METHODS FOR THE ANALYSIS OF SPATIO-TEMPORAL MULTI-STATE PROCESSES

by

Farouk S. Nathoo

M.Math (Statistics), University of Waterloo, 2000

B.Sc. (Mathematics and Statistics), University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
Fall 2005

APPROVAL

Name: Farouk S. Nathoo

Degree: Doctor of Philosophy

Title of thesis: Methods for the Analysis of Spatio-Temporal Multi-State Processes

Examining Committee: Dr. Richard Lockhart
Chair

Dr. C. Dean
Senior Supervisor

Dr. Alain Vanasse
Supervisory Committee

Dr. Tim B. Swartz
Supervisory Committee

Dr. Carl J. Schwarz
Internal External Examiner

Dr. Lance A. Waller
External Examiner,
Emory University

Date Approved: December 9, 2005

Abstract

Studies of recurring infection or chronic disease often collect longitudinal data on the disease status of subjects. Multi-state transitional models are commonly used for describing the development of such longitudinal data. In this setting, we model a stochastic process, which at any point in time will occupy one of a discrete set of states and interest centers on the transition process between states. For example, states may refer to the number of recurrences of an event or the stage of a disease.

Geographic referencing of data collected in longitudinal studies is progressively more common as scientific databases are being linked with GIS systems. This has created a need for statistical methods addressing the resulting spatial-longitudinal structure of the data. In this thesis, we develop hierarchical mixed multi-state models for the analysis of such longitudinal data when the processes corresponding to different subjects may be correlated spatially over a region. Methodological developments have been strongly driven by studies in forestry and spatial epidemiology.

Motivated by an application in forest ecology studying pine weevil infestations, the second chapter develops methods for handling mixtures of populations for spatial discrete-time two-state processes. The two-state discrete-time transitional model, often used for studying chronic conditions in human populations, is extended to settings where subjects are spatially arranged. A mixed spatially correlated mover-stayer model is developed. Here, clustering of infection is modelled by a spatially correlated random effect reflecting the density or closeness of the individuals under study. Analysis is carried out using maximum likelihood with a Monte Carlo EM algorithm for implementation and also using a fully Bayesian analysis.

The third chapter presents continuous-time spatial multi-state models. Here, joint

modelling of both the spatial correlation as well as correlation between different transition rates is required and a multivariate spatial approach is employed. A proportional intensities frailty model is developed where baseline intensity functions are modelled using both parametric Weibull forms as well as flexible representations based on cubic B-splines. The methodology is applied to a study of invasive cardiac procedure in Quebec examining readmission and mortality rates over a four-year period.

Finally, in the fourth chapter we return to the two-state discrete-time setting. An extension of the mixed mover-stayer model is motivated and developed within the Bayesian framework. Here, a multivariate conditional autoregressive (MCAR) model is incorporated providing flexible joint correlation structures. We also consider a test for the number of mixture components, quantifying the existence of a hidden subgroup of 'stayers' within the population. Posterior summarization is based on a Metropolis-Hastings sampler and methods for assessing the model goodness-of-fit are based on posterior predictive comparisons.

To my parents and my grandparents, thanks for everything.

Acknowledgments

I would like to begin by thanking my senior supervisor Dr. Charmaine Dean for her support and guidance in countless ways. I feel very fortunate to have had Dr. Dean as my supervisor. I also wish to thank Dr. Alain Vanasse, Dr. Théophile Niyonsenga, Dr. Abbas Hemiari, Dr. Josiane Courteau and all other members of our GEOIDE research team for their support and friendship.

Many thanks go to the faculty and staff of the Department of Statistics and Actuarial Science of Simon Fraser University for providing an excellent environment for graduate studies. I have many fond memories of my time here, especially playing on the departmental basketball team. A special thanks also goes to my friends and fellow students: Laurie Ainsworth, David Beaudoin, Suman Jiwani, Crystal Linkletter, Chunfang Lin, Jason Nielsen, Pritam Ranjan, Giovanni Silva, Darby Thompson and many more...

I would also like to acknowledge Dr. Bovas Abraham from the Department of Statistics and Actuarial Science at the University of Waterloo and Dr. Harry Joe and Dr. Jim Zidek from the Department of Statistics at the University of British Columbia all for encouraging me to pursue statistics at a higher level.

I am grateful for the financial support provided by Simon Fraser University and the GEOIDE network throughout the course of my studies.

Finally, and most importantly, I would like to thank my family. Especially my dad who has been a constant source of support and wisdom and Yasmin Vasanji for her love and patience.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Multi-state modelling is a powerful and convenient approach for describing the progression of longitudinal data. The framework is broad and encompasses techniques for the analysis of multivariate censored event-time data as well as methods for the analysis of longitudinal discrete data. In this thesis, multi-state transitional models are considered in a spatial setting. In essence, the work blends ideas adapted from longitudinal data analysis and spatial statistics. Methodological developments have been strongly motivated by studies in: 1) forest ecology, where interest lies in managing trees, forests and their associated resources for human benefit, and 2) epidemiologic studies, where investigators are interested in the spatial distribution of health-related states or events.

In the forest ecological setting, we have developed methods for analysis of data arising from a study of recurrent white pine weevil (*Pissodes strobi*) infestation in a white pine plantation in British Columbia. In this seven-year longitudinal study, conducted by the Ministry of Forests in British Columbia, each tree within the plantation was inspected each Fall for the presence of infection. Our main interest was to describe the pattern of weevil infestation throughout the area over the seven years of observation. White pine weevil infection poses a significant threat to British Columbia forests and there has been enormous investment recently on studying this disease.

In the spatial epidemiological setting, we have developed techniques for analysis of spatial data arising from a study of revascularization intervention in Quebec. In this four-year longitudinal study, patients hospitalized for acute coronary syndrome were followed over time and information regarding subsequent hospital readmissions and mortality was obtained. Additional demographic and treatment information was also obtained for each patient along with the local health unit in which the subject resides. The local health units serve as a geographical stratification of the subjects involved in the study. Here, interest lies in the identification of spatial heterogeneity in both mortality and readmission rates across the various local health units of the province.

We begin, in this section, with a review of some preliminary ideas that form the basis for model building and inference in later chapters. We then outline the remainder of the thesis in the next section.

1.1.1 Multi-State Models

In the multi-state modelling framework we assume that individuals in some population will occupy one of states $1, ..., k$ over a period of time. As subjects are observed over time, they may make changes from one state to another and we refer to such changes of state as *transitions*. Examining transitions can give insight into the dynamic aspects of the process under consideration. The state structure, often depicted graphically, specifies the states and which state-to-state transitions are possible. Figure 1.1 gives examples of two state structures which are often employed for models of chronic disease. The state structure depicted in Figure 1.1a will be employed in Chapters 2 and 4 where we consider models for recurring tree infection. The structure shown in Figure 1.1b, the so-called illness-death model (Hougaard 2000), is appropriate for modelling both disease and mortality simultaneously.

Models describing the evolution of a discrete-time process $Y(t)$, $t = 0, 1, 2, ...$ are typically specified through transition probabilities

$$p_{ij}(t) = Pr(Y(t) = j | Y(t-1) = i, H(t)) \tag{1.1}$$

where $H(t) = \{Y(u),\, u = 0, ..., t-1\}$ denotes the history of the process up to time

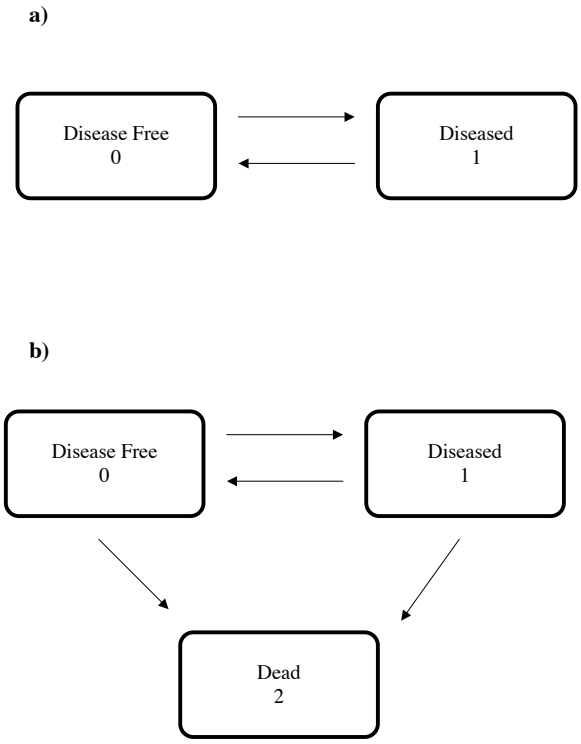


Figure 1.1: State structures commonly employed for modelling chronic diseases (a) a typical state structure used for modelling a recurring disease process; (b) a state structure for joint modelling of disease and mortality.

t. In the continuous-time setting, the transition process is governed by transition intensity functions which are defined by

$$q_{ij}(t) = \lim_{s \to 0} \frac{Pr(Y(t+s) = j | Y(t^-) = i, H(t))}{s} \tag{1.2}$$

where $H(t) = \{Y(u), 0 \leq u < t\}$. Typically, a Markov assumption is made where it is assumed that the entire history of the process is captured by the current state. In this case, (1.1) reduces to $p_{ij}(t) = Pr(Y(t) = j | Y(t-1) = i)$ (for a 1^{st} order chain) and (1.2) reduces to $q_{ij}(t) = \lim_{s \to 0} \frac{Pr(Y(t+s)=j|Y(t^-)=i)}{s}$. In more complex situations, Markov processes can be used as building blocks in a hierarchical framework used to specify *mixed* Markov models. This approach was considered by Cook and Ng (1997), Ng and Cook (1997) and Albert and Waclawiw (1998) in non-spatial settings.

1.1.2 Hierarchical Spatial Modelling

The modelling of non-Gaussian spatially correlated data typically proceeds in a hierarchical framework. Within such a framework, observations are assumed conditionally independent at the lowest level of the hierarchy and dependence is introduced at the second level through spatially correlated random effects. The random effects account for heterogeneity and, in many settings, represent covariates that are missing from the model. Within the realm of generalized linear models, the incorporation of spatial random effects has been studied extensively (see e.g., Besag et al., 1991, Bernardinelli and Montomoli, 1992, Best et al., 1999, Zhang 2002). In this thesis we adopt a similar approach, introducing random effects into the second level of hierarchical multi-state Markov processes.

A convenient distributional form for a vector of N spatially correlated random effects $\mathbf{b} = (b_1, ..., b_N)$ is the multivariate Gaussian with mean $\mathbf{0}$ and spatially structured covariance matrix $\mathbf{\Sigma}$. Here, associated with each b_i is either a point location $(x_i, y_i) \in \mathbf{R^2}$ or a position on a (possibly irregular) lattice. Typically, one of two approaches is adopted for specifying $\mathbf{\Sigma}$. A direct and simple approach, known as geostatistical modelling (Cressie, 1993, Diggle et al. 1998), requires knowledge of the point locations (x_i, y_i), $i = 1, ..., N$, and specifies $\mathbf{\Sigma}$ as a parametric function of these locations. A simple example is the exponential form which sets $\Sigma_{ij} = \sigma^2 \exp(-\rho d(i,j))$

and is based on two parameters $\rho > 0$ and $\sigma^2 > 0$ and the distance, $d(i,j)$, between points i and j. An alternative approach, known as conditional autoregressive modelling (CAR) (Besag, 1974, Cressie 1993, Carlin and Louis 1996) is the spatial analogue of autoregressive time-series modelling and specifies $\mathbf{\Sigma}$ indirectly through the set of conditional distributions $b_i|b_{j\neq i} \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., N$. Here, each conditional distribution is assumed univariate normal with conditional variance $\sigma_i^2 > 0$ and conditional mean $\mu_i = \sum_{j=1}^{N} w_{ij} b_j$ with $w_{ii} = 0$, $i = 1, ..., N$. The weight $w_{ij} \geq 0$ can be based on either the distance between units i and j or indicators for their adjacency on a lattice and reflects the influence of b_j on the conditional mean of b_i. With these conditional specifications, the results of Besag (1974) can be used to show that $\mathbf{\Sigma} = (\mathbf{I} - \mathbf{W})^{-1}\mathbf{M}$ where $\mathbf{W} = (w_{ij})$, $\mathbf{M} = \text{diag}\{\sigma_1^2, ..., \sigma_N^2\}$ and we impose the restriction $w_{ij}\sigma_j^2 = w_{ji}\sigma_i^2$ to ensure the symmetry of $\mathbf{\Sigma}$. A special case which has been used extensively in disease mapping is the intrinsic autoregression which sets $w_{ij} = \frac{C_{ij}}{C_{i\cdot}}$ and $\sigma_i^2 = \frac{\sigma^2}{C_{i\cdot}}$ where the C_{ij}'s are known user defined weights and $C_{i\cdot} = \sum_j C_{ij}$. This specification leads to a singular multivariate Gaussian distribution for $\mathbf{b}$.

Geostatistical models, due to their direct specification of $\mathbf{\Sigma}$ are easily interpreted; whereas, CAR models are most sensibly interpreted in a conditional sense. On the other hand, CAR models can be based on either point locations or derived at a lower spatial resolution using only the adjacency structure of a lattice. In addition, the conditional specification of CAR models makes them ideal for use with Markov chain Monte Carlo methods described in the next subsection. We adopt here the CAR modelling approach. Finally, we note that multivariate generalizations of the CAR modelling framework have been developed (see e.g. Kim et al. 2001, Carlin and Banerjee, 2002, Gelfand and Vounatsou 2003) which allow for the joint spatial modelling of $k > 1$ random effects associated with each spatial unit. We incorporate such joint spatial structures into our multi-state modelling framework in Chapters 3 and 4 of this thesis.

1.1.3 Markov chain Monte Carlo

Markov chain Monte Carlo is a collection of numerical simulation methods which allow the approximation of integrals that are analytically intractable. Even though the theory behind MCMC was developed much earlier (Metropolis et al. 1953), the techniques have become increasingly popular within the last decade, a result owing to the availability of cheap computing power.

The principle behind MCMC is the ergodic theorem applied to Markov chains. Inference with respect to some target distribution π is based on the construction of a Markov chain having π as its invariant distribution. The ergodic theorem then links expected values under π with observations, $\mathbf{x^0}$, $\mathbf{x^1}$, $\mathbf{x^2}$,..., from the Markov chain via

$$\lim_{J\to\infty} \frac{1}{J+1} \sum_{i=0}^{J} f(\mathbf{x^i}) = E_\pi[f(\mathbf{x})]$$

for any function f, integrable with respect to π. Expected values under π may then be approximated using realizations of the Markov chain. The technique is most useful when drawing realizations directly from π is not feasible and π is sufficiently high dimensional and complex so that importance sampling methods cannot be employed. This is typically the case with hierarchical spatial models involving large numbers of random effects.

The two most common MCMC algorithms are the Gibbs sampler (Geman and Geman 1984) and the Metropolis-Hastings algorithm (Hastings 1970). The Gibbs sampler is based on drawing from full conditional distributions. Suppose $\mathbf{x} = (x_1, ..., x_N)$ and it is feasible to obtain realizations from the full conditional distribution $\pi(x_j|\mathbf{x_{-j}})$ where $\mathbf{x_{-j}}$ denotes $\mathbf{x}$ with x_j removed. The Gibbs sampler changes the state of the chain $\mathbf{x^i}$ to $\mathbf{x^{i+1}}$ by updating each x_j, $j = 1, ..., N$, in turn by sampling the replacement value from the corresponding full conditional distribution $\pi(\cdot|x_1^{i+1}, ..., x_{j-1}^{i+1}, x_{j+1}^i, ..., x_N^i)$. The algorithm depends on the ability to draw from full conditional distributions. Often, the full conditional distributions will not take standard forms but the corresponding densities will be log-concave. In this case, adaptive rejection sampling (Gilks and Wild 1992) may be employed.

In situations where it is difficult to sample from full conditional distributions, the

Metropolis-Hastings algorithm, a generalization of the Gibbs sampler, may be used. In this case, the update from $\mathbf{x^i}$ to $\mathbf{x^{i+1}}$ proceeds by first generating a candidate $\mathbf{x'}$ from a proposal distribution $q(\cdot|\mathbf{x^i})$. The proposed value, $\mathbf{x'}$, is accepted as the new state of the chain, $\mathbf{x^{i+1}}$, with probability

$$\alpha(\mathbf{x^i}, \mathbf{x'}) = \min\{1, \frac{\pi(\mathbf{x'})q(\mathbf{x^i}|\mathbf{x'})}{\pi(\mathbf{x^i})q(\mathbf{x'}|\mathbf{x^i})}\} \tag{1.3}$$

Otherwise we set $\mathbf{x^{i+1}} = \mathbf{x^i}$ and the chain does not move. Note also that π only needs to be known up to a normalizing constant. When the proposal density is symmetric (1.3) reduces to $\alpha(\mathbf{x^i}, \mathbf{x'}) = \min\{1, \frac{\pi(\mathbf{x'})}{\pi(\mathbf{x^i})}\}$ and the resulting special case is referred to as the Metropolis algorithm. Often, it is not feasible to update the whole of $\mathbf{x}$ in one step. In this case, as in the Gibbs sampler, we divide $\mathbf{x}$ into components $\mathbf{x} = (x_1, ..., x_N)$ and apply a Metropolis-Hastings step to each component. This scheme includes, as a special case, the so-called hybrid samplers (Gilks et al. 1996) that update some components via Gibbs steps and others using Metropolis-Hastings steps.

In practice, an initial portion of the realized Markov chain is discarded as burn-in, a period required for the chain to 'forget' its initial state and converge to the stationary distribution. Determination of convergence is best assessed through running multiple chains, each initialized at different points in the sample space of π. Analysis then compares the output of each chain using diagnostics (see eg. Gelman and Rubin 1992) and through the examination of sample trace plots, most importantly plots which display the value of the $\log(\pi)$ (up to an additive constant) at each state of the chain.

In Chapter 2, we employ the Gibbs sampler within each iteration of an EM algorithm to approximate the conditional expectations required at each E-step. In Chapters 3 and 4 we use the Metropolis-Hastings algorithm to draw samples from posterior distributions arising from Bayesian model specifications.

1.1.4 Splines and Temporal Smoothing

Splines provide a conceptually simple approach for approximating complex nonlinear functions (De Boor 1978). In this thesis, splines of one variable are employed for

modelling temporal variation in the process governing the state-to-state transitions of a multi-state model. The basic idea behind splines is the representation of a possibly complicated curve through a combination of relatively simple smooth segments where each segment is represented by a polynomial of order D. To ensure smoothness of the composite curve, constraints are imposed on each segment at the joining points which are called inner knots. Given a set of L inner knots $t_1 < t_2 < ... < t_L$ a spline S of degree D may be written as

$$S(x) = \sum_{l=0}^{L} S_l(x) I(x \in [t_l, t_{l+1})) \tag{1.4}$$

where

$$S_l(x) = \sum_{d=0}^{D} a_d^l (x - t_l)^d,\ l = 0, ..., L$$

where $t_0 < t_{L+1}$ are boundary knots typically defined by the range of the data; and the a_d^ls are constrained to ensure that S has continuous derivatives of all degrees $\leq D-1$. As discussed by MacNab (1999), the collection of all functions taking the form (1.4) forms a linear space of dimension $D+L+1$. It is therefore spanned by any $D+L+1$ linearly independent members of the space forming a basis. The approximation of a function f over the interval $[t_0, t_{L+1}]$ using a spline of degree D and inner knots $t_1 < t_2 < ... < t_L$ is therefore given by

$$f(t) \approx \sum_{j=0}^{D+L} \alpha_j p_j(t) \tag{1.5}$$

where $\mathbf{p(t)} = \{p_0(t), ..., p_{D+L}(t)\}$ is any such basis and $\alpha_0, ..., \alpha_{D+L}$ are unknown parameters. A convenient choice, which we employ in this thesis, is the B-spline basis that is easily computed using the recursive algorithm of De Boor (1978). In addition, our functional approximations will take $D = 3$ and incorporate an intercept yielding

$$f(t) \approx \alpha_0 + \sum_{j=1}^{3+L} \alpha_j p_j(t)$$

where we exclude the first B-spline basis function $p_0(t)$ in order to identify the intercept.

1.2 Outline of Thesis

This thesis consists of three projects examining spatial variation in longitudinal multi-state processes. Chapters 2 and 4 deal with processes in discrete-time and methods developed therein are applied to the aforementioned forest ecological study. Chapter 3 develops methods for continuous-time processes in a spatial epidemiological setting. Each chapter is written in a style similar to that for publication and the summaries of the chapters provided below reflect the corresponding abstracts. As a result, some introductory material is repeated.

1.2.1 Chapter 2

Studies of recurring infection or chronic disease often collect longitudinal data on the disease status of subjects. Two-state transitional models are useful for analysis in such studies where, at any point in time, an individual may be said to occupy either a diseased or disease-free state and interest centers on the transition process between the two states. Here, two additional features are present. The data are spatially arranged and it is important to account for spatial correlation in the transitional processes corresponding to different subjects. In addition there are subgroups of individuals with different mechanisms of transitions. These subgroups are not known a priori and hence group membership must be estimated. Covariates modulating transitions are included in a logistic additive framework. Inference for the resulting mixture spatial Markov regression model is not straightforward. We develop here a Monte Carlo EM algorithm for maximum likelihood estimation and a Markov Chain Monte Carlo sampling scheme for summarizing the posterior distribution in a Bayesian analysis. The methodology is applied to a study of recurrent weevil infestation in British Columbia forests.

1.2.2 Chapter 3

Follow-up medical studies often collect longitudinal data on patients. Multi-state models can be employed for analysis in such studies where at any point in time, individuals may be said to occupy one of a discrete set of states and it is of interest to examine the process governing state-to-state transitions. For example, states may refer to the number of recurrences of an event, or the stage of a disease. We develop a hierarchical Bayesian model for the analysis of such longitudinal data when the processes corresponding to different subjects may be correlated spatially over a region. Continuous-time Markov chains incorporating spatially correlated random effects are introduced. Here, joint modelling of both spatial correlation as well as correlation between different transition rates is required and a multivariate spatial approach is employed. A proportional intensities frailty model is developed, where baseline intensity functions are modelled using both parametric Weibull forms and flexible representations based on cubic B-splines. The methodology is applied to a study of revascularization intervention in Quebec. We consider patients admitted for acute coronary syndrome throughout the 139 local health units of the province and examine readmission and mortality rates over a four-year period.

1.2.3 Chapter 4

In this final chapter we return to the discrete-time setting of Chapter 2 and develop an extended model with inference conducted from a Bayesian perspective. A joint spatial random effects model is incorporated into the transitional process of a hierarchical mover-stayer model. In this case, the random effects allow for two types of correlation. In addition to allowing for spatial correlation, we also permit correlation between subject specific transition probabilities. This flexible correlation structure is accommodated through a multivariate conditional autoregressive (MCAR) model. The chapter also develops a test for the number of mixture components, quantifying the existence of a hidden subgroup within the population. That is, we develop a test for 'stayers' in the mover-stayer model. The test is based on assigning a discrete mass

prior to the mixing probability. Testing of this point null hypothesis was of substantial interest to investigators in our forest ecological application. Inference is based on samples drawn from the posterior distribution using a Metropolis-Hastings algorithm. Finally, methods for assessing the model goodness-of-fit are developed based on posterior predictive comparisons.

1.2.4 Chapter 5

The thesis closes with a discussion of future work.

Chapter 2

A Discrete-Time Spatial Two-State Process

Studies of recurring infection or chronic disease often collect longitudinal data on the disease status of subjects. In many such studies, subjects are observed at regular time intervals and assessed for the presence/absense of a condition, such as a disease. Statistical analysis of the resulting longitudinal binary data is conveniently conducted through the use of two-state transitional models; in particular, when interest lies in the probabilities of transition between the diseased and disease-free states. In such analyses, it is typically assumed that individuals under observation are independent.

Markov chain modelling is a commonly used approach for describing a process which yields temporally dependent binary sequences. Inference in such models was considered in an early paper by Anderson and Goodman (1957) for the simple case where all subjects share the same transition probabilities. Muenz and Rubinstein (1985) allow the transition process to vary from subject to subject through regression modelling of the transition probabilities. In many scenarios there exists extra-variation which is not explained by the available covariates. To account for this extra variation, two stage, conditionally Markov processes can be employed. At the first stage, the data obtained from each subject are assumed to be drawn from a two-state Markov chain. At the second stage, continuous mixing distributions are used to model heterogeneity in transitions. Cook and Ng (1997) develop such a model incorporating

bivariate Gaussian random effects into transition probabilities and conduct maximum likelihood estimation using numerical integration. Albert and Waclawiw (1998) develop a similar model but specify only the first two moments of the independent random effects and conduct inference using generalized estimating equations.

An alternative approach to account for heterogeneity in a Markov Chain analysis is based on finite mixtures. In particular, this is useful when it is thought that a subgroup of individuals, known as 'stayers', will remain in their initial state throughout the course of observation. Such an approach can be employed for studying disease in populations where it is hypothesized that some members are resistant or immune to the condition studied. Models of this sort have been discussed by Frydman (1984), Fuchs and Greenhouse (1988) and Cook, Kalblfleisch and Yi (2002).

Independence between subjects is an assumption that is made in all models discussed above. In the application we consider, the subjects under observation are spatially arranged and it is of essence to describe the spatial correlation. Our motivating example is a study of recurrent weevil infestation in a white pine plantation in British Columbia. In this seven-year longitudinal study conducted by the Ministry of Forests in British Columbia each tree within the plantation was examined in the fall for the presence of infection. Of primary interest was to describe the pattern of weevil infestation throughout the area over the seven years of observation. White pine weevil infection poses a significant threat to British Columbia forests and there has been enormous investment recently on studying this disease.

In this chapter we present a transitional model for spatio-temporal two-state processes. There are several features of this model and our analysis which distinguish them from the usual two-state model analysis. Importantly, spatial random effects are incorporated into transition probabilities to accommodate correlation. In our study it was hypothesized that heterogeneity might arise through the presence of trees which were resistant to infection. In fact a major scientific objective in a follow-up analysis would be to identify and characterize such resistant trees with the goal of populating secondary forests with such qualities. To address this statistically, excess heterogeneity is accommodated by allowing for a subgroup of individuals whose initial state is absorbing. The resulting two-component model is of the mover-stayer

type with spatially correlated random effects introduced into the component associated with 'movers'. With the two layers of mixing distributions, one discrete and one involving a high dimensional spatial mixture, analytic tools for conducting inference need careful consideration. We develop these in both the likelihood and Bayesian frameworks and present a comparison of such methods in our analysis. Estimation is approached by maximum likelihood using a Monte Carlo EM algorithm in the classical framework and through Markov Chain Monte Carlo summarization of the posterior distribution in the Bayesian setting.

The remainder of the chapter is organized as follows. In Section 2.1 we specify the mixed spatially correlated mover-stayer model. Section 2.2 develops maximum likelihood inference for our model. A spatio-temporal analysis of weevil infestation in a white pine plantation in British Columbia is discussed in Section 2.3. In Section 2.4 we discuss extensions involving multivariate spatial processes and continuous time modelling.

2.1 Spatio-Temporal Mixed Two-State Model

Suppose there are N subjects, spatially arranged throughout some region and subject i is observed over a sequence of n_i equally spaced time points. Upon observation, each individual will occupy one of two possible states representing say, the presence or absence of some condition, for example an infection. We let state 1 denote the infected state and state 0, the infection-free state. Let $y_i(t)$ be the binary variable denoting the state occupied by subject i at time t and $\mathbf{y}_i = (y_i(0), ..., y_i(n_i - 1))'$ the sequence of states occupied by subject i, $i = 1, ..., N$.

The mixed mover-stayer model is specified hierarchically where, at the first stage of the model, we assume each response vector, $\mathbf{y}_i$, is independently drawn from a compartmental model having density

$$f_{MS}(\mathbf{y}_i|\mathbf{Z}, \mathbf{b_0}, \mathbf{b_1}) = \begin{cases} I(\mathbf{y}_i = \mathbf{0}) & \text{if } z_i = 0, \\ f_{MC}(\mathbf{y}_i|\boldsymbol{b_0}, \boldsymbol{b_1}) & \text{if } z_i = 1 \end{cases} \tag{2.1}$$

where $\mathbf{Z} = \{z_1, z_2, ..., z_N\}$ is a vector of latent variables with $z_i \in \{0, 1\}$ allocating

subject i into one of two mixture components and we adopt independent allocations to these components, $z_i \overset{ind}{\sim} \text{Bernoulli}(p_{M_i})$ $i = 1, ..., N$, where $p_{M_i} = Pr(z_i = 1)$. Extensions, allowing the components of $\mathbf{Z}$ to be spatially correlated, are considered in section 5. In (2.1), one mixture component places all its mass on the zero vector while the other component distributes mass according to the density $f_{MC}(\mathbf{y}_i|\boldsymbol{b_0}, \boldsymbol{b_1})$, which is that of a 1^{st} order, two-state Markov chain given by

$$f_{MC}(\mathbf{y}_i|\boldsymbol{b_0}, \boldsymbol{b_1}) = p_{I_i}^{y_i(0)}(1 - p_{I_i})^{1-y_i(0)} \prod_{t \in D_{0i}} p_{01_i}(t)^{y_i(t)}(1 - p_{01_i}(t))^{1-y_i(t)}$$
$$\times \prod_{t \in D_{1i}} p_{10_i}(t)^{1-y_i(t)}(1 - p_{10_i}(t))^{y_i(t)} \tag{2.2}$$

where $D_{li} = \{t > 0 | y_i(t-1) = l\}$, $l = 0, 1$, p_{I_i} is an initial state probability and $p_{01_i}(t)$ and $p_{10_i}(t)$ are transition probabilities. The transition probabilities are modelled using additive logistic specifications

$$\text{logit}(p_{01_i}(t)) = \boldsymbol{\beta_0}'\boldsymbol{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0i},$$
$$\text{logit}(p_{10_i}(t)) = \boldsymbol{\beta_1}'\boldsymbol{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1i}, \tag{2.3}$$

$i = 1, ..., N$, $t = 1, ..., n_i - 1$, where $\boldsymbol{x_i(t)}$ is a p-vector of covariates associated with subject i at time t; $\boldsymbol{\beta}$ a vector of regression parameters; $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ are functions of time describing temporal trends in transitions and b_{0i} and b_{1i} are random effects accounting for spatial correlation.

Several types of temporal trends can be considered. We allow for flexible forms using cubic B-splines. The cubic B-spline representations used here are given by

$$g_l(t, \boldsymbol{\alpha_l}) = \alpha_{l_0} + \sum_{j=1}^{K_l+3} \alpha_{l_j} p_{l_j}(t), \quad l = 0, 1 \tag{2.4}$$

where $\boldsymbol{\alpha_l} = (\alpha_{l_0}, ..., \alpha_{l_{K_l+3}})$, $l = 0, 1$, are vectors of unknown coefficients and $\{p_{l_1}(t), ..., p_{l_{K_l+3}}(t)\}$, $l = 0, 1$, are sets of known B-spline basis functions with K_l, $l = 0, 1$, representing the number of inner knots used in the representations. For the spatial random effects, $\boldsymbol{b_0} = (b_{01}, ..., b_{0N})'$ and $\boldsymbol{b_1} = (b_{11}, ..., b_{1N})'$, we adopt Gaussian intrinsic autoregressive (IAR) models based on conditional specifications of the form

$$b_{li}|\{b_{lj}, j \neq i\} \sim N\left(\frac{\sum_j C_{ij} b_{lj}}{\sum_j C_{ij}}, \frac{\sigma_{bl}^2}{\sum_j C_{ij}}\right), \quad l = 0, 1 \tag{2.5}$$

where C_{ij} are user-defined weights measuring the closeness or adjacency of subjects i and j ($C_{ii} = 0$) with $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ and $\mathbf{Z}$ assumed independent. As a result of (2.5), each vector $\boldsymbol{b_l}$, $l = 0, 1$, is distributed as $N(\mathbf{0}, \mathbf{\Sigma_l})$, with $\mathbf{\Sigma_l}$ having generalized inverse $\mathbf{\Sigma_l^{-1}} = \frac{1}{\sigma_{bl}^2}(\mathbf{D} - \mathbf{C})$, $l = 0, 1$; $\mathbf{C} = (C_{ij})$ is often called the neighbourhood matrix and $\mathbf{D} = \text{diag}\{C_{1.}, C_{2.}, ..., C_{N.}\}$ with $C_{i.} = \sum_{j=1}^{N} C_{ij}$.

Under this model, the marginal likelihood for $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$ takes the form

$$L(\mathbf{\Theta}) = E\Big\{\prod_{i=1}^{N} f_{MS}(\mathbf{y}_i|\mathbf{Z}, \mathbf{b_0}, \mathbf{b_1})\Big\} \tag{2.6}$$

where $\mathbf{\Theta} = \{\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \sigma_{b0}^2, \sigma_{b1}^2, \{p_{I_i}\}, \{p_{M_i}\}, \boldsymbol{\alpha_0}, \boldsymbol{\alpha_1}\}$ denotes the model parameters and the expectation in (2.6) is taken with respect to the distributions of $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ and $\mathbf{Z}$.

There are two situations where the above compartmental model may be considered. Empirically, the data may suggest that several individuals never change states over time; additionally, scientific considerations may point to a need to address the presence of subgroups even if this is not empirically obvious. The mixed mover-stayer model allows for a subgroup of subjects whose initial state is absorbing. These so called 'stayers' can represent individuals who are immune to infection and will therefore be observed in the disease-free state (state 0) at all times.

2.2 Inference Procedures

In this section we outline procedures for maximum likelihood inference. As the marginal likelihood function (2.6) is analytically intractable, we develop a Monte Carlo maximum likelihood scheme.

The EM algorithm (Dempster et al., 1977) is a popular tool for conducting maximum likelihood inference in situations involving missing data. We outline maximum likelihood procedures based on a Monte Carlo implementation of the algorithm (Wei and Tanner 1990), where the random effects and latent variables are treated as missing data. In situations where the E-step of the EM algorithm does not admit a closed form, Wei and Tanner (1990), among others, proposed that the E-step can be carried

out using Monte Carlo integration resulting in what has been called the MCEM algorithm. This algorithm is useful for estimation in our model since the expectation in (2.6) involves an integral of high dimension. Alternative approaches could use analytic approximations to these integrals. For example MacNab and Dean (2000) investigate the use of penalized quasi-likelihood (Breslow and Clayton, 1993) for estimation in spatial random effect models. Such an approach, while being less computationally intensive than MCEM, can yield severely biased estimates in the case of binary data (Lin and Breslow, 1993). Previous applications of the MCEM algorithm include Chan and Ledolter (1995), who study a model for count data incorporating temporally correlated random effects, and Chan and Kuk (1997) who examine probit-linear mixed models with correlated random effects. Most recently, Zhang (2002) developed a Monte Carlo version of the EM gradient algorithm, in a geostatistical setting.

Procedures for Bayesian inference are more carefully detailed in the following section where the application is considered. To simplify the presentation we assume $p_{I_i} = p_I$ and $p_{M_i} = p_M$, $i = 1, ..., N$. Permitting variation, for example, regression modelling of the initial probabilities is easily accommodated. Note that a prime focus here is on investigating transition probabilities so we direct attention to modelling these.

Under the mixed mover-stayer model, a sufficient statistic for $\mathbf{\Theta}$ is given by $\mathbf{T} = (\mathbf{Y_0}, \mathbf{Y_1})$ where

$$\mathbf{Y_0} = \{y_i(0),\ I\{\mathbf{y}_i = \mathbf{0}\},\ \{y_i(t)|t \in D_{0i}\};\ i = 1, ..., N\} \tag{2.7}$$

and

$$\mathbf{Y_1} = \{\{y_i(t)|t \in D_{1i}\};\ i = 1, ..., N\}. \tag{2.8}$$

The marginal likelihood function for $\mathbf{Y}$, given in (2.6), can be correspondingly factorized into two terms

$$L(\mathbf{\Theta}, \mathbf{Y}) = L_0(\mathbf{\Theta_0}, \mathbf{Y_0}) \times L_1(\mathbf{\Theta_1}, \mathbf{Y_1})$$

where $\mathbf{\Theta_0} = (\boldsymbol{\beta_0}, \sigma_{b0}, p_I, p_M, \boldsymbol{\alpha_0})'$ and $\mathbf{\Theta_1} = (\boldsymbol{\beta_1}, \sigma_{b1}, \boldsymbol{\alpha_1})'$ divide the model parameters into two disjoint sets. As a result, maximum likelihood estimates can be obtained by

maximizing $L_0(\cdot)$ and $L_1(\cdot)$ separately with

$$L_1(\cdot) = E\Big\{\prod_{i=1}^{N}\prod_{t\in D_{1i}}\frac{\exp(\boldsymbol{\beta_1}'\boldsymbol{x_i(t)}+g_1(t,\boldsymbol{\alpha_1})+b_{1i})^{1-y_i(t)}}{1+\exp(\boldsymbol{\beta_1}'\boldsymbol{x_i(t)}+g_1(t,\boldsymbol{\alpha_1})+b_{1i})}\Big\} \tag{2.9}$$

where the expectation in (2.9) is taken with respect to the distribution of $\boldsymbol{b_1}$ and

$$L_0(\cdot) = E\Big\{q(\mathbf{Y_0},\boldsymbol{Z},p_I)\prod_{i=1}^{N}\prod_{t\in D_{0i}}\frac{\exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i})^{z_iy_i(t)}}{[1+\exp(\boldsymbol{\beta_0}^t\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i})]^{z_i}}\Big\} \tag{2.10}$$

where

$$q(\mathbf{Y_0},\boldsymbol{Z},p_I) = p_I^{\sum_{i=1}^N z_iy_i(0)}(1-p_I)^{\sum_{i=1}^N z_i(1-y_i(0))}\prod_{\{i|z_i=0\}} I\{\mathbf{y}_i=\mathbf{0}\}$$

and the expectation in (2.10) is taken with respect to the distributions $\boldsymbol{b_0}$ and $\boldsymbol{Z}$. Both (2.9) and (2.10) are maximized using separate MCEM algorithms; however, we note that the form of (2.10) reduces to that of (2.9) when $\boldsymbol{Z}=\mathbf{1}$ and $q(\mathbf{Y_0},\boldsymbol{Z},p_I)\equiv 1$. We therefore outline the MCEM procedure for (2.10), maximization of (2.9) being a special case.

Treating $\boldsymbol{b_0}$ and $\boldsymbol{Z}$ as missing data, the complete-data loglikelihood associated with (2.10) takes the form:

$$l_c(\boldsymbol{\Theta_0},\mathbf{Y_0},\mathbf{b_0},\mathbf{Z}) = l_c^{(1)}(p_I,\mathbf{Y_0},\mathbf{Z}) + l_c^{(2)}(p_M,\mathbf{Z})$$

$$+l_c^{(3)}(\boldsymbol{\beta_0},\boldsymbol{\alpha_0},\mathbf{Y_0},\mathbf{b_0},\mathbf{Z}) + l_c^{(4)}(\sigma_{b0},\mathbf{b_0})$$

where,

$$l_c^{(1)}(p_I,\mathbf{Y_0},\mathbf{Z}) = \log(p_I)\sum_{i=1}^{N} z_iy_i(0) + \log(1-p_I)\sum_{i=1}^{N} z_i(1-y_i(0)),$$

$$l_c^{(2)}(p_M,\mathbf{Z}) = \log(p_M)\sum_{i=1}^{N} z_i + \log(1-p_M)(N-\sum_{i=1}^{N} z_i),$$

$l_c^{(3)}(\boldsymbol{\beta_0},\boldsymbol{\alpha_0},\mathbf{Y_0},\mathbf{b_0},\mathbf{Z}) =$

$$\sum_{i=1}^{N}\sum_{t\in D_{0i}} z_iy_i(t)[\boldsymbol{\beta_0}'\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i}] - z_i\log(1+\exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i})),$$

$$l_c^{(4)}(\sigma_{b0}, \mathbf{b_0}) = -(N-1)\log(\sigma_{b0}) - \frac{\mathbf{b_0}'(D-C)\mathbf{b_0}}{2\sigma_{b0}^2}$$

Starting with initial parameter estimates, $\mathbf{\Theta_0^{(0)}}$ and setting $h = 0$, the algorithm consists of four steps:

1. *Run the Gibbs sampler and generate J realizations, $\mathbf{b_0^{(1)}}, \mathbf{Z^{(1)}}, ..., \mathbf{b_0^{(J)}}, \mathbf{Z^{(J)}}$, from the conditional distribution $f(\mathbf{b_0}, \mathbf{Z}|\mathbf{Y_0}, \mathbf{\Theta_0^{(h)}})$*

2. *Calculate $Q(\mathbf{\Theta_0}; \mathbf{\Theta_0^{(h)}}) = \frac{1}{J}\sum_{k=1}^{J} l_c(\mathbf{\Theta_0}, \mathbf{Y_0}, \mathbf{b_0^{(k)}}, \mathbf{Z^{(k)}})$*

3. *Maximize $Q(\mathbf{\Theta_0}; \mathbf{\Theta_0^{(h)}})$ over $\mathbf{\Theta_0}$ to obtain $\mathbf{\Theta_0^{(h+1)}}$*

4. *Assess convergence. If convergence has been achieved then stop. Else Set $h = h+1$ and go to step 1*

In implementing this algorithm, two important issues arise: first is the choice of Monte Carlo sample size, J, to be used at each iteration and second is monitoring convergence of the algorithm. Wei and Tanner (1990) suggest using small values of J in the initial stages of the algorithm and increasing J as the algorithm moves closer to convergence. Regarding convergence, they recommend plotting estimates at each iteration of the algorithm. Convergence is then indicated by the stabilization of the process with random fluctuations about some fixed value.

Gibbs sampling at the $(h+1)^{st}$ iteration requires simulation from the full conditional distributions, $[b_{0i}|\mathbf{Y_0}, \mathbf{b_0^{(-i)}}, \boldsymbol{Z}, \mathbf{\Theta_0^{(h)}}]$ and $[z_i|\mathbf{Y_0}, \mathbf{Z^{(-i)}}, \boldsymbol{b_0}, \mathbf{\Theta_0^{(h)}}]$, $i = 1, ..., N$. These full conditional distributions are given by $f_i(b_{0i}|\mathbf{Y_0}, \mathbf{b_0^{(-i)}}, \boldsymbol{Z}, \mathbf{\Theta_0^{(h)}})$

$$\propto \begin{cases} \exp\left(-\frac{C_{i.}(b_{0i} - \frac{1}{C_{i.}}\sum_{j=1}^{n} C_{ij}b_{0j})^2}{2\sigma_{b0}^{(h)2}}\right) & \text{if } z_i = 0, \\ \exp\left(-\frac{C_{i.}(b_{0i} - \frac{1}{C_{i.}}\sum_{j=1}^{n} C_{ij}b_{0j})^2}{2\sigma_{b0}^{(h)2}}\right) \prod_{t \in D_{0_i}} \frac{\exp(\beta_0^{(h)} x_i(t) + g_0(t, \alpha_0^{(h)}) + b_{0i})^{y_i(t)}}{1 + \exp(\beta_0^{(h)} x_i(t) + g_0(t, \alpha_0^{(h)}) + b_{0i})} & \text{if } z_i = 1. \end{cases} \tag{2.11}$$

and $[z_i|\mathbf{Y_0}, \mathbf{Z^{(-i)}}, \boldsymbol{b_0}, \mathbf{\Theta_0^{(h)}}] \sim \text{Bernoulli}(p_{z_i})$ with

$$p_{z_i} = \begin{cases} 1 & \text{if } \mathbf{y}_i \neq \mathbf{0}, \\ \frac{p_M^{(h)}(1-p_I^{(h)}) \prod_{t=1}^{n_i - 1}[1 + \exp(\beta_0^{(h)} x_i(t) + g_0(t, \alpha_0^{(h)}) + b_{0i})]^{-1}}{(1-p_M^{(h)}) + p_M^{(h)}(1-p_I^{(h)}) \prod_{t=1}^{n_i - 1}[1 + \exp(\beta_0^{(h)} x_i(t) + g_0(t, \alpha_0^{(h)}) + b_{0i})]^{-1}} & \text{if } \mathbf{y}_i = \mathbf{0} \end{cases} \tag{2.12}$$

To sample from (2.11) we generate a Gaussian deviate if $z_i = 0$ or, if $z_i = 1$, realizations are obtained using rejection sampling (Ripley, 1987). Since the joint distribution of $\boldsymbol{b_0}$, specified conditionally via (2.5), is improper, we impose an identifying sum-to-zero constraint in our Gibbs sampler by recentering the simulated values of $\boldsymbol{b_0}$ around zero after each iteration.

Having generated values, $\mathbf{b_0^{(1)}}, \mathbf{Z^{(1)}}, ..., \mathbf{b_0^{(J)}}, \mathbf{Z^{(J)}}$, using the Gibbs sampler, the updated parameter estimates for p_I and p_M are given by: $p_I^{(h+1)} = \frac{\sum_{i=1}^{N} y_i(0)\bar{z_i}}{\sum_{i=1}^{N} \bar{z_i}}$ and $p_M^{(h+1)} = \frac{\sum_{i=1}^{N} \bar{z_i}}{N}$ where $\bar{z_i} = \frac{1}{J}\sum_{k=1}^{J} z_i^{(k)}$. To obtain $\boldsymbol{\beta_0}^{(h+1)}$ and $\boldsymbol{\alpha_0}^{(h+1)}$ we maximize the following objective function:

$$f(\boldsymbol{\beta_0}, \boldsymbol{\alpha_0}) = \sum_{i=1}^{N} \sum_{t \in D_{0i}} \bar{z_i} y_i(t)[\boldsymbol{\beta_0}' \boldsymbol{x_i(t)} + g_0(t, \boldsymbol{\alpha_0})]$$

$$-\frac{1}{J}\sum_{k=1}^{J}\sum_{i=1}^{N}\sum_{t \in D_{0i}} z_i^{(k)} \log(1 + \exp(\boldsymbol{\beta_0}' \boldsymbol{x_i(t)} + b_{0i}^{(k)} + g_0(t, \boldsymbol{\alpha_0})))$$

Numerical maximization is accomplished using a standard quasi-Newton routine (Fletcher 1987). The variance component is updated via: $\sigma_{b0}^{(h+1)} = [\frac{1}{(N-1)J}\sum_{k=1}^{J} \mathbf{b_0^{(k)\prime}}(D - C)\mathbf{b_0^{(k)}}]^{\frac{1}{2}}$.

2.2.1 Observed Information Matrix

Following Chan and Ledolter (1995), Chan and Kuk (1997) and Zhang (2002) we obtain standard errors using a Monte Carlo approximation of the observed information matrix. We describe the technique in a general setting. Let $\mathbf{Y}$ denote the observed data and $\mathbf{r}$ denote a vector of unobserved stochastic quantities (for example, random effects) which we treat as missing data in an EM framework to obtain maximum likelihood estimates of some parameter $\boldsymbol{\Theta}$. We let $l(\boldsymbol{\Theta}, \mathbf{Y})$ denote the observed data log-likelihood and $l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})$ the corresponding complete data log-likelihood. From Louis (1982) we have

$$\frac{\partial l(\boldsymbol{\Theta}, \mathbf{Y})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'} = E[\frac{\partial l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'}|\mathbf{Y}] + E[\frac{\partial l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})}{\partial \boldsymbol{\Theta}} \frac{\partial l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})'}{\partial \boldsymbol{\Theta}}|\mathbf{Y}]$$
$$-E[\frac{\partial l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})}{\partial \boldsymbol{\Theta}}|\mathbf{Y}]E[\frac{\partial l_c(\boldsymbol{\Theta}, \mathbf{Y}, \mathbf{r})}{\partial \boldsymbol{\Theta}}|\mathbf{Y}]' \qquad (2.13)$$

Having obtained an estimate, $\hat{\mathbf{\Theta}}$, from our MCEM algorithm, we run one final Gibbs sampler and obtain a set of realizations, $\mathbf{r^{(1)}}, ..., \mathbf{r^{(J)}}$, from the conditional distribution with density $f(\mathbf{r}|\mathbf{Y}, \hat{\mathbf{\Theta}})$. We then approximate the conditional expectations in (2.13) using ergodic averages and obtain an estimate of the observed information matrix evaluated at $\hat{\mathbf{\Theta}}$.

$$-\frac{\partial l(\hat{\mathbf{\Theta}}, \mathbf{Y})}{\partial \mathbf{\Theta} \partial \mathbf{\Theta}'} \approx -\frac{1}{J}\sum_{k=1}^{J} \frac{\partial l_c(\hat{\mathbf{\Theta}}, \mathbf{Y}, \mathbf{r}^{(k)})}{\partial \mathbf{\Theta} \partial \mathbf{\Theta}'} - \frac{1}{J}\sum_{k=1}^{J} \frac{\partial l_c(\hat{\mathbf{\Theta}}, \mathbf{Y}, \mathbf{r}^{(k)})}{\partial \mathbf{\Theta}} \frac{\partial l_c'(\hat{\mathbf{\Theta}}, \mathbf{Y}, \mathbf{r}^{(k)})}{\partial \mathbf{\Theta}}$$

$$+\{\frac{1}{J}\sum_{k=1}^{J} \frac{\partial l_c(\hat{\mathbf{\Theta}}, \mathbf{Y}, \mathbf{r}^{(k)})}{\partial \mathbf{\Theta}}\}\{\frac{1}{J}\sum_{k=1}^{J} \frac{\partial l_c(\hat{\mathbf{\Theta}}, \mathbf{Y}, \mathbf{r}^{(k)})}{\partial \mathbf{\Theta}}\}' \tag{2.14}$$

We apply this technique to the complete data log-likelihood functions associated with both (2.9) and (2.10) to obtain standard errors. In our analysis, the Monte Carlo sample size, J, is increased until the standard errors become stable to a desired number of decimal places.

2.3 Study of Weevil Infestation

The weevil infestation data were obtained over a seven year period beginning in 1996 and ending in 2002. The study region is a plantation in British Columbia covering an area of $21,960m^2$ and containing a population of $N = 2662$ trees susceptible to weevil attack. The positions of trees within the plantation are depicted in Figure 2.1. Each tree was inspected for the presence of weevil attack in the fall. In any given year, each tree is therefore classified into one of two states, either weevil-infected or not. The event history of the i^{th} tree is represented with a binary response vector $\mathbf{y}_i = (y_i(0), ..., y_i(6))$ where $y_i(t)$ indicates infection at year t. The purpose of the analysis is to provide a spatio-temporal description of the transition process between the infected and uninfected states.

For each year following 1996, the proportion of infected trees, conditioning on the infection status of the previous year, are shown in Figure 2.2. Two key features are evident. First, it seems that the state occupied at time t depends on the state of time

$t-1$ since, in each year, the raw estimates indicate

$$\Pr(\text{infected at } t|\text{infected at } t-1) > \Pr(\text{infected at } t|\text{not infected at } t-1).$$

Second, the process governing transitions seems to be inhomogeneous over time. It was hypothesized by ecologists that some pine trees may be resistant to weevil attack. This seemed reflected in our study where 27 percent of the 2662 trees under investigation were not infected during the 7 years of observation. Such resistant trees are accommodated by our spatial mover-stayer model which allows for a group of 'stayers' in the uninfected state. Conversely, only 1 tree had been infected every year; therefore, a model accounting for 'stayers' in the infected state seems unnecessary.

To investigate the possibility and scale of spatial correlation in transitions we fit a preliminary model where each $\mathbf{y}_i$ is assumed to be independently drawn from a two-state Markov chain incorporating independent random effects with transition probabilities given by $\text{logit}(p_{01_i}(t)) = g_0(t, \boldsymbol{\alpha_0}) + u_{0i}$ and $\text{logit}(p_{10_i}(t)) = g_1(t, \boldsymbol{\alpha_1}) + u_{1i}$, $i = 1, ..., 2662,\ t = 1, ..., 6$ with $u_{li} \overset{iid}{\sim} N(0, \sigma_l^2)\ l = 0, 1$. The terms $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ account for temporal variation and are modelled using cubic B-splines with one inner knot placed at $t = 4$ years. The random effects, $\boldsymbol{u_1} = (u_{1_1}, ..., u_{1_{2662}})$ and $\boldsymbol{u_0} = (u_{0_1}, ..., u_{0_{2662}})$, account for tree-to-tree variation in transitions. Figure 2.3 displays the fitted temporal trends $g_0(t, \hat{\boldsymbol{\alpha}}_\mathbf{0})$, $g_1(t, \hat{\boldsymbol{\alpha}}_\mathbf{1})$ as well as the empirical semivariograms of $\hat{\boldsymbol{u}}_\mathbf{1}$ and $\hat{\boldsymbol{u}}_\mathbf{0}$ based on a bin size of 1 meter. Examining the semivariograms, there appears to be spatial correlation in both $\hat{\boldsymbol{u}}_\mathbf{1}$ and $\hat{\boldsymbol{u}}_\mathbf{0}$ at the smaller distances, each having a range of about 5 to 10 meters. Dependence of this scale in our spatial mixed mover-stayer model is accommodated by setting the weights associated with the spatial random effects defined in (2.5) to $C_{ij} = I\{d(i,j) \leq 10m\}$ where $d(i,j)$ denotes the distance between trees i and j. In addition, we define spatially varying covariates $D_i = \sum_{j \neq i} C_{ij}$, a local measure of tree density and $A_i(t) = \sum_{j \neq i} C_{ij} y_j(t)$, a local measure of infection density at year t. The regression specifications are then

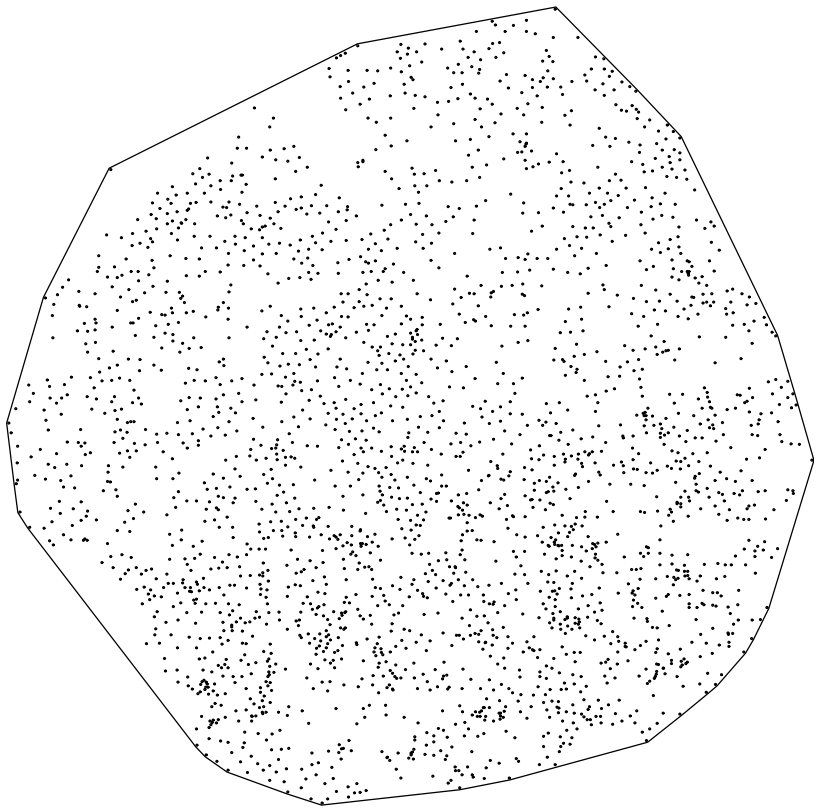


Figure 2.1: Positions of trees within the plantation. The boundary is taken to be the convex hull of these positions.

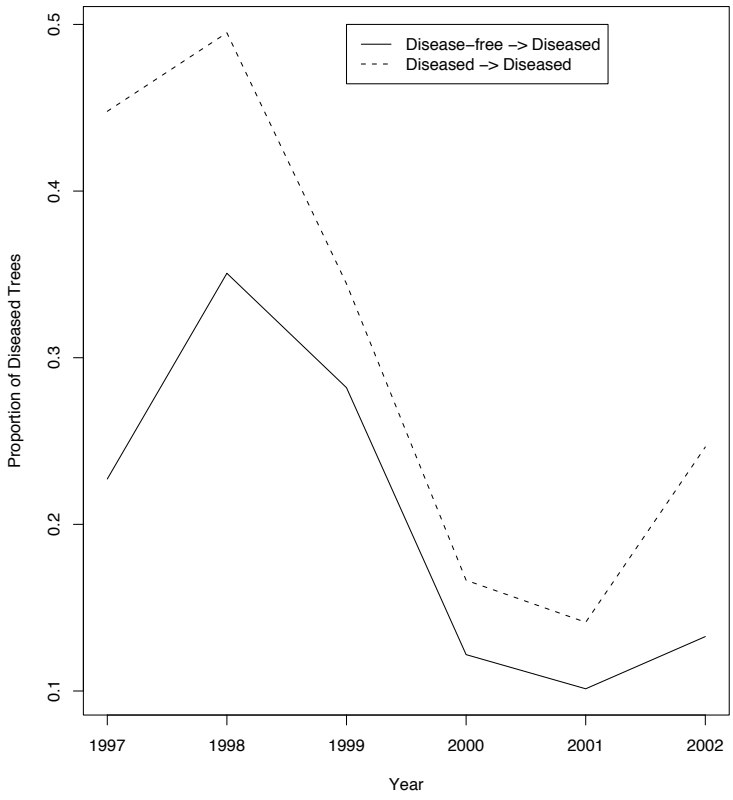


Figure 2.2: Raw estimates of the conditional probabilities of infection in each year.

given by

$$\text{logit}(p_{01_i}(t)) = \beta_{0_1} D_i + \beta_{0_2} A_i(t-1) + g_0(t, \boldsymbol{\alpha_0}) + b_{0i}$$

$$\text{logit}(p_{10_i}(t)) = \beta_{1_1} D_i + \beta_{1_2} A_i(t-1) + g_1(t, \boldsymbol{\alpha_1}) + b_{1i}$$

$$i = 1, ..., 2662, \ t = 1, ..., 6$$

where $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ are cubic splines with $K_l = 1$, $l = 0, 1$, and an inner knot at $t = 4$ years as in the preliminary analysis.

Starting values for $\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_0}$, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_0}$ in the MCEM algorithm were based on estimates from simpler models incorporating no spatial dependence. Starting values for p_I and p_M were based on the raw proportions obtained from the data. A burn-in of 1000 samples was used at each iteration of the algorithm. The Monte Carlo sample size was set to $J = 500$ for the first 100 iterations and then increased to $J = 1000$ until trace plots of the MCEM estimates indicate convergence had been achieved. The resulting trace plots of MCEM estimates are shown for several parameters in Figure 2.4. Upon convergence, an additional 100 iterations were run with $J = 5000$ to reduce Monte Carlo error.

For the purpose of comparison, a Bayesian analysis was also conducted. In this setting, the model specification is made complete by assigning a prior distribution to $\boldsymbol{\Theta}$. The resulting posterior distribution can be summarized using MCMC sampling. A program for drawing MCMC samples from the posterior distribution has been implemented in WinBUGS 1.4. which is freely available from the Medical Research Council Biostatistics Unit in Cambridge in the U.K. (www.mrc-bsu.cam.ac.uk/bugs). We employed weakly informative prior distributions for the model parameters. All regression coefficients and parameters of the cubic spline terms were assigned $N(0, 10^3)$ priors, except for the intercepts α_{0_0} and α_{1_0} which were assigned Uniform$(-\infty, \infty)$ distributions. Variance components were assigned Uniform$(0, 10)$ priors and Uniform$(0, 1)$ prior distributions were assigned to p_M and p_I. Sensitivity with respect to these prior distributions was assessed by comparisons from repeating the analysis with other weakly informative prior specifications. This comparison indicated results to be fairly robust over the forms of prior considered. Markov Chain Monte Carlo sampling was based on two chains run in parallel and convergence to the posterior distribution was

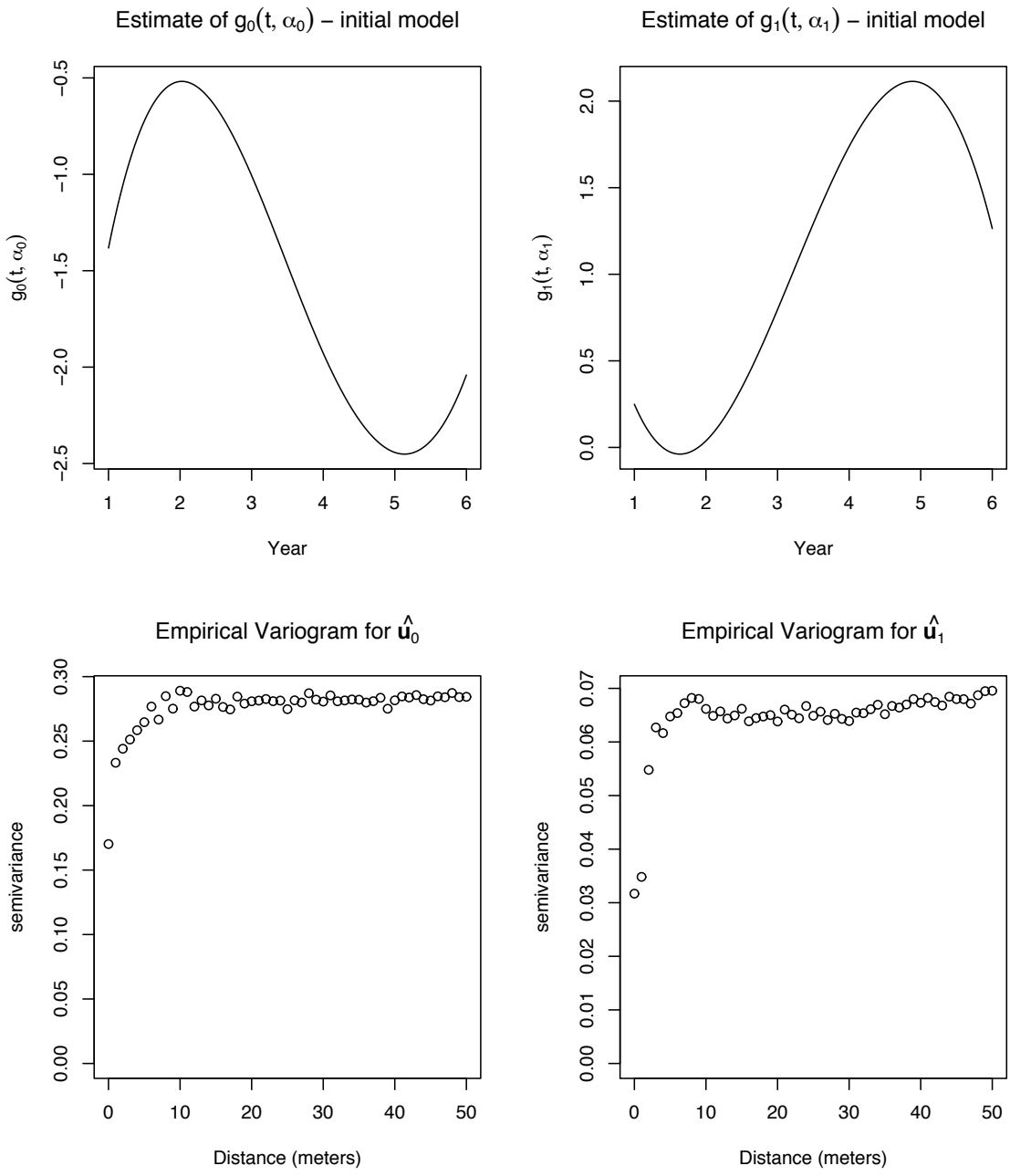


Figure 2.3: Estimates from initial model exploring spatial scale and spatial correlation.

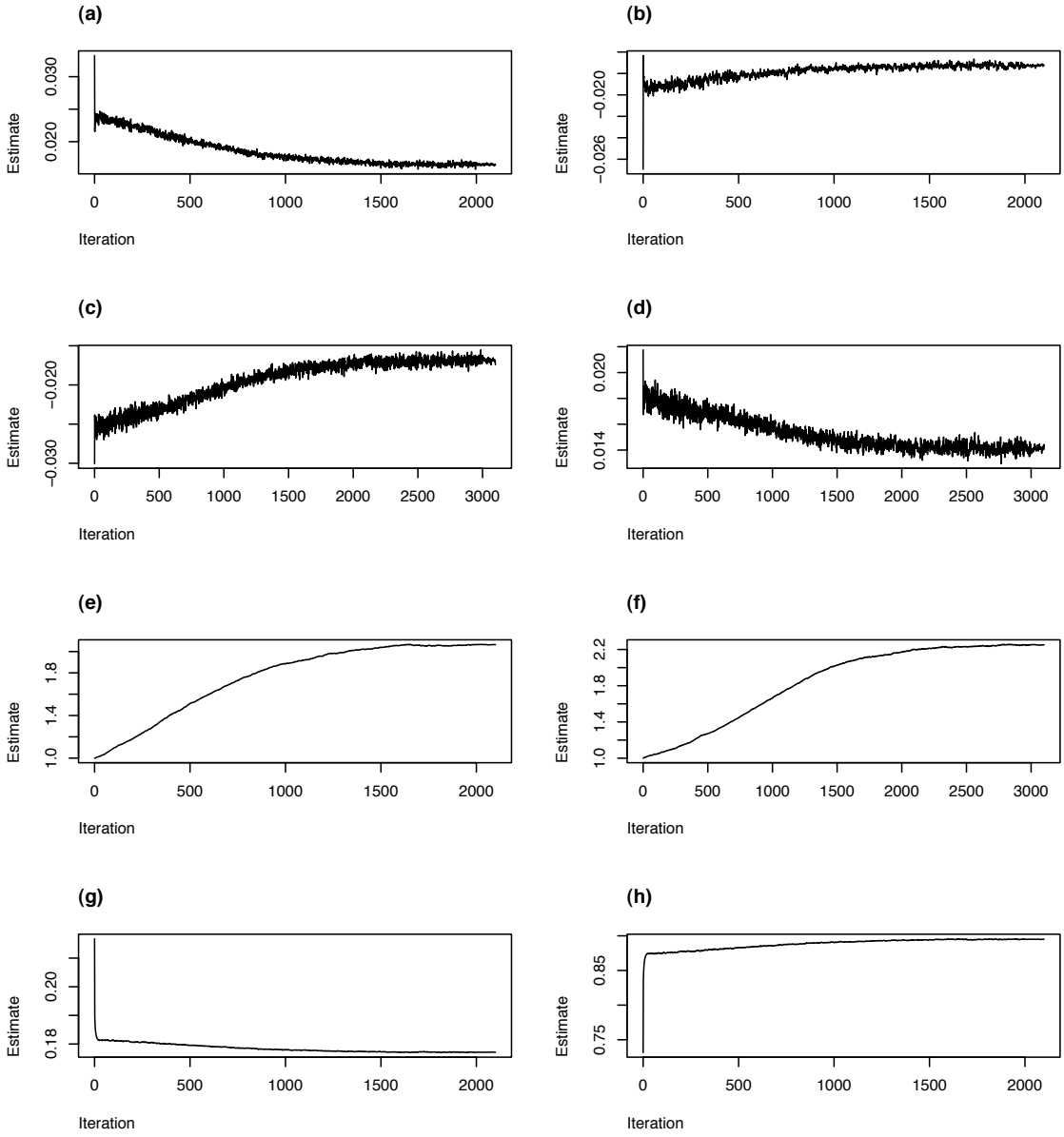


Figure 2.4: Monte Carlo EM estimates by iteration for selected parameters in the spatial mover-stayer model (a) β_{0_1}; (b) β_{0_2}; (c) β_{1_1}; (d)β_{1_2}; (e) σ_{b_0}; (f) σ_{b_1}; (g) P_I; (h) P_M.

assessed through the examination of sample trace plots and Gelman-Rubin (1992) statistics. A burn-in of 20,000 iterations was used with an additional 80,000 iterations from each chain retained for posterior summarization.

Parameter estimates of the regression coefficients and variance components are presented in Table 2.1. The Bayesian and maximum likelihood estimates are very similar. From the estimates of p_M, we see that roughly 10 percent of the trees under study are estimated to be resistant to weevil infection. Regarding covariate effects, local tree density, D_i, seems to have an effect on both types of transitions. Trees located in denser regions tend to have lower probability of transition into the infected state and higher probability of transition out of the infected state. Local attack density, $A_i(t-1)$, appears to affect transitions out of the uninfected state such that trees surrounded by a larger number of infected trees at time $t-1$ have a higher probability of making a transition from the uninfected to the infected state at time t.

	Uninfected → Infected		Infected → Uninfected	
Parameter	MLE (SE)	Bayes (Sd)	MLE (SE)	Bayes (Sd)
Intercept	-0.505 (0.142)	-0.496 (0.132)	-0.189 (0.195)	-0.169 (0.198)
D_i	-0.017 (0.003)	-0.018 (0.003)	0.014 (0.005)	0.014 (0.005)
$A_i(t-1)$	0.016 (0.008)	0.017 (0.008)	-0.0174 (0.011)	-0.015 (0.011)
σ_{b0}/σ_{b1}	2.064 (0.561)	1.98 (0.340)	2.25 (0.449)	2.619 (0.494)
p_I	0.177 (0.008)	0.178 (0.009)	-	-
p_M	0.895 (0.018)	0.894 (0.021)	-	-

Table 2.1: Parameter estimates for the spatial mover-stayer model. Bayes estimates are posterior means and standard deviations. Maximum likelihood estimates are obtained using the Monte Carlo EM algorithm.

Estimates of the temporal trends, $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$, are shown in Figure 2.5 and correspond with the patterns observed in the preliminary analysis. Posterior mean (from the Bayes analysis) and empirical Bayes (from the likelihood analysis) estimates of the random effects $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ are also obtained. To gain insight into tree-specific residual risks of transitions in and out of the infected state we have marked, in Figure 2.6, the locations within the plantation associated with the 100 largest and

smallest of these estimates. These maps of unmodeled heterogeneity can be useful for postulating additional covariates missing from the current model. Spatial clustering of the largest and smallest values is apparent. For $\boldsymbol{b_0}$, the largest estimates appear to be clustered in the far north of the plantation; whereas, the smallest estimates occur in two clusters lying on the boundary of the region, one located in the far west and the other occurring in the south. For $\boldsymbol{b_1}$, the largest estimates occur towards the south and south-east and the smallest estimates occur towards the far north and north-west. In Figure 2.7 we have located those 100 trees having the highest estimated posterior probability of resistance, $Pr(z_i = 0|\mathbf{Y})$, the majority of these being located in the mid-north section of the plantation. Maps identifying such trees can be useful in determining the characteristics which may lead to resistance.

2.4 Discussion

Some alternatives to our modelling above should be mentioned. First, other forms for the weights, C_{ij}, in the IAR model (2.5) could be entertained. Our choice of $C_{ij} = I\{d(i,j) \leq 10m\}$ could be expanded to forms such as $C_{ij} = I\{d(i,j) \leq w\} \times f(d(i,j))$ where $f(\cdot)$ is a positive, decreasing function of distance and $w > 0$ is an unknown parameter. The flexible neighbourhood structures proposed by Conlon and Waller (1998) may also be fruitfully employed here. As discussed by Besag and Kooperberg (1995) and Waller and Gotway (2005) the relationship between the neighbour weights C_{ij} in the IAR model and the spatial covariance range suggested by the semivariograms in Figure 2.3 is complex and the adequacy of determining the weights based on the semivariogram range will be investigated in future work. As a second alternative, one might consider Gaussian geostatistical models for the random effects. Indeed, examination of the empirical semivariograms in Figure 2.3 suggests a spherical correlation function might be employed (Cressie, 1993). Implementation of geostatistical covariance structures would increase the computational burden involved in model fitting whereas Markov random field models are ideally suited for computation within a Gibbs sampler. In addition, geostatistical models are typically used when prediction of random effects at unsampled sites is of interest which was not the case in our

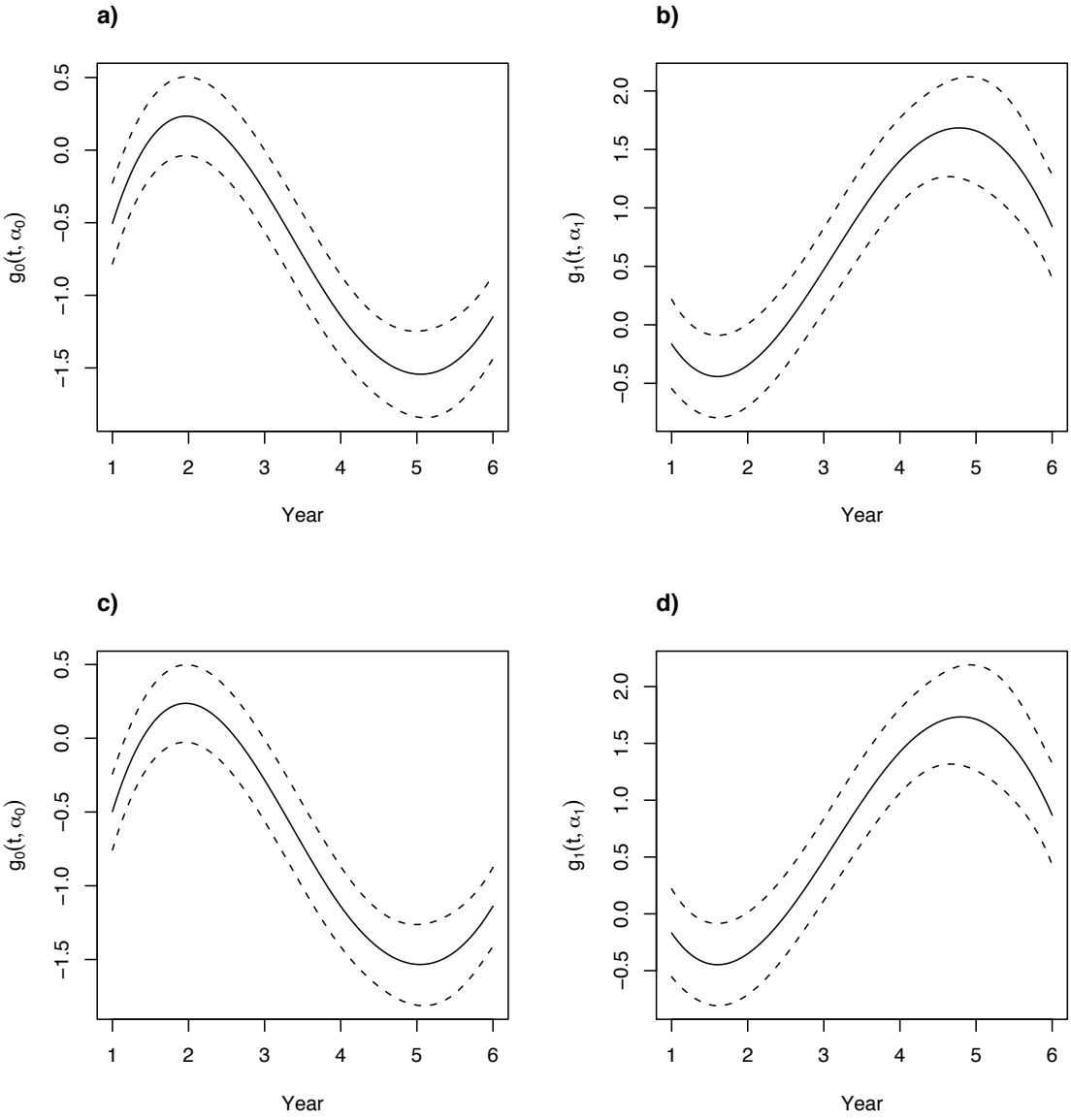


Figure 2.5: Maximum likelihood estimates of temporal trends from the spatial mover-stayer model with 95% confidence intervals (a) $g_0(t, \boldsymbol{\alpha_0})$; (b) $g_1(t, \boldsymbol{\alpha_1})$. Bayesian (posterior mean) estimates of temporal trends with 95% credible sets (c) $g_0(t, \boldsymbol{\alpha_0})$; (d) $g_1(t, \boldsymbol{\alpha_1})$.

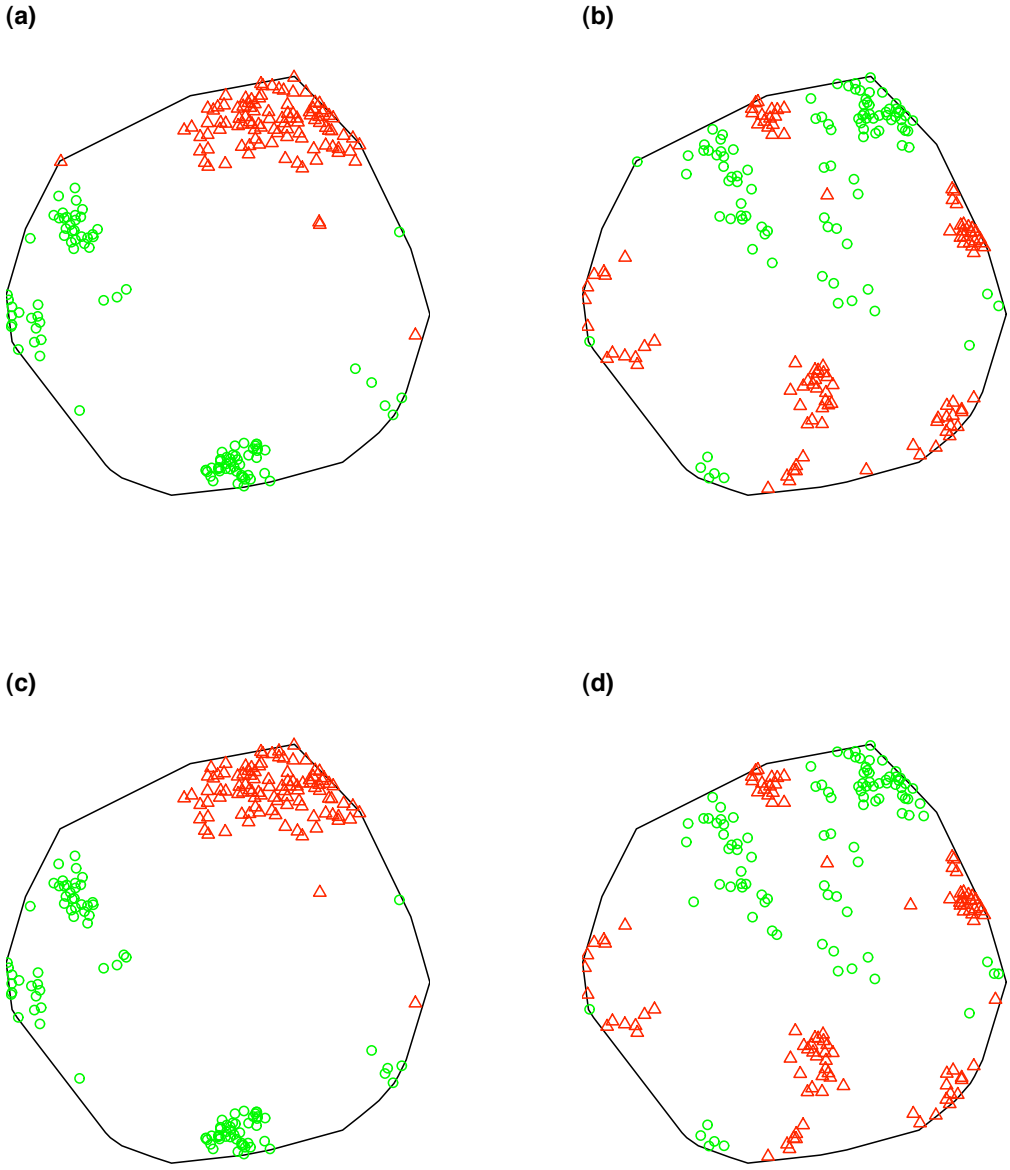


Figure 2.6: Locations of the 100 largest (triangles) and smallest (circles) estimated random effects from the spatial mover-stayer model (a) $\boldsymbol{b_0}$ - Likelihood Analysis; (b) $\boldsymbol{b_1}$ - Likelihood Analysis; (c) $\boldsymbol{b_0}$ - Bayesian Analysis; (d) $\boldsymbol{b_1}$ - Bayesian Analysis.

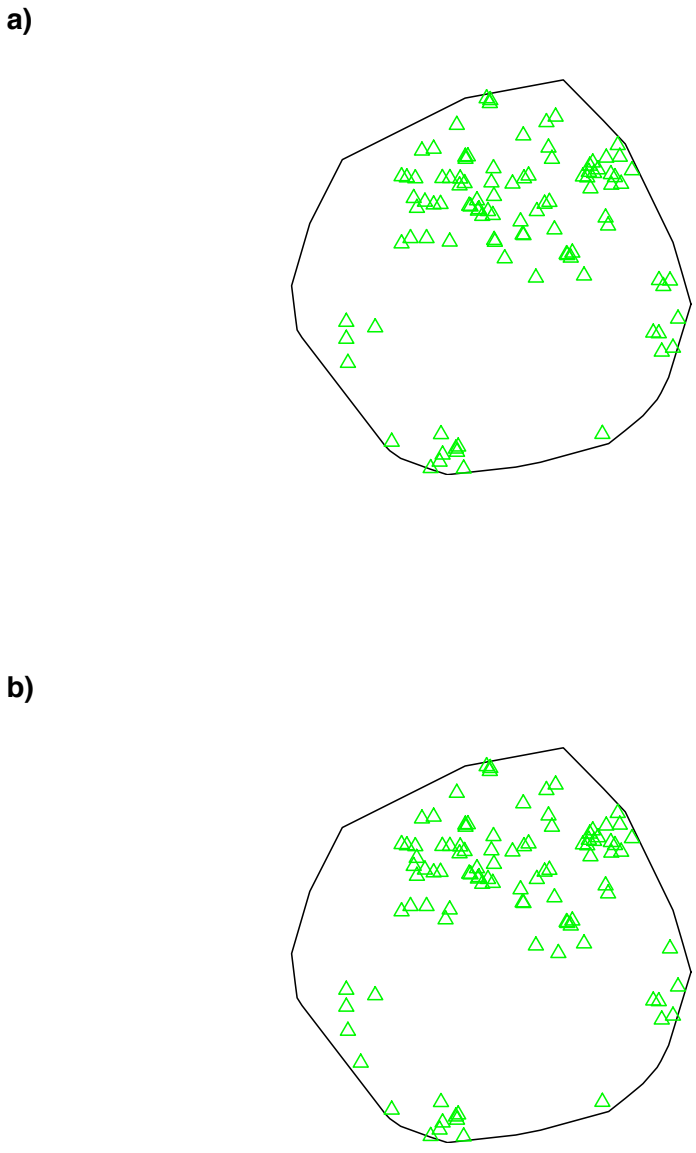


Figure 2.7: Locations of the 100 trees having the highest estimated posterior probability of resistance (a) Likelihood Analysis - $Pr(z_i = 0|\mathbf{Y}, \hat{\Theta})$; (b) Bayesian Analysis - $Pr(z_i = 0|\mathbf{Y})$.

analysis.

More flexible correlation structures for the random effects could be considered. In particular, the assumption of independence between $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ could be relaxed allowing for joint modelling of the spatial random effects associated with the two types of transitions. Such modelling will allow for the borrowing of information across spatial units as well as across different transition types within units. The required joint correlation structures can be accommodated using multivariate Markov random field models which have received some recent attention in the literature (Kim et al., 2001; Gelfand and Vounatsou, 2003; Carlin and Banerjee, 2002). Another useful extension would allow for spatial correlation in the mixture allocation variables, $\mathbf{Z}$, allowing for spatial clusters of resistant trees. In our application resistance was thought to be the result of genetic factors and spatial correlation in resistance was not expected. In other cases; however, such correlation might be postulated and would represent latent environmental factors contributing to resistance. Along these lines, Wu and Huffer (1997) and Huffer and Wu (1998) considered spatial autologistic models for describing the distribution of plant species. Their approach might be incorporated into our model by assuming that the latent variables, $\mathbf{Z}$, are drawn from an autologistic model.

Regarding the state structure, extended models allowing for more than two states are easily conceived. These extended models might prove useful in applications of infectious disease modelling where models typically incorporate three states: (1) a state representing individuals who are uninfected but susceptible to infection (2) a state representing individuals who are currently infected and infectious and (3) a state representing individuals who have recovered with lifelong immunity. Such models may be employed for examining infectious disease in animal studies; however, these applications require care for example when modelling spatial correlation due to the possibility of migration or in handling the mechanisms of the infectious transmission process.

Related models for continuous time data where spatially correlated random effects are incorporated into the transition intensities of a multi-state Markov process are reported in the next chapter. Here, joint modelling of both spatial correlation as well as correlation between transition rates is required and multivariate spatial approaches

need be employed.

Chapter 3

Continuous-Time Spatial Multi-State Processes

Clinical trials and medical studies are often concerned with the event histories of individuals in a population. In such studies, subjects are observed over time and significant events or changes in their health status are recorded. Statistical analysis of the resulting longitudinal data is conveniently conducted within the multi-state modelling framework. In this setting, we model a stochastic process, which at any point in time, will occupy one of a discrete set of states and interest centers on the transition process between states. For example, states may refer to the number of recurrences of an event, or the stages of a disease. Multi-state modelling provides a broad framework for the analysis of longitudinal data which encompasses, as a special case, methods for the analysis of censored event-time data including methods for the analysis of competing risks.

Most commonly employed are the Markov models which have found wide application in many fields including biostatistics, demography and the social sciences (see e.g. Clayton, 1988; Hougaard, 2000). In many situations, it is useful to account for heterogeneity in the state-to-state transition processes corresponding to different subjects. In the simplest of such cases, this heterogeneity can be explained entirely through covariates and Markov chain regression models are employed (Muenz and

Rubenstein 1985, Prentice et al. 1981). In cases where the extra-variation is not adequately explained by the available covariates, random effects can be introduced in a hierarchical setting. Several authors have examined such mixed-effects models for the analysis of heterogeneous multi-state data. Aalen (1982) explored mixing distributions on a Markov process in a general setting. More recently, Cook and Ng (1997), Ng and Cook (1997) and Albert and Waclawiw (1998) developed two-state mixed Markov models incorporating subject specific random effects. In these models, the focus has been accounting for within-subject heterogeneity and subjects were assumed independent. Extensions to handle bivariate two-state processes have been recently considered by Zeng and Cook (2004) in context of paired data arising from studies examining visual acuity. We consider a multivariate setting where subjects are clustered into geographical strata such as provincial districts or local health units. In this case, the assumptions of between-subject or between-cluster independence of processes may not be viable as subjects arising from the same or neighbouring strata may yield correlated outcomes. Such correlation can be attributed to latent, spatially-varying factors such as shared environmental influences. A thorough analysis must therefore address the possibility of spatial correlation in the longitudinal outcomes of different subjects in a broad multivariate setting and this is a fundamental focus of our work.

Generalized linear mixed models incorporating spatially correlated random effects have been studied extensively in disease mapping applications (see e.g., Besag et al., 1991, Bernardinelli and Montomoli, 1992, Best et al., 1999) where data typically take the form of spatially correlated counts or proportions. In a similar vein, several authors have recently developed random effect models for the analysis of spatially correlated survival data (Li and Ryan 2002, Henderson et al. 2002, Carlin and Banerjee 2002, Banerjee et al. 2003). In the more general multi-state setting, spatial models have received considerably less attention. A two-state model of this sort was developed in Chapter 2 in a discrete-time setting. There, a mixed mover-stayer model was developed for describing a spatio-temporal recurring disease process; however, more general continuous-time spatial multi-state models have not been considered previously and would serve usefully in many situations.

In this chapter we present spatial continuous-time multi-state models for the analysis of geographically referenced event history data. Mixed Markov transitional models are developed. Transition intensity functions are modelled through a proportional intensities frailty model that incorporates covariates and correlated random effects. Baseline transition intensity functions are modelled using both parametric Weibull forms as well as flexible representations based on cubic B-splines. Here, modelling of both spatial correlation across regions as well as correlation between different transition rates is desirable and a joint spatial model is required. In particular, we employ a multivariate Gaussian Markov random field as a prior distribution for the random effects operating across different transition rates. The use of a joint spatial model in this context is particularly novel as it allows for the borrowing of information, not only spatially, but also across different transition rates. When certain transitions correspond to rare events, the resulting joint correlation structure enables information to be obtained from other related events.

Our work is motivated by a study of invasive cardiac procedure in Quebec, in which patients hospitalized for acute coronary syndrome were followed over time and information regarding subsequent hospital readmissions and mortality was obtained. Additional demographic and treatment information was also obtained from each patient along with the local health unit (LHU) in which the subject resides. The local health units serve as the geographical strata in this study. Interest lies in the identification of spatial heterogeneity in the disease profile of subjects across the various local health units of the province. Our analysis reveals interesting trends in this heterogeneity.

In Section 3.1 we specify our models. We discuss joint spatial modelling of the random effects as well as two alternative representations for the baseline intensity functions. Model fitting and inference using Markov Chain Monte Carlo (MCMC) is also discussed. Section 3.2 presents an analysis of the Quebec cardiac study data. In Section 3.3 we summarize our findings and discuss future work.

3.1 Spatial Continuous-Time Multi-State Models

Suppose there are N subjects followed over time and each subject is located in one of J regions. Associated with the l^{th} subject is a continuous-time stochastic process, $Y_l(t),\ t \geq 0$, which takes values in a finite state-space, $S = \{1, ..., k\}$ and which we observe over a finite time interval $[0, C_l]$. In addition, we assume that associated with the l^{th} subject, $l = 1, ..., N$, there are two vectors of covariates $\mathbf{z_l}$ and $\mathbf{r_l}$ each described below.

We further assume that all N processes have the same state structure specifying which state-to-state transitions are allowable. Depending on the application, the state structure may not allow the maximum of $k(k-1)$ possible state-to-state transitions. This occurs, for example, when some states are absorbing or if the state structure is progressive. For any particular state structure, we define the set $T = \{(i,j) \in S \times S | Pr(i \to j) > 0\}$, so that T contains all ordered pairs, (i,j), $i \neq j$, corresponding to $i \to j$ transitions which are allowed by the state structure. We further assume that the elements (i,j) of T are sorted in ascending order in terms of the first index, i, and then by the second index, j, and we let N_T denote the cardinality of T which is the number of allowable transitions. As an example, the state structure employed for the modelling in our application is depicted in Figure 3.1. Here, there are $k = 4$ states, with $T = \{(1,2), (1,4), (2,3), (2,4), (3,4))\}$ and $N_T = 5$ allowable transitions.

To model the processes corresponding to all subjects, we work in a hierarchical Bayes framework. At the first level of the model, we assume that the processes, $Y_l(t),\ l = 1, ..., N$, are independent with each following a continuous-time Markov chain with state-space S and transition intensity functions $\{q_{lij}(t),\ (i,j) \in T\}$. To model the intensity functions, we assume that covariates and spatial terms have a multiplicative effect given by

$$q_{lij}(t) = q_{0ij}(t) \exp(\boldsymbol{\beta_{ij}}' \mathbf{z_l} + \boldsymbol{b_{ij}}' \mathbf{r_l}),\ \ (i,j) \in T \tag{3.1}$$

where, for each $(i,j) \in T$, $\mathbf{z_l}$ is a vector of covariates corresponding to fixed effects $\boldsymbol{\beta_{ij}}$ (including an intercept); $\mathbf{r_l}$ is a vector of covariates corresponding to random effects $\boldsymbol{b_{ij}}$; and $q_{0ij}(t)$ is a baseline transition intensity function common to all subjects. This

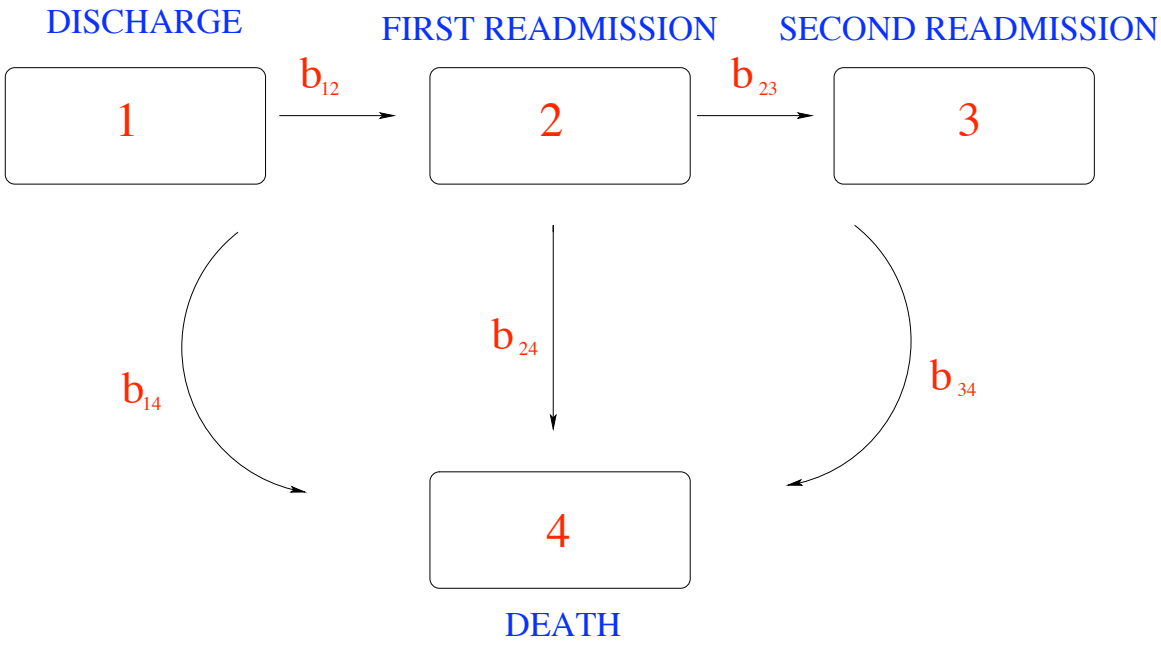


Figure 3.1: State structure for the Quebec cardiac study.

mixed transitional model is general and can be applied usefully in many situations where clustered multi-state processes are observed and there exists correlation both within and between clusters. This broad framework is similar to that adopted by Breslow and Clayton (1993) in their development of generalized linear mixed models. For spatial modelling in our application, we assume that $\boldsymbol{b_{ij}} = (b_{ij_1}, ..., b_{ij_J})'$ is a vector of region specific random effects accounting for spatial correlation in transitions from state i to state j. The vector of covariates $\mathbf{r_l}$ is taken to be a binary vector of length J, indicating the region associated with subject l and is defined by: $r_{l_i} = I\{\text{Subject } l \text{ is located in region } i\}$, $i = 1, ..., J$ with $I\{\cdot\}$ denoting the indicator function.

We consider two alternative representations for modelling the baseline intensity functions. The first allows for flexible forms using cubic B-splines. The B-spline

representation is given by

$$\log(q_{0ij}(t)) = \sum_{h=1}^{k_{ij}+3} \alpha_{ij_h} B_{ij_h}(t), \;\; (i,j) \in T \tag{3.2}$$

where, for each $(i,j) \in T$, $\boldsymbol{\alpha_{ij}} = (\alpha_{ij_1}, ..., \alpha_{ij_{k_{ij}+3}})'$ is a vector of unknown coefficients, $\boldsymbol{B_{ij}(t)} = (B_{ij_1}(t), ..., B_{ij_{k_{ij}+3}}(t))'$ is a set of known B-spline basis functions and k_{ij} represents the number of inner knots used in the representation. A second, simpler, representation assumes a standard parametric form, for example a Weibull form

$$q_{0ij}(t) = \rho_{ij} t^{\rho_{ij}-1}, \;\; (i,j) \in T \tag{3.3}$$

where $\rho_{ij} > 0$ is a shape parameter.

Upon observation of the l^{th} process, we denote by $E_l \geq 0$, the number of observed state-to-state transitions; $s_{l0} \in S$, the initial state occupied at time $t_{l0} = 0$; $t_{lm}, \; m = 1, ..., E_l$, the time at which the m^{th} transition occurs and $s_{lm} \in S$ the state entered at the m^{th} transition, $m = 1, ..., E_l$. From these, we define indicators

$$D_{lijm} = I\{s_{l(m-1)} = i, \; s_{lm} = j\}, \; m = 1, ..., E_l, \; (i,j) \in T$$

and risk set functions

$$R_{li}(t) = I\{Y_l(t^-) = i\} = \sum_{m=1}^{K_l} I\{s_{l(m-1)} = i, \; T_{l(m-1)} < t \leq T_{lm}\}, \;\; i = 1, ..., k$$

where $K_l = E_l + 1$ and $T_{lK_l} = C_l$. Letting $\boldsymbol{\Theta} = \{\{\boldsymbol{\alpha_{ij}}\}, \{\boldsymbol{\beta_{ij}}\}, \{\boldsymbol{b_{ij}}\}, \; (i,j) \in T\}$ for the spline model (3.2) and $\boldsymbol{\Theta} = \{\{\rho_{ij}\}, \{\boldsymbol{\beta_{ij}}\}, \{\boldsymbol{b_{ij}}\}, \; (i,j) \in T\}$ for the Weibull model (3.3), $\mathbf{X_l} = \{E_l, \{s_{lm}, \; m = 0, ..., E_l\}, \{T_{lm}, \; m = 0, ..., E_l + 1\}\}$ and $\mathbf{X} = \{\mathbf{X_l}, \; l = 1, ..., N\}$, the likelihood takes the form

$$L(\boldsymbol{X}|\boldsymbol{\Theta}) = \prod_{(i,j) \in T} L_{ij} \tag{3.4}$$

where

$$L_{ij} = [\prod_{l=1}^{N} \prod_{m=1}^{E_l} q_{lij}(T_{lm})^{D_{lijm}}] \times \exp(-\sum_{l=1}^{N} \int_0^{C_l} R_{li}(u) q_{lij}(u) du) \tag{3.5}$$

3.1.1 A Joint Spatial Model for Random Effects

The model contains a separate set of region specific random effects, $\boldsymbol{b_{ij}}$, for each allowable $i \rightarrow j$ transition. As a result, there are a total of N_T such sets, where the vector $\boldsymbol{b_{ij}} = (b_{ij_1}, ..., b_{ij_J})'$ contains a set of spatially correlated random variables operating on transitions from state i to state j. Examining Figure 3.1, we see that for modelling in our application, there are five such sets, $\boldsymbol{b_{12}}$, $\boldsymbol{b_{14}}$, $\boldsymbol{b_{23}}$, $\boldsymbol{b_{24}}$, $\boldsymbol{b_{34}}$, of local health unit specific random effects. In addition to allowing for spatial correlation across regions (within each $\boldsymbol{b_{ij}}$) we also wish to accommodate correlation in the different transition rates (across the $\boldsymbol{b_{ij}}$). For example, given the transition structure displayed in Figure 3.1, it was postulated that there may be positive correlation across the three sets of random effects, $\boldsymbol{b_{14}}$, $\boldsymbol{b_{24}}$, $\boldsymbol{b_{34}}$, operating on mortality rates. We thus seek a joint spatial model for the N_T sets of random effects $\boldsymbol{b_{ij}}$, $(i, j) \in T$.

To accommodate the joint correlation structure, we assume at the second level of the model, that the random effects are drawn from a multivariate generalization of the intrinsic Gaussian autoregression (Besag et al. 1991). Initially proposed by Mardia (1988) in an image processing context, multivariate conditional autoregressive models of this sort have received recent attention in the literature by Kim et al. (2001) and Knorr-Held and Rue (2002) for the joint mapping of several disease rates over a geographical region; by Gamerman et al. (2002) where they are employed as prior distributions in space-varying regression models and by Gelfand and Vounatsou (2003) for other applications in Bayesian hierarchical modelling. We let $\boldsymbol{b_h}' = \{b_{ij_h} |\ (i, j) \in T\}$, denote the vector of random effects, having length N_T, associated with region h, $h = 1, ..., J$, and $\boldsymbol{b} = (\boldsymbol{b_1}', ..., \boldsymbol{b_J}')'$ be the vector of all random effects, grouped by regions. The model for $\boldsymbol{b}$ is a Markov random field, where the joint distribution is determined through a set of local specifications. In particular, for each region h, we define a neighbourhood $\partial h = \{k | k \sim h\}$ where $k \sim h$ is typically meant to represent the adjacency of regions k and h. The model for $\boldsymbol{b}$ is then specified through the conditional distributions

$$\boldsymbol{b_h} | \{\boldsymbol{b_k},\ k \neq h\}, \boldsymbol{\Sigma} \sim N_{N_T}(\boldsymbol{\mu_h}, \boldsymbol{\Sigma_h}),\ \ h = 1, ..., J \tag{3.6}$$

where $\boldsymbol{\mu_h} = \frac{1}{n_h} \sum_{k \in \partial h} \mathbf{b_k}$, $\boldsymbol{\Sigma_h} = \frac{1}{n_h} \boldsymbol{\Sigma}$ and n_h denotes the number of neighbours of

region h. The conditional mean vector, $\boldsymbol{\mu_h}$, in (3.6) is taken as an average of the random effects associated with the neighbours of region h. The $N_T \times N_T$ positive definite and symmetric matrix $\boldsymbol{\Sigma}$ is a hyperparameter representing the (conditional) within region covariance of the random effects. With these conditional specifications, the joint distribution for $\boldsymbol{b}$ is a (singular) normal distribution, $\boldsymbol{b} \sim N_{J \times N_T}(\boldsymbol{0}, \boldsymbol{B})$ with $\boldsymbol{B}$ having generalized inverse $\boldsymbol{B^{-1}} = (\boldsymbol{D} - \boldsymbol{C}) \bigotimes \Sigma^{-1}$, where $\boldsymbol{D} = \text{diag}\{n_1, ..., n_J\}$, and $\boldsymbol{C}$, the so-called neighbourhood matrix, is defined by $C_{ij} = I\{j \sim i\}$ and $C_{ii} = 0$. As discussed by Gamerman et al. (2002), this distribution for $\boldsymbol{b}$ is improper, due to the rank deficiency of $\boldsymbol{B^{-1}}$. When employing this prior in our MCMC setting, we adopt the usual convention of working with the proper conditional distributions (3.6) and imposing a set of N_T identifying linear constraints, $\boldsymbol{b_{ij}}'\boldsymbol{1} = \boldsymbol{0}$, $(i, j) \in T$. Posterior propriety is then ensured by assigning a proper hyperprior to $\boldsymbol{\Sigma}$ (see eg. Sun et al. 1999 or Besag et al. 1995). Other approaches to dealing with this impropriety involve extending the model to incorporate so-called propriety parameters (see e.g. Carlin and Banerjee, 2002; Gelfand and Vounatsou, 2003) and such extensions will be considered in Chapter 4.

The model specification is made complete by assigning prior distributions to the remaining parameters: $\pi(\{\boldsymbol{\beta_{ij}}\})$, $\pi(\{\boldsymbol{\alpha_{ij}}\})$ (or $\pi(\{\rho_{ij}\})$, $(i, j) \in T$ and a hyperprior, $\pi(\boldsymbol{\Sigma})$. Our analysis in section 3 employs relatively vague priors for these parameters. The posterior distribution for the representation (3.2) is then

$$\pi(\boldsymbol{\Theta}, \boldsymbol{\Sigma}|\boldsymbol{X}) \propto L(\boldsymbol{X}|\boldsymbol{\Theta})\pi(\boldsymbol{b}|\boldsymbol{\Sigma})\pi(\{\boldsymbol{\alpha_{ij}}\})\pi(\{\boldsymbol{\beta_{ij}}\})\pi(\boldsymbol{\Sigma}) \tag{3.7}$$

where, $\pi(\boldsymbol{b}|\boldsymbol{\Sigma})$ is the density of the multivariate normal distribution associated with the random effects. When the Weibull representation (3.3) is used, the posterior replaces $\pi(\{\boldsymbol{\alpha_{ij}}\})$ with $\pi(\{\rho_{ij}\})$.

3.1.2 Computational Implementation

Posterior summarization is based on MCMC samples drawn from (3.7) using a component-wise Metropolis sampler. We have coded our algorithm in the C programming language. The full conditional distribution of each regression coefficient is easily

computed and is log-concave. As such, these are updated individually using Gibbs steps in conjunction with Adaptive Rejection Sampling. The shape parameters, ρ_{ij}, in the representation (3.3) are individually updated in a similar fashion. For the model based on (3.2), each vector $\boldsymbol{\alpha_{ij}}$ is block updated using a random walk Metropolis step employing a multivariate Gaussian proposal. The random effects are block updated by region in a similar manner. Finally, $\boldsymbol{\Sigma}$ is easily updated with a Gibbs step when a conditionally conjugate inverse-Wishart prior is employed. The variance of the proposal distributions in Metropolis steps are tuned in an adaptive phase to ensure acceptance rates of between 20 and 50 percent. Inference is then based on a second phase where these values are held fixed. When computing the likelihood terms (3.5) a large number of 1-dimensional integrals must be evaluated. For the Weibull model (3.3) these integrals have closed form expressions and are thus easily computed. This is not the case for the spline model (3.2) and we therefore use Romberg numerical integration for evaluation. This leads to a higher computational overhead when fitting the model based on splines. Details of the sampling algorithm for the spline model are given below. Modifications to the algorithm required for fitting the Weibull model are straightforward.

Denoting by $(\{\boldsymbol{\alpha_{ij}}\}, \{\boldsymbol{\beta_{ij}}\}, \boldsymbol{b}, \boldsymbol{\Sigma})$ the current state of the chain, we follow steps 1 to 4 below. One iteration of the sampler consists of a complete sweep through the four steps, at the end of which the new state is recorded.

1. Update spline coefficients $\boldsymbol{\alpha_{ij}}$, $(i,j) \in T$:

   We update each vector of spline coefficients, $\boldsymbol{\alpha_{ij}}$, separately. The full conditional distribution for $\boldsymbol{\alpha_{ij}}$ based on the prior $\boldsymbol{\alpha_{ij}} \overset{ind}{\sim} N_{k_{ij}+3}(\mathbf{0}, a^2\mathbf{I})$ has p.d.f. proportional to

   $$\exp(-\frac{1}{2a^2}\sum_{h=1}^{k_{ij}+3}\alpha_{ij_h}^2) \times L_{ij}$$

   where L_{ij} is given by (3.5). We use a random walk Metropolis step with candidate generated from a multivariate normal distribution.

2. Update regression coefficients $\boldsymbol{\beta_{ij}}$, $(i,j) \in T$:

   For each $(i,j) \in T$ and $k = 1, ..., p$, we update each scalar parameter β_{ij_k} separately. The full conditional distribution for β_{ij_k} based on the prior $\boldsymbol{\beta_{ij}} \overset{iid}{\sim} N_p(\mathbf{0}, a^2\mathbf{I})$ has p.d.f. proportional to

   $$\exp(-\frac{1}{2a^2}\beta_{ij_k}^2) \times L_{ij}$$

   and is easily shown to be log-concave. As such, we employ adaptive rejection sampling to obtain draws from this full conditional distribution in a Gibbs step.

3. Update random effects $\boldsymbol{b}$:

   We update the random effects associated with each region sequentially. Let $\boldsymbol{b_h}$ denote the vector of random effects, having length N_T, associated with region h, $h = 1, ..., J$. The full conditional distribution of $\boldsymbol{b_h}$ has p.d.f. proportional to

   $$[\prod_{(i,j)\in T} L_{ij_h}] \times \exp(-\frac{n_h}{2}(\boldsymbol{b_h} - \frac{1}{n_h}\sum_{k\in\partial h}\mathbf{b_k})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{b_h} - \frac{1}{n_h}\sum_{k\in\partial h}\mathbf{b_k}))$$

   where the terms L_{ij_h} are likelihood contributions associated with the set of subjects arising from region h. We use a random walk Metropolis step with candidate generated from a multivariate normal distribution.

4. Update variance components $\boldsymbol{\Sigma}$:

   A convenient prior for the $N_T \times N_T$ positive definite and symmetric matrix $\boldsymbol{\Sigma}$ is the conditionally conjugate inverse-Wishart having degrees of freedom $\nu \geq N_T$ and positive definite scale parameter $\boldsymbol{A}$. The resulting full conditional distribution is also inverse-Wishart with degrees of freedom $\nu' = \nu + J$ and scale

   $$\boldsymbol{A'} = (\sum_{i=1}^{J}\sum_{j=1}^{J}(\boldsymbol{D_{ij}} - \boldsymbol{C_{ij}})\boldsymbol{b_j}\boldsymbol{b_i}' + \boldsymbol{A^{-1}})^{-1}$$

   from which we can draw directly in a Gibbs step.

3.2 Study of Invasive Cardiac Procedure

The cardiac study involves patients, aged 25 years and older, who had been hospitalized for acute coronary syndrome (ACS) throughout the $J = 139$ local health units of Quebec during the period beginning January 1, 1996 and ending December 31, 1999. A total of 61,107 patients were recruited into the study, of which 40,031 were male. Upon discharge from index hospitalization, that corresponding to the first incidence of ACS, patients were followed over time and monitored for additional readmissions for the disease or death. Each patient was observed from discharge until December 31, 1999 or until the time of their death if this came first. All patients had either 0,1 or 2 readmissions for the disease and there were no subjects having more than two readmissions. Thus, at any time following discharge, each subject in the study can be said to occupy one of the four states depicted in Figure 3.1. Roughly one-third of the study subjects received some sort of invasive cardiac procedure during index hospitalization. This consisted of either angiography, angioplasty or aorto-coronary bypass. In addition to information on invasive cardiac procedure, other covariate information included the age (at discharge) and the gender of each subject. Of primary interest was the identification of *residual* spatial differences in mortality and readmission rates over the local health units, that is, spatial variation not accounted for by the covariates. A site-map of the local health unit structure is provided in Figure 3.2.

We fit the spatial continuous-time multi-state models described in the previous section with state space $S = \{1, 2, 3, 4\}$ depicted in Figure 3.1 and both Weibull and cubic B-spline representations for the baseline intensity functions. The B-spline representation for each $q_{0ij}(t)$ is based on $k_{ij} = 1$ inner knot placed at the median of the observed $i \to j$ transition times. The spline representations are then based on $\dim(\boldsymbol{\alpha_{ij}}) = 4$ parameters; whereas, the Weibull forms use only 1 parameter, ρ_{ij}, for each $(i, j) \in T$. Regarding prior distributions, we assume $\boldsymbol{\alpha_{ij}} \overset{iid}{\sim} N_4(\mathbf{0}, 10^2\mathbf{I})$ for the spline coefficients and $\rho_{ij} \overset{iid}{\sim} \text{Gamma}(0.1, 0.1)$ for the shape parameters in the Weibull model. For the remaining parameters, both models assume $\boldsymbol{\beta_{ij}} \overset{iid}{\sim} N_p(\mathbf{0}, 10^2\mathbf{I})$ with $p = 4$ (three covariates plus intercept) and $\boldsymbol{\Sigma}^{-1} \sim \text{Wishart}(\nu, \mathbf{A})$ where setting $\nu = 5$ and $\mathbf{A} = \mathbf{I}$ ensure a relatively vague prior. Each model was fit by running four

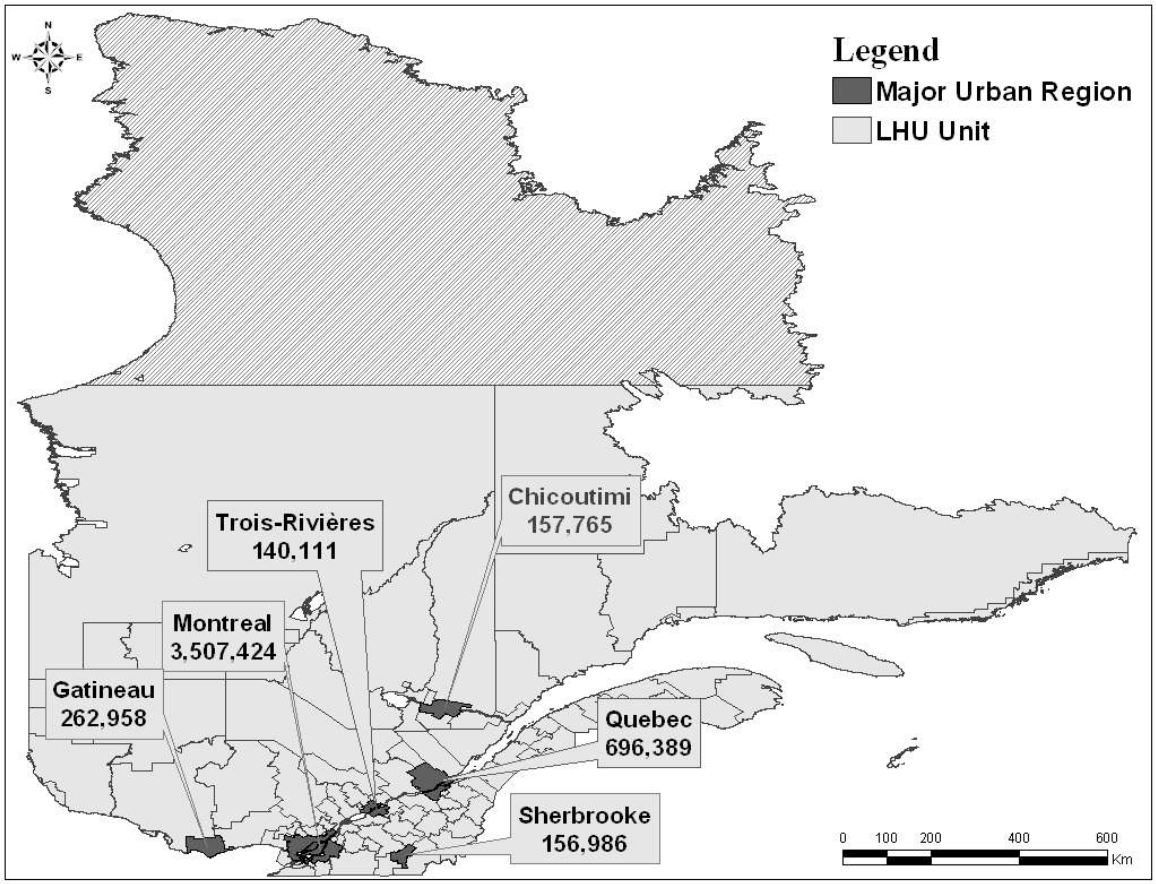


Figure 3.2: Site map of Quebec divided into 139 local health units.

MCMC chains in parallel and convergence of the samplers to posterior distributions was assessed through the examination of trace plots, log-posterior monitoring and Gelman-Rubin (1992) statistics. For the Weibull model, a burn-in of 7,000 iterations was followed by an additional 9,000 iterations from each chain yielding a total of 36,000 samples to be used for posterior summarization. The spline model required a greater number, 15,000, of burn-in iterations followed by another 11,000 iterations resulting in 44,000 posterior samples.

Defining cumulative baseline transition intensities: $Q_{0ij}(t) = \int_0^t q_{0ij}(u)du$, posterior summaries of $Q_{014}(t)$, $Q_{024}(t)$ and $Q_{034}(t)$, the cumulative intensities associated with mortality (state 4) are displayed in Figure 3.3 for both the spline and Weibull models. To assess and compare the fit of each model, Figure 3.3 also displays the semiparametric step-function estimates obtained from a simpler analysis. The simpler analysis ignores the random effects altogether and for each $(i, j) \in T$: estimates regression coefficients using partial likelihood (via the *coxph()* function in R) and obtains nonparametric estimates of the cumulative baseline intensity function using the Breslow generalization of the Nelson-Aalen estimator (via the *survfit()* function in R). Along the same lines, posterior summaries of the cumulative baseline intensities associated with first and second readmission, $Q_{012}(t)$ and $Q_{023}(t)$ are given in Figure 3.4. Comparing with the semiparametric estimates, the spline model appears to fit very well in all cases. In contrast, the simple 1-parameter Weibull representations do not perform as well particularly for estimation of $Q_{014}(t)$ (Figure 3.3b) and $Q_{012}(t)$ (Figure 3.4b).

In addition, we have also considered several submodels which make simplified assumptions on the spatial random effects, $\boldsymbol{b_{ij}}$, $(i, j) \in T$. In the first, we simplify the conditional specifications (3.6) by setting $\boldsymbol{\mu_h} = \mathbf{0}$ and $\boldsymbol{\Sigma_h} = \boldsymbol{\Sigma}$, $h = 1, ..., J$. As a result, we have $\boldsymbol{b_h} \overset{iid}{\sim} N_5(\mathbf{0}, \boldsymbol{\Sigma})$, $h = 1, ..., J$, for each 5-vector of region specific random effects. Such a model assumes spatial independence across regions but allows for within region dependence across transitions within each vector $\boldsymbol{b_h} = (b_{12_h}, b_{14_h}, b_{23_h}, b_{24_h}, b_{34_h})'$. In the second submodel, we assume independent Gaussian

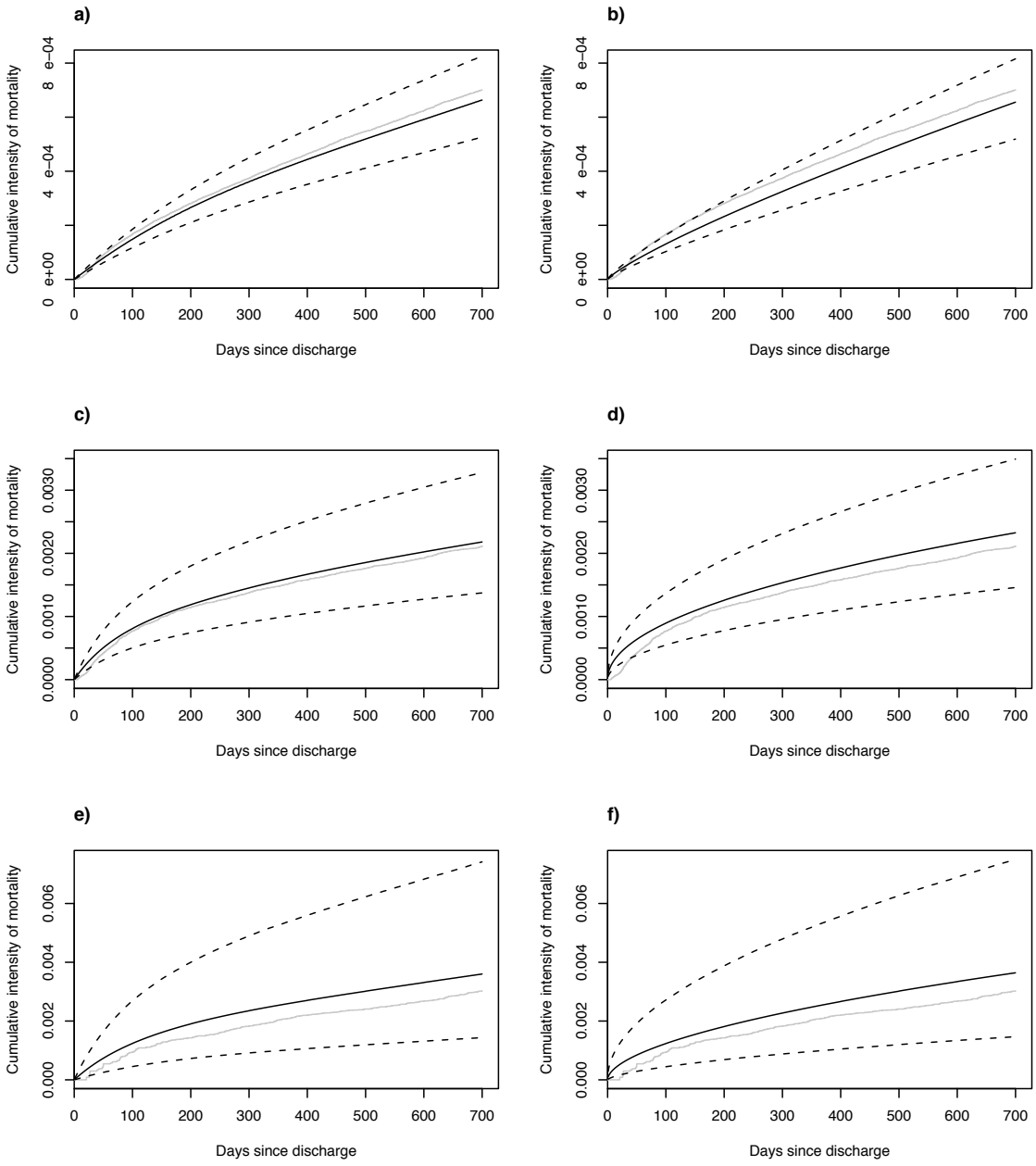


Figure 3.3: Estimated cumulative baseline intensities associated with mortality after 0, 1 and 2 readmissions (posterior means and 95% credible intervals) a) $Q_{014}(t)$ - Spline; b) $Q_{014}(t)$ - Weibull; c) $Q_{024}(t)$ - Spline; d) $Q_{024}(t)$ - Weibull; e) $Q_{034}(t)$ - Spline; f) $Q_{034}(t)$ - Weibull. For comparison, the step-function estimates obtained from the semiparametric analysis are indicated within each plot by the grey curve.

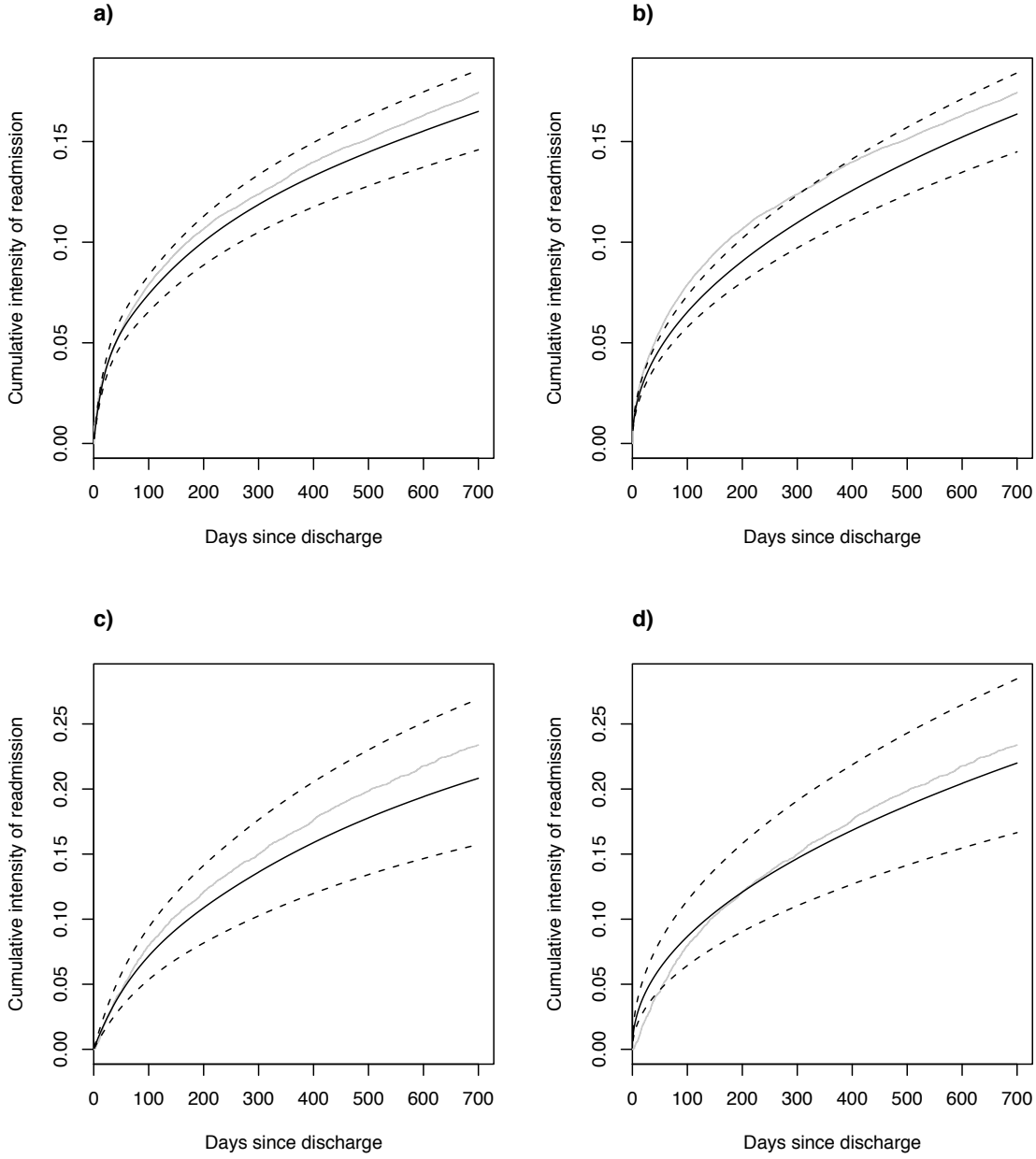


Figure 3.4: Estimated cumulative baseline intensities of first and second readmission (posterior means and 95% credible intervals) a) $Q_{012}(t)$ - Spline; b) $Q_{012}(t)$ - Weibull; c) $Q_{023}(t)$ - Spline; d) $Q_{023}(t)$ - Weibull. For comparison, the step-function estimates obtained from the semiparametric analysis are indicated within each plot by the grey curve.

intrinsic autoregressions for each $\boldsymbol{b_{ij}}$, $\boldsymbol{b_{ij}} \overset{ind}{\sim} IAR(\sigma_{ij})$, $(i,j) \in T$ which implies

$$\boldsymbol{\Sigma} = \text{diag}\{\sigma_{12}, \sigma_{14}, \sigma_{23}, \sigma_{24}, \sigma_{34}\}.$$

This model allows for spatial dependence across regions but assumes independence between the random effects operating on different types of transitions. Finally, our third submodel sets the random effects $\boldsymbol{b_{ij}} \equiv \mathbf{0}$, $\forall (i,j) \in T$. Such a model ignores the spatial aspect altogether and is a standard continuous-time Markov regression model. For each of the three submodels we examine both Weibull and spline representations for the intensity functions. In total, eight models are fit, representing different assumptions on the spatial effects and intensity functions.

To compare the various models we employ the deviance information criterion (DIC), a 'fit plus penalty' model selection tool, proposed by Spiegelhalter et al. (2002). The criterion is given by

$$\text{DIC} = \overline{D(\Theta)} + p_D$$

where the deviance, $D(\Theta)$, is defined in terms of (3.4) via $D(\Theta) = -2 \log L(\boldsymbol{X}|\boldsymbol{\Theta})$ and $\overline{D(\Theta)}$, the posterior mean of the deviance, is a measure of model fit. The penalty term, p_D, is defined by $p_D = \overline{D(\Theta)} - D(\bar{\Theta})$ where $\bar{\Theta}$ is the posterior mean of Θ. The term p_D measures the effective number of model parameters and will be lower when the data imposes a higher level of shrinkage in the random effects. Models with lower DIC scores are preferred as they achieve more optimal combinations of fit and parsimony.

Table 3.1 lists the DIC and p_D values for all eight models considered. The DIC scores for the B-spline models are uniformly lower than those for the Weibull models. For each row in Table 3.1, the p_D values for the B-spline models are greater than those for the Weibull models by about 16, a value slightly larger than the raw difference of 15 parameters. The spline representations prove to be more flexible however as they result in substantial improvement in model fit with the DIC scores differing by over 700 in each case. Comparing the various models for the random effects, the criterion seems to favor the joint spatial model defined in (3.6) over the various submodels. Among the submodels, submodel 1 which assumes spatial independence across regions but incorporates within region dependence in each $\boldsymbol{b_h}$ is preferred over

submodel 2 which assumes that each vector, $\boldsymbol{b_{ij}}$, of spatial effects is independently drawn from an $IAR(\sigma_{ij})$ model. Indeed, the data seem to encourage a greater level of shrinkage in the random effects for submodel 1 compared with submodel 2 as the p_D values of the former are smaller than those of the latter. Finally, the non-mixed models (submodel 3) preform the worst. As expected, the p_D values for these models are the lowest and are essentially equal to the raw parameter counts of 40 (spline model) and 25 (Weibull model); however, the resulting poor fits yield the largest DIC scores in both cases. Overall, the model which employs B-spline representations for intensity functions and the joint spatial model for the random effects is considered most optimal according to the criterion. We therefore summarize the results of this model further.

	B-Spline		Weibull	
Model for random effects	p_D	DIC	p_D	DIC
$\boldsymbol{b} \sim N_{5J}(\boldsymbol{0}, \boldsymbol{B}),\ \boldsymbol{B^{-1}} = (\boldsymbol{D} - \boldsymbol{C}) \bigotimes \Sigma^{-1}$	268	71	252	787
$\boldsymbol{b_h} \overset{iid}{\sim} N_5(\boldsymbol{0}, \boldsymbol{\Sigma}),\ h = 1, ..., J$	353	128	336	839
$\boldsymbol{b_{ij}} \overset{ind}{\sim} IAR(\sigma_{ij}),\ (i,j) \in T$	426	156	410	873
$\boldsymbol{b_{ij}} \equiv \boldsymbol{0},\ \forall (i,j) \in T$	40	612	24	1,336

Table 3.1: DIC scores (after subtracting 340,000) and p_D for the eight models considered for Quebec cardiac data.

Table 3.2 presents posterior summaries for the regression coefficients, $\boldsymbol{\beta_{14}}$, $\boldsymbol{\beta_{24}}$ and $\boldsymbol{\beta_{34}}$, associated with each of the three transitions related to mortality (transitions into state 4). Invasive cardiac procedure is associated with substantially lower mortality rates in all three cases. The posterior mean (95% CI) age and gender adjusted relative risks are 0.44 (0.40,0.47) for patients with no readmissions, 0.53 (0.45,0.62) for patients with one readmission and 0.50 (0.37,0.66) for those who have been admitted twice for the disease. Table 3.3 presents posterior summaries of the regression coefficients associated with the remaining two transitions, those related to readmission. We note that invasice cardiac procedure is associated with decreased rates of first readmission. The age and gender adjusted posterior mean (95 % CI) relative risk is 0.76 (0.73,0.79).

From Tables 3.2 and 3.3 we also note that event histories depend on the age of subjects. Older patients tend to have higher rates of mortality and readmission, an intuitive result.

Table 3.4 gives posterior summaries of the variances and correlations associated with the hyperparameter $\boldsymbol{\Sigma}$. Interpretation here is delicate as $\boldsymbol{\Sigma_h} = \frac{1}{n_h}\boldsymbol{\Sigma}$ is not directly interpretable as the variance matrix of $\boldsymbol{b_h}$. Rather, it is the variance of $\boldsymbol{b_h}$ conditional on the random effects at the neighbouring sites. Most striking in the table is the relatively large and positive conditional correlation between b_{12_h} and b_{23_h} suggesting that regions which have above average residual first readmission rates will also tend to have above average residual second readmission rates. The positive correlation also suggests that missing (area level) covariates, represented by the random effects, would have consistent effects on the two types of transitions (see e.g. Cook and Ng 1997). Figure 3.5 compares the estimates of each $\boldsymbol{b_{ij}}$ in a matrix scatter plot. The plots seem to indicate positive correlation among the estimates of b_{14_h}, b_{24_h} and b_{34_h}, the random effects which operate on mortality rates as well as positive correlation between estimates of b_{12_h} and b_{23_h}.

The random effects capture residual variation in mortality and readmission rates over the local health units. To examine the extent of this variation, Figure 3.6 displays boxplots summarizing the posterior samples of $\boldsymbol{b_{14}}$, $\boldsymbol{b_{24}}$ and $\boldsymbol{b_{34}}$. The posterior distributions of these spatial effects seem to vary considerably across the local health units, perhaps further justifying their inclusion in the model. The posterior means of these random effects are mapped in Figure 3.7. The three maps reveal similar patterns and indicate higher residual mortality risks towards the southern most regions of the province. The fact that residual mortality risks are higher in these predominantly urban areas might be further studied as in Kunzli et al. (2005) who report associations between air pollution and cardiovascular morbidity and mortality. The posterior samples of $\boldsymbol{b_{12}}$ and $\boldsymbol{b_{23}}$ are summarized and estimates mapped in Figures 3.8 and 3.9 respectively. Figure 3.8 reveals substantial differences in the posterior distributions across the local health units and the maps indicate higher risks of readmission towards the southwest and northeast regions of the province. These are primarily rural areas, in contrast with the higher mortality risks for urban areas noted above. The higher

Table 3.2: Posterior summaries of regression coefficients associated with each of the three transitions associated with mortality

	$1 \rightarrow 4$		$2 \rightarrow 4$		$3 \rightarrow 4$	
Parameter	Mean	95% CI	Mean	95% CI	Mean	95% CI
Intercept	−13.31	(-13.56,-13.07)	−11.24	(-11.72,-10.76)	−11.049	(-12.06,-10.059)
Revasc	−0.83	(-0.91,-0.75)	−0.63	(-0.79,-0.48)	−0.70	(-1.0,-0.41)
Age	0.073	(0.071,0.076)	0.072	(0.067,0.077)	0.070	(0.060,0.080)
Gender(Male)	0.20	(0.14,0.26)	0.0050	(-0.11,0.12)	0.19	(-0.0085,0.41)

Table 3.3: Posterior summaries of regression coefficients associated with each of the two transitions associated with readmission

	$1 \rightarrow 2$		$2 \rightarrow 3$	
Parameter	Mean	95% CI	Mean	95% CI
Intercept	-5.90	(-6.029,-5.77)	-6.77	(-7.11,-6.48)
Revasc	-0.27	(-0.32,-0.23)	-0.0056	(-0.10,0.088)
Age	0.0049	(0.0034,0.0065)	0.013	(0.0094,0.016)
Gender(Male)	-0.0085	(-0.050,0.033)	-0.15	(-0.24,-0.060)

Table 3.4: Posterior summaries for the conditional covariance matrix, $\mathbf{\Sigma}$, obtained from the final chosen model.

Parameter	Mean	95% CI
Σ_{11} (variance component b_{12_h})	0.12	(0.083, 0.17)
Σ_{22} (variance component b_{14_h})	0.086	(0.053,0.13)
Σ_{33} (variance component b_{23_h})	0.21	(0.12,0.34)
Σ_{44} (variance component b_{24_h})	0.17	(0.095, 0.30)
Σ_{55} (variance component b_{34_h})	0.16	(0.078, 0.31)
$\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$ (conditional correlation b_{12_h} and b_{14_h})	-0.20	(-0.47,0.090)
$\frac{\Sigma_{13}}{\sqrt{\Sigma_{11}\Sigma_{33}}}$ (conditional correlation b_{12_h} and b_{23_h})	0.52	(0.26,0.71)
$\frac{\Sigma_{14}}{\sqrt{\Sigma_{11}\Sigma_{44}}}$ (conditional correlation b_{12_h} and b_{24_h})	-0.43	(-0.67, -0.13)
$\frac{\Sigma_{15}}{\sqrt{\Sigma_{11}\Sigma_{55}}}$ (conditional correlation b_{12_h} and b_{34_h})	-0.25	(-0.62, 0.21)
$\frac{\Sigma_{23}}{\sqrt{\Sigma_{22}\Sigma_{33}}}$ (conditional correlation b_{14_h} and b_{23_h})	-0.33	(-0.60, 0.00082)
$\frac{\Sigma_{24}}{\sqrt{\Sigma_{22}\Sigma_{44}}}$ (conditional correlation b_{14_h} and b_{24_h})	0.39	(0.066, 0.64)
$\frac{\Sigma_{25}}{\sqrt{\Sigma_{22}\Sigma_{55}}}$ (conditional correlation b_{14_h} and b_{34_h})	0.18	(-0.27, 0.54)
$\frac{\Sigma_{34}}{\sqrt{\Sigma_{33}\Sigma_{44}}}$ (conditional correlation b_{23_h} and b_{24_h})	-0.35	(-0.66, 0.022)
$\frac{\Sigma_{35}}{\sqrt{\Sigma_{33}\Sigma_{55}}}$ (conditional correlation b_{23_h} and b_{34_h})	-0.21	(-0.63,0.35)
$\frac{\Sigma_{45}}{\sqrt{\Sigma_{44}\Sigma_{55}}}$ (conditional correlation b_{24_h} and b_{34_h})	0.34	(-0.16, 0.69)

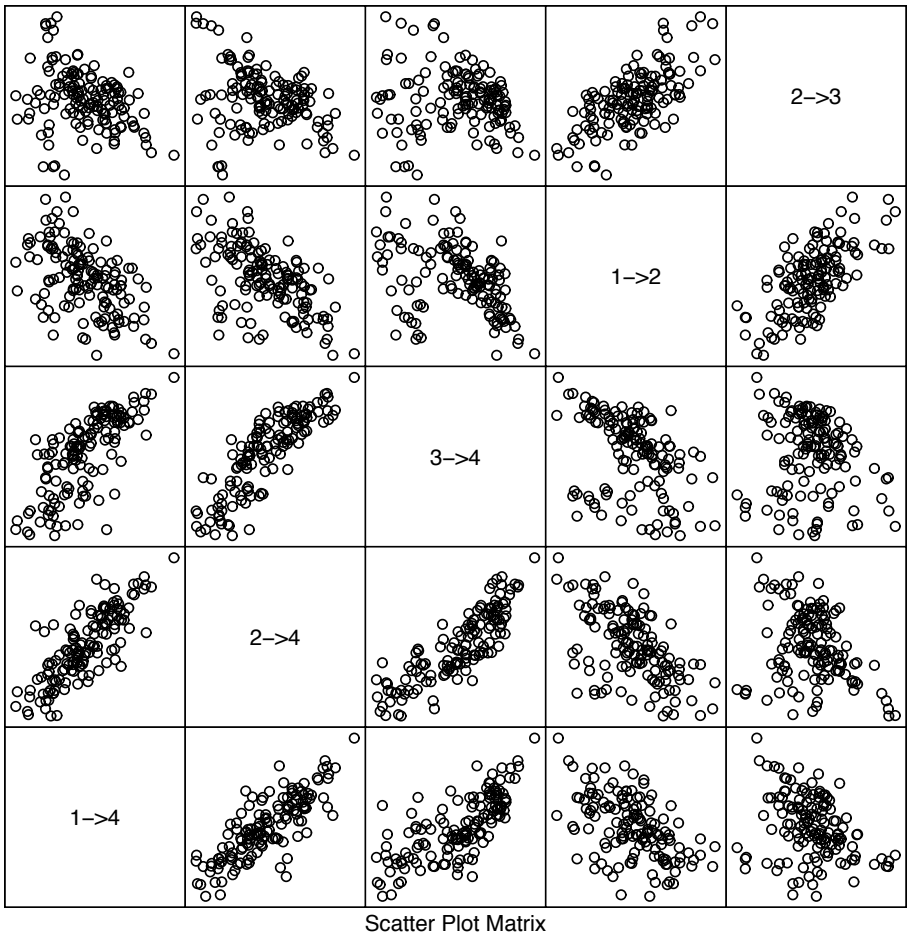


Figure 3.5: Matrix scatter plot comparing the posterior mean estimates of $\boldsymbol{b_{14}}$, $\boldsymbol{b_{24}}$, $\boldsymbol{b_{34}}$, $\boldsymbol{b_{12}}$ and $\boldsymbol{b_{23}}$.

residual readmission rates could be reflecting differing health policy in these regions. Geographic trends such as those revealed in Figures 3.7 and 3.9 can be useful as they may give clues regarding missing variables that underlie the autocorrelation observed in the maps.

3.3 Discussion

There are several directions for future work which are currently being investigated. First, the methodology presented here will be extended to accommodate processes under panel observation. In this scenario, subjects can only be observed at discrete and irregular time points and the exact timing of events is unknown. With panel data, the complexity of the likelihood increases significantly compared with the form (3.4) considered here. The Bayesian framework adopted here, allows for data augmentation which may prove useful in such a missing data setting.

In our modelling we have allowed the spatial effects, $\boldsymbol{b_{ij}}$, to be correlated across different transition types $(i, j) \in T$; whereas, the regression coefficients, $\boldsymbol{\beta_{ij}}$, are assumed independent. This assumption could be relaxed allowing for correlation in the covariate effects across transitions. For example, examination of Table 3.2 reveals that both invasive cardiac procedure and age have similar effects on all transitions into state 4. In addition, models containing subject specific (in addition to the region specific) frailty terms are also being investigated.

Techniques for the modelling of intensity functions using splines will also be refined by addressing the crucial problem of choosing the number and positions of the knots. In situations where the transition intensities follow more complicated functional forms, a larger number of knots and advanced knot selection techniques may prove useful. For example, a close examination of Figures 3.3 and 3.4 suggests the need for an additional knot at some point during early time periods to model the 'dip' apparent when compared to the semiparametric estimates. We are examining several approaches including the use of penalized splines (P-splines) as well as adaptive knot selection where inference is conducted within a reversible jump MCMC setting.

Finally, moving beyond the spatial epidemiologic setting considered here, it is

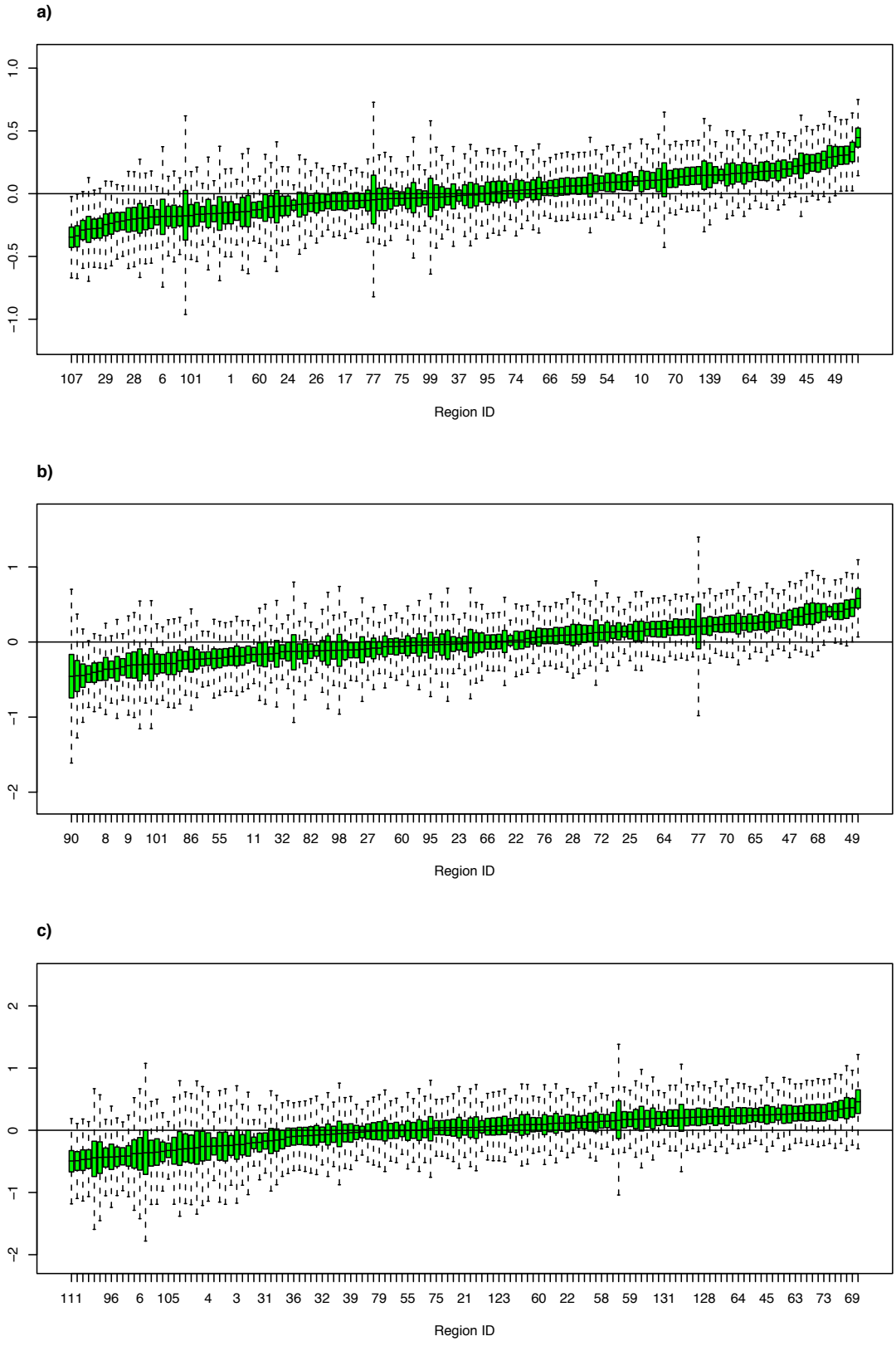
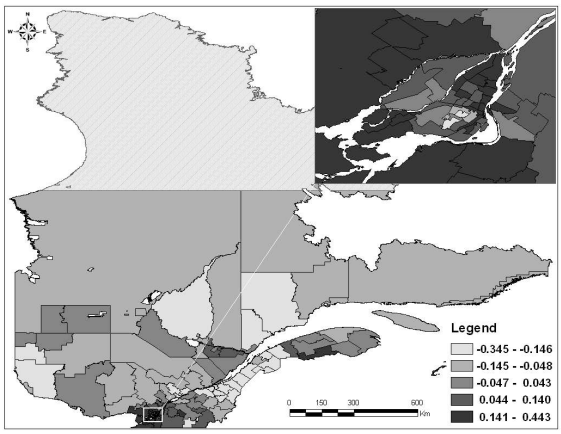
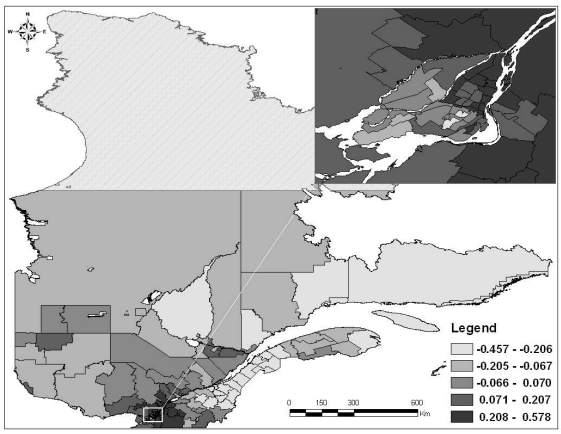


Figure 3.6: Boxplots (arranged in increasing order by posterior median) obtained from posterior samples of random effects associated with mortality a) $\boldsymbol{b_{14}}$; b) $\boldsymbol{b_{24}}$; c)$\boldsymbol{b_{34}}$.

a)



b)



c)

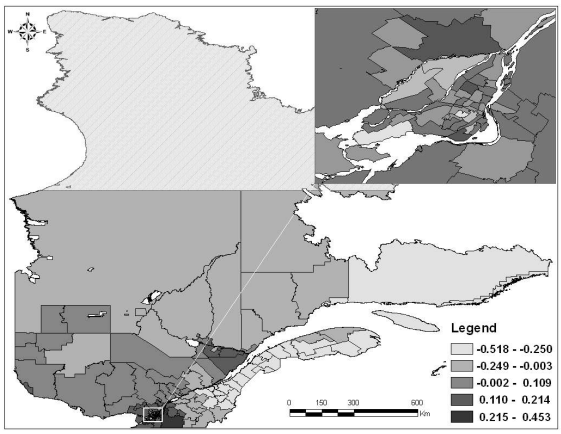


Figure 3.7: Posterior mean maps of random effects associated with mortality a) $\boldsymbol{b_{14}}$; b) $\boldsymbol{b_{24}}$; c)$\boldsymbol{b_{34}}$.

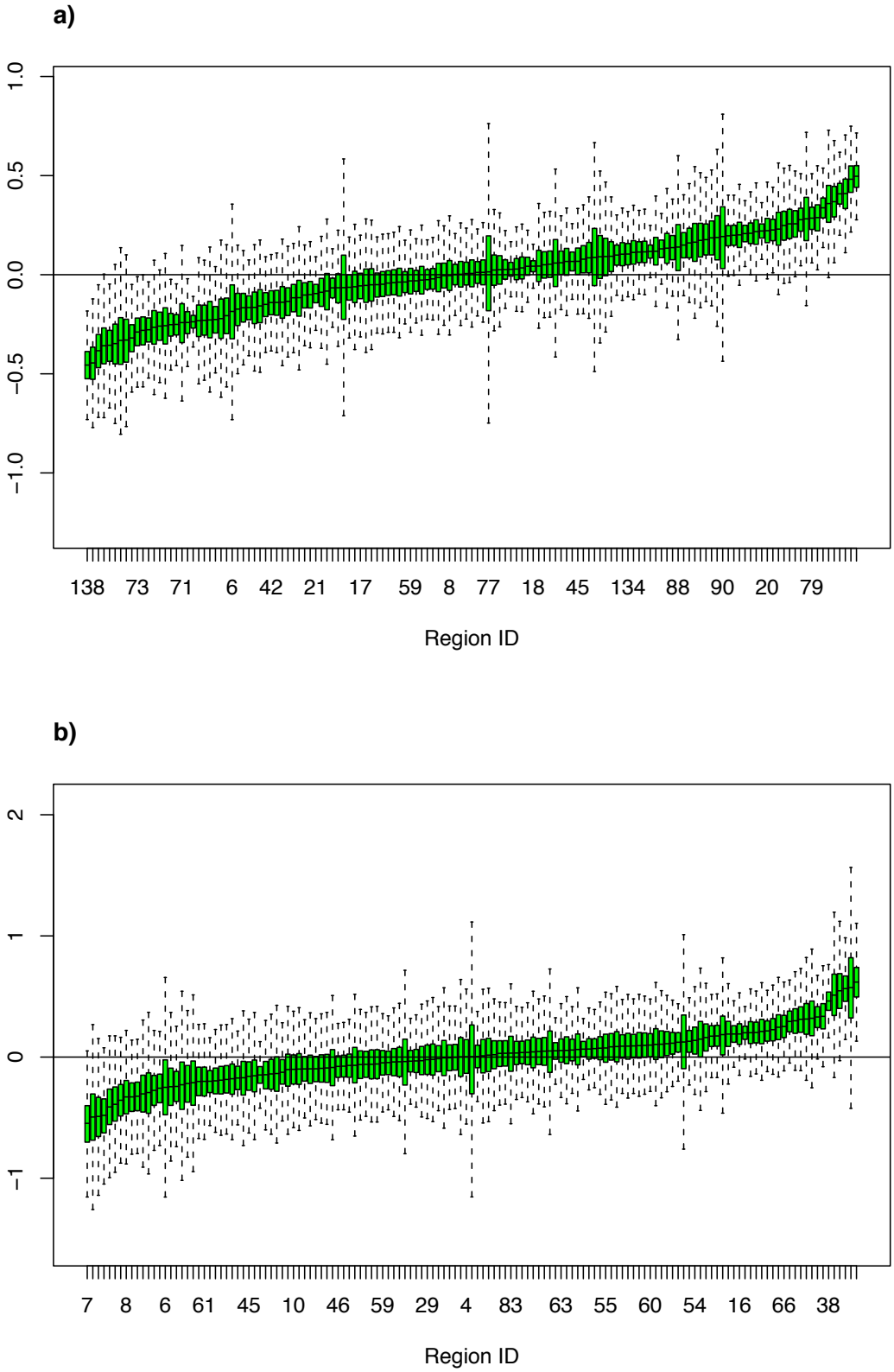


Figure 3.8: Boxplots (arranged in increasing order by posterior median) obtained from posterior samples of random effects associated with readmission a) $\boldsymbol{b_{12}}$; b) $\boldsymbol{b_{23}}$.

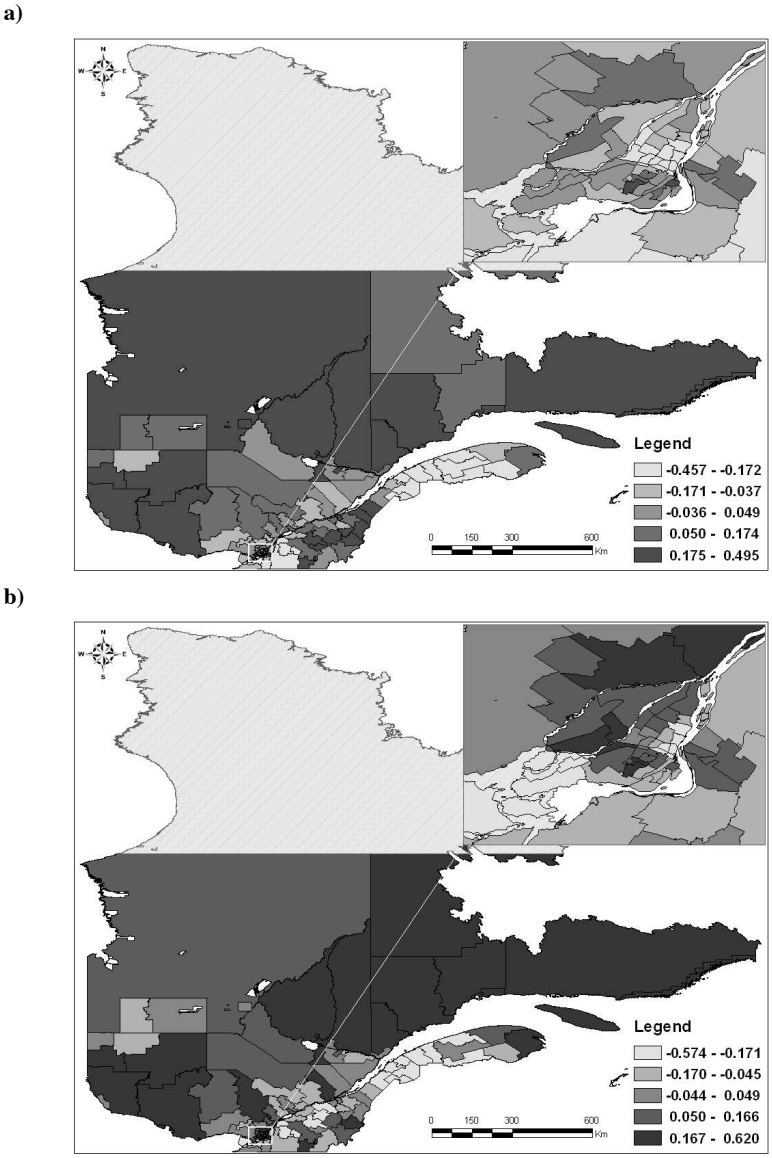


Figure 3.9: Posterior mean maps of those random effects associated with first and second readmission a) $\boldsymbol{b_{12}}$; b) $\boldsymbol{b_{23}}$.

envisioned that our methodology can be applied in other areas where multi-state data are collected and spatial aspects are present but perhaps more subtle. For example, this may be the case in longitudinal studies in physiology, where several joints on the human body are examined and dependence between joints is a factor.

Chapter 4

Extending the Spatial Mover-Stayer Model

In this Chapter we return to the discrete-time setting and refocus our attention on the spatially correlated mover-stayer model developed in Chapter 2. There, methodological developments focussed primarily on inference through maximum likelihood although a supplementary Bayesian analysis was carried out for comparison. In this Chapter, we focus on Bayesian inference and develop methodological extensions within this framework.

We begin in Section 4.1 by motivating an expansion of the model. Such an expansion incorporates a joint spatial structure for the random effects. As with the multivariate intrinsic autoregression employed in Chapter 3, the model for the random effects, which we present in Section 4.2, allows for both spatial correlation as well as correlation across transition probabilities. Section 4.2 also takes up the issue of hypothesis testing with respect to the number of mixture components. Specifically, we develop a test for 'stayers' in our mover-stayer model. Such a test is based on the assignment of a discrete mass prior to the mixing probability. A Metropolis-Hastings algorithm is then developed to produce posterior summaries including the posterior probability associated with the hypothesis of interest. We illustrate the techniques through an analysis of synthetic data in Section 4.3 and through a re-analysis of the weevil infestation data in Section 4.4. Section 4.4 also considers model validation

based on posterior predictive methods (Gelman et al. 1996).

4.1 Motivation

The mover-stayer model presented in Chapter 2 allowed for spatial correlation through the incorporation of two vectors of random effects: $\boldsymbol{b_0} = (b_{01}, ..., b_{0N})'$ and $\boldsymbol{b_1} = (b_{11}, ..., b_{1N})'$. For the i^{th} subject, b_{0i} and b_{1i} operate on transition probabilities accounting for spatial correlation in transitions out of states 0 and 1 respectively. Each vector of spatial effects $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ was assumed independently drawn from a Gaussian intrinsic autoregression: $\boldsymbol{b_l} \overset{ind}{\sim} IAR(\sigma_l),\ l = 0, 1$. While allowing for spatial correlation, this prior assumes independence within each pair $\boldsymbol{b_{(i)}} = (b_{0i}, b_{1i})'$ and thus the prior can be factorized $\pi(\boldsymbol{b_0}, \boldsymbol{b_1}|\sigma_0, \sigma_1) = \pi(\boldsymbol{b_0}|\sigma_0)\pi(\boldsymbol{b_1}|\sigma_1)$. Upon observation of the data, $\boldsymbol{Y}$, this prior independence, when combined with the likelihood which factorizes in a similar manner, results in a separable posterior

$$\pi(\boldsymbol{b_0}, \boldsymbol{b_1}, \boldsymbol{\Theta}|\boldsymbol{Y}\ ) = \pi(\boldsymbol{b_0}, \boldsymbol{\Theta_0}|\boldsymbol{Y_0}) \times \pi(\boldsymbol{b_1}, \boldsymbol{\Theta_1}|\boldsymbol{Y_1}\ ) \tag{4.1}$$

where $\boldsymbol{\Theta} = \boldsymbol{\Theta_0} \cup \boldsymbol{\Theta_1}$ denotes the remaining model parameters separated into two disjoint sets, and $\mathbf{Y_0}$ and $\mathbf{Y_1}$ are defined in (2.7) and (2.8) respectively. The resulting a posteriori independence of $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$, *when justified*, can be used advantageously for model fitting allowing the use of standard software. In this case, the posterior factorization (4.1) implies that two separate and independent MCMC samplers may be used: one to draw posterior samples of $\boldsymbol{b_0}$ and $\boldsymbol{\Theta_0}$ and another to draw posterior samples of $\boldsymbol{b_1}$ and $\boldsymbol{\Theta_1}$. Sampling from the individual posterior factors $\pi(\boldsymbol{b_0}, \boldsymbol{\Theta_0}|\boldsymbol{Y_0})$ and $\pi(\boldsymbol{b_1}, \boldsymbol{\Theta_1}|\boldsymbol{Y_1}\ )$ can be accomplished using the WinBugs software.

Figure 4.1a compares the posterior mean estimates of b_{0i} and b_{1i} in a scatter plot obtained from the analysis of the weevil infestation data in Chapter 2. Under the model (4.1), these estimates should not display correlation (as posterior samples should be independent). Nevertheless, the scatter plot appears to indicate a residual negative correlation within each pair $\boldsymbol{b_{(i)}} = (b_{0i}, b_{1i})'$ of estimates. Considering the context of the analysis, the negative correlation is intuitive. It implies that those trees having an above average probability of transition into the infected state (large b_{0i}) will

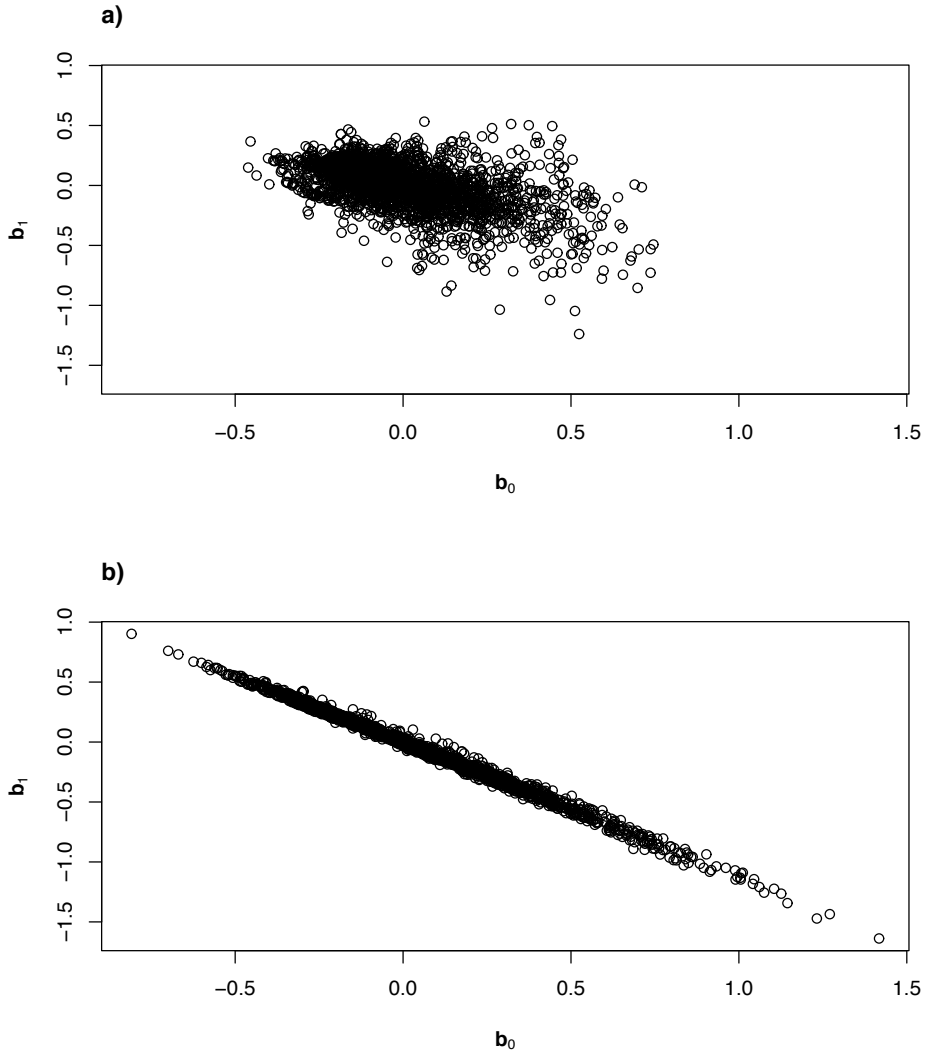


Figure 4.1: Scatter plot comparing posterior mean estimates of $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$: a) Estimates obtained in Chapter 2 assuming $\boldsymbol{b_l} \overset{ind}{\sim} IAR(\sigma_l),\ l = 0, 1$; b) Estimates obtained from expanded model assuming $\boldsymbol{b} \sim 2CAR(\kappa, \boldsymbol{\Sigma})$.

have a correspondingly lower than average probability of transition into the uninfected state (small b_{1i}). An expanded model, accommodating such correlation is presented in the next section. Moving beyond our specific application, such an extended model will serve usefully in a wide variety of scenarios where random effects represent heterogeneity arising from missing covariates. The within-pair correlation accommodated by the extended model can lend insight into the nature of this heterogeneity.

4.2 An Extended Spatial Mover-Stayer Model

Following the notation of Chapter 2, we assume there are N spatially arranged subjects and each subject is observed over a set of equally spaced time points. As before, we let $\mathbf{y}_i = (y_i(0), ..., y_i(n_i - 1))'$, $i = 1, ..., N$, denote the response vector of binary values obtained from subject i, where $y_i(t)$ indicates the state occupied by subject i at time t and $\boldsymbol{x_i(t)}$ is a corresponding vector of covariates.

We let $\boldsymbol{\Theta}$ denote the collection of all model parameters (now including the random effects). At the first level of the model, we assume, conditional on $\boldsymbol{\Theta}$, that each $\mathbf{y_i}$ is independently drawn from a two-component mixture having density $f_{MS}(\mathbf{y_i}|\boldsymbol{\Theta})$ given by

$$f_{MS}(\mathbf{y_i}|\boldsymbol{\Theta}) = p_M f_{MC}(\mathbf{y_i}|\boldsymbol{\Theta}) + (1 - p_M) I(\mathbf{y_i} = \mathbf{0}) \tag{4.2}$$

where $p_M \in [0, 1]$ is a mixing probability; and $f_{MC}(\mathbf{y_i}|\boldsymbol{\Theta})$ is the density of a 1^{st} order two-state Markov chain, which, upon adopting the mixed logistic specifications (2.3) takes the form

$$f_{MC}(\mathbf{y_i}|\boldsymbol{\Theta}) = p_I^{y_i(0)}(1 - p_I)^{1-y_i(0)} \times L_{0i} \times L_{1i} \tag{4.3}$$

where

$$L_{0i} = \prod_{t \in D_{0i}} \frac{\exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0i})^{y_i(t)}}{1 + \exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0i})} \tag{4.4}$$

$$L_{1i} = \prod_{t \in D_{1i}} \frac{\exp(\boldsymbol{\beta_1}'\boldsymbol{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1i})^{1-y_i(t)}}{1 + \exp(\boldsymbol{\beta_1}'\boldsymbol{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1i})} \tag{4.5}$$

and, as before, p_I is an initial state probability; $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ are temporal trends modelled using the B-spline representations (2.4); and

$D_{li} = \{t > 0 | y_i(t-1) = l\}$, $l = 0, 1$. At this lowest level of the hierarchy, the model is the same as that presented in Chapter 2. Changes in the model structure occur at the second level, where we assign a prior distribution to the random effects.

To jointly model the two vectors of spatial random effects, $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$, we adopt a bivariate conditional autoregressive model. We let $\mathbf{b}$ denote the vector of all random effects, grouped by subjects $\mathbf{b} = (\mathbf{b}'_{(1)}, ..., \mathbf{b}'_{(N)})'$. As with the joint spatial model employed in Chapter 3, the prior for $\mathbf{b}$ is a Markov random field specified through a set of conditional distributions:

$$\left(\begin{array}{c} b_{0i} \\ b_{1i} \end{array} \right) \Bigg| \left\{ \left(\begin{array}{c} b_{0j} \\ b_{1j} \end{array} \right) \right\}_{j \neq i} \sim N_2(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}), \quad i = 1, ..., N \tag{4.6}$$

where

$$\boldsymbol{\mu_i} = \frac{\kappa}{C_{i\cdot}} \sum_j C_{ij} \mathbf{b_{(j)}} \tag{4.7}$$

and $\boldsymbol{\Sigma_i} = \frac{1}{C_{i\cdot}} \boldsymbol{\Sigma}$. Here, as in Chapter 2, C_{ij} is a scalar weight measuring the closeness or adjacency of subjects i and j ($C_{ii} = 0$); $C_{i\cdot} = \sum_j C_{ij}$; and κ is a so-called propriety parameter. With these conditional specifications, the joint distribution for the vector of all random effects (grouped by subjects) is given by $\mathbf{b} \sim N_{2N}(\mathbf{0}, \boldsymbol{B})$ with $\boldsymbol{B}$ having generalized inverse $\boldsymbol{B^{-1}} = (\boldsymbol{D} - \kappa \boldsymbol{C}) \bigotimes \Sigma^{-1}$, where $\mathbf{D} = \text{diag}\{C_{1\cdot}, C_{2\cdot}, ..., C_{N\cdot}\}$. Note that by taking $\kappa = 1$ we obtain a 2-dimensional version of the multivariate intrinsic autoregression employed in Chapter 3. Along the same lines, taking $\kappa = 1$ and $\boldsymbol{\Sigma} = \text{diag}\{\sigma^2_{b0}, \sigma^2_{b1}\}$ we obtain the spatial structure used in Chapter 2. As discussed by Cressie (1993) and Sun et al. (1999) in the univariate spatial setting and by Carlin and Banerjee (2002) and Gelfand and Vounatsou (2003) in the multivariate setting, restricting κ to an appropriate range ensures that the joint prior for the random effects is proper. In particular, taking $\kappa \in [0, 1)$ will result in propriety (a wider range allowing negative values is allowable but not desirable). Within this range, lower values of κ imply a lesser degree of spatial dependence with $\kappa = 0$ corresponding to spatial independence. Following Carlin and Banerjee (2002) we denote this distribution by $2CAR(\kappa, \boldsymbol{\Sigma})$.

The model specification is made complete by assigning prior distributions to the

remaining components of $\boldsymbol{\Theta} = \{\mathbf{b}, \kappa, \boldsymbol{\Sigma}, \boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\alpha_0}, \boldsymbol{\alpha_1}, p_M, p_I\}$. The posterior distribution then takes the form

$$\pi(\boldsymbol{\Theta}|\mathbf{Y}) \propto [\prod_{i=1}^{N} f_{MS}(\mathbf{y_i}|\boldsymbol{\Theta})] \times \pi(\mathbf{b}|\kappa, \boldsymbol{\Sigma})\pi(\kappa)\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\beta_0})\pi(\boldsymbol{\beta_1})\pi(\boldsymbol{\alpha_0})\pi(\boldsymbol{\alpha_1})\pi(p_M)\pi(p_I) \tag{4.8}$$

In section 4.2.2 we outline two separate Metropolis-Hastings algorithms to draw samples from this distribution.

4.2.1 Hypothesis Testing for Stayers

Consider testing the point-null hypothesis: $H_0 : p_M = 1$ against $H_1 : p_M \neq 1$ corresponding to the mixing probability. In this case, H_0 corresponds to a particular submodel of interest: that model which reduces the number of mixture components from two to one. In the forest ecological setting, such a test is of scientific interest and relates to the existence of *infection resistant* trees. Examining (4.2) it is clear that $p_M = 1$ implies that the Markov component, $f_{MC}(\mathbf{y_i}|\boldsymbol{\Theta})$, completely describes the transition process and the point mass, $I(\mathbf{y_i} = \mathbf{0})$, corresponding to stayers, has no role to play.

Within the frequentist framework, testing for stayers in the mover-stayer model was considered by Albert (1999) in a purely longitudinal setting. There, a likelihood ratio (LR) test was employed. Such a test is slightly non-standard as the null hypothesis occurs on the boundary of the parameter space. The resulting asymptotic distribution of the LR statistic, under the null, was presented as a mixture of the standard chi-squared (with 1 d.f.) and a point mass at zero. An alternative testing procedure might have employed methods based on the score statistic. In our more complicated spatial-longitudinal setting, such an approach might be integrated within a Monte Carlo maximum likelihood framework such as that developed in Chapter 1. For example, a simulated version of the score test could be conceived. While such an approach warrants further investigation, we proceed here within the Bayesian framework.

The Bayesian approach to hypothesis testing is less formal and is typically based

on the posterior probability $Pr(H_0|\mathbf{Y}) = \int_{H_0} d\pi(\mathbf{\Theta}|\mathbf{Y})$, with large values providing evidence in favor of H_0. The posterior probability is an appealing measure of evidence; it is easily interpreted and communicated to subject matter specialists. A closely related approach employs the Bayes factor (Kass and Raftery 1995) which is defined as the posterior odds in favor of H_0 over the prior odds in favor of H_0: $BF = \frac{Pr(H_0|\mathbf{Y})}{Pr(H_1|\mathbf{Y})} / \frac{Pr(H_0)}{Pr((H_1)}$. In fact, when a fair prior, $Pr(H_0) = Pr(H_1) = \frac{1}{2}$, is employed (as we shall do here), the two approaches are equivalent. Assuming such a fair prior, Table 4.1, adapted from Raftery (1996), gives a scale for interpreting the posterior probability as a summary of evidence for H_0 provided by the data.

$Pr(H_0\|\mathbf{Y})$	evidence for H_0
$< \frac{1}{2}$	negative (supports H_1)
$\frac{1}{2}$ to $\frac{3}{4}$	barely worth mentioning
$\frac{3}{4}$ to $\frac{12}{13}$	positive
$\frac{12}{13}$ to $\frac{150}{151}$	strong
$> \frac{150}{151}$	very strong

Table 4.1: Calibration of the posterior probability assuming a fair prior $Pr(H_0) = Pr(H_1) = \frac{1}{2}$. The table is adapted from Raftery (1996), where it was use to calibrate the Bayes factor.

A critical issue is the choice of prior for p_M. A continuous prior distribution such as the Uniform$(0, 1)$ which we employed in Chapter 2, or more generally, a beta distribution are not compatible with a point-null hypothesis. Such continuous priors would imply that H_0 is a priori impossible. The data will not modify this absolute information and the posterior probability of H_0 will also be zero. As a solution, we adopt a discrete mass approach adopting the prior

$$p_M \sim \omega I(p_M = 1) + (1 - \omega)Beta(\alpha, \beta) \tag{4.9}$$

which is a mixture of a beta distribution with a point mass on the null value. Such discrete mass priors were considered by Geweke (1996) and George and McCullough

(1997) for variable selection problems in regression. Here, we set $\omega = 0.5$ corresponding to the fair prior mentioned above. By choosing a prior for p_M with a point mass at 1, prior probability is allocated to the submodel with only one mixture component.

Finally, we note that a simpler approach would avoid testing the point-null hypothesis altogether and simply examine the posterior distribution of p_M under the full mover-stayer model. Indeed, such an approach is advocated by Gelman et al. (2003 p.339) in a related problem. Along these lines, our analysis of the weevil infestation data in Chapter 2 produced a posterior 95% credible interval for p_M of $(0.86, 0.93)$. From this, we might conclude that the posterior places a sufficiently small amount of mass in the neighbourhood of $p_M = 1$ and there no need to consider the simpler model. Nevertheless, in other applications of our model, the posterior interval may not be so clearly bounded away from $p_M = 1$ and the choice between the full model and the submodel will be less clear. In addition, the scientific importance of the null hypothesis in our application seems to warrant a more rigorous approach.

4.2.2 Computational Implementation

Posterior summarization is based on MCMC samples drawn from (4.8) using a component-wise Metropolis-Hastings sampling algorithm. We present two such algorithms which have both been coded in the C programming language. Following the data augmentation strategy of Chapter 2, we introduce N latent allocation variables, $z_i \overset{iid}{\sim}$ Bernoulli(p_M), $i = 1, ..., N$, where each z_i allocates subject i into one of two mixture components

$$f_{MS}(\mathbf{y}_i|\mathbf{Z}, \boldsymbol{\Theta}) = \begin{cases} I(\mathbf{y}_i = \mathbf{0}) & \text{if } z_i = 0, \\ f_{MC}(\mathbf{y}_i|\boldsymbol{\Theta}) & \text{if } z_i = 1 \end{cases} \tag{4.10}$$

This alternative representation of the mixture model through the use of hidden allocation variables simplifies computation.

The full conditional distribution of each regression and B-spline coefficient is log-concave. As such, we employ adaptive rejection sampling obtaining draws from each full conditional distribution in a Gibbs update step. The initial state probability, p_I, is assigned a conditionally conjugate beta prior and is thus also updated in a Gibbs

step. The random effects are blocked into subject pairs and updated using a random walk Metropolis step employing a bivariate Gaussian proposal. As in Chapter 3, we assign the hyper-parameter, $\mathbf{\Sigma}$, a conditionally conjugate inverse-Wishart prior and draw directly from the full conditional distribution. The hyper-parameter κ is updated using a Metropolis-Hastings step employing a logit-normal proposal. In other words, we transform κ to the real line and update via random walk Metropolis using a Gaussian proposal (after re-transforming the proposal is not symmetric and hence this is a Metropolis-Hastings step). Each allocation variable, z_i, is easily updated by drawing directly from the corresponding full conditional distribution. Finally, careful consideration must be given to the update of the mixing probability, p_M, due to the discrete mass prior (4.9). We consider two different methods for updating p_M and this leads to two different algorithms. The first is based on a Metropolis-Hastings update where the allocation variables, z_i, $i = 1, ..., N$, have been integrated out. The second is based on a Gibbs update which includes the allocation variables.

A Metropolis-Hastings update may be employed with careful choice of the proposal distribution. For example, a standard random walk Metropolis step is not available here. Gottardo and Raftery (2004) discuss the measure theoretic details of defining a Markov transition kernel in this setting and more general settings involving Bayesian computation with mixtures of singular distributions. In our case, we must employ a proposal distribution which places an atom at $p_M = 1$. In doing so, the usual Metropolis-Hastings formula for the acceptance probability will yield a valid update step. This leads to our first algorithm, which we present below in full detail.

Algorithm 1

Denoting by $(p_M, \{z_i\}, p_I, \boldsymbol{b}, \mathbf{\Sigma}, \kappa, \boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\alpha_0}, \boldsymbol{\alpha_1})$ the current state of the chain, we follow steps 1 to 8 below. One iteration of the sampler consists of a complete sweep through the eight steps, at the end of which the new state is recorded.

1. Update p_M:

   If the current value of p_M does *not* equal 1, then follow step (a); otherwise, follow step (b).

(a) $\mathbf{p_M \neq 1}$: Propose a new value $p_M^* = 1$. The Metropolis-Hastings acceptance probability for this move, assuming the prior (4.9) is $p = \text{Min}(1, A)$ where

$$A = \frac{1}{p_M^{N-N_0}} \times \frac{\omega}{1-\omega} \times \frac{f(p_M|a,b)}{f(p_M|\alpha,\beta)} \times \prod_{\{i|\mathbf{y_i}=\mathbf{0}\}} \frac{(1-p_I)L_{0i}}{p_M(1-p_I)L_{0i} + (1-p_M)} \tag{4.11}$$

(b) $\mathbf{p_M = 1}$: Propose a new value p_M^* from a Beta(a,b) distribution. The Metropolis-Hastings acceptance probability for this move, assuming the prior (4.9) is $p = \text{Min}(1, A)$ where

$$A = (p_M^*)^{N-N_0} \times \frac{1-\omega}{\omega} \times \frac{f(p_M^*|\alpha,\beta)}{f(p_M^*|a,b)} \times \prod_{\{i|\mathbf{y_i}=\mathbf{0}\}} \frac{p_M^*(1-p_I)L_{0i} + (1-p_M^*)}{(1-p_I)L_{0i}} \tag{4.12}$$

In (4.11) and (4.12), $N_0 = \#\{i|\mathbf{y_i} = \mathbf{0}\}$; $f(\cdot|x,y)$ is the density of a beta distribution with mean $\frac{x}{x+y}$ and variance $\frac{xy}{(x+y)^2(x+y+1)}$; a and b are tuning parameters; and finally, L_{0i} is a likelihood contribution given by (4.4).

Overall, starting from p_M, the proposal sets the candidate value p_M^* to 1 if $p_M \neq 1$. Otherwise, the proposal generates a candidate $p_M^* \neq 1$ from a Beta(a, b) distribution. From theorem 1 of Gottardo and Raftery (2004) the density of this proposal is given by

$$q(p_M^*|p_M) = [1 - I(p_M = 1)]I(p_M^* = 1) + I(p_M = 1)f(p_M^*|a,b)[1 - I(p_M^* = 1)]$$

To set the tuning parameters a and b, we run a pilot analysis and fit a simpler model that assumes a uniform$(0, 1)$ prior for p_M. The values of a and b are then chosen so that the mean and variance of the corresponding beta distribution match the posterior mean and variance of p_M obtained from the pilot run.

2. Update z_i, $i = 1, ..., N$:

   If $p_M = 1$ set $z_i = 1$, $i = 1, ..., N$. Otherwise, the full conditional distribution of

each z_i is Bernoulli(p_{z_i}) where

$$p_{z_i} = \begin{cases} 1 & \text{if } \mathbf{y}_i \neq \mathbf{0}, \\ \frac{p_M(1-p_I)\prod_{t=1}^{n_i-1}[1+\exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i})]^{-1}}{(1-p_M)+p_M(1-p_I)\prod_{t=1}^{n_i-1}[1+\exp(\boldsymbol{\beta_0}'\boldsymbol{x_i(t)}+g_0(t,\boldsymbol{\alpha_0})+b_{0i})]^{-1}} & \text{if } \mathbf{y}_i = \mathbf{0} \end{cases} \tag{4.13}$$

We draw directly from this distribution in a Gibbs step.

3. Conditional update of p_M:

   If $p_M = 1$ we skip this step. Otherwise, we perform another update of p_M, by obtaining a draw from the full conditional distribution of p_M *given* $p_M \neq 1$. This full conditional is a Beta($\alpha + \sum_{i=1}^{N} z_i, N + \beta - \sum_{i=1}^{N} z_i$) distribution from which we draw directly.

   This conditional update, is not strictly necessary for the Markov chain to be ergodic; nevertheless, it is included with the aim of improving mixing of the sampler.

4. Update p_I:

   Assuming a uniform$(0, 1)$ prior, the full conditional distribution for p_I is Beta$(1+\sum_{i=1}^{N} y_i(0)z_i, 1+\sum_{i=1}^{N} z_i[1-y_i(0)])$ from which we draw directly in a Gibbs step.

5. Update random effects $\boldsymbol{b}$:

   We update the pair of random effects associated with each subject sequentially. The full conditional distribution of each pair, $\boldsymbol{b_{(i)}}$, $i = 1, ..., N$, has p.d.f. proportional to

$$L_{0i}^{z_i} \times L_{1i}^{z_i} \times \exp(-\frac{C_{i\cdot}}{2}(\boldsymbol{b_{(i)}} - \frac{\kappa}{C_{i\cdot}}\sum_j C_{ij}\mathbf{b_{(j)}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{b_{(i)}} - \frac{\kappa}{C_{i\cdot}}\sum_j C_{ij}\mathbf{b_{(j)}}))$$

   where L_{0i} and L_{1i} are given by (4.4) and (4.5) respectively. We use a random walk Metropolis step with candidate generated from a bivariate normal distribution.

6. Update variance components $\mathbf{\Sigma}$:

   A convenient prior for the 2×2 positive definite and symmetric matrix $\mathbf{\Sigma}$ is the conditionally conjugate inverse-Wishart having degrees of freedom $\nu \geq 2$ and positive definite scale parameter $\boldsymbol{A}$. The resulting full conditional distribution is also inverse-Wishart with degrees of freedom $\nu' = \nu + N$ and scale

$$\boldsymbol{A}' = (\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{D_{ij}} - \kappa \boldsymbol{C_{ij}})\boldsymbol{b_{(j)}}\boldsymbol{b_{(i)}}' + \boldsymbol{A^{-1}})^{-1}$$

   from which we can draw directly in a Gibbs step.

7. Update κ:

   We employ a uniform(0,0.99) prior for κ. The resulting full conditional distribution then has p.d.f. proportional to

$$f_c(\kappa) = [\prod_{i=1}^{N}(1 - \kappa\lambda_i)]$$

$$\times \exp(-\frac{1}{2}\text{trace}\{\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{D_{ij}} - \kappa \boldsymbol{C_{ij}})\boldsymbol{b_{(j)}}\boldsymbol{b_{(i)}}'\mathbf{\Sigma^{-1}}\}) \times I(0 \leq \kappa \leq 0.99)$$

   where $\lambda_1, ..., \lambda_N$ are the eigenvalues of $\mathbf{D^{-1}C}$. A candidate value κ^* is obtained by transforming κ to the real line and applying a random walk Metropolis step as follows:

   (a) Obtain a draw, x, from a normal$(0, c)$ distribution, where $c > 0$ is a tuning parameter.

   (b) Let $y^* = \log(\frac{\kappa}{0.99-\kappa}) + x$

   (c) Let $\kappa^* = 0.99 * \frac{\exp(y^*)}{1+\exp(y^*)}$

   The Metropolis-Hastings acceptance probability for this move, is $p = \text{Min}(1, A)$ where

$$A = \frac{f_c(\kappa^*)}{f_c(\kappa)} \times \frac{\kappa^*(0.99 - \kappa^*)}{\kappa(0.99 - \kappa)} \tag{4.14}$$

   The second term on the right hand side of (4.14) arises from the Jacobian of the transformation.

8. Update regression coefficients $\boldsymbol{\beta_0}$:

   We update each scalar parameter β_{0_k} separately. The full conditional distribution for β_{0_k} based on the prior $\beta_{0_k} \overset{iid}{\sim} N(\mathbf{0}, a^2)$ has p.d.f. proportional to

   $$\exp(-\frac{1}{2a^2}\beta_{0_k}^2) \times \prod_{\{i|z_i=1\}} L_{0i}$$

   and is easily shown to be log-concave. As such, we employ adaptive rejection sampling to obtain draws from this full conditional distribution in a Gibbs step. The parameters $\boldsymbol{\beta_1}, \boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_0}$ are updated analogously.

We note a connection between this algorithm and the reversible jump MCMC methodology proposed by Green (1995) for Bayesian model determination. In the reversible jump setting, a Markov chain is constructed for sampling over both parameter and model space where the dimension of the parameter space is allowed to vary from one model to another. In our problem, there are two competing models: the full spatial mover-stayer model and the model with only one mixture component obtained by setting $p_M = 1$. Within the reversible jump literature, step 1 of our algorithm would be referred to as a between-model move and all other steps would be termed within-model moves. As the two models are nested, the only dimension change necessary is the addition or deletion of p_M from the parameter vector. As such, the standard Metropolis-Hastings formula holds provided the proposal incorporates a point mass at $p_M = 1$. The reversible jump formalism is therefore not explicitly needed. Nevertheless, if such an approach were employed, we would obtain the same algorithm.

Algorithm 2

In step 1 of our first algorithm, the allocation variables, z_i, $i = 1, ..., N$, were integrated out. If they are included, the full conditional distribution for p_M is available

$$\text{full conditional } [p_M] \equiv p^* I(p_m = 1) + (1 - p^*)\text{Beta}(\alpha + \sum_{i=1}^{N} z_i, N + \beta - \sum_{i=1}^{N} z_i) \quad (4.15)$$

where p^* is the full conditional probability of H_0 and is given by

$$p^* = \frac{\omega I(z_i = 1, \forall i = 1, ..., N)}{\omega I(z_i = 1, \forall i = 1, ..., N) + (1 - \omega)\frac{B(\alpha + \sum_{i=1}^{N} z_i, N + \beta - \sum_{i=1}^{N} z_i)}{B(\alpha, \beta)}}$$

where $B(\cdot,\cdot)$ denotes the beta function. Our second algorithm, is obtained as a modification of the first by replacing step 1 with a Gibbs update. That is, we draw p_M directly from the full conditional (4.15). In addition, step 3, the conditional update of p_M in algorithm 1, is removed. All other steps remain the same.

To test our sampling algorithms, we analyze three synthetic datasets in Section 4.3. We then conduct a reanalysis of the weevil infestation data in Section 4.4.

4.3 Analysis of Synthetic Data

To test our MCMC algorithms, we fit our model to synthetic data. We simulate data from the spatial mover-stayer model under three different scenarios. Spatially, we assume that data are collected on a 20×20 regular lattice corresponding to 400 subjects under observation; where, associated with each subject is a position on the lattice given by coordinates (x, y), $x = 1, ..., 20$, $y = 1, ..., 20$. Data are generated for each subject over seven time points corresponding to $t = 0, ..., 6$. For all three simulation scenarios, we assume an initial state probability of $p_I = 0.15$ and time-homogeneous subject specific transition probabilities given by $\text{logit}(p_{01_i}(t)) = \beta_0 + b_{0i}$ and $\text{logit}(p_{10_i}(t)) = \beta_1 + b_{1i}$, $i = 1, ..., 400$, $t = 1, ..., 6$. Transitions are therefore governed by two intercepts, β_0 and β_1 and random effects which we assume follow the $2CAR(\kappa, \mathbf{\Sigma})$ model. We set the intercepts to $\beta_0 = -0.1$ and $\beta_1 = 0.1$. For the random effects, we assume a 1^{st} order neighbourhood system. That is, we set the weights, C_{ij}, in (4.7) to $C_{ij} = I\{d(i, j) \leq 1\}$ where $d(i, j)$ denotes the distance between subjects i and j. Depending on their position, each subject will therefore have either 4, 3 or 2 neighbours. The hyperparameters in the 2CAR model are set to $\kappa = 0.9$ and

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix}$$

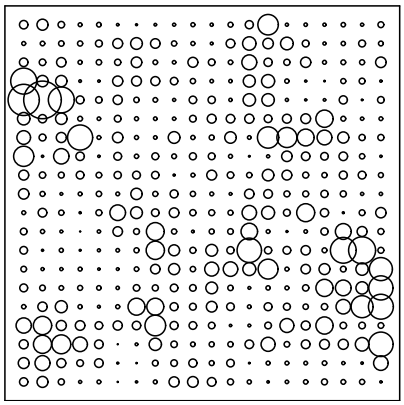
Using these values of the hyperparameters in conjunction with the full conditionals (4.6), we ran the Gibbs sampler to obtain a realization of the random effects $\mathbf{b}$. This realization is displayed in Figure 4.2, where we have mapped the simulated random effects on their corresponding lattice positions, and in Figure 4.3 where a scatter plot

illustrates the negative correlation between each pair $\mathbf{b_{(i)}} = (b_{0i}, b_{1i})'$ of simulated values. Finally, the three different data sets correspond to three different values for the mixing probability: $p_M = 0.90$, $p_M = 0.93$ and $p_M = 1.0$.

For each simulated dataset, we fit the extended spatial mover-stayer model assuming: $g_0(t, \boldsymbol{\alpha_0}) \equiv 0$ and $g_1(t, \boldsymbol{\alpha_1}) \equiv 0$; independent $N(0, 10^3)$ priors for the two intercepts; a Uniform$(0, 1)$ prior for p_I; a Uniform$(0, 0.99)$ prior for κ; a vague inverse-Wishart$(\nu, \mathbf{A})$ prior $\boldsymbol{\Sigma}$, with $\nu = 2$ and $\mathbf{A} = \mathbf{I}$; and finally, we take $\alpha = \beta = 1$ in the discrete mass prior (4.9) assigned to p_M. We apply both of our sampling algorithms to obtain draws from the corresponding posterior distribution. For each algorithm, we run one chain for 50,000 burn-in iterations, followed by an additional 200,000 iterations used for posterior summarization. Posterior 95% credible intervals for each parameter are given in Table 4.2 along with estimates of $Pr(p_M = 1|\mathbf{Y})$. These estimates are obtained from the Monte Carlo output as $\hat{Pr}(H_0|\mathbf{Y}) = \frac{1}{J}\sum_{j=1}^{J} I(p_M^{(j)} = 1)$, where $p_M^{(j)}$, $j = 1, ..., J$, are the posterior draws of p_M. The Monte Carlo error for this estimate is computed based on the method of batch means.

In each case, both algorithms give similar results. Regarding the model parameters, the results are in line with our expectations as all 95% credible intervals cover the true values. When data are simulated with a true value $p_M = 1$ (simulation 3), support for the null hypothesis $H_0 : p_M = 1$, as measured by the posterior probability $Pr(p_M = 1|\mathbf{Y})$, is strong. In simulation 2, where the true value is set to $p_M = 0.93$, the evidence for H_0 is negative as the posterior supports $p_M \neq 1$. Finally, we note that in simulation 1, where the true value of p_M was set to its smallest value of 0.90, neither of the two algorithms produce a single iteration where $p_M = 1$ (resulting in $\hat{Pr}(p_M = 1|\mathbf{Y}) = 0$). This is presumably due to an essentially negligible amount of posterior mass placed on this value for data simulated under this scenario. In all three cases, results seem reasonable and intuitive given the simulation settings. As our two algorithms employ vastly different mechanisms for updating p_M, we are encouraged by the similarity of the estimates, $\hat{Pr}(H_0|\mathbf{Y})$, obtained from each one.

a)



b)

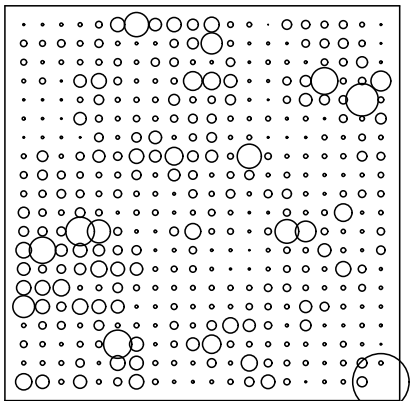


Figure 4.2: Maps of simulated random effects. Each map plots the exponential of the simulated random effect as a circle, where larger circles correspond to larger values (a) $\exp(b_{0i})$, $i = 1, ..., 400$; (b) $\exp(b_{1i})$, $i = 1, ..., 400$.

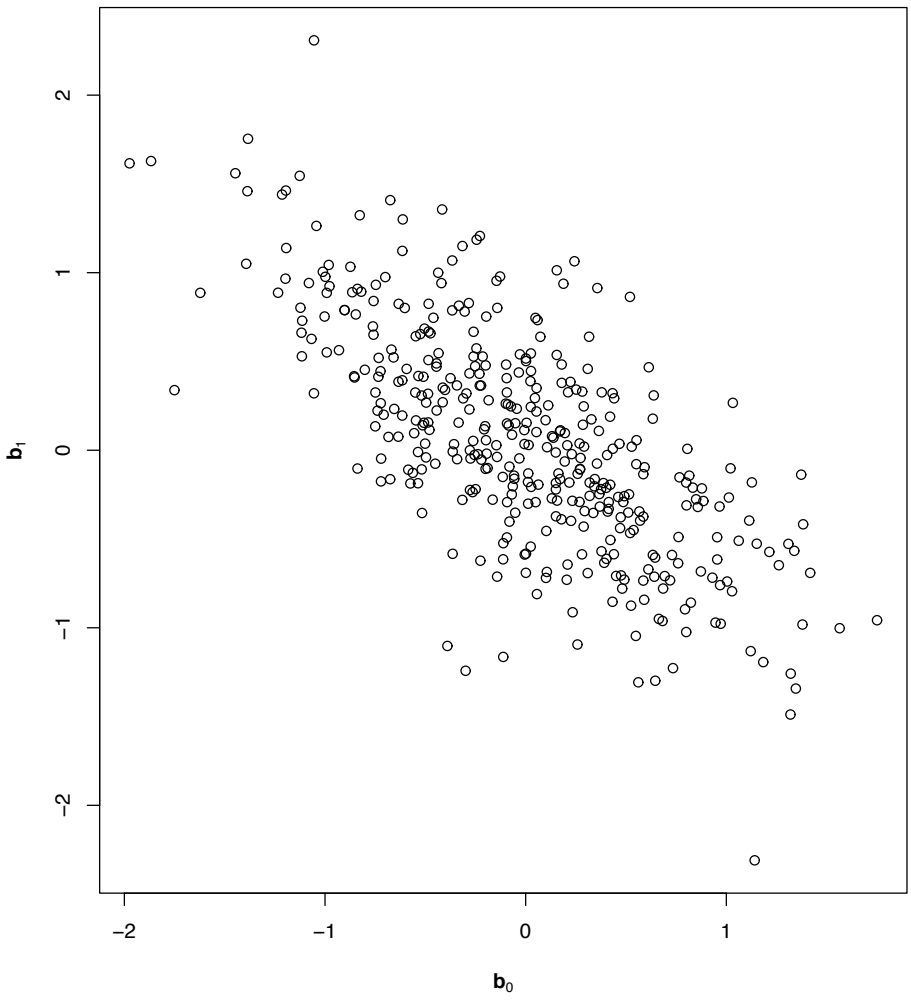


Figure 4.3: Scatter plot comparing simulated values of $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$.

Parameter	Simulation 1 ($p_M = 0.90$)		Simulation 2 ($p_M = 0.93$)		Simulation 3 ($p_M = 1$)	
	Algorithm 1	Algorithm 2	Algorithm 1	Algorithm 2	Algorithm 1	Algorithm 2
Σ_{11}	(0.48,1.46)	(0.51,1.49)	(0.48,1.51)	(0.51,1.51)	(0.60,1.28)	(0.58,1.29)
Σ_{22}	(0.63,1.49)	(0.61,1.52)	(0.50,1.25)	(0.47,1.20)	(0.73,1.57)	(0.70,1.63)
Σ_{12}	(-0.90,0.20)	(-0.87,0.18)	(-0.84,0.16)	(-0.83,0.11)	(-0.89,-0.14)	(-0.90,-0.15)
κ	(0.25,0.96)	(0.23,0.96)	(0.53,0.99)	(0.58,0.99)	(0.51,0.97)	(0.51,0.97)
β_0	(-0.24,0.13)	(-0.26,0.12)	(-0.38,0.13)	(-0.41,0.12)	(-0.25,0.13)	(-0.26,0.11)
β_1	(-0.25,0.15)	(-0.25,0.18)	(-0.28,0.21)	(-0.25,0.20)	(-0.10,0.37)	(-0.091,0.40)
p_I	(0.15,0.23)	(0.15,0.23)	(0.10,0.17)	(0.10,0.17)	(0.12,0.19)	(0.12,0.19)
p_M	(0.83,0.91)	(0.83,0.91)	(0.93,1.0)	(0.93,1.0)	(1.0,1.0)	(1.0,1.0)
$Pr(p_M = 1\|\mathbf{Y})$	0 (0)	0 (0)	0.39 (0.034)	0.35 (0.04)	0.99 (3.4×10^{-4})	0.99 (9.9×10^{-4})

Table 4.2: Posterior summaries obtained from the analysis of three simulated datasets. For each parameter we give the 95% credible interval. Estimates of $Pr(p_M = 1|\mathbf{Y})$ are also given along with Monte Carlo standard errors (computed using the method of batch means employing 100 batches each of size 2000).

4.4 Analysis of Tree Infection data

In this section we build upon the analysis of the weevil infestation data conducted in Chapter 2 by fitting the extended spatial mover-stayer model developed in this chapter. We employ the same regression regression specifications as Chapter 2

$$\text{logit}(p_{01_i}(t)) = \beta_{0_1} D_i + \beta_{0_2} A_i(t-1) + g_0(t, \boldsymbol{\alpha_0}) + b_{0i}$$

$$\text{logit}(p_{10_i}(t)) = \beta_{1_1} D_i + \beta_{1_2} A_i(t-1) + g_1(t, \boldsymbol{\alpha_1}) + b_{1i}$$

$$i = 1, ..., 2662, \;\; t = 1, ..., 6$$

where, D_i is a local measure of tree density; $A_i(t-1)$ is a local measure of infection density; and $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ are cubic B-splines with one inner knot placed at $t = 4$ years. For the random effects, we define the weights in the $2CAR(\kappa, \boldsymbol{\Sigma})$ model by $C_{ij} = I\{d(i,j) \leq 10m\}$, as justified in Chapter 2. Regarding prior distributions, we assume a Uniform$(0, 1)$ prior for p_I and assign independent $N(0, 10^3)$ priors for the regression coefficients and parameters of cubic spline terms. The hyperparameters, κ and $\boldsymbol{\Sigma}$, are assigned Uniform$(0, 0.99)$ and inverse-Wishart$(\nu, \mathbf{A})$ priors respectively where we again set $\nu = 2$ and $\mathbf{A} = \mathbf{I}$. Finally, for the discrete mass prior assigned to p_M, we take $\alpha = \beta = 1$. The prior is then a mixture of a point mass and a Uniform$(0, 1)$ distribution.

We fit the model using both of our MCMC algorithms. For each algorithm, we ran two initially overdispersed sampling chains in parallel. With regards to initial values for p_M, one chain was initialized at $p_M = 1$ while the other was initialized at $p_M = 0.89$, the point estimate obtained from the Chapter 2 analysis. Each chain was run for an initial 50,000 burn-in iterations followed by an additional 500,000 iterations. This resulted in a total of 1,000,000 iterations to be used for posterior inference. As with our analysis of the simulated data in Section 4.3, the results obtained from both algorithms are virtually identical. As such, we only present the results obtained from algorithm 1. Posterior summaries of model parameters are given in Table 4.3. Results regarding regression coefficients and the initial probability, p_I, are very similar to those obtained from the simpler model fit in Chapter 2 (see Table 2.1 for comparison). Estimates of the temporal trends are shown in Figure 4.4. These are also similar to

	Uninfected → Infected		Infected → Uninfected	
Parameter	Mean	Sd	Mean	Sd
Intercept	−0.52	0.15	0.027	0.21
D_i	−0.018	0.0032	0.0092	0.0047
$A_i(t-1)$	0.013	0.0079	−0.0062	0.011
σ_{b0}/σ_{b1}	2.86	0.35	3.26	0.48
p_I	0.18	0.0084	–	–
p_M	0.90	0.017	–	–
ρ	−0.96	0.030	–	–
κ	0.98	0.015	–	–

Table 4.3: Posterior summaries obtained from fitting the extended spatial mover-stayer model to the weevil infestation data. Here, we have defined $\sigma_{b0} = \sqrt{\Sigma_{11}}$, $\sigma_{b1} = \sqrt{\Sigma_{22}}$ and $\rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$.

the estimates obtained from the simpler model (see Figure 2.5 for comparison). The estimate of κ is close to its upper bound, indicating that an intrinsic autoregression, which sets $\kappa = 1$, is likely sufficient in this case. Regarding variance components, the posterior distribution of the conditional correlation parameter, $\rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$, indicates a strong negative correlation within each pair $\boldsymbol{b_{(i)}} = (b_{0i}, b_{1i})$. A 95% credible interval for ρ is given by $(-0.99, -0.88)$. This strong correlation is also evident in Figure 4.1b where we compare posterior mean estimates of b_{0i} and b_{1i}. Allowing for such correlation has attenuated point estimates of the variance components, $\sqrt{\Sigma_{11}}$ and $\sqrt{\Sigma_{22}}$, as both estimates are larger compared to those obtained in Chapter 2. Thus it seems that the bivariate spatial model, by allowing for the borrowing of information across the two transition types, indicates a greater level of heterogeneity than was initially estimated in Chapter 2, where the simplifying independence assumption was made. This increased variability is also evident when comparing Figures 4.1a and 4.1b as the estimates of $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ obtained from the extended model encompass a larger range than those obtained in Chapter 2. In Figure 4.5 we have mapped the extreme random effects, locating those trees having the 100 largest and 100 smallest estimates of $\boldsymbol{b_0}$ (4.5a) and $\boldsymbol{b_1}$ (4.5b). As before, spatial clustering of the largest and smallest values is apparent; however, compared to the maps obtained in Chapter 2 (see Figure

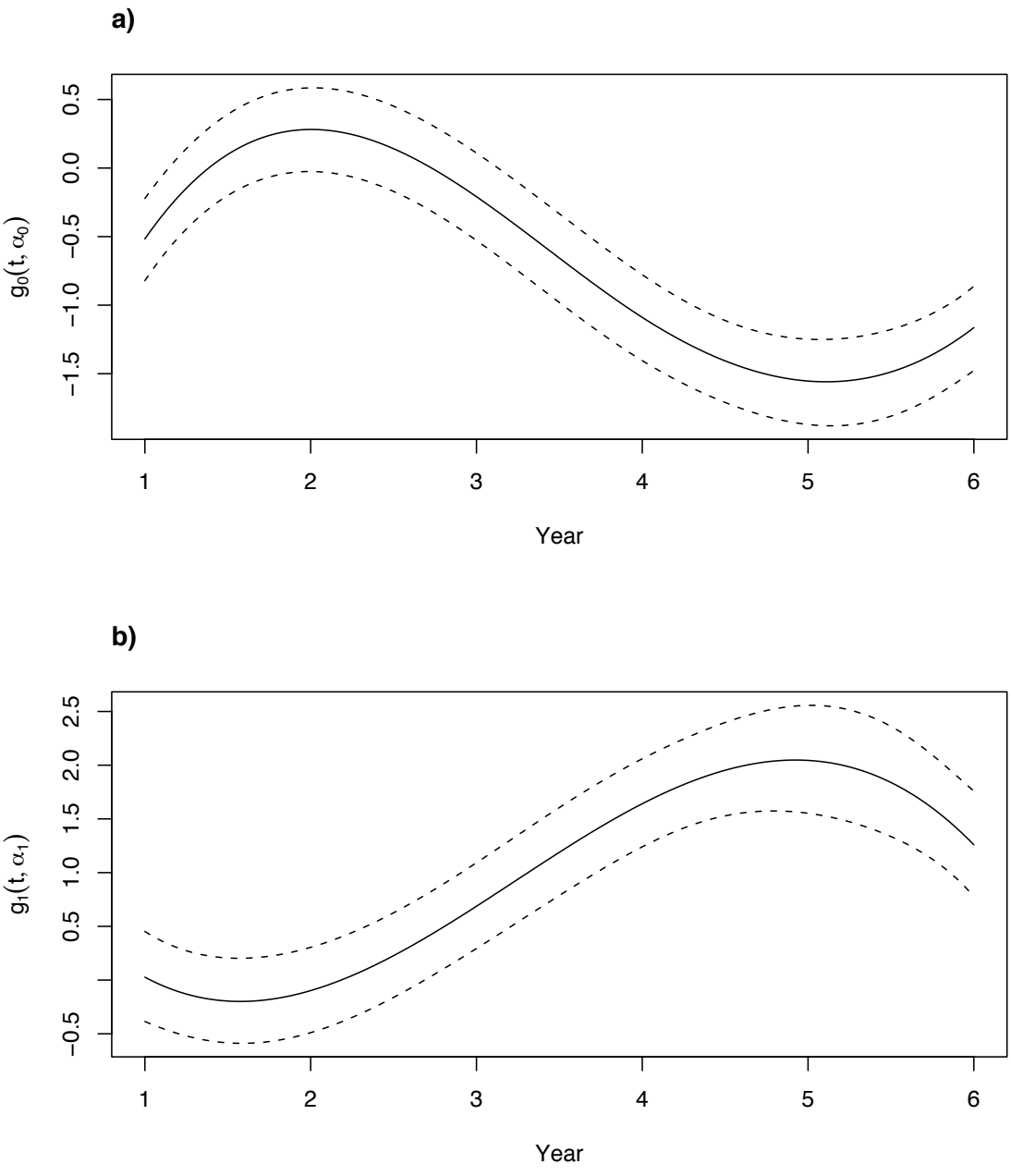


Figure 4.4: Posterior mean estimates of the temporal trends with 95% credible sets (a) $g_0(t, \boldsymbol{\alpha_0})$; (b) $g_1(t, \boldsymbol{\alpha_1})$.

2.6 for comparison), the map of extreme $\hat{\boldsymbol{b}_0}$ corresponds more closely to that of $\hat{\boldsymbol{b}_1}$; a result owing to the correlation we have allowed between the two. In Figure 4.6 we locate those trees which may be resistant. Figure 4.6a locates those 100 trees which, based on our model, have the highest estimated posterior probability of resistance, $Pr(z_i = 0|\mathbf{Y})$, and Figure 4.6b simply locates all trees which were never infected. The latter is easily obtained but not very informative; whereas, the former selects a subset of trees whose known characteristics can not adequately explain a long period without infection. These trees are the prime candidates for further study.

For testing the hypothesis $H_0 : p_M = 1$, we obtain an estimate of the posterior probability $\hat{Pr}(p_M = 1|\mathbf{Y}) = 2.6\times10^{-5}$ with a Monte Carlo standard error of 2.5×10^{-5} (based on a batch means estimate employing 100 batches each of size 10,000). Even after accounting for Monte Carlo error, it is clear that the posterior lends very little support to the null hypothesis.

Finally, we compare the $2CAR(\kappa, \boldsymbol{\Sigma})$ specification for the random effects with two submodels. The first sets $\kappa = 1$ leading to a bivariate intrinsic autoregression. The second sets $\kappa = 1$ and $\boldsymbol{\Sigma} = \text{diag}\{\sigma^2_{b0}, \sigma^2_{b1}\}$ leading to the original model presented in Chapter 2. As in Chapter 3, we use the DIC criterion to compare across random effect specifications. Table 4.4 lists the DIC and p_D values for the three models considered. The lowest DIC score belongs to the full model; however, the DIC score associated with the intrinsic version which sets $\kappa = 1$ is very close and the two models can not be distinguished. This is in agreement with our estimate of κ in Table 4.3 which was very close to its upper bound. Finally, allowing for a joint spatial structure appears to be beneficial as both models are preferred over the original model which assumes $\boldsymbol{b_l} \overset{ind}{\sim} IAR(\sigma_l),\ l = 0, 1$.

4.4.1 Model Validation

In this section we discuss how to employ Bayesian model checking for our spatial mover-stayer model using the posterior predictive distribution. As before, we let $\mathbf{Y}$ denote the observed data, $L(\mathbf{Y}|\boldsymbol{\Theta}) = \prod_{i=1}^{N} f_{MS}(\mathbf{y_i}|\boldsymbol{\Theta})$ denote the (normalized) likelihood of the data under our model and $\pi(\boldsymbol{\Theta}|\mathbf{Y})$ denote the posterior distribution.

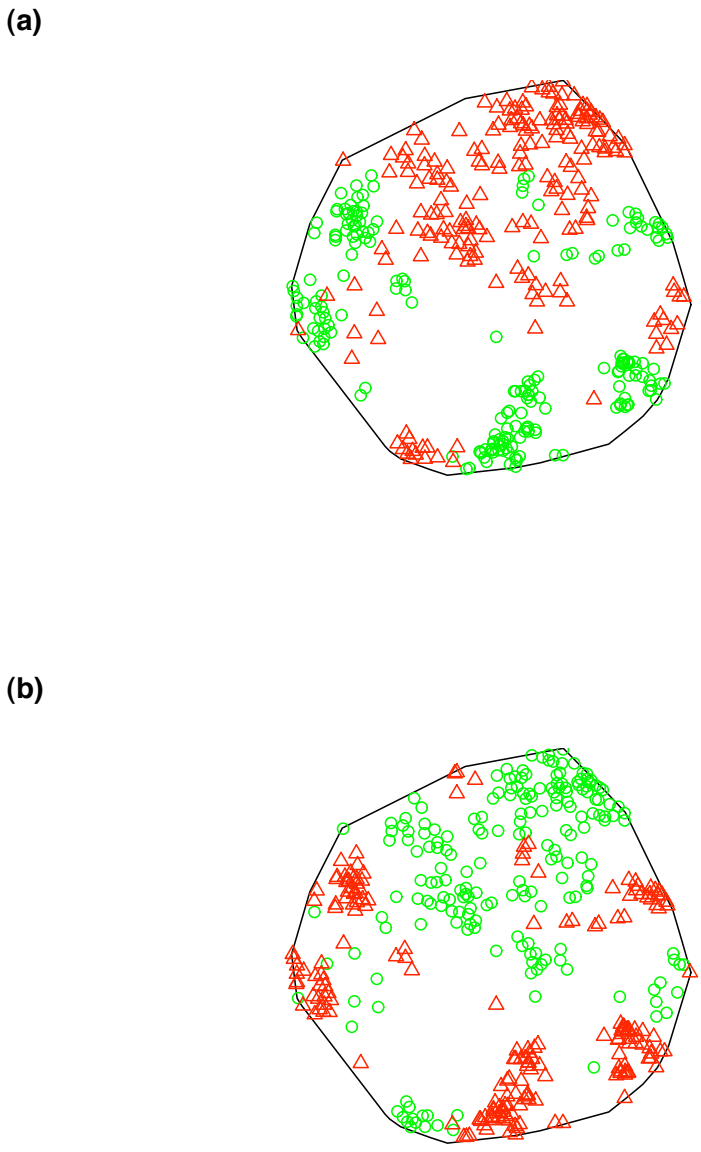


Figure 4.5: Locations of the 100 largest (triangles) and smallest (circles) estimated random effects from the extended spatial mover-stayer model (a) $\boldsymbol{b_0}$; (b) $\boldsymbol{b_1}$.

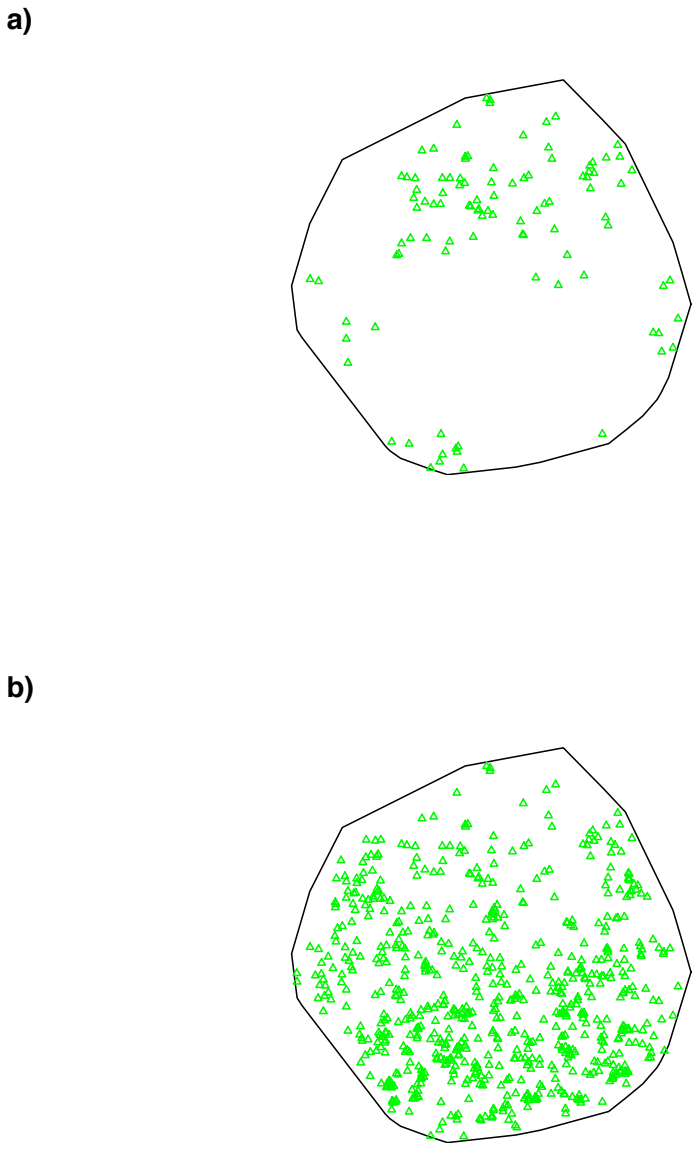


Figure 4.6: Locations of trees which may be resistant (a) Those 100 trees having the highest posterior probability of resistance $Pr(z_i = 0|\mathbf{Y})$; (b) Those 715 trees which were never infected.

Model for random effects	p_D	DIC
$2CAR(\kappa, \boldsymbol{\Sigma})$	469	679
$2CAR(1, \boldsymbol{\Sigma})$	406	683
$\boldsymbol{b_l} \overset{ind}{\sim} IAR(\sigma_l),\ l = 0, 1$	254	759

Table 4.4: DIC scores and p_D for three models fit to the weevil infestation data which consider different structures for the random effects.

Posterior predictive model checking is based upon the notion of a hypothetical replication of the data, $\mathbf{Y^{rep}}$, drawn from the sampling distribution assumed by our model, $L(\mathbf{Y}|\boldsymbol{\Theta})$, and under the same conditions as the observed data. If it were possible to simulate $\mathbf{Y^{rep}}$ from $L(\mathbf{Y^{rep}}|\boldsymbol{\Theta})$ we could compare the observed data, $\mathbf{Y}$, to the data simulated under our model, $\mathbf{Y^{rep}}$, to assess how well the hypothesized model replicates the observed data. Unfortunately, $\boldsymbol{\Theta}$ is unknown. A natural approach within the Bayesian setting is to integrate out $\boldsymbol{\Theta}$ against the posterior $\pi(\boldsymbol{\Theta}|\mathbf{Y})$. This yields

$$P(\mathbf{Y^{rep}}|\mathbf{Y}) = \int L(\mathbf{Y^{rep}}|\boldsymbol{\Theta}) d\pi(\boldsymbol{\Theta}|\mathbf{Y}) \tag{4.16}$$

the posterior predictive distribution.

To simulate from (4.16) we first obtain J draws from the posterior $\boldsymbol{\Theta_1}, ..., \boldsymbol{\Theta_J}$. Using these draws, we then simulate J replicate datasets, $\mathbf{Y_1^{rep}}, ..., \mathbf{Y_J^{rep}}$, where $\mathbf{Y_i^{rep}}$ is drawn from the sampling distribution assumed by our model, given the simulated parameters $\boldsymbol{\Theta_i}$. As discussed by Gelman et al. (1996), if the model is reasonably accurate, the replications should look similar to the observed data.

To compare the observed data to the data simulated under the model, we define a test statistic, $T(\mathbf{Y})$, and compare the observed value to the simulated values $T(\mathbf{Y_1^{rep}}), ..., T(\mathbf{Y_J^{rep}})$. Such comparisons can be accomplished visually by comparing the histogram of simulated values to the observed value. If the data are in conflict with the proposed model, the observed value, $T(\mathbf{Y})$, will lie in or beyond the tails of the histogram. Several test statistics can be chosen, each reflecting a particular aspect of the model we wish to check. To examine the fit of our model, we define four

test statistics

$$T_1(\mathbf{Y}) = \sum_{i=1}^{N} \sum_{t=1}^{n_i-1} [1 - y_i(t-1)] y_i(t)$$

$$T_2(\mathbf{Y}) = \sum_{i=1}^{N} \sum_{t=1}^{n_i-1} y_i(t-1)[1 - y_i(t)]$$

$$T_3(\mathbf{Y}) = T_1(\mathbf{Y}) + T_2(\mathbf{Y})$$

$$T_4(\mathbf{Y}) = \sum_{i=1}^{N} I(\mathbf{y_i} = \mathbf{0})$$

Here, $T_1(\mathbf{Y})$ is the total number of transitions from state 0 to state 1 (summed over all subjects); $T_2(\mathbf{Y})$ is the total number of transitions from state 1 to state 0; $T_3(\mathbf{Y})$ is the overall total number of transitions; and $T_4(\mathbf{Y})$ is the total number of subjects who remain in state 0 throughout the course of observation.

To check the fit of our extended mover-stayer model to the weevil infestation data, we drew $J = 500$ replicate datasets from the posterior predictive distribution, $\mathbf{Y_1^{rep}}, ..., \mathbf{Y_{500}^{rep}}$, and obtained the corresponding histograms for each of the four test statistics. The histograms are displayed in Figure 4.7 along with the observed values. Comparing the observed values to the histograms, we see no indication of a lack of fit. For the purpose of comparison, we fit the submodel obtained by setting $p_M = 1$ with all other aspects of the model remaining the same. Posterior predictive comparisons for this model, based on $J = 500$ replications are shown in Figure 4.8. It is clear from Figure 4.8d that the submodel is a poor fit with respect to the total number of trees which were never infected. In addition, by not allowing for trees that never make transitions between the diseased and disease-free states ('stayers'), the predicted number of $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions under the submodel is substantially greater than the number of such transitions in the observed data. These results are consistent with our results concerning the posterior probability $Pr(p_M = 1|\mathbf{Y})$.

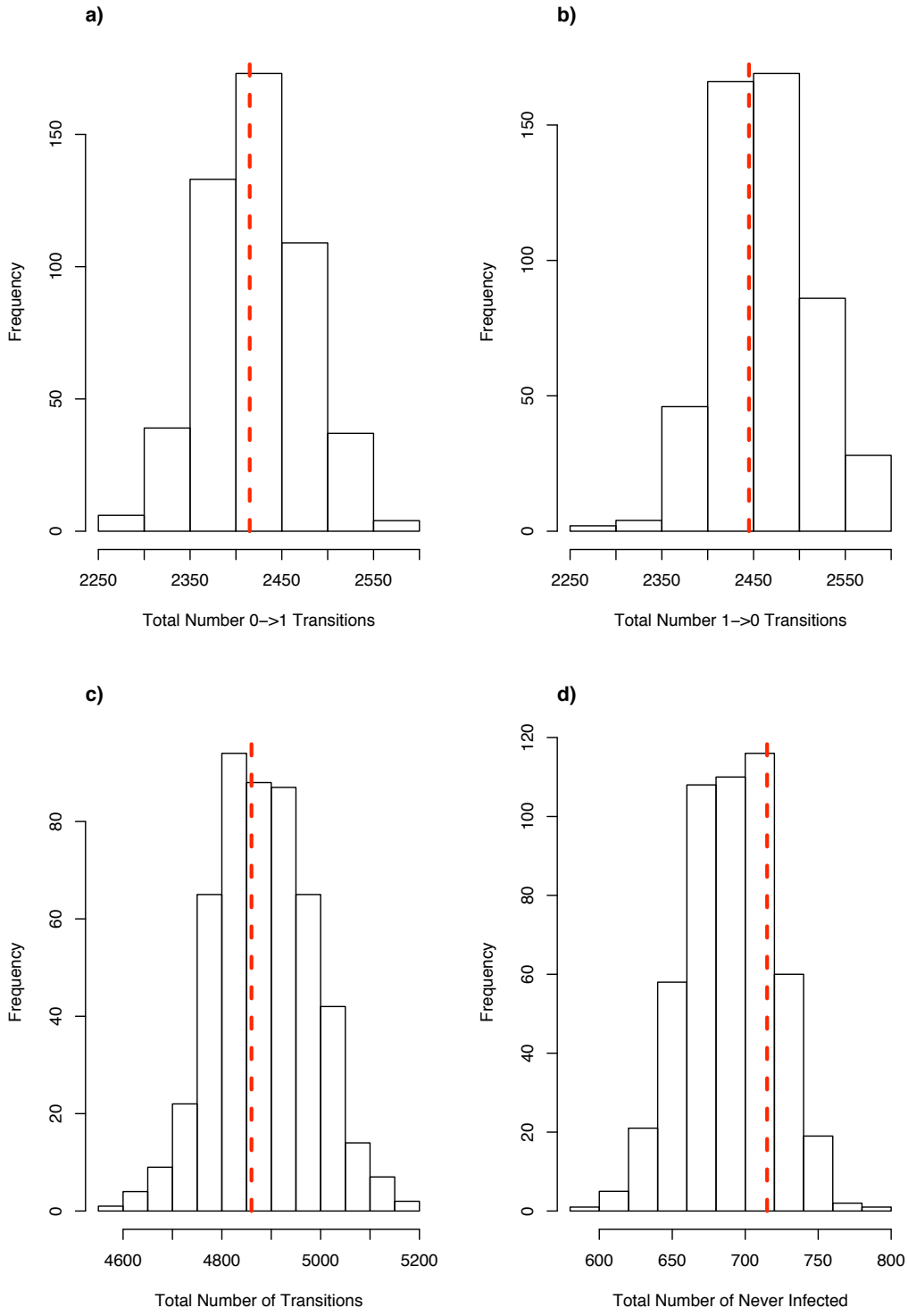


Figure 4.7: Posterior predictive distributions obtained from the extended spatial mover-stayer model (a) $T_1(\mathbf{Y})$ - the total number of $0 \rightarrow 1$ transitions; (b) $T_2(\mathbf{Y})$ - the total number of $1 \rightarrow 0$ transitions; (c) $T_3(\mathbf{Y})$ - the overall total number of transitions; (d) $T_4(\mathbf{Y})$ - total number of trees which were never infected; the dashed vertical line on each histogram represents the observed value of the test statistic.

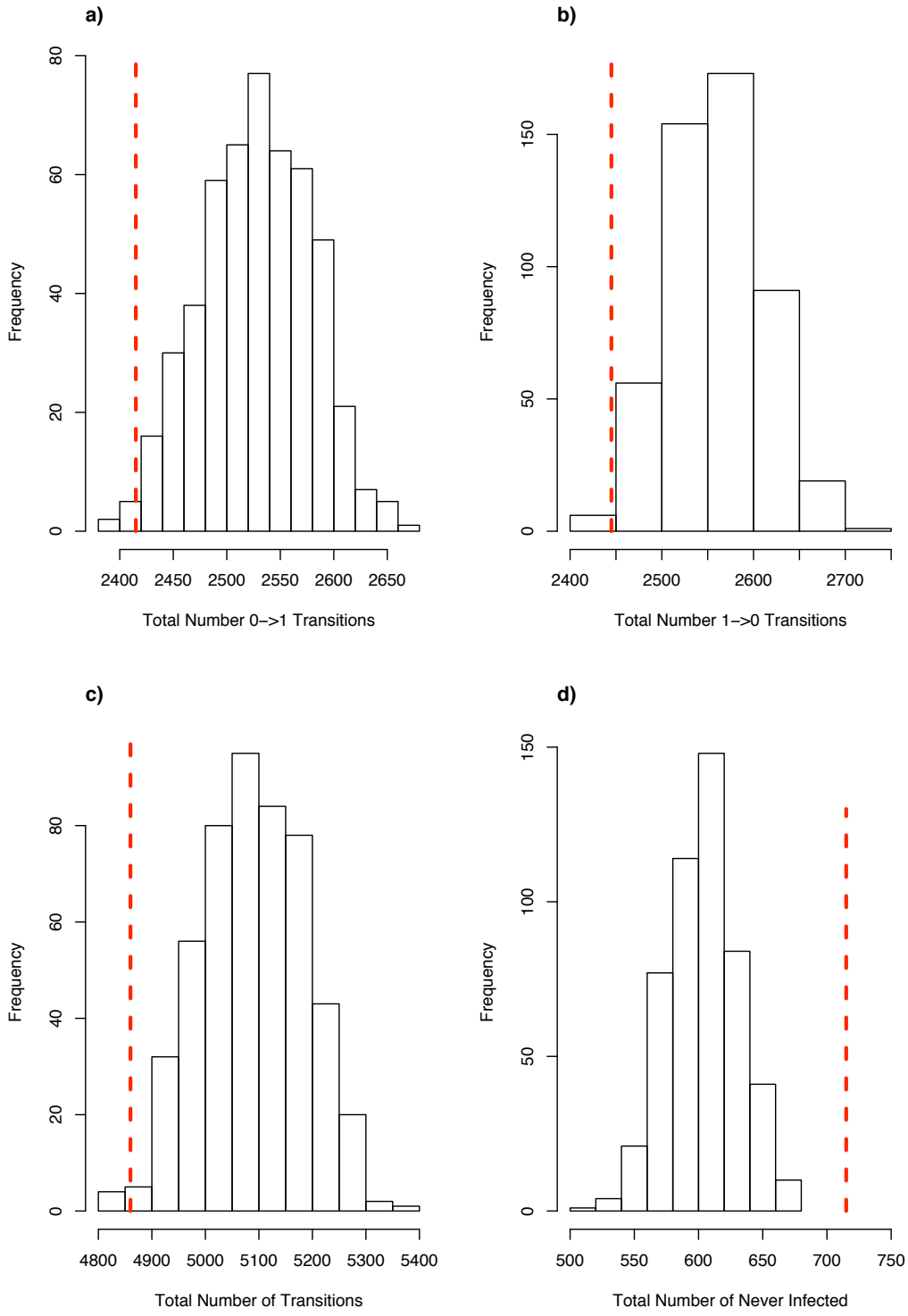


Figure 4.8: Posterior predictive distributions obtained from the submodel which sets $p_M = 1$ (a) $T_1(\mathbf{Y})$ - the total number of $0 \rightarrow 1$ transitions; (b) $T_2(\mathbf{Y})$ - the total number of $1 \rightarrow 0$ transitions; (c) $T_3(\mathbf{Y})$ - the overall total number of transitions; (d) $T_4(\mathbf{Y})$ - total number of trees which were never infected; the dashed vertical line on each histogram represents the observed value of the test statistic.

4.5 Discussion

This Chapter has presented an extension of the spatial mover-stayer model initially proposed in Chapter 2. A joint spatial model has been incorporated, allowing for correlation within each pair of subject specific random effects. In addition, a hypothesis test for stayers was developed. In our application this test relates to the existence of infection resistance trees. A discrete mass prior was incorporated for the mixing probability and this enabled the calculation of the required posterior probability. Two MCMC algorithms were developed to draw samples from the posterior distribution of our model. A reanalysis of the weevil infestation data was conducted based on the extended model. This analysis revealed a strong negative correlation within each pair of tree specific random effects. From this, we can conclude that missing covariates may which underlie the spatial autocorrelation are likely to have opposite effects on the two types of transitions. Applying our hypothesis test, we found overwhelming evidence in favor of the model containing two mixture components and thus it seems viable that some trees may be resistant to infection. Finally, we discussed a goodness-of-fit procedure based on the posterior predictive distribution. Applying this to our analysis of the weevil infestation data, we found no evidence indicating a lack of fit for the mover-stayer model.

The mover-stayer model developed for our forestry application assumes that those trees which make transitions ('movers') between the infected and uninfected states do so without restriction. Future work will consider extensions allowing for subjects who, upon infection, are no longer susceptible to reinfection. Tests for such monotonicity in infection might also be developed using the discrete-mass prior approach. Moving beyond the random intercepts employed here, a more general model allowing covariate effects to vary from subject to subject would be interesting to explore. Other extensions will incorporate spatial adaptive splines and allow for spatial correlation in mover-stayer allocations. The next chapter discuss these extensions in detail.

Chapter 5

Future Work

Directions for future work will consider refinements of the multi-state models developed here as well as new directions in the arena of mixture models.

5.1 Spatially Correlated Mover-Stayer Allocations

In Chapters 2 and 4 of this thesis we have examined methods for handling mixtures of populations for spatial discrete-time two-state processes. There, spatial correlation was imbedded into the transitional process of a two-state Markov chain. This, in turn, was then fused into a two-component mixture model. Allocation of subjects to each of the mixture components was assumed spatially independent. While this was suited to the forestry application considered, more general forms of mixture allocation could be constructed which would serve usefully in many other situations, for example, incorporating spatial clustering into the mixture allocations. In the forest ecological setting, this relates to spatial clusters of infection-resistant trees. In the spirit of Fernandez and Green (2002), such spatial allocation to mixture components could be accomplished through a logistic transformation of an autoregressive Gaussian process. Alternatively, a latent autologistic process may be used to govern mixture allocations. Tests for spatial correlation in the mixture allocations will be useful in this context. Methods based on Bayes factors (Kass and Raftery 1995) could be employed. As discussed by Sinharay and Stern (2005), computation of Bayes factors

in high dimensional spatial settings is challenging. Several techniques for computing Bayes factors in this context might be considered: Bridge sampling (Meng and Wong 1996), Reversible jump MCMC (Green 1995), Importance sampling and the MCMC based approaches proposed by Chib and Jeliazkov (2001).

5.2 Spatial Adaptive Splines and P-Splines

Regression splines are an attractive approach for modelling nonlinear smooth trends. The methodology is based on the representation of such trends through a sequence of piecewise polynomials. A crucial problem with regression splines is the choice of the number and positions of the knots defining the piecewise polynomials. In this thesis, where splines have been used for modelling temporal trends and intensity functions, we have addressed this problem primarily through exploratory methods, fixing the number of knots and their positions. Alternative, more rigorous, approaches will incorporate sophisticated techniques developed in the smoothing literature.

In the context of additive models, Denison et al. (1998) set up a joint distribution over the number and location of the knots and conduct inference in the Bayesian framework using Reversible Jump MCMC (Green 1995). The resulting adaptive splines avoid over-fitting resulting from the incorporation of too many knots while maintaining sufficient flexibility. An alternative approach is based on penalized regression splines (P-splines) and was considered in the context of Bayesian additive mixed models by Lang and Brezger (2004). In the P-spline approach, a moderate number of equally spaced knots are chosen and sufficient smoothness of the fitted curve is achieved through a difference penalty imposed on adjacent spline coefficients. Both techniques will be investigated and incorporated into methodology for spatial multi-state processes. In addition, extensions will allow temporal trends to vary by location in a smooth manner. That is, region specific temporal trends will be incorporated by assuming that the spline coefficients are themselves a realization of a spatial process. Such an extension would prove useful for analysis in longitudinal agricultural studies where growth curves or incidence rates vary substantially across the region under study. As discussed by Gelfand et al. (2003) and Gamerman et al. (2002),

who considered similar ideas in the context of space-varying regression models for Gaussian data, a multivariate spatial process would be required to jointly model all spline coefficients. Methods for adaptive space-varying splines based on multivariate spatial processes have not been considered previously. Such methods would represent a significant contribution to the spatio-temporal modelling literature.

5.3 Accelerated Failure Time Models with Spatial Frailties

The modelling of clustered or multivariate event-time data arises in many applications of statistics including, but not limited to: health, biology, ecology, demography as well as industrial applications (see e.g., Hougaard 2000). A popular approach for modelling such data introduces random effects or 'frailties' into the widely used proportional hazards regression model. Recently, several authors have extended the proportional hazards frailty model to the spatial setting by incorporating spatial correlation in the random effects. These include Li and Ryan (2002) who develop the proportional hazards spatial frailty model in the classical framework and Henderson et al. (2002) and Banerjee et al. (2003) who develop Bayesian methods. These methods were then further extended by Carlin and Banerjee (2002) and Jin and Carlin (2005) who develop joint spatial models based on multivariate spatial mixing distributions. Along these lines, the continuous-time models presented in Chapter 4 of this thesis are an extension of these methods to the more general state-space encompassed by multi-state models. All such work has been based upon the proportional hazards regression framework.

An important alternative to the proportional hazards model is the accelerated failure time (AFT) model (Kalbfleisch and Prentice 1980). Models in this framework are appealing due to their ease of interpretability. The basic models in this class assume observations are independent and adopt parametric distributional forms. More flexible AFT models adopt a semi-parametric approach and avoid distributional assumptions. In the Bayesian setting, semiparametric AFT regression models for univariate survival

times have been considered by, among others, Christenson and Johnson (1988) and Kuo and Mallick (1998), who develop methods based on the Dirichlet process and Walker and Mallick (1998) who employ a Polya tree prior for the error distribution. More recently, Komrek and Lesaffre (2004) develop a semiparametric approach which models the error distribution as a normal mixture with an unknown number of components and further allow for multivariate event-times through the inclusion of a random effect. Accelerated failure time models for multivariate event-time data have not been considered in the spatial setting. These spatial AFT models would serve usefully in forest ecological applications for examining the survival times of white spruce trees with respect to infection by the white pine weevil. Indeed, ecologists (see e.g., He and Alfaro 2000) have used simplified parametric AFT survival models for precisely this purpose but have ignored the possibility of spatial correlation. Extensions which develop semiparametric AFT models incorporating spatial frailties would prove extremely useful for the analysis of forest ecological time-to-event data. Various specifications for the spatial random effects might be explored and compared including: Gaussian Markov random fields, non-Gaussian Markov random fields (pairwise interaction random fields employing absolute value or log cosh potential functions) and Gaussian geostatistical forms.

5.4 Spatial Finite Mixtures

In this thesis, spatial correlation in the transition processes of multi-state models has been represented by spatially correlated random effects. Continuous mixture models of this sort have been widely used for modelling spatial and spatio-temporal correlation in many other settings. In the context of generalized linear mixed models, a broad framework for spatio-temporal analysis has been developed over the last decade through numerous publications in the literature: see for example, Waller et al. 1997, Diggle et al. 1998, Zhang 2002, Zhu et al. 2005 and references therein. In the spatial epidemiological setting, the random effects-Poisson model introduced by Besag et al. (1991) has been used extensively. Indeed, application of these models has been

made relatively straightforward through the freely available and user-friendly WinBugs software. What is often ignored in the spatial setting is the possibility that the random effects distribution, typically assumed multivariate normal, has been misspecified. This is particularly problematic when the underlying spatial heterogeneity may be subject to sharp boundaries, as random effect models do not naturally allow for such discontinuities. As discussed by Knorr-Held and Rasser (2000), the resulting oversmoothing may mask such boundaries and these may be of substantial interest.

Finite mixtures provide a robust alternative to random effect models. While methodology for finite mixtures has been studied extensively (see e.g., Lindsay 1995), they have received limited attention in the spatial realm. Often, as in Chapters 2 and 4 of this thesis, finite mixtures are used to represent hidden subpopulations corresponding to different models for the quantity of interest. Alternatively, they may be viewed as a flexible, semiparametric specification for a mixing distribution, particularly when the number of mixture components is left unspecified (Richardson and Green 1997). In recent ground-breaking work, Fernandez and Green (2002) and Green and Richardson (2002) have developed finite mixture Poisson models for the analysis of spatially indexed count data. Future work will consider the development of a flexible class of spatial finite mixtures, providing an alternative to random effect models. Spatial allocation to mixture components could be based on discrete valued latent processes. Several latent process models will be considered including the Gibbs random fields, which have been employed in statistical mechanics and image processing (Geman and Geman 1984). As discussed in Qian and Titterington (1991), estimation for models containing latent Gibbs fields is extremely challenging. Both Bayesian and Likelihood based inferential methods could be developed, with the former being based on reversible jump MCMC and the later achieved through Monte Carlo maximum likelihood techniques (Geyer and Thompson 1992). Within the spatial finite mixture context, several extensions could also be investigated including: 1) spatio-temporal data 2) multivariate spatial structures for joint modelling of several response variables and 3) censored event-time data.

Bibliography

Aalen, O. O. (1987). Mixing distributions on a Markov chain. *Scandinavian Journal of Statistics* 14 281–289.

Albert, P. S. (1999). A mover-stayer model for longitudinal marker data. *Biometrics* 55 1252–1257.

Albert, P. S. & Waclawi, M. A. (1998). A two-state Markov chain for heterogeneous transitional data: A quasi-likelihood approach. *Statistics in Medicine* 17 1481–1493.

Anderson, T. W. & Goodman, L. A. (1957). Statistical inference about Markov chains. *Annals of Mathematical Statistics* 28 89–110.

Banerjee, S. & Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics* 60 268–275.

Banerjee, S., Wall, M. & Carlin, B. P. (2003). Frailty modelling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics* 4 123–142.

Bernardinelli, L., Clayton, D. & Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* 14 2411–2431.

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

Besag, J., Green, P. J., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science* 10 3–66.

Besag, J., York, J. C. & Mollié, A. (1991). Bayesian image restoration, with application in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43 1–59.

Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A. & Conlon, E. M. (1999). Bayesian models for spatially correlated disease and exposure data. *Bayesian Statistics* 6 131–156.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88 9–25.

Carlin, B. P. & Banerjee, S. (2002). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). *Bayesian Statistics* 7 45–63.

Carlin, B. P. & Louis, T. A. (1996). *Bayes and Empircal Bayes Methods for Data Analysis*. Chapman and Hall: London.

Chan, J. K. & Kuk, A. C. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics* 88 9–25.

Chan, K. S. & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* 90 242–252.

Chib, S. & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-hastings output. *Journal of the American Statistical Association* 96 270–281.

Christensen, R. & Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika* 75 693–704.

Conlon, E. M. & Waller, L. A. (1998). Flexible neighborhood structures in hierarchical models for disease mapping. Tech. rep., University of Minnesota, Division of Biostatistics.

Cook, R. J., Kalbfleisch, J. D. & Yi, G. Y. (2002). A generalized mover-stayer model for panel data. *Biostatistics* 3 407–420.

Cook, R. J. & Ng, E. M. (1997). A logistic-bivariate normal model for overdispersed two-state Markov processes. *Biometrics* 52 358–364.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

De Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag: New York.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B* 39 1–38.

Denison, D., Mallick, B. & Smith, A. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society, Series B* 60 333–350.

Diggle, P. J., Tawn, J. A. & Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* 47 299–350.

Fernandez, C. & Green, P. J. (2002). Modelling spatially correlated data via mixtures. *Journal of the Royal Statistical Society, Series B* 64 805–825.

Fletcher, R. (1987). *Practical methods of optimization*. John Wiley and Sons.

Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* 79 632–638.

Fuchs, C. & Greenhouse, J. B. (1988). The EM algorithm for maximum likelihood estimation in the mover-stayer model. *Biometrics* 44 605–613.

Gamerman, D., Moreire, A. R. B. & Rue, H. (2003). Space-varying regression models: specifications and simulation. *Computational statistics and data analysis* 42 513–533.

Gelfand, A. E., Kim, H., Sirmans, C. F. & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98 387–396.

Gelfand, A. E. & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4 11–25.

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

Gelman, A., Meng, X. L. & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* 6 733–807.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7 457–511.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* 6 721–742.

George, E. I. & McCulloch, R. E. (1997). Approaches for variable selection. *Statistica Sinica* 7 339–374.

Geweke, J. (1996). Variable selection and model comparison in regression. *Bayesian statistics* 5 609–620.

Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B* 54 657–699.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In W. R. Gilks, S. Richardson & D. J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 1–19.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41 337–348.

Gottardo, R. & Raftery, A. E. (2004). Markov chain Monte Carlo with mixtures of singular distributions. Tech. rep., Univertsity of Washington, Department of Statistics.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82 711–732.

Green, P. J. & Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association* 97 1055–1070.

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

He, F. & Alfaro, R. I. (2000). White pine weevil attack on white spruce: a survival time analysis. *Ecological Applications* 10 35–42.

Henderson, R., Shimakura, S. & Gorst, D. (2002). Modelling spatial variation in leukemia survival data. *Journal of the American Statistical Association* 97 965–972.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer Verlag.

Huffer, F. W. & Wu, H. L. (1998). Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 54 509–524.

Jin, X. & Carlin, B. P. (2005). Multivariate parametric spatiotemporal models for county level breast cancer survival data. *Lifetime Data Analysis* 11 5–27.

Jin, X., Carlin, B. P. & Banerjee, S. (2004). Generalized hierarchical multivariate CAR models for areal data. Tech. rep., University of Minnesota, Division of Biostatistics.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley and Sons.

Kass, R. & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* 90 773–795.

Kim, H., Sun, D. & Tsutakawa, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association* 96 1506–1521.

Knorr-Held L., R. G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* 56 13–21.

Komarek, A. & Lesaffre, E. (2004). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. Tech. rep., Katholieke Universiteit Leuven, Biostatistical Centre.

Kunzli, N., Jerret, M., Mack, W. J., Beckerman, B., LaBree, B., Gilliland, F., Thomas, D., Peters, J. & Hodis, H. N. (2005). Ambient air pollution and atherosclerosis in los angelas. *Environmental Health Perspectives* 113 201–206.

Kuo, L. & Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics* 25 457–472.

Lang, S. & Brezger, A. (2004). Bayesian P-Splines. *Journal of computational and graphical statistics* 13 183–21.

Li, Y. & Ryan, L. (2002). Modelling spatial survival data using semiparametric frailty models. *Biometrics* 58 287–297.

Lin, X. & Breslow, N. E. (1993). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* 91 1007–1016.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications NSF-CBMS regional conference series in probability and statistics, Vol. 5.* IMS, Hayward, CA.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* 44 226–233.

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

MacNab, Y. C. (1999). *Spatio-Temporal Modeling of Rates*. Ph.D. thesis, Simon Fraser University.

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

Meng, X. L. & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity. *Statistica Sinica* 6 831–860.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21 1087–1091.

Muenz, L. R. & Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics* 41 91–101.

Ng, E. M. & Cook, R. J. (1997). Modelling two-state disease processes with random effects. *Lifetime Data Analysis* 3 315–335.

Qian, W. & Titterington, D. M. (1991). Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society of London* 337 407–428.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson & D. J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 163–187.

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

Sinharay, S. & Stern, H. S. (2005). An empirical comparison of methods for computing Bayes factors in generalized linear mixed models. *Journal of Computational and Graphical Statistics* 14 415–435.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* 64 583–639.

Sun, D., Tsutakwa, R. K. & Speckman, P. L. (1999). Bayesian inference for CAR(1) models with noninformative priors. *Biometrika* 86 341–350.

Walker, S. G. & Mallick, B. K. (1999). A Bayesian accelerated failure time model. *Biometrics* 55 477–483.

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

Waller, L. A. & Gotway, C. A. (1993). *Statistics for Spatial Data*. Wiley, New York.

Wu, H. L. & Huffer, F. W. (1997). Modelling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics* 4 49–64.

Zeng, L. & Cook, R. J. (2004). Transition models for multivariate longitudinal binary data. Tech. rep., Univertsity of Waterloo, Department of Statistics and Actuarial Science.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics* 58 129–136.

Zhu, J., Eickhoff, J. C. & Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* 61 674–683.