Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

11-21-2016 12:00 AM

Joint Models for Spatial and Spatio-Temporal Point Processes

Alisha Albert-Green, *The University of Western Ontario*

Supervisor: Dr. Charmaine Dean, *The University of Western Ontario*
Joint Supervisor: Dr. W. John Braun, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

© Alisha Albert-Green 2016

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Statistics and Probability Commons

Recommended Citation
Albert-Green, Alisha, "Joint Models for Spatial and Spatio-Temporal Point Processes" (2016). *Electronic Thesis and Dissertation Repository*. 4306.
https://ir.lib.uwo.ca/etd/4306

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

In biostatistics and environmetrics, interest often centres around the development of models and methods for making inference on observed point patterns assumed to be generated by latent spatial or spatio-temporal processes. Such analyses, however, are challenging as these data are typically hierarchical with complex correlation structures. In instances where data are spatially aggregated by reporting region and rates are low, further complications may result from zero-inflation.

In this research, motivated by the analysis of spatio-temporal storm cell data, we generalize the Neyman-Scott parent-child process to account for hierarchical clustering. This is accomplished by allowing the parents to follow a log-Gaussian Cox process thereby incorporating correlation and facilitating inference at all levels of the hierarchy. A primary focus for these data is to jointly model storm cell detection and trajectories. To do so, storm cell duration, speed and direction are included in a marked point process framework. The thesis also proposes a general approach for the joint modelling of multivariate spatially aggregated point processes with the observed outcomes being zero-inflated count random variables. For such models, we incorporate correlation between the random field assumed to generate events and mean event counts. This is applied to lung and bronchus cancer incidence by public health unit in Ontario and a study of Comandra blister rust infection of lodgepole pine trees in British Columbia.

The key contributions from this thesis include the following: 1) developing a spatio-temporal hierarchical cluster process that incorporates correlation at all levels of the hierarchy, 2) joint modelling of a hierarchical cluster process and multivariate marks, 3) extending the framework for the joint modelling of multivariate lattice data to enable decomposition of the sources of shared spatial structure and 4) investigating aspects of the partial misspecification of joint spatial structure for multivariate lattice data.

Keywords: joint modelling, spatio-temporal point processes, Neyman-Scott process, log-Gaussian Cox process, zero-inflation, marked point processes, generalized additive models, disease mapping, conditional autoregressive models.

Co-Authorship Statement

This work was completed under the supervision of Dr. John Braun and Dr. Charmaine Dean. All papers resulting from this thesis will be co-authored with Drs. Braun and Dean.

Acknowledgements

This thesis would not have been completed without the generous support of my supervisors, Charmaine Dean and John Braun. They have provide me with many interesting avenues for research and I am thankful for their guidance and encouragement throughout all aspects of this degree.

I would also like to express my gratitude to Doug Woolford for his continued support and mentorship. He has always gone out of his way to listen and provide advice and for that I am grateful.

I wish to thank Patrick Brown, Valerie Isham and Reg Kulperger for many useful discussions related to point processes and Craig Miller for providing the storm cell data and for answering all related questions. Thanks also to Cindy Feng for her insight related to shared component models. I am grateful to my committee members Simon Bonner and Francis Zwiers, in addition to Reg Kulperger and Craig Miller, for sharing their time and expertise.

Many thanks are due to the graduate students in the Department of Statistical and Actuarial Sciences, and in particular my fellow lab members, who have made my time at Western so enjoyable.

Financial support provided by the Canadian Statistical Sciences Institute as well as the Natural Sciences and Engineering Research Council of Canada and Queen Elizabeth II scholarship programs is gratefully acknowledged.

Finally, I would like to thank my family for their unwavering love, support and encouragement.

Contents

Abstract **i**

Co-Authorship Statement **ii**

Acknowledgements **iii**

List of Figures **vi**

List of Tables **ix**

1 Introduction **1**

1.1 Outline of Thesis 3

2 Background **4**

2.1 Point Processes 4

2.2 Cox Processes 6

2.2.1 Log-Gaussian Cox Processes 6

2.2.2 Neyman-Scott Processes 7

2.2.3 Parameter Estimation 9

2.3 Marked Point Processes 11

2.4 Aggregated Point Processes 12

2.4.1 Zero-Heavy Data 12

2.4.2 Random Effects 14

2.4.3 Spline Smoothing 16

3 A Spatio-Temporal Cluster Process for Modelling Storm Cells **19**

3.1 Data Description 20

3.2 A Point Process with Multiple Levels of Clustering 22

3.2.1 Definition of Cluster Process Model 23

3.2.2 Moment Properties 24

3.2.3 Additional Assumptions 26

3.2.4 Parameter Estimation 27

3.2.5 Identifying Multiple Levels of Clustering 30

3.2.6 Confidence Intervals 32

3.2.7 Goodness of Fit 33

3.3 Application to Storm Cell Data 34

3.3.1 Goodness of fit 36

3.4 Discussion 37

4 A Joint Model for a Hierarchical Cluster Process with Evolving Marks 43
4.1 Data Description 44
4.2 Joint Models for Speed and Direction 44
4.3 Spatio-Temporal Marked Point Process Model 50
4.3.1 Marked Point Process 50
4.3.2 Parameter Estimation and Inference 53
4.4 Application to Storm Cell Trajectories 55
4.4.1 Results 55
4.4.2 Data Errors 73
4.5 Discussion 74

5 A General Framework for the Joint Modelling of Aggregated Spatial Point Patterns Subject to Clustering 77
5.1 Shared Component Models for Aggregated Point Patterns 78
5.2 Shared Component Model Framework 79
5.2.1 Model Description 79
5.2.2 Parameter Estimation and Inference 82
5.3 Applications 83
5.3.1 Ontario Lung and Bronchus Cancer 83
5.3.2 Comandra Blister Rust Infection of Lodgepole Pine Trees 90
5.4 Misspecification of Spatial Structure 96
5.4.1 Advantages of Including Between-Component Correlation 97
5.4.2 Effect of Misspecification of Spatial Structure 100
5.5 Discussion 103

6 Future Work 106
6.1 A Joint Log-Gaussian Cox Process for Multivariate Aggregated Point Patterns 107
6.2 A Composite Likelihood Approach to Parameter Estimation of a Spatio-Temporal Point Process 107
6.3 Joint Estimation for Spatio-Temporal Point Processes with Evolving Marks 108

Appendix A Supplementary Material for Chapter 3 110
A.1 Storm Cell Identification and Tracking Algorithm 110
A.2 Verification of Campbell's Theorem 112

Appendix B Supplementary Material for Chapter 4 114

Appendix C Supplementary Material for Chapter 5 138

Bibliography 141

Curriculum Vitae 148

List of Figures

3.1 Initial location of storm cells detected in April 2003 at the Bismarck, North Dakota radar station. 20
3.2 Locations of initial detection for April 2003 storm cells. 21
3.3 Locations of initial detection for May 2003 storm cells. 21
3.4 Locations of initial detection for June 2003 storm cells. 22
3.5 Locations of initial detection for July 2003 storm cells. 22
3.6 Locations of initial detection for August 2003 storm cells. 23
3.7 Empirical pair correlation function of the temporal projection process from April 2003. 31
3.8 Temporal and spatial empirical and fitted pair correlation functions for the hierarchical cluster process, the Neyman-Scott process and the log-Gaussian Cox process fit by month. 42

4.1 Storm cell duration, speed and direction from April 2003. 45
4.2 Storm cell duration, speed and direction from May 2003. 46
4.3 Storm cell duration, speed and direction from June 2003. 47
4.4 Storm cell duration, speed and direction from July 2003. 48
4.5 Storm cell duration, speed and direction from August 2003. 49
4.6 Fitted storm cell mean duration, speed and direction from April 2003. . . 60
4.7 Normalized quantile residuals for the duration, speed and direction components of the April 2003 mark models. 61
4.8 Fitted storm cell mean duration, speed and direction from May 2003. . . 63
4.9 Normalized quantile residuals for the duration, speed and direction components of the May 2003 mark models. 64
4.10 Fitted storm cell mean duration, speed and direction from June 2003. . . 66
4.11 Normalized quantile residuals for the duration, speed and direction components of the June 2003 mark models. 67
4.12 Fitted storm cell mean duration, speed and direction from July 2003. . . 69
4.13 Normalized quantile residuals for the duration, speed and direction components of the July 2003 mark models. 70
4.14 Fitted storm cell mean duration, speed and direction from August 2003. . 72
4.15 Normalized quantile residuals for the duration, speed and direction components of the August 2003 mark models. 73

5.1 Standardized incidence ratios of Ontario lung and bronchus cancer incidence for females and males. 90

5.2 Median posterior estimates of the conditional autoregressive random effects for all outcomes and model components when fit to the lung and bronchus cancer data. 91
5.3 Median posterior estimates and 95% credible intervals for the unstructured random effects for females and males. 92
5.4 Level plots and histograms for lesion and host plant counts. 96
5.5 Linearly interpolated posterior median estimates of the shared random effect. 97
5.6 Linearly interpolated independent random effects for lesions and host plants. 98

B.1 Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from April 2003. 115
B.2 Estimated spatial partial effect for the scale parameter in the duration component from April 2003. 116
B.3 Estimated spatial and temporal partial effects for the location parameter in the speed component from April 2003. 116
B.4 Estimated spatial partial effect for the scale parameter in the speed component from April 2003. 117
B.5 Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from April 2003. 118
B.6 Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from May 2003. 119
B.7 Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from May 2003. 120
B.8 Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from May 2003. 121
B.9 Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from May 2003. 122
B.10 Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from May 2003. 123
B.11 Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from June 2003. 124
B.12 Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from June 2003. 125
B.13 Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from June 2003. 126
B.14 Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from June 2003. 127
B.15 Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from June 2003. 128
B.16 Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from July 2003. 129
B.17 Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from July 2003. 130

B.18 Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from July 2003. 131
B.19 Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from July 2003. 132
B.20 Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from July 2003. 133
B.21 Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from August 2003. 134
B.22 Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from August 2003. 135
B.23 Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from August 2003. 136
B.24 Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from August 2003. 137

List of Tables

3.1 Parameter estimates and 95% confidence intervals for the spatio-temporal hierarchical cluster process fit by month. 38
3.2 Parameter estimates and 95% confidence intervals for the spatio-temporal Neyman-Scott process fit by month. 39

4.1 Summary from the four component model for storm cell trajectory in April 2003. 59
4.2 Summary from the four component model for storm cell trajectory in May 2003. 62
4.3 Summary from the four component model for storm cell trajectory in June 2003. 65
4.4 Summary from the four component model for storm cell trajectory in July 2003. 68
4.5 Summary from the four component model for storm cell trajectory in August 2003. 71

5.1 Posterior median estimates and 95% credible intervals from the joint spatial zero-inflated Poisson model for female and male cancer counts. 87
5.2 Proportion of spatial variability by model component and outcome. . . . 88
5.3 Comparison of deviance information criterion and effective number of parameters for competing models in the analysis of Ontario lung and bronchus cancer incidence. 88
5.4 Description of the shared random effects for models compared in Table 5.3. 89
5.5 Posterior median estimates and 95% credible intervals from the joint spatial zero-inflated Poisson model for lesions and host plants. 94
5.6 Comparison of deviance information criterion and effective number of parameters for competing models in the analysis of lesions and host plants. 95
5.7 Relative bias and relative root mean square error for scaling parameters and variances from models J2, J3 and S1 at different levels of the variance ratios. 101
5.8 Relative bias and relative root mean square error for scaling parameters and variances from models J2, J3 and S1 at different levels of the variance ratios when varying the neighbourhood structure. 104

C.1 Relative bias and relative root mean square error for intercept and threshold parameters from the simulation study in Section 5.4.1. 139

C.2 Relative bias and relative relative root mean square error for intercept and threshold parameters from the simulation study in Section 5.4.2. 140

Chapter 1

Introduction

A point process is defined as a set of points or events that are generated stochastically and distributed in d-dimensional space. A realization of a point process, referred to as a point pattern, can arise in many fields of science, including medicine, meteorology, forestry, ecology and seismology. In the most basic point process, a Poisson process, events occur randomly and independently; this represents a standard against which other processes may be compared. For example, if the occurrence of an event makes it more likely, relative to a Poisson process, that another event will occur in a close neighbourhood, this is referred to as a clustered process. Conversely, a process is considered to be regular if the occurrence of an event makes it less likely that another event will occur. Examples of clustered patterns include earthquake occurrences (e.g. Ogata, 1998) while in ecology, plants that compete for resources may be regularly distributed (e.g. Yau and Loh, 2012).

Often events in a point process are indistinguishable other than by their location. However, in some applications, additional characteristics can be recorded along with the point. For example, in seismology, we may be interested in earthquake magnitude as well as location. This extra information is referred to as a mark and the resulting point process is called a marked point process. In some instances, such as when anonymity is a concern, exact event locations are censored in space or space-time and instead counts aggregated by region are recorded. In the spatial statistics literature, these data structures are commonly referred to as lattice data (Cressie, 1993). Further, it is not uncommon that when the underlying point process is clustered or non-randomly thinned there may be more zeros than expected under the assumed distribution; this is referred to as zero-inflation.

Methods for *joint modelling* of several outcomes measured on a single observation have recently undergone rapid development, stemming from longitudinal studies with the goal of understanding the relationship between an observed trajectory and a time-to-event outcome (Diggle et al., 2008). In these types of analyses, the longitudinal and survival random variables are assumed to depend on latent random effects common to both outcomes. For point process data, joint modelling may refer to methods for multivariate point processes or to modelling a point and a mark process. In the context of aggregated or lattice patterns, joint modelling via the so-called shared component model is employed for increased power and facilitates the testing of a common spatial structure.

This thesis consists of three projects focussed on developing joint models for spatial and spatio-temporal point processes. The first two projects are motivated by the analysis of storm cell data from Bismarck, North Dakota. These types of data are known to be hierarchically clustered with a group of storm cells referred to as a storm and a storm system consisting of a cluster of storms (Mohee and Miller, 2010). Storm cells are also dynamic and accordingly evolve over space and time. In the first project, we develop a spatio-temporal hierarchical cluster process for the analysis of detected storm cells that accounts for correlation and facilitates inference at all levels of the hierarchy. We then extend this, in the second project, to jointly model storm cell detection and movement through the use of multivariate marks corresponding to duration, speed and direction. These models are expected to be used in simulations for understanding the impact of stresses on power systems. The third project proposes a general framework for the joint modelling of multivariate, zero-inflated spatial outcomes that arise due to the aggregation of clustered or non-randomly thinned multivariate point processes. This approach provides a clear conceptual interpretation to the generation of these random variables and accounts for correlation between the random field generating the outcomes and the mean of the observed outcomes. This joint model is applied to two data sets, the first being an analysis of lung and bronchus cancer incidence in Ontario and the second being a study of Comandra blister rust infection of lodgepole pine trees from British

Columbia. The key contributions of this thesis are as follows:

1. Generalization of the Neyman-Scott parent-child process by allowing the parents to follow a log-Gaussian Cox process thereby incorporating a hierarchical clustering structure and facilitating inference at all levels of the hierarchy.
2. Extension of the hierarchical cluster process to a marked point process which incorporates multivariate and evolving marks to jointly model storm cell detection and movement.
3. Development of a general framework for the joint modelling of multivariate spatially aggregated point processes resulting in zero-inflated outcomes which incorporates correlation between the random field assumed to generate events and mean count outcomes and facilitates inference on the types of shared spatial structure across all outcomes and components.
4. Investigation of aspects of partial misspecification of spatial structure for lattice data.

1.1 Outline of Thesis

The rest of this document is organized as follows: Chapter 2 provides background on aspects of spatial and spatio-temporal point processes as they relate to the work contained in this thesis. The development and application of the hierarchical cluster process for storm cell data is provided in Chapter 3 and Chapter 4 extends this work by incorporating the multivariate storm cell trajectories in a marked point process. Chapter 5 then shifts the focus to proposing a general framework for the joint modelling of multivariate spatially aggregated point processes and investigating common aspects of misspecification of the spatial structure in shared component models. We conclude in Chapter 6 by discussing extensions emerging from methods developed throughout this thesis.

Chapter 2

Background

This chapter provides background on the key concepts employed throughout the thesis. We begin with a brief introduction to point processes before building to Cox processes, specifically the log-Gaussian Cox process and the Neyman-Scott process. We then consider extensions to marked point processes and modelling considerations for point patterns aggregated on a lattice. Throughout this chapter we also highlight key pieces of literature. Diggle (2003) and Illian et al. (2008) provide a thorough development of spatial point processes. For temporal point processes Daley and Vere-Jones (2003) and Daley and Vere-Jones (2008) are excellent references and for spatio-temporal patterns Diggle (2014) is useful.

2.1 Point Processes

In this thesis, we are concerned with spatial ($d = 2$) and spatio-temporal ($d = 3$) point processes. When describing these processes, we are often interested in the first- and second-order intensities with the former related to the density of points and the latter affiliated with patterns of clustering or regularity. The first-order intensity at x, $\lambda(x)$, corresponds to the probability of an event occurring in a small area containing x. In the case of a homogeneous point process, $\lambda(x) = \lambda$. A point process is said to be inhomogeneous if $\lambda(x)$ is not constant. The second-order intensity, $\lambda^{(2)}(x_1, x_2)$, represents the probability of simultaneously getting points in small areas containing x_1 and x_2. If

$\lambda^{(2)}(x_1, x_2) = \lambda^{(2)}(x_2 - x_1)$ or $\lambda^{(2)}(x_1, x_2) = \lambda^{(2)}(||x_2 - x_1||)$, with $||\cdot||$ denoting the Euclidean norm, the corresponding point process is said to be stationary (translation invariant) or isotropic (translation and rotation invariant), respectively.

Further summaries of second-order properties for stationary and isotropic point processes include the K-function and the pair correlation function. The K-function, $K(r)$, is defined as $\lambda^{-1}E[N_0(r)]$ where $N_0(r)$ represents the number of further events within a distance r of an arbitrary event while the pair correlation function is

$$g(r) = \frac{\lambda^{(2)}(r)}{\lambda^2}.$$

If $r = ||x_2 - x_1||$, this is defined as the probability of events occurring simultaneously near x_1 and x_2 relative to what is expected from a Poisson process with first-order intensity λ. Hence, for a Poisson process $g(r) = 1$, while $g(r) > 1$ and $g(r) < 1$ corresponds to clustered and regular patterns, respectively. However, more complicated patterns also exist. For example, Yau and Loh (2012) introduce a spatial point process that exhibits clustering at small scales and regularity at large scales.

The Poisson process is the most basic form of a point process and is a situation in which manipulation of the likelihood function is tractable. It may be expressed as the product of two densities: the first corresponding to the Poisson$\left(\int_A \lambda(x)\mathrm{d}x\right)$ distribution for the mean number of events on a bounded d-dimensional region A and the second representing the density of the set of independent locations $\{x_i,\ i = 1, 2, \ldots, N\}$, $\lambda(x_i)/\int_A \lambda(x)\mathrm{d}x$. With this type of process the log-likelihood simplifies to

$$\ell(\lambda) = \sum_{i=1}^{N} \log[\lambda(x_i)] - \int_A \lambda(x)\mathrm{d}x. \tag{2.1}$$

For a homogeneous Poisson process, this function can be maximized analytically with respect to the parameter λ and the maximum likelihood estimate, $\hat{\lambda}$, is $N/|A|$.

2.2 Cox Processes

Cox or doubly stochastic processes refer to a class of point processes in which the observed point pattern is an inhomogeneous Poisson process conditional on a stochastic intensity, $\Lambda(x)$. They are defined by the following two postulates:

1. The intensity $\left\{\Lambda(x) : x \in \mathbb{R}^d\right\}$ is a non-negative stochastic process.
2. Conditional on $\left\{\Lambda(x) = \lambda(x) : x \in \mathbb{R}^d\right\}$, the observed pattern is an inhomogeneous Poisson process with intensity $\lambda(x)$.

If the stochastic intensity is both stationary and isotropic, the resulting Cox process possesses these properties. That is, if $\lambda = E[\Lambda(x)]$ a process is said to be stationary and if both stationary and isotropic $\lambda^{(2)}(r) = E[\Lambda(x_1)\Lambda(x_2)]$ where again $r = ||x_2 - x_1||$. A weaker form of stationarity, known as second-order intensity reweighted stationarity, has been developed for spatial (Baddeley et al., 2000) and spatio-temporal (Gabriel and Diggle, 2009) processes where the assumption of a constant first-order intensity is relaxed. As mentioned previously, the two specific forms of Cox processes that we focus on in this thesis are the log-Gaussian Cox process (Møller et al., 1998) and the Neyman-Scott process (Neyman and Scott, 1958).

2.2.1 Log-Gaussian Cox Processes

Log-Gaussian Cox processes are natural models for point patterns driven by environmental factors, for example non-infectious diseases. In this scenario, the resulting events are due to exposure of observed and possibly unobserved environmental covariates. This is in contrast to processes driven at least partially by interaction amongst the points, such as infectious diseases. Furthermore, log-Gaussian Cox processes are completely characterized by their first-order intensity and their pair correlation function (see proof in Møller et al., 1998).

Specifically, we assume that the observed pattern is driven by a Gaussian process $\mathcal{Z} = \left\{Z(x) : x \in \mathbb{R}^d\right\}$ where $E[Z(x)] = \mu$ and $\text{Cov}[Z(x_1), Z(x_2)] = \sigma^2\rho(x_2 - x_1; \phi)$ being positive semi-definite with $\rho(0) = 1$. The variance parameter, σ^2, represents the amount of clustering present with a larger value indicating greater clustering in the observed process, while the scale parameter, ϕ, corresponds to the range of correlation in the Gaussian process. Common covariance structures include Matérn, exponential and Gaussian. As the name implies, the stochastic intensity generating the log-Gaussian Cox process, $\Lambda(x)$, is equal to $\exp\{Z(x)\}$. This process has first-order intensity $E[\Lambda(x)] = \exp\{\mu + 0.5\sigma^2\}$ and second-order intensity $E[\Lambda(x_1)\Lambda(x_2)] = [\exp\{\mu+0.5\sigma^2\}]^2\exp\{\sigma^2\rho(x_2-x_1;\phi)\}$ which may be calculated directly from the properties of the log-normal distribution. The pair correlation function for the log-Gaussian Cox process has the form $g(x_2 - x_1) = \exp\{\sigma^2\rho(x_2 - x_1;\phi)\}$.

A convenient reparameterization of the process that we utilize in this thesis allows for separation between the first- and second-order properties. As suggested in Diggle et al. (2013) this may be obtained if we let $\Lambda(x) = \exp\{\beta^* + Z(x)\}$ with $E[Z(x)] = -0.5\sigma^2$. In this case, $E[\exp\{Z(x)\}] = 1$ and $E[\Lambda(x)] = \exp\{\beta^*\} = \lambda$.

Covariates may be incorporated into the log-Gaussian process by replacing λ with $\lambda(x) = \lambda\{\varphi(x);\boldsymbol{\beta}\}$ if φ is a function of covariates indexed by location and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ resulting in a second-order intensity-reweighted stationary pattern. Log-Gaussian Cox processes may also be employed for data where the region A is discretized into a regular lattice. The number of events per region may then be modelled as Poisson random variables (e.g. Diggle et al., 2013).

2.2.2 Neyman-Scott Processes

The Neyman-Scott process is a parent-child process commonly employed for modelling clustered point patterns. It is defined as follows:

1. Parent points, $\mathcal{P} = \{p_1, p_2, \ldots\} \in \mathbb{R}^d$ form a Poisson process with intensity λ_p.
2. Each parent generates an independent and identically distributed random number of offspring, N_j, with mean $\alpha > 0$.
3. For the jth parent, the displacement of the offspring $\boldsymbol{X}_j = \{x_{1j}, x_{2j}, \ldots, x_{N_j j}\}$ relative to their parent is independent and identically distributed according to the d-dimensional distribution, $k(\cdot)$.

The observed pattern in this case consists only of the offspring. Typically, the distribution generating the number of offspring is Poisson. For a planar point process, if the density $k(\cdot)$ follows a bivariate normal distribution, this specification is referred to as a modified Thomas process (Thomas, 1949) while a Matérn cluster process (Matérn, 1960) results when the offspring are randomly scattered in a disc ($d = 2$) or ball ($d = 3$). More generally, if $d = 3$ and $k(\cdot) = k_s(\cdot)k_t(\cdot)$, where $k_s(\cdot)$ and $k_t(\cdot)$ represent the spatial and temporal displacement distributions, $k(\cdot)$ is said to be separable in space-time. Commonly, $k_s(\cdot)$ is bivariate normal and $k_t(\cdot)$ may be normal, exponential or half-normal depending on the application. However, note that the case where $k_t(\cdot)$ is normal corresponds to a scenario in which the offspring may appear prior to the arrival of their parent. The displacement distribution is parameterized in terms of a standard deviation parameter, ω; the radius, 2ω, represents the typical size of the cluster in space or time (Wiegand et al., 2007).

Writing this as a Cox process, the distribution of the offspring conditional on the parents is said to follow an inhomogeneous Poisson process with intensity $\Lambda_{X|P}(x) = \sum_{j:p_j \in \mathcal{P}} \alpha k(p_j, x; \omega)$. Therefore, the Neyman-Scott process is a Cox process generated by the stochastic intensity $\Lambda_{X|P}(x)$. As written, this process is stationary with unconditional first-order intensity $E[\Lambda_{X|P}] = \alpha\lambda_p$, which is calculated by integrating out the unobserved parents, and isotropic so long as the displacement distribution, $k(\cdot)$, is symmetric. The corresponding second-order intensity and pair correlation function are

$$\lambda^{(2)}(r) = \alpha^2 \lambda_p k * k(r; \omega) + \alpha^2 \lambda_p^2$$

and

$$g(r) = \frac{k * k(r; \omega)}{\lambda_p} + 1,$$

where $*$ is the convolution operator.

Extensions to second-order intensity reweighted stationarity are regularly employed in order to incorporate covariates into the offspring distribution (e.g. Henrys and Brown, 2009; Waagepetersen and Guan, 2009; Waagepetersen, 2007).

2.2.3 Parameter Estimation

For a non-Poisson point process, evaluation of the likelihood treats the unobserved intensity as missing data and utilizes Monte Carlo methods for maximization as it is not available in closed form (Møller and Waagepetersen, 2004); this is typically computationally prohibitive for Cox and cluster processes (Waagepetersen and Guan, 2009). With log-Gaussian Cox processes specifically, integrated nested Laplace approximation may be employed to quickly and accurately approximate the posterior marginal distributions (Lindgren et al., 2011; Rue et al., 2009). For general stationary and isotropic spatial point processes, parameter estimation is performed either by maximizing a composite likelihood (e.g. Guan, 2006) or via minimum contrast estimation (e.g. Diggle, 2003). As described in Guan (2006) the former derives a likelihood by summing log-likelihoods, with each element being a valid marginal or conditional density. Specifically, they define

$$\frac{\lambda^{(2)}(x_2 - x_1)}{\int\int_S \lambda^{(2)}(x_2 - x_1)\mathrm{d}x_1\mathrm{d}x_2}$$

as the joint distribution utilized for parameter estimation where S is a two-dimensional region. Minimum contrast estimation is a non-parametric least squares approach to parameter estimation which minimizes the discrepancy between an empirical function of the data and a theoretical function based on the proposed process. Applied in the context of point process modelling, the pair correlation function (e.g. Prokešová and Dvořák, 2014) and the K-function (e.g. Henrys and Brown, 2009) are commonly utilized for optimization. The remainder of this section focusses on the use of the pair correlation function, which is employed throughout the thesis. This approach to estimating the clustering parameters is conditional on the first-order intensity, which may be estimated by solving an estimating equation that comes from differentiating an objective function, such as the log-likelihood in Equation (2.1). As shown in Schoenberg (2005), estimates based on this function are consistent even for non-Poisson processes. The clustering parameters may then be estimated by minimizing

$$D = \int_0^{r_{\text{corr}}} \left[g(u)^c - \hat{g}(u)^c\right]^2 \mathrm{d}u.$$

where $g(u)$ is the theoretical pair correlation function and $\hat{g}(u)$ is the empirical pair correlation function, which in d dimensions is

$$\hat{g}(r) = \frac{\sum_{i=1}^{N} \sum_{j \neq i} \kappa_{\epsilon}(r - ||x_i - x_j||) v_{ij}}{\upsilon |A| \hat{\lambda}^2}. \tag{2.2}$$

In the above, $\hat{\lambda}$ is the estimated first-order intensity, κ is a kernel, such as the Epanechnikov, with bandwidth ϵ, v_{ij} is an edge correction factor (Gabriel, 2014) and $\upsilon = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1}$. For a thorough overview of kernel density estimation, please see (Silverman, 1998). Further, r_{corr}, termed the range of correlation, is the value above which the empirical pair correlation function is equal to one. Finally, c is a constant which stabilizes the variance inherent in $\hat{g}(\cdot)$. For minimum contrast estimation, the variance of the empirical estimates increases with r. Therefore, the value of c is chosen to reduce the influence of large values on the estimated parameters. Typically $c = 0.5$ for regular patterns while a more severe transformation of $c = 0.25$ is suggested for clustered patterns. For additional discussion on the choice of constant, we direct interested readers to Diggle (2003). Although the composite likelihood approach has not yet been extended to the spatio-temporal realm, the minimum contrast technique is known to be unstable if extended directly to three dimensions (Prokešová and Dvořák, 2014). Accordingly, parameter estimation for spatio-temporal point processes is still considered to be in its infancy. Briefly, Onof et al. (2000) describe a spatio-temporal point pattern for modelling rainfall that employs a generalized method of moments estimator. More recently, the use of minimum contrast estimation for spatio-temporal cluster processes via the lower dimensional spatial and temporal projection processes has been advocated (Møller and Ghorbani, 2012; Prokešová and Dvořák, 2014).

Following this, obtaining confidence intervals for the parameters requires the use of Monte Carlo methods, for example the parametric bootstrap (Davison and Hinkley, 1997) or the non-parametric bootstrap (e.g. Braun and Kulperger, 1998; Loh, 2008; Loh and Stein, 2004, 2008).

2.3 Marked Point Processes

Sections 2.1 and 2.2 dealt with unmarked point processes only concerned with event positions. However, sometimes points may also be characterized by additional variables associated with the events, referred to as marks. That is, for a point process $\mathcal{X}$ in $A \subset \mathbb{R}^d$, $\mathcal{X}_m = \{[x, m(x)] : x \in X\}$ is a marked point process where $m(x) \in \mathcal{M}$ is a mark associated with the point x in mark space $\mathcal{M}$. An example of a well known marked point process is the Neyman-Scott parent process. In Section 2.2.2, we defined the parent process as a point process on $\mathbb{R}^d \times (0, \infty)$. This can also be considered a marked point process where the points are the parent process and the marks are the corresponding number of offspring.

If we let $\mathcal{X}$ denote the unmarked point process and $\mathcal{M}$ denote the marks, the goal for a marked point process is to model the joint distribution of the events and the marks, $[\mathcal{X}, \mathcal{M}]$. For example, the epidemic-type aftershock model (Ogata, 1998) considers data on locations of earthquakes or aftershock as well as marks related to the corresponding magnitude. However, if the point and marks processes are separable, it suffices to model $[\mathcal{X}, \mathcal{M}] = [\mathcal{X}][\mathcal{M}]$ where $[\mathcal{X}]$ and $[\mathcal{M}]$ denote the distribution of the point and mark processes, respectively. That is, we can model the marks and points independently. For spatial point processes, this amounts to employing point process techniques for modelling the events and point-referenced methods for the marks. Tests for separability of points and marks include Schoenberg (2004), for example. If $\mathcal{X}$ and $\mathcal{M}$ are not separable, the dependence between these processes should be taken into account. This may also be done through conditional analyses where $[\mathcal{X}, \mathcal{M}] = [\mathcal{X}][\mathcal{M} \mid \mathcal{X}] = [\mathcal{M}][\mathcal{X} \mid \mathcal{M}]$. The former decomposition corresponds to a scenario in which the events are generated according to a point process and the marks are modelled conditional on locations while the latter corresponds to an analysis of the events conditional on the marks.

A marked point process may also be employed to reduce model complexity. For example, a three-dimensional spatio-temporal point process can be fit as a spatial point process with time as a mark or for modelling a non-simple point process in which there are coincident points, with the mark corresponding to the multiplicity. They may also be utilized for multivariate point patterns with a categorical mark for the type of event.

2.4 Aggregated Point Processes

As mentioned previously, point process applications are plentiful, and in these instances, joint modelling techniques are often concerned with modelling the joint distribution of the points and the marks. However, for applications in which anonymity is a concern, as it commonly is for health administrative purposes with public health data, point patterns may only be available in the form of aggregated counts. In fact, even given exact locations, patterns may be discretized as likelihood-based techniques are then feasible for parameter estimation. Joint models for discretized marked point processes have been successfully fit. For example, Illian et al. (2012) fit a spatial log-Gaussian Cox process with two spatially correlated marks in a shared component framework.

For aggregated patterns, joint modelling via shared component models is also regularly utilized for multivariate outcomes hypothesized to have a common spatial structure (e.g. Feng and Dean, 2012). A further complication is that zero-heaviness may arise as a result of clustered patterns or due to non-random thinning. In what follows we introduce methods for the analysis of zero-heavy count data as well as techniques employed to account for spatial and spatio-temporal autocorrelation in these types of processes.

2.4.1 Zero-Heavy Data

Accounting for zero-heaviness in aggregated point patterns is traditionally done through the use of zero-inflated models (e.g. zero-inflated Poisson regression models as originally proposed by Lambert, 1992) or so-called hurdle models (e.g. Welsh et al., 1996). However, the choice of model is often guided by scientific objectives. In zero-inflated models, zeros may arise either from the distribution for the counts or from the structural zero component, the latter being represented by a distribution which takes on the value 0 with probability one incorporated through a Bernoulli model, typically in a logistic framework. This component is then mixed with a count distribution. The hurdle model can be envisioned as a two-stage process with the first stage being a Bernoulli trial where “success” might equate to exceedance of the hurdle, and the second stage being the generation of an outcome with positive support. For zero-heavy data, the hurdle corresponds to zero and the conditional model would typically be a count distribution, truncated at zero.

Letting $Y_1, Y_2, \ldots, Y_n$ denote n count random variables, if assumed to follow a zero-inflated Poisson regression model,

$$\begin{aligned} Y_i &= 0, \text{ with probability } \pi_i \\ &\sim \text{Poisson}(\lambda_i), \text{ with probability } 1-\pi_i \end{aligned}$$

so

$$Y_i = \begin{cases} 0, & \text{with probability } \pi_i + (1-\pi_i)e^{-\lambda_i} \\ k, & \text{with probability } (1-\pi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad k = 1, 2, \ldots . \end{cases}$$

The zero-inflation probability and the Poisson mean may be modelled as functions of covariates, as follows:

$$\varrho_1(\pi_i) = \boldsymbol{G}_i\boldsymbol{\gamma} \tag{2.3}$$

and

$$\varrho_2(\lambda_i) = \boldsymbol{B}_i\boldsymbol{\beta} \tag{2.4}$$

where $\varrho_1(\cdot)$ and $\varrho_2(\cdot)$ are link functions for the zero-inflation and count components, $\boldsymbol{G}_i$ and $\boldsymbol{B}_i$ are the covariate vectors for the ith observation incorporated into the two terms, respectively, with $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{p_1})^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p_2})^T$ being the corresponding parameters. As advocated in Lambert (1992) and Hall (2000), the EM algorithm may be employed for parameter estimation. However, if random effects are incorporated to account for autocorrelated data or to link multivariate outcomes in a shared component framework, this approach may quickly become computationally intensive.

For the hurdle model,

$$\begin{aligned} Y_i &= 0, \text{ with probability } 1-\pi_i \\ &\sim \text{truncated Poisson}(\lambda_i), \text{ with probability } \pi_i \end{aligned}$$

such that

$$Y_i = \begin{cases} 0, & \text{with probability } 1-\pi_i \\ k, & \text{with probability } \pi_i \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i![1-e^{-\lambda_i}]}, \quad k = 1, 2, \ldots \end{cases}$$

where π_i and λ_i can be modelled as functions of covariates, as shown in Equations (2.3) and (2.4), respectively. In the hurdle framework, unlike the zero-inflated Poisson model, it is fully efficient to estimate the two components separately and so standard iteratively re-weighted least squares techniques may be employed to estimate parameters within each component (Welsh et al., 1996). Zero-inflated distributions, however, are not constrained to count data. For example, Li et al. (2011) employed a zero-inflated model for the analysis of log-normal random variables with a point mass at zero.

2.4.2 Random Effects

Zero-heavy regression models are regularly employed for modelling autocorrelated data whether they be spatial (e.g. Agarwal et al., 2002; Neelon et al., 2013; Recta et al., 2012) or spatio-temporal (e.g. Tzala and Best, 2008; Richardson et al., 2006). If spatial data are aggregated over a lattice, accounting for correlation is often accomplished through the use of a conditional autoregressive random effect (Besag et al., 1991). These structures can also be employed to approximate what is likely a correlation based on distances as there may be computational advantages for large n. Additionally, distance-based effects including the aforementioned Matérn or exponential structures may be utilized for spatial or spatio-temporal point-referenced data or even lattice data based on the coordinates of region centroids, for example.

In this thesis, when accounting for spatial correlation in lattice data, we employ conditional autoregressive random effects. Let $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)^T$ denote the vector of spatial random effects for n regions and let $\boldsymbol{W} = (w_{ii'})$ represent the spatial proximity matrix where $w_{ii'} = 1$ if regions i and i' are neighbours (denoted $i \sim i'$), $w_{ii'} = 0$ otherwise. This random effect is then specified through a series of conditional distributions, assumed to be normally distributed, where

$$E[b_i \mid b_{i' \neq i}] = \frac{1}{w_{i+}} \sum_{i' \sim i} b_{i'}$$

denotes the conditional expectation with $w_{i+} = \sum_{i'} w_{ii'}$ and

$$\text{Var}[b_i \mid b_{i' \neq i}] = \frac{\sigma_b^2}{w_{i+}}$$

is the conditional variance with σ_b^2 representing the variance. This random effect smooths locally as $E[b \mid b_{i' \neq i}]$ is the average effect over the neighbours of region i and $\text{Var}[b \mid b_{i' \neq i}]$ is larger for regions with fewer numbers of neighbours. As shown in Besag (1974), the vector $\boldsymbol{b}$ has a joint multivariate normal distribution where $\boldsymbol{b} \sim \text{MVN}(\boldsymbol{0}, \Sigma)$ and $\Sigma = \sigma_b^2(\boldsymbol{D} - \boldsymbol{W})^{-1}$ with $\boldsymbol{D} = \text{diag}(w_{1+}, w_{2+}, \ldots, w_{n+})$. This formulation is termed the intrinsic conditional autoregressive structure and, due to its conditional specification, facilitates Markov chain Monte Carlo methods. Briefly, Markov chain Monte Carlo techniques (Gelman et al., 2004) are a class of algorithms for sampling from a potentially high dimensional target distribution, referred to as a posterior distribution, which may be unavailable in closed form. For parameters, $\boldsymbol{\theta}$, if $P(\boldsymbol{\theta} \mid \boldsymbol{Y})$ denotes the posterior distribution and $\boldsymbol{Y}$ represents the random variables, then through Bayes theorem:

$$P(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \frac{P(\boldsymbol{Y} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{\int P(\boldsymbol{Y} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}$$

where $P(\boldsymbol{Y} \mid \boldsymbol{\theta})$ is the likelihood and $P(\boldsymbol{\theta})$ represents the prior distribution. For the conditional autoregressive random effect, Gibbs sampling, a type of Markov chain Monte Carlo method in which parameters are estimated by sampling from their full conditional distributions, is particularly convenient. However, for the intrinsic conditional autoregressive random effect specifically, because $(\boldsymbol{D} - \boldsymbol{W})$ is singular, this is an improper prior and therefore a sum-to-zero constraint (i.e. $\sum_{i=1}^{n} b_i = 0$) is often imposed at each iteration of the Markov chain Monte Carlo sampler (Eberly and Carlin, 2000).

2.4.3 Spline Smoothing

Splines provide an alternative approach to accounting for autocorrelation in spatial or spatio-temporal data. While random effects are stochastic representations of smooth functions, a spline is deterministic. However, Cressie (1993) showed that a thin plate regression spline of order m_s may be viewed as a realization of a Gaussian process model with generalized covariance, $C(r) \propto r^{2m_s-2}\log(r)$. Paciorek (2007) provides an overview of competing models used to account for spatial autocorrelation, including splines. Splines will play an important role in some of the modelling methodology in this thesis; for a thorough discussion of terminology and results, please see Wood (2006).

Specifically, splines are semi-parametric functions incorporated in linear or generalized linear regression models as a flexible approach to accounting for non-linear covariate effects (Wood, 2006; Hastie and Tibshirani, 1990). In our context, these effects would include spatial coordinates and event times. The resulting model is referred to as a generalized additive model.

Suppose $Y_1, Y_2, \ldots, Y_n$ denotes random variables with conditional mean λ_i where

$$\varrho(\lambda_i) = \boldsymbol{B}_i\boldsymbol{\beta} + \sum_{j=1}^{J} \varphi_j(x_{ji}).$$

Here, $\varrho(\cdot)$ is the known link function, $\boldsymbol{B}_i$ are the covariates for the ith observation which are linearly related to $\varrho(\lambda_i)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ are the corresponding coefficients and φ_j is the jth smoother or piecewise polynomial for the covariate x_{ji}. Note that x_{ji} need not be scalar; it may have two components $x_{ji} = (x_{1ji}, x_{2ji})$ or, for spatio-temporal data, it may be the triplet $x_{ij} = (x_{1ji}, x_{2ji}, x_{3ji})$ corresponding to two-dimensional space and time. Suppose, for example, that x_j is scalar and

$$\varphi_j(x_j) = \sum_{k=1}^{q_j} \psi_{jk}\eta_{jk}(x_j), \tag{2.5}$$

where $k = 1, 2, \ldots, q_j$ indexes the knots for the jth smoother, $\eta_{jk}(x_j)$ is the value of the basis function for the jth covariate at the kth knot and ψ_{jk} is the corresponding coefficient. When utilizing splines, the number and location of knots needs to be carefully selected. In a generalized additive model, this is accomplished by fitting a model with more knots than required and adding a penalty term in the likelihood to control for overfitting. Specifically, we maximize the following penalized log-likelihood

$$\ell_{\text{pen}}(\boldsymbol{\theta} \mid \boldsymbol{y}) = \ell(\boldsymbol{\theta} \mid \boldsymbol{y}) - \frac{1}{2}\sum_{j=1}^{J} \zeta_j \int \boldsymbol{\psi}_j^T \boldsymbol{\Omega}_j(x_j)\boldsymbol{\psi}_j \mathrm{d}x_j \tag{2.6}$$

to estimate the parameters $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_j$ corresponding to the elements of $\boldsymbol{\theta}$ belonging to the jth spline and $\boldsymbol{\Omega}_j(x_j) = \sum_{i=1}^{n} \eta''_{j\ell}(x_{ji})\eta''_{j\ell'}(x_{ji})$ for $\ell = 1, 2, \ldots, q_j$ and $\ell' = 1, 2, \ldots, q_{j'}$. This is performed conditional on the smoothing parameters, ζ_j, which controls the trade-off between smoothness and fit. That is, if $\zeta_j = 0$, $\varphi_j(x_j)$ would closely follow the data while as $\zeta_j \to \infty$, $\varphi_j(x_j)$ becomes increasingly smooth. Smoothing parameter selection is accomplished through generalized cross validation or unbiased risk estimation (Wood, 2006). Maximization of the likelihood may be accomplished through a penalized iteratively re-weighted least squares algorithm (Wood, 2006) or a backfitting algorithm (Rigby and Stasinopoulos, 2005). One advantage of the latter approach is that, unlike in the former, it does not require the response to be from within the exponential family.

Common forms of basis functions include tensor product splines, which are scale invariant, and isotropic thin plate regression splines, which are rotation invariant. The latter is recommended for spatial data if isotropy is a reasonable assumption while the former may be useful if different amounts of smoothing for space and time are desired with spatio-temporal data. Briefly, for thin plate regression splines, the smoother may be written as

$$\varphi(x) = \sum_{k=1}^{q} \psi_{1k}\eta_{1m_d d}(||x - x_k^*||) + \sum_{k=1}^{M} \psi_{2k}\eta_{2k}(x).$$

Here, x_k^* represents the location of the kth knot, $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ contain unknown parameters and η_{2k} are linearly independent polynomials. If m_d is the order of the derivative measuring flexibility of the spline (commonly two) and d is the dimension of the smoother,

$M = \binom{m_d+d-1}{d}$ and

$$\eta_{1m_dd}(r) = \begin{cases} \frac{(-1)^{m_d+1+d/2}}{2^{2m_d-1}\pi^{d/2}(m_d-1)!(m_d-d/2)!} r^{2m_d-d}\log(r), & \text{if } d \text{ even} \\ \frac{\Gamma(d/2-m_d)}{2^{2m_d}\pi^{d/2}(m_d-1)!} r^{2m_d-d}, & \text{if } d \text{ odd.} \end{cases}$$

The corresponding penalty term is

$$J_{m_dd}(\varphi) = \int \cdots \int_{\mathbb{R}^d} \sum_{m^*_{d1}+\cdots+m^*_{dd}=m_d} \frac{m_d!}{m^*_{d1}!\cdots m^*_{dd}!} \left(\frac{\partial^{m_d}\varphi}{\partial x_1^{m^*_{d1}} \cdots \partial x_d^{m^*_{dd}}} \right)^2 \mathrm{d}x_1 \cdots \mathrm{d}x_d$$

where $m^*_{d1}, m^*_{d2}, \ldots, m^*_{dd}$ represents the order of the derivative for the respective covariate. For a spatial smoother ($d = 2$), if we measure model flexibility by the squared second derivative of the basis function, $m_d = 2$, $M = 3$, $\eta_{21}(x_1, x_2) = 1$, $\eta_{22}(x_1, x_2) = x_1$ and $\eta_{23}(x_1, x_2) = x_2$, then the basis function, $\eta_{m_dd}(r)$, is $\frac{1}{8\pi}r^2\log(r)$ with penalty

$$J_{22}(\varphi) = \int\int \left[\left(\frac{\partial^2\varphi}{\partial x_1^2} \right) + 2\left(\frac{\partial^2\varphi}{\partial x_1 \partial x_2} \right) + \left(\frac{\partial^2\varphi}{\partial x_2^2} \right) \right] \mathrm{d}x_1\mathrm{d}x_2.$$

For further discussion related to thin plate regression splines as well as various other types of smoothers and their properties, Wood (2006) provides a good overview.

Chapter 3

A Spatio-Temporal Cluster Process for Modelling Storm Cells

This chapter deals with the development of models for possibly hierarchically clustered spatio-temporal point patterns, motivated by the analysis of storm cell data.

On September 15, 1996 a severe thunderstorm brought down Manitoba Hydro electricity transmission line towers (Mohee and Miller, 2010). The accompanying winds lead to numerous tower failures and an interruption in electricity supply from the generation plants in Northern Manitoba to the distribution networks in North Dakota for the two subsequent weeks; it caused a financial loss for Manitoba Hydro and was disruptive to residents of Manitoba and North Dakota. This prompted research investigations led by Manitoba Hydro on the modelling and prediction of the failure of transmission lines caused by high intensity winds. Our focus here is understanding the clustering of storm cells with hopes that this information could be utilized by power system operators who monitor power flow on transmission lines. Accordingly, our goal for this chapter is to develop a model for storm cell detection which facilitates inference on both storms (clusters of storm cells) and storm systems (clusters of storms).

3.1 Data Description

The storm cell data used in this analysis are from the Bismarck, North Dakota radar station which has a maximum detection range of 460 kilometers (km). Figure 3.1 displays the location of this radar station with the storm cells from April 2003 identified. Scans are performed every 4.5-6 minutes in precipitation mode and in clean-air mode they occur in 10 minute intervals (Mohee and Miller, 2010).

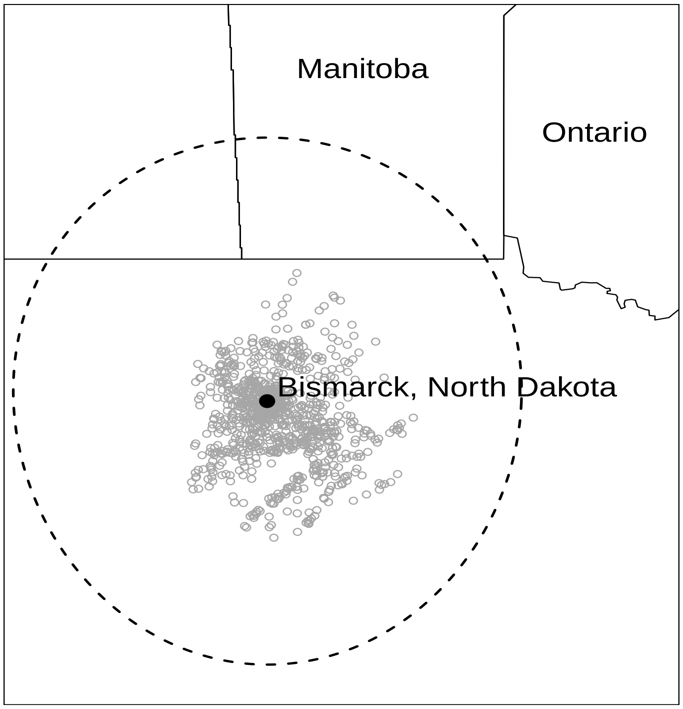
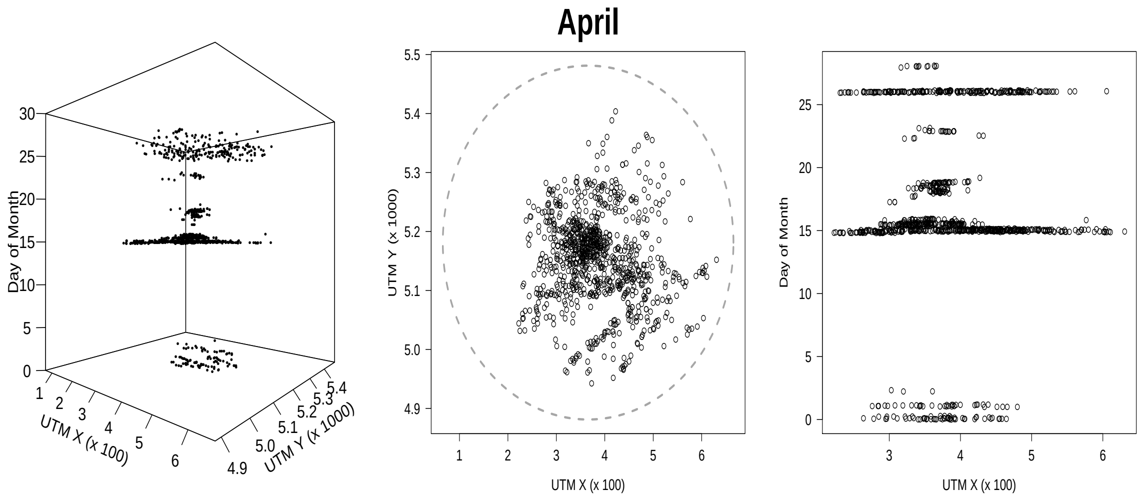


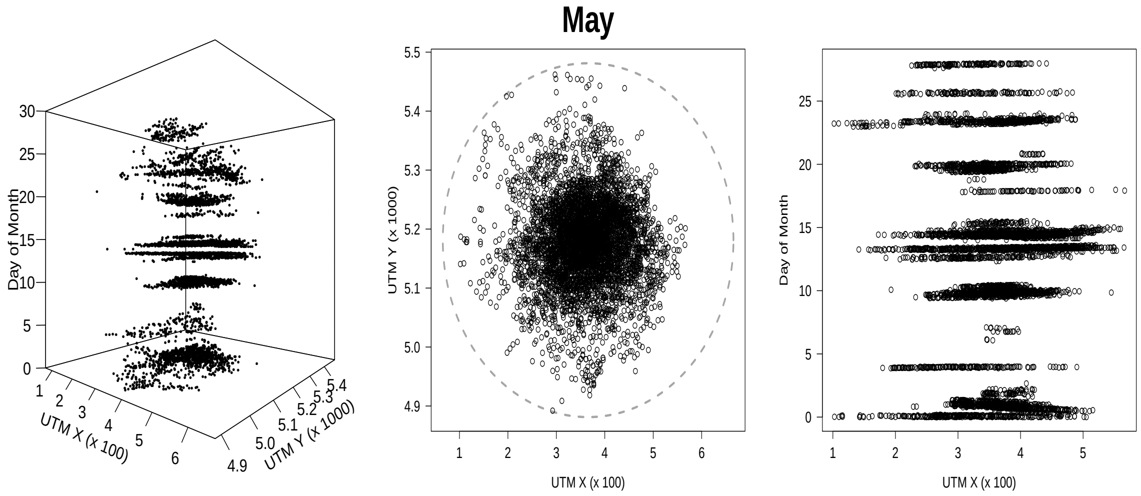
Figure 3.1: Initial location of storm cells detected in April 2003 at the Bismarck, North Dakota radar station.

This analysis focusses on modelling the initial location of detected storm cells from April 2003 - August 2003 which have been identified by a Doppler radar called Weather Surveillance Radar-1988 and pre-processed according to the Storm Cell Identification and Tracking algorithm outlined in Appendix A.1. For each storm cell, we have UTM X and UTM Y coordinates of the mass-weighted centroids, precise to 10^{-4} km, and time, measured in Julian days, accurate to within one second (U.S. Deparment of Commerce/National Oceanic and Atmospheric Administration, 2006). Figures 3.2 - 3.6 dis-



(a) Spatio-temporal process. (b) Spatial projection process. (c) Temporal projection process.

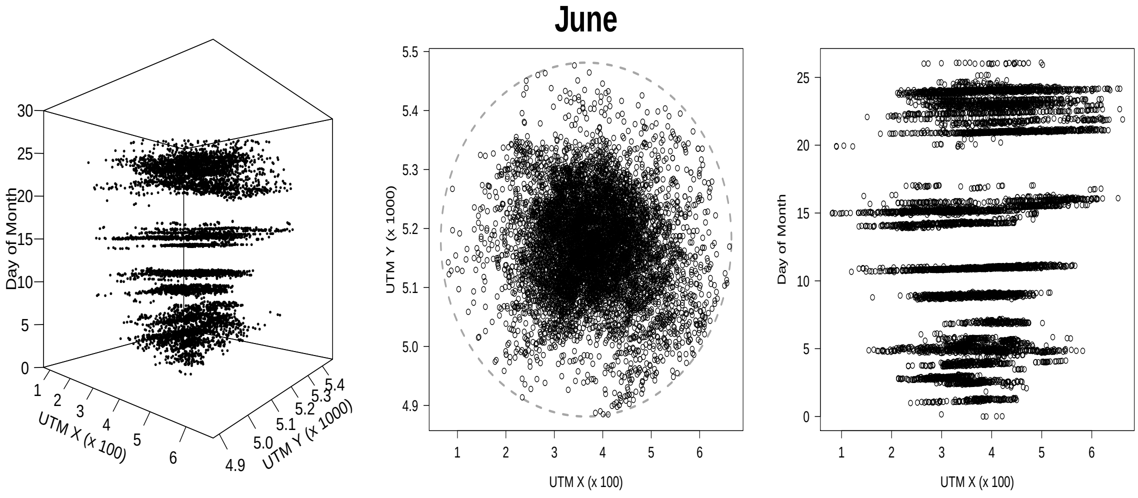
Figure 3.2: Locations of initial detection for April 2003 storm cells.



(a) Spatio-temporal process. (b) Spatial projection process. (c) Temporal projection process.

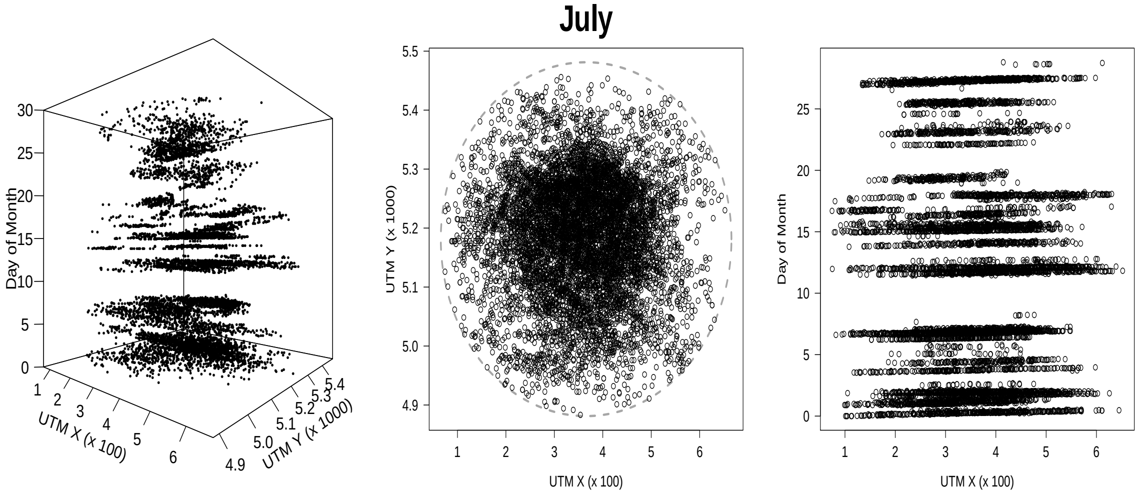
Figure 3.3: Locations of initial detection for May 2003 storm cells.

play spatio-temporal processes, spatial projection processes and temporal projection processes, by month for April - August, 2003. These figures all display patterns of spatial and temporal clustering.



(a) Spatio-temporal process. (b) Spatial projection process. (c) Temporal projection process.

Figure 3.4: Locations of initial detection for June 2003 storm cells.

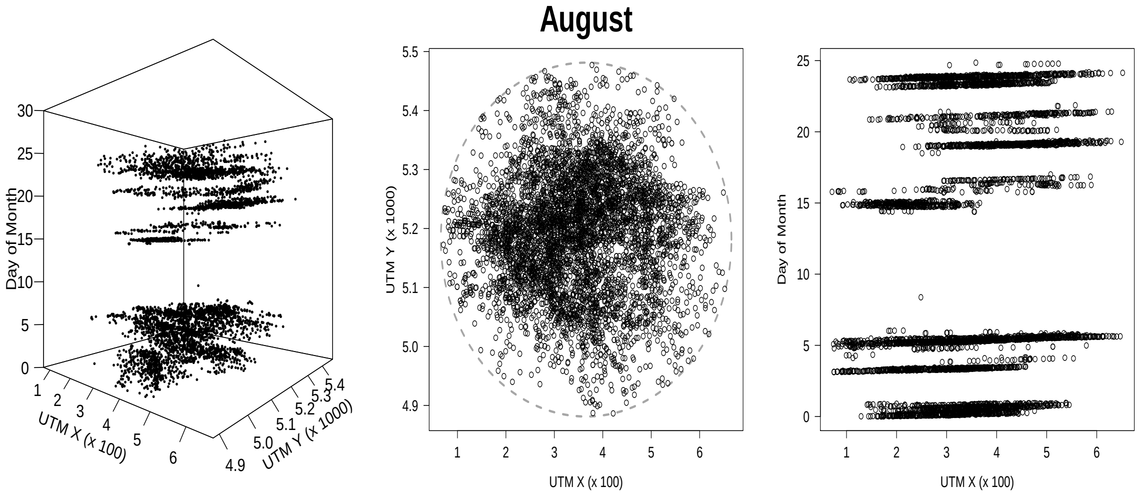


(a) Spatio-temporal process. (b) Spatial projection process. (c) Temporal projection process.

Figure 3.5: Locations of initial detection for July 2003 storm cells.

3.2 A Point Process with Multiple Levels of Clustering

Rather than explicitly modelling storm systems, we assume that they are represented by a Gaussian process which governs the generation of storms (parents) according to a log-Gaussian Cox process. The corresponding storm cells (offspring) are distributed around



(a) Spatio-temporal process. (b) Spatial projection process. (c) Temporal projection process.

Figure 3.6: Locations of initial detection for August 2003 storm cells.

the parents following a Neyman-Scott structure. This proposed process incorporates hierarchical clustering with spatio-temporal correlation at both levels.

This section starts by describing our proposed process along with the first- and second-order intensities and corresponding pair correlation function. We then outline parameter estimation and uncertainty quantification.

3.2.1 Definition of Cluster Process Model

Let $\mathcal{P}$ denote an unobserved point process (the "parent" process) on $\mathbb{R}^2 \times \mathbb{R}$ where a point $(u, v) \in \mathcal{P}$ corresponds to an event $u \in \mathbb{R}^2$ in space occurring at time $v \in \mathbb{R}$. Conditional on a Gaussian process, $\mathcal{Z} = \{Z(u, v)\}$, we assume $\mathcal{P}$ follows an inhomogeneous Poisson process with intensity $\Lambda_P(u, v)$ where

$$\Lambda_P(u, v) \quad = \quad \exp\{\beta^* + Z(u, v)\} \tag{3.1}$$

with β^* being an intercept parameter. The Gaussian process is characterized in terms of its mean $E[Z(u, v)] = -0.5\sigma^2$ and covariance $\text{Cov}[Z(u_1, v_1), Z(u_2, v_2)] = \sigma^2 \rho(u_2 - u_1, v_2 - v_1; \phi)$ with variance σ^2 and scale ϕ. In general, any positive semidefinite covariance function in which $\rho(0) = 1$ may be utilized and we return to discuss the specific form employed at the end of this section.

Letting $N_j,\ j = 1, 2, \ldots$ be the number of offspring generated by the jth parent, we assume that N_j follows a Poisson(α) distribution. Suppose $\mathcal{X}$ represents another point process (the "offspring" process) with a point $(s,t) \in \mathcal{X}$ where $s \in S \subset \mathbb{R}^2$ and $t \in T \subset \mathbb{R}$, and conditional on $\mathcal{P}$, $\mathcal{X}$ follows an inhomogeneous Poisson process with intensity $\Lambda_{X|P}(x)$. If $k(\cdot)$ is the displacement distribution of the offspring from the parents, for example trivariate normal with covariance $\omega^2 \boldsymbol{I}_3$ and $\boldsymbol{I}_3$ being a 3×3 identity matrix, then $\Lambda_{X|P}(x) = \sum_{j:p_j \in \mathcal{P}} \alpha k(x - p_j; \omega)$. Note that $N = \sum_{j:p_j \in \mathcal{P}} N_j$ is the total number of offspring in $\mathcal{X}$.

3.2.2 Moment Properties

The unconditional first-order intensity, $\lambda(x) = E[\Lambda_{X|P}(x)]$, can be rigorously derived as follows. Let $N_X(A)$ denote the number of points of the process $\mathcal{X}$ in a compact subset A. Conditioning on the parent process, $\mathcal{P}$, we observe that

$$E[N_X(A)] = E\{E[N_X(A)|\mathcal{P}]\} = E\left[\int_A \Lambda_{X|P}(u)du\right] = \int_A E[\Lambda_{X|P}(u)]du.$$

The last equality follows from Tonelli's theorem (Jacod and Protter, 2000), and we see that the first-order intensity of $\mathcal{X}$ is $E[\Lambda_{X|P}(x)]$. Conditioning on the Gaussian process that underlies the parent process, we have

$$E[\Lambda_{X|P}(x)] = \alpha E\left\{E\left[\sum_{j:p_j \in \mathcal{P}} k(p_j - x; \omega)|\mathcal{Z}\right\}\right]. \tag{3.2}$$

Applying Theorem A.2.1 in Appendix A.2 to the inner expectation of (3.2), we have

$$E[\Lambda_{X|P}(x)] = \alpha E\left[\int k(u - x; \omega)\Lambda_{X|P}(u)du\right].$$

The conditions of the Campbell's theorem (Daley and Vere-Jones, 2003) in Appendix A.2 hold, including boundedness of the expectation of $\Lambda_{X|P}(x)$; the expected value is e^{β^*}. Since k is a probability density function, we conclude that the first-order intensity is given by

$$E[\Lambda_{X|P}(x)] = \alpha \int k(u - x; \omega) e^{\beta^*} du = \alpha e^{\beta^*}.$$

Therefore, this process is homogeneous. Similarly, we can derive the second-order intensity, $\lambda^{(2)}(x_1, x_2)$, by evaluating

$$\begin{aligned} E\left[\Lambda_{X|P}(x_1)\Lambda_{X|P}(x_2)\right] &= E\left[\sum_{i:p_i \in \mathcal{P}} \alpha k\left(x_1 - p_i; \omega\right) \sum_{j:p_j \in \mathcal{P}} \alpha k\left(x_2 - p_j; \omega\right)\right] \\ &= E\left[\alpha^2 \sum_{i:p_i \in \mathcal{P}} k\left(x_1 - p_i; \omega\right) k\left(x_2 - p_i; \omega\right)\right] + \qquad (3.3) \\ & \qquad E\left[\alpha^2 \sum_{i:p_i \in \mathcal{P}} \sum_{j \neq i} k\left(x_1 - p_i; \omega\right) k\left(x_2 - p_j; \omega\right)\right] \qquad (3.4) \end{aligned}$$

where line (3.3) corresponds to the contribution of offspring from the same parent and line (3.4) corresponds to that from different parents. This can be shown to equal

$$\begin{aligned} \lambda^{(2)}(x_1, x_2) &= \alpha^2 e^{\beta^*} k * k\left(x_2 - x_1; \omega\right) + \qquad (3.5) \\ & \left(\alpha e^{\beta^*}\right)^2 \int\int \exp\{\sigma^2 \rho(u_2 - u_1; \phi)\} k\left(x_1 - u_1; \omega\right) k\left(x_2 - u_2; \omega\right) \mathrm{d}u_1 \mathrm{d}u_2. \end{aligned}$$

To derive the second-order intensity, we use a similar approach, noting that we need Equation (A.2) of Theorem A.2.1 to conclude that

$$E\left[\sum_{j:p_j \in \mathcal{P}} k(x_1 - p_j; \omega) k(x_2 - p_j; \omega) | \Lambda_P\right] = \int k(x_1 - u; \omega) k(x_2 - u; \omega) \Lambda_P(u) du.$$

In the notation of the theorem, $g(u) = k(x_1 - u)$ and $h(u) = k(x_2 - u)$. The remaining details of the derivation of the second-order intensity are slightly more involved, but do not yield new insights into the process and are therefore omitted. Following this, the pair correlation function can be written as:

$$g(x_1, x_2) = \frac{k * k\left(x_2 - x_1; \omega\right)}{e^{\beta^*}} + \int\int \exp\left\{\sigma^2 \rho(u_2 - u_1; \phi)\right\} k\left(x_1 - u_1; \omega\right) k\left(x_2 - u_2; \omega\right) \mathrm{d}u_1 \mathrm{d}u_2.$$

So long as $\sigma^2\rho(u_2 - u_1;\phi)$ is stationary and isotropic, the corresponding hierarchical cluster process will inherit these properties.

Notice that this process contains Neyman-Scott as a special case. If $\sigma^2 = 0$,

$$\int\int \exp\left\{\sigma^2\rho(u-v;\phi)\right\} k\left(x_1 - u;\omega\right) k\left(x_2 - v;\omega\right) \mathrm{d}u\mathrm{d}v = 1$$

and the unconditional second-order intensity corresponds to that of the Neyman-Scott process. See also related work by Møller and Torrisi (2005) who develop a class of point processes, which they refer to as generalized shot noise Cox processes. In addition to deriving the moment properties, they show that their general class contains many of the well known point processes, such as the Neyman-Scott. In particular, they present an extension of this process, termed the generalized Neyman-Scott process, in which parents are only assumed to follow a stationary point process with a finite intensity.

3.2.3 Additional Assumptions

In our application, for the Gaussian process covariance we use an additive exponential structure where $\text{Cov}[Z(u_1,v_1), Z(u_2,v_2)] = \sigma_s^2\exp(-||u_2-u_1||/\phi_s)+\sigma_t^2\exp(-|v_2-v_1|/\phi_t)$ and therefore $E[Z(u,v)] = -0.5(\sigma_s^2 + \sigma_t^2)$. Moreover, we assume that the displacement distribution of the offspring process is separable in space and time. That is, $k\left(x - p_j;\omega\right) = k_s\left(s - u_j;\omega_s^2\right) k_t\left(t - v_j;\omega_t\right)$ where $k_s(\cdot)$ is a bivariate normal density with mean 0 and covariance $\omega_s^2\boldsymbol{I}_2$ and $k_t(\cdot)$ is a normal density with mean 0 and standard deviation ω_t. Note that our choice of $k_t(\cdot)$ implies that offspring (storm cells) may be observed prior to the corresponding unobserved parent (storm centre). This is well aligned with our application as the parent location represents the storm centre in space and time and therefore we would expect that offspring will appear before and after the storm centre. These parameters all have clear interpretations in relation to our process. The level of spatial and temporal clustering of storms within storm systems is represented by σ_s^2 and σ_t^2 in space and time, respectively, while ϕ_s and ϕ_t represents the size of the storm. Finally, $2\omega_s$ and $2\omega_t$ are the size of the storm in space and time. Note that this formulation leaves the form of the first-order intensity unchanged, but the second order intensity and pair correlation function can be written, respectively, as

$$\begin{aligned}\lambda^{(2)}(x_1,x_2) &= \alpha^2 e^{\beta^*} k_s * k_s\left(s_2-s_1;\omega_s^2\right) k_t * k_t\left(t_2-t_1;\omega_t\right) + \left(\alpha e^{\beta^*}\right)^2 \times \\ &\quad \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2-u_1||/\phi_s)}\right\} k_s\left(s_1-u_1;\omega_s^2\right) k_s\left(s_2-u_2;\omega_s^2\right) \mathrm{d}u_1\mathrm{d}u_2 \times \\ &\quad \int_{\mathbb{R}}\int_{\mathbb{R}} \exp\left\{\sigma_t^2 e^{(-|v_2-v_1|/\phi_t)}\right\} k_t\left(t_1-v_1;\omega_t\right) k_t\left(t_2-v_2;\omega_t\right) \mathrm{d}v_1\mathrm{d}v_2\end{aligned}$$

and

$$\begin{aligned}g(x_1,x_2) &= \frac{k_s * k_s(s_2-s_1;\omega_s^2) k_t * k_t(t_2-t_1;\omega_t)}{e^{\beta^*}} + \\ &\quad \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2-u_1||/\phi_s)}\right\} k_s\left(s_1-u_1;\omega_s^2\right) k_s\left(s_2-u_2;\omega_s^2\right) \mathrm{d}u_1\mathrm{d}u_2 \times \\ &\quad \int_{\mathbb{R}}\int_{\mathbb{R}} \exp\left\{\sigma_t^2 e^{(-|v_2-v_1|/\phi_t)}\right\} k_t\left(t_1-v_1;\omega_t\right) k_t\left(t_2-v_2;\omega_t\right) \mathrm{d}v_1\mathrm{d}v_2.\end{aligned}$$

3.2.4 Parameter Estimation

Because the first-order intensity, λ, is constant, using Equation (2.1) λ is estimated as $N/|S \times T|$. For the second-order parameters we follow Prokešová and Dvořák (2014) who proposed minimum contrast estimation via the spatial and temporal projection processes. Letting $\mathcal{X}_s$ denote the spatial projection process, $\mathcal{X}_s = \{s : (s,t) \in \mathcal{X} \cap (S \times T)\}$ and similarly for the temporal projection process $\mathcal{X}_t = \{t : (s,t) \in \mathcal{X} \cap (S \times T)\}$. Using the lower dimensional processes for estimation requires that both the spatial and temporal projections be simple.

The first-order intensity of the spatial projection process can be estimated by integrating time out of the unconditional first-order intensity and analogously integrating out space for the temporal projection. That is, $\lambda_s = \int_T \lambda \mathrm{d}t = \lambda|T|$ and for time, $\lambda_t = \int_S \lambda \mathrm{d}s = \lambda|S|$ where $\lambda = \alpha e^{\beta^*}$. The second-order intensities of the projection processes can be derived in a similar manner. Specifically, for the spatial projection,

$$\begin{aligned}
\lambda_s^{(2)}(s_1, s_2) &= \int_T \int_T \lambda^{(2)}[(s_1, t_1), (s_2, t_2)] \mathrm{d}t_1 \mathrm{d}t_2 \\
&= \alpha\lambda k_s * k_s \left(s_2 - s_1; \omega_s^2\right) \int_T \int_T k_t * k_t \left(t_2 - t_1; \omega_t\right) \mathrm{d}t_1 \mathrm{d}t_2 + \\
&\quad \lambda^2 \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2 - u_1||/\phi_s)}\right\} k_s \left(s_1 - u_1; \omega_s^2\right) k_s \left(s_2 - u_2; \omega_s^2\right) \mathrm{d}u_1 \mathrm{d}u_2 \times \\
&\quad \int_T \int_T \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left\{\sigma_t^2 e^{(-|v_2 - v_1|/\phi_t)}\right\} k_t \left(t_1 - v_1; \omega_t\right) k_t \left(t_2 - v_2; \omega_t\right) \mathrm{d}v_1 \mathrm{d}v_2 \mathrm{d}t_1 \mathrm{d}t_2 \\
&= \alpha\lambda k_s * k_s \left(s_2 - s_1; \omega_s^2\right) C_s' + \\
&\quad \lambda^2 \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2 - u_1||/\phi_s)}\right\} k_s \left(s_1 - u_1; \omega_s^2\right) k_s \left(s_2 - u_2; \omega_s^2\right) \mathrm{d}u_1 \mathrm{d}u_2 C_s''
\end{aligned}$$

with C_s' and C_s'' being the following constants:

$$C_s' = \int_T \int_T k_t * k_t \left(t_2 - t_1; \omega_t\right) \mathrm{d}t_1 \mathrm{d}t_2$$

and

$$C_s'' = \int_T \int_T \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left\{\sigma_t^2 e^{(-|v_2 - v_1|/\phi_t)}\right\} k_t \left(t_1 - v_1; \omega_t\right) k_t \left(t_2 - v_2; \omega_t\right) \mathrm{d}v_1 \mathrm{d}v_2 \mathrm{d}t_1 \mathrm{d}t_2.$$

The second-order intensity of the temporal projection process can be derived analogously:

$$\begin{aligned}
\lambda_t^{(2)}(t_1, t_2) &= \int_S \int_S \lambda^{(2)}[(s_1, t_1), (s_2, t_2)] \mathrm{d}s_1 \mathrm{d}s_2 \\
&= \alpha\lambda k_t * k_t \left(t_2 - t_1; \omega_t\right) C_t' + \\
&\quad \lambda^2 \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left\{\sigma_t^2 e^{(-|v_2 - v_1|/\phi_t)}\right\} k_t \left(t_1 - v_1; \omega_t\right) k_t \left(t_2 - v_2; \omega_t\right) \mathrm{d}v_1 \mathrm{d}v_2 C_t''
\end{aligned}$$

with the constants

$$C_t' = \int_S \int_S k_s * k_s \left(s_2 - s_1; \omega_s^2\right) \mathrm{d}s_1 \mathrm{d}s_2$$

and

$$C_t'' = \int_S \int_S \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2 - u_1||/\phi_s)}\right\} k_s \left(s_1 - u_1; \omega_s^2\right) k_s \left(s_2 - u_2; \omega_s^2\right) \mathrm{d}u_1 \mathrm{d}u_2 \mathrm{d}s_1 \mathrm{d}s_2.$$

The corresponding theoretical pair correlation function for the spatial projection process is

$$\begin{aligned}
& g_s(||s_1 - s_1||) \\
= \; & \frac{\lambda_s^{(2)}(s_1, s_2)}{\lambda_s(s_1)\lambda_s(s_2)} \\
= \; & k_s * k_s\left(s_2 - s_1; \omega_s^2\right) C_s'^* + \\
& \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp\left\{\sigma_s^2 e^{(-||u_2-u_1||/\phi_s)}\right\} k_s\left(s_1 - u_1; \omega_s^2\right) k_s\left(s_2 - u_2; \omega_s^2\right) \mathrm{d}u_1 \mathrm{d}u_2 C_s''^*
\end{aligned}$$

where $C_s'^* = C_s'/(e^{\beta^*}|T|^2)$ and $C_s''^* = C_s''/|T|^2$. For the temporal projection, $g_t(|t_2 - t_1|)$, is analogously $\frac{\lambda_t^{(2)}(t_1,t_2)}{\lambda_t(t_1)\lambda_t(t_2)}$.

Minimum contrast estimation is then accomplished by minimizing the following criteria:

$$D_s = \int_0^{r_{\text{corr}}} \left\{[g_s(u)]^c - [\hat{g}_s(u)]^c\right\}^2 \mathrm{d}u \tag{3.6}$$

and

$$D_t = \int_0^{t_{\text{corr}}} \left\{[g_t(v)]^c - [\hat{g}_t(v)]^c\right\}^2 \mathrm{d}v \tag{3.7}$$

where $\hat{g}_s(u)$ and $\hat{g}_t(v)$ are the empirical spatial and temporal pair correlation functions, as written in Equation (2.2). The range of correlation for the spatial and temporal projection processes are represented by r_{corr} and t_{corr} and we set $c = 0.25$. As mentioned in Section 2.2.3, Diggle (2003) suggested this as a variance stabilizing constant for clustered patterns.

From Equation (3.6), we are able to estimate σ_s^2, ϕ_s, ω_s^2, $C_s'^*$ and $C_s''^*$ and the analogous temporal parameters can be estimated by Equation (3.7). The first-order parameters, β^*

and α, may be calculated from either of the projection processes as

$$\begin{aligned}
\widehat{\exp\{\beta_s^*\}} &= \frac{1}{|T|^2 \hat{C}_s'^*} \int_T \int_T k_t * k_t \left(t_2 - t_1; \hat{\omega}_t\right) \mathrm{d}t_1 \mathrm{d}t_2 \\
\hat{\beta}_s^* &= \log\{\widehat{e^{\beta_s^*}}\} \\
\hat{\alpha}_s &= \hat{\lambda} / \widehat{\exp\{\beta_s^*\}}
\end{aligned}$$

or

$$\begin{aligned}
\widehat{\exp\{\beta_t^*\}} &= \frac{1}{|S|^2 \hat{C}_t'^*} \int_S \int_S k_s * k_s \left(s_2 - s_1; \hat{\omega}_s^2\right) \mathrm{d}s_1 \mathrm{d}s_2 \\
\hat{\beta}_t^* &= \log\{\widehat{e^{\beta_t^*}}\} \\
\hat{\alpha}_t &= \hat{\lambda} / \widehat{\exp\{\beta_t^*\}}.
\end{aligned}$$

We return to address this again in Section 3.3.

3.2.5 Identifying Multiple Levels of Clustering

In hierarchical cluster processes identifiability issues arise if the scales of the small- and large-scale clustering do not differ. For our proposed hierarchical cluster process, this implies that the scale of clustering in the offspring process must be smaller than that of the parent process. To identify these two levels of clustering, we follow the two-stage approach of Wiegand et al. (2007), which is done separately for the spatial and temporal projection processes. Using the temporal projection process from April 2003 to illustrate, in Figure 3.7 consider the point labelled τ_t. (We discuss estimation strategies for τ_t below.) This represents the division between small- and large-scale clustering. Parameter estimation proceeds as follows: first, we estimate parameters corresponding to the parent log-Gaussian Cox process by minimizing Equation (3.7), but rather than integrating over the range of correlation (0 to t_{corr}), we integrate over the range of the large-scale clustering (τ_t to t_{corr}). In this step, the theoretical pair correlation function corresponds to that of the log-Gaussian Cox process:

$$g_t(v) = \sigma_t^2 \exp\left(-|v|/\phi_t\right).$$

Note that for the log-Gaussian Cox process, because the full spatio-temporal pair correlation function is separable (due to the additive space-time covariance structure in the Gaussian process), as discussed in Prokešová and Dvořák (2014), we can employ minimum contrast estimation utilizing the lower dimensional spatial and temporal processes, rather than the projection processes, without much loss of accuracy. Second, conditional on $\hat{\sigma}_t^2$ and $\hat{\phi}_t$, we estimate the remaining parameters by minimizing D_t, as written in Equation (3.7). Depending on the application, the value of τ_t may be known a priori. However, as in our motivating example, we have no knowledge of these values. We therefore treat these as tuning parameters and, as such, try a range of values and select the minimizer of D_t.

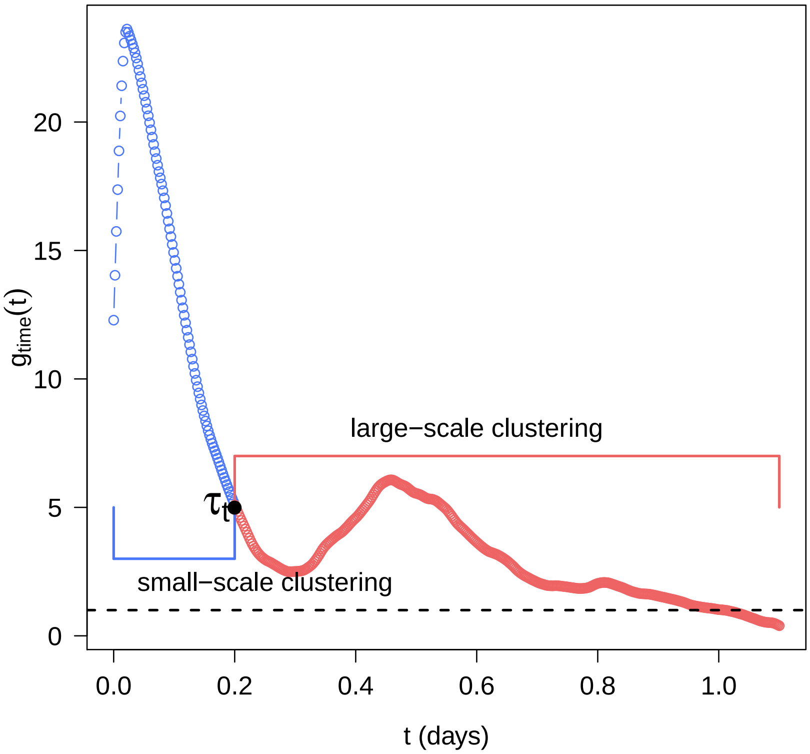


Figure 3.7: Empirical pair correlation function of the temporal projection process from April 2003.

3.2.6 Confidence Intervals

As mentioned previously, for minimum contrast estimation resampling-based methods are required to obtain confidence intervals for the clustering parameters. Specifically, we employ the non-parametric marked point bootstrap proposed in Loh (2008). The idea is that the empirical pair correlation function for each of the projection processes can be decomposed into their observation-specific contributions. Using the temporal projection as an example, when $\hat{\lambda}_t = N/|T|$,

$$\begin{aligned} g(v) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\frac{T}{2N}v_{ij}\kappa_{\epsilon}(v-|t_i-t_j|) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\chi_{t_i,t_j}(v). \end{aligned}$$

That is, associated with the ith observation is the quantity $\sum_{j\neq i}\chi_{t_i,t_j}(v)$, termed a mark, which depends on the distance between t_i and t_j and represents its contribution to the empirical pair correlation function.

Letting $m_i(v) = \sum_{j\neq i}\chi_{t_i,t_j}(v)$, we can obtain an estimate of the empirical pair correlation function using the marks $\hat{g}(v) = \frac{1}{N}\sum_{i=1}^{N} m_i(v)$. For the r_1st resample, $r_1 = 1, 2, \ldots, R_1$, if $m_i^{*(r_1)}$ represents the marks, which are block resampled from $m_i(v)$ using fixed or random blocks, we can obtain bootstrapped estimates of the empirical pair correlation function, $\hat{g}^{*(r_1)}(v) = \frac{1}{N^{*(r_1)}}\sum_{i=1}^{N^{*(r_1)}} m_i^{*(r_1)}(v)$ with $N^{*(r_1)} = \sum_i N_i^{*(r_1)}$ and $N_i^{*(r_1)}$ being the number of times the ith observation is resampled. By substituting $\hat{g}^{*(r_1)}(v)$ into Equation (3.7) we can obtain bootstrap-based confidence intervals for the point process parameters. This was successfully applied to a generalized shot noise Cox process in Yau and Loh (2012). For large point pattern data sets, this approach has the advantage of being considerably less computationally demanding than a parametric bootstrap. However, because the properties of this non-parametric bootstrap have not been studied theoretically we perform a double bootstrap with a length adjustment for calibration. As shown in Han and Braun (2015), this technique improves coverage, which for dependent data may be low relative to the nominal rate, and decreases confidence interval length relative to the standard percentile bootstrap. Specifically, for each of the $r_1 = 1, \ldots, R_1$ resamples performed in the bootstrap, we take $r_2 = 1, 2, \ldots, R_2$ re-resamples. Letting $m_i^{**(r_1,r_2)}(v)$ represent the re-resampled marks, we calculate $\hat{g}^{**(r_1,r_2)}(v) = \frac{1}{N^{**(r_1,r_2)}} \sum_{i=1}^{N^{**(r_1,r_2)}} m_i^{**(r_1,r_2)}(v)$ where $N^{**(r_1,r_2)}$ is the number of samples for the r_2nd re-resample of the r_1st resample. To obtain parameter estimates based on the re-resampled observations, we can again substitute $\hat{g}^{**(r_1,r_2)}(v)$ into Equation (3.7). If $\hat{\theta}$ denotes the parameter estimate from the original point pattern, $\hat{\hat{\theta}}_{r_1}$ denotes that of the r_1st resample and $\hat{\hat{\hat{\theta}}}_{r_1,r_2}$ from the (r_1, r_2) re-resample, the α-level confidence interval for $\hat{\theta}$ can be calculated as

$$\left(\hat{\hat{\theta}} + \hat{\tau}_1 + \hat{c}_1,\ \hat{\hat{\theta}} + \hat{\tau}_2 + \hat{c}_2\right)$$

where $\hat{\tau}_1$ is the $\alpha/2$ sample percentile of the distribution of $\hat{\hat{\hat{\theta}}}_{r_1,r_2} - \hat{\hat{\theta}}_{r_1}$ and $\hat{\tau}_2$ is the $1-\alpha/2$ sample percentile. Using the independence approximation described in Han and Braun (2015), $\hat{c}_1$ may be obtained from

$$\sqrt{1-\alpha} \doteq \hat{P}\left(\hat{\hat{\theta}} + \hat{\tau}_1 + \hat{c}_1 \leq \hat{\theta}\right) \tag{3.8}$$

and $\hat{c}_2$ from

$$\sqrt{1-\alpha} \doteq \hat{P}\left(\hat{\hat{\theta}} + \hat{\tau}_2 + \hat{c}_2 \geq \hat{\theta}\right). \tag{3.9}$$

3.2.7 Goodness of Fit

Goodness of fit for this process was assessed by comparing the fitted pair correlation functions for the spatial and temporal projection processes to that of the empirical. We also compared this with simpler, commonly employed cluster processes, namely the Neyman-Scott process and the log-Gaussian Cox process. Additionally, we compared the estimated values of storm size ($2\hat{\omega}_t$ in time and $2\hat{\omega}_s$ in space) with τ_t and τ_s, as appropriate, to see how well we could differentiate between the two levels of clustering in these data.

3.3 Application to Storm Cell Data

In this section, we summarize results from our analysis of the 2003 Bismarck, North Dakota storm cell data. The spatial region employed was a disc of radius 300 km centred at the Bismarck radar station. Although the radar is able to detect storm cells up to 460 km, when visualizing these data, detection issues were obvious after approximately 300 km. In three dimensions this pattern constitutes a simple point process. However, because these data are measured with error, as mentioned in Section 3.1, the spatial and temporal projection processes contain some multiplicities. Therefore, in order to employ methods for simple point processes, we jitter the UTM X, UTM Y and Julian dates uniformly according to this measurement error; Baddeley et al. (2015) suggest this as a simple technique for modelling non-simple point processes.

To estimate the empirical pair correlation functions for the spatial projection processes, we used the Epanechnikov kernel with a bandwidth of $\delta_s/\sqrt{\hat{\lambda}_s}$ where $0.1 \leq \delta_s \leq 0.25$, as suggested by Stoyan and Stoyan (1996) and the translation edge correction. For the temporal projection processes, the Epanechnikov kernel was again employed with a bandwidth of $\delta_t/\hat{\lambda}_t$, as suggested by Vio et al. (2007) where δ_t was chosen by trial and error, but was similar in magnitude to δ_s. The temporal isotropic edge correction was used here. To estimate $\hat{c}_1$ and $\hat{c}_2$ as in Equations (3.8) and (3.9), we performed a double bootstrap with $R_1 = 100$ first-level resamples and $R_2 = 50$ second level re-resamples with random blocking. Once $\hat{c}_1$ and $\hat{c}_2$ were estimated, we then ran 1000 first-level bootstrap runs to calculate the final confidence intervals. For the spatial projections, we utilized circular blocks with a radius of 100 km at the first level and for the temporal bootstrap, we used four blocks corresponding to sizes of 7.5 days for April and June (months with 30 days) and 7.75 days for the remaining months with 31 days. At the second-level, block radius or block length was halved.

Table 3.1 displays the estimated point process parameters which all have clear interpretations in terms of our application. The parent process variances σ_t^2 and σ_s^2 represent the amount of temporal and spatial clustering or the intensity of storms within storm systems with larger estimates indicative of stronger clustering. The scale parameters for the parent process may be interpreted in terms of storm system size over time and space with larger values associated with longer lasting and larger storm systems, respectively. Finally, the variances or standard deviations in the offspring distributions are functions of storm size; an average storm is estimated to have a radius of $2\omega_s$ km in space and a duration of $2\omega_t$ days. In this table, the estimates of the first-order parameters are based on the results from the temporal projection processes.

As can be seen in Table 3.1, regardless of month, storms are short lasting with $2\hat{\omega}_t$ averaging between 0.10 and 0.18 days for all months. However, the average storm size ($2\hat{\omega}_s$) increases with month. Therefore, although there is no considerable difference in the average storm duration, their size increases between April and August. Note that storm sizes in July and August are quite large in comparison with the size of the disc and so edge effects associated with storm complexes moving into or exiting the disc may be substantial. As well, the average number of storm cells per storm is smallest in April and considerably larger during the remaining four months. This is aligned with what we can see in Figures 3.2 - 3.6. Storm systems in April and August are the shortest with ϕ_t estimated to be less than one day, specifically 0.68 (0.25, 0.75) and 0.85 (0.52, 1.76), respectively, with the numbers in brackets representing the 95% confidence intervals. In May and July, storm systems last longer with $\hat{\phi}_t$ being 1.45 (1.13, 4.86) and 1.06 (0.76, 4.67). Storm systems in June have the longest duration with ϕ_t estimated as 3.59 (2.85, 8.77). Based on this, it would be interesting to understand if June is climatologically different or if this is specific to June 2003. Storm system size is smallest in May when $\hat{\phi}_s$ is 87.06 (69.82, 112.05), but is of similar size for all remaining months with estimates between 101.48 (84.80, 128.02) and 115.32 (93.39, 152.72), corresponding to the months of June and August, respectively. A similar level of temporal clustering of storms within storm systems is found across all months, as shown by the magnitudes of $\hat{\sigma}_t^2$ with the exception of June when $\hat{\sigma}_t^2$ is 0.99 (0.28, 1.41); this level of intensity is smaller than the other months, which all have estimates greater than two. All months have similar magnitudes for $\hat{\sigma}_s^2$, although this is largest in May which is estimated as 3.51 (2.75, 4.57) and smallest in August, estimated as 1.80 (0.79, 2.87). Therefore, with regards to the temporal process, the month of June, which has the longest lasting storm systems, also has the least intense. Conversely for the spatial process in May, although the amount of clustering is high, the storm systems are the smallest. We note that increasing c to 0.5 changed the parameters estimates slightly, but these were within the confidence intervals displayed in Table 3.1.

3.3.1 Goodness of fit

Figure 3.8 displays the empirical and fitted pair correlation functions by month for our proposed hierarchical cluster process as well as those of the Neyman-Scott and log-Gaussian Cox processes. For the temporal projection processes, the Neyman-Scott is clearly inferior to the log-Gaussian Cox and the hierarchical cluster processes as they consistently underestimate the small-scale clustering and tend to overestimate the moderate-scale clustering. The log-Gaussian Cox processes, however, are a considerable improvement as the fitted pair correlation functions more closely match that of the empirical estimates, although quite often they still underestimate the small-scale clustering. Our proposed cluster process has the flexibility required to capture these hierarchical trends and closely matches the empirical pair correlation functions. By examining the temporal projections in the left panel of Figure 3.8, it is evident that we appear to be differentiating between the small- and large-scale clustering as τ_t always occurs between the two peaks. This mimics what is shown in Table 3.1 with $2\hat{\omega}_t$ always well below both $\hat{\tau}_t$ and $\hat{\phi}_t$.

For the spatial projection processes, the log-Gaussian Cox is inferior to the Neyman-Scott and the hierarchical clustering processes with the exception of April when the results are satisfactory; for May to August it consistently overestimates the small-scale clustering and underestimates the large-scale clustering. The results from the Neyman-Scott processes are comparable to what we obtain from the hierarchical cluster processes. This is to be expected for a number of reasons. First, in the empirical spatial pair correlation functions (right panel of Figure 3.8), unlike for the temporal projection processes, it is difficult to differentiate between the two scales of clustering. Although $\hat{\omega}_s$ is consistently less than $\hat{\phi}_s$, $2\hat{\omega}_s$ is not always less than $\hat{\tau}_s$.

For comparison, Table 3.2 summarizes the parameter estimates and 95% confidence intervals when fitting the Neyman-Scott process to these data. To be consistent with Table 3.1, the first-order parameters from Table 3.2 are based on estimates from the temporal projection processes. For the temporal processes, cluster size ($2\hat{\omega}_t$) is consistently larger than what was estimated in the hierarchical cluster processes. Similarly for that of the spatial processes ($2\hat{\omega}_s$), with the exception of July and August. Moreover, for all months, the estimated number of offspring per parent ($\hat{\alpha}$) is much larger than what we observe in the hierarchical cluster process. This emphasizes that, as was shown in Figure 3.8, the Neyman-Scott process is picking up on the large-scale clustering and missing the small-scale behaviour.

3.4 Discussion

The Neyman-Scott process, as outlined in Section 2.2.2, is a widely applicable parent-child cluster process for modelling spatial and spatio-temporal point patterns. However, its use is limited to modelling data with one level of clustering. That is, data in which the process concludes after an unobserved parent process is assumed to generate the offspring process. Extensions to double or “multigeneration” cluster processes (e.g. Wiegand et al., 2007) have been developed where offspring from the first generation process become parents in the second generation which produce further offspring. Superposed Neyman-Scott processes have also been employed for scenarios in which two sizes of clusters are present, but the data do not have the hierarchical structure required of the double cluster process (e.g. Wiegand et al., 2007; Tanaka and Ogata, 2014; Stoyan and Stoyan, 1996). For our application, the superposed Neyman-Scott process is inappropriate as it does not incorporate the known storm system hierarchy as shown in the temporal projection processes in Figures 3.2 - 3.6 and as discussed in Mohee and Miller (2010). Meanwhile, double cluster processes are inadequate as the first-generation parent process is homogeneous, which again, based on knowledge of storm systems is too strong an assumption. Therefore, in this chapter, we generalized the Neyman-Scott cluster process by allowing the parents to follow a log-Gaussian Cox process, rather than restricting them to be homogeneous Poisson. Not only does this permit spatio-temporal correlation in the

		First-Order			Temporal			Spatial	
		Est.	CI		Est.	CI		Est.	CI
April									
	β^*	-13.89	(-14.81, -13.30)	σ_t^2	2.25	(0.57, 5.31)	σ_s^2	2.42	(1.91, 2.93)
	α	118.18	(93.26, 276.53)	ϕ_t	0.68	(0.25, 0.75)	ϕ_s	112.27	(99.73, 149.70)
				ω_t	0.05	(0.01, 0.27)	ω_s^2	2707.25	(1915.06, 3727.86)
May									
	β^*	-14.03	(-14.07, -7.91)	σ_t^2	2.08	(0.93, 7.53)	σ_s^2	3.51	(2.75, 4.57)
	α	771.07	(427.16, 945.53)	ϕ_t	1.45	(1.13, 4.86)	ϕ_s	87.06	(69.82, 112.05)
				ω_t	0.08	(0.02, 0.26)	ω_s^2	2983.18	(2478.61, 3838.90)
June									
	β^*	-13.73	(-14.37, -13.68)	σ_t^2	0.99	(0.28, 1.41)	σ_s^2	2.63	(1.85, 3.53)
	α	779.28	(759.65, 1383.75)	ϕ_t	3.59	(2.85, 8.77)	ϕ_s	101.48	(84.80, 128.02)
				ω_t	0.09	(0.06, 0.11)	ω_s^2	5129.08	(4439.32, 6720.97)
July									
	β^*	-13.80	(-15.20, -3.56)	σ_t^2	2.07	(0.75, 4.53)	σ_s^2	2.15	(1.25, 3.45)
	α	874.47	(500.01, 4545.01)	ϕ_t	1.06	(0.76, 4.67)	ϕ_s	105.75	(78.17, 137.35)
				ω_t	0.08	(0.01, 0.54)	ω_s^2	8183.81	(6690.32, 9809.72)
August									
	β^*	-13.90	(-15.33, -6.44)	σ_t^2	3.02	(2.44, 8.95)	σ_s^2	1.80	(0.79, 2.87)
	α	616.04	(350.35, 5742.19)	ϕ_t	0.85	(0.52, 1.76)	ϕ_s	115.32	(93.39, 152.72)
				ω_t	0.06	(1.26×10^{-4}, 0.13)	ω_s^2	9621.95	(7915.58, 11846.99)

Table 3.1: Parameter estimates (Est.) and 95% confidence intervals (CIs) for the spatio-temporal hierarchical cluster process fit by month. Note that ϕ_t and ω_t are measured in days while σ_t^2 has the units of squared days. Similarly ϕ_s has the units of km, and σ_s^2 and ω_s^2 are in km^2.

		Est.	CI
April			
	λ_p $(\times 10^{-7})$	3.69	$(4.53\times10^{-5},\ 1.71\times10^{4})$
	α	296.79	(271.51, 794.64)
	ω_t	0.25	(0.01, 0.97)
	ω_s^2	3326.24	(2508.07, 4814.12)
May			
	λ_p $(\times 10^{-7})$	2.35	(1.18, 3.82)
	α	2663.95	(1858.99, 4681.32)
	ω_t	0.51	(0.36, 0.81)
	ω_s^2	3245.03	(2445.18, 4516.75)
June			
	λ_p $(\times 10^{-7})$	4.28	(2.39, 5.55)
	α	1981.05	(1659.93, 3389.69)
	ω_t	0.58	(0.21, 0.80)
	ω_s^2	5343.94	(3681.92, 7433.54)
July			
	λ_p $(\times 10^{-7})$	2.90	(2.05, 6.11)
	α	3074.49	(1551.60, 4189.27)
	ω_t	0.34	(0.25, 0.52)
	ω_s^2	7703.45	(5445.72, 10595.18)
August			
	λ_p $(\times 10^{-7})$	1.74	(1.49, 4.21)
	α	3236.89	(2154.65, 3629.63)
	ω_t	0.24	(0.15, 0.34)
	ω_s^2	9101.59	(5958.85, 12868.84)

Table 3.2: Parameter estimates (Est.) and 95% confidence intervals (CIs) for the spatio-temporal Neyman-Scott process fit by month. Note that ω_t has the units of days and ω_s^2 is in km^2.

parent process, but it may be employed to model hierarchically clustered point patterns. As we demonstrated, this type of process also enables inference on the unobserved storms and storm systems.

The utility of this model was shown through an analysis of monthly storm cell data from the Bismarck, North Dakota radar station in 2003. Parameter estimation was accomplished through minimum contrast estimation of the lower dimensional spatial and temporal projection processes and we advocated the use of a two-stage technique due to identifiability issues that arose. For this approach, the pairwise distance separating small- and large-scale clustering was treated as a tuning parameter and we note that the resulting parameter estimates were not sensitive to this choice so long as it was greater than the size of the storms and smaller than the storm system scale.

Goodness of fit was assessed by comparing the fitted pair correlation functions from our proposed process to the empirical functions and also to simpler point processes employed for clustered data. In general, we saw that for the temporal projection processes, in which a hierarchical structure was obvious from the empirical pair correlation functions, this process offered an improved fit in addition to the ability to make inference on the unobserved storms and storm systems. In the spatial projection processes, when the two levels of clustering were not as pronounced, the results were comparable to that of the Neyman-Scott. We believe this is due to the use of spatial and temporal projection processes for parameter estimation which makes it difficult to discern the different levels of spatial clustering. However, as discussed, the results from the Neyman-Scott process were not completely satisfactory either.

In order to utilize the lower dimension projection processes for parameter estimation, an additive space-time covariance structure in the Gaussian process was required. However, developing an approach to parameter estimation using the full spatio-temporal process is of interest to get improved estimates and relax this additivity requirement within the Gaussian process. We return to this in Chapter 6. Further extensions include allowing the mean number of storm cells per storm and cluster size to vary as a function of space and time as well as the inclusion of covariates into either the parent or the offspring process. The former would result in a non-stationary process, although it could help to decipher any inhomogeneity from clustering. Extensions to account for limitations in these data would also be important. For example, there is likely a diurnal cycle of cell development and dissipation with peak activity in the late afternoon as well as the possibility of detection limitations (e.g. a cell will be more difficult to detect at the edges of the radar field of view than in the central part of its field of view). Modifications to include a cyclical element to the cell development process as well as probabilities of inclusions to data based on distance from the radar centre require further scientific input.

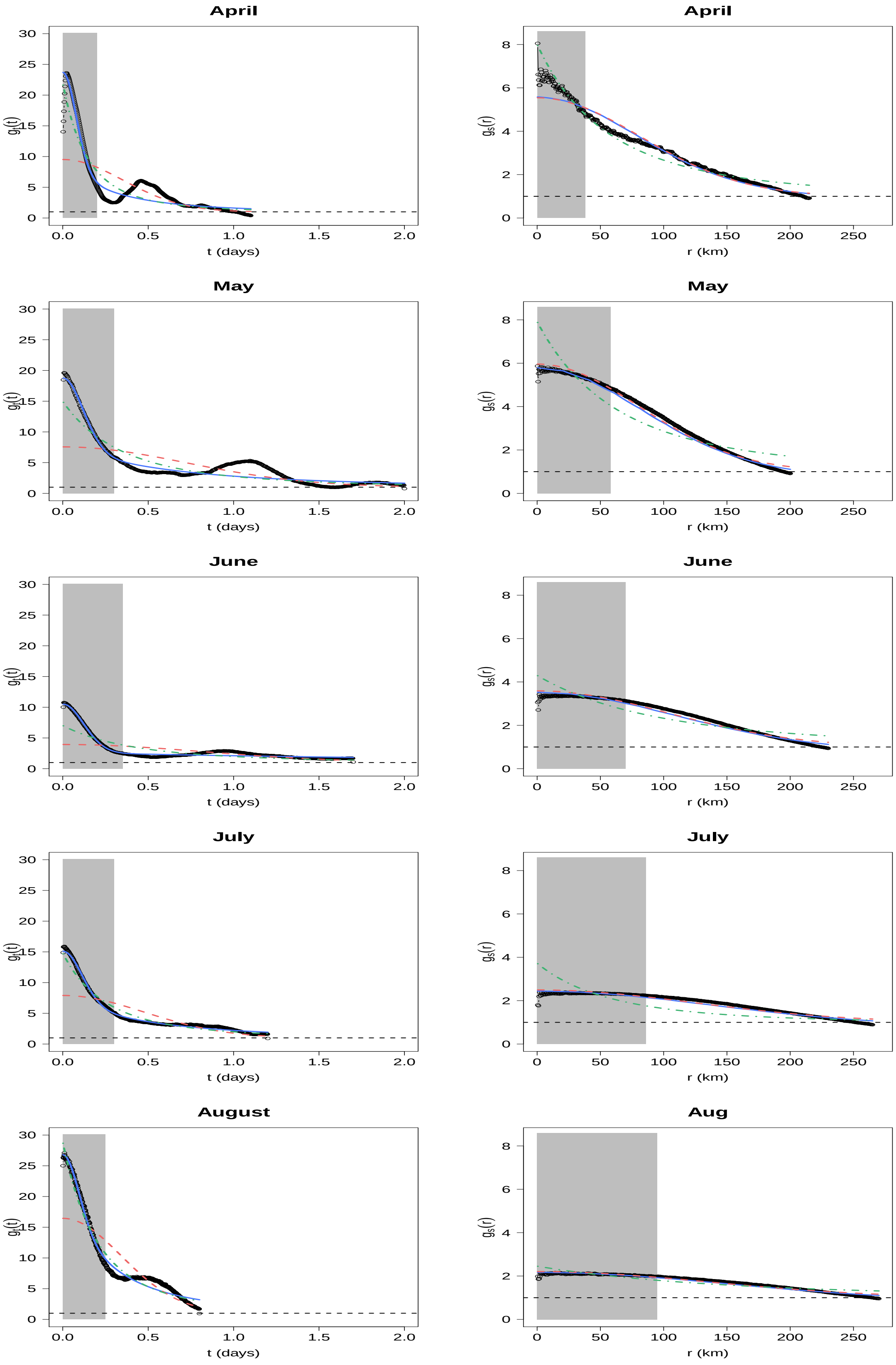


Figure 3.8: Temporal (left panel) and spatial (right panel) empirical (black points) and fitted pair correlation functions for the hierarchical cluster process (blue lines), the Neyman-Scott process (red dashed lines) and the log-Gaussian Cox process (green dotted-dashed lines) fit by month. The shaded region corresponds to the small-scale clustering.

Chapter 4

A Joint Model for a Hierarchical Cluster Process with Evolving Marks

This work builds on the hierarchical cluster process developed previously by incorporating storm cell movement through a marked point process and thereby modelling a storm cell's complete trajectory. Specifically, in this chapter we develop a joint model for storm cell detection and evolution by incorporating multivariate marks for duration, speed and direction into the hierarchical cluster process. In conjunction with Chapter 3, these results may be employed by power system operators who, in real time, monitor power flow on transmission lines and perform simulations of different contingencies in case of failure. Knowing that over a period of time weather is more likely to take out power lines allows transmission operators the time required to initiate defensive strategies which, while imposing costs, minimize the impact of power system interruptions. It is our hope that this understanding of storm systems and storm cell movement could be utilized to reduce the need to operate in such sub-optimal modes.

4.1 Data Description

Once storm cells are identified by the Storm Cell Identification and Tracking algorithm (see Appendix A.1 for details) they are assigned a unique identifier and their trajectories are tracked. Based on a storm cell's recorded trajectory, we calculate its duration in hours and its speed in km/hour. The radians between the first and last recorded observation of a storm cell starting at zero and proceeding counter clockwise are used as a storm cell's direction. Radar scans only occur every 4.5 to 6 minutes during precipitation events. Because of this, 39% of the storm cells are only observed once by the radar. Figures 4.1 - 4.5 display storm cell trajectories from April 2003 - August 2003 colour coded by duration, speed and direction. There appear to be some spatio-temporal trends in duration. Storm cells within a storm appear to have a similar direction and speed. This is expected as these quantities are related to prevailing wind speed and direction.

4.2 Joint Models for Speed and Direction

Joint models for vector fields (speed and direction) are commonly utilized for modelling hurricane surface wind fields, for example Modlin et al. (2012) and Reich and Fuentes (2007). There are two common approaches to this. The first being to model the so-called u- and v-components, representing the west-east and north-south elements of a vector field. These may be jointly normally distributed and modelled in a so-called shared component model in which two outcomes are assumed to have a common spatial random effect. The advantages of this technique include increased efficiency when estimating the joint spatial structure (Feng and Dean, 2012) and being able to avoid directly modelling circular data which pose difficulties as standard distributions are no longer applicable. This may be challenging, however, if the resulting u- and v-components are heavy-tailed. The second method, which we utilize, employs modelling direction following a circular distribution and speed conditional on a function of direction. Joint modelling in this scenario is simplified as it employs ecological regression methods (e.g. Held et al., 2005) where a function of one explanatory variable (direction) serves as a predictor of the other

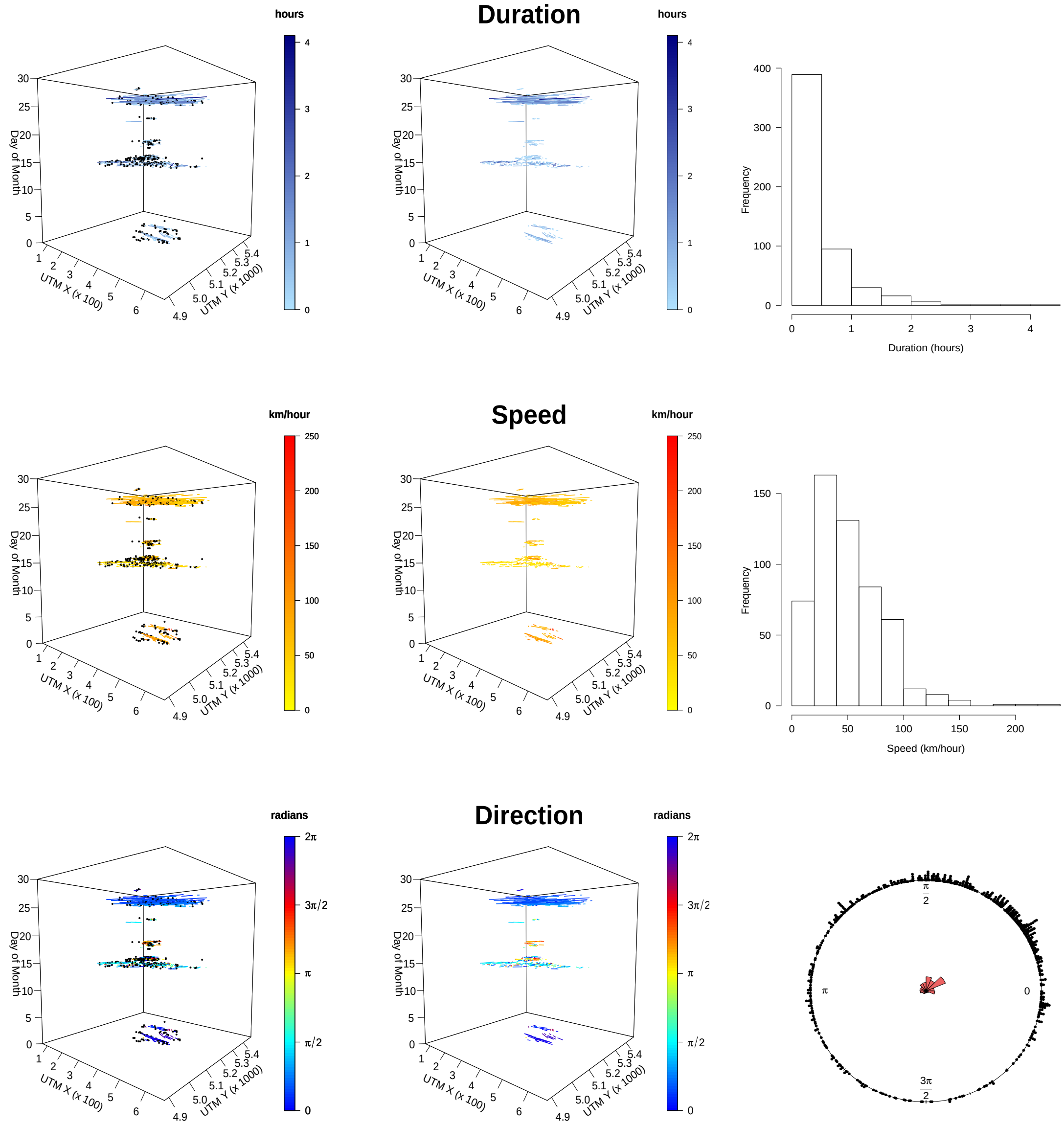


Figure 4.1: Three dimensional plots and histograms of storm cell duration (top row), speed (middle row) and direction (bottom row) from April 2003. In the left panel, black points represent storm cells only observed on one occasion.

(speed). Such a model does not require joint estimation techniques. With the ecological regression approach, Modlin et al. (2012) accounted for spatial correlation in speed using a conditional autoregressive model and arrived at a circular conditional autoregressive structure when assuming direction to be distributed according to a wrapped normal. For hurricane surface wind fields, Reich and Fuentes (2007) employed a stick breaking prior

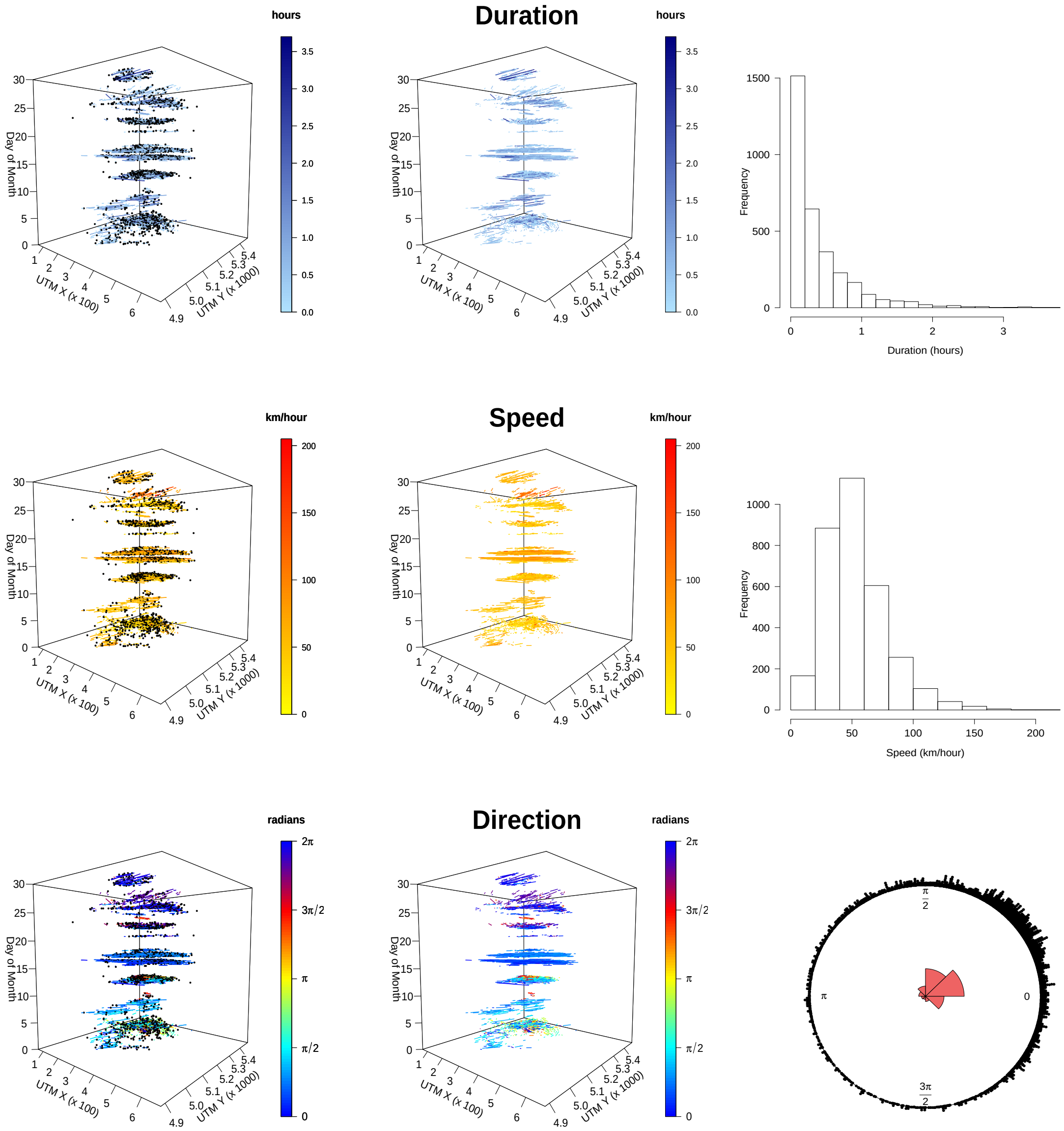


Figure 4.2: Three dimensional plots and histograms of storm cell duration (top row), speed (middle row) and direction (bottom row) from May 2003. In the left panel, black points represent storm cells only observed on one occasion.

to account for erratic behaviour in the u- and v-components. Wang and Gelfand (2014) proposed a projected Gaussian process for spatial and spatio-temporal wave direction.

In general vector field data may be heavy-tailed and highly variable. To describe these data, models such as the class of generalized additive models for location, scale and shape (Rigby and Stasinopoulos, 2005) or double hierarchical generalized linear models

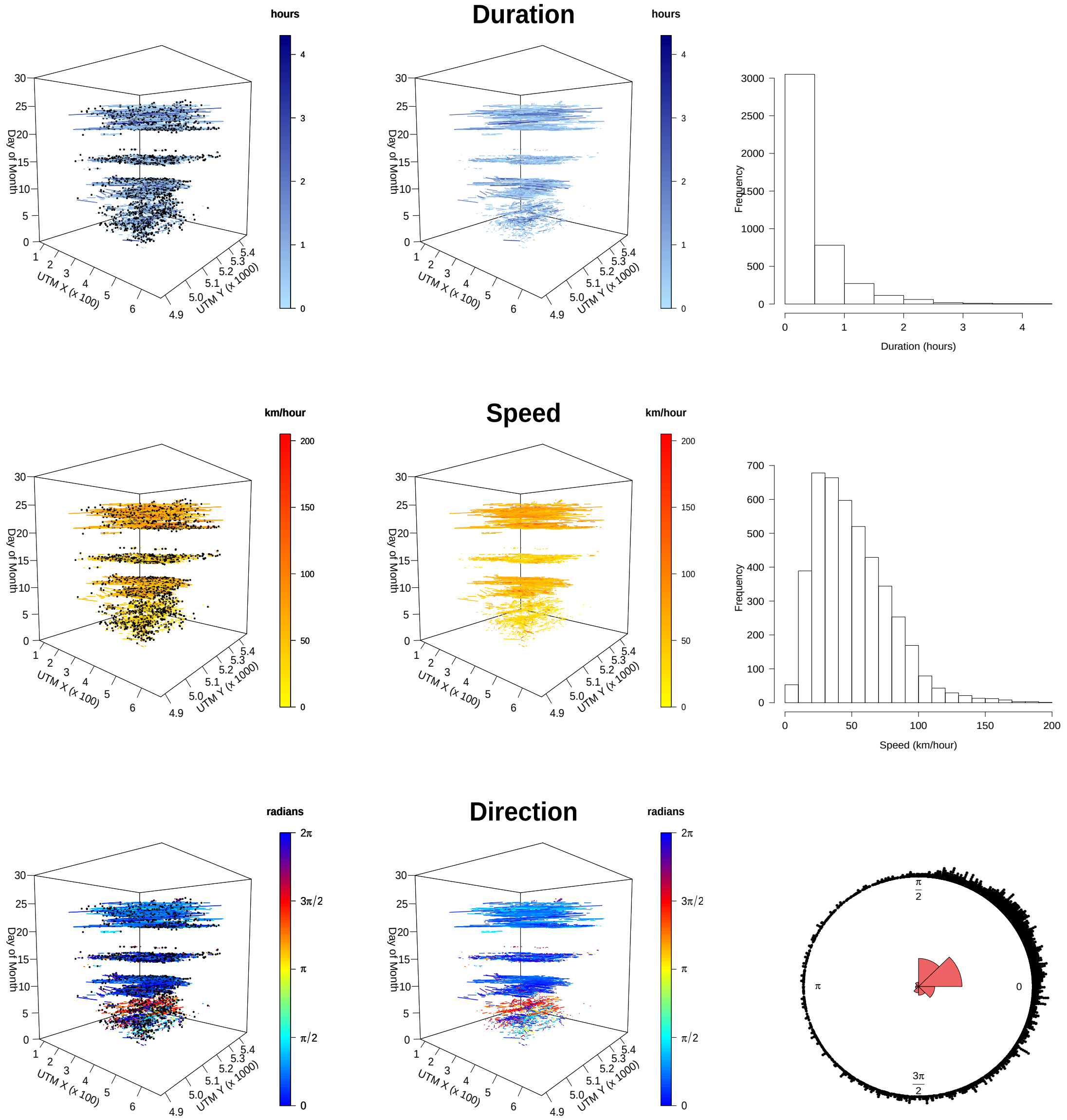
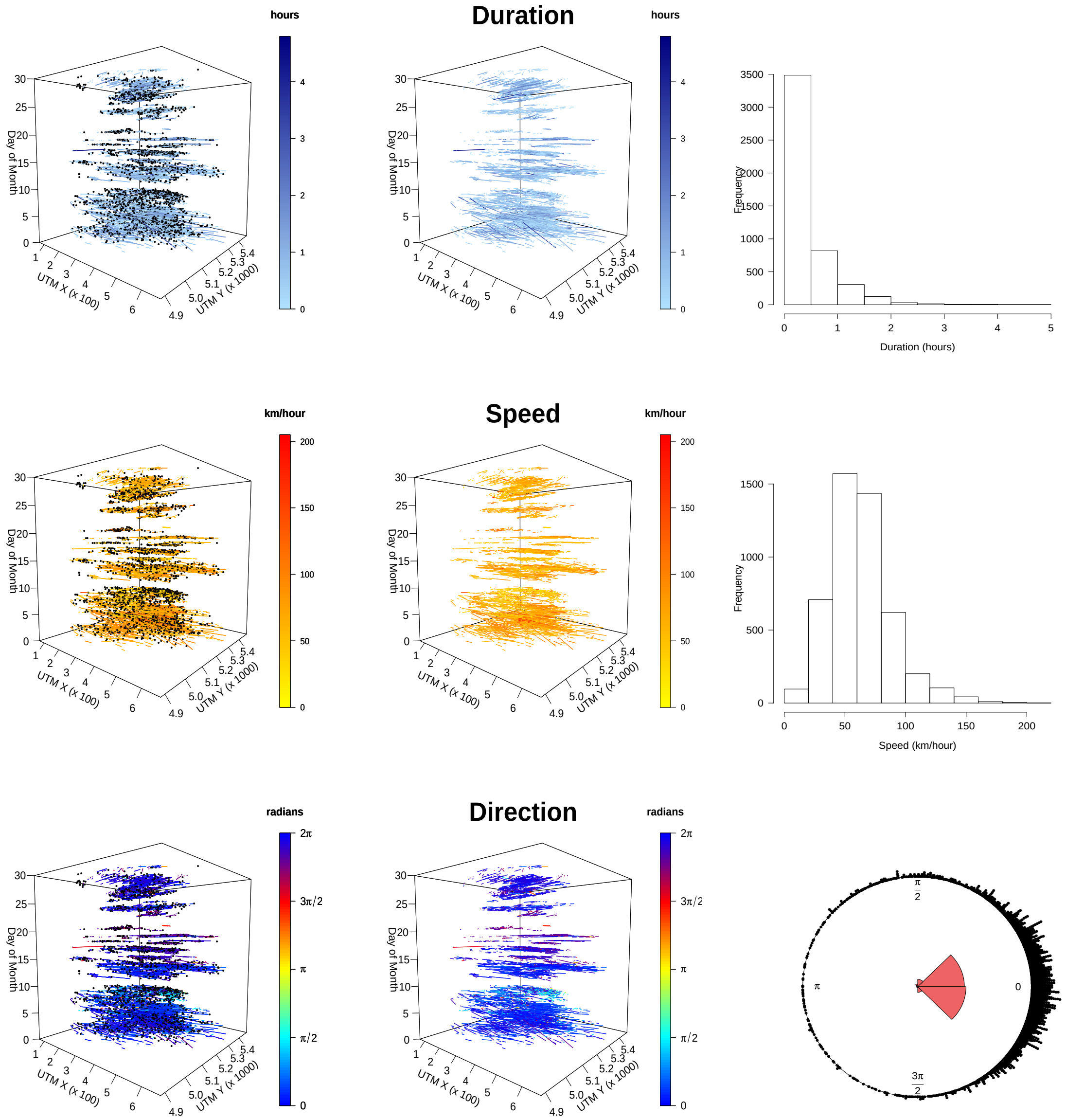


Figure 4.3: Three dimensional plots and histograms of storm cell duration (top row), speed (middle row) and direction (bottom row) from June 2003. In the left panel, black points represent storm cells only observed on one occasion.

(Lee and Nelder, 2006) may be utilized. Double hierarchical generalized linear models are extensions of generalized linear mixed models which allow both the mean and dispersion to be modelled as functions of random effects. Generalized additive models for location, scale and shape parameters, in the same spirit, allow not only the location and dispersion to be functions of covariates and possibly random effects, but depending on the assumed


Duration

Speed

Direction

Figure 4.4: Three dimensional plots and histograms of storm cell duration (top row), speed (middle row) and direction (bottom row) from July 2003. In the left panel, black points represent storm cells only observed on one occasion.

distribution, also the shape parameter(s). Additionally, splines may be incorporated into any of these terms, fit with a penalized likelihood. Therefore, this flexible class of models contains double hierarchical generalized linear models as well as generalized additive models, introduced in Section 2.4.3.

For this project, not only are we interested in modelling storm cell speed and direction,

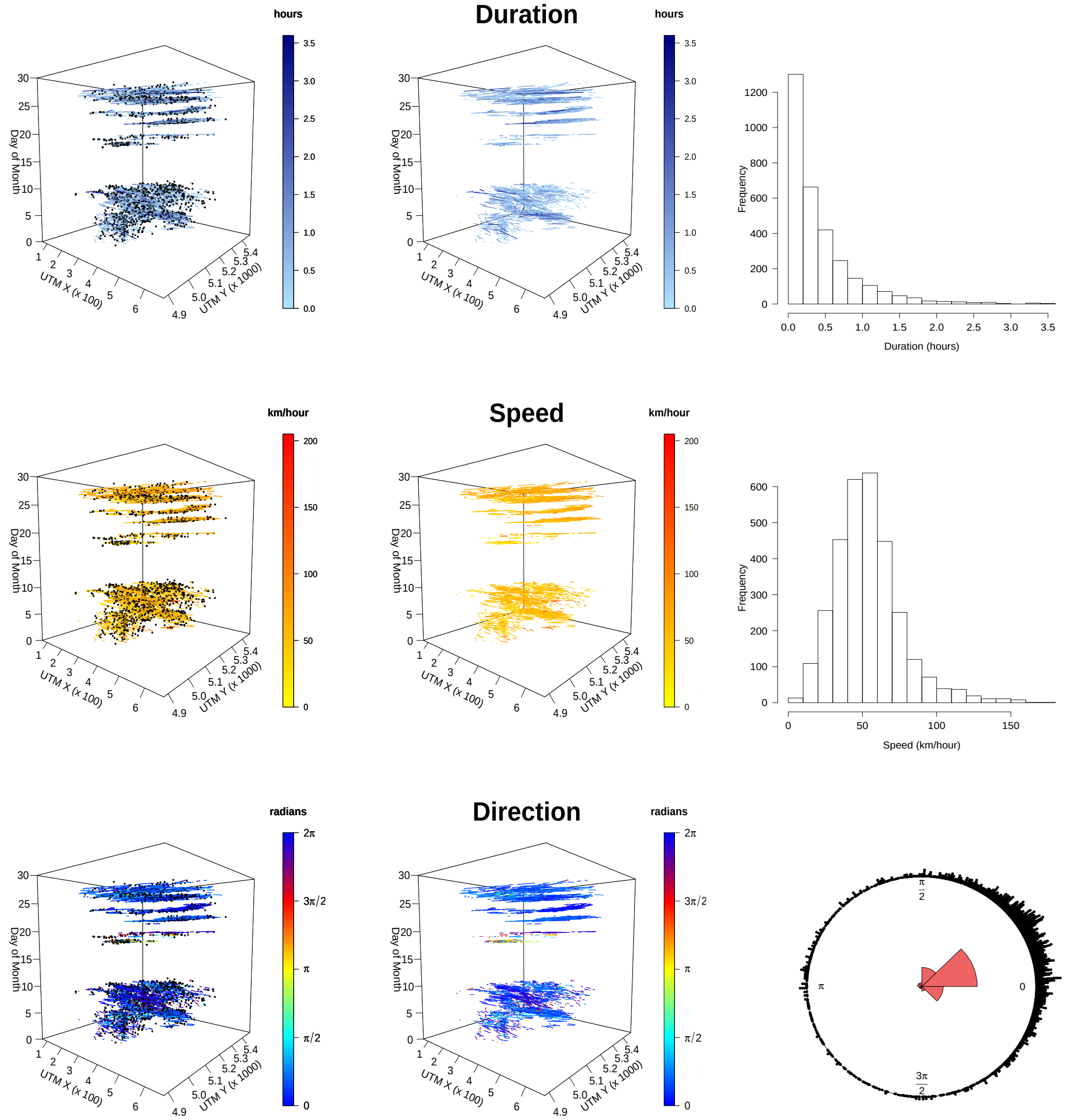


Figure 4.5: Three dimensional plots and histograms of storm cell duration (top row), speed (middle row) and direction (bottom row) from August 2003. In the left panel, black points represent storm cells only observed on one occasion.

but each observation also has a random duration that is of interest. This is further complicated by the fact that not all storm cells have a recorded duration. Finally, not only are we interested in modelling these marks, but we wish to jointly model the point pattern of storm cell detection along with storm cell trajectory. In the remainder of this chapter we extend the aforementioned ecological regression approach for modelling storm cell trajectories in two ways. First, to account for storm cells without a recorded trajectory we utilize the hurdle framework described in Section 2.4.1. Second, to incorporate random storm cell duration we also model this outcome in a similar manner as we do with speed. That is, we model it conditional on a function of direction, following a log-normal distribution. Direction is assumed to follow the von Mises distribution (Fisher and Lee, 1992). Also referred to as circular normal, the von Mises distribution is parameterized in terms of a location or mean, μ, and a concentration parameter, ϑ. The case where $\vartheta = 0$ corresponds to the uniform distribution and as ϑ increases, the resulting distribution becomes concentrated about the angle μ.

4.3 Spatio-Temporal Marked Point Process Model

This section describes our model for multivariate marks as well as the connection between the point and mark processes before providing details on parameter estimation.

4.3.1 Marked Point Process

Storm cell trajectory is characterized by duration, speed and direction. This section outlines a four component model for these quantities that distinguishes between the mechanisms that determine whether or not a storm cell is observed more than once and storm cell duration, speed and direction. To do this, we utilize a hurdle model that has the flexibility needed to account for storm cells without a complete trajectory and provides a simple approach to link the point and mark processes.

For the ith storm cell, let $Z(s_i, t_i)$ denote the Gaussian process described in Section 3.2.1 with intercept β^*, mean $E[Z(s_i, t_i)] = -0.5(\sigma_s^2 + \sigma_t^2)$ and covariance $\Sigma = \text{Cov}[Z(s_i, t_i), Z(s_j, t_j)] = \sigma_s^2 \text{exp}(-||s_j - s_i||/\phi_s) + \sigma_t^2 \text{exp}(-|t_j - t_i|/\phi_t)$. Suppose

$$Y^*(s_i, t_i) = \beta^* + \boldsymbol{G}_i \boldsymbol{\gamma} + Z(s_i, t_i) \tag{4.1}$$

for $i = 1, 2, \ldots, n$. Therefore, $Y^*(s_i, t_i)$ is a Gaussian process with the same covariance structure as $Z(s_i, t_i)$ and a mean shifted by $\boldsymbol{G}_i \boldsymbol{\gamma}$ where $\boldsymbol{G}_i$ represents the covariate vector for the ith storm cell and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)^T$ are the corresponding parameters. This implies that storm cells detected near larger values of the Gaussian parent process are more likely to last longer. Let Y_{i1} denote the indicator for whether the ith storm cell is observed on more than one occasion and τ be a threshold parameter. Then,

$$Y_{i1} = \begin{cases} 0, & \text{if } Y^*(s_i, t_i) \leq \tau \\ 1, & \text{if } Y^*(s_i, t_i) > \tau \end{cases}$$

and $\pi_{i1} = P[Y^*(s_i, t_i) > \tau] = E[Y_{i1}]$ represents the probability that the ith storm cell at location (s_i, t_i) is observed at least twice by the radar. Therefore, if $Y_{i1} = 1$, we are interested in the distribution of its duration as well as its speed and direction represented by Y_{i2}, Y_{i3} and Y_{i4}, respectively. That is, we assume the Gaussian process that generates storms and the resulting storm cells is also related to how long a storm cell will last. Given that a storm cell has a recorded trajectory, its corresponding speed and direction will be related to wind. To provide sufficient flexibility required for capturing the potentially highly variable spatio-temporal outcomes we employ generalized additive models with the location and scale parameters being smooth functions of space and time. For our purposes, this accounts for the spatio-temporal correlation present in these data and also for the potentially heavy-tailed distributions.

For $j = 2, 3$, we assume $Y_{ij} \sim \text{log-normal}(\mu_{ij}, \vartheta_{ij}^2)$ where

$$\mu_{ij} = \boldsymbol{B}_{ij1}\boldsymbol{\beta}_{j1} + \varphi_{j1}(s_{ji}, t_{ji})$$

and

$$\log(\vartheta_{ij}) = \boldsymbol{B}_{ij2}\boldsymbol{\beta}_{j2} + \varphi_{j2}(s_{ji}, t_{ji})$$

with $\boldsymbol{B}_{ij1}$ representing the covariate vector for the ith observation and jth outcome that is linearly related to μ_{ij}, $\boldsymbol{\beta}_{j1} = (\beta_{j10}, \beta_{j11}, \ldots, \beta_{j1p_j})^T$ denoting the corresponding coefficient vector and $\varphi_{j1}(s_{ji}, t_{ji})$ being the space-time smoother as in Equation (2.5). The terms $\boldsymbol{B}_{ij2}$, $\boldsymbol{\beta}_{j2}$ and $\varphi_{j2}(s_{ji}, t_{ji})$, in relation to $\log(\vartheta_{ij})$, are defined analogously to that of μ_{ij}. To model direction we assume $Y_{i4} \sim$ von Mises$(\mu_{i4}, \vartheta_{i4})$ where

$$\tan(\mu_{i4}/2) = \boldsymbol{B}_{i41}\boldsymbol{\beta}_{41} + \varphi_{41}(s_{ji}, t_{ji})$$

and

$$\log(\vartheta_i) = \boldsymbol{B}_{i42}\boldsymbol{\beta}_{42} + \varphi_{42}(s_{ji}, t_{ji})$$

with $\boldsymbol{B}_{i41}$, $\boldsymbol{B}_{i42}$, $\boldsymbol{\beta}_{41}$, $\boldsymbol{\beta}_{42}$, $\varphi_{41}(s_{ji}, t_{ji})$ and $\varphi_{42}(s_{ji}, t_{ji})$ being analogous to the models for $j = 2$ and 3. The likelihood for this model, conditional on the covariates, is proportional to

$$\begin{aligned} \mathcal{L}(\cdot \mid \boldsymbol{Y}) &= \prod_{i=1}^{n} [1 - \Phi(\pi_{i1})]^{I(Y_{i1}=0)} \left[\frac{\Phi(\pi_{i1})}{Y_{i2}\vartheta_{i2}} \exp\left\{-\frac{[\log(Y_{i2}) - \mu_{i2}]^2}{2\vartheta_{i2}^2}\right\}\right]^{I(Y_{i1}=1)} \times \\ & \left[\frac{1}{Y_{i3}\vartheta_{i3}} \exp\left\{-\frac{[\log(Y_{i3}) - \mu_{i3}]^2}{2\vartheta_{i3}^2}\right\} \frac{1}{I_0(\vartheta_{i4})} \exp\left\{\vartheta_{i4} \cos(Y_{i4} - \mu_{i4})\right\}\right]^{I(Y_{i1}=1)} \end{aligned} \quad (4.2)$$

where Φ is the standard normal cumulative distribution function and $I_0(\vartheta)$ is the zeroth order modified Bessel function of the first kind, expressed as

$$I_0(\vartheta) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{\vartheta \cos(y)\right\} \mathrm{d}y.$$

In this chapter, the form of the spatio-temporal smoother employed is a tensor product of a bivariate thin plate regression spline (Wood, 2003), which is an isotropic spatial smoother, and a univariate thin plate spline for the temporal dimension, as described in Section 2.4.3. This type of smoother is mathematically convenient for spatio-temporal data as $\varphi(s_i) + \varphi(t_i)$ is strictly nested within $\varphi(s_i, t_i)$ so model comparisons with lower dimensional smoothers may be easily performed. Tensor product splines are also scale invariant and enable different levels of smoothing across each dimension.

4.3.2 Parameter Estimation and Inference

Simultaneously estimating the point process parameters as well as those from the evolving mark model is a challenging task. Therefore, to facilitate parameter estimation, we model the mark process conditional on the point process. This permits the use of point-referenced techniques, as described above, for modelling the marks. Moreover, because of the form of the likelihood in Equation (4.2), it is fully efficient to estimate the parameters from all model components separately.

In the probit component, we assume β^*, σ_t^2, σ_s^2, ϕ_t and ϕ_s are known and fixed at the estimates from the hierarchical cluster process (see Table 3.1). Therefore, to fully specify it suffices to estimate $\boldsymbol{\gamma}$ and τ, which can be done using the Markov chain EM algorithm as proposed by Chib and Greenberg (1998). Note that γ_0 and τ cannot be simultaneously estimated because of identifiability issues and therefore, we set $\gamma_0 = 0$. If $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \tau)$, at the rth iteration, we can estimate the conditional expectation of the complete data log-likelihood as:

$$\begin{aligned} & \frac{1}{M^*}\sum_{j=1}^{M^*} \log\left\{f\left(\boldsymbol{Y}^{*(j)} \mid \boldsymbol{\theta}\right)\right\} \\ = \; & -\frac{1}{2}|\Sigma| - \frac{1}{M^*}\sum_{j=1}^{M^*}\left(\boldsymbol{Y}^{*(j)} - \beta^* - \boldsymbol{G}\boldsymbol{\gamma} + \tau\right)^T \Sigma^{-1}\left(\boldsymbol{Y}^{*(j)} - \beta^* - \boldsymbol{G}\boldsymbol{\gamma} + \tau\right) \end{aligned} \tag{4.3}$$

where $\boldsymbol{Y}^{*(j)} = \left(Y_1^{*(j)}, Y_2^{*(j)}, \ldots, Y_n^{*(j)}\right)^T$ are draws from a multivariate normal distribution truncated to be in the range $(-\infty, 0]$ if $Y_{i1} = 0$ and $(0, \infty)$ otherwise. Updates of $\boldsymbol{\theta}^{(r)}$ may be calculated analytically by maximizing Equation (4.3) where

$$\boldsymbol{\theta}^{(r+1)} = \left(\boldsymbol{G}^T\Sigma^{-1}\boldsymbol{G}\right)^{-1}\left(\boldsymbol{G}^T\Sigma^{-1}\bar{\boldsymbol{Y}}^*\right)^{-1}$$

and $\bar{\boldsymbol{Y}}^* = \frac{1}{M^*}\sum_{j=1}^{M^*}\boldsymbol{Y}^{*(j)}$ is the average over M^* draws from $\boldsymbol{Y}^*$. Standard errors of the estimated parameters are calculated using the observed information matrix

$$-E\left\{\frac{\partial^2 \log\left[f(\boldsymbol{Y}^* \mid \boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\} - \mathrm{Var}\left\{\frac{\partial \log f(\boldsymbol{Y}^* \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\}.$$

Parameter estimation for the duration ($j = 2$), speed ($j = 3$) and direction ($j = 4$) model components is performed by maximizing the penalized likelihood,

$$\ell_{\text{pen}}(\boldsymbol{\theta}_j \mid \boldsymbol{y}_j) = \ell(\boldsymbol{\theta}_j \mid \boldsymbol{y}_j) - \frac{1}{2}\sum_{k=1}^{p_j}\sum_{\ell=1}^{L_{jk}} \zeta_{jk\ell} \int \boldsymbol{\psi}_{jk\ell}^T \boldsymbol{\Omega}_{jk\ell}(x_{jk\ell})\boldsymbol{\psi}_{jk\ell} \mathrm{d}x_{jk\ell}$$

where $p_j = 1$ corresponds to the location parameter for the jth component and $p_j = 2$, the scale. This is accomplished via the backfitting algorithm outlined in Appendix B of Rigby and Stasinopoulos (2005). Briefly, for the location parameter, at the rth iteration, we evaluate the partial residuals for the linear covariates $\left(\boldsymbol{\beta}_{j1}^{(r)}\right)$. These are regressed against the design matrix $\boldsymbol{B}_{j1}$ to obtain $\left(\boldsymbol{\beta}_{j1}^{(r+1)}\right)$. For the non-linear terms, the partial residuals are calculated and smoothed in order to update the parameters from the function $\varphi_{j1}(\cdot)$. This is repeated to convergence for the location parameter and then performed in a similar manner for the scale. In `R`, the `gamlss` package may be employed to maximize the penalized likelihood. For non-standard distributions, such as the von Mises, this package allows users to specify their own distribution by simply providing the appropriate link, likelihood and score functions.

Model selection for the duration, speed and direction components is performed by likelihood ratio tests for nested models and a comparison of model Akaike information criterion, otherwise. The likelihood ratio is calculated as $-2\ell_{\text{pen}}(\boldsymbol{\theta}_{j0}) + 2\ell_{\text{pen}}(\boldsymbol{\theta}_{j1})$, where the subscript 0 indicates the null model. This statistic is asymptotically χ^2 with the degrees of freedom being the difference between the error degrees of freedom in the two models. The Akaike information criterion is calculated as the fitted global deviance plus a penalty of 2 times the effective degrees of freedom. Goodness of fit is assessed by an analysis of the normalized quantile residuals (Dunn and Smyth, 1996). Specifically, for the random variable Y_{ij}, letting $u_{ij} = F_j(Y_{ij} \mid \hat{\boldsymbol{\theta}}_{ij})$ where $F_j(\cdot)$ is the cumulative distribution function of the jth outcome and $\hat{\boldsymbol{\theta}}_{ij}$ are the corresponding parameter estimates, the normalized quantile residuals may be calculated as $\hat{r}_{ij} = \Phi^{-1}(u_{ij})$. If the model is appropriate, r_{ij}, will follow a standard normal distribution.

4.4 Application to Storm Cell Trajectories

This section presents the results for the storm cell trajectory (mark) models from the Bismarck, North Dakota radar station.

4.4.1 Results

As in Chapter 3, these were fit monthly to account for possible seasonal effects. The duration and speed components included the covariates cos(direction) and vertical integrated liquid, a measure of storm cell intensity, assumed to be linearly related to the mean on the link scale. Incorporating direction as a covariate allows us to make inference on its relationship with speed and duration; the cos function is utilized to account for direction being a circular covariate, as suggested by Modlin et al. (2012). Furthermore, since vertical integrated liquid is employed in the Storm Cell Identification and Tracking algorithm, it may explain some of the observed trends in the speed and duration random variables and, again, enables us to explore the relationship between storm cell intensity and storm cell speed and duration. As described above, non-linear functions of covariates accounting for spatio-temporal effects were included in models for the location and, as appropriate, scale parameters. In this section, we focus on the main modelling results for all months. Additional figures displaying the partial effects for the splines included in the location and scale terms are provided in Appendix B. Note that because the covariance structure in the probit component is fixed at the values from Chapter 3, the parameter estimates from the spatio-temporal Gaussian process are the same as presented previously; the only parameter estimated from within this component was the threshold.

Table 4.1 summarizes the results for the four component mark model fit to April storm cells. Here, both the duration and direction components include a spatio-temporal smoother in the mean accounting for an interaction effect between space and time. In the speed submodel, an additive space-time smoother suffices. Additional spatial structure is incorporated in the scale term for duration and speed. In the duration component, direction is a significant covariate with the coefficient estimated as 0.094 (0.006, 0.182), the numbers in brackets representing 95% confidence intervals, indicating that storms cells moving east have longer lifetimes. At a direction of 2π radians, storm cell duration is estimated to be 1.099 hours longer than those moving in a direction of $\frac{3\pi}{2}$ or $\frac{\pi}{2}$ radians with all other elements of the model remaining constant. For storm cell speed, neither direction nor vertical integrated liquid were significant. Figure 4.6 displays the three-dimensional plots of the fitted means for the duration, speed and direction components as well as histograms for these quantities. These subfigures display similar spatio-temporal trends as were observed in Figure 4.1. However, notice that the ranges of the histograms for the duration and speed components are much smaller than that of corresponding subfigures in Figure 4.1; this motivates the use of additional smoothers in the scale parameters to account for the heavy-tailed distribution. Histograms, variograms and autocorrelation function plots of the normalized quantile residuals for these model components are displayed in Figure 4.7 indicating no lack of fit is detected, as the distribution of the normalized quantile residuals appears to be normally distributed, and that we are accounting for the spatial and temporal autocorrelation.

To model May storm cell trajectories, as shown in Table 4.2, spatio-temporal splines are included in the location terms for duration, speed and direction. The latter two components also require spatio-temporal smoothers in the scale parameters to adequately model these quantities. In the duration submodel, the longest lasting observations correspond to more intense cells moving towards the east with the estimated coefficients for cos(direction) and vertical integrated liquid being 0.190 (0.140, 0.240) and 0.030 (0.023, 0.037), respectively. There is also a significant association between a storm cell's speed and its intensity and direction; lower values of vertical integrated liquid are correlated with faster moving storm cells and the corresponding estimated coefficient is -0.004 (-0.007, -0.002). However, storm cells with larger values of cos(direction) are associated with faster speeds as the estimated parameter is 0.090 (0.066, 0.114). Figure 4.8 displays the fitted mean values for duration, speed and direction for May, which show similar spatio-temporal trends as Figure 4.2. Notice that storm cells within a cluster have a tendency to travel at the same speed and in the same direction. However, as before, we are underestimating large values of speed and the direction model is not fitting well to storm cells moving south and west; this again motivates the inclusion of the spatio-temporal splines in the scale terms. As with the April results, the normalized quantile residuals for the three components of storm cell trajectory are displayed in Figure 4.9. Although there does appear to be some unexplained temporal correlation in the speed component, there do not seem to be any severe problems with model fit.

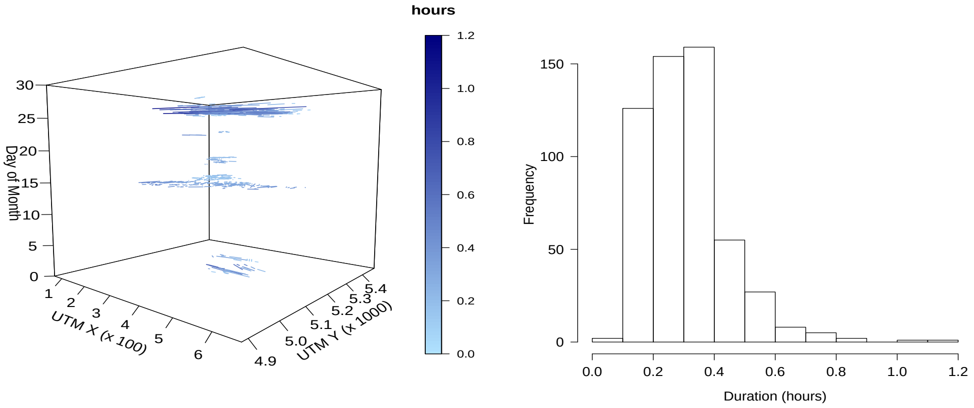
The results, by model component, for June storm cell trajectories can be seen in Table 4.3. Similar to the May results, spatio-temporal smoothers are included in mean duration as well as both the location and scale parameters for speed and direction. Storm cells with the largest measure of vertical integrated liquid and moving east have the longest lifetimes while less intense storm cells are slower moving. The estimated coefficients for direction and vertical integrated liquid in the duration component are 0.296 (0.244, 0.348) and 0.012 (0.009, 0.015), respectively, and in the speed submodel, the parameter estimate for intensity is -0.0012 (-0.0016, -0.0008). Figures 4.10 and 4.11 display the fitted values and goodness of fit diagnostics for all model components. In the direction submodel, the mean correctly identifies most storm cells as moving between 0 and $\pi/2$ radians. Meanwhile, there is not a large variability in duration with the mean value for all storm cells being less than one hour. In the fitted speed component, storm cells within a storm are shown to travel at similar speeds; this distribution appears to be bimodal, perhaps suggesting that for this model there are not enough events to capture average storm cell behaviour. However, goodness of fit diagnostics do not indicate any severe lack of fit.

A summary of July storm cell trajectories can be seen in Table 4.4. For these components, the spatio-temporal smoothers employed are the same as in May and June. As before, more intense storm cells with larger values of cos(direction) are associated with longer lifetimes, as the parameter estimates for these covariates are 0.014 (0.011, 0.016) and 0.483 (0.417, 0.549), respectively. Storm cells moving east are also the fastest with the estimated coefficient being 0.159 (0.132, 0.186). Figure 4.10 displays the mean fitted duration, speed and direction for all July storm cells. The spatio-temporal trends observed in Figure 4.3 are similar to those displayed in Figure 4.10. However, again, the ranges of the fitted means are smaller than in the observed data, prompting our use of splines in the scale parameters. Goodness of fit diagnostics, as provided in Figure 4.11, do not indicate problems with model fit.

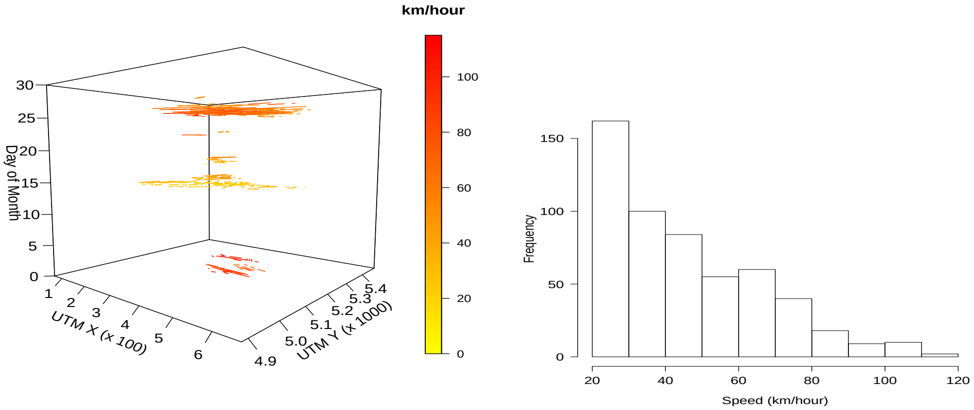
Finally, Table 4.5 and Figure 4.14 summarize the model fitting results for August storm cell trajectories. As with July, storm cells moving towards the east have longer lifetimes and faster speeds with the estimated coefficients in the duration and speed components being 0.446 (0.389, 0.503) and 0.094 (0.072, 0.116), respectively. More intense storm cells, as measured by higher values of vertical integrated liquid, are longer lasting with the corresponding coefficient estimated as 0.010 (0.007, 0.013). Similar spatio-temporal trends as shown in Figure 4.5 can be seen in the fitted means displayed in Figure 4.14. Furthermore, Figure 4.15 does not indicate any problems with model fit.

		Linear Effects		Non-linear Effects		
		Est.	CI		EDF	p-value
Probit						
	β^*	-13.89	(-14.81, -13.30)			
	σ_t^2	2.25	(0.57, 5.31)			
	σ_s^2	2.42	(1.91, 2.93)			
	ϕ_t	0.68	(0.25, 0.75)			
	ϕ_s	112.27	(99.73, 149.70)			
	τ	-14.92	(-15.84, -13.99)			
Duration						
μ_2	β_{210}	-1.298	(-1.375, -1.220)	$\varphi(s,t)$	38.02	<0.001
	$\cos(Y_4)$	0.094	(0.006, 0.182)			
	VIL	0.007	(-0.008, 0.021)			
ϑ_2	β_{220}	-0.257	(-0.317, -0.197)	$\varphi(s)$	23.10	<0.001
Speed						
μ_3	β_{310}	3.757	(3.711, 3.804)	$\varphi(s)$	16.03	<0.001
	$\cos(Y_4)$	-0.021	(-0.072, 0.030)	$\varphi(t)$	8.81	<0.001
	VIL	-0.005	(-0.015, 0.005)			
ϑ_3	β_{320}	-0.866	(-0.926, -0.807)	$\varphi(s)$	6.05	0.037
Direction						
μ_4	β_{410}	0.382	(0.342, 0.422)	$\varphi(s,t)$	210.31	<0.001
ϑ_4	β_{420}	1.115	(1.012, 1.219)			

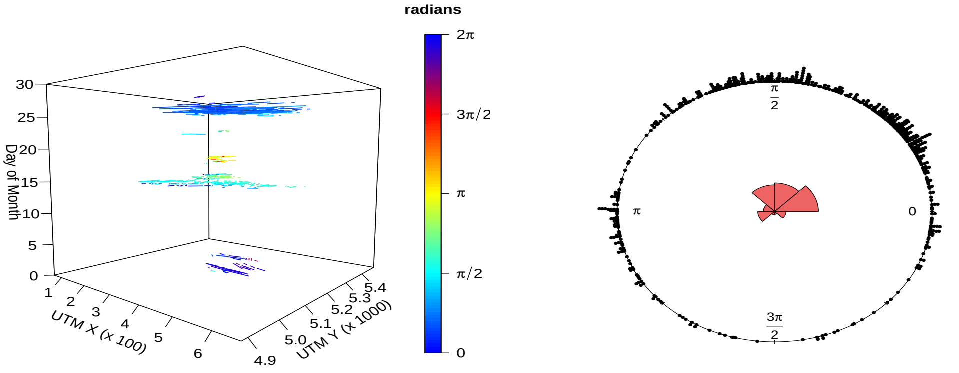
Table 4.1: Summary from the four component model for storm cell trajectory in April 2003. For the linear effects, parameter estimates (Est.) and 95% confidence intervals (CIs) of vertical integrated liquid (VIL) and cos(direction) are provided, and effective degrees of freedom (EDF) and p-values are given for the non-linear spatio-temporal effects.



(a) Fitted mean duration. (b) Fitted mean duration.



(c) Fitted mean speed. (d) Fitted mean speed.



(e) Fitted mean direction. (f) Fitted mean direction.

Figure 4.6: Three dimensional plots and histograms of fitted storm cell mean duration (top row), speed (middle row) and direction (bottom row) from April 2003.

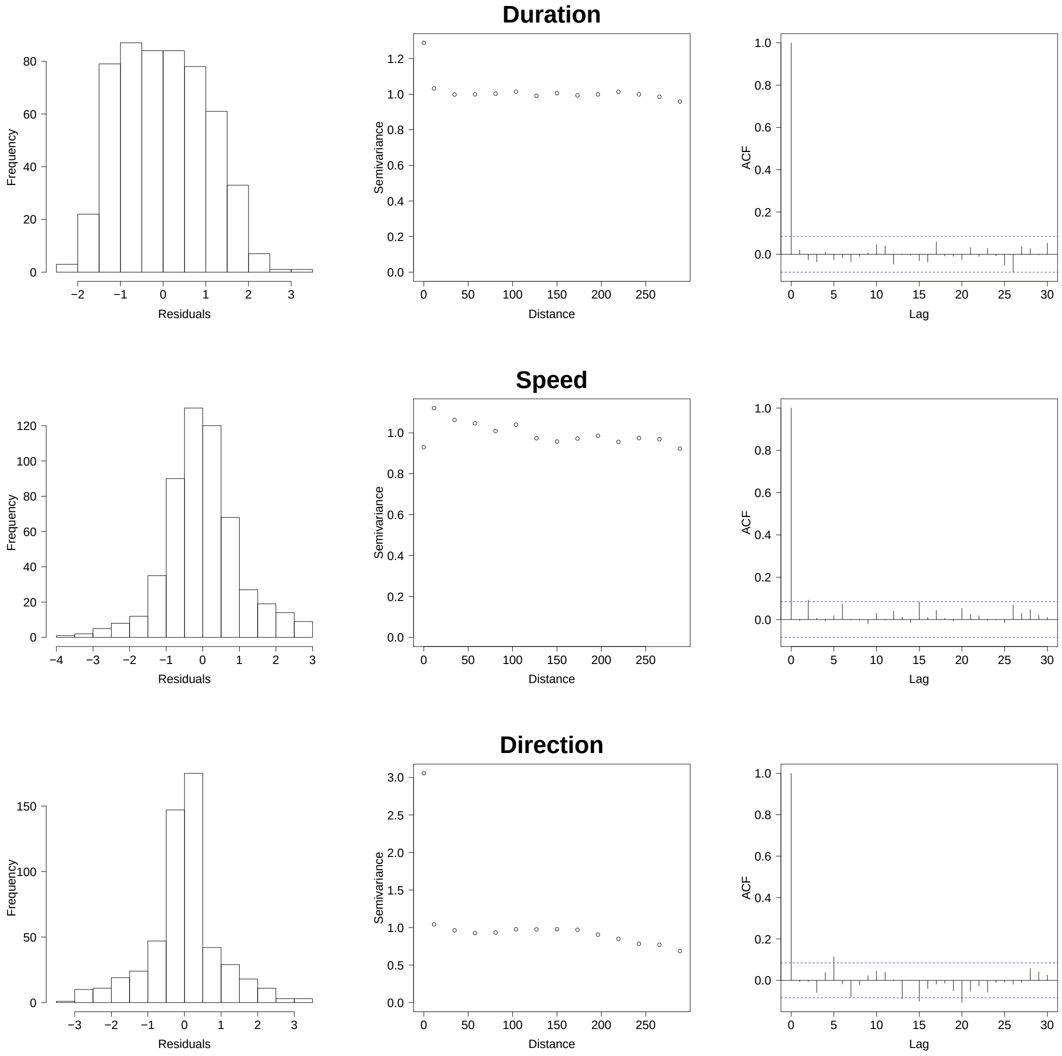
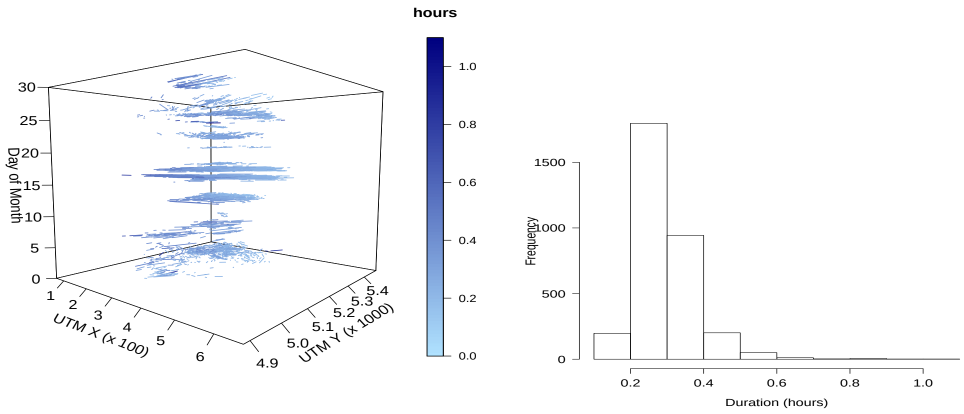


Figure 4.7: Histograms (left), variograms (middle) and autocorrelation function plots (right) of the normalized quantile residuals for the duration (top), speed (middle) and direction (bottom) components of the April 2003 mark models.

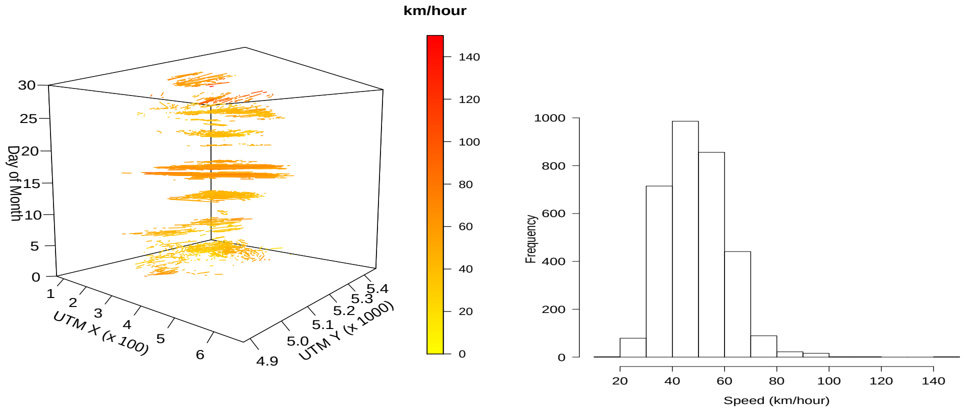
		Linear Effects		Non-linear Effects		
		Est.	CI		EDF	p-value
Probit						
	β^*	-14.03	(-14.07, -7.91)			
	σ_t^2	2.08	(0.93, 7.53)			
	σ_s^2	3.51	(2.75, 4.57)			
	ϕ_t	1.45	(1.13, 4.86)			
	ϕ_s	87.06	(69.82, 112.05)			
	τ	-16.56	(-18.06, -15.04)			
Duration						
μ_2	β_{210}	-1.446	(-1.489, -1.404)	$\varphi(s,t)$	29.84	<0.001
	$\cos(Y_4)$	0.190	(0.140, 0.240)			
	VIL	0.030	(0.023, 0.037)			
ϑ_2	β_{220}	-0.148	(-0.172, -0.124)			
Speed						
μ_3	β_{310}	3.838	(3.817, 3.859)	$\varphi(s,t)$	113.60	<0.001
	$\cos(Y_4)$	0.090	(0.066, 0.114)			
	VIL	-0.004	(-0.007, -0.002)			
ϑ_3	β_{320}	-0.932	(-0.957, -0.908)	$\varphi(s,t)$	47.13	<0.001
Direction						
μ_4	β_{410}	0.281	(0.271, 0.292)	$\varphi(s,t)$	211.35	<0.001
ϑ_4	β_{420}	1.020	(0.974, 1.067)	$\varphi(s,t)$	37.22	<0.001

Table 4.2: Summary from the four component model for storm cell trajectory in May 2003. For the linear effects, parameter estimates (Est.) and 95% confidence intervals (CIs) of vertical integrated liquid (VIL) and cos(direction) are provided, and effective degrees of freedom (EDF) and p-values are given for the non-linear spatio-temporal effects.



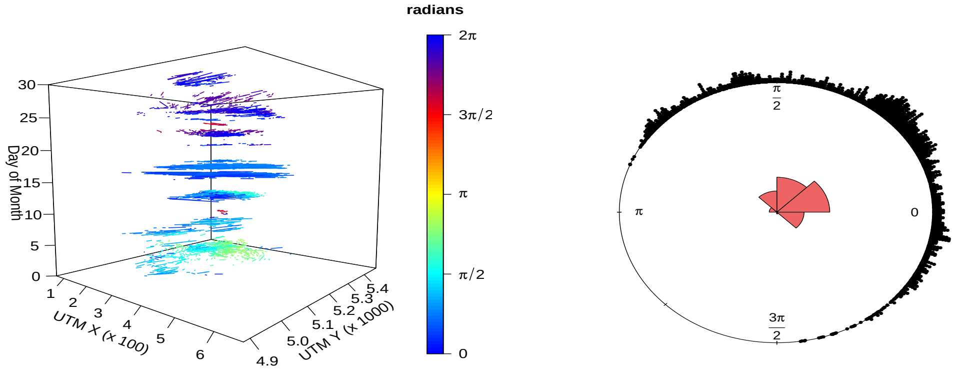
(a) Fitted mean duration.

(b) Fitted mean duration.



(c) Fitted mean speed.

(d) Fitted mean speed.



(e) Fitted mean direction.

(f) Fitted mean direction.

Figure 4.8: Three dimensional plots and histograms of fitted storm cell mean duration (top row), speed (middle row) and direction (bottom row) from May 2003.

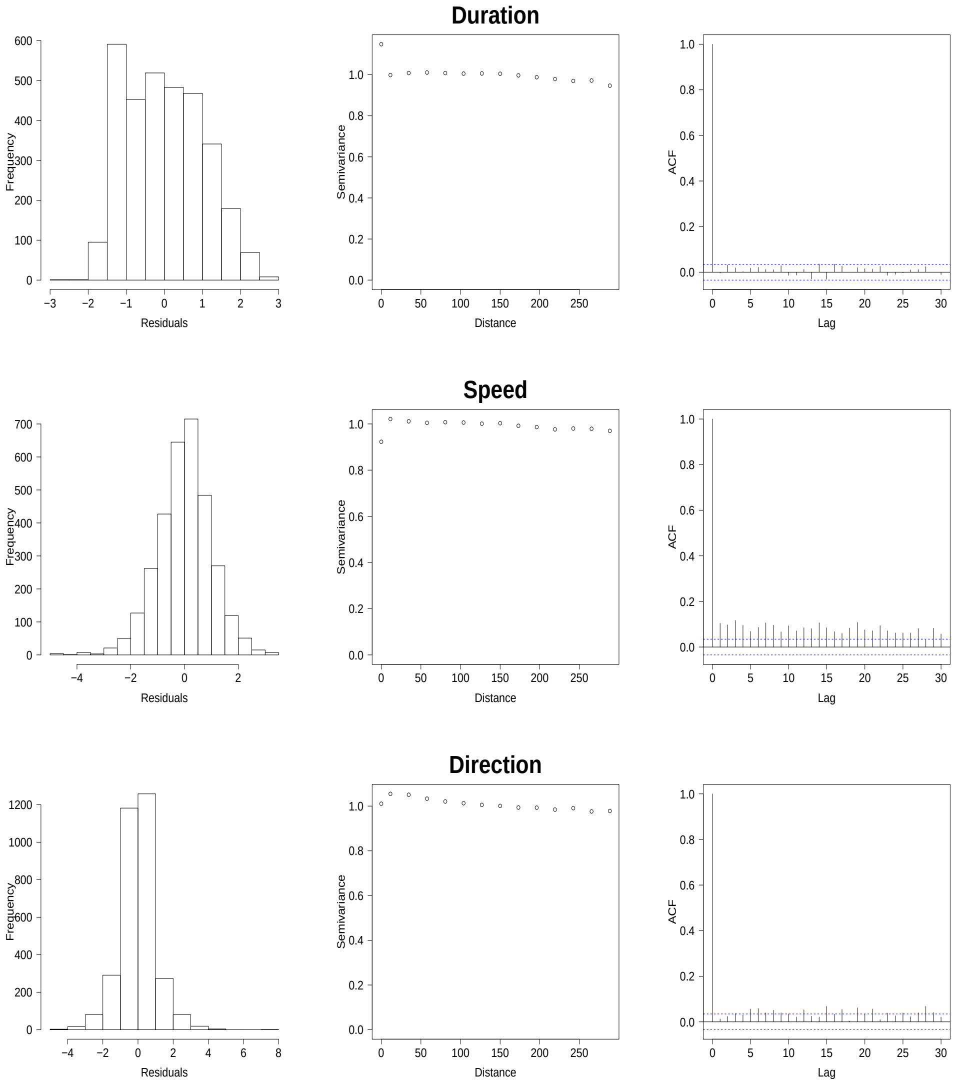
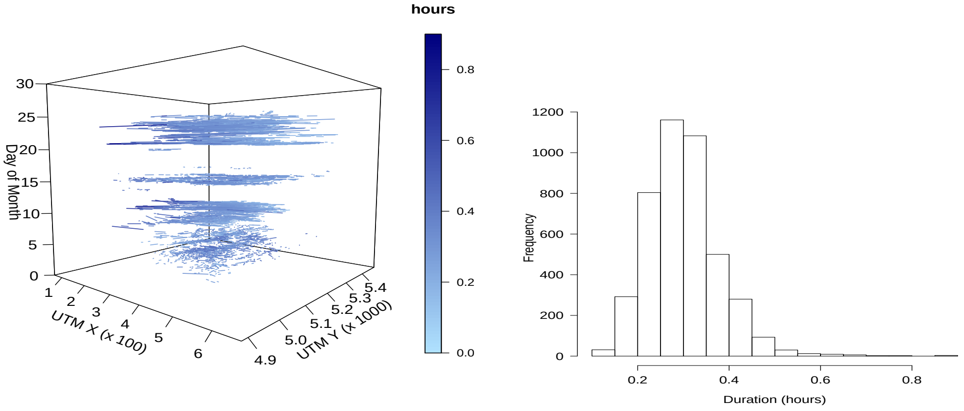


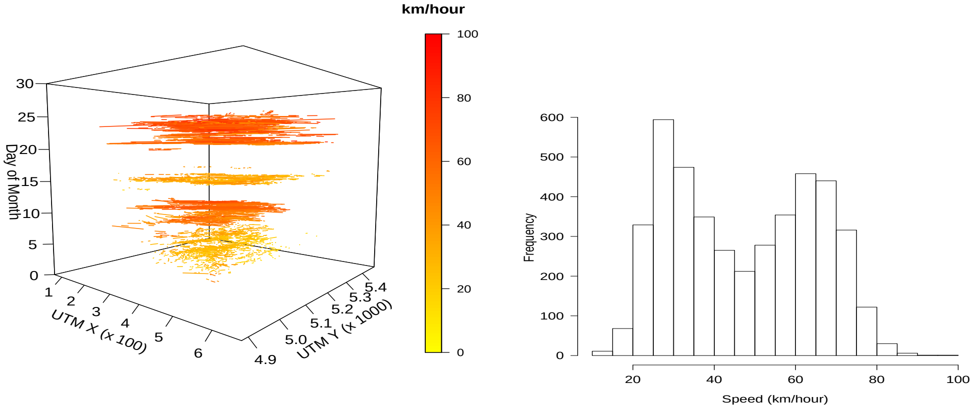
Figure 4.9: Histograms (left), variograms (middle) and autocorrelation function plots (right) of the normalized quantile residuals for the duration (top), speed (middle) and direction (bottom) components of the May 2003 mark models.

		Linear Effects		Non-linear Effects		
		Est.	CI		EDF	p-value
Probit						
	β^*	-13.73	(-14.37, -13.68)			
	σ_t^2	0.99	(0.28, 1.41)			
	σ_s^2	2.63	(1.85, 3.53)			
	ϕ_t	3.59	(2.85, 8.77)			
	ϕ_s	101.48	(84.80, 128.02)			
	τ	-15.47	(-16.86, -14.08)			
Duration						
μ_2	β_{210}	-1.459	(-1.500, -1.418)	$\varphi(s,t)$	59.40	<0.001
	$\cos(Y_4)$	0.296	(0.244, 0.348)			
	VIL	0.012	(0.009, 0.015)			
ϑ_2	β_{220}	-0.120	(-0.141, -0.099)			
Speed						
μ_3	β_{310}	3.790	(3.773, 3.806)	$\varphi(s,t)$	226.94	<0.001
	$\cos(Y_4)$	0.010	(-0.011, 0.031)			
	VIL	-0.0012	(-0.0016, -0.0008)			
ϑ_3	β_{320}	-0.996	(-1.017, -0.975)	$\varphi(s,t)$	175.81	<0.001
Direction						
μ_4	β_{410}	0.151	(0.142, 0.160)	$\varphi(s,t)$	381.83	<0.001
ϑ_4	β_{420}	1.221	(1.183, 1.259)	$\varphi(s,t)$	48.22	<0.001

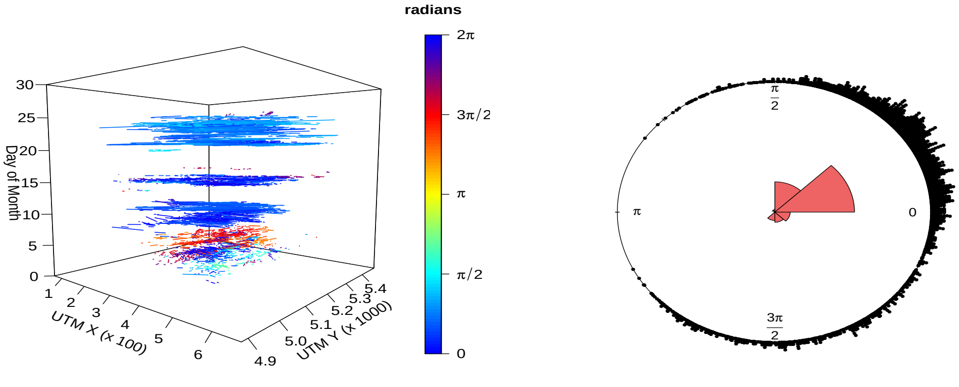
Table 4.3: Summary from the four component model for storm cell trajectory in June 2003. For the linear effects, parameter estimates (Est.) and 95% confidence intervals (CIs) of vertical integrated liquid (VIL) and cos(direction) are provided, and effective degrees of freedom (EDF) and p-values are given for the non-linear spatio-temporal effects.



(a) Fitted mean duration. (b) Fitted mean duration.



(c) Fitted mean speed. (d) Fitted mean speed.



(e) Fitted mean direction. (f) Fitted mean direction.

Figure 4.10: Three dimensional plots and histograms of fitted storm cell mean duration (top row), speed (middle row) and direction (bottom row) from June 2003.

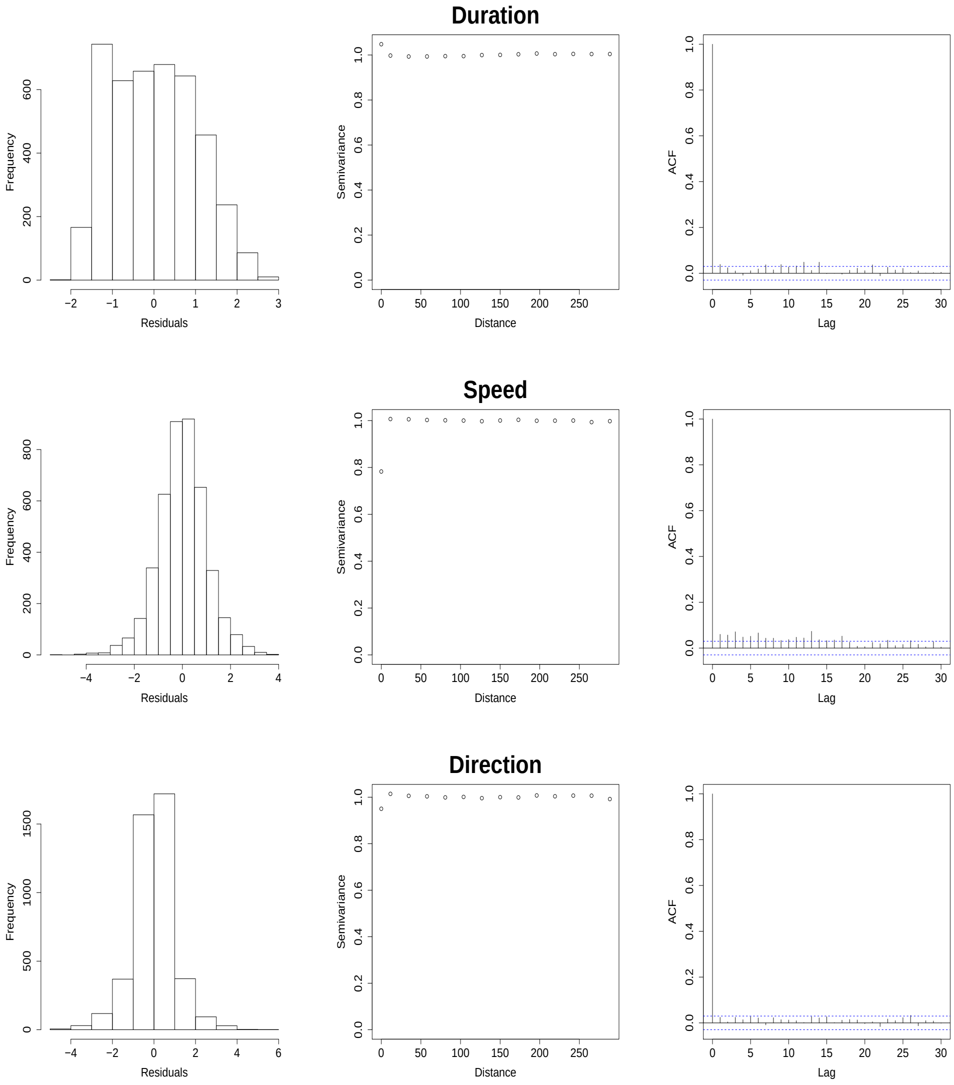
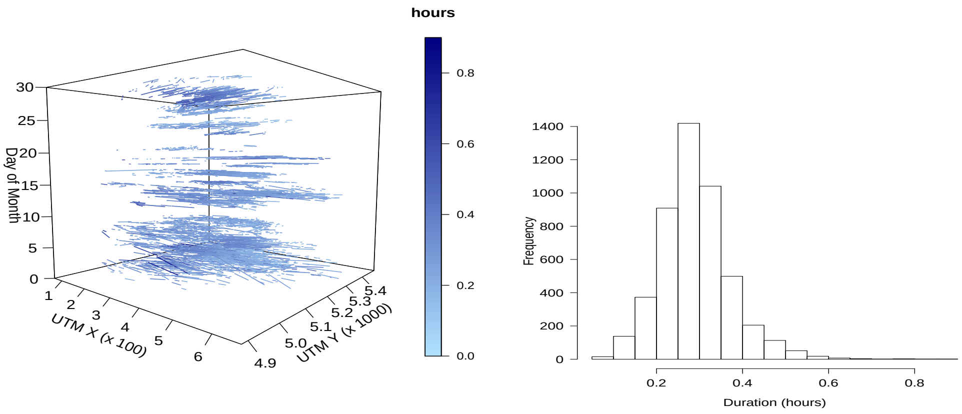


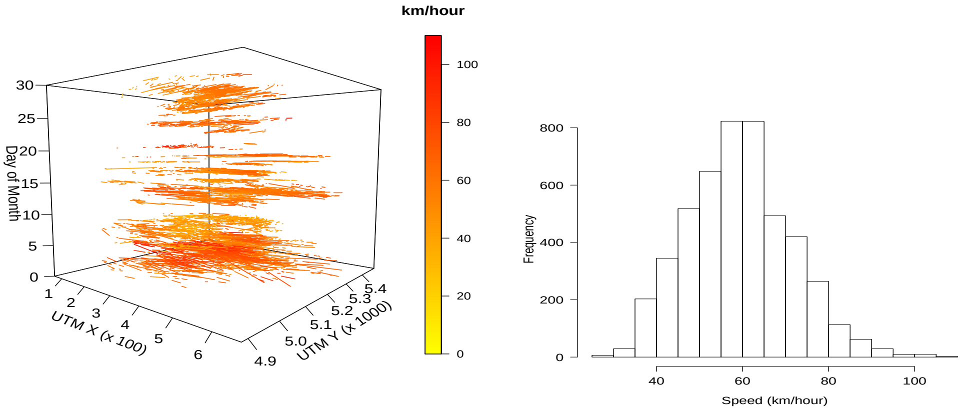
Figure 4.11: Histograms (left), variograms (middle) and autocorrelation function plots (right) of the normalized quantile residuals for the duration (top), speed (middle) and direction (bottom) components of the June 2003 mark models.

		Linear Effects			Non-linear Effects	
		Est.	CI		EDF	p-value
Probit						
	β^*	-13.80	(-15.20, -3.56)			
	σ_t^2	2.07	(0.75, 4.53)			
	σ_s^2	2.15	(1.25, 3.45)			
	ϕ_t	1.06	(0.76, 4.67)			
	ϕ_s	105.75	(78.17, 137.35)			
	τ	-16.13	(-17.36, -14.89)			
Duration						
μ_2	β_{210}	-1.752	(-1.812, -1.691)	$\varphi(s,t)$	78.74	<0.001
	$\cos(Y_4)$	0.483	(0.417, 0.549)			
	VIL	0.014	(0.011, 0.016)			
ϑ_2	β_{220}	-0.156	(-0.176, -0.136)			
Speed						
μ_3	β_{310}	3.936	(3.911, 3.962)	$\varphi(s,t)$	229.28	<0.001
	$\cos(Y_4)$	0.159	(0.132, 0.186)			
	VIL	0.0003	(-0.0004, 0.0010)			
ϑ_3	β_{320}	-1.099	(-1.119, -1.079)	$\varphi(s,t)$	188.62	<0.001
Direction						
μ_4	β_{410}	0.030	(0.025, 0.035)	$\varphi(s,t)$	201.02	<0.001
ϑ_4	β_{420}	1.693	(1.656, 1.731)	$\varphi(s,t)$	152.78	<0.001

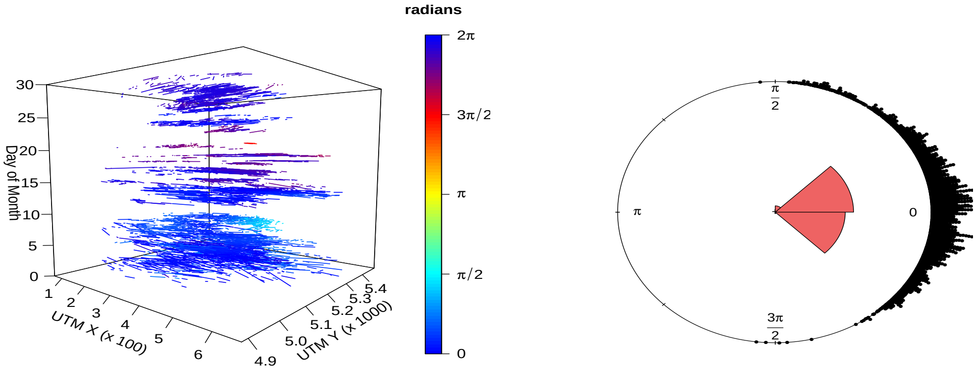
Table 4.4: Summary from the four component model for storm cell trajectory in July 2003. For the linear effects, parameter estimates (Est.) and 95% confidence intervals (CIs) of vertical integrated liquid (VIL) and cos(direction) are provided, and effective degrees of freedom (EDF) and p-values are given for the non-linear spatio-temporal effects.



(a) Fitted mean duration. (b) Fitted mean duration.



(c) Fitted mean speed. (d) Fitted mean speed.



(e) Fitted mean direction. (f) Fitted mean direction.

Figure 4.12: Three dimensional plots and histograms of fitted storm cell mean duration (top row), speed (middle row) and direction (bottom row) from July 2003.

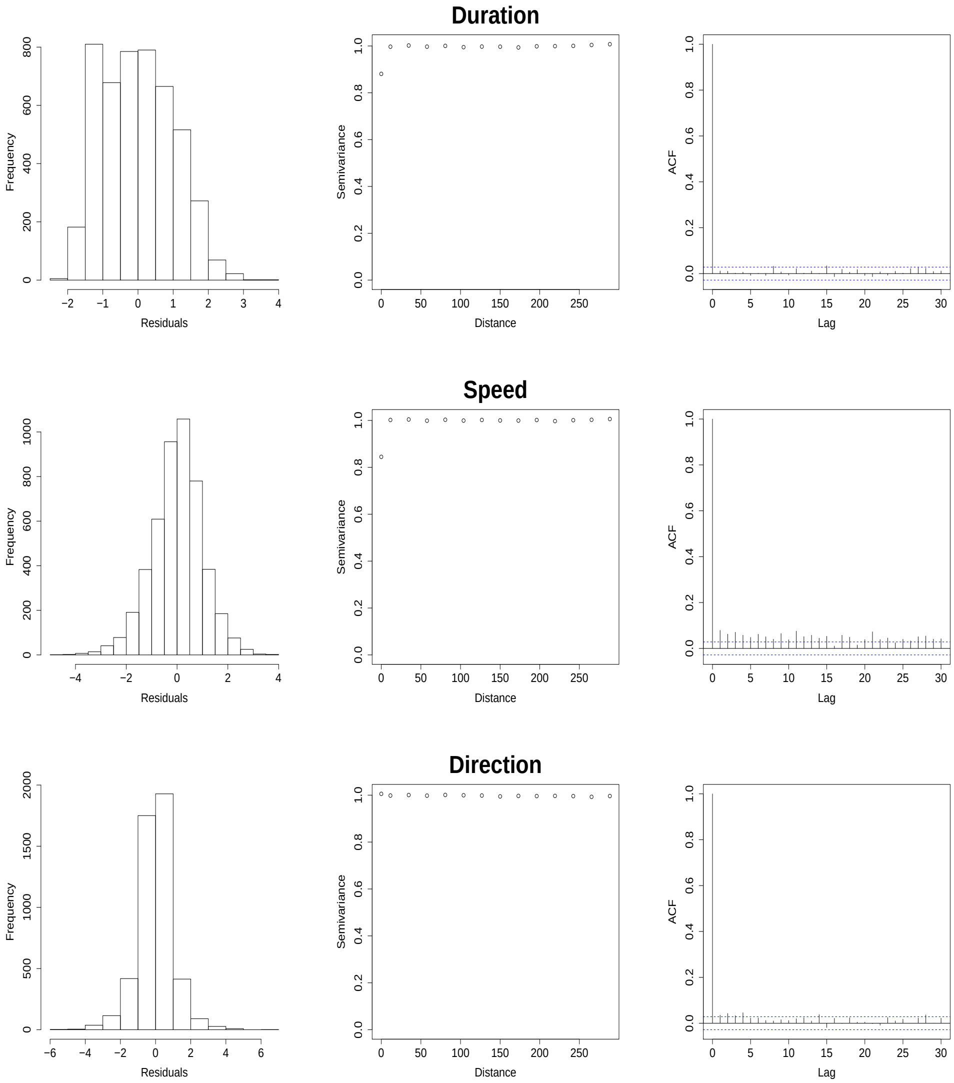
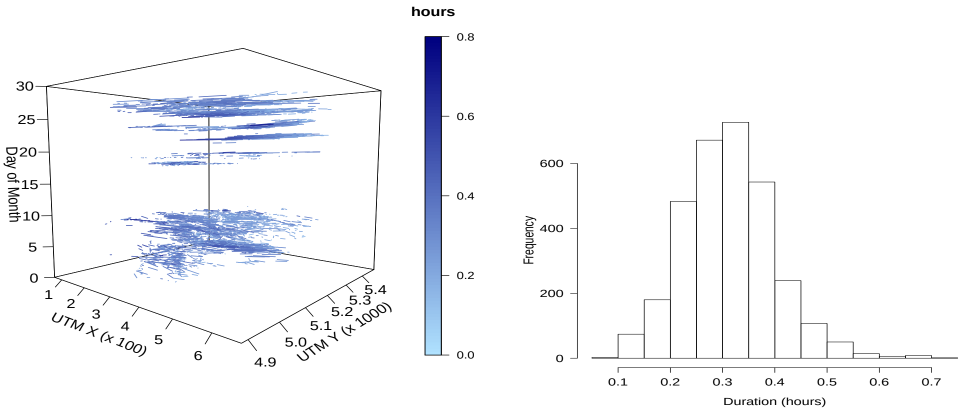


Figure 4.13: Histograms (left), variograms (middle) and autocorrelation function plots (right) of the normalized quantile residuals for the duration (top), speed (middle) and direction (bottom) components of the July 2003 mark models.

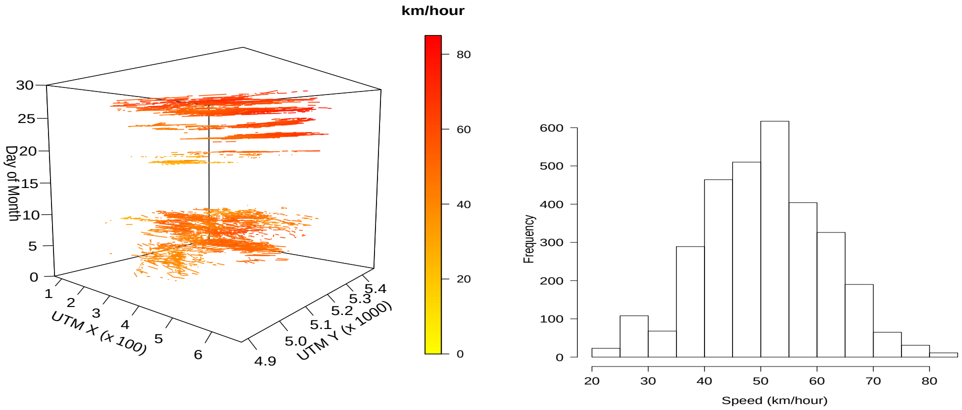
		Linear Effects			Non-linear Effects	
		Est.	CI		EDF	p-value
Probit						
	β^*	-13.90	(-15.33, -6.44)			
	σ_t^2	3.02	(2.44, 8.95)			
	σ_s^2	1.80	(0.79, 2.87)			
	ϕ_t	0.85	(0.52, 1.76)			
	ϕ_s	115.32	(93.39, 152.72)			
	τ	-16.42	(-17.81, -15.03)			
Duration						
μ_2	β_{210}	-1.556	(-1.609, -1.504)	$\varphi(s,t)$	74.10	<0.001
	$\cos(Y_4)$	0.446	(0.389, 0.503)			
	VIL	0.010	(0.007, 0.013)			
ϑ_2	β_{220}	-0.163	(-0.188, -0.138)			
Speed						
μ_3	β_{310}	3.844	(3.824, 3.864)	$\varphi(s,t)$	101.46	<0.001
	$\cos(Y_4)$	0.094	(0.072, 0.116)			
	VIL	0.0003	(-0.0009, 0.0014)			
ϑ_3	β_{320}	-1.064	(-1.089, -1.039)	$\varphi(s,t)$	109.44	<0.001
Direction						
μ_4	β_{410}	0.195	(0.183, 0.207)	$\varphi(s,t)$	434.24	<0.001
ϑ_4	β_{420}	1.225	(1.181, 1.268)			

Table 4.5: Summary from the four component model for storm cell trajectory in August 2003. For the linear effects, parameter estimates (Est.) and 95% confidence intervals (CIs) of vertical integrated liquid (VIL) and cos(direction) are provided, and effective degrees of freedom (EDF) and p-values are given for the non-linear spatio-temporal effects.



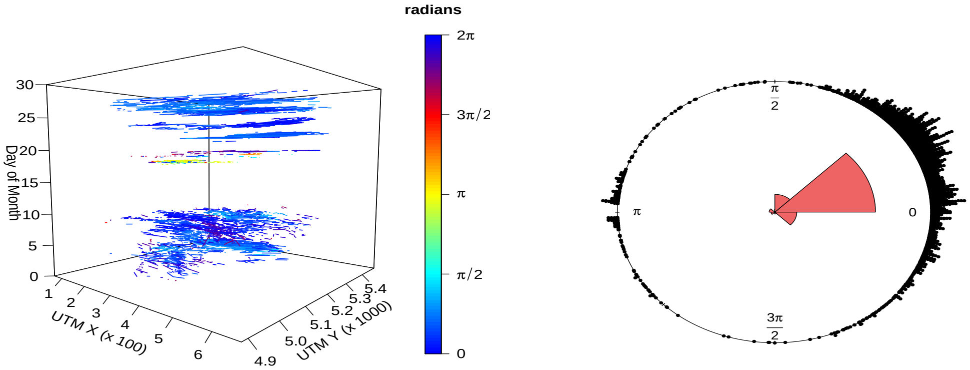
(a) Fitted mean duration.

(b) Fitted mean duration.



(c) Fitted mean speed.

(d) Fitted mean speed.



(e) Fitted mean direction.

(f) Fitted mean direction.

Figure 4.14: Three dimensional plots and histograms of fitted storm cell mean duration (top row), speed (middle row) and direction (bottom row) from August 2003.

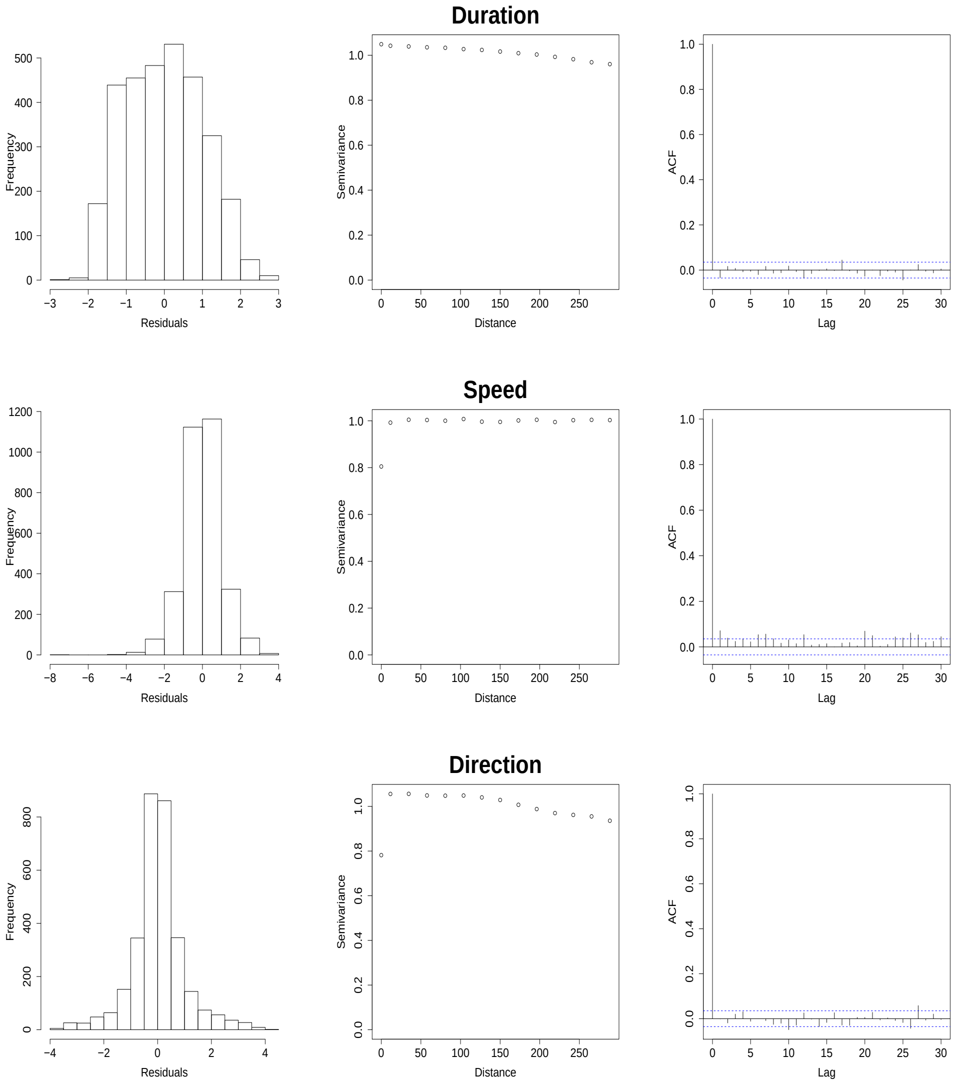


Figure 4.15: Histograms (left), variograms (middle) and autocorrelation function plots (right) of the normalized quantile residuals for the duration (top), speed (middle) and direction (bottom) components of the August 2003 mark models.

4.4.2 Data Errors

Out of the 26584 unique storm cells recorded between April and August 2003, 0.12% of them had time between consecutive scans recorded as more than 20 minutes. We believe that, based on the Storm Cell Identification and Tracking algorithm, these are errors in the data. Hence, they have been removed for the purpose of this analysis. However, we performed a sensitivity analysis by comparing the results presented in this section to models fit with these observations included and, in general, the resulting spatio-temporal partial effects and fitted values were not sensitive to this. There were also 0.62% of storm cells which were observed more than once with the same UTM X and UTM Y coordinates recorded at their first and last observations. We also believe these to be errors in the data based on how storm cell movement is tracked and, again, these observations were excluded.

4.5 Discussion

In this chapter, we extended the hierarchical cluster process developed in Chapter 3 to incorporate storm cell trajectories in a marked point process. This was done by building on methods typically employed for vector field data to the scenario in which the data are spatio-temporal with a potentially unobserved trajectory. Specifically, we modelled a storm cell's duration, speed and direction within a hurdle framework to account for the point mass at zero corresponding to observations which were only observed by radar at one instance. To link storm cell detection and movement we assumed that the Gaussian process that generated storms and the associated storm cells was also related to the distribution of storm cell duration. Parameter estimation was accomplished by modelling the mark process conditional on the point process which enabled techniques utilized for point-referenced data to be employed. Conditional on a storm cell having an observed trajectory, duration and speed were modelled as log-normal random variables with spatio-temporal splines incorporated into the location and scale parameters, as necessary. For these outcomes, vertical integrated liquid and a function of direction were also included as covariates. The von Mises distribution was employed for modelling the circular random variable direction with spatio-temporal splines again included in the mean and scale parameters in a generalized additive model framework. Including smoothing splines in the scale term allowed us to account for heavy-tailed distributions and also for correlation not modelled by the mean. Joint modelling in the ecological regression framework enabled us to explore and quantify the relationships between the outcomes duration and speed and the covariates direction and vertical integrated liquid. Chapter 5 builds on this idea of jointly estimating parameters.

For all months, no strong spatio-temporal trends in duration were observed, as indicated by the inclusion of only a mean smoother for all months except in April which also included a spatial smoother in the scale parameter. Clear spatio-temporal trends in speed were seen with clusters of storm cells moving at similar speeds. In the direction components, storm cells tended to move between $3\pi/2$ and $\pi/2$ radians and, typically, clusters of storm cells evolved in similar directions. For many of these models, we noted that more intense storm cells, as measured by their vertical integrated liquid, tended to be longer lasting, but slower moving. Meanwhile storm cells moving east tended to be longer lasting and faster moving.

We noted that 39% of storm cells were only observed on one occasion and that this may be due to limitations of the detection mechanisms and warrants further investigation. Storm cell trajectories have complicated spatio-temporal correlation structures, which we believe to be related to local weather conditions, such as wind speed and direction. Incorporating this into our modelling framework is of interest and could be done in a straightforward manner within the class of generalized additive models for location, scale and shape parameters. Incorporating wind speed and direction as covariates would allow Manitoba Hydro to identify likely locations of transmission line failures dynamically by simulating from these models when monitoring power flow and develop a decision rule to optimize cost. This would also require forecasting the necessary covariates. As we expect trends in speed and direction to be directly related to wind, joint modelling in a shared component framework could be employed to shed light on this joint spatio-temporal structure. As we saw in Figures 4.1-4.5, storm cells within a cluster have a tendency to move at the same speed and in the same direction. This was picked up by our models as displayed in Figure 4.6, 4.8, 4.10, 4.12 and 4.14. However, extending these findings to facilitate direct inference on storms and storm systems would be of interest. Additionally, further work to incorporate anisotropy could also be useful as covariance structures may have preferred orientations dictated by the mean circulation and nature of temperature gradients. This may increase efficiency when modelling storm cell trajectories. Further exploring the functional relationship between vertical integrated liquid and storm cell speed and direction is also of interest. Other extensions include jointly estimating the parameters in both the point and mark processes and dynamically modelling storm cell trajectories. We return to discuss these topics in more detail in Chapter 6.

Chapter 5

A General Framework for the Joint Modelling of Aggregated Spatial Point Patterns Subject to Clustering

Chapters 3 and 4 focussed on the joint modelling of point and marked point process data. As parameter estimation for spatio-temporal marked point patterns of this form is difficult and not well developed, joint modelling was performed via a two-stage approach where the events were first modelled and, subsequently, the marks were modelled conditional on the events. However, for multivariate aggregated point patterns, joint modelling using a joint parameter estimation scheme is a simpler task as likelihood-based techniques are computationally feasible. We therefore consider extensions of such processes in the context of aggregated data. This chapter develops a general framework for the joint modelling of multivariate zero-inflated count data, arising from multivariate aggregated point patterns.

5.1 Shared Component Models for Aggregated Point Patterns

In this chapter, we develop a general framework for the joint modelling of spatially aggregated multivariate point processes where the resulting counts exhibit zero-heaviness. For multivariate spatial outcomes, the use of so-called shared component or common factor models are often employed. These types of models assume that correlation exists between several outcomes at the same location as well as across locations for a given outcome. To account for this, a shared random effect is typically employed with scaling or factor loading parameters governing the outcome-specific strength of this effect. Such a model results in improved relative risk estimates by allowing all outcomes to borrow strength from each other and facilitates hypothesis testing for a shared spatial structure, which is often an important aspect of model interpretation. Common factor models for exponential family distributions with spatially correlated outcomes were developed in Wang and Wall (2003). Knorr-Held and Best (2001) developed a joint model for counts of two types of cancer with shared and disease-specific components where a cluster model was used for the underlying risk surface to incorporate a spatially-varying level of smoothing. Held et al. (2005) extended the idea of the shared component model to more than two outcomes that are functions of several random effects shared between all or a subset of the outcomes. Spatio-temporal joint models were developed for counts of six types of cancer in Tzala and Best (2008) and for male and female lung cancer incidence in Richardson et al. (2006). Feng (2015) proposed joint ecological regression models for bivariate counts by including each outcome as a covariate of the other as well as a shared spatial surface. Finally, Feng and Dean (2012) extended the idea of joint modelling to zero-inflated spatial counts by including two sets of shared random effects: one across the logistic components and one across the Poisson components.

In this project, we develop an overarching framework for the joint modelling of multivariate zero-inflated spatial outcomes using a shared component model. This formulation is unifying in that it brings together many special cases that exist in the literature. We provide a clear conceptual interpretation by assuming that, for each outcome, there exists an underlying spatial random field that governs the linkages across outcomes and across the zero-inflation and Poisson model components. Where these random fields exceed some threshold, corresponding outcomes following distributions with non-negative support are observed; otherwise only zeros are observed. To account for correlation across outcomes as well as model components within outcomes we include shared spatial random effects with outcome- and component-specific variances. The proposed model permits the use of correlated random fields for each outcome and it also accounts for an association between values of the random fields at each location with the means of the aggregated patterns. The random fields and possibly also the thresholds can be constrained to be the same across outcomes, if warranted by the application. More generally, we argue that the use of an underlying random field with a threshold provides a useful interpretation to the binary model component and along with that, insight into the unobserved spatial structure. For example, it enables us to explore spatial trends between and across components of the multivariate outcomes and identify whether or not these distributions are the same. Furthermore, it provides the scaling parameters in the binary component with a meaningful interpretation as the ratio of the spatial to unstructured variability. We illustrate our framework on two data sets exhibiting zero-inflation: the first being female and male Ontario lung and bronchus cancer incidence and the second being counts of lesions and host plants from a study of Comandra blister rust infection in lodgepole pine trees.

5.2 Shared Component Model Framework

5.2.1 Model Description

Assume that for each outcome there exists a latent random field, the mean of which may be modelled as a function of both observed and unobserved covariates and, conditional on these covariates, is normally distributed. These random fields typically represent environmental conditions which affect the generation of the observed outcomes. For example, in our cancer application, the random field may correspond to the underlying regional health status related to environmental factors, with higher values indicating poorer levels of health; for the Comandra blister rust data, it represents habitat suitability with larger values indicating areas with conditions conducive to the growth of infections. Specifically,

$$Y_{ij}^{*} = \boldsymbol{G}_{ij}\boldsymbol{\gamma}_j + \xi_{Y^*j}a_{Y^*i} + \xi_j b_{ij} + \varepsilon_{ij} \tag{5.1}$$

represents the latent field for regions $i = 1, \ldots, n$ and outcomes $j = 1, \ldots, J$. Here, $\boldsymbol{G}_{ij}$ is a vector containing covariate information corresponding to region i and outcome j, $\boldsymbol{\gamma}_j$ is the covariate effect, and $(\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{nj})^T \sim N(\mathbf{0}, \sigma_j^2 \boldsymbol{I}_n)$. Further, a_{Y^*i} and b_{ij} are random effects accounting for spatial structure across the random fields. That is, $(a_{Y^*1}, a_{Y^*2}, \ldots, a_{Y^*n})^T \sim N(\mathbf{0}, \Sigma_{a_{Y^*}})$ and similarly $(b_{1j}, b_{2j}, \ldots, b_{nj})^T \sim N(\mathbf{0}, \Sigma_{b_j})$ where $\Sigma_{a_{Y^*}}$ and Σ_{b_j} are spatial covariance matrices. We return to discuss these components as they apply to our framework later in this section, however, in general, the form of these matrices will depend on the type of data being analyzed. For example, the conditional autoregressive formulation, as described in Section 2.4.2, is natural for lattice data, but it may also be a suitable approximation to what is likely a correlation based on distances. The terms ξ_{Y^*j} and ξ_j denote the scaling or factor loading parameters for the jth outcome corresponding to the random effects a_{Y^*i} and b_{ij}, respectively. These parameters allow the magnitudes of the common spatial factor to vary across outcomes and acknowledges that, although two outcomes may have a common spatial surface they can have different scales and hence the influence of the common spatial factor may differ across outcomes.

If the latent random field exceeds an outcome-specific threshold, random variables Y_{ij} are assumed to follow an exponential family distribution f_j with non-negative support; otherwise Y_{ij} is identically zero. That is,

$$Y_{ij} \sim \begin{cases} 0, & \text{if } Y_{ij}^{*} \leq \tau_j \\ f_j(\lambda_{ij}), & \text{if } Y_{ij}^{*} > \tau_j \end{cases}$$

where τ_j represents the threshold above which we observe outcomes from the distribution f_j and λ_{ij} represents the mean of that distribution. Letting $P\left(Y_{ij}^{*} \leq \tau_j\right)$ be denoted π_{ij}, as a consequence of assuming the Gaussian distribution in Equation (5.1), we have

$$\pi_{ij} \quad = \quad \Phi\left(\frac{\tau_j - [\boldsymbol{G}_{ij}\boldsymbol{\gamma}_j + \xi_{Y^*j}a_{Y^*i} + \xi_j b_{ij}]}{\sigma_j}\right). \tag{5.2}$$

Further, to model the mean parameter, λ_{ij}, we formulate

$$\varrho_j(\lambda_{ij}) \quad = \quad \boldsymbol{B}_{ij}\boldsymbol{\beta}_j + \nu_{\lambda j}a_{\lambda i} + \nu_j b_{ij} + h_{ij} \tag{5.3}$$

where $\varrho_j(\cdot)$ is a known link function, $\boldsymbol{B}_{ij}$ and $\boldsymbol{\beta}_j$ are the covariates and parameters influencing the mean component for the jth outcome and $a_{\lambda i}$ is a spatial random effect where $(a_{\lambda 1}, a_{\lambda 2}, \ldots, a_{\lambda n})^T \sim N(\mathbf{0}, \Sigma_{a_\lambda})$. Again, $\nu_{\lambda j}$ and ν_j are scaling parameters. Across outcomes, the random effects a_{Y^*i} and $a_{\lambda i}$ account for the correlation of the underlying surfaces and the means of the f_j components, respectively, whereas the random effects b_{ij}, link the underlying surface and f_j components for the jth outcome. The outcome-specific terms, b_{ij}, are important to include as, in many scientific contexts, it is likely that the underlying random fields governing the presence of the outcomes are correlated with the means of the observed outcomes. Note that with the model parameterized as in Equations (5.2) and (5.3), with all else held constant, a larger value of the random effect, b_{ij}, corresponds to a lower probability of observation (i,j) belonging to the structural zero component (or larger probability of being in the Poisson component) and larger Poisson mean. Finally, $(h_{1j}, h_{2j}, \cdots, h_{nj})^T \sim N(0, \sigma^2_{hj}\boldsymbol{I}_n)$ provides additional flexibility to account for any excess variability in the jth outcome not accounted for through the shared spatial random effects.

This framework allows us to account for: 1) regional effects that operate on the zero-inflation components for all outcomes, 2) regional effects that operate on the means for the non-negative components, 3) outcome-specific regional components that operate on the zero-inflation and non-negative mean components and 4) additional unstructured variability in the means of the non-negative components. In some applications, it may be plausible that all outcomes and components are correlated through the same latent spatial surface. In such cases, a single shared spatial random effect can be used resulting in a simplified structure. Additionally, where outcomes are very closely linked a simpler model with $\tau_j = \tau,\ j = 1, 2, \ldots, J$ may be preferred.

Two important models may be viewed as special cases. The first is that of Rathbun and Fei (2006). This is a univariate zero-inflated Poisson model applied to point referenced data with the underlying random field having a Matérn covariance structure and corresponds to the special case of our model in which all outcomes and components are independent. Our proposed framework simplifies to the model developed in Feng and Dean (2012) if no between-component correlation is present in the data and the logit link is employed in the zero-inflation submodel.

5.2.2 Parameter Estimation and Inference

This joint model may be fit in a Bayesian framework with the conditional likelihood

$$\mathcal{L}(\boldsymbol{Y} \mid \cdot) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \left[\pi_{ij} + (1 - \pi_{ij})\mathrm{e}^{-\lambda_{ij}}\right]^{I(y_{ij}=0)} \left[(1 - \pi_{ij}) \frac{\mathrm{e}^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}\right]^{I(y_{ij}>0)} \right\} \tag{5.4}$$

by calling `OpenBUGS` (Sturtz et al., 2005) through `R` (R Core Team, 2016). In Equation (5.4), $I(A)$ denotes an indicator variable for the event A with π_{ij} and λ_{ij} as defined in Equations (5.2) and (5.3), respectively. We employ standard normal priors for the parameters $\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}_j$ and τ_j and gamma(1,1) prior distributions are used on the precision parameters from the independent normal random effect terms. Appropriate priors for the scaling parameters include gamma (e.g. Feng and Dean, 2012), log-normal (e.g. Held et al., 2005) and half-normal (e.g. Gelman, 2006). Finally, intrinsic conditional autoregressive priors are placed on the spatial random effects.

Convergence is assessed via trace plots and the Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992). By examining histograms similar to those suggested in Gelman (2006) where the prior density curve was overlaid on a histogram of the posterior Markov chain Monte Carlo samples, we investigate how informative our choices of priors are. Model assessment is based on the posterior predictive distribution (Meng, 1994). Specifically, we calculate Bayesian p-values based on Pearson and deviance residuals to assess lack of fit. The deviance information criterion is employed for model comparison

(Spiegelhalter et al., 2002). This quantity is calculated based on the posterior mean as well as the posterior median as estimates of the effective number of parameters may not be invariant to the quantity used to calculate it. Furthermore, we adopt the guidelines suggested by Spiegelhalter et al. (2002) for model comparison: models within one or two units of the best model (lowest deviance information criterion) should be considered, while those within three and seven units are deemed to be considerably inferior.

Identifiability issues arise in our setting. First, for a shared random effect, the variance parameter in addition to all scaling parameters are not jointly identifiable. We therefore follow the suggestion of Wang and Wall (2003) to fix the variance parameter at one, estimating all scaling parameters. Second, due to the identifiability constraints in the probit component, we set the standard deviation of the error term in the underlying random field, σ_j, to one. Finally, and as was done in Chapter 4, since the zero-inflation component intercept is intrinsically linked to the threshold parameter we set the intercept to zero and estimate the threshold. Together, the first two identifiability conditions in the zero-inflation component give the scaling parameters, ξ_{Y^*j} and ξ_j, a useful interpretation as the ratio of the standard deviation of the structured (spatial) term to that of the unstructured (independent) term in the random field.

5.3 Applications

5.3.1 Ontario Lung and Bronchus Cancer

We consider lung and bronchus cancer incidence for females and males ages 50-59 in 2010 across the 49 public health units in Ontario, Canada. Ontario public health units are health agencies composed of rural and urban municipalities responsible for health promotion and disease prevention programs (Ontario Ministry of Health and Long-Term Care). Analyzing the spatial distribution of lung and bronchus cancer rates from this age group is of particular interest because studies have shown a relationship between lung cancer rates and Ontario miners (Kusiak et al., 1993), an occupation performed primarily in Northern Ontario (Ontario Mining Association). The 50-59 age group is of interest here as we suspect this group might have had enough exposure to carcinogenic environmental conditions that the resulting symptoms may be present.

As is common with spatially explicit public health data for rare diseases, these counts have been randomly rounded to the nearest five to maintain anonymity (rounded to the nearest five with some probability). Maps of the standardized incidence ratios, defined as the ratio of observed to expected incidence, which consist of 24.5% and 20.4% zeros for females and males, respectively, are displayed in Figure 5.1. Although there are some differences across males and females in Southern Ontario, the contrasts are striking in Northern Ontario where males have relatively large values in comparison to females. Even though we are not asserting causality here, this perhaps warrants investigation. It is important to note, however, that since these regions have low populations, the corresponding standardized incidence ratios are subject to high variability.

When fitting this model, no covariates are included in the zero-inflation component while an offset representing the log of the expected number of cases, denoted E_{ij}, is included in the Poisson component. Parameter estimation is based on one long chain with 75000 iterations, the first 25000 being discarded as burn-in and log-normal(0,1) priors distributions for the scaling parameters. As mentioned previously, of most interest in these types of models are the spatial random effects and scaling parameters that allow us to explore the spatial clustering across outcomes as well as components and to test for common spatial structures. In our analysis, we focus on the posterior median estimates and 95% highest posterior density credible intervals for these terms. In the zero-inflation components, the spatial variability is dominated by the common factor connecting these two components as demonstrated by the magnitudes of $\hat{\xi}_{Y^*1}$ and $\hat{\xi}_{Y^*2}$, which are 1.101 (0.185, 3.530) and 1.331 (0.216, 3.780), with the numbers in brackets representing the 95% credible intervals. Recall that these parameters represent the ratio of the spatial to unstructured standard deviation. Therefore, with these credible intervals including the value one, we cannot detect a difference in these quantities. Furthermore, the posterior estimate of the difference $\widehat{\xi_{Y^*1} - \xi_{Y^*2}}$ is -0.168 (-2.590, 2.157); the hypothesis of the difference being zero is not rejected indicating that we cannot detect a difference in this spatial structure across females and males. The scaling parameters for $\hat{\xi}_1$ and $\hat{\xi}_2$ are 0.635

(0.113, 2.496) and 0.757 (0.135, 3.017) for females and males, respectively. In the Poisson components, $\hat{\nu}_{\lambda 1}$ and $\hat{\nu}_{\lambda 2}$ correspond to the scaling parameters for the component-specific random effects for females and males; their posterior estimates are 0.317 (0.082, 0.712) and 0.270 (0.078, 0.591), respectively, and once again the hypothesis that these scaling parameters are not different cannot be rejected with a median posterior estimate of the difference $\widehat{\nu_{\lambda 1} - \nu_{\lambda 2}}$ being 0.044 (-0.320, 0.454). The estimated value of ν_1 is 0.275 (0.072, 0.659) and ν_2 is 0.248 (0.070, 0.572). Finally, $\hat{\sigma}^2_{h1}$ and $\hat{\sigma}^2_{h2}$ are 0.225 (0.121, 0.441) and 0.163 (0.091, 0.308), respectively.

We next fit a simplified version of this model in which $\xi_{Y^*1} = \xi_{Y^*2}$ and $\nu_{\lambda 1} = \nu_{\lambda 2}$; the parameter estimates and their corresponding 95% credible intervals are displayed in Table 5.1. Based on this, the estimated scaling parameters, $\hat{\xi}_{Y^*}$ and $\hat{\nu}_{\lambda}$, are equal to 1.501 (0.219, 3.559) and 0.241 (0.067, 0.539), respectively. For females, these terms account for 81.2% and 43.2% of the estimated empirical spatial variability across the zero-inflation and Poisson components calculated from the Markov chain Monte Carlo chains. The analogous percentages for males are 79.4% and 49.0%. (See Table 5.2 for a complete summary of the proportion of estimated spatial variability by outcome and model component.) The remaining spatial variability is explained through the sex-specific random effects, which are joint across the zero-inflation and Poisson components. For females the term $\hat{\xi}_1$ is 0.659 (0.121, 2.314) and $\hat{\nu}_1$ is 0.279 (0.073, 0.660). The analogous terms for males ($\hat{\xi}_2$ and $\hat{\nu}_2$) are similar in magnitude to females and have posterior estimates of 0.728 (0.126, 2.689) and 0.254 (0.069, 0.577). In the zero-inflation components, the magnitudes of these estimates are consistently smaller than that of the corresponding component-specific random effects. Therefore, this component-specific random effect describes more of the joint spatial structure. In the Poisson components, all scaling parameters have similar magnitudes. Finally, the unstructured variance term for females is estimated as 0.228 (0.122, 0.438) and for males is 0.162 (0.091, 0.312).

Figure 5.2 maps the posterior median estimates of the shared conditional autoregressive random effects from the simplified model although we note that maps of the posterior estimates of the spatial random effects based on the first model are almost identical. The top left panel displays the sex-specific shared random effect across the zero-inflation components ($\boldsymbol{a}_{Y^*}$) where the prominent trend is across Southern Ontario with the posterior estimates decreasing from west to east. For the sex-specific random effect across the Poisson components ($\boldsymbol{a}_{\lambda}$), there is similarly a trend in Southern Ontario where the posterior estimates of the random effect slightly increased from west to east. In Figure 5.2, we notice that the sex-specific random effects, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, tend to pick up spatial differences in Northern Ontario where males have larger posterior estimates than females. This difference was also highlighted in the maps of standardized incidence ratios (Figure 5.1). Posterior estimates of the unstructured random effects are displayed in Figure 5.3. These do not show any residual spatial structure, confirming that the structured variability is accounted for through the spatially explicit terms.

Overall, when we overlaid the prior distribution on the histogram of the posterior, the scaling parameters from the zero-inflation components were constrained by this choice. However, this same observation was not true for the other parameters. As mentioned above, model goodness of fit was assessed via posterior predictive p-values based on Pearson and deviance residuals; no strong evidence of lack of fit was detected. Table 5.3 displays the deviance information criterion and effective number of parameters, p_D, for our proposed model as well as the simplified version presented here, identified as models J1A and J1B, along with competing models, which we describe in detail as this discussion continues. Note that Table 5.4 offers further description of the joint models considered. Overall, the deviance information criterion estimates from the joint models (identified as models beginning with “J”) are smaller than those of the separate models (models beginning with “S”), indicating a better fit to these data. Models S1 and S2 assume that female and male processes are independent with correlation structure between the model components in the former via a shared random effect and for the latter, all components and outcomes are assumed to be independent. The difference in deviance information criterion between models S1 and S2 is negligible, regardless of whether a mean- or median-based criterion is used. The three alternative joint structures considered, J2-J4, assume a simplified shared spatial structure, and are described as follows: model J2 includes a single shared random effect across all components and outcomes, model J3 considers only a shared structure across the zero-inflation and Poisson components and model

	Females (j=1)		**Males (j = 2)**	
	median	CI	median	CI
τ_j	-1.411	(-2.526, -0.727)	-1.511	(-2.508, -0.826)
ξ_{Y^*}	1.501	(0.219, 3.559)	1.501	(0.219, 3.559)
ξ_j	0.659	(0.121, 2.314)	0.728	(0.126, 2.689)
β_j	0.161	(-0.051, 0.362)	0.158	(-0.026, 0.335)
ν_λ	0.241	(0.067, 0.539)	0.241	(0.067, 0.539)
ν_j	0.279	(0.073, 0.660)	0.254	(0.069, 0.577)
σ^2_{hj}	0.228	(0.122, 0.438)	0.162	(0.091, 0.312)

Table 5.1: Posterior median estimates and 95% credible intervals (CIs) from the joint spatial zero-inflated Poisson model for female and male cancer counts.

J4 is an extension of J2 to also include outcome-specific random effects connecting the zero-inflation and Poisson components. Of the joint models, J2 has the largest deviance information criterion indicating that a single shared random effect is too restrictive a structure. The most appropriate model based on this criterion is the simplified version of our proposed model, J1B. When comparing deviance measures based on the mean posterior estimates, the deviance information criterion is similar for models J1A, J3 and J4 with J1B being an improvement over all joint models. However, the median-based estimate suggests that both models J1A and J1B offer a considerable improvement over J3.

In addition to an improved fit, models J1A and J1B provide insight into the shared spatial structure beyond the previously mentioned models currently available in the literature. Specifically, we are able to visualize and estimate the sex-specific correlation between the zero-inflation and Poisson components, which allows us to capture differences in the spatial structure across the multivariate outcomes; this type of association was not examined in Rathbun and Fei (2006). Further, in the Feng and Dean (2012) model (J3) this spatial structure would likely be picked up in the structured variability term. This is especially important in exploring sex-specific trends across Ontario, as we pointed out in Figure 5.2.

	Zero-Inflation		**Poisson**	
	Outcome-Specific	Component-Specific	Outcome-Specific	Component-Specific
Females	18.8	81.2	56.8	43.2
Males	20.6	79.4	51.0	49.0

Table 5.2: Proportion of spatial variability by model component and outcome.

Model	Components	Mean DIC	Mean p_D	Median DIC	Median p_D
Joint					
J1A	$Y^*_{ij} = \xi_{Y^*j} a_{Y^*i} + \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_{\lambda j} a_{\lambda i} + \nu_j b_{ij} + h_{ij}$	517.2	69.1	499.5	51.6
J1B	$Y^*_{ij} = \xi_{Y^*} a_{Y^*i} + \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_{\lambda} a_{\lambda i} + \nu_j b_{ij} + h_{ij}$	514.5	68.4	500.8	54.8
J2	$Y^*_{ij} = \xi_j b_i + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_j b_i + h_{ij}$	520.5	63.3	509.6	52.4
J3	$Y^*_{ij} = \xi_{Y^*j} a_{Y^*i} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_{\lambda j} a_{\lambda i} + h_{ij}$	517.9	63.2	508.0	53.4
J4	$Y^*_{ij} = \xi_j b_i + \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_j b_i + \nu_j b_{ij} + h_{ij}$	517.8	68.8	498.6	49.8
Separate					
S1	$Y^*_{ij} = \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_j b_{ij} + h_{ij}$	523.7	61.6	513.2	51.2
S2	$Y^*_{ij} = \xi_{Y^*j} b_{Y^*ij} + h_{ij}$ $\log(\lambda_{ij}) = \beta_j + \log E_{ij} + \nu_{\lambda j} b_{\lambda ij} + h_{ij}$	523.2	60.6	513.4	50.9

Table 5.3: Comparison of deviance information criterion (DIC) and effective number of parameters (p_D) for competing models in the analysis of Ontario lung and bronchus cancer incidence. For models J1A, J1B and J3, in the ith region, a_{Y^*i} and $a_{\lambda i}$ refer to shared random effects across the zero-inflation and Poisson components, respectively. In models J1A, J1B, J4 and S1, b_{ij} represents the shared random effect across model components within the jth outcome. The term b_i in models J2 and J4 represents a shared random effect across all outcomes and components and in S2, b_{Y^*ij} represents a random effect for the jth outcome in the zero-inflation component and similarly for $b_{\lambda ij}$ in the Poisson component. Note that Table 5.4 offers further description of the joint models considered.

Model	Joint Effect	Description
J1A	b_{ij}	- links the zero-inflation components with the Poisson components
		- different for each individual and outcome
	a_{Y^*i}	- individual-specific random effect operating on all outcomes in the zero-inflation component
	$a_{\lambda i}$	- individual-specific random effect operating on all outcomes in the Poisson component
J1B	b_{ij}	- links the zero-inflation components with the Poisson components
		- different for each individual and outcome
	a_{Y^*i}	- individual-specific random effect operating on all outcomes in the zero-inflation component
		- this is the same as model J1A with the scaling parameter being the same across outcomes
	$a_{\lambda i}$	- individual-specific random effect operating on all outcomes in the Poisson component
		- this is the same as model J1A with the scaling parameter being the same across outcomes
J2	b_i	- random effect linking zero-inflation components with Poisson components
		- the difference between this and J1A is that this random effect is the same for all outcomes
J3		- no random effect linking the zero-inflation and Poisson components
J4	b_i	- links zero-inflation and Poisson components, is common to all outcomes
	b_{ij}	- links zero-inflation and Poisson components, differs across outcomes

Table 5.4: Description of the shared random effects for models compared in Table 5.3.

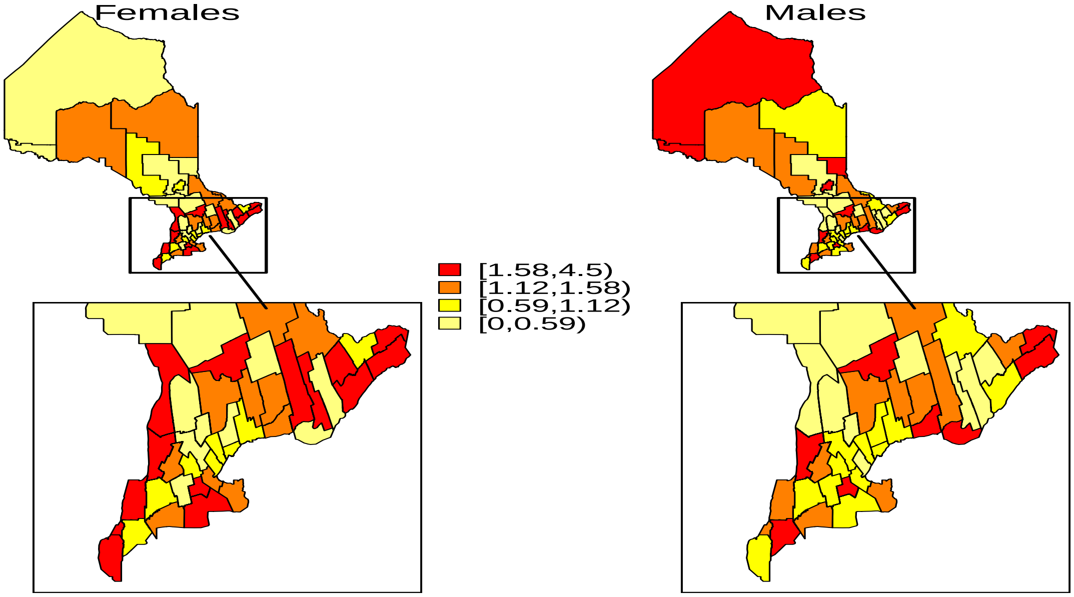


Figure 5.1: Standardized incidence ratios of Ontario lung and bronchus cancer incidence for females (left) and males (right).

5.3.2 Comandra Blister Rust Infection of Lodgepole Pine Trees

In hard pine trees, fungus growing on the inner bark has the potential to cause growth reduction, stem deformation and mortality, in addition to a disease referred to as Comandra blister rust. To spread from one tree to another Comandra blister rust requires an alternative host plant, known as bastard toad flax. Within a tree, these data may be viewed as a point pattern for the location of lesions resulting from infections and host plants promoting infection. However, interest here focusses on examining the shared spatial distribution in the counts of lesions and host plants by tree. Our data, from the Ministry of Forests, Lands and Natural Resource Operations, contain counts of lesions and host plants by tree. Each tree is located at the centre of a 1.5 squared meter (m^2) cell of a 124 × 64 grid. In this analysis, we analyze counts of 1000 randomly sampled cells from this grid, as considered in Feng and Dean (2012). This random sampling was due to constraints with the software as using all recorded data was computationally prohibitive. These data consist of 66.6% and 81.1% zeros for the counts of lesions and host plants within grid cells, respectively. Figure 5.4 displays maps of the sampled lesions and host plants which have been linearly interpolated using the `akima` (Akima and Gebhardt,

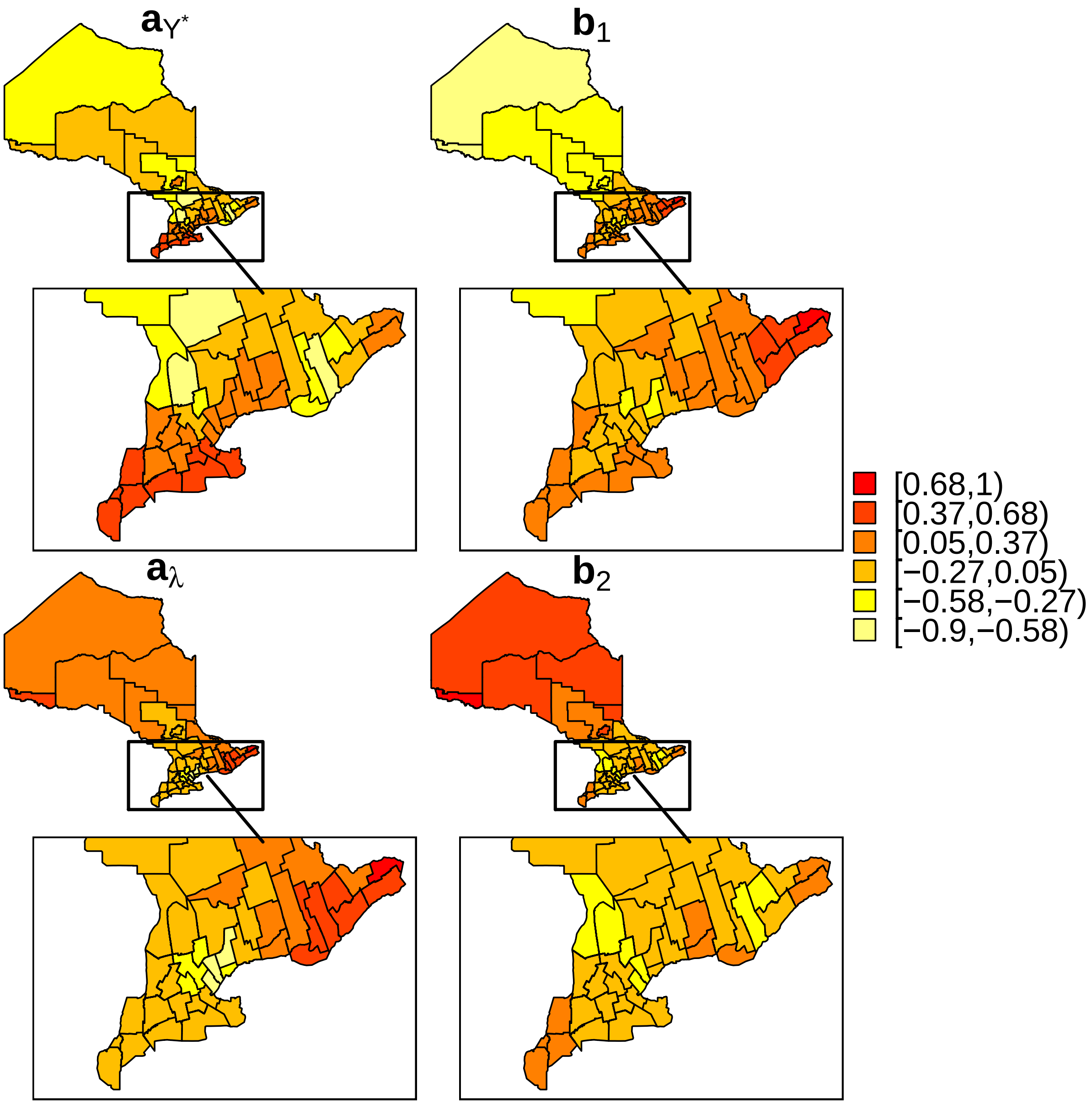


Figure 5.2: Median posterior estimates of the conditional autoregressive random effects for all outcomes and model components when fit to the lung and bronchus cancer data. This model contains four joint random effects: 1) across the zero-inflation components ($\boldsymbol{a}_{Y^*}$), 2) across the Poisson components ($\boldsymbol{a}_\lambda$), 3) across components for the female outcome ($\boldsymbol{b}_1$) and 4) across components for the male outcome ($\boldsymbol{b}_2$).

2015) `R` package as well as histograms displaying the zero-heavy nature of these data.

Here we follow Feng and Dean (2012) who analyzed these data with a conditional autoregressive structure and assumed two cells are neighbours if the Euclidean distance between them is less than or equal to 20 m; this was based on an estimate derived

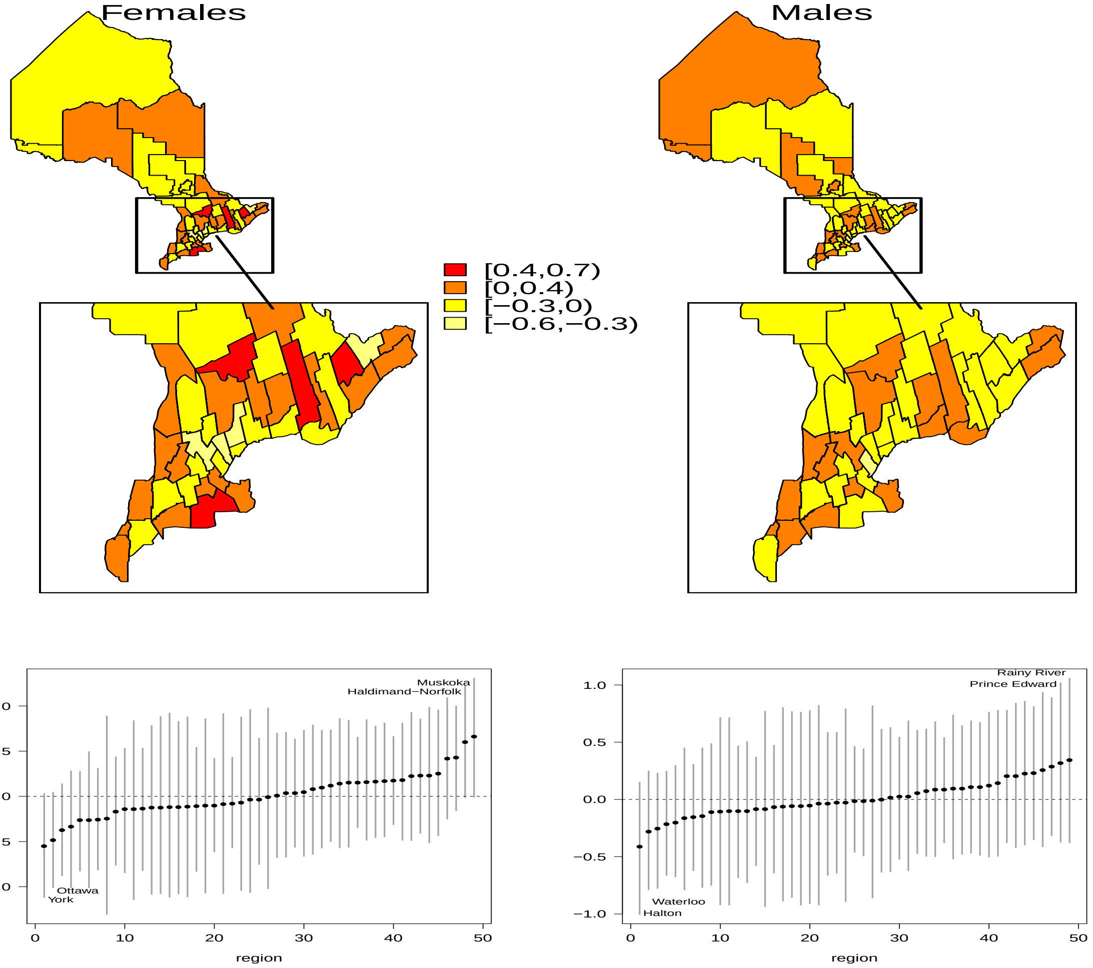


Figure 5.3: Median posterior estimates and 95% credible intervals for the unstructured random effects for females (left) and males (right), ordered by increasing median estimate.

from the analysis of an empirical semivariogram as well as scientific considerations. To estimate model parameters a half-normal(0,1) prior distribution is used for the scaling parameters and we run one long chain with 200000 iterations, the first 50000 of which we discard as burn-in and we thin every twentieth. This results in 7500 posterior samples from which to base our inference.

Table 5.5 provides a complete summary of parameter estimates and 95% credible intervals for the final model in which only one shared random effect is required to adequately account for the spatial correlation within and between outcomes. For lesions, spatial variability is much stronger in the Poisson component than in the zero-inflation component; this is evident through an examination of the estimated scaling parameters which are 1.598 (0.201, 3.175) and 3.955 (2.895, 5.071) in the zero-inflation and Poisson components, respectively. Furthermore, the estimated unstructured variability term, $\hat{\sigma}^2_{h1}$, is 0.261 (0.128, 0.483). For the host plants, the posterior estimate of ξ_2 is 4.334 (3.152, 5.591), and in the Poisson component $\hat{\nu}_2$ is 2.220 (0.634, 3.829) and $\hat{\sigma}^2_{h2}$ is 1.170 (0.811, 1.674). As was done in the previous analysis, we can test for a common spatial structure within each component. When doing this we get an estimate of -2.745 (-4.643, -0.755) for $\widehat{\xi_1 - \xi_2}$ in the zero-inflation component and for $\widehat{\nu_1 - \nu_2}$ in the Poisson component we get 1.729 (-0.129, 3.598). Therefore, we only detect a significant difference between the scaling parameters in the zero-inflation component.

Posterior median estimates of the conditional autoregressive random effects, with variance one, are displayed in Figure 5.5. From this, we can see smaller posterior estimates, indicating smaller Poisson means in the north-west quadrant, with these estimates being larger in the surrounding areas. Maps of the posterior median estimates of the unstructured random effects, as shown in Figure 5.6, do not display spatial structure with the exception of an absence of host plants in the north-west quadrant.

For this model, no lack of fit is detected when examining the Pearson and deviance-based posterior predictive p-values. We note that the spatial trends, as shown in Figure 5.5, are not dependent on the choice of prior distribution. As was done in Section 5.3.1, Table 5.6 displays estimates of the deviance information criterion and p_D using both the posterior mean and median estimates for joint and separate models. Again, in the separate models, S1 and S2, the latter assuming independence across all outcomes and components and the former incorporating dependence between the zero-inflation and Poisson components for each outcome, are clearly inferior to the joint models. For the joint models, J1, which is the general framework detailed in Section 5.2.1, and J3, which assumes dependence between the zero-inflation components for lesions and host plants and also across the Poisson components, a negligible difference in the mean-based deviance information criterion is observed, while the median-based criterion indicates a clear improvement of model J1 over J3. Regardless as to which parameterization is employed, model J2, with a single shared random effect across all model components, has the smallest estimated deviance information criterion. In addition, we perform a

	Lesions (j=1)		**Host Plants (j = 2)**	
	median	CI	median	CI
τ_j	-0.710	(-1.491, -0.326)	0.971	(0.829, 1.117)
ξ_j	1.598	(0.201, 3.175)	4.334	(3.152, 5.591)
β_j	-0.692	(-0.998, -0.434)	1.127	(0.794, 1.430)
ν_j	3.955	(2.895, 5.071)	2.220	(0.634, 3.829)
σ^2_{hj}	0.261	(0.128, 0.483)	1.170	(0.811, 1.674)

Table 5.5: Posterior median estimates and 95% credible intervals (CIs) from the joint spatial zero-inflated Poisson model for lesions and host plants.

sensitivity analysis as to the choice of neighbourhood structure utilized in the conditional autoregressive random effect. Specifically, we consider three situations: two observations are assumed to be neighbours if their Euclidean distance is less than 1) 10 m, 2) 25 m and 3) 30 m. We found that in the latter two scenarios the parameter estimates as well as the distribution of the posterior spatial random effect are similar to the results presented in this section. However, in the first scenario, the scaling parameters are considerably smaller indicating smaller estimates of the spatially-structured variability. We further investigate the effect of misspecifying the spatial structure in terms of imposed bias and mean squared error in Section 5.4.2.

As mentioned previously, one of the advantages of being able to decompose the spatial variability between and across the multivariate components is the potential to identify situations in which the spatial structure is shared across all components, as is the case for these data. This results in a more parsimonious fit than the common models currently employed; it also enables a simpler interpretation of model components in terms of the applications. These results also clearly indicate that there is considerable correlation between the random field generating the counts and the mean count; an association which is often overlooked.

		Mean		Median	
Model	Components	DIC	p_D	DIC	p_D
Joint					
J1	$Y^*_{ij} = \xi_{Y^*j} a_{Y^*i} + \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \nu_{\lambda j} a_{\lambda i} + \nu_j b_{ij} + h_{ij}$	3746.9	390.3	3664.6	307.1
J2	$Y^*_{ij} = \xi_j b_i + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \nu_j b_i + h_{ij}$	3676.3	414.6	3646.6	383.7
J3	$Y^*_{ij} = \xi_{Y^*j} a_{Y^*i} + \varepsilon_{ij}$ $\log(\lambda_{ij}) + \beta_j + \nu_{\lambda j} a_{\lambda i} + h_{ij}$	3745.9	378.0	3716.7	348.3
Separate					
S1	$Y^*_{ij} = \xi_j b_{ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \nu_j b_{ij} + h_{ij}$	3775.8	402.4	3737.0	361.9
S2	$Y^*_{ij} = \xi_{Y^*j} b_{Y^*ij} + \varepsilon_{ij}$ $\log(\lambda_{ij}) = \beta_j + \nu_{\lambda j} b_{\lambda ij} + h_{ij}$	3779.0	411.6	3749.8	380.5

Table 5.6: Comparison of deviance information criterion (DIC) and effective number of parameters (p_D) for competing models in the analysis of lesions and host plants. For models J1 and J3, in the ith region, a_{Y^*i} and $a_{\lambda i}$ refer to shared random effects across the zero-inflation and Poisson components, respectively. In models J1 and S1, b_{ij} represents the shared random effect across model components within the jth outcome. The term b_i in model J2 represents a shared random effect across all outcomes and components and in S2, b_{Y^*ij} represents a random effect for the jth outcome in the zero-inflation component and similarly for $b_{\lambda ij}$ in the Poisson component.

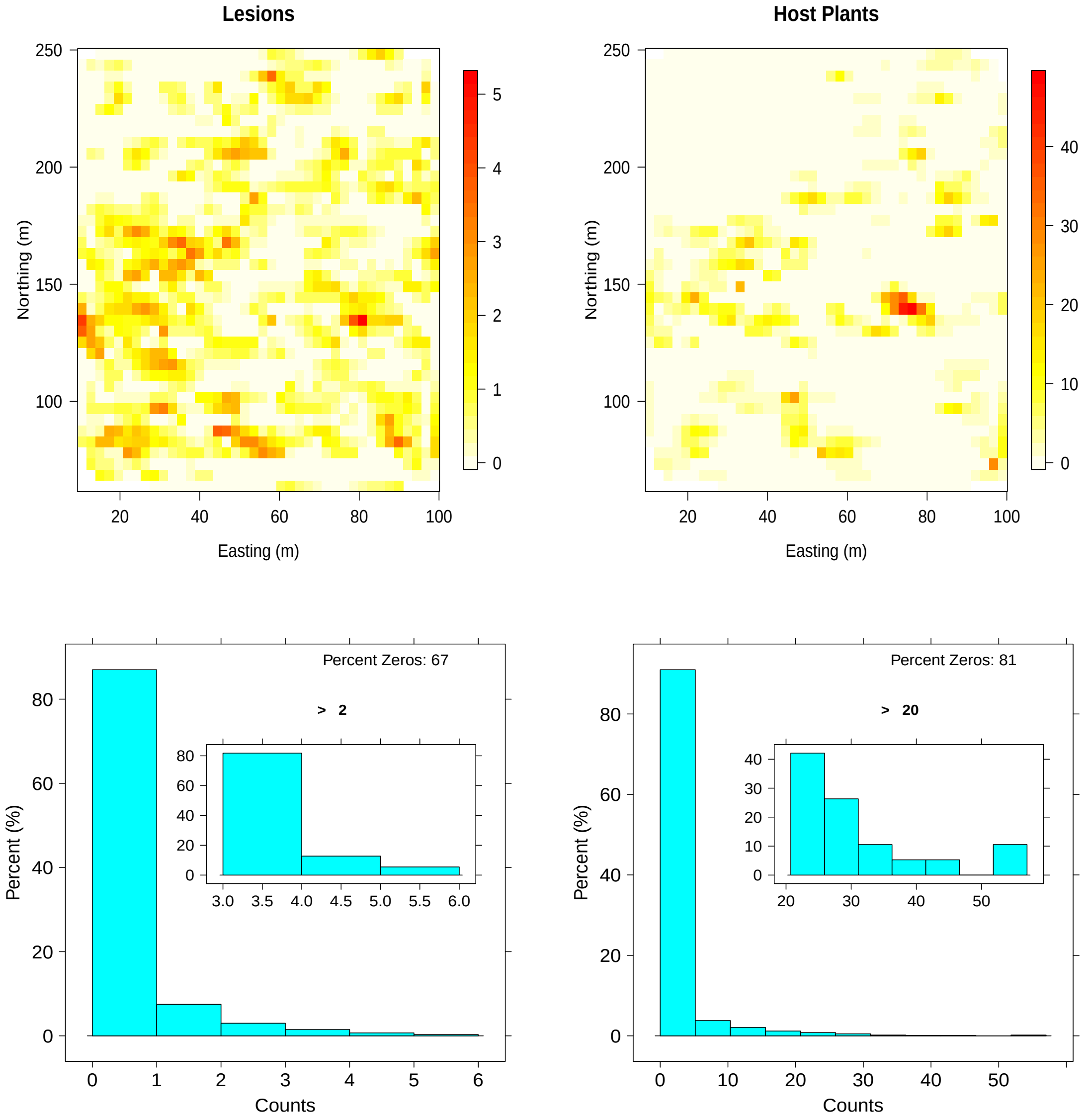


Figure 5.4: Level plots (top row) and histograms (bottom row) for lesion and host plant counts. Note that the Easting and Northing coordinates have been translated by a constant for the purpose of visualizing these data.

5.4 Misspecification of Spatial Structure

In this section, we evaluate two aspects of fit that arose from this work. The first investigates the relative bias and relative root mean square error when between-component spatial correlation is present in the data, but not accounted for in the model. We then explore robustness to the form of the covariance structure; this is applicable when a spatial covariance structure for lattice data is used as an approximation to a distance-based

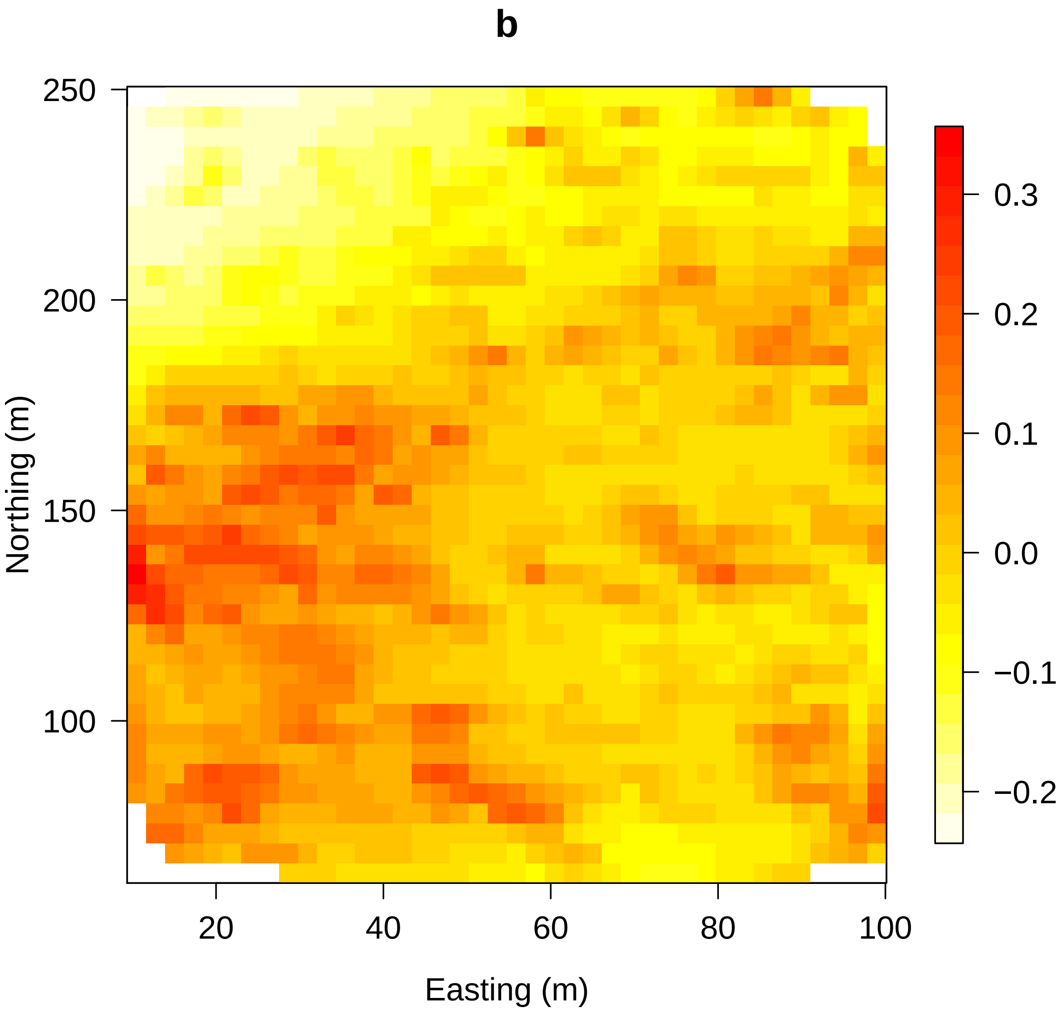


Figure 5.5: Linearly interpolated posterior median estimates of the shared random effect (***b***). Note that the Easting and Northing coordinates have been translated by a constant for the purpose of visualizing these data.

structure, as was done in Section 5.3.2.

5.4.1 Advantages of Including Between-Component Correlation

As mentioned in Section 5.2.1, with the interpretation of the random field as representing unobserved environmental factors related to health status or habitat suitability, incorporating correlation between this term and the mean of the Poisson component is intuitively appealing. However, this type of correlation is seldom accounted for. To investigate the

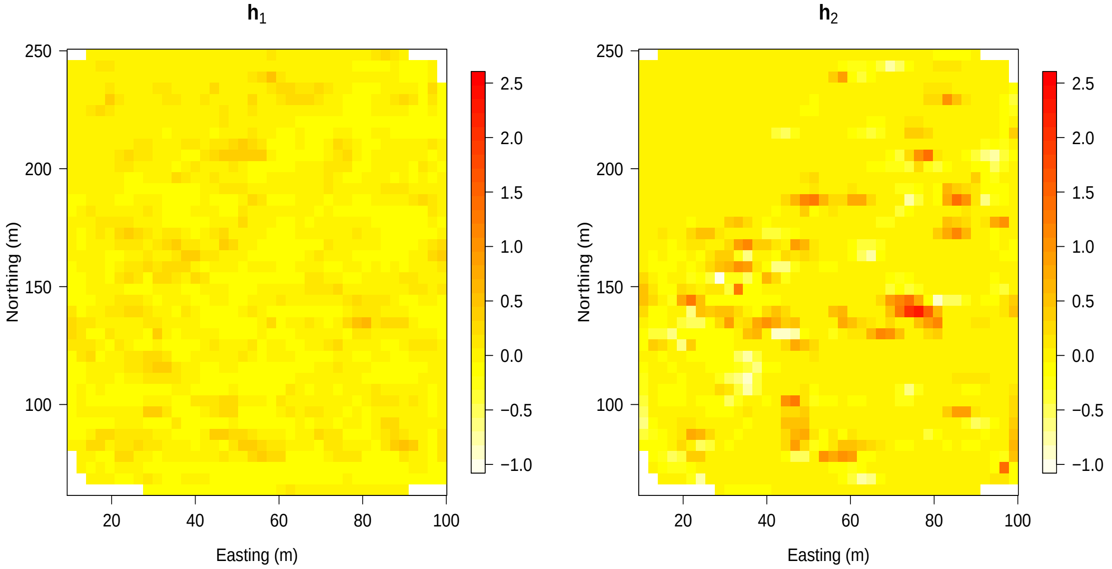


Figure 5.6: Linearly interpolated independent random effects for lesions and host plants. Note that the Easting and Northing coordinates have been translated by a constant for the purpose of visualizing these data.

advantages of incorporating between-component correlation, we compare results from fitting models J2, J3 and S1 with data generated from J2. Recall that model J2 assumes dependence across all components and outcomes through a single spatial random effect, model J3 only incorporates dependence across the zero-inflation components and across the Poisson components while model S1 assumes independent outcomes with dependence between components. In this simulation study, we investigate model fit under partial misspecification, as described above, while varying the ratio of spatial to unstructured variability, hereafter referred to as the variance ratio, corresponding to $(\xi_j/\sigma_j)^2$ and $(\nu_j/\sigma_{hj})^2$ in the zero-inflation and Poisson components, respectively. This is done based on the results of a simulation study by Feng and Dean (2012) who showed that an increase in this ratio results in increased power to detect the scaling parameters.

We consider a scenario similar to the Comandra blister rust data with $J = 2$ zero-heavy outcomes. Specifically, at the rth replication, we simulate $\left(b_1^{(r)}, b_2^{(r)}, \ldots, b_n^{(r)}\right)^T \sim MVN(\mathbf{0}, (\boldsymbol{D}-\boldsymbol{W})^{-1})$, where $\boldsymbol{W}$ is the neighbourhood matrix as defined in Section 5.3.2, and $\left(h_{1j}^{(r)}, h_{2j}^{(r)}, \ldots, h_{nj}^{(r)}\right)^T \sim N(\mathbf{0}, \sigma_{hj}^2 \boldsymbol{I}_n)$, $j = 1, 2$. We then simulate the zero-inflated Poisson random variables where the probability of being in the zero-inflation component is $\pi_{ij}^{(r)} = \Phi\left(\tau_j - \xi_j b_i^{(r)}\right)$ and with the mean of the Poisson component being $\lambda_{ij}^{(r)} = \exp\left\{\beta_j + \nu_j b_i^{(r)} + h_{ij}^{(r)}\right\}$, for $i = 1, \ldots, n$ and $r = 1, \ldots, R$, with $n = 500$ and $R = 100$. Here, we set $\boldsymbol{\tau} = (-0.5, 0.15)^T$, $\boldsymbol{\beta} = (3.5, 3)^T$ and $\boldsymbol{\sigma}_{\boldsymbol{h}}^2 = (0.1, 0.2)^T$. The variance ratios are set to four levels: 10, 7, 1 and 0.8. For each scenario, we fit models J2, J3 and S1 via Markov chain Monte Carlo using two chains with 125000 iterations including 25000 burn-in and retaining every 100th observation; this results in a total of 2000 samples from which inference is based. In the Markov chain Monte Carlo algorithm, we employ standard normal priors for the intercept parameters, and log-normal(0,1) and gamma(1,1) prior distributions on the scaling and independent normal precision parameters, respectively. Performance is assessed in terms of the relative bias and relative root mean square error (RMSE) of the scaling parameters. These quantities are calculated, respectively, as:

$$\text{relative bias}(\theta) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\theta}^{(r)} - \theta}{\theta}\right)$$

and

$$\text{relative RMSE}(\theta) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\frac{\left(\hat{\theta}^{(r)} - \theta\right)^2}{\theta^2}}$$

for a parameter θ where $\hat{\theta}$ is the posterior median estimate and θ is the true value.

The results from these simulations are summarized in Table 5.7. The intercept parameters from both components are well estimated under all model scenarios. Given that we are interested in the scaling and variance parameters, summaries of the intercepts have been omitted from the body of the thesis, but may be seen in Tables C.1 and C.2 in Appendix C.

All models tend to underestimate the spatial variability and overestimate the unstructured variability when the variance ratios are large (i.e. 7 and 10) and spatial variability dominates the map; this is not apparent when the variance ratios are small or moderate (i.e. 0.8 or 1). The relative biases in the zero-inflation components tend to be smaller (in absolute value) than in the Poisson components while the relative root mean square errors are consistently larger. For the large variances ratios, we see both a smaller relative bias, in absolute value, and a smaller relative root mean square error for the second outcome, where the magnitudes of the scaling parameters are large. This also holds true in all scenarios for the unstructured variance term.

In general, for models J3 and S1, when the variance ratios are large, the increase in the relative bias and relative root mean square error is larger than in the true models. However, this increase is more pronounced when the two outcomes are treated as independent, as was the case for model S1. For moderate or small variance ratios, the effect of partially misspecifying the joint spatial structure is not as severe.

Overall, both types of partial misspecification of the joint spatial structure have a more pronounced effect on the absolute relative bias and relative root mean square error when the map is dominated by spatial variability. However, this increase is more severe for the scaling parameters and unstructured variances when the outcomes are treated as independent (model S1) than when the components are treated as independent (model J3). This is logical as assuming independence between outcomes is essentially halving the sample size utilized to estimate the scaling parameters.

5.4.2 Effect of Misspecification of Spatial Structure

As mentioned previously, conditional autoregressive covariance structures may be employed for the analysis of spatial data when the neighbourhood structure and therefore the structure of the corresponding weights is not immediately obvious. For example, as described in Section 2.4.2, for the intrinsic conditional autoregressive structure, when employed for lattice data, the weights $w_{ii'} = 1$ if region i and i' are neighbours and 0 otherwise. If this is the desired structure for point-referenced spatial data, the standard definition of a neighbour no longer applies. In such instances, two observations will typically be considered neighbours if their Euclidean distance is less than or equal to some value. Alternatively, approaches such as the weights being inversely proportional to Euclidean distance may be employed (e.g. Earnest et al., 2007). In this section, we

		J2		**J3**		**S1**	
	True Value	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**
Variance Ratio: 10							
ξ_1	$\sqrt{10.00}$	-0.083	0.276	-0.052	0.360	-0.211	0.460
ξ_2	$\sqrt{10.00}$	-0.003	0.225	-0.046	0.267	-0.152	0.469
ν_1	$\sqrt{1.00}$	-0.141	0.212	-0.290	0.335	-0.304	0.363
ν_2	$\sqrt{2.00}$	-0.111	0.202	-0.275	0.326	-0.281	0.378
σ^2_{h1}	0.10	0.170	0.215	0.214	0.250	0.243	0.274
σ^2_{h2}	0.20	0.117	0.172	0.153	0.195	0.187	0.234
Variance Ratio: 7							
ξ_1	$\sqrt{7.00}$	-0.072	0.295	-0.105	0.335	-0.189	0.429
ξ_2	$\sqrt{7.00}$	-0.045	0.272	-0.084	0.333	-0.093	0.506
ν_1	$\sqrt{0.70}$	-0.179	0.265	-0.283	0.346	-0.290	0.365
ν_2	$\sqrt{1.40}$	-0.131	0.253	-0.265	0.340	-0.246	0.352
σ^2_{h1}	0.10	0.143	0.179	0.163	0.202	0.179	0.211
σ^2_{h2}	0.20	0.087	0.152	0.114	0.172	0.130	0.184
Variance Ratio: 1							
ξ_1	1.00	-0.056	0.397	-0.131	0.385	-0.112	0.421
ξ_2	1.00	-0.017	0.497	-0.031	0.505	-0.010	0.608
ν_1	$\sqrt{0.10}$	0.042	0.265	0.015	0.215	0.049	0.206
ν_2	$\sqrt{0.20}$	0.157	0.379	0.099	0.318	0.113	0.296
σ^2_{h1}	0.10	0.089	0.125	0.088	0.123	0.087	0.124
σ^2_{h2}	0.20	0.039	0.125	0.042	0.121	0.042	0.120
Variance Ratio: 0.8							
ξ_1	$\sqrt{0.80}$	0.045	0.491	-0.012	0.419	-0.007	0.404
ξ_2	$\sqrt{0.80}$	0.059	0.485	-0.004	0.403	0.015	0.467
ν_1	$\sqrt{0.08}$	0.202	0.377	0.154	0.329	0.201	0.368
ν_2	$\sqrt{0.16}$	0.184	0.389	0.144	0.327	0.189	0.333
σ^2_{h1}	0.10	0.082	0.127	0.082	0.126	0.080	0.124
σ^2_{h2}	0.20	0.046	0.121	0.047	0.121	0.044	0.117

Table 5.7: Relative bias (RBIAS) and relative root mean square error (RRMSE) for scaling parameters and variances from models J2, J3 and S1 at different levels of the variance ratios.

investigate the affect of misspecifying the spatial covariance structure in the conditional autoregressive random effect. To do this, we simulate joint zero-inflated Poisson models with conditional autoregressive random effects having adjacency matrices based on 10 m, 25 m, 30 m and an inverse distance-based scheme. All models are fit using the adjacency matrix with a Euclidean distance of 20 m. Therefore, for comparison, we also simulate from the "true" neighbourhood structure of 20 m. Performance is assessed in terms of relative bias and relative root mean square error of the parameters. Other than the conditional autoregressive formulation, the parameter settings are identical to that from Section 5.4.1.

The results for the scaling parameters and variances are summarized in Table 5.8. Note that the scenarios in which the neighbourhood structure is based on a Euclidean distance of 20 m are the same as in Table 5.7 under model J2. Regardless of the variance ratio, the relative bias and relative root mean square error for all scaling parameters based on a 10 m neighbourhood definition are considerably larger in comparison to the model simulated from the 20 m neighbourhood definition and we consistently overestimate all parameters. In the 25 m and 30 m scenarios, when the variance ratio is large, we tend to have increased estimates of the relative bias (in absolute value) and relative root mean square error. However, in such scenarios, this observation does not hold when the variance ratios are small and, in fact, in some instances a slight decrease in these quantities is observed. Finally, when the spatial covariance structure is generated from an inverse distance weighting scheme, as with the 10 m neighbourhood structure, we observe a positive increase in the relative bias and relative root mean square error, especially when the variance ratios are large. Although this increase is larger than when the data are simulated from the 25 m or 30 m neighbourhood structures, it is much smaller than the 10 m neighbourhood structure. In some instances we again observe a slight decrease in the relative bias and relative root mean square error for the small or moderate variance ratios. Overall, the variance parameters from the unstructured random effects tend to be well estimated and, as in the previous simulation, we have a smaller relative bias and relative root mean square error for the outcomes in which the true value has a larger magnitude.

Therefore, in situations where the variance ratio is large and the map is dominated by spatially structured variability, care must be taken to use an appropriate neighbourhood structure by, for example, examining a semivariogram or in-depth investigations into the scientific contexts. Further, contrasting model fits from several covariance structures would provide an indication as to the sensitivity of results to the assumed structure. This is important as we showed that using an incorrect structure has the potential to induce considerable bias and increased variance. However, this is not as great of a concern when the spatial and unstructured variability have similar magnitudes or when unstructured variability dominates.

5.5 Discussion

In this chapter, we presented a general modelling framework for multivariate zero-inflated spatial processes. The utility of this model was demonstrated on two data sets: counts of male and female lung and bronchus cancer incidence in Ontario as well as counts of lesions and host plants from an ecological study of Comandra blister rust infection of lodgepole pine trees in British Columbia. In these analyses, we showed that our framework allowed us to decompose the shared spatial structure across model components within multivariate outcomes and, through a comparison of model deviance information criterion, we have an informal approach to determine if the same spatial structure is shared across all outcomes and components. The use of a random field is an advancement over the current models in the literature (e.g. Feng and Dean, 2012; Rathbun and Fei, 2006) as it more closely follows the data generating process while incorporating correlation between the random field generating the counts and the observed mean count, and permits substantial model flexibility. This is important as it has the potential to aid investigators in understanding the spatial distribution of disease etiology and results in more flexibility to detect outcome-specific hotspots and lowspots, which are often of interest to investigators. We also showed how this approach provides a useful interpretation to the scaling parameters in the binary component as the ratio of the standard deviation of the spatial terms to that of the unstructured which does not lend itself to the standard

	Ratio: 10		**Ratio: 7**		**Ratio: 1**		**Ratio: 0.8**	
	RBIAS	**RRMSE**	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**
Neighbourhood Structure: $\leq$ 20 m								
ξ_1	-0.083	0.276	-0.072	0.295	-0.056	0.397	0.045	0.491
ξ_2	-0.003	0.225	-0.045	0.272	-0.017	0.497	0.059	0.485
ν_1	-0.141	0.212	-0.179	0.265	0.042	0.265	0.202	0.377
ν_2	-0.111	0.202	-0.131	0.253	0.157	0.379	0.184	0.389
σ^2_{h1}	0.170	0.215	0.143	0.179	0.089	0.125	0.082	0.127
σ^2_{h2}	0.117	0.172	0.087	0.152	0.039	0.125	0.046	0.121
Neighbourhood Structure: $\leq$ 10 m								
ξ_1	1.497	1.539	1.678	1.724	2.084	2.178	2.144	2.262
ξ_2	1.706	1.771	1.733	1.800	2.040	2.139	2.063	2.197
ν_1	1.509	1.527	1.524	1.551	1.781	1.847	1.795	1.878
ν_2	1.628	1.647	1.621	1.652	1.853	1.945	1.856	1.988
σ^2_{h1}	0.172	0.212	0.185	0.226	0.095	0.164	0.088	0.143
σ^2_{h2}	0.030	0.152	0.059	0.164	0.041	0.153	0.061	0.153
Neighbourhood Structure: $\leq$ 25 m								
ξ_1	-0.297	0.388	-0.368	0.449	-0.162	0.394	-0.123	0.473
ξ_2	-0.285	0.371	-0.301	0.430	-0.181	0.408	-0.096	0.461
ν_1	-0.377	0.406	-0.429	0.468	-0.017	0.213	0.025	0.203
ν_2	-0.333	0.383	-0.350	0.405	0.067	0.309	0.127	0.398
σ^2_{h1}	0.168	0.209	0.157	0.191	0.085	0.129	0.073	0.112
σ^2_{h2}	0.104	0.184	0.082	0.165	0.033	0.114	0.043	0.135
Neighbourhood Structure: $\leq$ 30 m								
ξ_1	-0.507	0.558	-0.522	0.566	-0.205	0.423	-0.107	0.457
ξ_2	-0.447	0.500	-0.465	0.545	-0.207	0.430	-0.153	0.450
ν_1	-0.492	0.507	-0.505	0.529	-0.073	0.207	-0.037	0.189
ν_2	-0.494	0.516	-0.484	0.518	-0.023	0.281	0.104	0.272
σ^2_{h1}	0.150	0.179	0.117	0.149	0.070	0.120	0.080	0.129
σ^2_{h2}	0.103	0.151	0.076	0.139	0.024	0.113	0.017	0.118
Neighbourhood Structure: Inverse Distance								
ξ_1	0.524	0.634	0.439	0.549	0.033	0.596	0.066	0.547
ξ_2	0.817	0.947	0.744	0.841	0.086	0.548	0.238	0.726
ν_1	0.380	0.428	0.314	0.391	0.106	0.306	0.168	0.334
ν_2	0.503	0.571	0.409	0.508	0.224	0.479	0.177	0.364
σ^2_{h1}	0.241	0.271	0.246	0.278	0.130	0.161	0.110	0.158
σ^2_{h2}	0.101	0.167	0.158	0.224	0.081	0.159	0.074	0.142

Table 5.8: Relative bias (RBIAS) and relative root mean square error (RRMSE) for scaling parameters and variances from models J2, J3 and S1 at different levels of the variance ratios when varying the neighbourhood structure.

logistic framework. Although we applied this to two specific zero-heavy count data sets, we emphasize that this model is broadly applicable. For example, it could be applied to multivariate outcomes not necessarily following the same distribution, such as understanding linkages between environmental outcomes and related health effects including forest fire smoke and respiratory problems (e.g. Wan et al., 2011).

Through simulation, we investigated the effects of misspecifying the spatial structure in two ways: first by partially misspecifying the shared structure across outcomes and components and second by misspecifying the adjacency matrix in the conditional autoregressive random effects. Results from the former suggest that when maps are dominated by spatial variability, partially misspecifying the shared structure (e.g. models J3 and S1) results in increased relative bias and relative root mean square error in the scaling parameters. Note that larger relative biases and relative root mean square errors were seen when excluding between-outcome correlation (model S1) than when between-component correlation (model J3) was excluded as this was essentially halving the number of observations used in estimating the scaling parameters. Similarly for the second simulation study, when the ratio of spatial to unstructured variability was large, we saw a considerable increase in the relative bias and relative root mean square error. These studies suggest that care must be taken to correctly identify the joint spatial and neighbourhood structures for these types of models. This is especially important when spatial variability dominates as results are typically more robust when the variance ratio is small or moderate.

Where rates change over time, it would be important to incorporate temporal trends; examples where the threshold depends on time may be rare. Though not considered here, tests for simpler structures could also be developed based on this framework. For example, it would be interesting to test whether the same spatial random field is generating the multivariate outcomes.

Chapter 6

Future Work

In this thesis, we developed new frameworks and inference for the joint modelling of spatial and spatio-temporal point processes. We began by developing a spatio-temporal hierarchical cluster process for storm cell data by generalizing the Neyman-Scott process and allowing the parents to follow a log-Gaussian Cox process. Not only did this account for the different levels of clustering in these data, but it incorporated correlation at each level. Joint modelling of storm cell detection (point process) and storm cell evolution (multivariate mark process) is a challenging problem and we addressed this by modelling the marks conditional on the events. This enabled the use of methods for point-referenced data to account for correlation in the duration, speed and direction of these trajectories. We also demonstrated joint modelling for multivariate zero-inflated count data arising as a result of aggregated spatial point patterns. In this project, we incorporated correlation between the Gaussian process assumed to generate events and mean event counts in a flexible framework. These developments suggest many avenues for future work and we conclude this thesis by detailing three such extensions.

6.1 A Joint Log-Gaussian Cox Process for Multivariate Aggregated Point Patterns

For a zero-inflated Poisson model, data are assumed to arise as a mixture of a point mass at zero and a Poisson component with component membership for observations being unobserved. The interpretation of such a model is that random variables in the zero-inflation component, referred to as structural zeros, arise from a "perfect" state in which positive counts cannot be observed. For counts of cancer cases aggregated by public health unit, the assumption of an entire region being immune and hence generating a structural zero is unrealistic. A more reasonable model would be a two-component Poisson mixture with the first component having a lower mean than the second, representing regions with fewer environmental carcinogens, for example. However, with randomly rounded public health data, the zero-inflation component serves as an approximation the lower mean component. Future work could build on the use of a log-Gaussian Cox process for these types of data. Li et al. (2012) utilized a log-Gaussian Cox process to model an unobserved point process conditional on aggregated count data using a Gibbs sampling Markov chain Monte Carlo algorithm with a data augmentation step. Extending this idea to the joint modelling framework is of interest. Following this, we would no longer have the zero-inflated Poisson model interpretation of the structural and random zeros.

6.2 A Composite Likelihood Approach to Parameter Estimation of a Spatio-Temporal Point Process

As discussed in this thesis, parameter estimation for spatio-temporal point processes is a challenging task and one which is not yet well developed. However, considerably more effort has been devoted to parameter estimation for spatial point patterns. Guan (2006), for example, developed a composite likelihood approach by deriving a valid density at locations x_1 and x_2 as follows:

$$f(x_1, x_2; \boldsymbol{\theta}) = \frac{\lambda^{(2)}(x_2 - x_1; \boldsymbol{\theta})}{\int \int_S \lambda^{(2)}(u - v; \boldsymbol{\theta}) \mathrm{d}u \mathrm{d}v}$$

with $\boldsymbol{\theta}$ representing the vector of parameters and S begin the spatial observation window. The corresponding composite log-likelihood may then be expressed as

$$\ell(x_1, x_2; \boldsymbol{\theta}) = \log \left[\lambda^{(2)}(x_2 - x_1; \boldsymbol{\theta})\right] - \log \left[\int \int_S \lambda^{(2)}(u - v; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{u} dv\right].$$

Guan (2006) also proved that under mild conditions, this yields consistent and asymptotically normal estimators.

Extensions to spatio-temporal point processes may be considered by defining the density and composite log-likelihood analogously to the above with $\lambda^{(2)}(x_1, x_2; \boldsymbol{\theta})$ being the second-order intensity of a spatio-temporal point process and x_1 and x_2 being events in space-time. For hierarchically clustered point patterns, such as our storm cell data, to properly identify small- and large-scale clustering, the nearest neighbour distance property may be employed, as was in done in Tanaka and Ogata (2014). Adding the additional temporal dimension may pose challenges in implementing inferential techniques. However, this is likely a feasibly approach and could be explored as an alternative to methods developed herein.

6.3 Joint Estimation for Spatio-Temporal Point Processes with Evolving Marks

Developing methods to jointly estimate the parameters in a hierarchically clustered point process with multivariate evolving marks is of interest. Models for spatio-temporal point processes with temporally evolving marks were developed and refined in Renshaw and Särkkä (2001) and Särkkä and Renshaw (2006) for forestry data related to tree location and size. In the proposed process, trees or "immigrants" arrived stochastically following a Poisson process with rate α, according to an immigration-death process with uniformly distributed locations, and died via a death process with a constant probability μ. The marks were then assumed to evolve according to a deterministic growth function which, at time t, depended on size at time $t-1$, an observation-specific growth function and a spatial interaction function. This growth function was incorporated into a Gibbs process, characterized by the density $f(\psi) = \frac{1}{V}\exp\{-U(\psi)\}$ with V being the normalizing constant and $U(\cdot)$ representing the energy function. Defined in terms of pairwise interactions, $U(\cdot)$ is a function of the event locations and marks.

This type of process is both novel and interesting and could be extended to jointly model storm cells and their corresponding trajectories, as follows: storm cells might arrive occur according to a cluster process, rather than the Poisson process, to account for the spatio-temporal clustering inherent in these data. Rather than having a single mark, we would have multivariate marks for speed and direction, modelled autoregressively in space and time with wind speed and direction included as covariates. We could also assume a mixture distribution for the death process, where storm cells are assumed to either die shortly after their arrival or have a duration following a time-to-event distribution, such as the log-normal. This sophisticated analysis may pose challenges for estimation and inference, but would be based on a conceptually appealing modelling framework.

Appendix A

Supplementary Material for Chapter 3

A.1 Storm Cell Identification and Tracking Algorithm

Storm modelling is done via the smallest and only detectable unit of a storm producing system, termed a storm cell. At the Bismarck, North Dakota radar station, storm cell data are collected by a Doppler radar called Weather Surveillance Radar-1988. In particular, a storm cell identification and tracking algorithm is employed for identifying storm cells and tracking their trajectories. Johnson et al. (1998) contains a thorough description of this algorithm; what follows is a brief summary highlighting the key aspects as they relate to this thesis.

In order to identify storm cells, Doppler radar uses reflectivity, measured in decibels (dBZ). This is defined by the National Oceanic and Atmospheric Administration as the amount of transmitted power returned to the radar after hitting precipitation. The storm cell identification and tracking algorithm starts by identifying one-dimensional segments, which are runs of reflectivities above 30 dBZ along a radial with $1^\circ \times$ 0.54 nautical mile bins. At the 30 dBZ threshold, contiguous bins having reflectivity at or above this value are grouped together until a smaller reflectivity is encountered; if this value is no more than 5 dBZ below the threshold for a maximum of two adjoining bins the process continues. When either a value more than 5 dBZ below the threshold or more than two bins having reflectivities within 5 dBZ below the threshold are encountered, this process is terminated. A storm segment is saved if its length is greater than 1.9 km. This is repeated for reflectivities between 30 dBZ and 60 dBZ in increments of five. Segments defined by the various thresholds may overlap, but only those with the largest reflectivities are kept.

Two dimensional components are then constructed by merging one-dimensional segments based on spatial proximity, a process referred to as "horizontal association". Segments are combined if they are within 1.5° of each other azimuthally and overlap in range by 2 km. A component must be composed of at least two segments and have an area larger than 10 km^2. If the centre of a component of higher-reflectivity falls within an area of lower-reflectivity, the component at the lower threshold is discarded.

Lastly, to identify three-dimensional storm cells, a vertical association is performed. This is an iterative process beginning at the lowest elevation angle: associations are made between components on consecutive elevation scans whose centroids are within 5 km of one another horizontally. In instances where more than one association is possible, it is made with the two-dimensional component having the largest mass. For the non-associated components, the search radius is increased to 7.5 km and then finally 10 km. A cell must consist of at least two components.

Using centroid locations from the previous scan, a storm cell's current location is predicted based on a linear least squares fit of its speed and direction at up to eleven previous scans, if available. For cells which were identified at the previous scan, a location is calculated based on a default projected velocity, either from an average of all velocities at the previous scan or based on user input. Storm cells are ranked by their intensity as measured by their vertical integrated liquid and starting with the most intense, its centroid is compared to all predicted centroid positions based on the previous scan. The storm cell at the current scan which is closest (within a threshold) to its projected location is considered to be the same observation and given the same identifier. This tracking algorithm is then employed for all remaining storm cells in order of decreasing vertical integrated liquid. If no projected centroids are located within a threshold range, a storm cell is given a new unique identifier. Temporal associations are not considered if more than 20 minutes has elapsed between consecutive scans. For further details related to storm cells, please see Mohee and Miller (2010) who developed a climatology of North Dakota thunderstorms between 2002 and 2006 at three radar stations.

A.2 Verification of Campbell's Theorem

Theorem A.2.1. *Suppose $g(y)$ and $h(y)$ are probability density functions, with $h(y)$ bounded, and $\Lambda(y)$ is a random nonnegative function with bounded expectation. If, conditional on Λ, $\mathcal{Y}$ is an inhomogeneous Poisson process with intensity $\Lambda(y)$, then*

1. $$E\left[\sum_{j:y_j\in\mathcal{Y}} g(y_j)|\Lambda\right] = \int g(y)\Lambda(y)dy, \text{ w.p. 1.} \tag{A.1}$$

2. $$E\left[\sum_{j:y_j\in\mathcal{Y}} g(y_j)h(y_j)|\Lambda\right] = \int g(y)h(y)\Lambda(y)dy, \text{ w.p. 1.} \tag{A.2}$$

Proof. The conclusion of this theorem follows from an application of Campbell's theorem (Daley and Vere-Jones, 2003), applied to the elements of the probability space on which $\Lambda(y)$ is defined, subject to verification that the following condition holds, with probability 1:

$$\int \min(|g(y)|, 1)\Lambda(y)dy < \infty. \tag{A.3}$$

Indeed, if we take the expectation of the left hand side of (A.3), apply Tonelli's theorem (Jacod and Protter, 2000), which holds, since the integrand is nonnegative, note that

$$\min(|g(y)|, 1) \leq g(y)$$

for all y, and observe that $E[\Lambda(y)] \leq L$, for some positive constant L, we have

$$E\left[\int \min(|g(y)|, 1)\Lambda(y)dy\right] \leq \int g(y)E[\Lambda(y)]dy \leq L\int g(y)dy = L < \infty.$$

The result at (A.1) now follows.

Furthermore, boundedness of $h(y)$ by B, say, implies that

$$E\left[\int \min(|g(y)h(y)|, 1)\Lambda(y)dy\right] \leq BE\left[\int \min(|g(y)|, 1)\Lambda(y)dy\right] \leq LB < \infty$$

from which we deduce (A.2). □

Appendix B

Supplementary Material for Chapter 4

These are a sequence of figures representing the partial effects for the models developed in Chapter 4. Although these figures are not of great importance to our discussion of the results, we have included them for completeness.

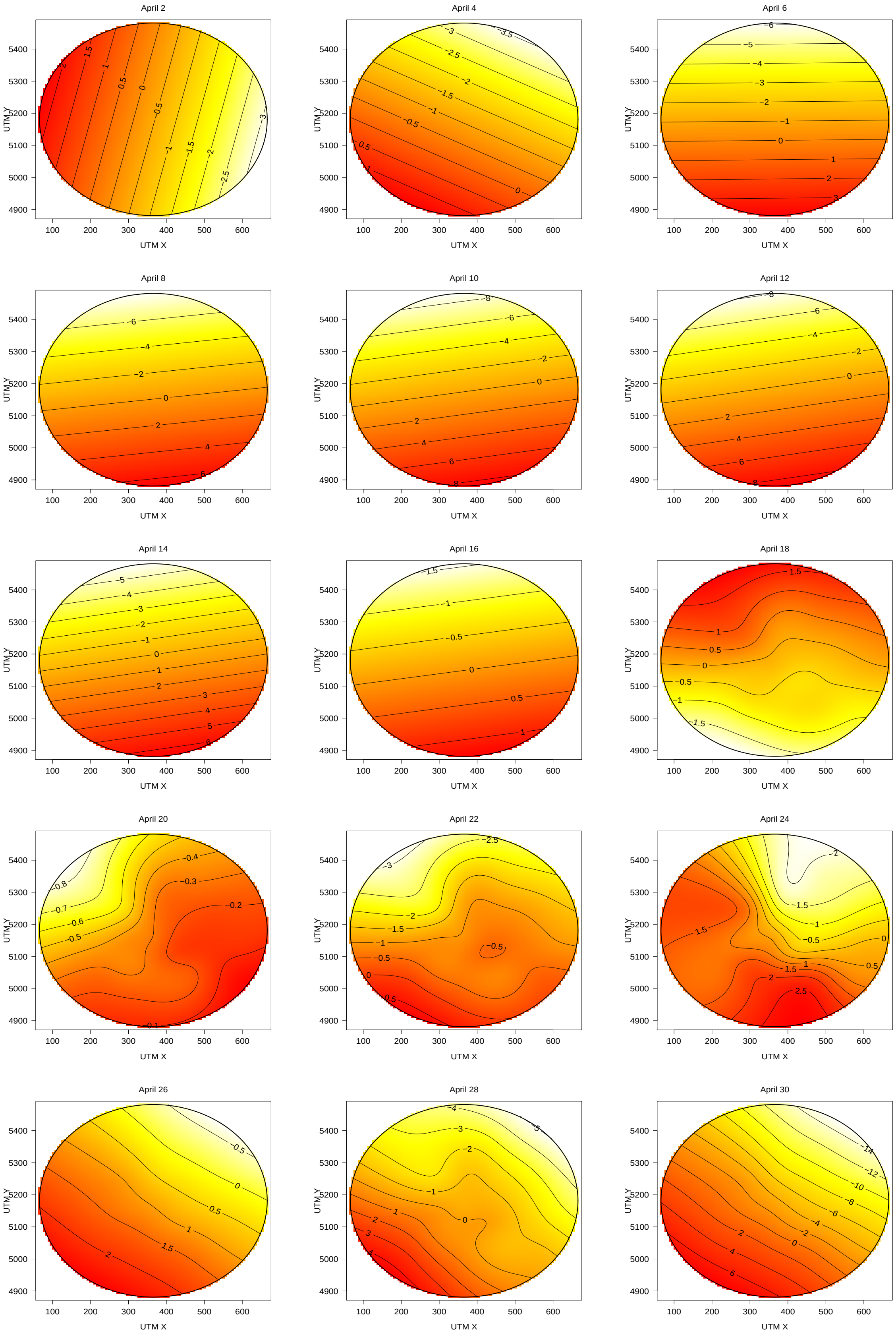


Figure B.1: Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from April 2003.

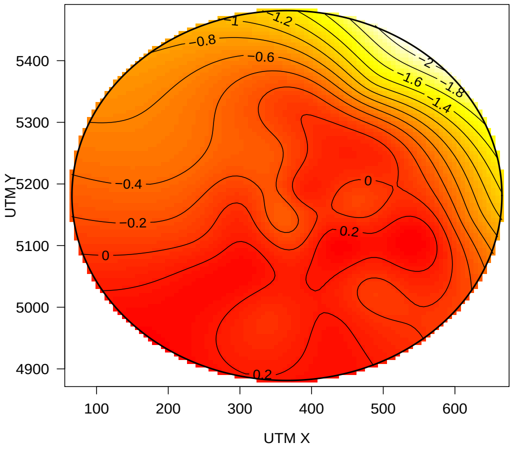
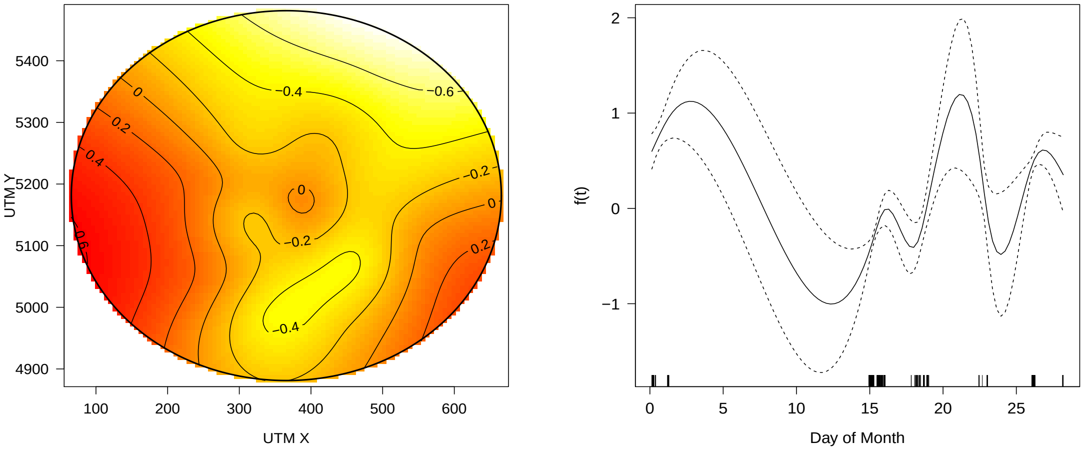


Figure B.2: Estimated spatial partial effect for the scale parameter in the duration component from April 2003.



(a) Spatial partial effect.

(b) Temporal partial effect.

Figure B.3: Estimated spatial and temporal partial effects for the location parameter in the speed component from April 2003.

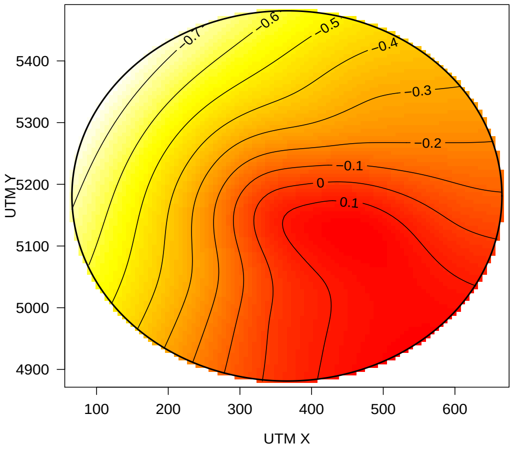


Figure B.4: Estimated spatial partial effect for the scale parameter in the speed component from April 2003.

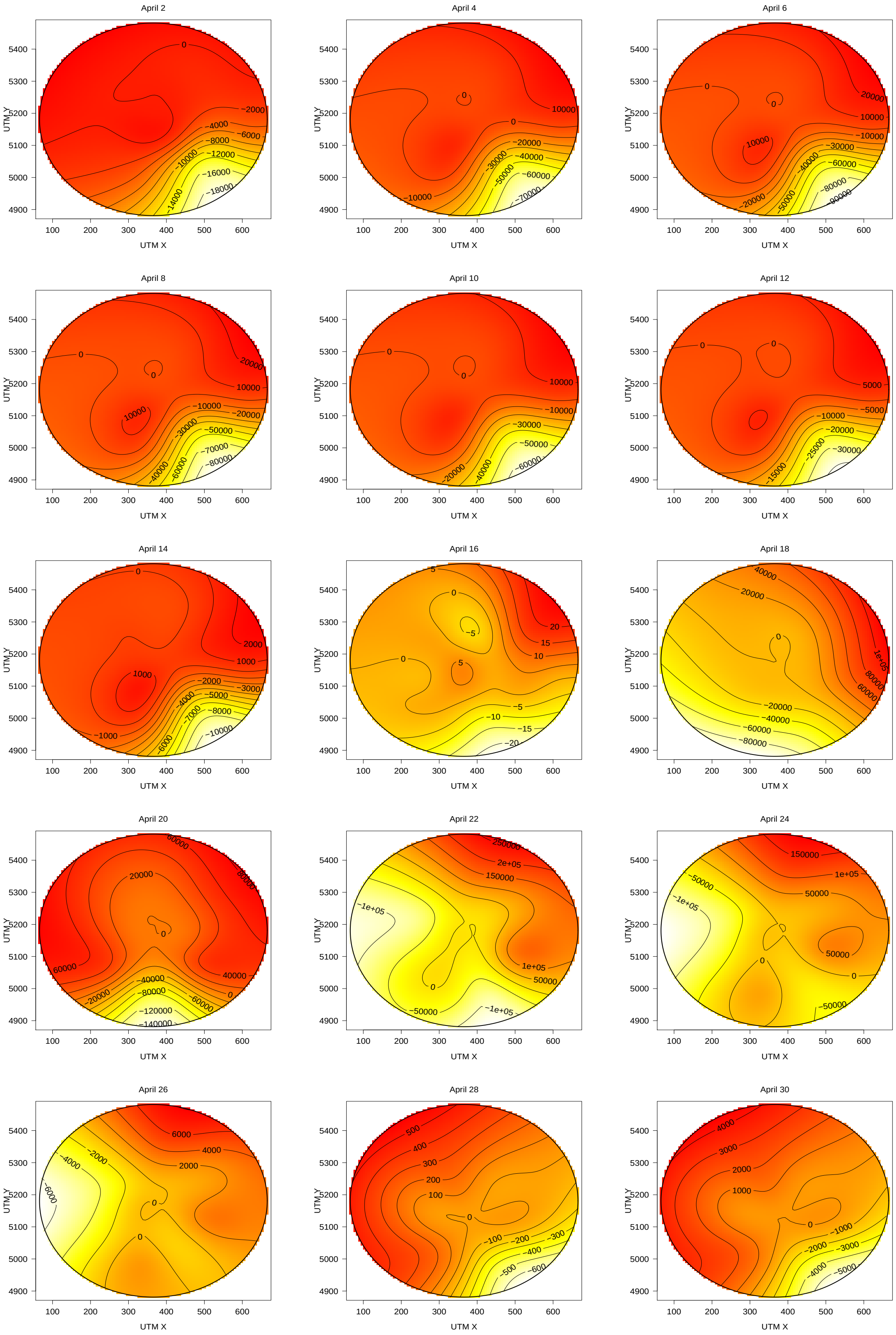

April 2
April 4
April 6
April 8
April 10
April 12
April 14
April 16
April 18
April 20
April 22
April 24
April 26
April 28
April 30

Figure B.5: Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from April 2003.

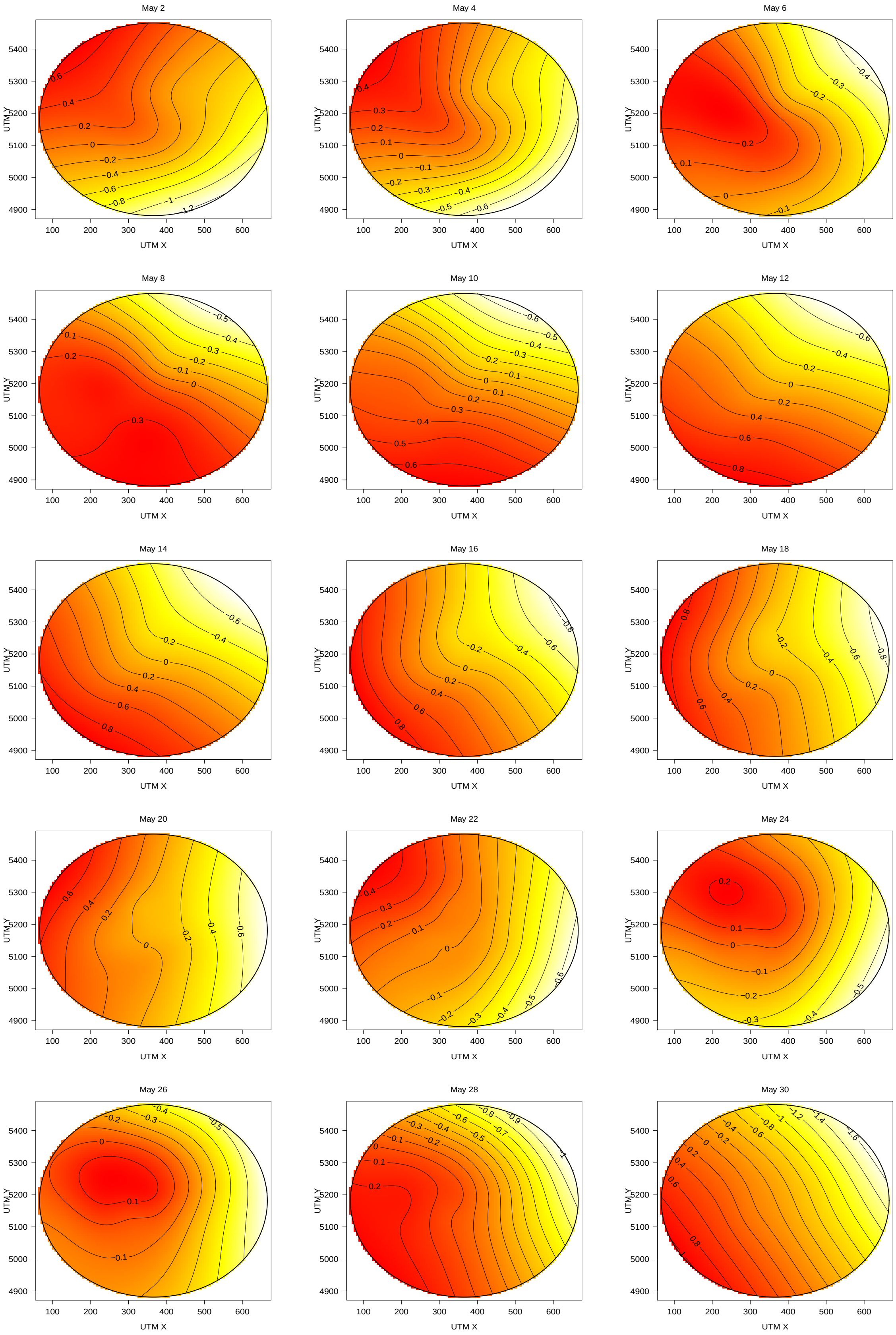


Figure B.6: Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from May 2003.

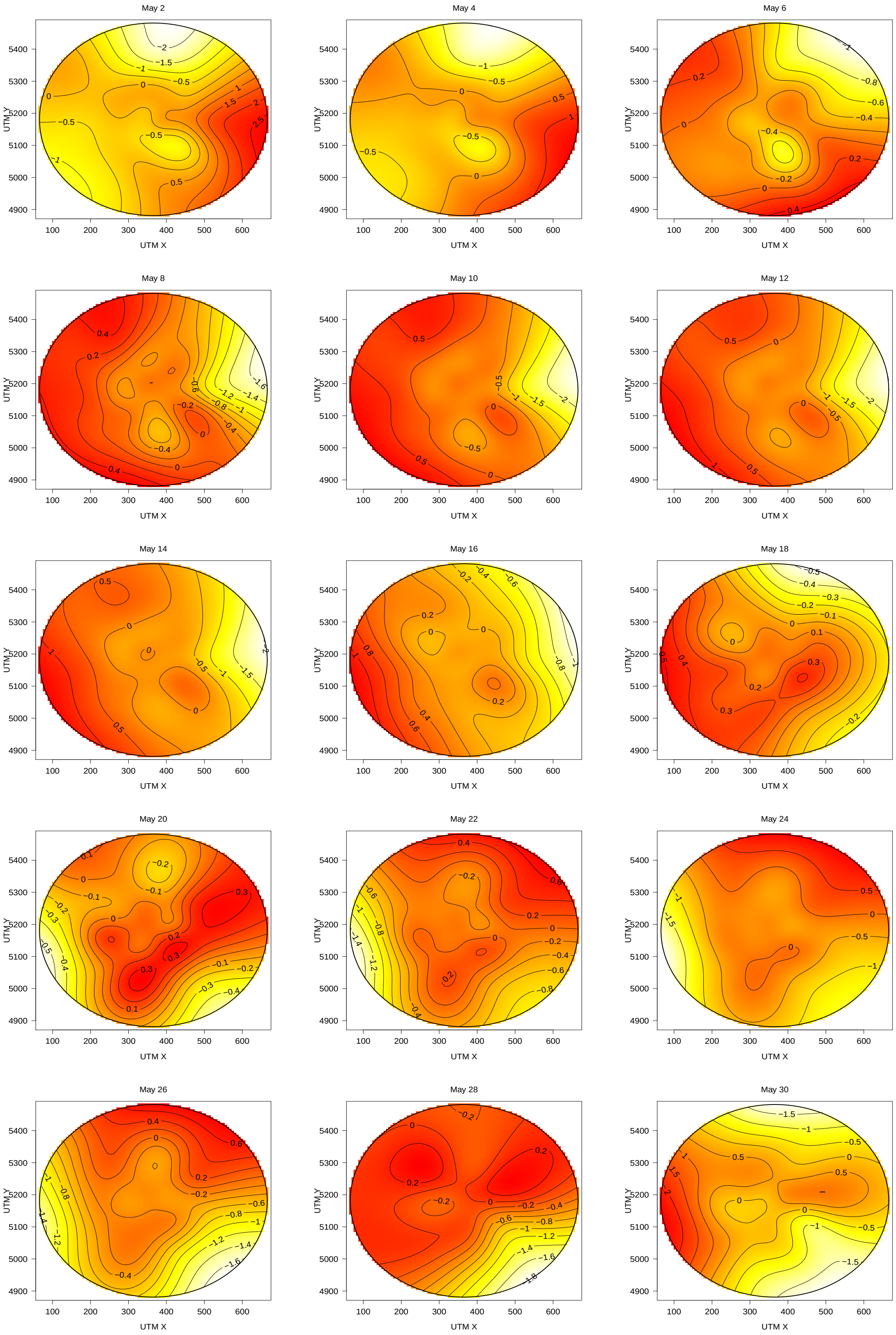


Figure B.7: Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from May 2003.

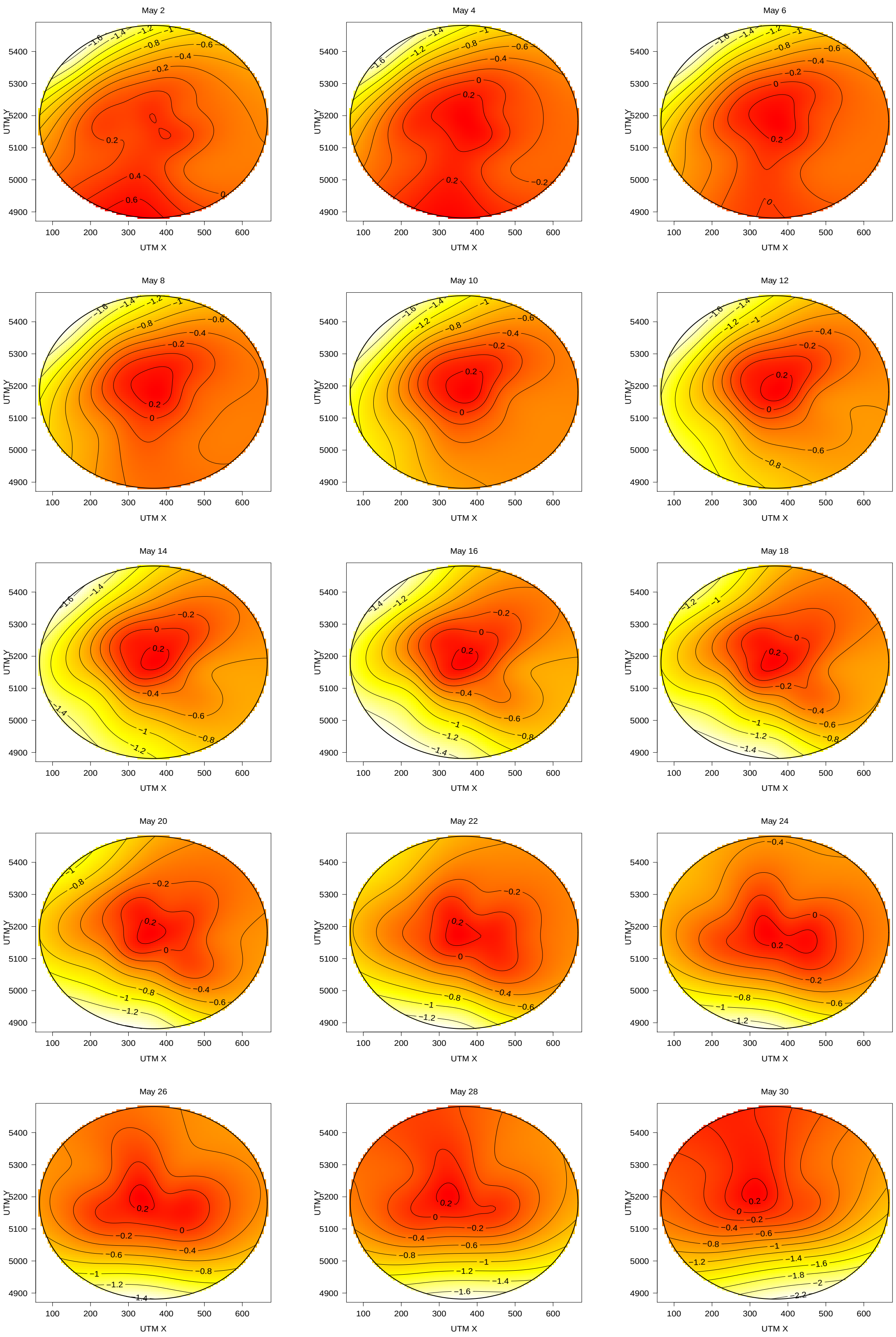

May 2
May 4
May 6
May 8
May 10
May 12
May 14
May 16
May 18
May 20
May 22
May 24
May 26
May 28
May 30

Figure B.8: Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from May 2003.

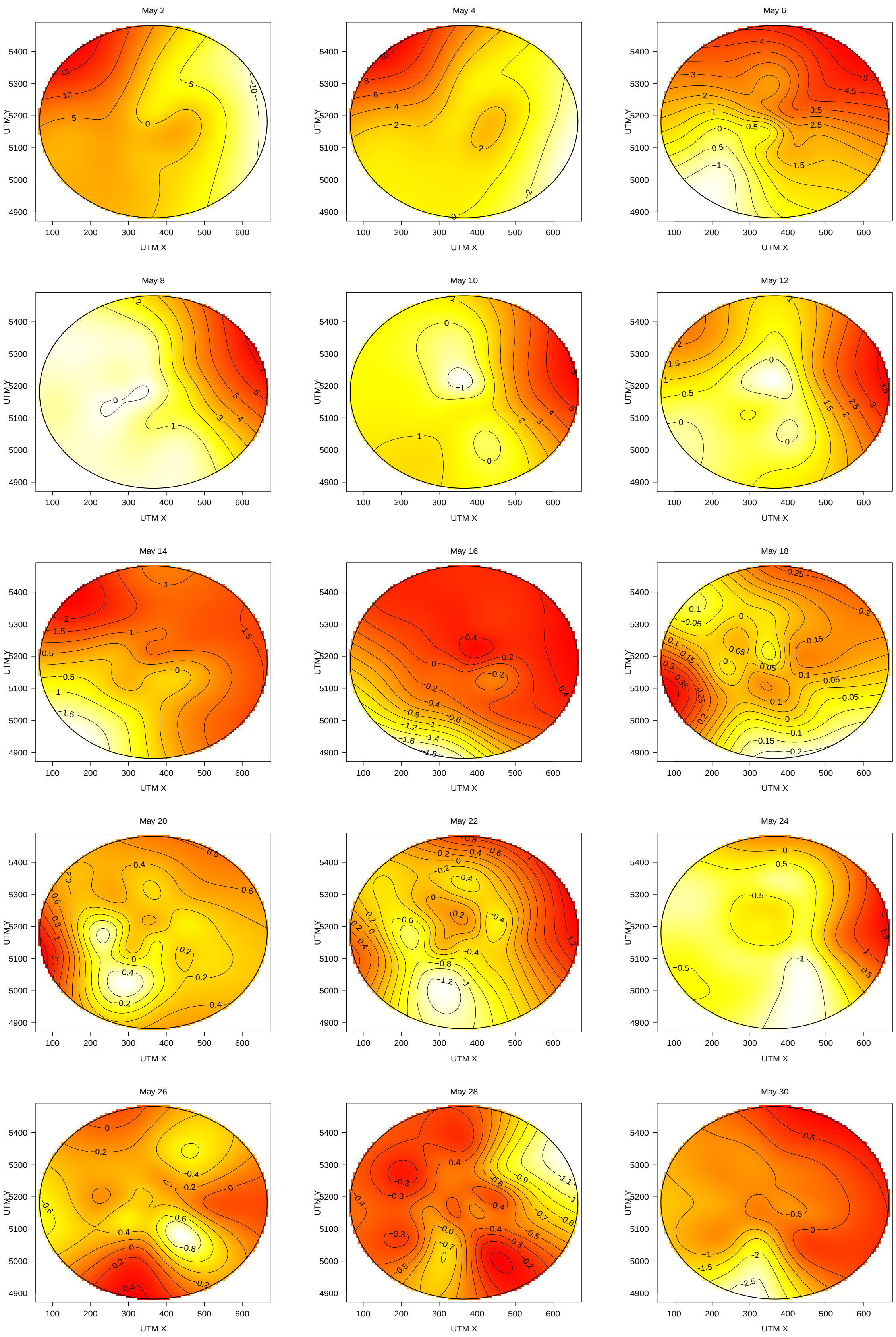


Figure B.9: Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from May 2003.

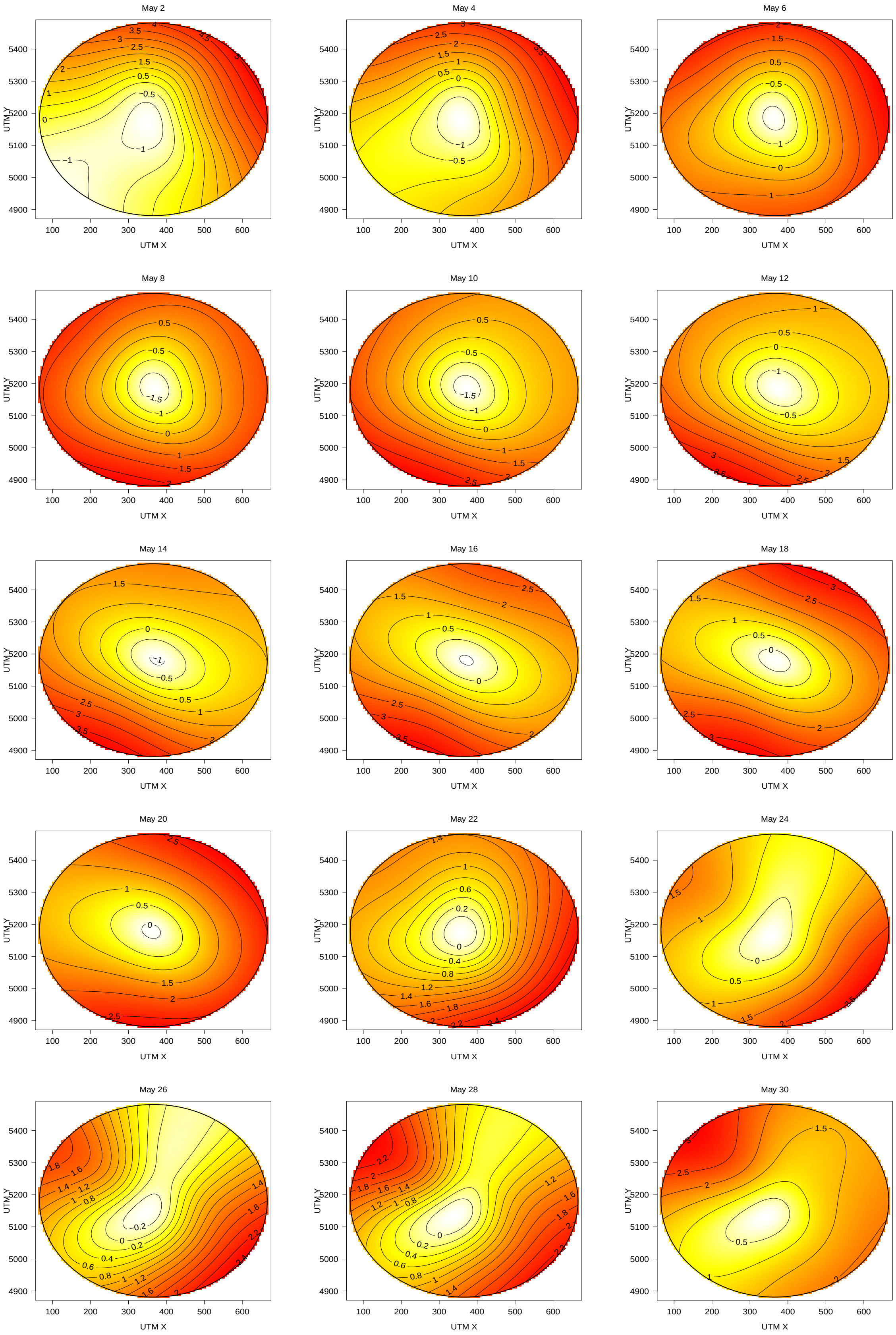
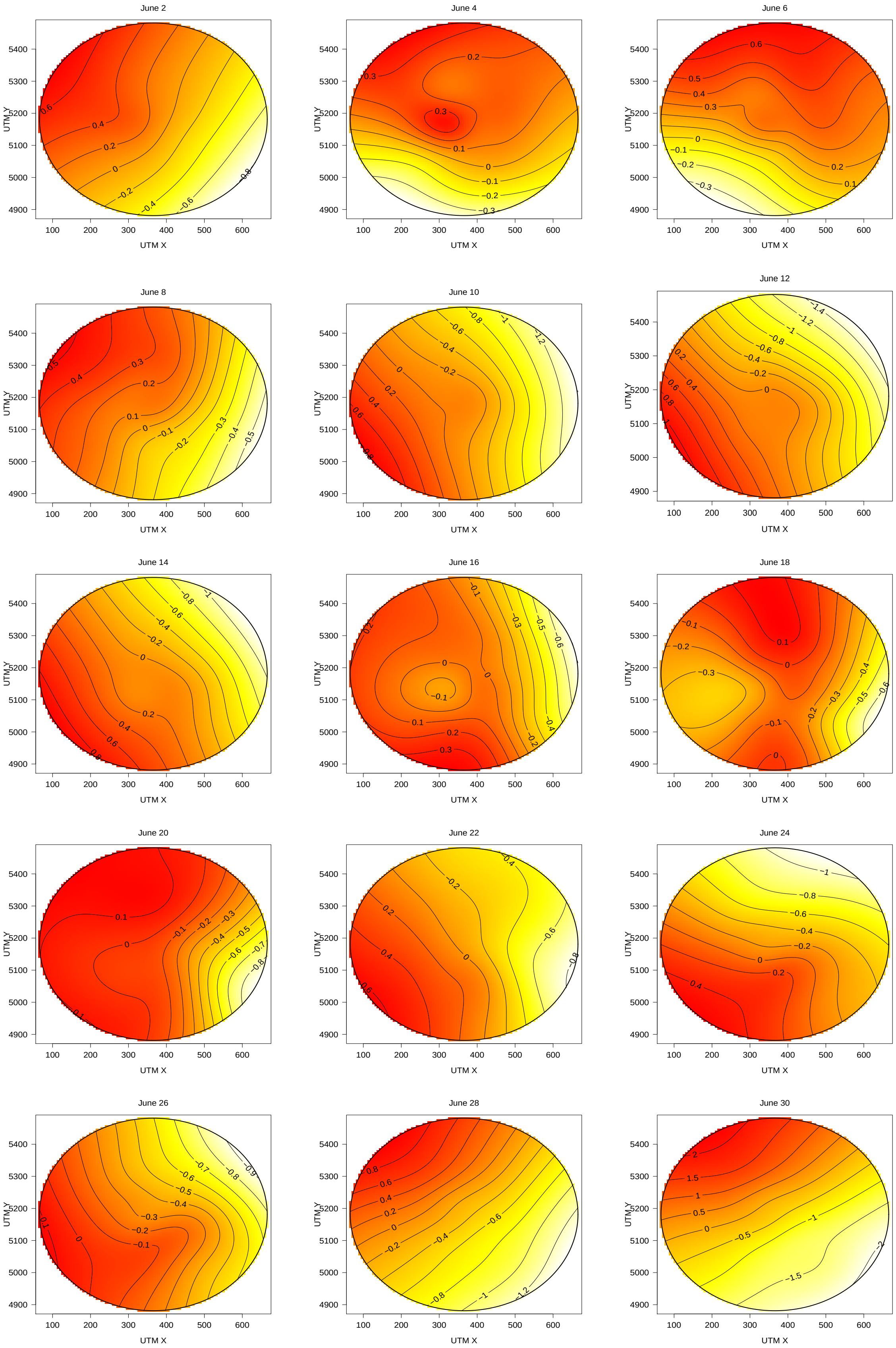

May 2 May 4 May 6 May 8 May 10 May 12 May 14 May 16 May 18 May 20 May 22 May 24 May 26 May 28 May 30

Figure B.10: Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from May 2003.


June 2

June 4

June 6

June 8

June 10

June 12

June 14

June 16

June 18

June 20

June 22

June 24

June 26

June 28

June 30

Figure B.11: Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from June 2003.

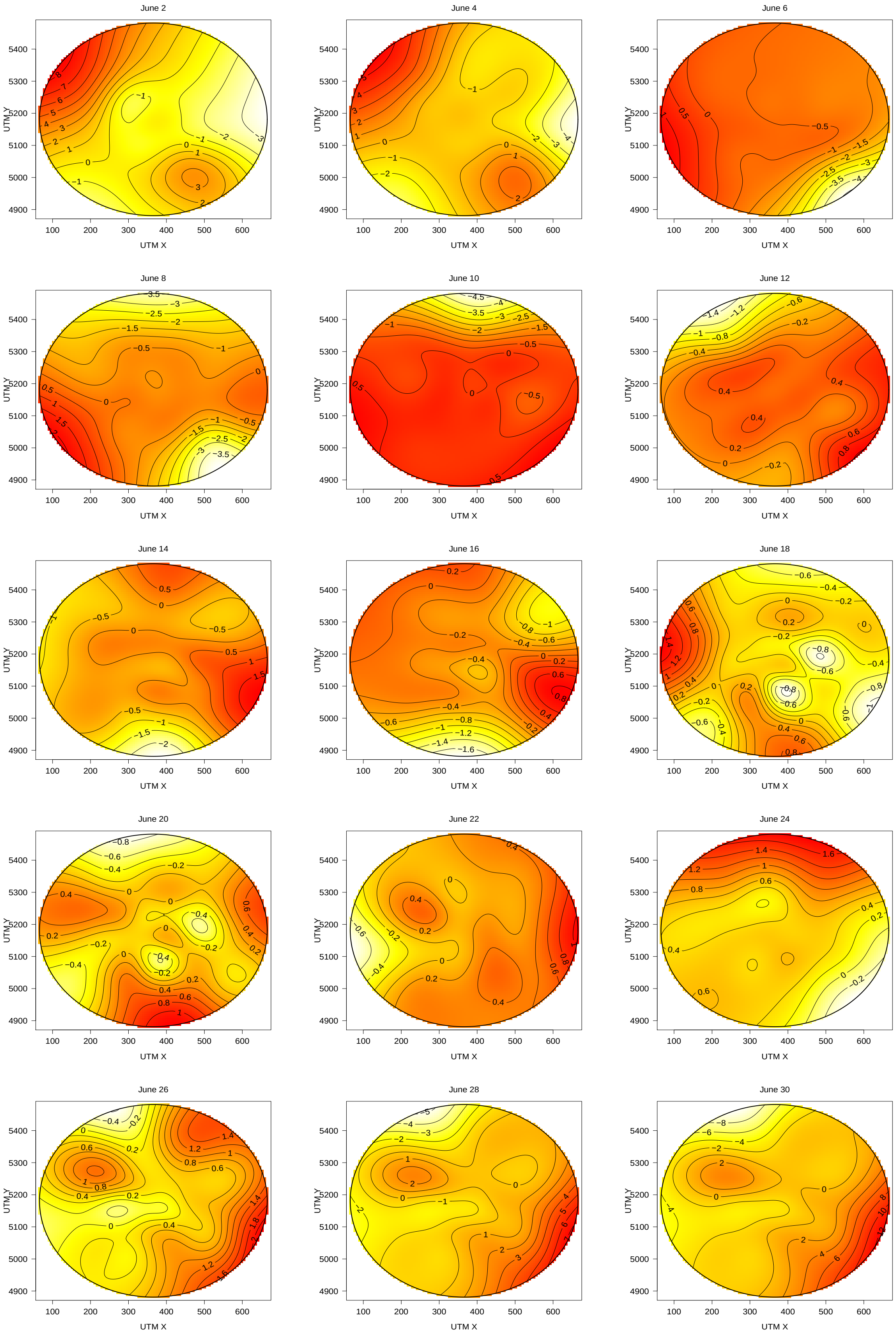


Figure B.12: Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from June 2003.

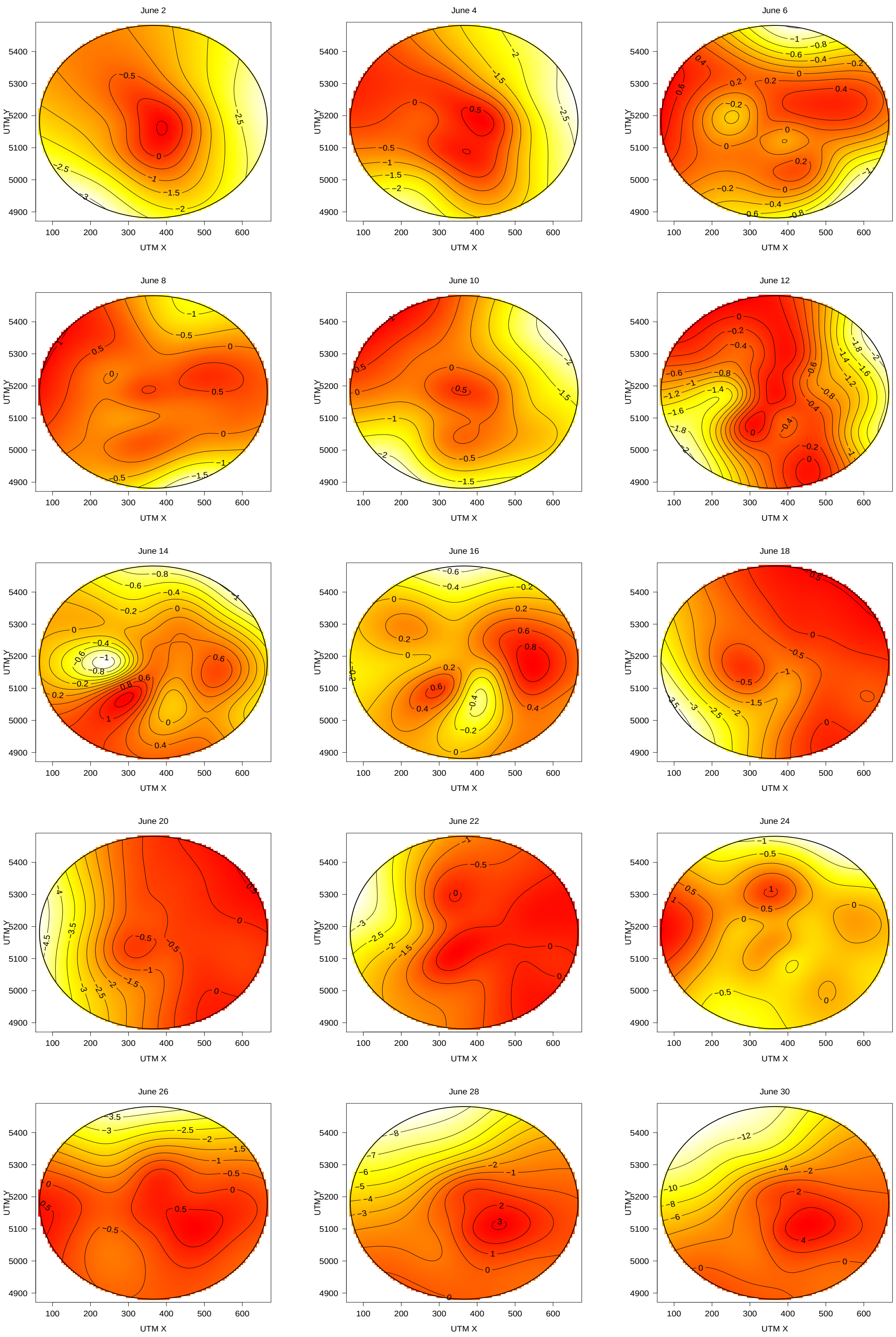


Figure B.13: Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from June 2003.

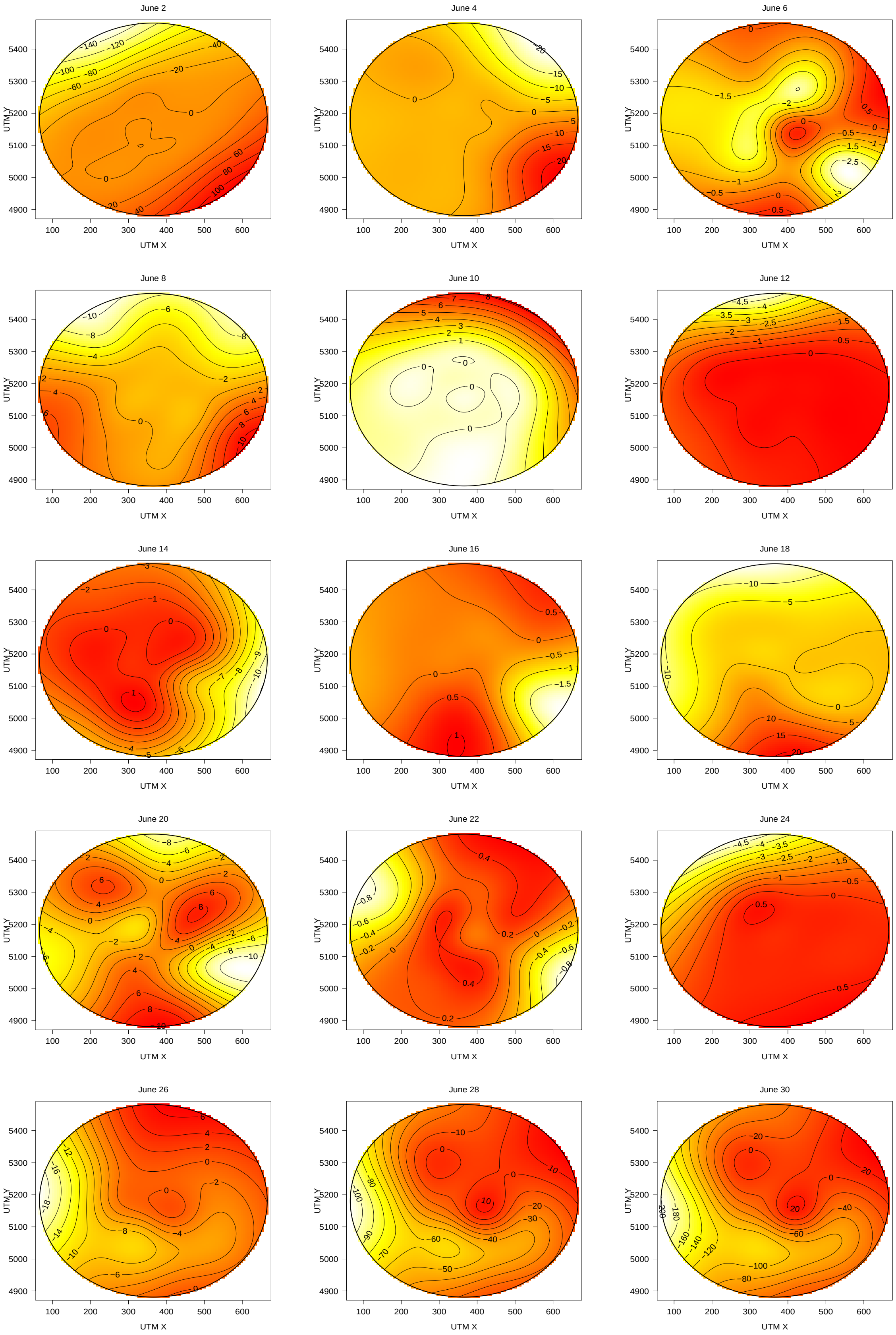


Figure B.14: Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from June 2003.

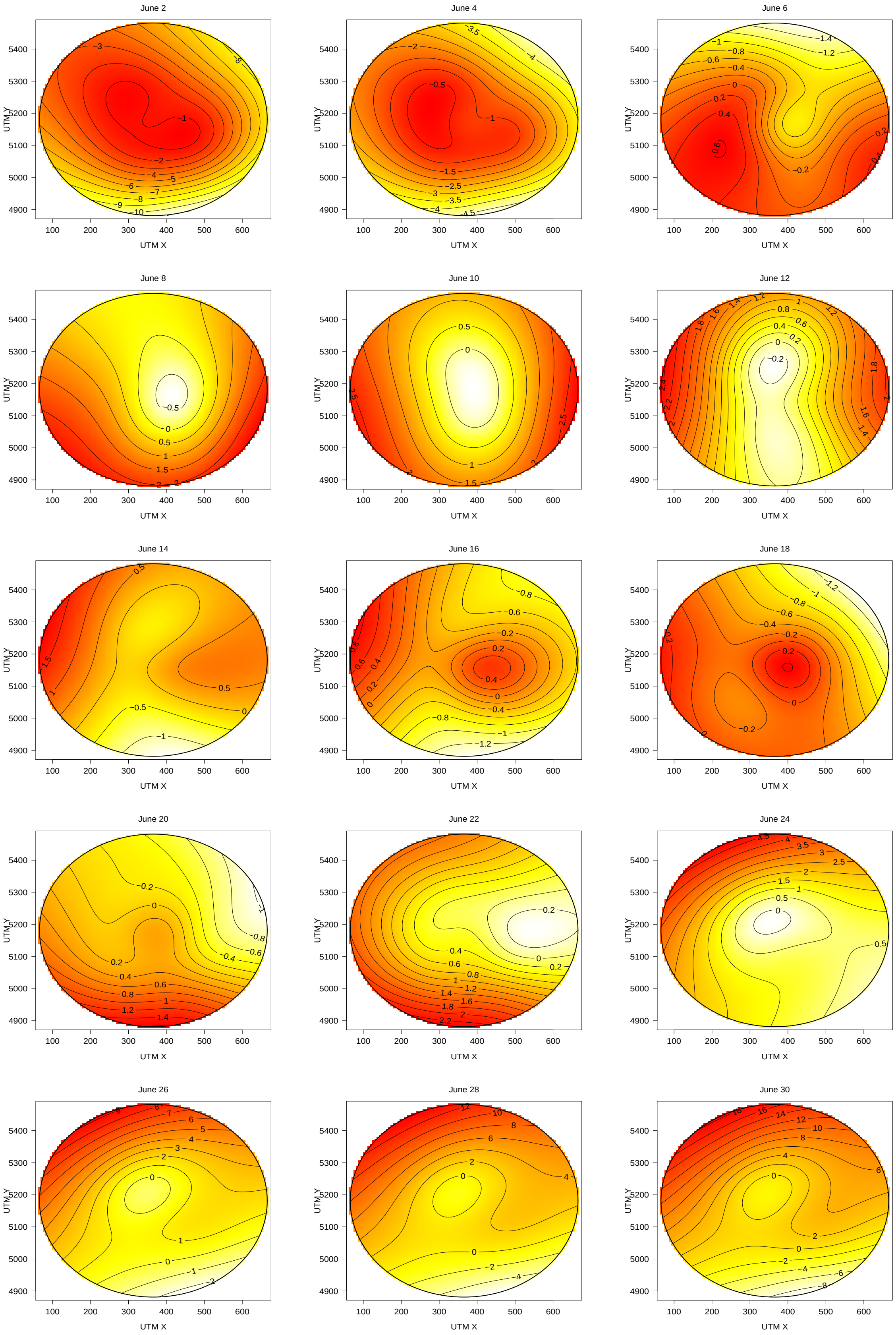


Figure B.15: Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from June 2003.

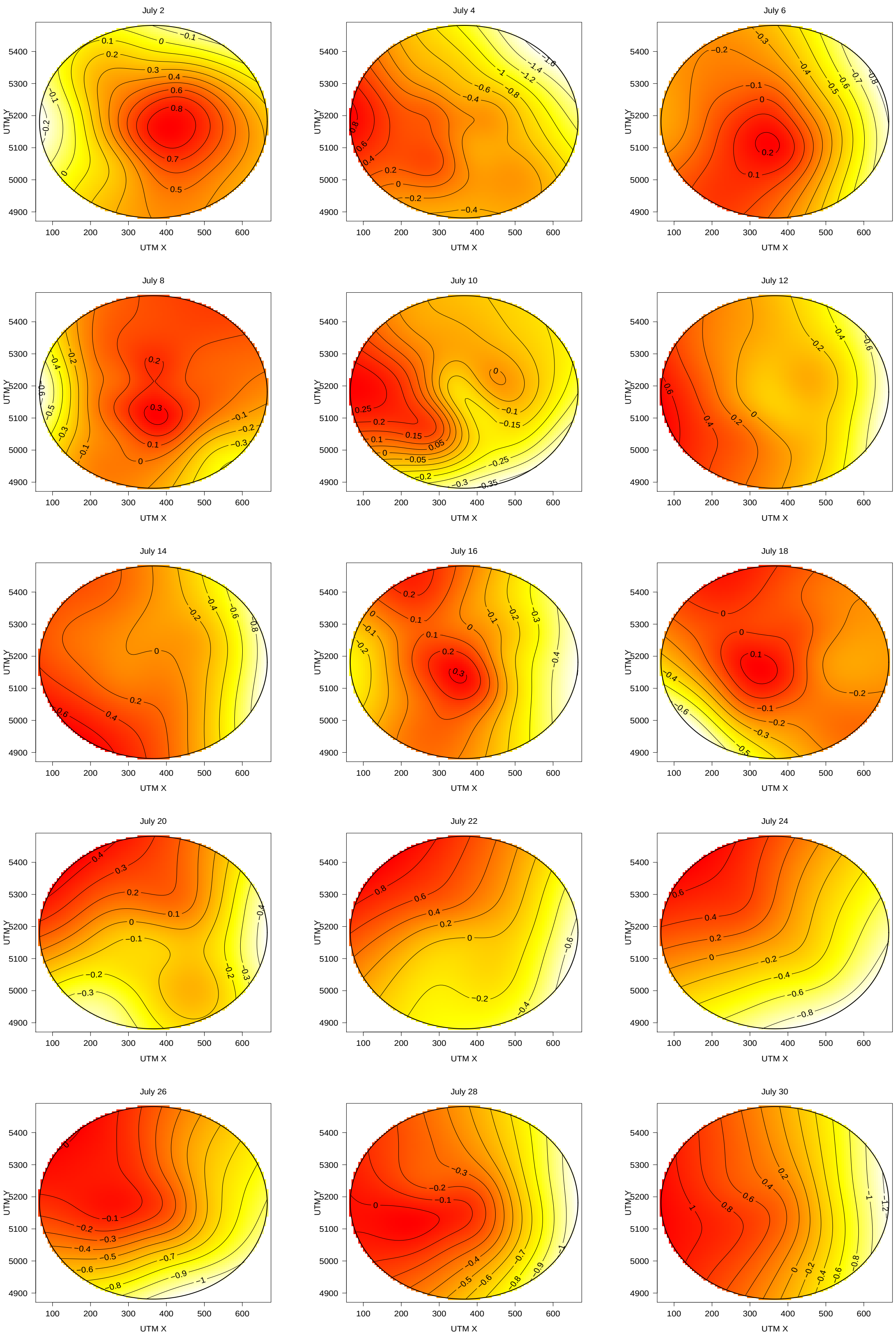


Figure B.16: Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from July 2003.

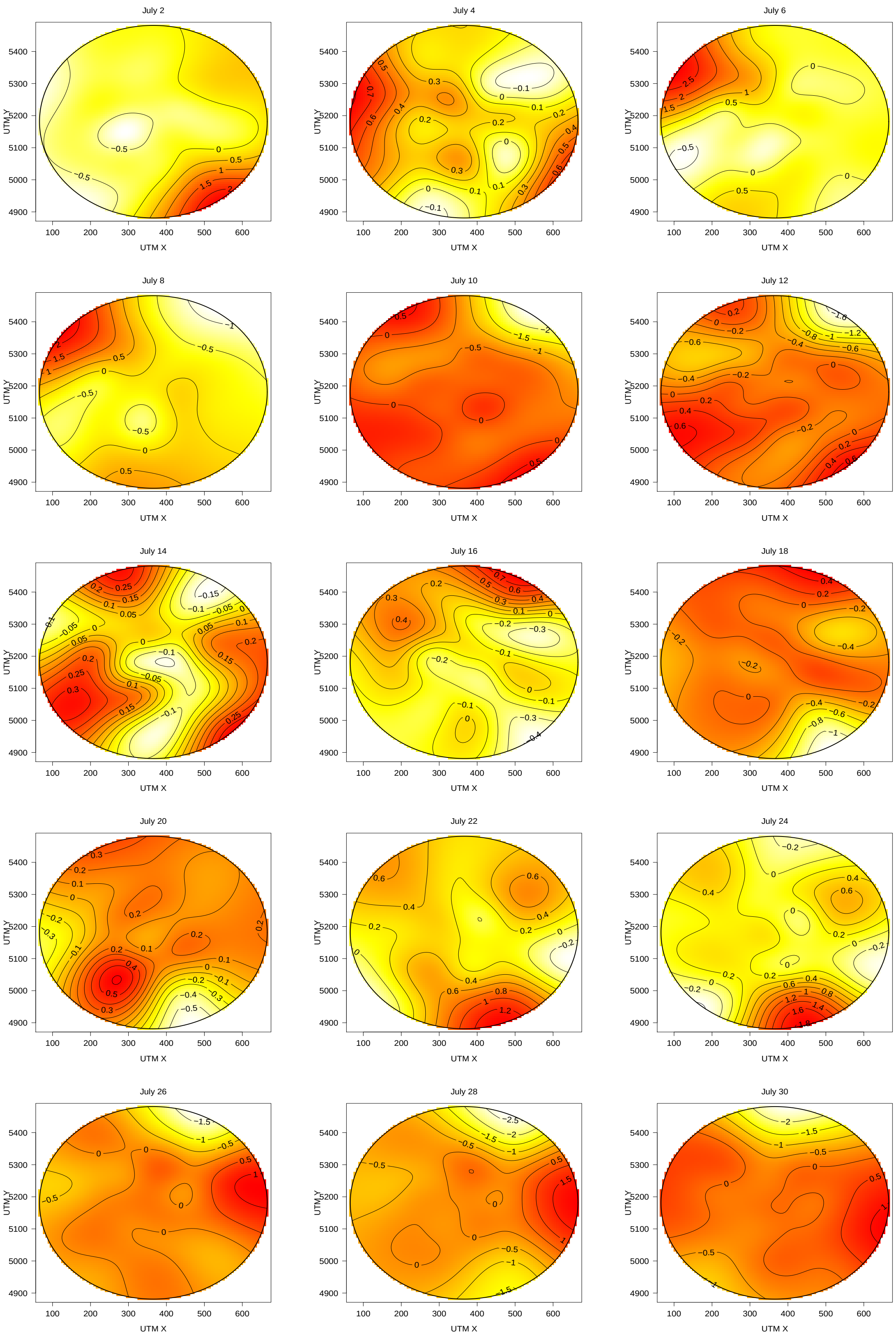


Figure B.17: Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from July 2003.

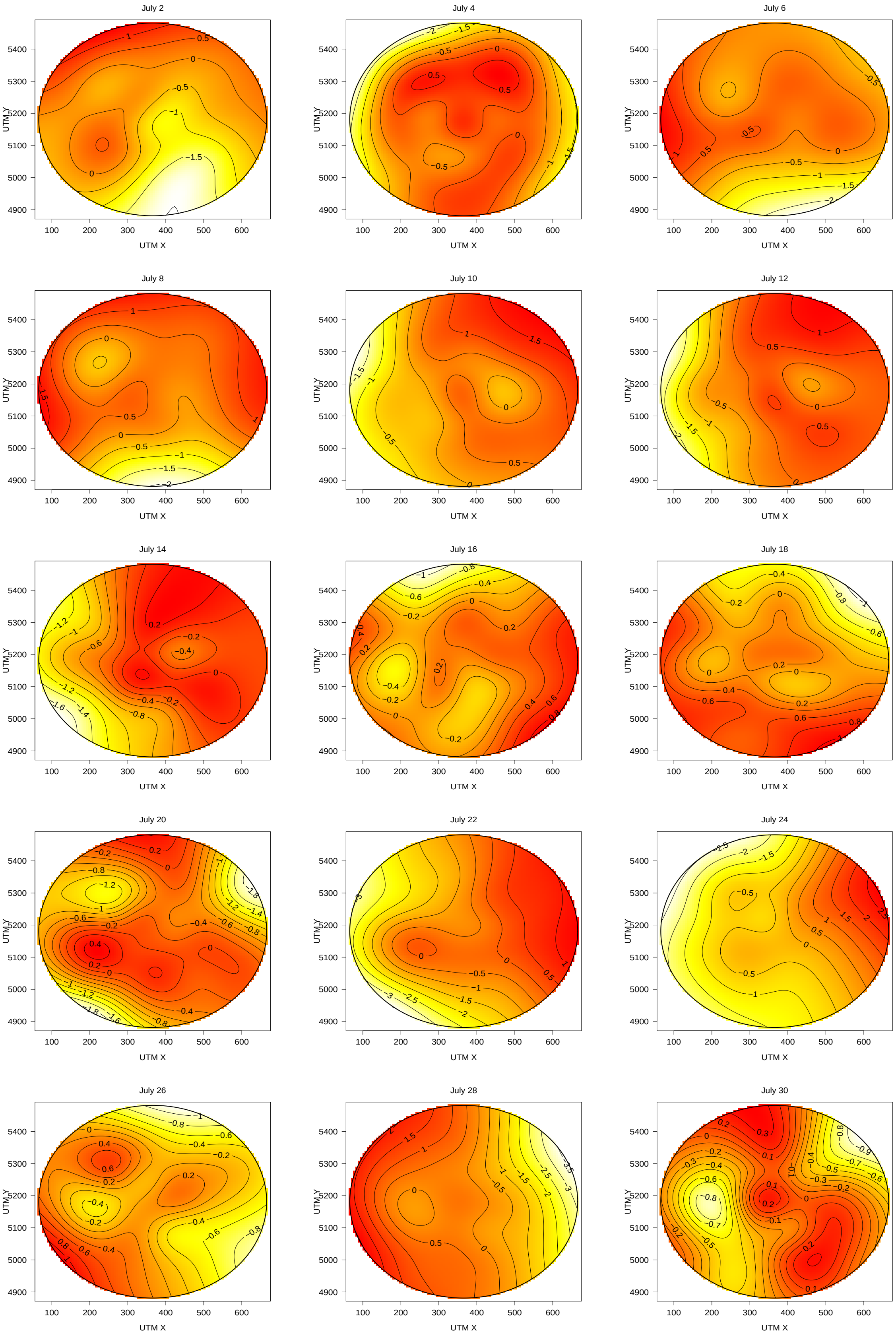


Figure B.18: Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from July 2003.

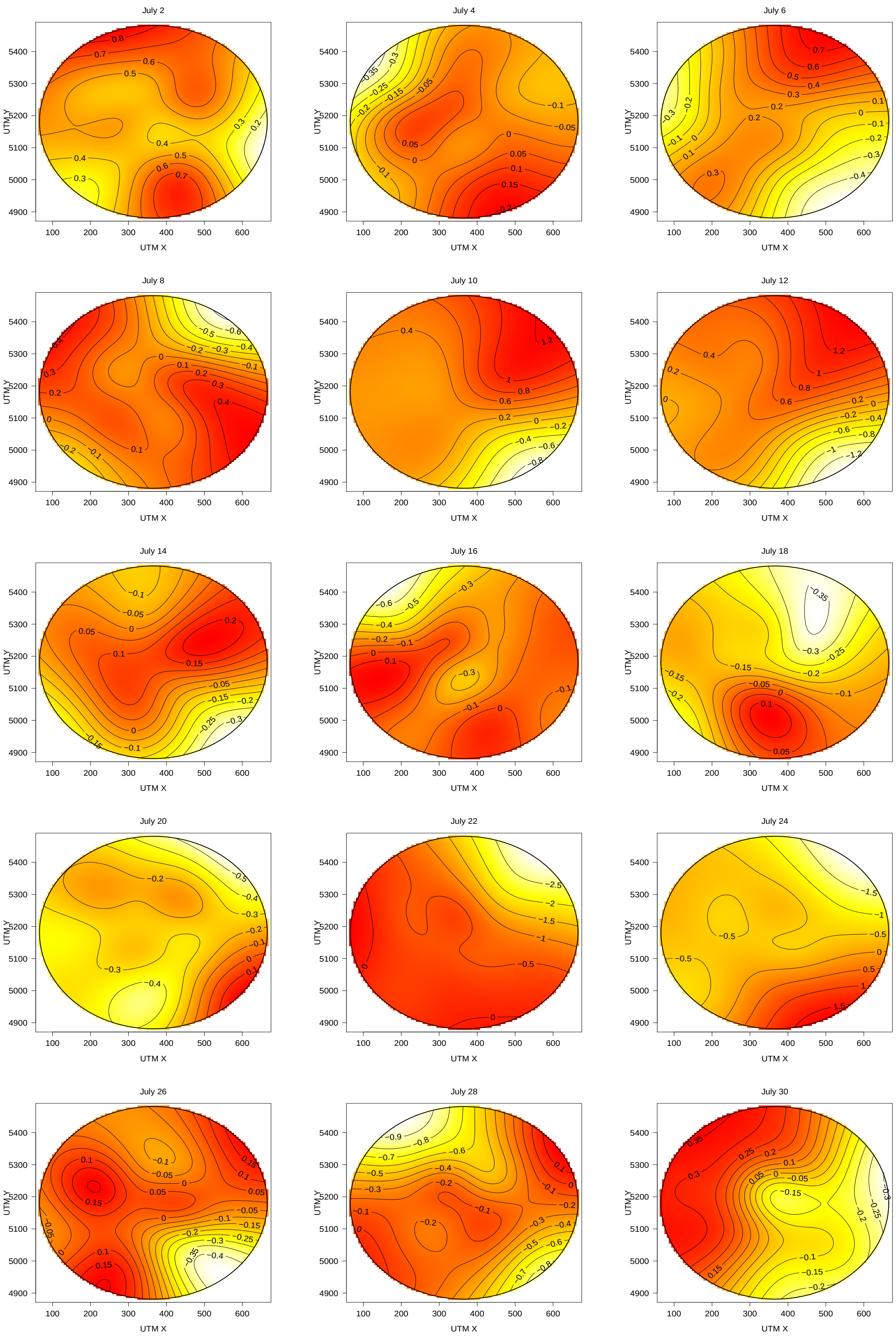


Figure B.19: Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from July 2003.

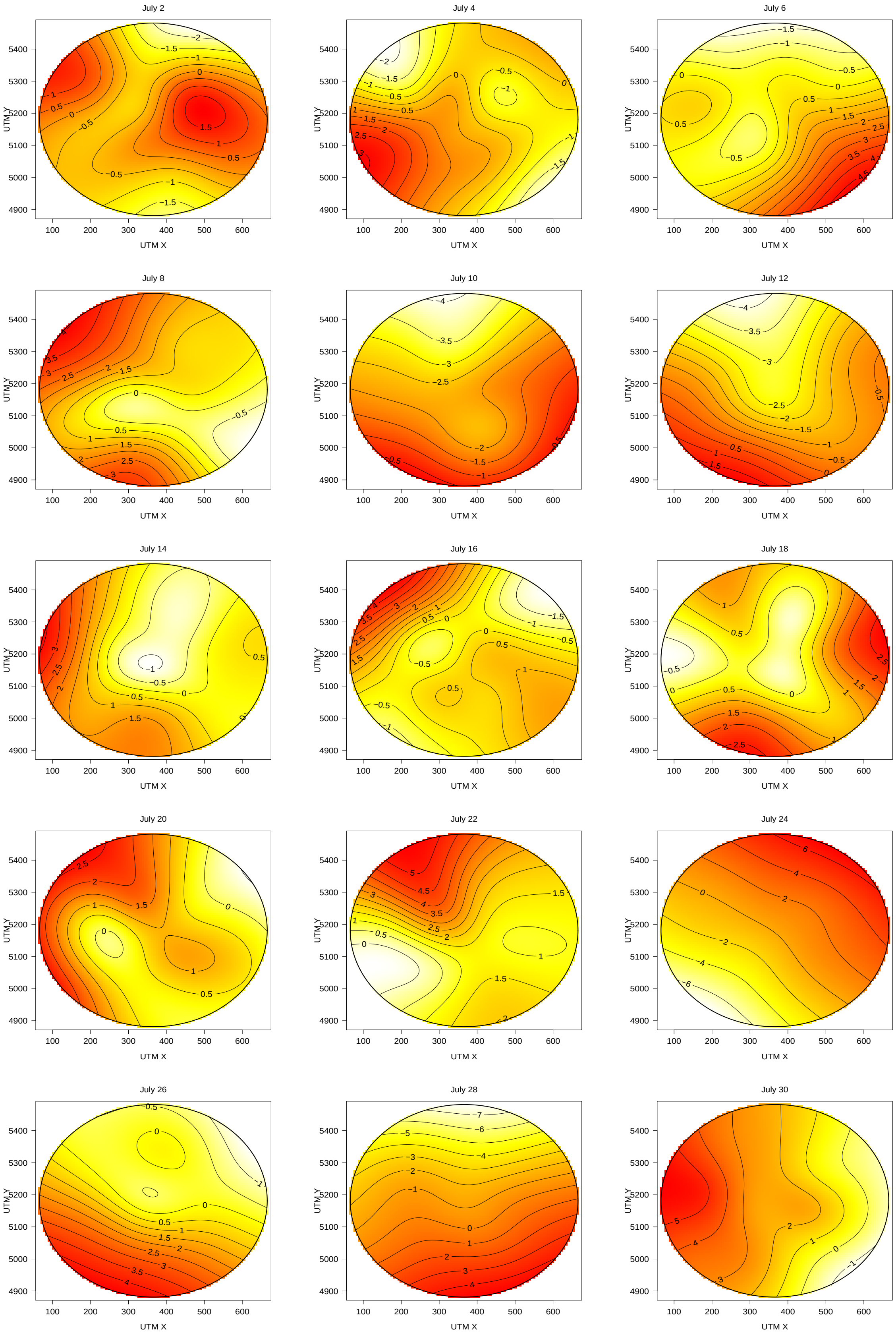


Figure B.20: Estimated spatio-temporal partial effect on select days for the scale parameter in the direction component from July 2003.

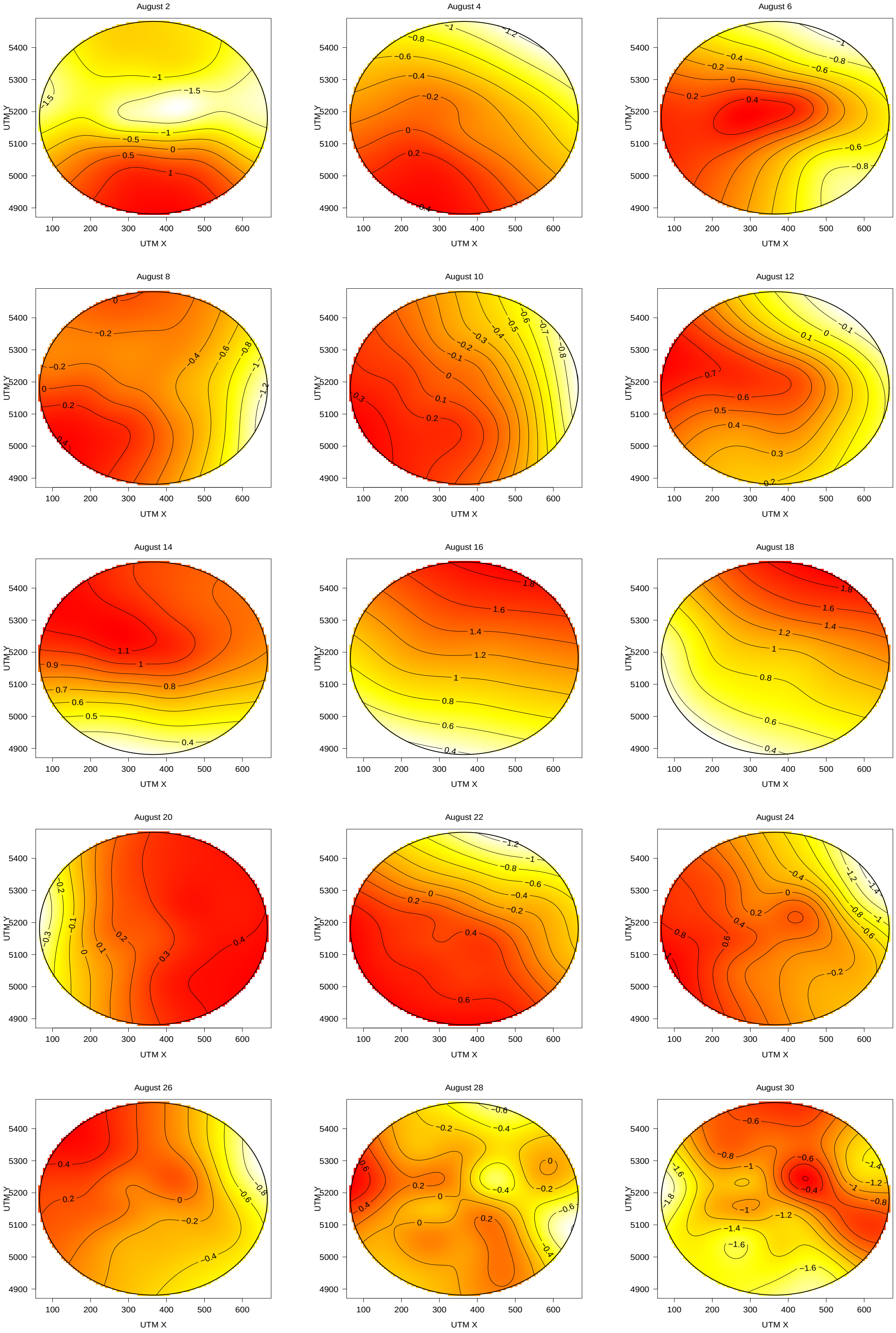


Figure B.21: Estimated spatio-temporal partial effect on select days for the location parameter in the duration component from August 2003.

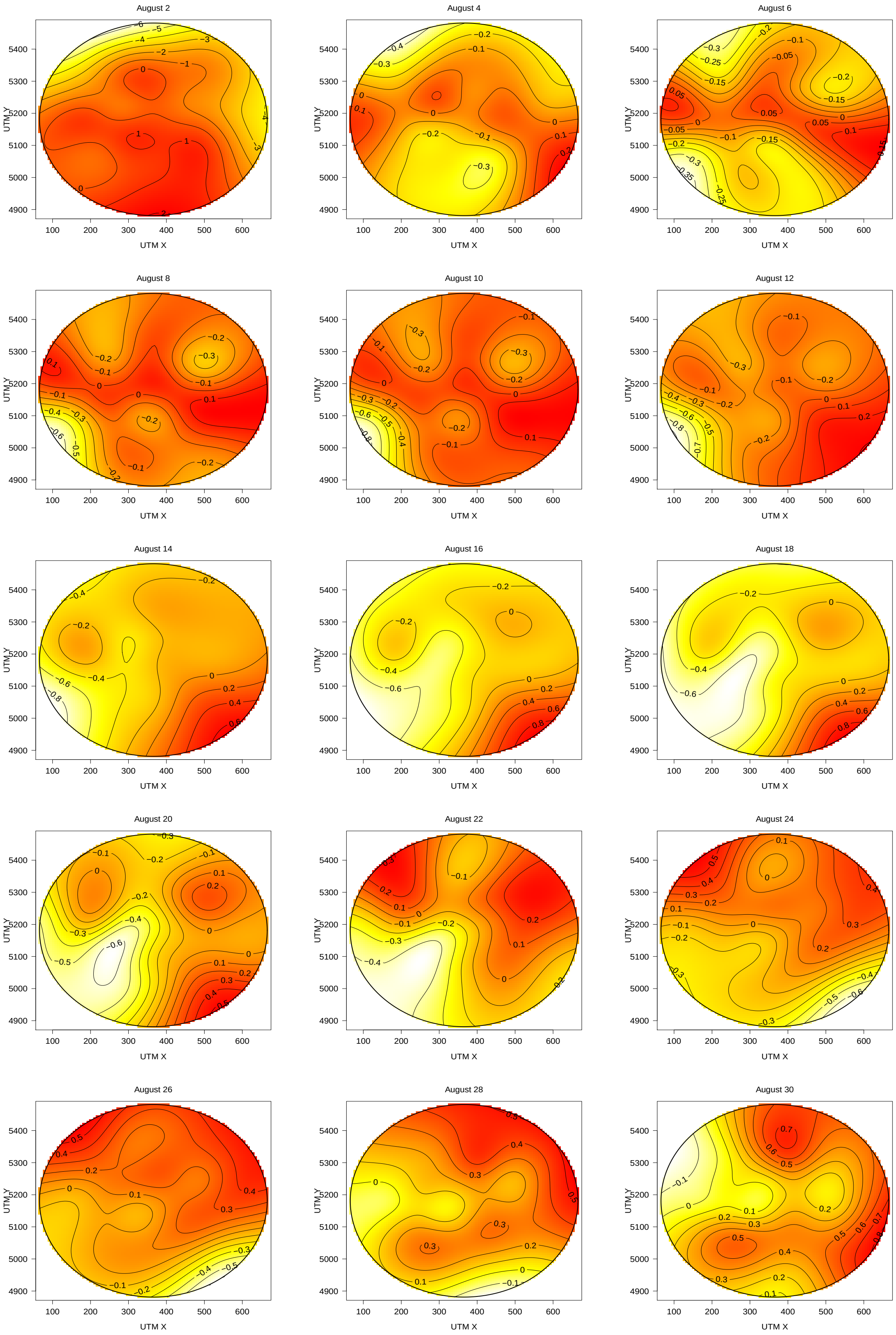

August 2, August 4, August 6, August 8, August 10, August 12, August 14, August 16, August 18, August 20, August 22, August 24, August 26, August 28, August 30

Figure B.22: Estimated spatio-temporal partial effect on select days for the location parameter in the speed component from August 2003.

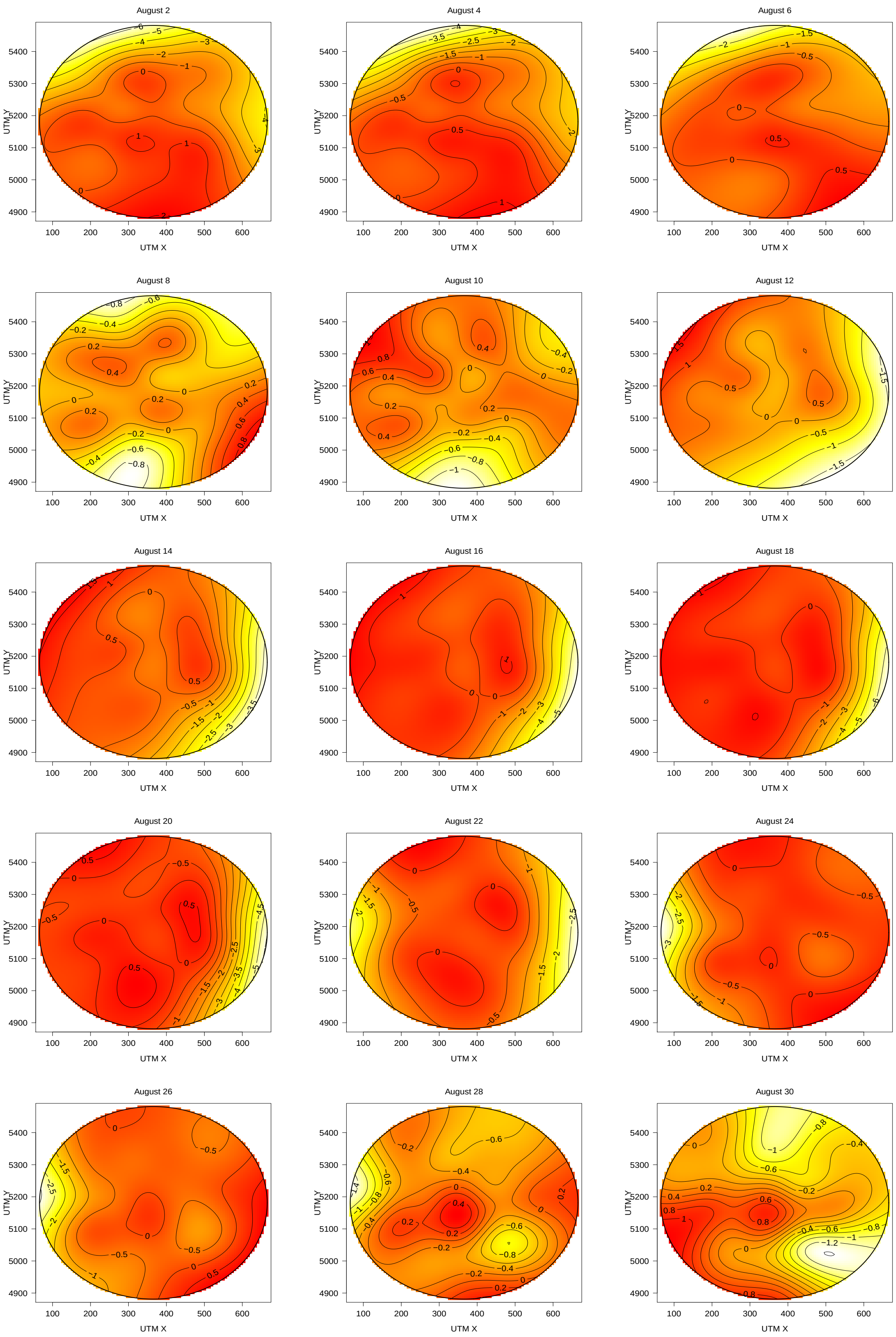

August 2, August 4, August 6, August 8, August 10, August 12, August 14, August 16, August 18, August 20, August 22, August 24, August 26, August 28, August 30

Figure B.23: Estimated spatio-temporal partial effect on select days for the scale parameter in the speed component from August 2003.

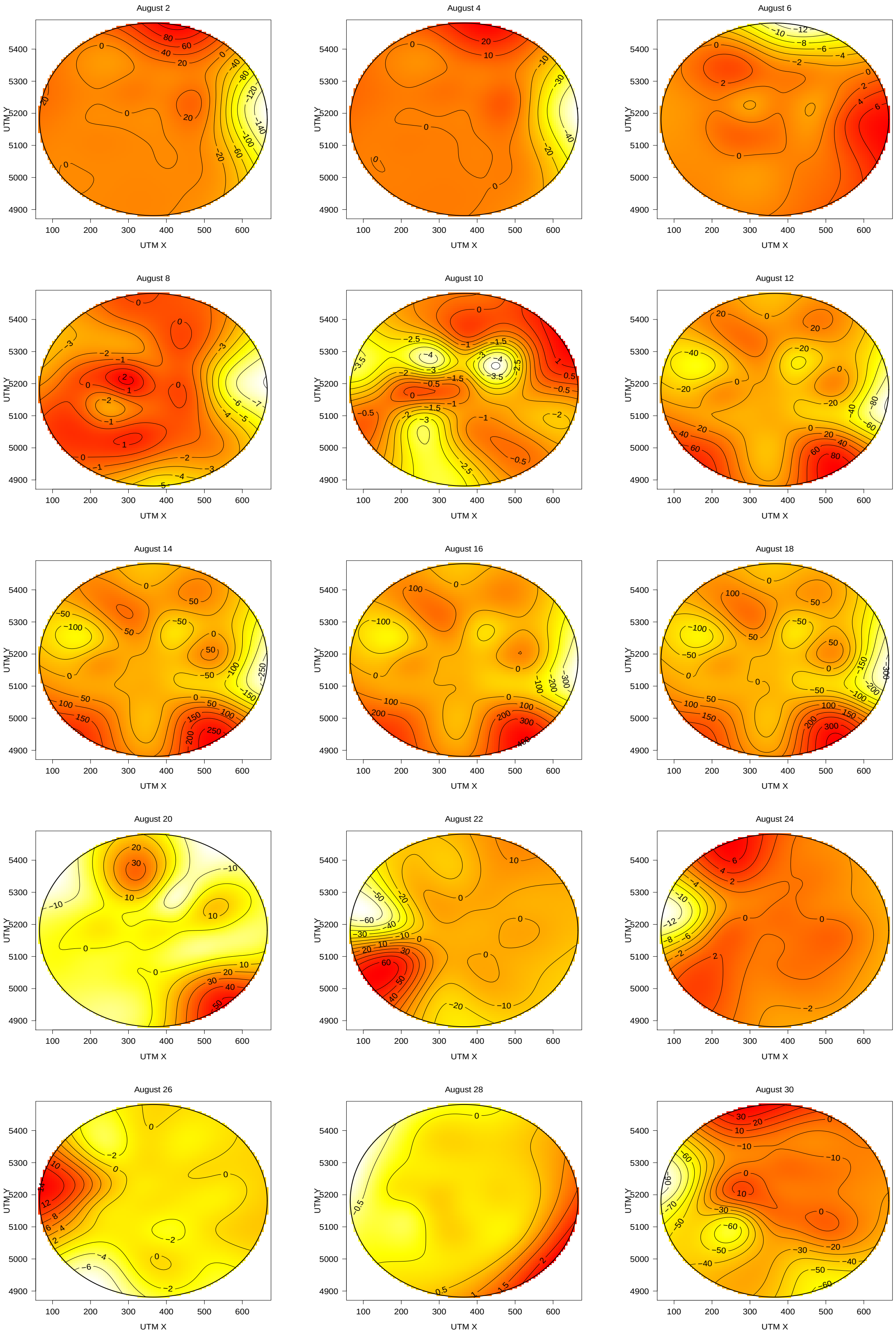


Figure B.24: Estimated spatio-temporal partial effect on select days for the location parameter in the direction component from August 2003.

Appendix C

Supplementary Material for Chapter 5

These tables summarize the simulation study results for the threshold and intercept parameters in Sections 5.4.1 and 5.4.2 of Chapter 5. As our main interest in these simulation studies was in the relative bias and relative root mean square error of the scaling and variance parameters, these results have been omitted form the body of the thesis. They have been included here for completeness.

	True Value	J1 RBIAS	J1 RRMSE	J2 RBIAS	J2 RRMSE	S1 RBIAS	S1 RRMSE
Variance Ratio: 10							
τ_1	-0.50	-0.034	0.152	-0.020	0.179	-0.046	0.182
τ_2	0.15	0.029	0.435	0.026	0.430	0.018	0.433
β_1	3.50	0.001	0.007	0.014	0.015	0.005	0.009
β_2	3.00	0.002	0.016	0.036	0.038	0.013	0.024
Variance Ratio: 7							
τ_1	-0.50	0.027	0.148	0.024	0.152	0.018	0.160
τ_2	0.15	0.024	0.463	0.023	0.461	0.034	0.476
β_1	3.50	0.001	0.007	0.009	0.011	0.003	0.008
β_2	3.00	0.001	0.016	0.025	0.028	0.006	0.019
Variance Ratio: 1							
τ_1	-0.50	0.002	0.107	0.0001	0.104	0.003	0.102
τ_2	0.15	-0.015	0.396	-0.015	0.394	-0.007	0.397
β_1	3.50	-0.001	0.006	0.001	0.006	-0.001	0.006
β_2	3.00	-0.003	0.013	0.003	0.012	-0.003	0.013
Variance Ratio: 0.8							
τ_1	-0.50	0.005	0.120	0.002	0.121	0.005	0.120
τ_2	0.15	-0.009	0.347	-0.011	0.346	-0.008	0.345
β_1	3.50	-0.002	0.007	-0.0002	0.006	-0.002	0.007
β_2	3.00	-0.004	0.012	0.002	0.012	-0.004	0.013

Table C.1: Relative bias (RBIAS) and relative root mean square error (RRMSE) for intercept and threshold parameters from the simulation study in Section 5.4.1.

	Ratio: 10		**Ratio: 7**		**Ratio: 1**		**Ratio: 0.8**	
	RBIAS	**RRMSE**	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**	**RBIAS**	**RRMSE**
Neighbourhood Structure: $\leq$ 20m								
τ_1	-0.034	0.152	0.027	0.148	0.002	0.107	0.005	0.120
τ_2	0.029	0.435	0.024	0.463	-0.015	0.396	-0.009	0.347
β_1	0.001	0.007	0.001	0.007	-0.001	0.006	-0.002	0.007
β_2	0.002	0.016	0.001	0.016	-0.003	0.013	-0.004	0.012
Neighbourhood Structure: $\leq$ 10m								
τ_1	-0.048	0.241	0.061	0.237	0.051	0.156	0.024	0.169
τ_2	0.127	0.789	-0.032	0.616	0.061	0.481	-0.002	0.395
β_1	0.001	0.012	0.001	0.011	-0.002	0.008	-0.0005	0.007
β_2	-0.010	0.026	-0.007	0.022	-0.004	0.021	-0.005	0.016
Neighbourhood Structure: $\leq$ 25m								
τ_1	0.001	0.150	-0.014	0.123	0.011	0.133	0.046	0.135
τ_2	0.052	0.360	0.004	0.407	-0.036	0.392	0.045	0.419
β_1	0.002	0.008	0.001	0.006	-0.003	0.006	-0.001	0.005
β_2	0.004	0.016	0.004	0.012	-0.005	0.015	-0.005	0.012
Neighbourhood Structure: $\leq$ 30m								
τ_1	-0.015	0.128	0.016	0.128	0.021	0.129	0.018	0.117
τ_2	-0.058	0.416	0.035	0.436	0.043	0.406	0.026	0.375
β_1	0.002	0.007	0.001	0.006	-0.001	0.006	-0.002	0.006
β_2	0.005	0.014	0.002	0.013	-0.005	0.013	-0.003	0.013
Neighbourhood Structure: Inverse Distance								
τ_1	-0.006	0.146	-0.037	0.127	-0.016	0.126	-0.016	0.122
τ_2	0.043	0.566	0.094	0.491	-0.013	0.358	-0.052	0.340
β_1	0.006	0.010	0.005	0.009	0.001	0.005	0.001	0.006
β_2	0.002	0.021	0.003	0.019	0.003	0.013	0.001	0.012

Table C.2: Relative bias (RBIAS) and relative relative root mean square error (RRMSE) for intercept and threshold parameters from the simulation study in Section 5.4.2.

Bibliography

Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-Inflated Models with Application to Spatial Count Data. *Environmental and Ecological Statistics*, 9(4):341–355.

Akima, H. and Gebhardt, A. (2015). *akima: Interpolation of Irregularly and Regularly Spaced Data*. R package version 0.5-12.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC, Boca Raton.

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns. *Statistica Neerlandica*, 54(3):329–350.

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

Besag, J., York, J., and Mollié, A. (1991). Bayesian Image Restoration, with Two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Braun, W. J. and Kulperger, R. J. (1998). A Bootstrap for Point Processes. *Journal of Statistical Computation and Simulation*, 60(2):129–155.

Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, 85(2):347–361.

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

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, New York, second edition.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Springer, New York, second edition.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Oxford University Press, New York.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman and Hall/CRC, Boca Raton.

Diggle, P. J., Moraga, P., Rowlingson, B., and Taylor, B. M. (2013). Spatial and Spatio-temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science*, 28(4):542–563.

Diggle, P. J., Sousa, I., and Chetwynd, A. G. (2008). Joint Modelling of Repeated Measurements and Time-To-Event Outcomes: The Fourth Armitage Lecture. *Statistics in Medicine*, 27(16):2981–2998.

Dunn, P. K. and Smyth, G. K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244.

Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., and Beard, J. (2007). Evaluating the Effect of Neighbourhood Weight Matrices on Smoothing Properties of Conditional Autoregressive (CAR) Models. *International Journal of Health Geographics*, 6:54.

Eberly, L. E. and Carlin, B. P. (2000). Identifiability and Convergence Issues for Markov Chain Monte Carlo Fitting of Spatial Models. *Statistics in Medicine*, 19:2279–2294.

Feng, C. X. (2015). Bayesian Joint Modeling of Correlated Counts Data with Application to Adverse Birth Outcomes. *Journal of Applied Statistics*, 42(6):1206–1222.

Feng, C. X. and Dean, C. B. (2012). Joint Analysis of Multivariate Spatial Count and Zero-Heavy Count Outcomes Using Common Spatial Factor Models. *Environmetrics*, 23(6):493–508.

Fisher, N. I. and Lee, A. J. (1992). Regression Models for an Angular Response. *Biometrics*, 48(3):665–677.

Gabriel, E. (2014). Estimating Second-Order Characteristics of Inhomogeneous Spatio-Temporal Point Processes. *Methodology and Computing in Applied Probability*, 16(2):411–431.

Gabriel, E. and Diggle, P. J. (2009). Second-Order Analysis of Inhomogeneous Spatio-temporal Point Process Data. *Statistica Neerlandica*, 63(1):43–51.

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, second edition.

Gelman, A. and Rubin, D. B. (1992). Inference From Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.

Guan, Y. (2006). A Composite Likelihood Approach in Fitting Spatial Point Process Models. *Journal of the American Statistical Association*, 101(476):1502–1512.

Hall, D. B. (2000). Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, 56(4):1030–1039.

Han, L. S. and Braun, W. J. (2015). Block Bootstrap Calibration with Application to the Fire Weather Index. *Communications in Statistics-Simulation and Computation*, 44(3):647–665.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall/CRC, Boca Raton.

Held, L., Natário, I., Fenton, S. E., Rue, H., and Becker, N. (2005). Towards Joint Disease Mapping. *Statistical Methods in Medical Research*, 14(1):61–82.

Henrys, P. A. and Brown, P. E. (2009). Inference for Clustered Inhomogeneous Spatial Point Processes. *Biometrics*, 65(2):423–430.

Illian, J., Penttiten, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Anaylsis and Modelling of Spatial Point Patterns*. Wiley, West Sussex.

Illian, J. B., Sørbye, S. H., and Rue, H. (2012). A Toolbox for Fitting Complex Spatial Point Process Models Using Integrated Nested Laplace Approximation (INLA). *The Annals of Applied Statistics*, 6(4):1499–1530.

Jacod, J. and Protter, P. (2000). *Probability Essentials*. Springer, Italy.

Johnson, J., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D., and Thomas, K. W. (1998). The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. *Weather and Forecasting*, 13(2):263–276.

Knorr-Held, L. and Best, N. G. (2001). A Shared Component Model for Detecting Joint and Selective Clustering of Two Diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):73–85.

Kusiak, R. A., Ritchie, A. C., Muller, J., and Springer, J. (1993). Mortality From Lung Cancer in Ontario Uranium Miners. *British Journal of Industrial Medicine*, 50(10):920–928.

Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1):1–14.

Lee, Y. and Nelder, J. A. (2006). Double Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2):139–185.

Li, N., Elashoff, D. A., Robbins, W. A., and Xun, L. (2011). A Hierarchical Zero-Inflated Log-Normal Model for Skewed Responses. *Statistical Methods in Medical Research*, 20(3):175–189.

Li, Y., Brown, P., Gesink, D. C., and Rue, H. (2012). Log Gaussian Cox Processes and Spatially Aggregated Disease Incidence Data. *Statistical Methods in Medical Research*, 21(5):479–507.

Lindgren, F., Rue, H., and Lindström, J. (2011). An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Loh, J. M. (2008). A Valid and Fast Spatial Bootstrap for Correlation Functions. *The Astrophysical Journal*, 681(1).

Loh, J. M. and Stein, M. L. (2004). Bootstrapping a Spatial Point Process. *Statistica Sinica*, 14(1):69–101.

Loh, J. M. and Stein, M. L. (2008). Spatial Bootstrap with Increasing Observations in a Fixed Domain. *Statistica Sinica*, 18(2):667–688.

Matérn, B. (1960). Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations. *Meddelanden från Statens Skogsforskningsinstitut*, 49(5).

Meng, X.-L. (1994). Posterior Predictive p-Values. *The Annals of Statistics*, 22(3):1142–1160.

Modlin, D., Fuentes, M., and Reich, B. (2012). Circular Conditional Autoregressive Modeling of Vector Fields. *Environmetrics*, 23(1):46–53.

Mohee, F. M. and Miller, C. (2010). Climatology of Thunderstorms for North Dakota, 2002-06. *Journal of Applied Meteorology and Climatology*, 49(9):1881–1890.

Møller, J. and Ghorbani, M. (2012). Aspects of Second-Order Analysis of Structured Inhomogeneous Spatio-Temporal Point Processes. *Statistica Neerlandica*, 66(4):472–491.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox Processes. *Scandinavian Journal of Sstatistics*, 25(3):451–482.

Møller, J. and Torrisi, G. L. (2005). Generalised Shot Noise Cox Processes. *Advances in Applied Probability*, 37(1):48–74.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton.

Neelon, B., Ghosh, P., and Loebs, P. F. (2013). A Spatial Poisson Hurdle Model for Exploring Geographic Variation in Emergency Department Visits. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):389–413.

Neyman, J. and Scott, E. L. (1958). Statistical Approach to Problems of Cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):1–43.

Ogata, Y. (1998). Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., and Isham, V. (2000). Rainfall Modelling using Poisson-Cluster Processes: A Review of Developments. *Stochastic Environmental Research and Risk Assessment*, 14(6):384–411.

Ontario Mining Association. Ontario Mining Operations 2015. `http://www.oma.on.ca/en/ontariomining/Map.asp`. Accessed: 2015-08-29.

Ontario Ministry of Health and Long-Term Care. Public Health Units. `http://www.health.gov.on.ca/en/common/system/services/phu/`. Accessed: 2015-08-29.

Paciorek, C. J. (2007). Computational Techniques for Spatial Logistic Regression with Large Data Sets. *Computational Statistics and Data Analysis*, 51(8):3631–3653.

Prokešová, M. and Dvořák, J. (2014). Statistics for Inhomogeneous Space-Time Shot-Noise Cox Processes. *Methodology and Computing in Applied Probability*, 16(2):433–449.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rathbun, S. L. and Fei, S. (2006). A Spatial Zero-Inflated Poisson Regression Model for Oak Regeneration. *Environmental and Ecological Statistics*, 13(4):409–426.

Recta, V., Haran, M., and Rosenberger, J. L. (2012). A Two-Stage Model for Incidence and Prevalence in Point-Level Spatial Count Data. *Environmetrics*, 23(2):162–174.

Reich, B. J. and Fuentes, M. (2007). A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields. *The Annals of Applied Statistics*, 1(1):249–264.

Renshaw, E. and Särkkä, A. (2001). Gibbs Point Processes for Studying the Development of Spatial-Temporal Stochastic Processes. *Computational Statistics & Data Analysis*, 36(1):85–105.

Richardson, S., Abellan, J. J., and Best, N. (2006). Bayesian Spatio-Temporal Analysis of Joint Patterns of Male and Female Lung Cancer Risks in Yorkshire (UK). *Statistical Methods in Medical Research*, 15(4):385–407.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Särkkä, A. and Renshaw, E. (2006). The Analysis of Marked Point Patterns Evolving Through Space and Time. *Computational Statistics & Data Analysis*, 51(3):1698–1718.

Schoenberg, F. (2004). Testing Separability in Spatial-Temporal Marked Point Processes. *Biometrics*, 60(2):471–481.

Schoenberg, F. P. (2005). Consistent Parametric Estimation of the Intensity of a Spatial–Temporal Point Process. *Journal of Statistical Planning and Inference*, 128(1):79–93.

Silverman, B. W. (1998). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, Boca Raton.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stoyan, D. and Stoyan, H. (1996). Estimating Pair Correlation Functions of Planar Cluster Processes. *Biometrical Journal*, 38(3):259–271.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16.

Tanaka, U. and Ogata, Y. (2014). Identification and Estimation of Superposed Neyman–Scott Spatial Cluster Processes. *Annals of the Institute of Statistical Mathematics*, 66(4):687–702.

Thomas, M. (1949). A Generalization of Poisson's Binomial Limit for Use in Ecology. *Biometrika*, 36(1/2):18–25.

Tzala, E. and Best, N. (2008). Bayesian Latent Variable Modelling of Multivariate Spatio-Temporal Variation in Cancer Mortality. *Statistical Methods in Medical Research*, 17:97–118.

U.S. Deparment of Commerce/National Oceanic and Atmospheric Administration (2006). *Doppler Radar Meteorological Observations Part C: WSR-88D Products and Algorithms*. Washington, DC.

Vio, R., D'Odorico, V., Stoyan, H., and Stoyan, D. (2007). Ly-$\{\alpha\}$ Forest: Efficient Unbiased Estimation of Second-Order Properties with Missing Data. *Astronomy & Astrophysics*, 466(1):403–411.

Waagepetersen, R. and Guan, Y. (2009). Two-Step Estimation for Inhomogeneous Spatial Point Processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):685–702.

Waagepetersen, R. P. (2007). An Estimating Function Approach to Inference for Inhomogeneous Neyman–Scott Processes. *Biometrics*, 63(1):252–258.

Wan, V., Braun, W. J., Dean, C. B., and Henderson, S. (2011). A Comparison of Classification Algorithms for the Identification of Smoke Plumes from Satellite Images. *Statistical Methods in Medical Research*, 20(2):131–156.

Wang, F. and Gelfand, A. E. (2014). Modeling Space and Space-Time Directional Data Using Projected Gaussian Processes. *Journal of the American Statistical Association*, 109(508):1565–1580.

Wang, F. and Wall, M. M. (2003). Generalized Common Spatial Factor Model. *Biostatistics*, 4(4):569–582.

Welsh, A. H., Cunningham, R. B., Donnelly, C. F., and Lindenmayer, D. B. (1996). Modelling the Abundance of Rare Species: Statistical Models for Counts with Extra Zeros. *Ecological Modelling*, 88(1):297–308.

Wiegand, T., Gunatilleke, S., Gunatilleke, N., and Okuda, T. (2007). Analyzing the Spatial Structure of a Sri Lankan Tree Species with Multiple Scales of Clustering. *Ecology*, 88(12):3088–3102.

Wood, S. N. (2003). Thin Plate Regression Splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton.

Yau, C. Y. and Loh, J. M. (2012). A Genralization of the Neyman-Scott process. *Statistica Sinica*, 22(4):1717–1736.

Curriculum Vitae

Name:	Alisha Albert-Green
Education:	Ph.D., Statistics, University of Western Ontario (UWO), 2012 - 2016 M.Sc., Statistics, Simon Fraser University (SFU), 2009 - 2011 B.Sc. (Hon.), Statistics, UWO, 2005 - 2009
Awards:	Natural Sciences and Engineering Research Council (NSERC), Canada Graduate Scholarship - Doctoral, 2014 - 2016 Statistical Society of Canada Student Research Presentation Award (Oral), 2015 Ontario Graduate Scholarship (declined), 2014 Queen Elizabeth II Graduate Scholarship in Science and Technology, 2013 PhD Entrance Scholarship, UWO, 2012 Graduate Fellowship, SFU, 2010, 2011 Targeted Special Graduate Entrance Scholarship, SFU, 2009 Northern Life Assurance Gold Medal Award, UWO, 2009 NSERC Undergraduate Student Research Award, 2009
Work Experience:	Instructor, An Introduction to R and its Applications, Western Summer School in Longitudinal Data Analysis, Faculty of Social Science, UWO, 2013 Senior Biostatistician, Biostatistics Department, Princess Margaret Cancer Centre, 2011-2012

Peer-Reviewed Publications:

Morin, A.A., Albert-Green, A., Woolford, D.G. and Martell, D.L. (2015). The Use of Survival Analysis Methods to Model the Control Time of Forest Fires in Ontario, Canada. *International Journal of Wildland Fire, 24(7)*, 964-973. DOI: 10.1071/WF14158.

Albert-Green, A., Braun, W.J., Martell, D.L. and Woolford, D.G. (2014). Visualization Tools for Assessing the Markov Property: Sojourn Times in the Forest Fire Weather Index in Ontario. *Environmetrics, 25(6)*, 417-430. DOI: 10.1002/env.2237.

Albert-Green, A., Dean, C.B., Martell, D.L. and Woolford, D.G. (2013). A Methodology for Investigating Trends in Changes in the Timing of the Fire Season with Applications to Lightning-Caused Forest Fires in Alberta and Ontario, Canada. *Canadian Journal of Forest Research, 43(1)*, 39-45. DOI: 10.1139/cjfr-2011-0432.

Gupta, A.A., Edelstein, K., Albert-Green, A. and D'Agostino, N. (2013). Assessing Information and Service Needs of Young Adults with Cancer at a Single Institution: The Importance of Information on Cancer Diagnosis, Fertility Preservation, Diet, and Exercise. *Supportive Care in Cancer, 21(9)*, 2477-2484. DOI: 10.1007/s00520-013-1809-4.

Krema, H., Herrmann, E., Albert-Green, A., Payne, D., Laperriere, N. and Chung, C. (2013). Orthovoltage Radiotherapy in the Management of Medial Canthal Basal Cell Carcinoma. *British Journal of Ophthalmology, 97*, 730-4. DOI:10.1136/bjophthalmol-2012-302991.

Technical Reports:

Albert-Green, A., Martell, D.L. and Braun, W.J. (2008). A Hidden Markov Model of the Fire Weather Index. *Technical Report TR-08-10*. Department of Statistical and Actuarial Sciences, UWO, London, ON.