit applications on non-communicable diseases.

The outcomes of chapter 3 contribute to the CRD literature by providing considerable amount of information as to the behaviour of the disease in space and time. Asthma and COPD attract a lot of interest from public health researchers and policy makers, who wish to take appropriate measures to address the burden of the disease. By comparing the behaviour of the disease across different data sources underlining different condition severity through maps and temporal plots, several

hypotheses as to the aetiology may be generated and examined. In addition, certain detected areas are expected to be further investigated, such as the Isle of Wight which was found to exhibit an alarming temporal mortality trend.

Chapter 4 contributes to the state-of-the-art by establishing a modelling framework for disease mapping and detection, using the BaySTDetect model as the foundation. The extensive simulation study that was conducted showed great improvements of the proposed model over BaySTDetect, suggesting the potential of the model for space-time detection within public health surveillance. The two case studies performed, on road traffic accidents and COPD data, showed that the framework can help with epidemiological investigations of various public health problems characterised by specific spatio-temporal structures.

Besides the aforementioned contributions, the contents of this thesis may serve as the basis for further research within the Bayesian spatio-temporal setting for disease mapping and detection, an active area of research which keeps growing steadily.

5.3 Future directions

A number of gaps in the literature were addressed in this thesis, while open problems that offer potential for future research were also identified.

First, within public health surveillance there is increased importance of early warning detection, that is, to be able to locate abnormal behaviour at the earliest possible time. Prospective modelling approaches are therefore required, where the data are fitted sequentially, one time point at a time. Syndromic data, including ambulance calls, school absenteeism and over-the-counter prescription sales may be used in this setting as they provide an additional level of information often

leading to detection even before the aberration occurs (Corberán-Vallet and Lawson, 2014). While the applicability of such approaches is not obvious in the context of non-communicable diseases where changes usually occur over longer time periods, a routine chronic disease surveillance system could nevertheless benefit certain conditions associated to risk factors that change rapidly with time (Yiannakoulias et al., 2009). For instance, asthma is known to be influenced by infectious disease trends and environmental factors, such as air pollution and allergens. Routine and ongoing spatial surveillance is also critical in the context of road safety, where there is a need for improved traffic safety assessment techniques in order to allow for instant response when abnormal events occur. Little has been done in this area, with most advances coming from the engineering field (Krishnaveni and Hemalatha, 2011; You et al., 2017).

Second, the computational burden increases exponentially with the dimension of the data. It is therefore essential to consider alternatives to the standard MCMC algorithms. A recent addition to the literature is the Integrated Nested Laplace Approximations (INLA), a new inferential tool used to approximate a wide range of Bayesian models with a latent Gaussian structure (Rue et al., 2009; Lindgren and Rue, 2015; Blangiardo and Cameletti, 2015). The main advantage of INLA over MCMC methods is the low computational cost, however due to the underlying assumption of latent Gaussian random field mixture models, these cannot be readily implemented. Recent work by Gómez-Rubio and Rue (2017) shows that there is a potential of expressing a mixture model as a conditional latent Gaussian model and fit it using INLA within MCMC, yet the computational gain of this approach over MCMC is not apparent yet.

Third, despite the proven potential of Bayesian hierarchical spatio-temporal models in the context of disease mapping and lately in the detection of disease,

they often have limitations in their ability to handle large datasets, and to discover relationships which are not pre-specified. For example, a fully parametric model such as the ones that were used in this thesis requires the inclusion of parameters accounting for different sources of probability in the data. This can be particularly challenging when analysing several conditions simultaneously with complex spatio-temporal structures. Given the increasing availability of healthcare data, the development of more powerful approaches is thus necessary.

An alternative to the standard Bayesian models is represented by non-parametric machine learning methods, which use infinite-dimensional parameter space; given the absence of a parametric structure, these are extremely flexible in learning the underlying relationships of complex datasets. These methods have been widely applied in disciplines known for their high-dimensional nature of data, yet research in the area of epidemiology and small area health data has only scratched the surface. A notable exemption are the Dirichlet Process (DP) mixture models (Ferguson, 1973) which have been recently extended to several directions, incluiding the spatial DP by Gelfand et al. (2005) for spatial data. A key example is the approach suggested by (Kottas et al., 2008), which is a mixture model for space-time disease data, with potential directions to disease detection. Bayesian non-parametric spatio-temporal models seem to be promising within the disease mapping and detection context and further research is required to explore this.

A final research direction that is worth considering concerns the importance of communicating information effectively to public health researchers and policy makers. The development of user friendly tools such as web applications and reporting services that allow for easy implementation of advanced methodologies and interactive data visualisations are therefore essential; practitioners who may lack adequate programming skills and statistical knowledge to carry out complex analyses will be

able to implement sophisticated methods which provide an important opportunity to improve public health.

Bibliography

Abellan, J. J., S. Richardson, and N. Best (2008). Use of space - time models to investigate the stability of patterns of disease. *Environmental Health Perspectives 116*(8), 1111.

Ackaah, W. and M. Salifu (2011). Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. *IATSS Research 35*(1), 34 – 40.

Aguero-Valverde, J. (2011). Direct spatial correlation in crash frequency models. In *3rd International Conference on Road Safety and Simulation*.

Aguero-Valverde, J. and P. Jovanis (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention 38*(3), 618–625.

Aguero-Valverde, J. and P. Jovanis (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record: Journal of the Transportation Research Board 2136*(1), 82–91.

Aguero-Valverde, J. and P. P. Jovanis (2007). Identifying road segments with high risk of weather-related crashes using full Bayesian hierarchical models. In *86th Annual Meeting of the Transportation Research Board, Washington, DC*.

Al-Osh, M. A. and A. A. Alzaid (1987). First-Order Integer-Valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis 8*(3), 261–275.

Aldor-Noiman, S., L. D. Brown, E. B. Fox, and R. A. Stine (2013). Spatio-temporal low count processes with application to violent crime events. *arXiv preprint arXiv:1304.5642*.

Andersen, Z. J., M. Hvidberg, S. S. Jensen, M. Ketzel, S. Loft, M. Sørensen, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen (2011). Chronic obstructive pulmonary disease and long-term exposure to traffic-related air pollution: a cohort study. *American journal of respiratory and critical care medicine 183*(4), 455–461.

Arsenault, J., P. Michel, O. Berke, A. Ravel, and P. Gosselin (2013). How to choose geographical units in ecological studies: proposal and application to campylobacteriosis. *Spatial and spatio-temporal epidemiology 7*, 11–24.

Banks, D., G. Datta, A. Karr, J. Lynch, J. Niemi, and F. Vera (2012). Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice. *Inf. Fusion 13*(2), 105–116.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions 53*, 370–418.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Bennett, J. E., G. Li, K. Foreman, N. Best, V. Kontis, C. Pearson, P. Hambly, and M. Ezzati (2015). The future of life expectancy and life expectancy inequalities in england and wales: Bayesian spatiotemporal forecasting. *The Lancet 386*(9989), 163–170.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine 14*(21-22), 2433–2443.

Bernardo, J. M. and A. F. M. Smith (1994). Bayesian theory Wiley. *New York*.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 36*(2), 192–236.

Besag, J., J. York, and A. Mollie (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*(1), 1–20.

Best, N., S. Richardson, and A. Thomson (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research 14*(1), 35–59.

Best, N. and J. Wakefield (1999). Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 162*(3), 363–382.

Bijleveld, F., J. Commandeur, S. J. Koopman, and K. v. Montfort (2010). Multivariate non-linear time series modelling of exposure and risk in road safety research. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 59*(1), 145–161.

Bijleveld, F. D. (2005). The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis & Prevention 37*(4), 591–600.

Blangiardo, M. and M. Cameletti (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Box, G. E. and G. M. Jenkins (1976). *Time series analysis: forecasting and control, revised ed*. Holden-Day.

Briët, O. J., P. H. Amerasinghe, and P. Vounatsou (2013). Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. *PloS one 8*(6), e65761.

Brijs, T., D. Karlis, F. Van den Bossche, and G. Wets (2007). A Bayesian model for ranking hazardous road sites. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(4), 1001–1017.

Brooks, S. P. and G. O. Roberts (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing 8*(4), 319–335.

Brown, R. E. (2015). Avoidable mortality in Cumbria: A review of 73 fatal road traffic collissions. `http://www.cph.org.uk/wp-content/uploads/2015/04/A-Review-of-73-Fatal-Road-Traffic-Collisions-in-Cumbria_RBrown-2.pdf`. Accessed: 2017-09-11.

Carlin, B. P., A. E. Gelfand, and A. F. Smith (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied statistics*, 389–405.

Christiansen, C., C. Morris, and O. Pendleton (1992). A hierarchical Poisson model with beta adjustments for traffic accident analysis. *Center for Statistical Sciences Technical Report 103*.

Clayton, D. and J. Kaldor (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671–681.

Coleman, M., M. Coleman, A. M. Mabuza, G. Kok, M. Coetzee, and D. N. Durrheim (2009). Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malaria Journal 8*(1), 68.

Congdon, P. (1996). Suicide and parasuicide in London: a small-area study. *Urban studies 33*(1), 137–158.

Corberán-Vallet, A. and A. B. Lawson (2014). Prospective analysis of infectious disease surveillance data using syndromic information. *Statistical methods in medical research 23*(6), 572–590.

Cowles, M. K. and B. P. Carlin (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association 91*(434), 883–904.

Cressie, N. and A. S. Mugglin (2000). Spatio-temporal hierarchical modeling of an infectious disease from (simulated) count data. In *Compstat*, pp. 41–52. Springer.

Cringoli, G., L. Rinaldi, V. Veneziano, and V. Musella (2005). Disease mapping and risk assessment in veterinary parasitology: some case studies. *Parassitologia 47*(1), 9–25.

Critchley, K. and M. Whitfield (2015). TIIG Cumbria: Themed Report. `http://www.cph.org.uk/wp-content/uploads/2015/02/TIIG-Cumbria-Road-traffic-collisions-across-Cumbria-2011-12-to-2013-14.pdf`. Accessed: 2017-09-05.

Dass, S. C., C. Y. Lim, T. Maiti, and Z. Zhang (2015). Clustering curves based on change point analysis: A nonparametric Bayesian approach. *Statistica Sinica*, 677–708.

de Hoogh, K., M. Wang, M. Adam, C. Badaloni, R. Beelen, M. Birk, G. Cesaroni, M. Cirach, C. Declercq, A. Dėdelė, E. Dons, A. de Nazelle, M. Eeftens, K. Eriksen, C. Eriksson, P. Fischer, R. Gražulevičienė, A. Gryparis, B. Hoffmann, M. Jerrett, K. Katsouyanni, M. Iakovides, T. Lanki, S. Lindley, C. Madsen, A. Mölter, G. Mosler, G. Nádor, M. Nieuwenhuijsen, G. Pershagen, A. Peters, H. Phuleria, N. Probst-Hensch, O. Raaschou-Nielsen, U. Quass, A. Ranzi, E. Stephanou, D. Sugiri, P. Schwarze, M.-Y. Tsai, T. Yli-Tuomi, M. J. Varrò, D. Vienneau, G. Weinmayr, B. Brunekreef, and G. Hoek (2013). Development of land use regression models for particle composition in twenty study areas in Europe. *Environmental Science & Technology 47*, 5778–5786.

Dempster, A. P. (1997). The direct use of likelihood for significance testing. *Statistics and Computing 7*(4), 247–252.

Department for Transport (2010). Road Accident Statistics Branch, Road Accident Data, 2009 [computer file]. Technical report, Colchester, Essex: UK Data Archive [distributor].

Drazen, J. M., D. S. Postma, and K. F. Rabe (2015). The asthma–COPD overlap syndrome. *New England Journal of Medicine 373*(13), 1241–1249.

Duncan, E. W., N. M. White, and K. Mengersen (2016). Bayesian spatiotemporal modelling for identifying unusual and unstable trends in mammography utilisation. *BMJ open 6*(5), e010253.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Number 38. Oxford University Press.

Eeftens, M., M.-Y. Tsai, C. Ampe, B. Anwander, R. Beelen, T. Bellander, G. Cesaroni, M. Cirach, J. Cyrys, K. d. Hoogh, A. D. Nazelle, F. d. Vocht, C. Declercq,

A. Dėdelė, K. Eriksen, C. Galassi, R. Gražulevičienė, G. Grivas, J. Heinrich, B. Hoffmann, M. Iakovides, A. Ineichen, K. Katsouyanni, M. Korek, U. Krämer, T. Kuhlbusch, T. Lanki, C. Madsen, K. Meliefste, A. Mölter, G. Mosler, M. Nieuwenhuijsen, M. Oldenwening, A. Pennanen, N. Probst-Hensch, U. Quass, O. Raaschou-Nielsen, A. Ranzi, E. Stephanou, D. Sugiri, O. Udvardy, v. Vaskövi, G. Weinmayr, B. Brunekreef, and G. Hoek (2012). Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 - results of the ESCAPE project. *Atmospheric Environment 62*(0), 303 – 317.

Eksler, V. (2008). Exploring spatial structure behind the road mortality of regions in Europe. *Applied Spatial Analysis and Policy 1*(2), 133–150.

Elliott, P., J. Wakefield, N. Best, and D. Briggs (2001). *Spatial Epidemiology*. Oxford University Press.

Elliott, P. and D. Wartenberg (2004). Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect 112*(9), 998–1006.

Fahrmeir, L. and S. Lang (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 50*(2), 201–220.

Farrington, C., N. Andrews, A. Beale, and M. Catchpole (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 159*(3), 547–563.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 209–230.

Fleming, D. M., R. J. Taylor, R. L. Lustig, C. Schuck-Paim, F. Haguinet, D. J. Webb, J. Logie, G. Matias, and S. Taylor (2015). Modelling estimates of the burden of respiratory syncytial virus infection in adults and the elderly in the United Kingdom. *BMC infectious diseases 15*(1), 443.

Fridstrom, L., J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis & Prevention 27*(1), 1–20.

Friedman, L. S., P. Barach, and E. D. Richter (2007). Raised speed limits, case fatality and road deaths: a six year follow-up using ARIMA models. *Injury Prevention 13*(3), 156–161.

Gaudry, G., R. Long, and T. Qian (1993). A martingale proof of L2 boundedness of Clifford-valued singular integrals. *Annali di Matematica Pura ed Applicata 165*(1), 369–394.

Gelfand, A. E., A. Kottas, and S. N. MacEachern (2005). Bayesian nonparametric spatial modeling with dirichlet process mixing. *Journal of the American Statistical Association 100*(471), 1021–1035.

Gelfand, A. E. and A. F. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*(410), 398–409.

Gelfand, A. E. and P. Vounatsou (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics 4*(1), 11–15.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis 1*(3), 515–534.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). Bayesian Data Analysis, Chapman & Hall/CRC.

Gelman, A., J. Hill, and M. Yajima (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness 5*(2), 189–211.

Gelman, A. and D. B. Rubin (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian statistics 4*, 625–631.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (6), 721–741.

Ghosh, M., K. Natarajan, T. W. F. Stroud, and B. P. Carlin (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association 93*(441), 273–282.

Gilks, W., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis.

Goh, B. H. (2005). The dynamic effects of the Asian financial crisis on construction demand and tender price levels in Singapore. *Building and Environment 40*(2), 267–276.

Gómez-Rubio, V. and H. Rue (2017). Markov chain Monte Carlo with the integrated nested Laplace approximation. *arXiv preprint arXiv:1701.07844*.

Graham, D. J. and S. Glaister (2003). Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix. *Urban Studies 40*(8), 1591–1607.

Hacking, J. M., S. Muller, I. E. Buchan, et al. (2011). Trends in mortality from 1965 to 2008 across the English north-south divide: comparative observational study. *BMJ 342*.

Hansell, A., J. Hollowell, R. McNiece, T. Nichols, and D. Strachan (2003). Validity and interpretation of mortality, health service and survey data on COPD and asthma in England. *European respiratory journal 21*(2), 279–286.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Held, L., I. NatÃąrio, S. E. Fenton, H. Rue, and N. Becker (2005). Towards joint disease mapping. *Statistical Methods in Medical Research 14*(1), 61–82. PMID: 15691000.

Holt, J. B., X. Zhang, L. Presley-Cantrell, and J. B. Croft (2011). Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among medicare beneficiaries in the United States. *International journal of chronic obstructive pulmonary disease 6*, 321.

Houston, D. J. and L. E. Richardson (2002). Traffic safety and the switch to a primary seat belt law: the california experience. *Accident Analysis & Prevention 34*(6), 743–751.

HSCIC (2015). Statistics on Smoking: England, 2011. The Health and Social Care Information Centre. `http://www.hscic.gov.uk/catalogue/PUB00693/smok-eng-2011-rep.pdf`. Accessed: 2015-10-05.

Hughes, G. and R. Gorton (2013). An evaluation of SaTScan for the prospective detection of space-time campylobacter clusters in the north east of England. *Epidemiology & Infection 141*(11), 2354–2364.

Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of Quality Technology 18*(4), 203–210.

Jin, X., B. P. Carlin, and S. Banerjee (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics 61*(4), 950–961.

Johnson, G. (2008). Prospective spatial prediction of infectious disease: Experience of New York State (USA) with west nile virus and proposed directions for improved surveillance. *Environmental and Ecological Statistics 15*(3), 293–311.

Jones, A. P., R. Haynes, V. Kennedy, I. M. Harvey, T. Jewell, and D. Lea (2008). Geographical variations in mortality and morbidity from road traffic accidents in England and Wales. *Health & Place 14*(3), 519–535.

Joo, M. J., T. A. Lee, and K. B. Weiss (2007). Geographic variation in chronic obstructive pulmonary disease exacerbation rates. *Journal of general internal medicine 22*(11), 1560–1565.

Jovanis, P. P. and H.-L. Chang (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record: Journal of the Transportation Research Board 1068*, 42–51.

Kass, R. E. and R. Natarajan (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Analysis 3*, 535–542.

Kelsall, J. and J. Wakefield (1999). Discussion of 'Bayesian models for spatially correlated disease and exposure data', by Best et al. *Bayesian statistics 6*, 151.

Kim, K. and E. Yamashita (2002). Motor vehicle crashes and land use: Empirical analysis from Hawaii. *Transportation Research Record: Journal of the Transportation Research Board 1784*(-1), 73–79.

Kleinman, K. (2005). Generalized linear models and generalised linear mixed models for small-area surveillance. In *Spatial and Syndromic Surveillance for Public Health*, pp. 77–93. West Sussex: John Wiley.

Kleinman, K., R. Lazarus, and R. Platt (2004). A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology 159*(3), 217–224.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine 19*(17-18), 2555–2567.

Knorr-Held, L. and J. Besag (1998). Modelling risk from a disease in time and space. *Statistics in Medicine 17*(18), 2045–2060.

Knorr-Held, L. and N. G. Best (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 164*(1), 73–85.

Knorr-Held, L. and S. Richardson (2003). A hierarchical model for space–time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 52*(2), 169–183.

Knox, E. and M. Bartlett (1964). The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 13*(1), 25–30.

Kottas, A., J. A. Duan, and A. E. Gelfand (2008). Modeling disease incidence data with spatial and spatio temporal dirichlet process mixtures. *Biometrical Journal 50*(1), 29–42.

Krishnaveni, S. and M. Hemalatha (2011). A perspective analysis of traffic accident using data mining techniques. *International Journal of Computer Applications 23*(7), 40–48.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and methods 26*(6), 1481–1496.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 164*(1), 61–72.

Kulldorff, M., R. Heffernan, J. Hartman, R. Assunção, and F. Mostashari (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS medicine 2*(3), e59.

Kulldorff, M. and N. Nagarwalla (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine 14*(8), 799–810.

Laurent, O., G. Pedrono, L. Filleul, C. Segala, A. Lefranc, C. Schillinger, E. Rivière, and D. Bard (2009). Influence of socioeconomic deprivation on the relation between air pollution and β-agonist sales for asthma. *CHEST Journal 135*(3), 717–723.

Lawson, A. B. (2013). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lawson, A. B. and K. Kleinman (2005). *Spatial and syndromic surveillance for public health*. John Wiley & Sons.

Lawson, A. B. and H.-R. Song (2010). Bayesian hierarchical modeling of the dynamics of spatio-temporal influenza season outbreaks. *Spatial and spatio-temporal epidemiology 1*(2), 187–195.

Leroux, B. G., X. Lei, N. Breslow, M. Halloran, and B. D. Elizabeth (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical models in epidemiology, the environment, and clinical trials*, 179–191.

Levin-Rector, A., E. L. Wilson, A. D. Fine, and S. K. Greene (2015). Refining historical limits method to improve disease cluster detection, New York City, New York, USA. *Emerging infectious diseases 21*(2), 265.

Li, G., N. Best, A. L. Hansell, I. Ahmed, and S. Richardson (2012). BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics 13*(4), 695–710.

Li, G., R. Haining, S. Richardson, and N. Best (2014). Space–time variability in burglary risk: a Bayesian spatio-temporal modelling approach. *Spatial Statistics 9*, 180–191.

Li, G., S. Richardson, L. Fortunato, I. Ahmed, A. Hansell, M. Toledano, and N. Best (2011). Data mining cancer registries: retrospective surveillance of small area time trends in cancer incidence using BaySTDetect. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pp. 885–890. IEEE.

Lindebeg, A., A. Bjerg-Backlund, E. Ranmark, et al. (2006). Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking: report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med 100*, 264–72.

Lindgren, F. and H. Rue (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software 63*(19).

Lopez, A., K. Shibuya, C. Rao, C. Mathers, A. Hansell, L. Held, V. Schmid, and S. Buist (2006). Chronic obstructive pulmonary disease: current burden and future projections. *European Respiratory Journal 27*(2), 397–412.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter (2012). *The BUGS book: A practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC.

Ma, J. and K. Kockelman (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record: Journal of the Transportation Research Board 1950*(1), 24–34.

Ma, J., K. M. Kockelman, and P. Damien (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention 40*(3), 964–975.

MacNab, Y. and C. Dean (2000). Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. *Statistics in Medicine 19*(17-18), 2421–2435.

MacNab, Y. C. (2004). Bayesian spatial and ecological models for small-area accident and injury analysis. *Accident Analysis & Prevention 36*(6), 1019–1028.

MacNab, Y. C. (2011). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research 20*(1), 49–68.

MacNab, Y. C. and C. B. Dean (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics 57*(3), 949–956.

MacNab, Y. C. and P. Gustafson (2007). Regression b-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. *Statistics in medicine 26*(24), 4455–4474.

Maiti, T. (1998). Hierarchical Bayes estimation of mortality rates for disease mapping. *Journal of Statistical Planning and Inference 69*(2), 339–348.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research 27*(2 Part 1), 209–220.

Mardia, K. V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis 24*(2), 265–284.

Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika 100*(3), 539–553.

Masoli, M., D. Fabian, S. Holt, and R. Beasley (2004). The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy 59*(5), 469–478.

McKenzie, E. (1985). Some simple models for discrete variable time series. *Water Resources Bulletin 21*, 645–650.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics 21*(6), 1087–1092.

Miaou, S. P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus Negative Binomial regressions. *Accident Analysis & Prevention 26*(4), 471–482.

Miaou, S. P. and H. Lum (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention 25*(6), 689 – 709.

Miaou, S. P. and J. J. Song (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention 37*(4), 699–720.

Miaou, S. P., J. J. Song, and B. K. Mallick (2003). Roadway traffic crash mapping: A space-time modeling approach. *Journal of Transportation and Statistics 6*(1).

Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe (2005). Alternative risk models for ranking locations for safety improvement. *Transportation Research Record: Journal of the Transportation Research Board 1908*(1), 1–8.

Muller, P., G. Parmigiani, and K. Rice (2006). FDR and Bayesian multiple comparisons rules. In *Proceedings Valencia*.

Nandram, B., J. Sedransk, and L. W. Pickle (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association 95*(452), 1110–1118.

Napier, G., D. Lee, C. Rob ertson, A. Lawson, and K. G. Pollock (2016). A model to estimate the impact of changes in MMR vaccine uptake on inequalities in measles susceptibility in Scotland. *Statistical methods in medical research 25*(4), 1185–1200.

Naus, J. I. (1965a). Clustering of random points in two dimensions. *Biometrika 52*, 263–267.

Naus, J. I. (1965b). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association 60*(310), 532–538.

Newton, M. A., A. Noueiry, D. Sarkar, and P. Ahlquist (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics 5*(2), 155–176.

NICE (2015). National Institute for Health and Clinical Excellence, Kings Fund, Paul Jerram (IOW PCT). `https://www.nice.org.uk/sharedlearning/isle-of-wight-respiratory-inhaler-project`. Accessed: 2015-10-05.

Noland, R. B. and M. A. Quddus (2004). A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention 36*(6), 973–984.

Olsen, S. F., M. Martuzzi, and P. Elliott (1996). Cluster analysis and disease mapping–why, when, and how? a step by step guide. *BMJ: British Medical Journal 313*(7061), 863.

Page, E. (1954). Continuous inspection schemes. *Biometrica*, 100–115.

Park, E. S. and D. Lord (2007). Multivariate Poisson lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record: Journal of the Transportation Research Board 2019*(1), 1–6.

Prescott, E. and J. Vestbo (1999). Socioeconomic status and chronic obstructive pulmonary disease. *Thorax 54*(8), 737–741.

Qin, X., J. N. Ivan, and N. Ravishanker (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention 36*(2), 183–191.

Quddus, M. A. (2008a). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention 40*(4), 1486–1497.

Quddus, M. A. (2008b). Time series count data models: an empirical application to traffic accidents. *Accident Analysis & Prevention 40*(5), 1732–1741.

Richardson, S., J. J. Abellan, and N. Best (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire(UK). *Statistical methods in Medical Research 15*(4), 385–407.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*(4), 731–792.

Robertson, C., T. A. Nelson, Y. C. MacNab, and A. B. Lawson (2010). Review of methods for space–time disease surveillance. *Spatial and spatio-temporal epidemiology 1*(2), 105–116.

Rodrigues, E. C. and R. Assunção (2012). Bayesian spatial models with a mixture neighborhood structure. *Journal of Multivariate Analysis 109*, 88–102.

Rogerson, P. A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine 16*(18), 2081–2093.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology) 71*(2), 319–392.

Rushworth, A., D. Lee, and R. Mitchell (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and spatio-temporal epidemiology 10*, 29–38.

Schlüter, P., J. Deely, and A. Nicholson (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.

Scott, J. G. and J. O. Berger (2006). An exploration of aspects of Bayesian multiple testing. *Journal of statistical planning and inference 136*(7), 2144–2162.

Scott E, P. (2016a). More than 400 pedestrians hurt in road accidents in surrey last year. `http://www.getsurrey.co.uk/news/surrey-news/more-400-pedestrians-hurt-road-11706152`. Accessed: 2017-09-11.

Scott E, P. (2016b). Worrying increase in lorry and van crashes according to new figures. `http://www.getsurrey.co.uk/news/surrey-news/worrying-increase-lorry-van-crashes-11819808`. Accessed: 2017-09-11.

Shankar, V., F. Mannering, and W. Barfield (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention 27*(3), 371–389.

Sherman, R. L., K. A. Henry, S. L. Tannenbaum, D. J. Feaster, E. Kobetz, and D. J. Lee (2014). Peer reviewed: Applying spatial analysis tools in public health: An example using SaTScan to detect geographic targets for colorectal cancer screening interventions. *Preventing chronic disease 11*.

Shewhart, W. (1931). *Economic Control of Quality of Manufactured Product*. Bell Telephone Laboratories series. American Society for Quality Control.

Sin, D. D., N. R. Anthonisen, J. B. Soriano, and A. Agusti (2006). Mortality in COPD: role of comorbidities. *European Respiratory Journal 28*(6), 1245–1257.

Snow, J. (1855). *On the mode of communication of cholera*. John Churchill.

Sofianopoulou, E., S. P. Rushton, P. J. Diggle, and T. Pless-Mulloli (2013). Association between respiratory prescribing, air pollution and deprivation, in primary health care. *Journal of Public Health 35*(4), 502–509.

Song, J. J., M. Ghosh, S. Miaou, and B. Mallick (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis 97*(1), 246 – 273.

Southampton City Council (2016). Cycling Southampton: A plan for our city 2016-2026. `https://www.southampton.gov.uk/Images/Cycling-Southampton-2016-2026-DRAFT-Nov%202016_tcm63-389640.pdf`. Accessed: 2017-09-11.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical report, Research Report, 98-009.

Stern, H. S. and N. Cressie (2000). Posterior predictive model checks for disease mapping models. *Statistics in medicine 19*(17-18), 2377–2397.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics 31*(6), 2013–2035.

Stroup, D. F., G. D. Williamson, J. L. Herndon, and J. M. Karon (1989). Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine 8*(3), 323–329.

Sun, D., R. K. Tsutakawa, H. Kim, Z. He, et al. (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine 19*(15), 2015–2035.

Surrey Police (2015). The increase in fatal road traffic collissions. `http://www.surrey-pcc.gov.uk/wp-content/uploads/2015/01/03_The-Increase-in-Fatal-RTCs-final.pdf`. Accessed: 2017-09-05.

Tegnér, G. and V. Loncar-Lucassi (1997). Demand for road use, accidents and their gravity in Stockholm: Measurement and analysis of the dennis package. *Transek AB, Stockholm*.

Thacker, S. B. and D. F. Stroup (1994). Future directions for comprehensive public health surveillance and health information systems in the United States. *American Journal of Epidemiology 140*(5), 383–397.

Thomas, A., N. Best, D. Lunn, R. Arnold, and D. Spiegelhalter (2004). GeoBugs user manual. *Cambridge: Medical Research Council Biostatistics Unit*.

Torre, G. L., E. V. Beeck, G. Quaranta, A. Mannocci, and W. Ricciardi (2007). Determinants of within-country variation in traffic accident mortality in Italy: a geographical analysis. *International Journal of Health Geographics 6*(1), 49.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of statistics 31*(3), 221–229.

Tzala, E. and N. Best (2008). Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. *Statistical methods in medical research 17*(1), 97–118.

Van den Bossche, F., G. Wets, and T. Brijs (2006). Predicting road crashes using calendar data. In *Transportation Research Board 85th Annual Meeting*, Number 06-2016.

Vegni, F. E., B. Castelli, F. Auxilia, and P. Wilkinson (2005). Air pollution and respiratory drug use in the city of Como, Italy. *European journal of epidemiology 20*(4), 351–358.

Ventrucci, M., E. M. Scott, and D. Cocchi (2010). Multiple testing on standardized mortality ratios: a Bayesian hierarchical model for FDR estimation. *Biostatistics 12*(1), 51–67.

Wakefield, J., N. Best, and L. Waller (2000). Bayesian approaches to disease mapping. In *Spatial Epidemiology*, pp. 104–27. Oxford University Press.

Walker, P., P. Mitchell, F. Diamantea, C. Warburton, and L. Davies (2006). Effect of primary-care spirometry on the diagnosis and management of COPD. *European Respiratory Journal 28*(5), 945–952.

Waller, L. A., B. P. Carlin, H. Xia, and A. E. Gelfand (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association 92*(438), 607–617.

Wang, C., M. A. Quddus, and S. G. Ison (2011). Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Analysis & Prevention 43*(6), 1979–1990.

Wang, Y. and K. M. Kockelman (2013). A Poisson-lognormal conditional auto-regressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention 60*, 71–84.

Weiss, K. B. and D. K. Wagener (1990). Geographic variations in US asthma mortality: small-area analyses of excess mortality, 1981–1985. *American Journal of Epidemiology 132*(supp1), 107–115.

Wells, C. and E. Gordon (2008). Geographical variations in premature mortality in England and Wales, 1981–2006. *Health statistics quarterly 38*, 6–18.

Whittemore, A. S. (2007). A Bayesian false discovery rate for multiple testing. *Journal of Applied Statistics 34*(1), 1–9.

Wilkinson, T., M. North, and S. C. Bourne (2014). Reducing hospital admissions and improving the diagnosis of COPD in Southampton city: methods and results of a 12-month service improvement project. *NPJ primary care respiratory medicine 24*.

Wolfe, A. C. (1982). The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis & Prevention 14*, 337–340.

Woodall, W. H. and M. M. Ncube (1985). Multivariate CUSUM quality-control procedures. *27*(3), 285–292.

World Health Organization (2009a). World health statistics 2009.

World Health Organization (2009b). Global status report on road safety: time for action.

World Health Organization (2012). Prevention and control of noncommunicable diseases: Guidelines for primary health care in low resource settings.

World Health Organization (2014). Global status report on alcohol and health.

Yiannakoulias, N., L. W. Svenson, and D. P. Schopflocher (2009). An integrated framework for the geographic surveillance of chronic disease. *International Journal of Health Geographics 8*(1), 69.

You, J., J. Wang, and J. Guo (2017). Real-time crash prediction on freeways using data mining and emerging techniques. *Journal of Modern Transportation*, 1–8.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika 75*, 621–629.

Zhou, H. and A. B. Lawson (2008). EWMA smoothing and Bayesian spatial modeling for health surveillance. *Statistics in medicine 27*(28), 5907–5928.

Zou, J., A. F. Karr, D. Banks, M. J. Heaton, G. Datta, J. Lynch, and F. Vera (2012). Bayesian methodology for the analysis of spatial-temporal surveillance data. *Statistical Analysis and Data Mining 5*(3), 194–204.

Appendices

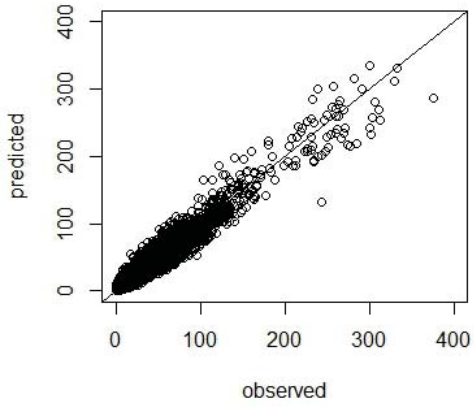
A Supplementary material for Chapter 2

Observed versus predicted data

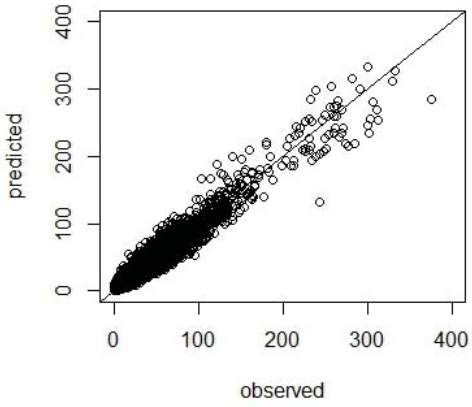
Figures A.1 and A.2 illustrate the observed versus predicted data summed across years for low and high severity under all models. In general, for low severity, there are no systematic differences between observed and predicted data (posterior means) and the results for the different models look very similar to each other. On the other hand, for high severity, the variability is clearly larger, with Models C and D (*Common* space effects) shown to be worse, a fact which might be well explained by the smaller numbers of counts for high severity compared to low severity.

Maps of p_{Bayes} values

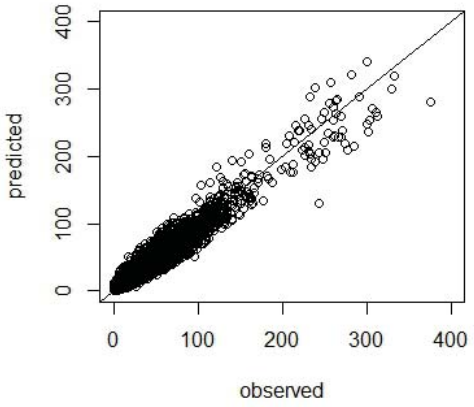
To assess the adequacy of Model E, which is favoured by the DIC, maps of p_{Bayes} values across years and areas are presented. The maps in Figures A.3 and A.4 show in orange the areas in which the predicted number of accident counts was substantially lower than the number observed ($p_{Bayes} \leq 0.1$) while in purple the opposite is true ($p_{Bayes} \geq 0.9$).



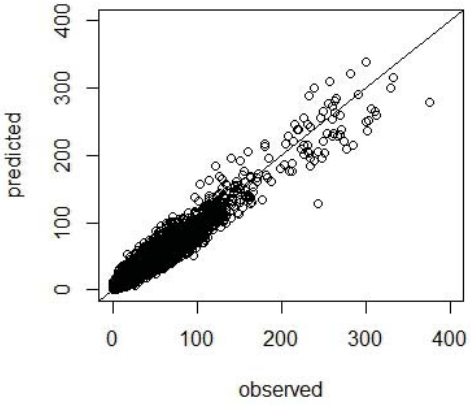
(a) Model A



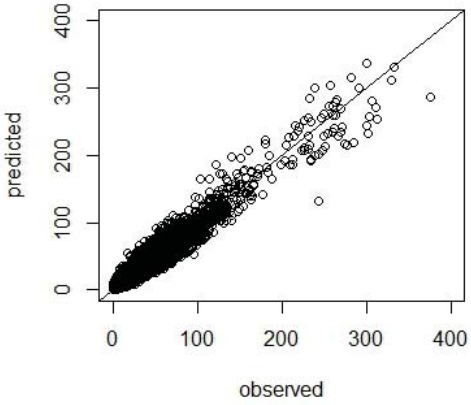
(b) Model B



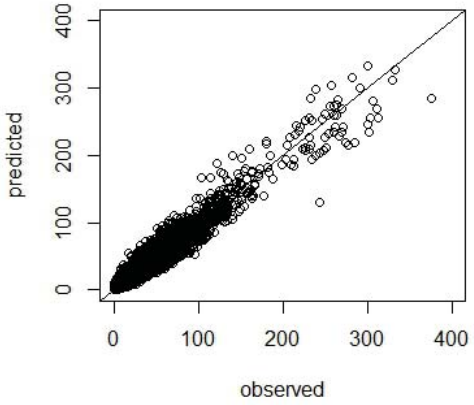
(c) Model C



(d) Model D

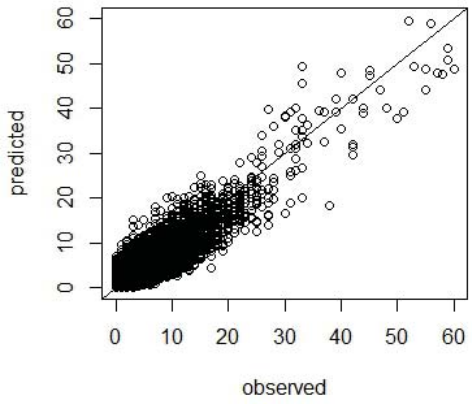


(e) Model E

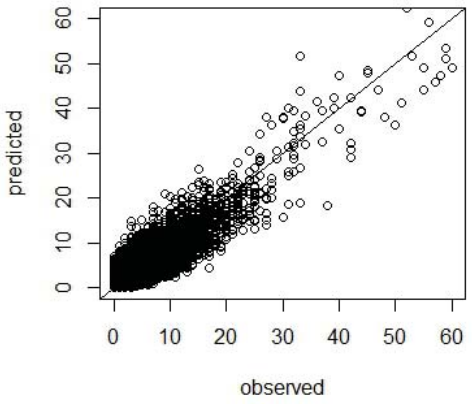


(f) Model F

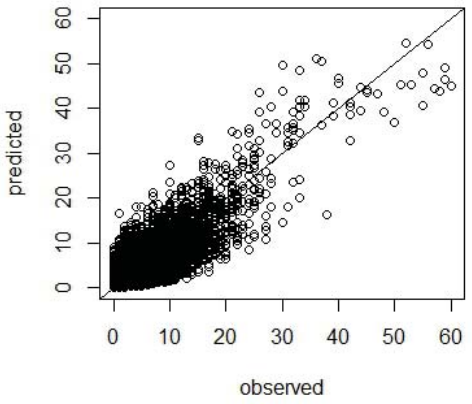
Figure A.1: Plots showing observed versus predicted (posterior means) data for low severity across all years under each of the models considered in the study.



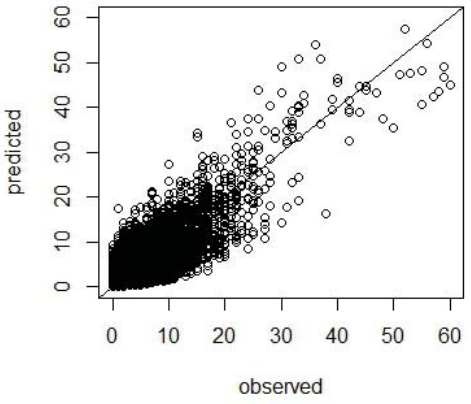
(a) Model A



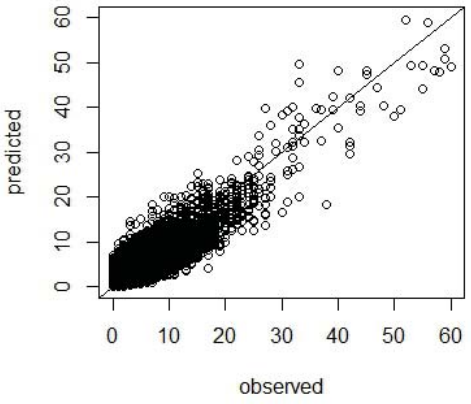
(b) Model B



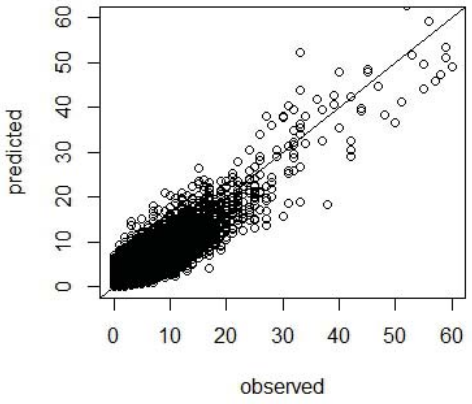
(c) Model C



(d) Model D

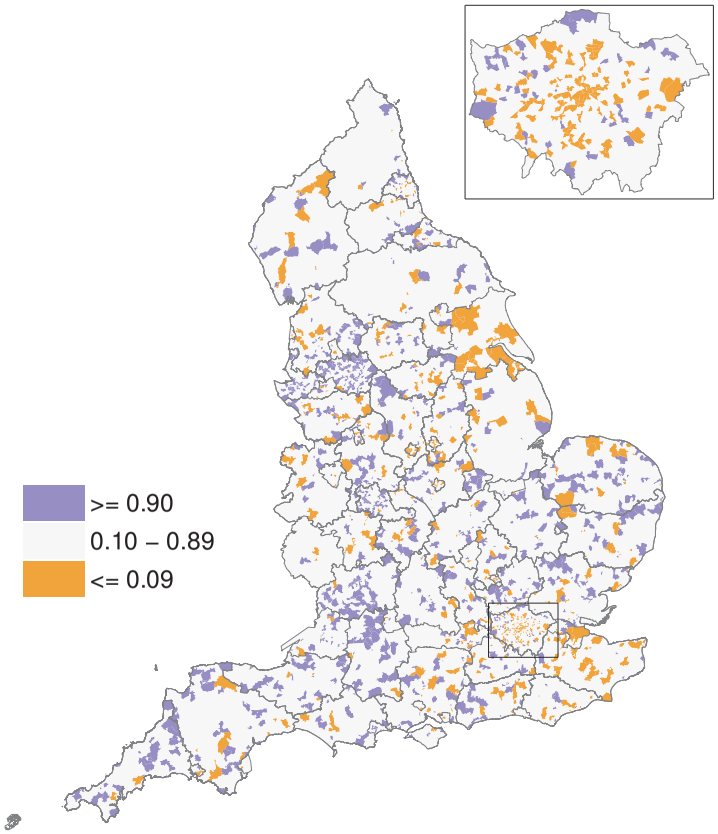


(e) Model E

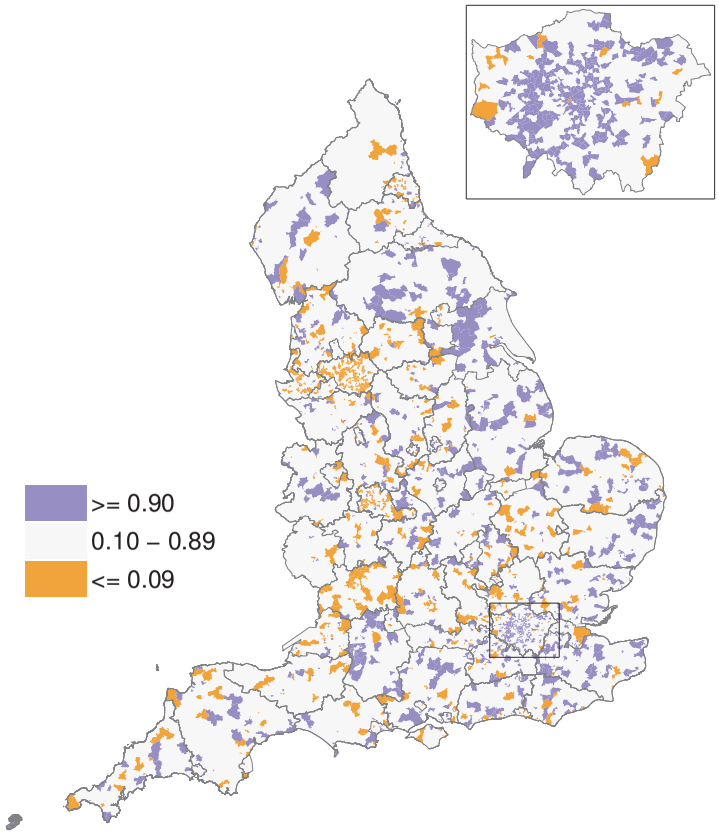


(f) Model F

Figure A.2: Plots showing observed versus predicted (posterior means) data for high severity across all years under each of the models considered in the study.

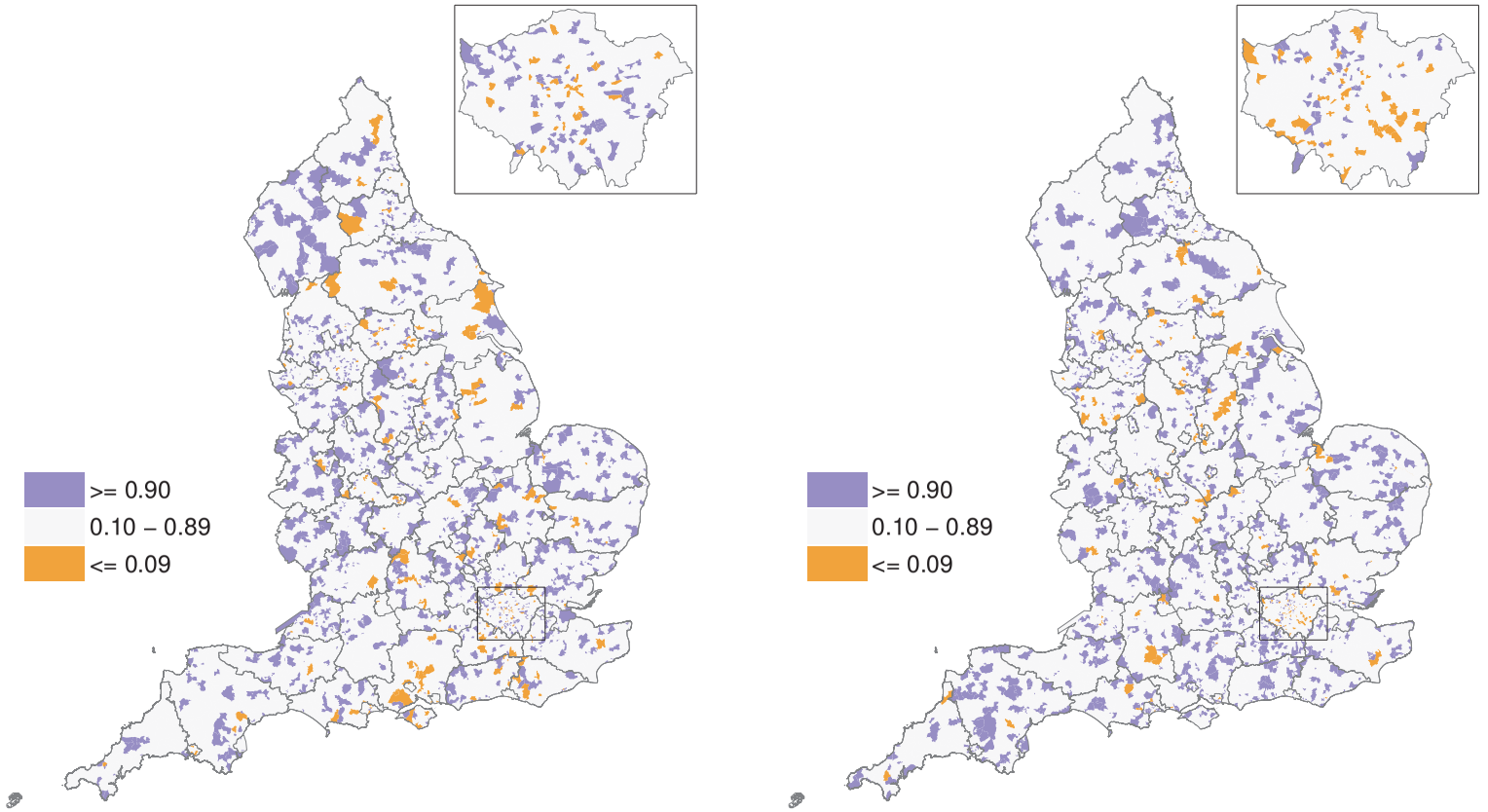


(a) Year 2005



(b) Year 2013

Figure A.3: Maps showing areas with extreme p_{Bayes} values under Model E for low severity for years 2005 and 2013; the insets are enlarged views of London.



(a) Year 2005

(b) Year 2013

Figure A.4: Maps showing areas with extreme p_{Bayes} values under Model E for high severity for years 2005 and 2013; the insets are enlarged views of London.

Sensitivity Analysis

The results of the sensitivity analysis that was conducted for Model E are presented in Figures A.5 to A.8. Three alternative Wishart priors are considered for the multivariate normal and multivariate CAR distributions, with diagonals 1, 500, and 1000. The off-diagonals are set to 0.0005. Figures A.5 and A.6 are for the spatial component of the estimated posterior rates $(\theta_i + \phi_i)$ for a sample of 800 areas. Figure A.7 shows the corresponding comparison between priors 1000 and 500 for all areas, and Figure A.8 shows the temporal component of the estimated posterior rates $\exp(\delta_t)$. As can be seen from the plots the results mostly overlap therefore the model is robust to the prior specification.

Road network

The map of the posterior rates for high severity is linked with the shapefile of the motorways in order to explore relationships and help interpret the results. As can be seen in Figure A.9, motorways in England appear to have lower rates compared to the surrounding areas given the traffic volume, making them safer than the rest of the roads. This suggests that information on motorway may be considered as an explanatory variable in future analyses.

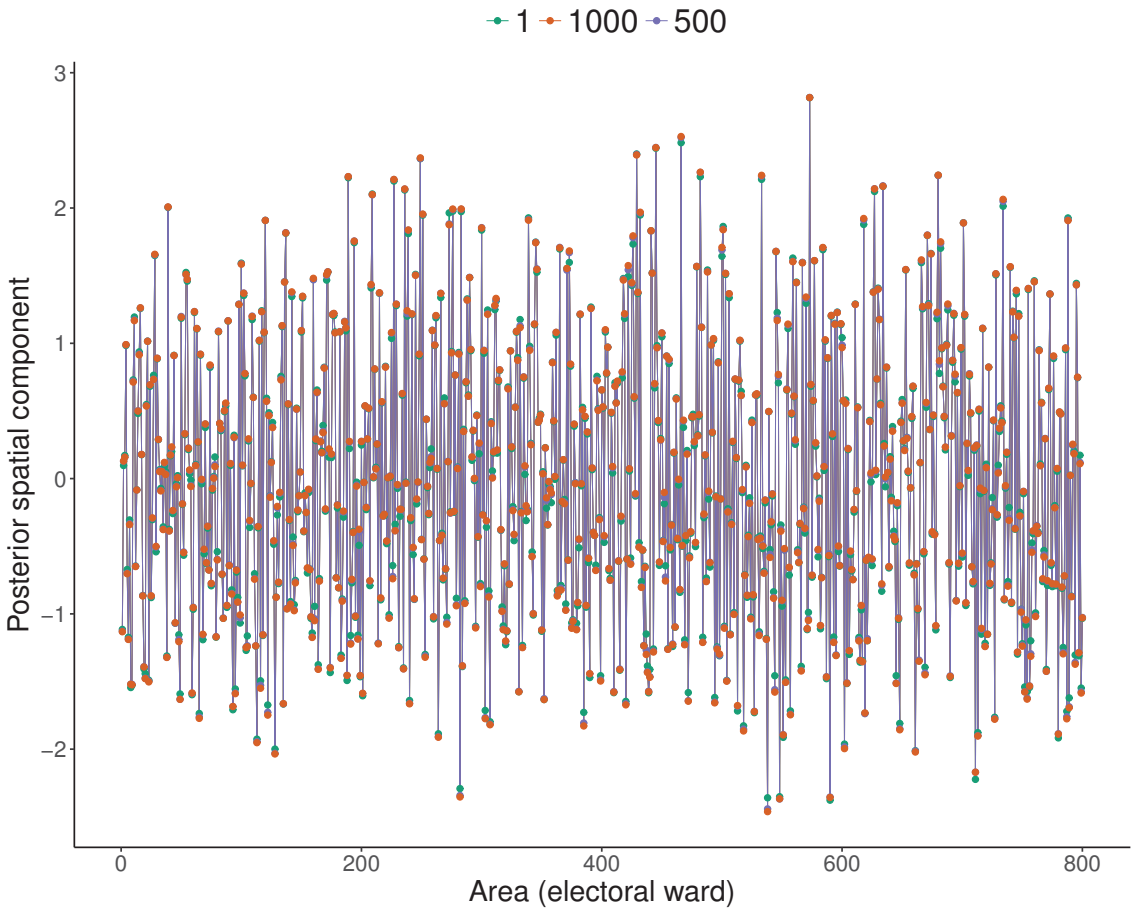


Figure A.5: Posterior spatial effects $\exp(\theta_i + \phi_i)$ under three alternative Wishart priors for a sample of 800 areas for low severity.

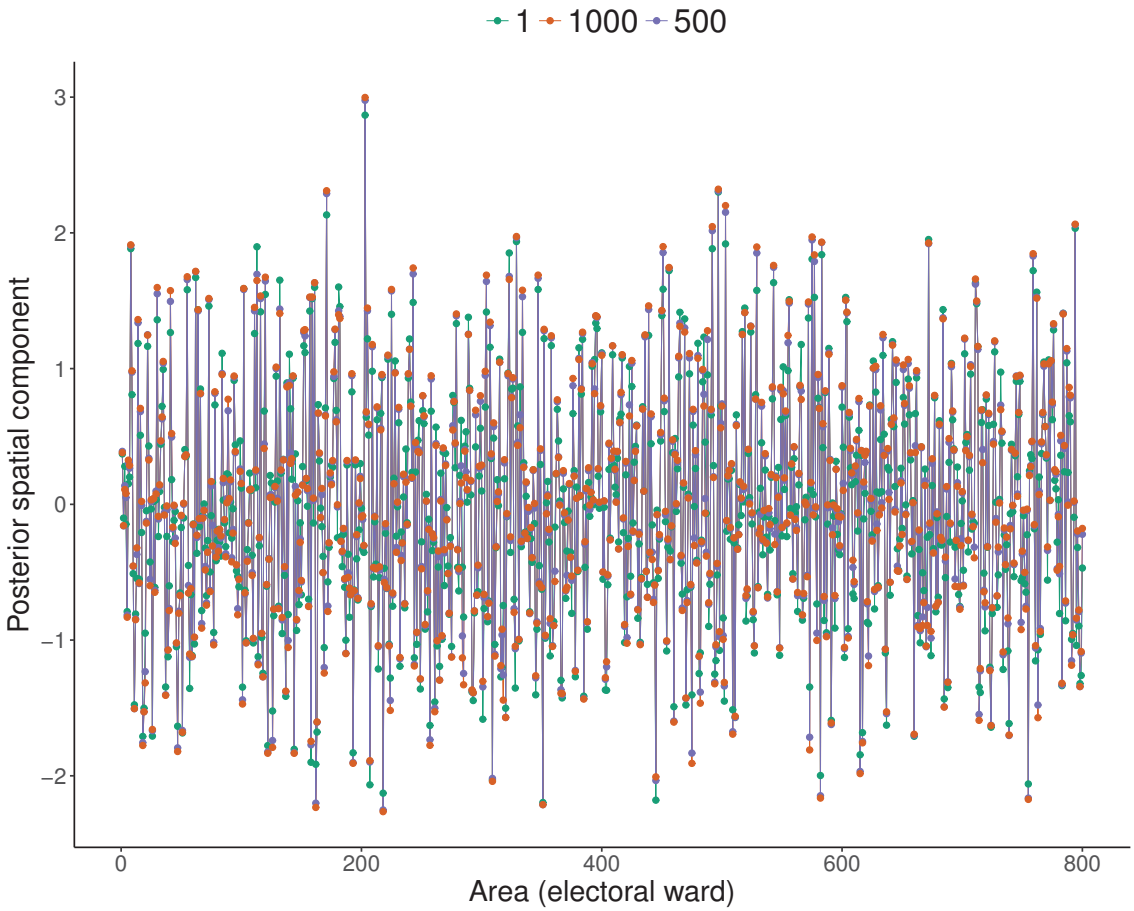
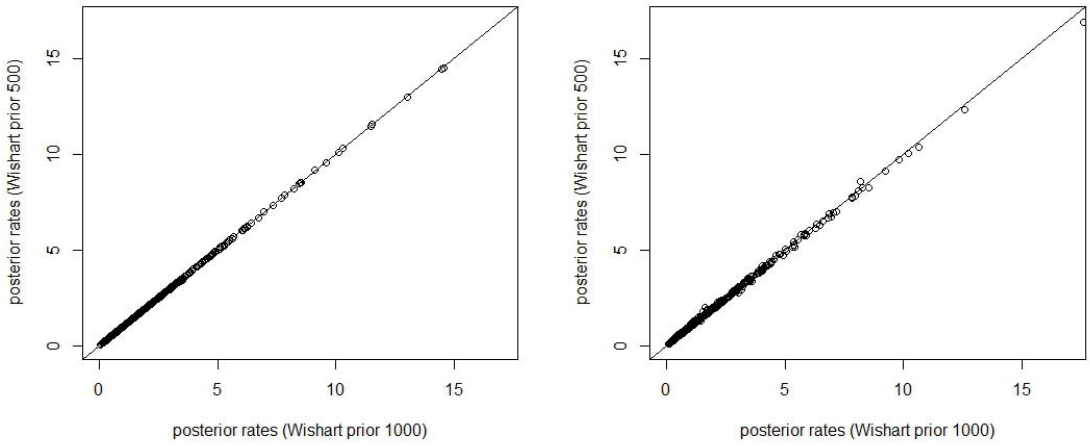


Figure A.6: Posterior spatial effects $\exp(\theta_i + \phi_i)$ under three alternative Wishart priors for a sample of 800 areas for high severity.



(a) low severity (b) high severity

Figure A.7: Posterior accident rates under two alternative Wishart priors for low and high severity level.

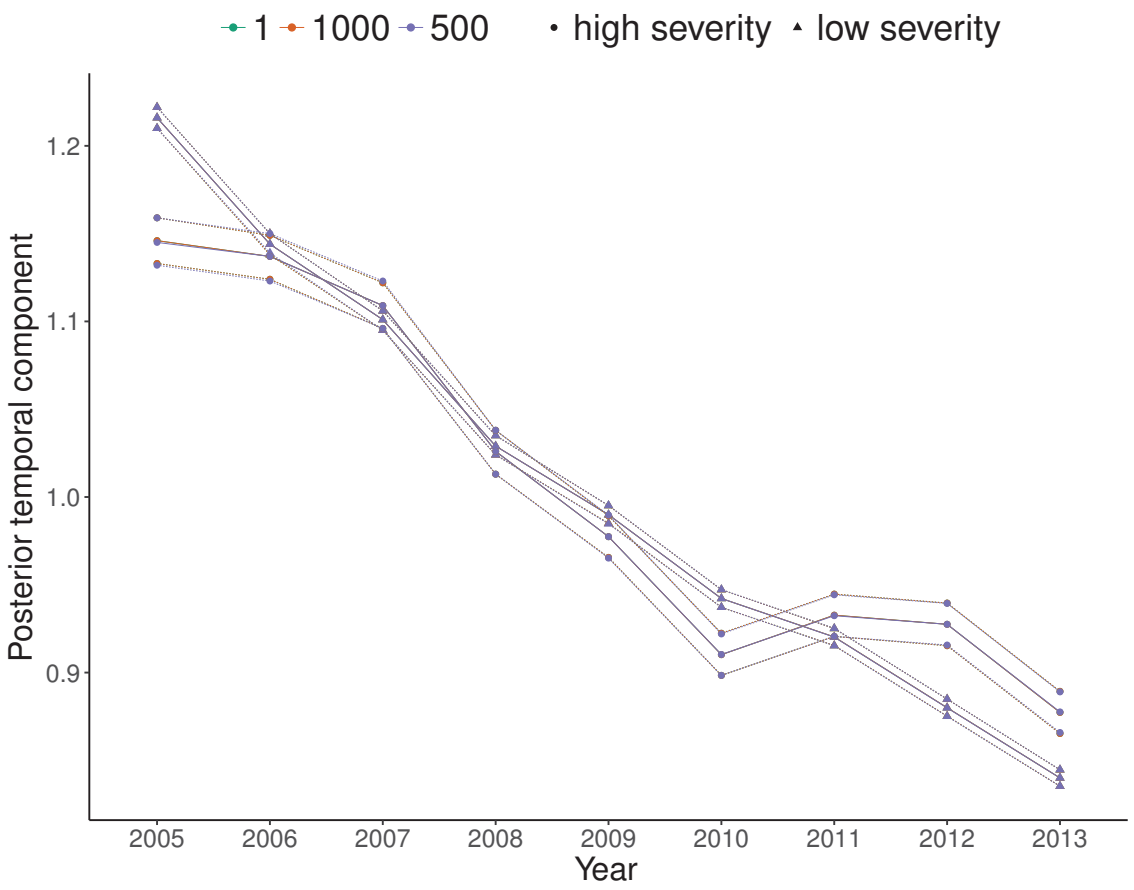


Figure A.8: Posterior temporal effects $\exp(\delta_t)$ under three alternative priors for low and high severity; the solid lines show the means and the dashed lines represent the corresponding 95% CI values.

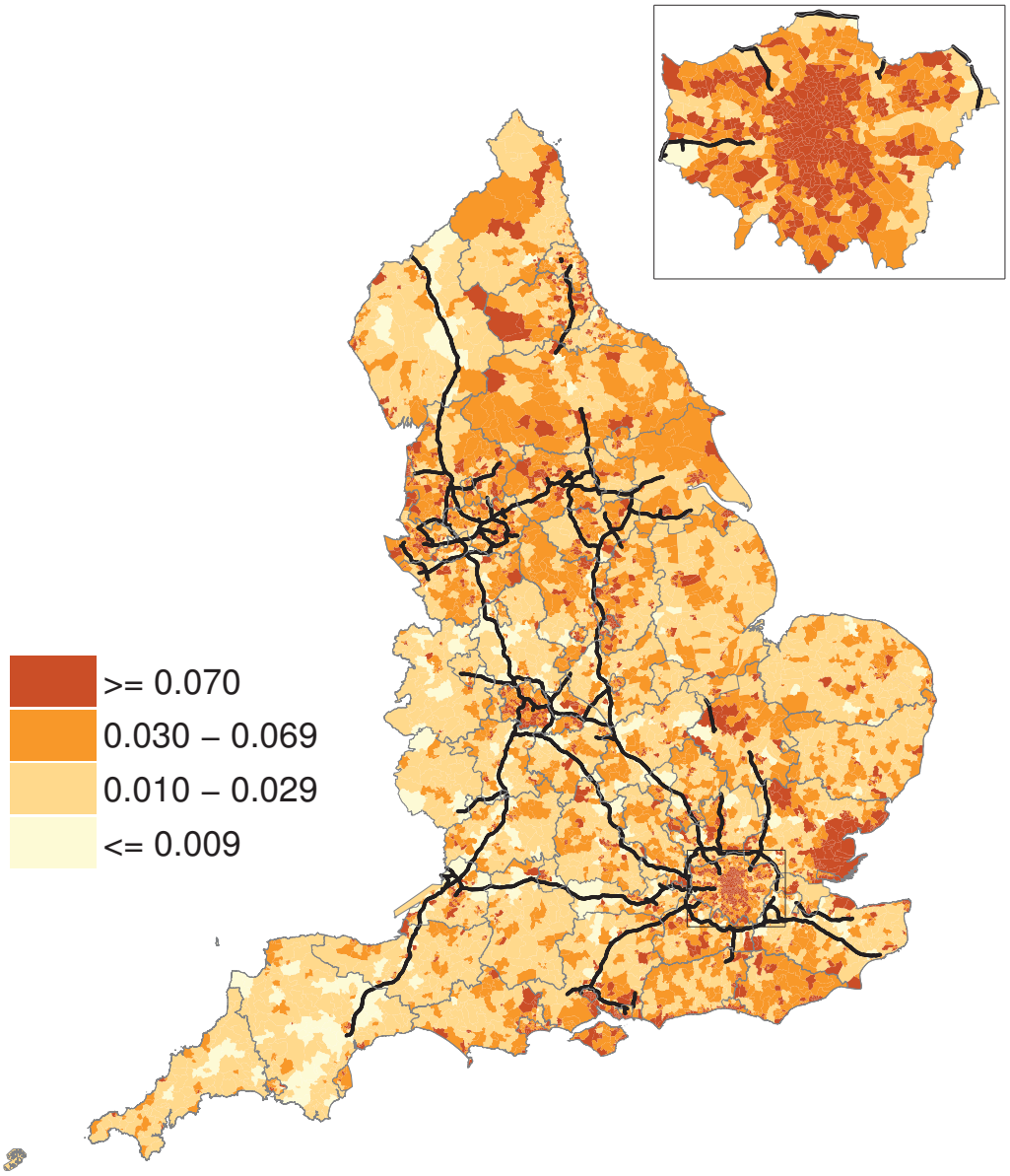


Figure A.9: Map of posterior accident rates for high severity linked with the shapefile of the motorways; the insets are enlarged views of London.

BUGS Model

```
Model {
for (i in 1:Nareas) {
  for (j in 1:Nlevels) {
    for (t in 1:T) {
      Y[i, j, t] ~ dpois(mu[i, j, t])
      log(mu[i, j, t]) <- alpha[j] + log(Off[i]) + U[j, i] + V[i, j] + W[j, t]
      # Posterior rates
      resRR[i, j, t] <- exp(U[j, i] + V[i, j] + W[j, t])
      # Predictions
      Y.rep[i, j, t] ~ dpois(mu[i, j, t])
      P.rep[i, j, t] <- 1-step(Y[i, j, t] - Y.rep[i, j, t])
} } }

####    MVNormal prior for the spatially unstructured random effects
for (i in 1:Nareas) { V[i, 1:Nlevels] ~ dmnorm(zero[], omikron[, ]) }

####    MVCAR prior for the spatially structured random effects
U[1:Nlevels, 1:Nareas] ~ mv.car(adj[], weights[], num[], omega[, ])
for (i in 1:sumNumNeigh) { weights[i] <- 1 }

####    RW1 prior for the temporal effects
for (j in 1:Nlevels) {
  W[j, 1:T] ~ car.normal(adj2[], weights2[], num2[], tau.w[j])
}
# Posterior temporal effects
for (j in 1:Nlevels) {
  for (t in 1:T) {
    resRR.W[j, t] <- exp(W[j, t])
} }

#  Weights and adjacency matrix for the RW1 prior
for (t in 1:1) {
  weights2[t] <- 1; adj2[t] <- t+1; num2[t] <- 1
}
for (t in 2:(T-1)) {
  weights2[2+(t-2)*2] <- 1; adj2[2+(t-2)*2] <- t-1
  weights2[3+(t-2)*2] <- 1; adj2[3+(t-2)*2] <- t+1; num2[t] <- 2
}
for (t in T:T) {
```

```
  weights2[(T-2)*2 + 2] <- 1; adj2[(T-2)*2 + 2] <- t-1; num2[t] <- 1
}

####    Hyperpriors
for (j in 1:Nlevels){
  tau.w[j] ~ dgamma(0.5, 0.0005)
  sigma.w[j] <- 1/tau.w[j]
  alpha[j] ~ dflat()
  mean.RR[j] <- exp(alpha[j])
}

# Precision matrix of MVCAR
omega[1:Nlevels, 1:Nlevels] ~ dwish(R[, ], Nlevels)
# Covariance matrix of MVCAR
sigmaB[1:Nlevels, 1:Nlevels] <- inverse(omega[, ])
sigma[1] <- sigmaB[1, 1]
sigma[2] <- sigmaB[2, 2]
# Within-area conditional correlation between severities
corr <- sigmaB[1, 2] / (sqrt(sigma[1]) * sqrt(sigma[2]))

# Precision matrix of MVNormal
omikron[1:Nlevels, 1:Nlevels] ~ dwish(Q[, ], Nlevels)
# Covariance matrix of MVNormal
sigmaB.V[1:2, 1:2] <- inverse(omikron[, ])
sigma.V[1] <- sigmaB.V[1, 1]
sigma.V[2] <- sigmaB.V[2, 2]
# Within-area correlation between severities
corr.V <- sigmaB.V[1, 2] / (sqrt(sigma.V[1]) * sqrt(sigma.V[2]))

# Within-area conditional correlation between severities for total spatial effects
corr.sum <- (sigmaB[1, 2] + sigmaB.V[1, 2]) /
(sqrt(sigmaB[1, 1] + sigmaB.V[1, 1]) * sqrt(sigmaB[2, 2] + sigmaB.V[2, 2]))
} # end of model
```

B Supplementary material for Chapter 3

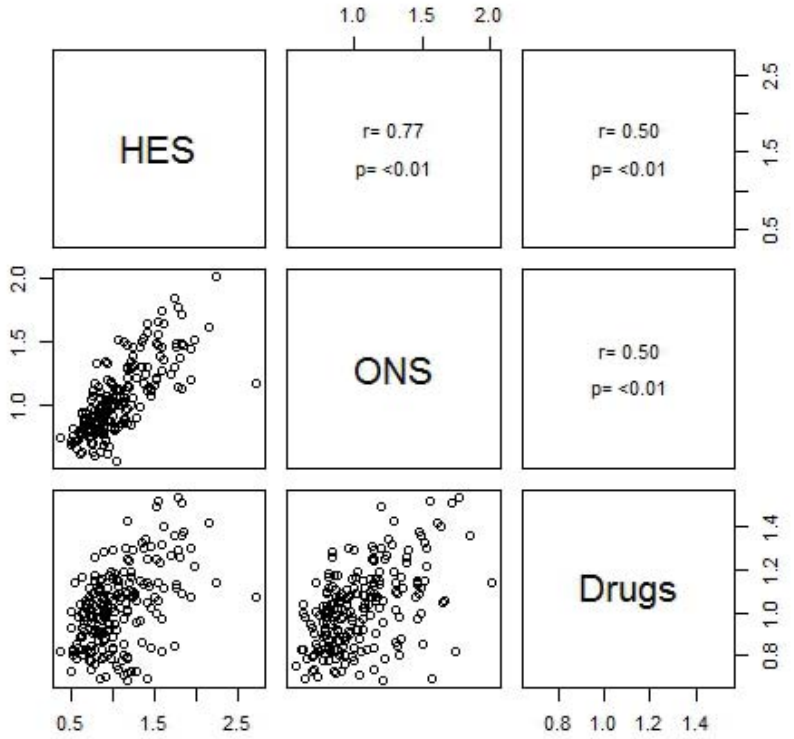


Figure B.1: Correlation among admission, mortality, and prescription data on CRD.

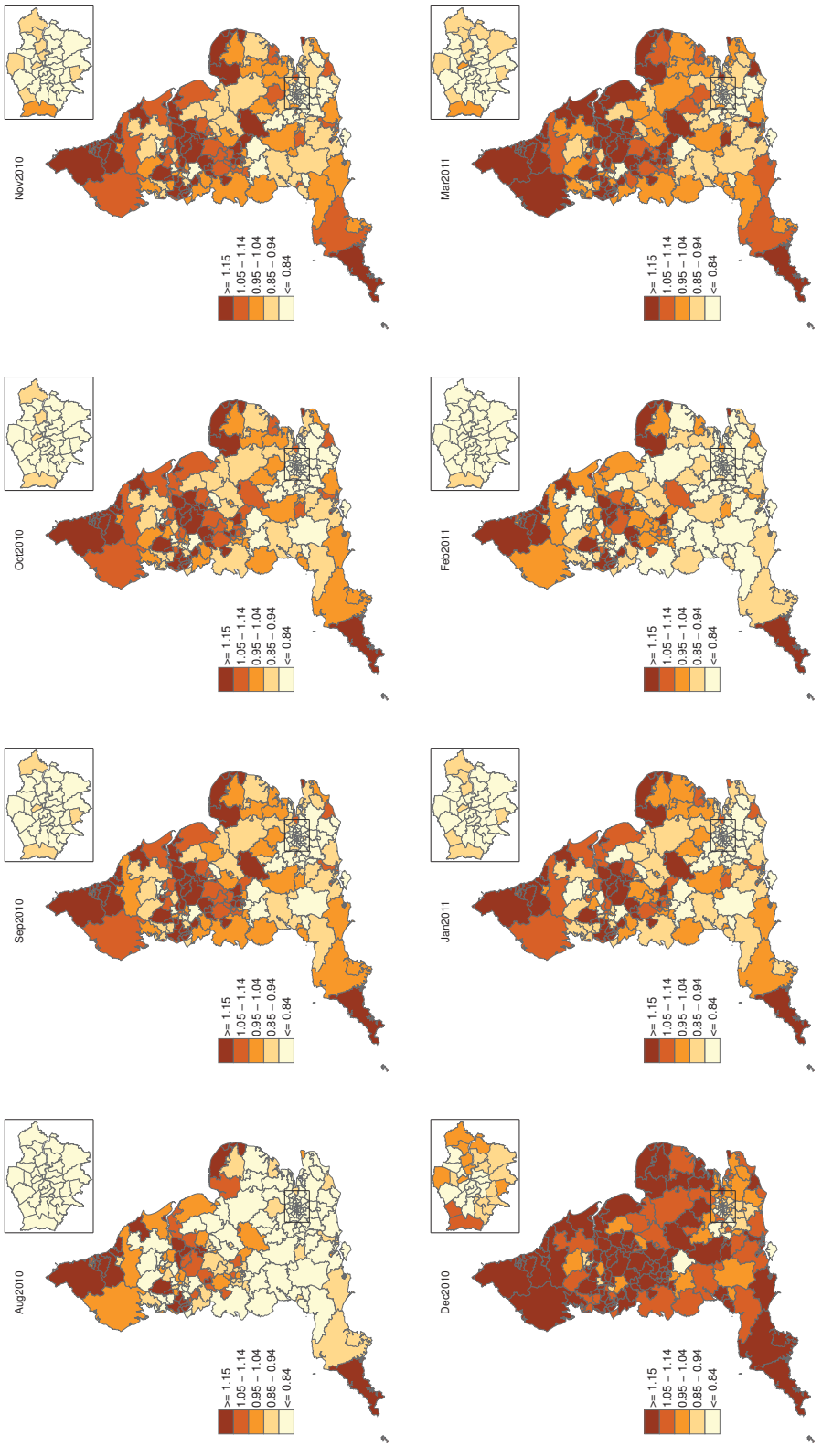


Figure B.2: Maps of crude rates for GP drug prescription data on CRD across months August 2010 to March 2011; the insets are enlarged views of London.

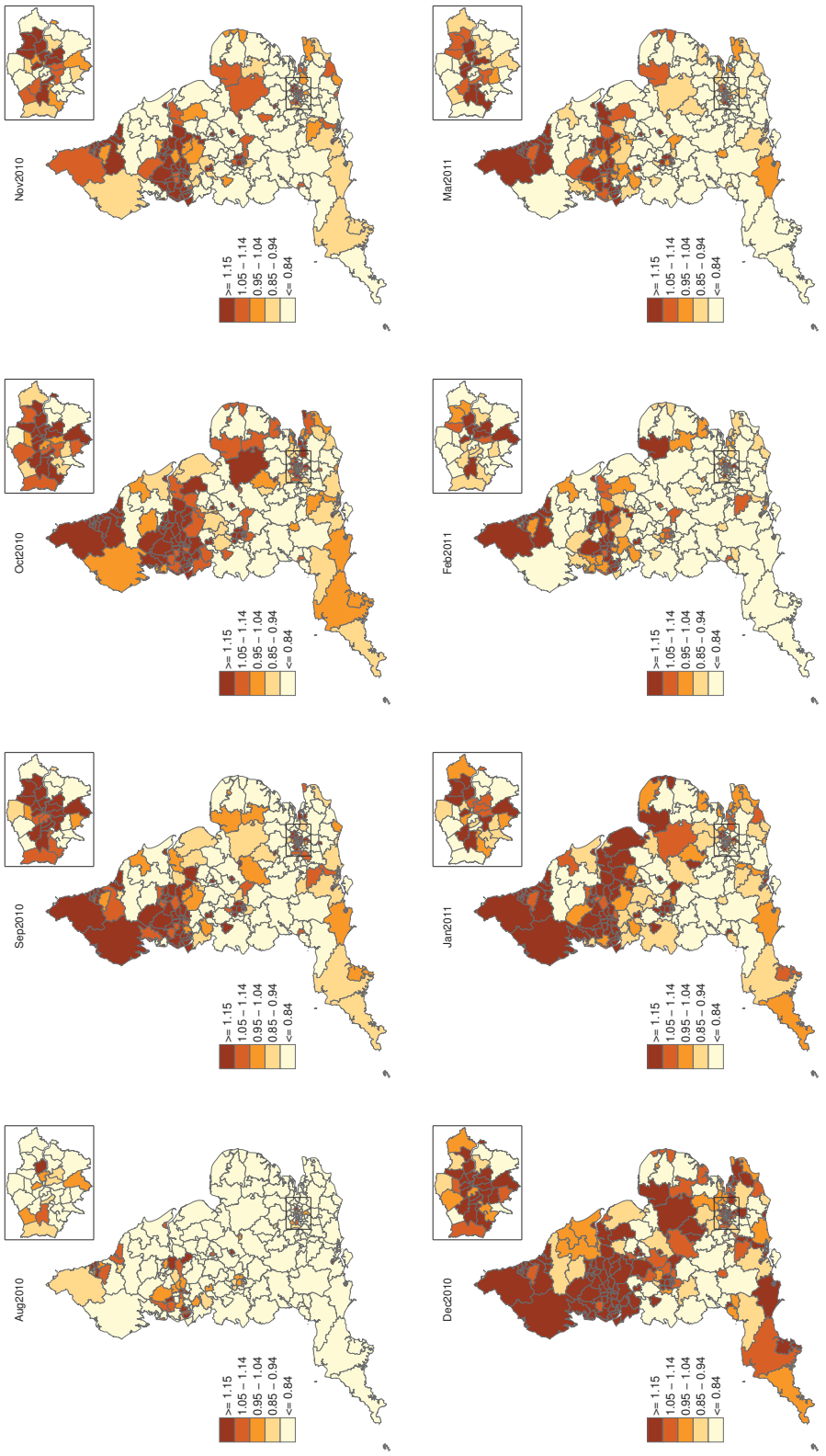


Figure B.3: Maps of crude rates for HES admission data on asthma and COPD across months August 2010 to March 2011; the insets are enlarged views of London.

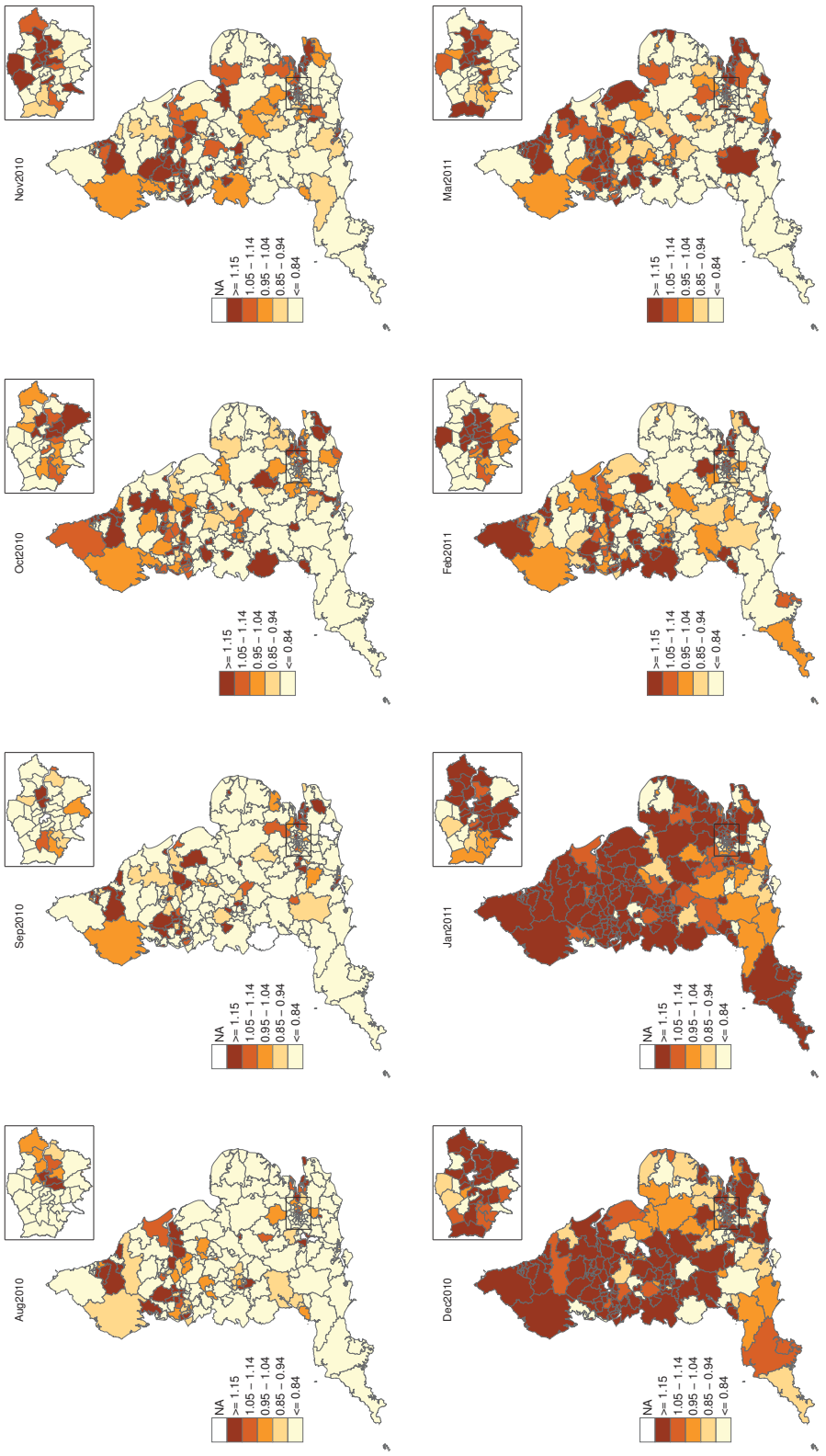


Figure B.4: Maps of crude rates for mortality data on asthma and COPD across months August 2010 to March 2011; the insets are enlarged views of London.

C Supplementary material for Chapter 4

Simulation study

Selection of 15 unusual areas

Unusual areas were selected under two different spatial scenarios: i) isolated and ii) clustered. In order to obtain a good representation of areas, these were selected based on a range of expected cases and spatial risk as the levels of these are expected to affect the ability of the model to detect a particular area.

Following Li et al. (2012), for scenario i, 3 areas were selected from each of the percentiles 10th, 25th, 50th, 75th and 90th. At each percentile, each of the 3 areas corresponds to one of the 3 levels of the overall spatial risk: low (within the 10th-30th percentiles), medium (within the 45th-55th percentiles) and high (within the 70th-90th percentiles). For spatial scenario ii, 3 clusters were selected, consisting of 4, 5, and 6 areas respectively. For each of these clusters, areas with a variety of levels in their expected cases and spatial risk were selected. Under each scenario, 15 areas were selected to be unusual.

Construction of unusual temporal patterns

If γ_t, for $t = 1, \ldots, 15$, is the common time trend, and γ_t^* is the time trend corresponding to the unusual areas, then the following 4 temporal scenarios were considered:

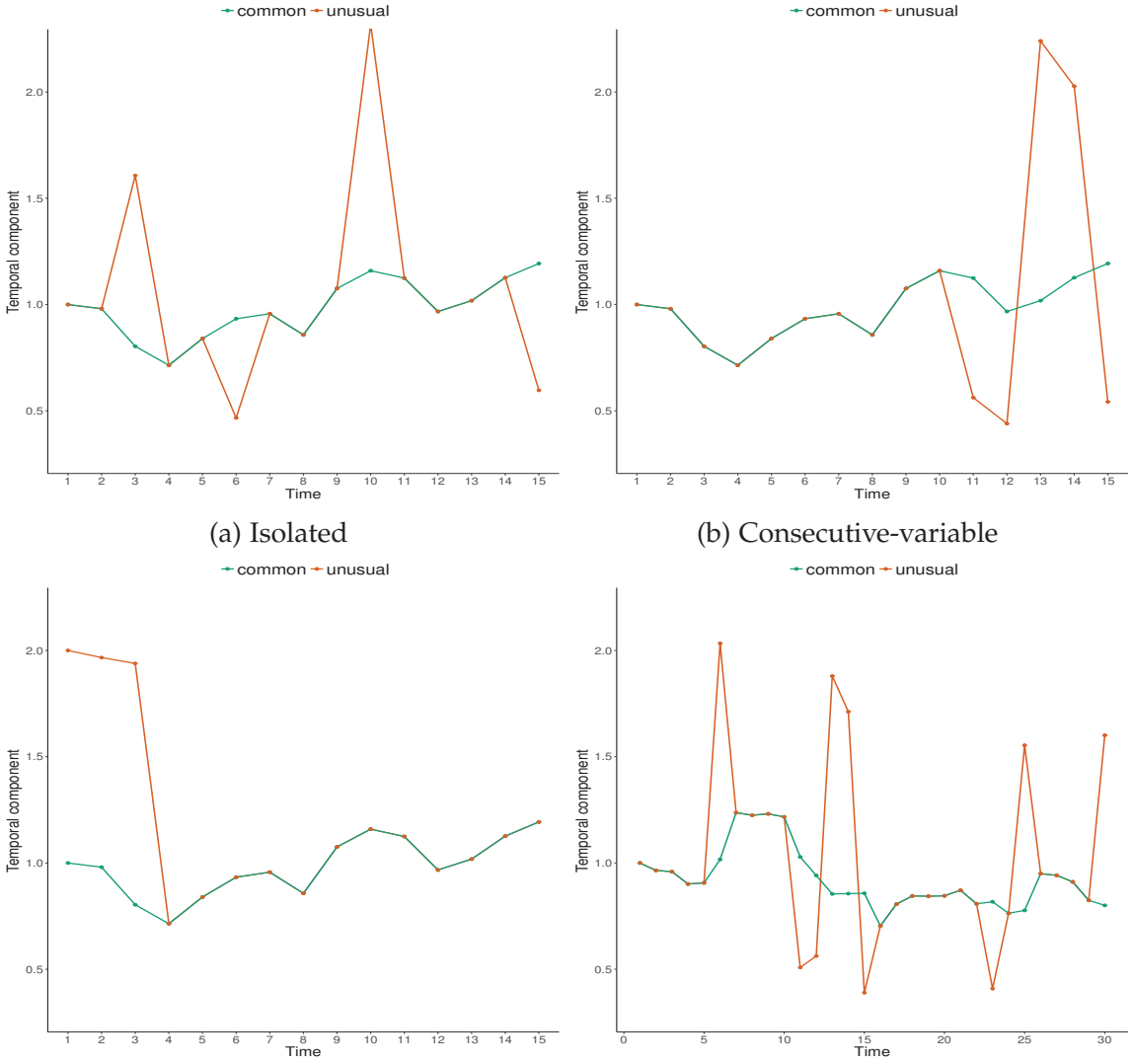
1. Isolated: $\gamma_t^* = \gamma_t + \log(2)$, for t = {3, 10}, $\gamma_t^* = \gamma_t - \log(2)$, for t = {6, 12, 15}, and $\gamma_t^* = \gamma_t$ otherwise.

2. Consecutive-variable: $\gamma_t^* = \gamma_t - \log(2)$ for t = 11, $\gamma_t^* = \gamma_t - \log(2.2)$ for t = {12, 15}, $\gamma_t^* = \gamma_t + \log(2.2)$ for t = 13, $\gamma_t^* = \gamma_t + \log(1.8)$ for t = 14, and $\gamma_t^* = \gamma_t$ otherwise.

3. Consecutive-stable: $\gamma_t^* = \gamma_t + \log(2)$, for t = 1, $\gamma_t^* = \gamma_{t-1} + \epsilon$, for t = {2, 3}, where ϵ represents small random noise, and $\gamma_t^* = \gamma_t$ otherwise.

4. Longer time series: $\gamma_t^* = \gamma_t + \log(2)$, for t = {6, 14, 25, 30}, $\gamma_t^* = \gamma_t - \log(2)$, for t = {11, 23} and $\gamma_t^* = \gamma_t + \log(2.2)$, for t = 13, $\gamma_t^* = \gamma_t - \log(2.2)$, for t = {12, 15} and $\gamma_t^* = \gamma_t$ otherwise.

Figure C.1 shows the unusual temporal patterns corresponding to the above scenarios.



(a) Isolated

(b) Consecutive-variable

(c) Consecutive-stable

(d) Longer time series

Figure C.1: Temporal scenarios for the main simulation study (15 and 30 time points).

Summary statistics for the expected cases

Expected counts	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Normal	8	22.00	29	33.20	39.00	108
Increased	16	44.00	58	66.41	78.00	216
Reduced	4	11.00	14	16.60	19.50	54

Table C.1: Summary statistics for counts of asthma hospitalisations used as expected counts in the simulation study.

Simulation scenario: short time series

In addition to the above simulation study, one with short time series, where 8 time points are considered, is also performed following closely the simulation design of the original paper by Li et al. (2012). The spatial scenario with isolated unusual areas was used.

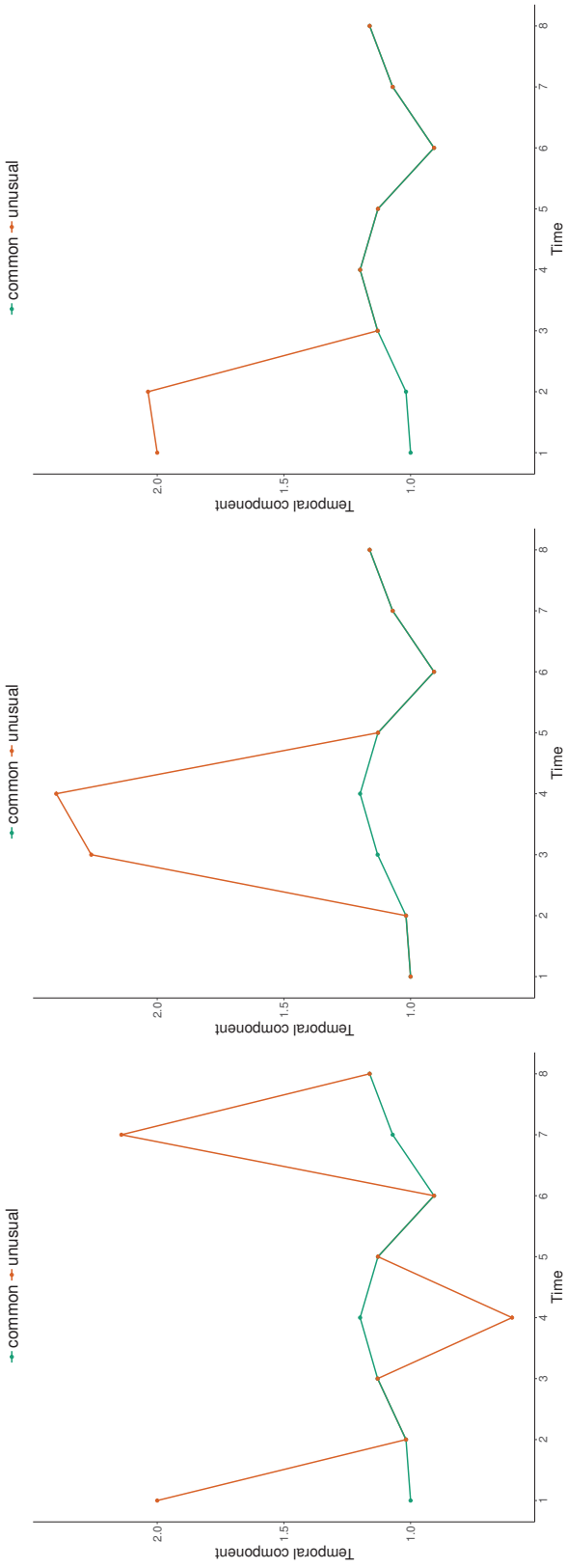
If γ_t, for $t = 1, \ldots, 8$, is the common time trend and γ_t^* is the time trend corresponding to the unusual areas, we assumed the following 3 temporal scenarios:

1. Scenario 1: $\gamma_t^* = \gamma_t + \log(2)$ for t = {1, 7}, $\gamma_t^* = \gamma_t - \log(2)$ for t = 4, and $\gamma_t^* = \gamma_t$ otherwise.

2. Scenario 2: $\gamma_t^* = \gamma_t + \log(2)$ for t = {3, 4}, and $\gamma_t^* = \gamma_t$ otherwise.

3. Scenario 3: $\gamma_t^* = \gamma_t + \log(2)$ for t = {1, 2}, and $\gamma_t^* = \gamma_t$ otherwise.

Figure C.2 shows the unusual temporal patterns plotted against the common one under the short time series simulation. Results can be seen in Table C.2.



(a) Pattern 1 (b) Pattern 2 (c) Pattern 3

Figure C.2: Temporal scenarios for the simulation study with short time series (8 time points).

Scenario	%FP	Sensitivity	Specificity	Global error
Baseline model				
Scenario 1	0.079 (0.062, 0.125)	0.885 (0.867, 0.933)	0.994 (0.990, 0.995)	0.014 (0.009, 0.014)
Scenario 2	0.089 (0.064, 0.133)	0.889 (0.867, 0.933)	0.993 (0.990, 0.995)	0.014 (0.009, 0.019)
Scenario 3	0.059 (0.00, 0.083)	0.844 (0.750, 0.933)	0.996 (0.995, 1.000)	0.015 (0.009, 0.019)
Proposed model				
Scenario 1	0.030 (0.025, 0.054)	0.733 (0.689, 0.778)	0.996 (0.999, 0.999)	0.011 (0.007, 0.008)
Scenario 2	0.159 (0.121, 0.192)	0.667 (0.600, 0.733)	0.998 (0.997, 0.998)	0.008 (0.007, 0.009)
Scenario 3	0.080 (0.046, 0.115)	0.715 (0.642, 0.767)	0.999 (0.998, 0.999)	0.006 (0.006, 0.007)

Table C.2: Posterior means of performance measures for the simulation study with short time series; credible intervals are in parentheses.

Case study 1: Road traffic accidents

Year	Min	1st Quartile	Median	Mean	3rd Quartile	Max
2005	0	44.00	61	74.77	89.50	442
2006	0	44.00	59	74.22	89.00	435
2007	0	43.00	59	72.49	87.00	426
2008	0	40.00	55	66.90	79.00	391
2009	0	38.50	53	63.86	75.50	391
2010	0	34.50	49	59.29	72.50	373
2011	0	34.00	50	60.91	72.50	422
2012	0	35.50	49	60.61	70.00	372
2013	0	34.50	49	57.20	66.50	372
2014	1	37.00	50	60.04	70.00	380
2015	0	36.00	49	58.45	69.00	407

Table C.3: Summary statistics for counts of road traffic accidents at district level by year.

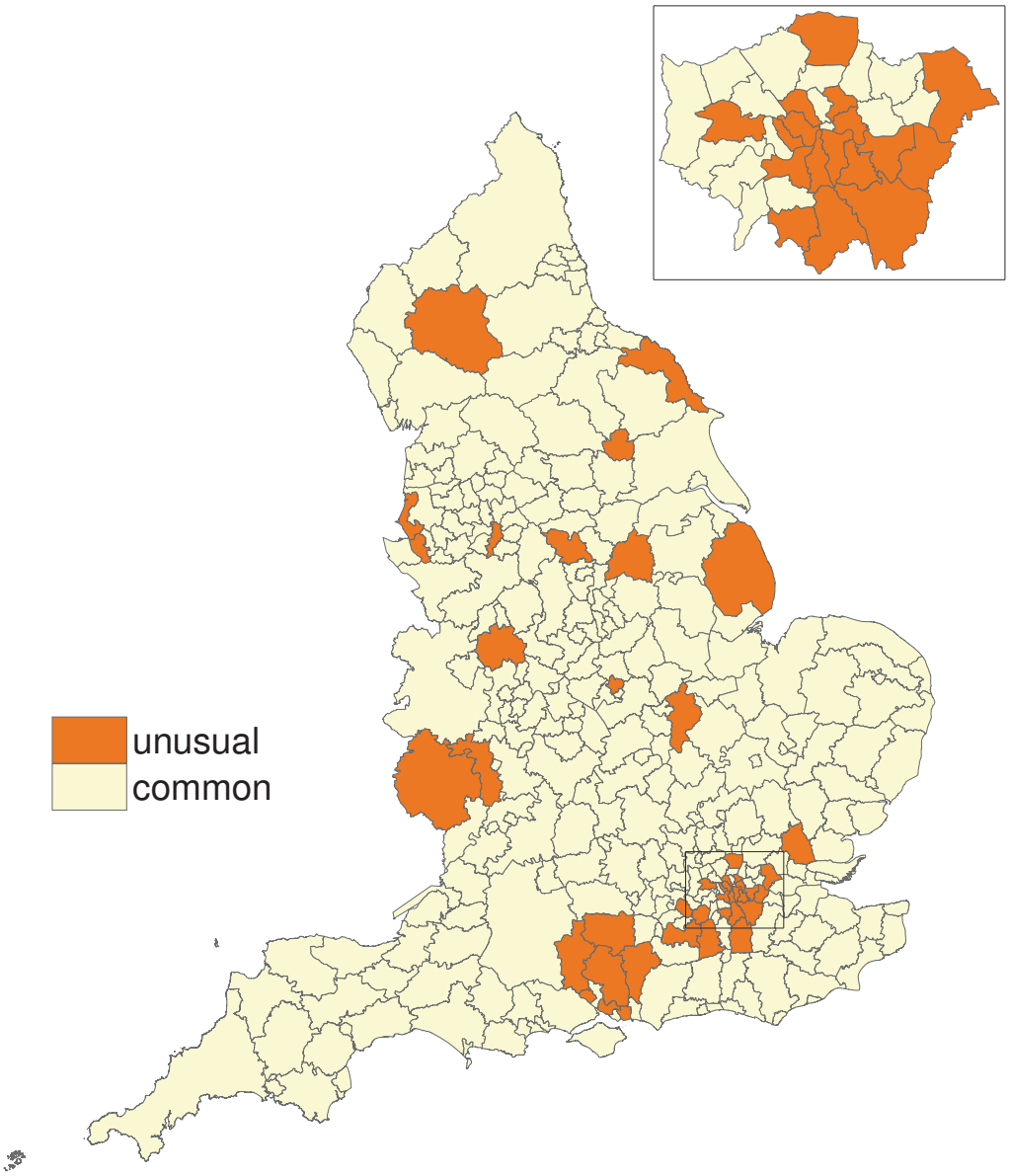


Figure C.3: Map of England with common and unusual districts in terms of road traffic accident risk; the insets are enlarged views of London.

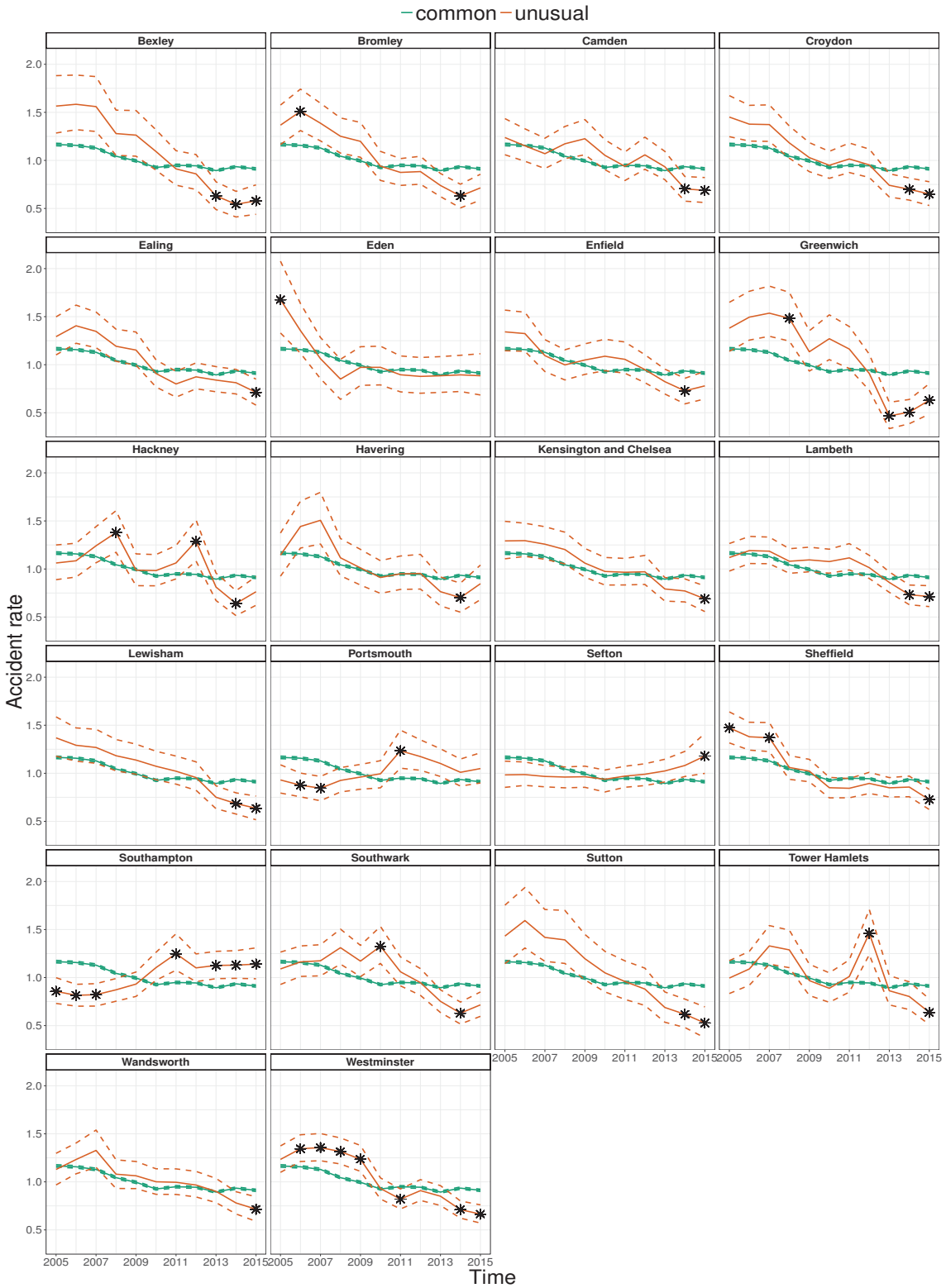


Figure C.4: Area-specific time trends plotted against the national trend for road traffic accidents in England; the first half of the total 44 unusual areas are plotted.

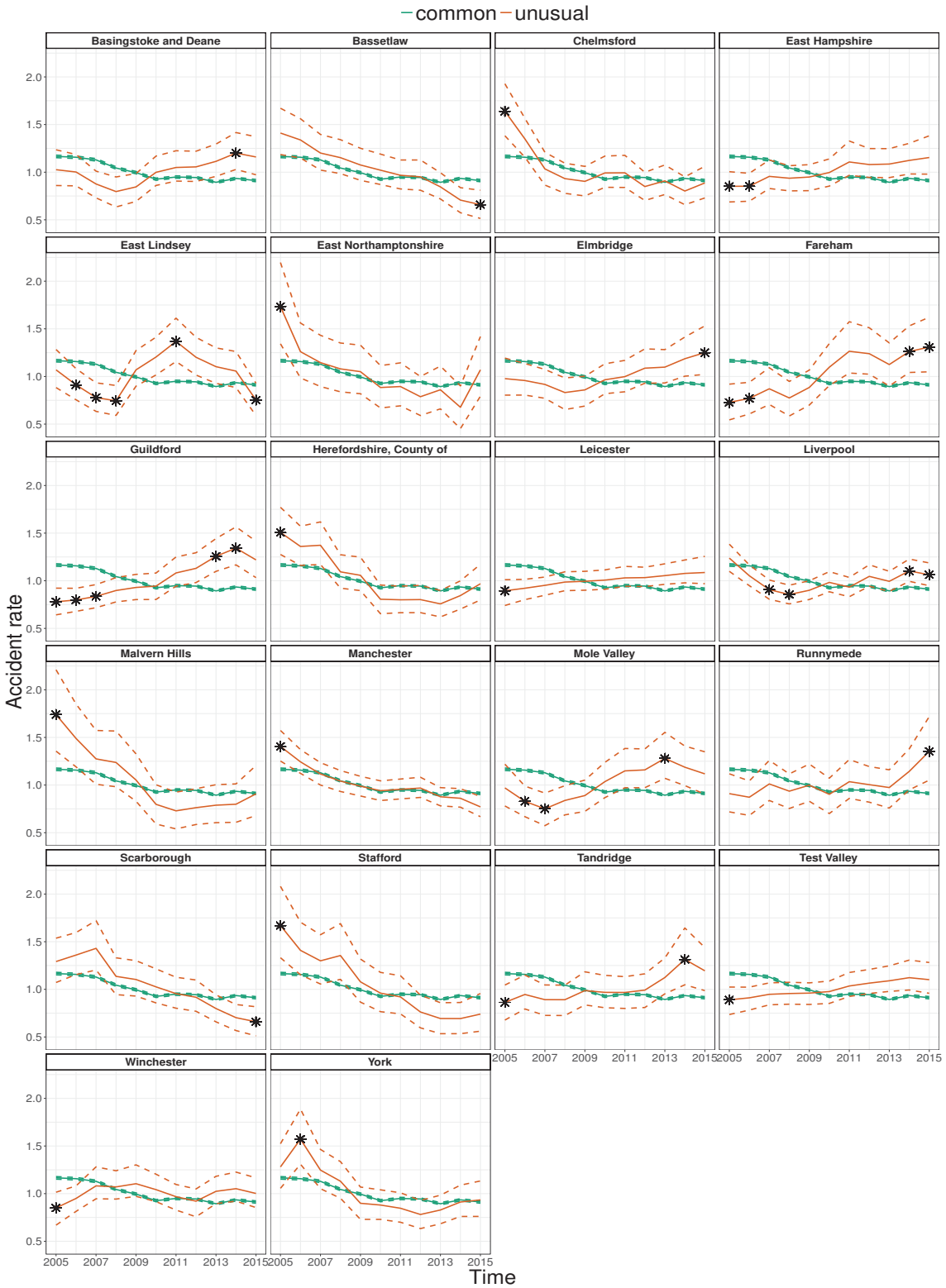


Figure C.5: Area-specific time trends plotted against the national trend for road traffic accidents in England; the second half of the total 44 unusual areas are plotted.

Case study 2: COPD admissions

Year	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Apr 2010	14	41.00	59	71.28	89.50	266
May 2010	9	43.50	63	73.00	93.50	286
Jun 2010	8	39.00	54	64.06	79.00	245
Jul 2010	12	35.00	50	58.33	73.50	247
Aug 2010	8	35.00	50	59.16	74.00	224
Sep 2010	14	49.00	70	81.36	102.00	354
Oct 2010	16	53.50	77	89.79	114.00	367
Nov 2010	16	50.00	70	80.33	99.00	337
Dec 2010	16	66.00	93	106.00	135.50	373
Jan 2011	15	58.00	79	94.00	119.00	330
Feb 2011	12	44.00	64	74.19	90.00	260
Mar 2011	14	44.50	67	76.63	95.50	276

Table C.4: Summary statistics for counts of HES admissions on COPD at CCG level by month.

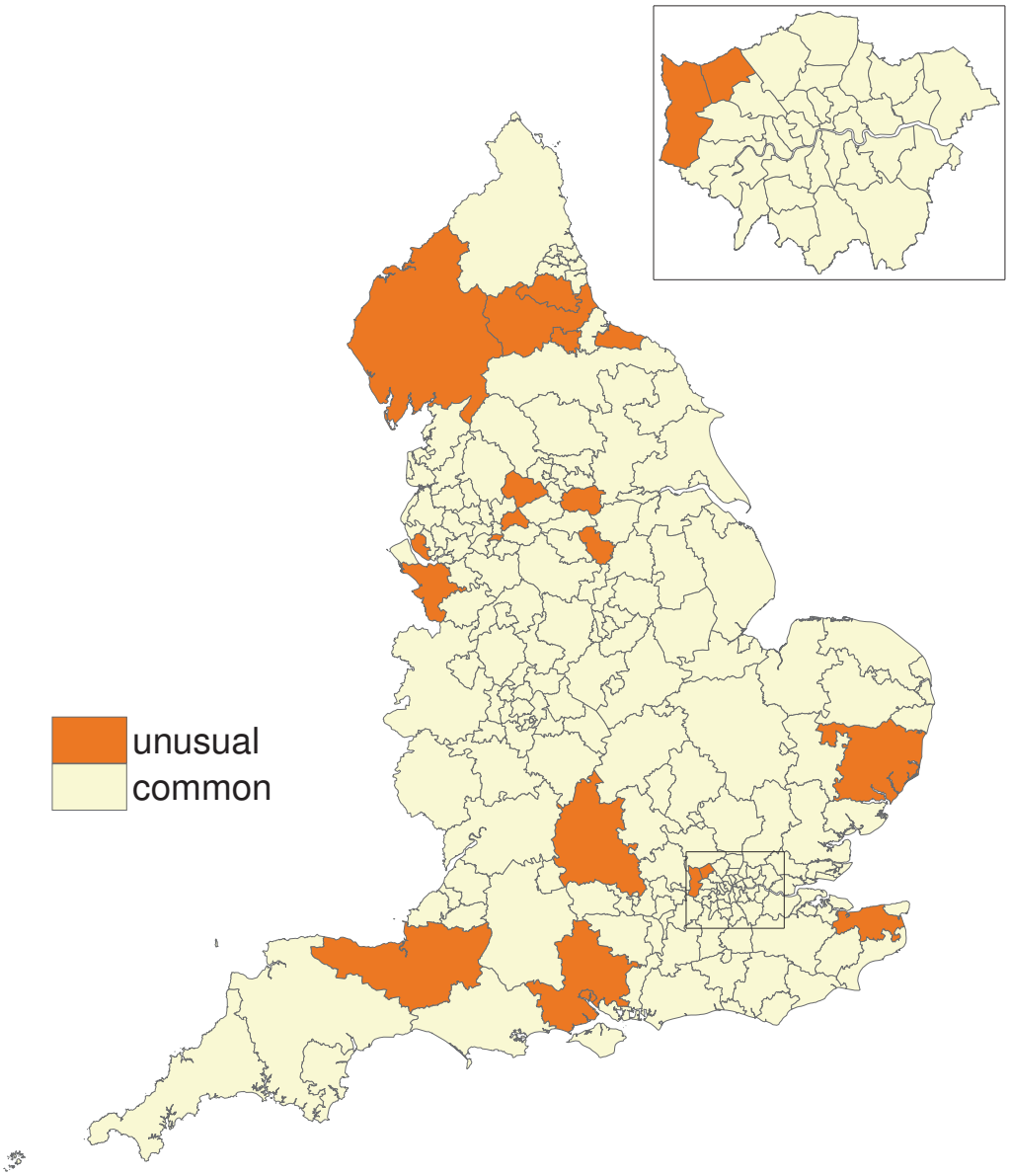


Figure C.6: Map of England with common and unusual CCGs in terms of COPD admissions relative risk; the insets are enlarged views of London.

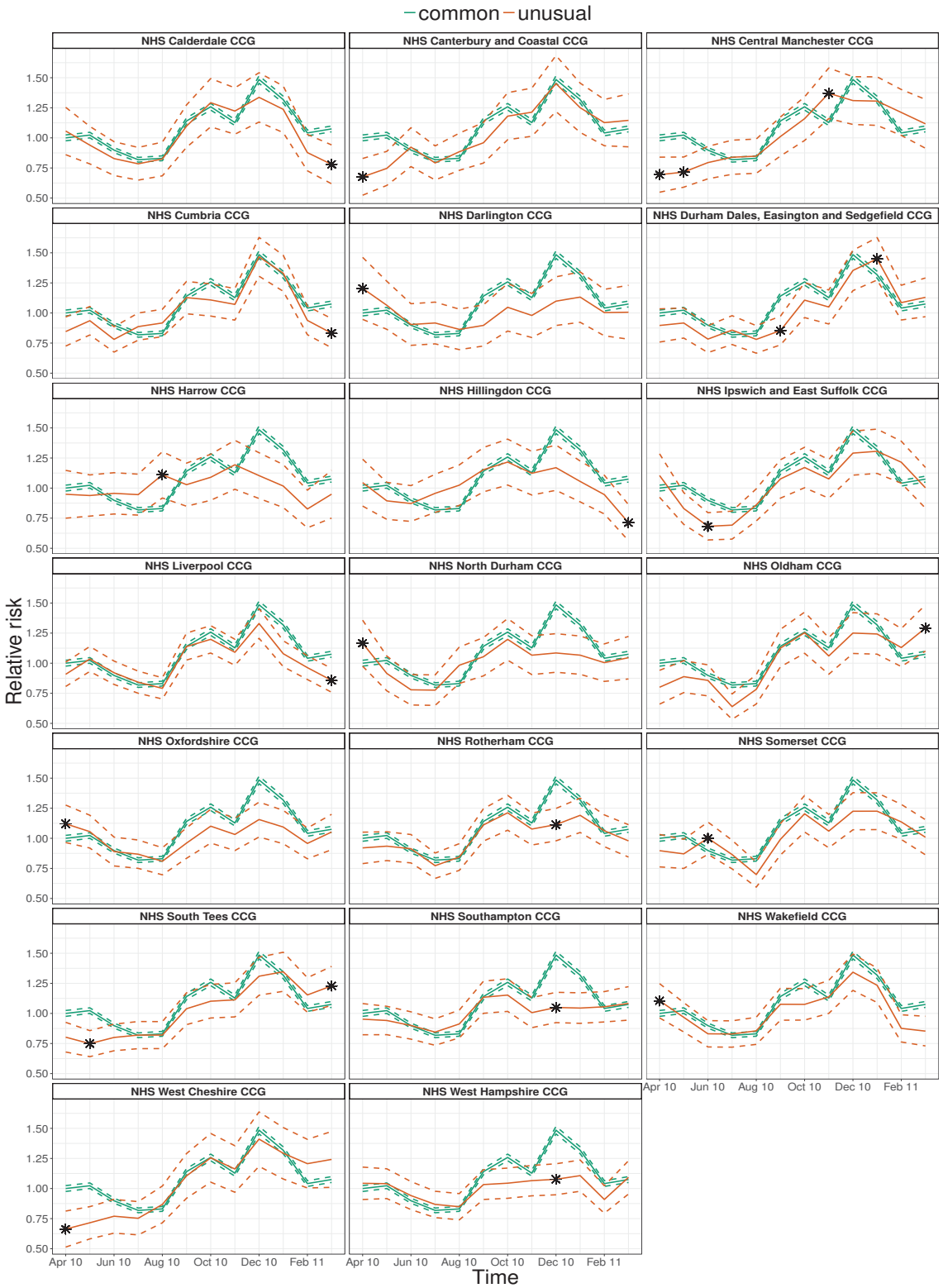


Figure C.7: Area-specific time trends plotted against the national trend for COPD admissions in England.

BUGS Model

```
Model {
for (i in 1:N) {
  for (t in 1:T) {
    Y[i, t] ~ dpois(m[i, t])
    log(m[i, t]) <- log.mu[i, t] + log(Off[i])
    log.mu[i, t] <-  z[i, t] * common[i, t] + (1 - z[i, t]) * specific[i, t]
    z[i, t] ~ dbern(p[i, t])
    logit(p[i, t]) <- beta[i] + wz[t] + logit(phi)
} }

##    Priors for the probability model
phi ~ dunif(0.9, 1)
beta[1:N] ~ car.normal(adj[], weights[], num[], prec.beta)
wz[1:T] ~ car.normal(adj.tm[], weights.tm[], num.tm[], prec.wz)

####    Common-trend model
for (i in 1:N) {
  for (t in 1:T) {
    Y1[i, t] ~ dpois(mu1[i, t])
    log(mu1[i, t]) <- alpha + log(Off[i]) + V[i] + W[t]
    temp1[i, t] <-  alpha + V[i] + W[t]
    common[i, t] <- cut(temp1[i, t])
  }
  # Posterior residual spatial effects
  resRR[i] <- exp(V[i])
}
# Posterior residual temporal effects
for (t in 1:T) {
  Time_common[t] <- exp(W[t])
}

##  Priors for the Common-trend model
# Convolution prior for the spatial effects
for (i in 1:N) {
  V[i] ~ dnorm(U[i], prec.v)
}
U[1:N] ~ car.normal(adj[], weights[], num[], prec.u)
# RW1 prior for the temporal effects
W[1:T] ~ car.normal(adj.tm[], weights.tm[], num.tm[], prec.w)
```

```
## Hyperpriors
alpha ~ dflat()
meanRR <- exp(alpha)
prec.w <- pow(sigma.w,-2)
sigma.w ~ dnorm(0,1)I(0,)
prec.wz <- pow(sigma.wz,-2)
sigma.wz ~ dnorm(0,1)I(0,)
prec.v <- pow(sigma.v,-2)
sigma.v ~ dnorm(0,1)I(0,)
prec.u <- pow(sigma.u,-2)
sigma.u ~ dnorm(0,1)I(0,)
prec.beta <- pow(sigma.beta,-2)
sigma.beta ~ dnorm(0,1)I(0,)

####    Area-specific trend model
for (i in 1:N) {
  for (t in 1:T) {
    Y2[i, t] ~ dpois(mu2[i, t])
    log(mu2[i, t]) <-  log(off[i]) + u[i] + xi[i, t]
    temp2[i, t] <-  u[i] + xi[i, t]
    specific[i, t] <- cut(temp2[i, t])
  }
  #  area-specific trends
  xi[i, 1:T] ~ car.normal(adj.tm[], weights.tm[], num.tm[], prec.xi[i])
  #  area-specific intercepts
  u[i] ~ dnorm(0, 0.001)
  # hyperpriors
  prec.xi[i] <- pow(sigma.xi[i], -2)
  sigma.xi[i] ~ dnorm(0,1)I(0,)
}

# Posterior residual area-specific temporal effects
for ( i in 1:N) {
  for (t in 1:T) {
    Time_specific[i, t] <- exp(xi[i, t])
} }

# Weights and adjacency matrix for the RW1 prior
for(t in 1:1) {
  weights.tm[t] <- 1; adj.tm[t] <- t+1; num.tm[t] <- 1
}
for (t in 2:(T-1)) {
```

```
  weights.tm[2+(t-2)*2] <- 1; adj.tm[2+(t-2)*2] <- t-1
  weights.tm[3+(t-2)*2] <- 1; adj.tm[3+(t-2)*2] <- t+1; num.tm[t] <- 2
}
for (t in T:T) {
  weights.tm[(T-2)*2 + 2] <- 1; adj.tm[(T-2)*2 + 2] <- t-1; num.tm[t] <- 1
}
} # end of model
```