

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/93636>

Copyright and reuse:

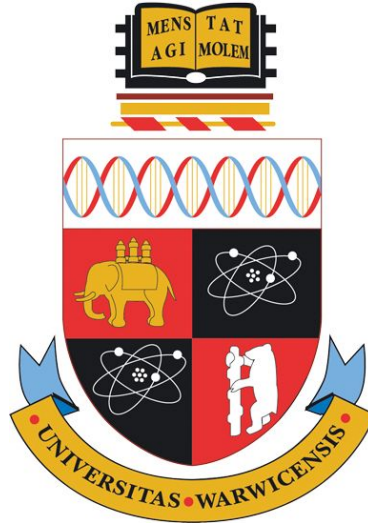
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Spatial Point Process Models for MRI Lesion Data in Multiple Sclerosis

by

Bernd Taschler

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Centre for Complexity Science

February 2017

THE UNIVERSITY OF
WARWICK

Contents

List of Tables	v
List of Figures	vii
Acknowledgments	ix
Declarations	x
Abstract	xi
Abbreviations	xii
Chapter 1 Introduction	1
Chapter 2 Background	4
2.1 Neuroimaging: Magnetic Resonance Imaging	4
2.1.1 Preprocessing steps	6
2.2 Medical background: Multiple Sclerosis	7
2.2.1 Disease pathology	8
2.2.2 Diagnosis and MS subtypes	9
2.2.3 Therapy	11
2.2.4 MRI criteria	12
2.2.5 Quantitative analysis	13
2.3 Methodological background: Spatial point processes	14
2.3.1 Spatial Poisson point processes	15
2.3.2 The log-Gaussian Cox process	17
2.3.3 Poisson/Gamma random fields	18
2.3.4 Inverse Lévy measure construction of a Gamma random field	19
2.3.5 Data augmentation: Auxiliary points	20
2.3.6 The independent PGRF model	21

2.3.7	Multiple realisations	22
2.3.8	Marked point processes	24
2.4	Methodological background: Classification and prediction	25
2.4.1	Cross-validation	25
2.4.2	A Bayesian classifier for multi-type point patterns	26
Chapter 3	Two clinical data sets	29
3.1	The GeneMSA data set	29
3.2	The BENEFIT data set	36
Chapter 4	Comparison of two machine learning approaches and two spatially informed models for MS subtype classification	40
4.1	Introduction	40
4.2	A naïve Bayesian classifier	41
4.3	Support vector machines based on lesion-specific features	42
4.3.1	Support Vector Machines	42
4.3.2	Minkowski functionals and lesion geometry	46
4.3.3	The feature set	47
4.3.4	Principal component analysis	50
4.3.5	Model evaluation	51
4.4	A Bayesian Spatial Generalised Linear Mixed Model	52
4.5	A Bayesian log-Gaussian Cox process model	53
4.6	Application: GeneMSA data	55
4.6.1	SVM feature inference	56
4.6.2	Posterior probability and intensity maps	59
4.6.3	Prediction Accuracies	62
4.7	Discussion	66
Chapter 5	Poisson/Gamma random field models for spatial point data	68
5.1	Introduction	68
5.2	The hierarchical PGRF model	69
5.3	Including covariates	71
5.3.1	Poisson regression	72
5.4	Marked hierarchical PGRF models	73
5.4.1	The marked HPGRF model with additional covariates	74
5.4.2	Assuming an independent mark distribution	75
5.4.3	The intensity-marked HPGRF model	76

5.4.4	Variance-stabilised marks	77
5.5	Posterior approximation and sampling algorithm	77
5.6	Simulation study	78
5.6.1	Simulated spatial data	79
5.6.2	Simulated covariates	80
5.6.3	Simulated mark process	80
5.6.4	Simulation setup	81
5.6.5	Simulation results and model assessment	82
5.7	Discussion	94
5.7.1	Limitations	95
5.7.2	Further extensions	96
Chapter 6 Application of spatial point process models to MS lesion data		98
6.1	Application: GeneMSA data	99
6.1.1	Algorithmic details and posterior computation	99
6.1.2	Posterior results and prediction	100
6.1.3	Model assessment	108
6.1.4	A GLM analysis	109
6.2	Application: BENEFIT data	111
6.2.1	Algorithmic details and posterior computation	111
6.2.2	Posterior results and prediction	112
6.2.3	Model assessment	116
6.2.4	A GLM analysis	117
6.3	Discussion	118
Chapter 7 Conclusions		122
7.1	Contributions	122
7.2	Future work	124
Appendix A imHPGRF supplements		126
A.1	Posterior computation	126
A.1.1	Sampling of mark parameters	127
A.1.2	Sampling of regression coefficients	129
A.1.3	Sampling of Gamma random fields	130
A.1.4	Sampling of kernel parameters	135
A.1.5	Intensity computation	136

Appendix B Application supplements	138
B.1 Supplement to the SVM classifier	138
B.2 Supplement to the imHPGRF model: GeneMSA data	141
B.3 Supplement to the imHPGRF model: BENEFIT data	148

List of Tables

2.1	Summary of MS lesion types.	13
3.1	Summary statistics of the GeneMSA data set including the PRMS subtype.	33
3.2	Summary statistics of the GeneMSA data set excluding the PRMS subtype.	34
3.3	Summary statistics of the BENEFIT data set.	39
4.1	Summary of feature sets for the SVM classifier.	51
4.2	Parameter estimates for the LGCP model.	55
4.3	Confusion matrices and prediction accuracies for different classifiers (T1).	64
4.4	Confusion matrices and prediction accuracies for different classifiers (T2).	65
5.1	Parameter specification for simulated data sets.	79
5.2	Summary statistics of simulated points and covariates.	80
5.3	Summary statistics of simulated mark values.	81
5.4	Mean posterior estimates of average count per group.	87
5.5	Summary of IMSE, IWMSE and mark residuals.	91
5.6	Ground truth and mean posterior estimates of mark parameters. . .	92
5.7	Ground truth and mean posterior estimates of regression coefficients. .	92
5.8	Simulation study: Summary of confusion matrices and prediction accuracies.	93
6.1	GeneMSA results: Posterior parameter estimates of type-specific parameters for the imHPGRF model including covariates.	103
6.2	GeneMSA results: Posterior parameter estimates of shared parameters for the imHPGRF model including covariates.	105

6.3	GeneMSA results: Confusion matrices for different model variants. .	107
6.4	GeneMSA results: Posterior estimates of the average number of lesions per subject for each MS subtype.	109
6.5	GeneMSA results: Results from fitting a GLM to subject-specific covariates.	110
6.6	BENEFIT results: Posterior parameter estimates of type-specific parameters for the imHPGRF model including covariates.	114
6.7	BENEFIT results: Posterior parameter estimates of shared parameters for the imHPGRF model including covariates.	114
6.8	BENEFIT results: Confusion matrices for different model variants. .	116
6.9	BENEFIT results: Posterior estimates of the average number of lesions per subject.	117
6.10	BENEFIT results: Summary of results from fitting a GLM to subject-specific covariates.	118

List of Figures

2.1	Example MRI of a standard brain template and MS lesions.	7
2.2	Schematic depiction of disease progression for the four clinical sub- types of multiple sclerosis.	11
3.1	GeneMSA data set: Box plots of lesion count versus demographic and clinical covariates.	32
3.2	GeneMSA data set: Scatter plots of lesion volume versus lesion count.	35
3.3	BENEFIT data set: Scatter plots of lesion volume versus lesion count.	37
3.4	BENEFIT data set: Box plots of lesion count versus demographic and clinical covariates.	38
4.1	Visualisation of two brain segmentations into regions of interest. . . .	49
4.2	Standardised SVM weights for one pairwise classifier.	58
4.3	Quadratic means of SVM weights for pairwise classifiers.	59
4.4	Results for the BSGLMM and LGCP model fit (axial).	60
4.5	Results for the BSGLMM and LGCP model fit (sagittal).	61
4.6	Standardised coefficient maps of subject-level covariates for the BS- GLMM model fit.	62
5.1	Simulation study: Posterior traceplots of scalar model parameters. .	83
5.2	Simulation study: Autocorrelation of scalar model parameters. . . .	84
5.3	Simulation setup and estimated posterior mean spatial intensities. .	85
5.4	True and estimated posterior mean intensity of the simulated mark process.	86
5.5	Results of the posterior predictive check for second order properties of the point process.	88
5.6	Simulation study: Mark residuals versus predicted values.	89
5.7	Simulation study: Histograms of mark residuals.	90

6.1	GeneMSA results: Empirical and estimated posterior intensity maps per MS subtype (axial).	101
6.2	GeneMSA results: Empirical and estimated posterior intensity maps per MS subtype (sagittal).	102
6.3	GeneMSA results: Mark residuals of posterior estimates of individual lesion volume.	104
6.4	GeneMSA results: Covariate residuals from Poisson regression of posterior estimates of lesion count.	104
6.5	GeneMSA results: GLM fit.	109
6.6	BENEFIT results: Estimated posterior intensity maps (axial).	112
6.7	BENEFIT results: Estimated posterior intensity maps (sagittal).	113
6.8	BENEFIT results: Mark residuals of posterior estimates of individual lesion volume.	115
6.9	BENEFIT results: Covariate residuals of posterior estimates of lesion count per subject.	115
6.10	BENEFIT results: GLM fit.	118
B.1	Standardised SVM weights for five pairwise classifiers (I).	139
B.2	Standardised SVM weights for five pairwise classifiers (II).	140
B.3	GeneMSA results: Traceplots for parameters of the imHPGRF model.	142
B.4	GeneMSA results: Autocorrelation plots for parameters of the imH-PGF model	143
B.5	GeneMSA results: Square-root transformed posterior mean intensity for MS subtype CIS.	144
B.6	GeneMSA results: Square-root transformed posterior mean intensity for MS subtype RRMS.	145
B.7	GeneMSA results: Square-root transformed posterior mean intensity for MS subtype SPMS.	146
B.8	GeneMSA results: Square-root transformed posterior mean intensity for MS subtype PPMS.	147
B.9	BENEFIT results: Traceplots for parameters of the imHPGRF model.	149
B.10	BENEFIT results: Autocorrelation plots for parameters of the imH-PGRF model.	150
B.11	BENEFIT results: Square-root transformed posterior mean intensity for CIS.	151
B.12	BENEFIT results: Square-root transformed posterior mean intensity for CDMS.	152

Acknowledgments

I would like to express my heartfelt gratitude and appreciation to my supervisor Tom Nichols. His excellent support and guidance, his ability to motivate, his deep interest in and knowledge of neuroimaging statistics, combined with lots of patience and a sense of humour have helped me in countless ways over the course of my Ph.D. and, ultimately, to become a better researcher.

I want to thank Tim Johnson for numerous interesting discussions and for sharing his wide-ranging insights of various areas of statistics.

I am especially grateful to Ernst-Wilhelm Radü, Jens Würfel and the Medical Image Analysis Center in Basel for financially supporting and thereby allowing me to carry out this work. I would like to thank Kerstin Bendfeldt for a great research collaboration and everybody at MIAC for their hospitality and the welcoming atmosphere every time I came to visit.

Many thanks are due to my dear friends and colleagues, especially Silvia and Pantelis and everybody else in Tom's group. Our discussions whether about research or other matters are invaluable to me. The same holds true for my office mates Ellen, Ed, Fede, Joe and Illiana with whom I had the joy and privilege to share an office for several years. I want to thank everybody at the Centre for Complexity Science for creating and enabling such an interesting, supportive and joyful research environment.

My deepest gratitude goes to my parents, Gerhard and Manuela, and my wonderful sister Lisa, for all their love, support and encouragement. I would not be who I am today without them. Finally, I am deeply grateful to Joana for her love, her understanding, and her friendship. *Joana, eu gosto muito, muito de ti!*

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- The GeneMSA and BENEFIT data sets used in Chapter 4 and Chapter 6 were provided by the Medical Image Analysis Center in Basel, Switzerland.
- The implementation of the LGCP model in Chapter 4 was done by Timothy Johnson. The implementation of the BSGLMM model in Chapter 4 was in large parts done by Tian Ge. The work in Chapter 4 has been published in Taschler et al. [2014].

Abstract

Over the past three decades neuroimaging techniques in general and magnetic resonance imaging (MRI) in particular have made large contributions to the understanding of human brain function and to the diagnosis and treatment of neurological diseases. One area of wide-spread clinical use of MRI is in the assessment of multiple sclerosis (MS). MS patients present with lesions – areas of decreased neuronal conductivity, akin scarred tissue – that occur across the brain and spinal cord. There has been growing interest to use quantitative measures of lesion incidence, exact lesion location and the shape of lesions in the analysis of MRI-based lesion data. Our objective is to address some of the limitations of current methods which rely on particular assumptions about the data and mostly ignore any spatial correlation and structure in the data. In this work we explore several ways of incorporating multiple sources of information into models that can be used for classification and prediction purposes. We compare and assess different machine learning and Bayesian spatial models in a classification task based on MS lesion data. Furthermore, we propose an extended doubly-stochastic spatial point process model based on Gamma random fields that includes non-imaging data as well as location-specific measures attached to xyz -coordinates, and use both simulated and real data to evaluate our methods.

Abbreviations

BSGLMM Bayesian spatial generalised linear mixed model

CDMS clinically definite multiple sclerosis

CIS clinically isolated syndrome

GM grey matter

HPGRF hierarchical Poisson/Gamma random field

imHPGRF intensity-marked hierarchical Poisson/Gamma random field

LGCP log-Gaussian Cox process

MCMC Markov chain Monte Carlo

MRI magnetic resonance imaging

MS multiple sclerosis

PGRF Poisson/Gamma random field

PPMS primary-progressive multiple sclerosis

PRMS progressive-relapsing multiple sclerosis

RRMS relapsing-remitting multiple sclerosis

SPMS secondary-progressive multiple sclerosis

SVM support vector machine

WM white matter

ABBREVIATIONS

The facts of science, as they appeared to him [Heraclitus], fed the flame in his soul, and in its light, he saw into the depths of the world by the reflection of his own dancing swiftly penetrating fire.

—Bertrand Russell, *Mysticism and Logic* (1918)

I am a brain, Watson. The rest of me is a mere appendix. Therefore, it is the brain I must consider.

—Arthur Conan Doyle, *The Adventure of the Mazarin Stone* (1921)

Chapter 1

Introduction

Magnetic resonance imaging (MRI) has transformed the study of the human brain, and especially brain diseases, since the first human scan was carried out in 1977. MRI relies on differences in the magnetic properties of molecules that make up brain tissue. It is highly effective at detecting small anatomical structures and changes in the physiology of brain tissue. The availability and wide-spread use of MRI in the clinical setting continues to have a significant impact on the detection and management of chronic diseases such as Alzheimer’s [Frisoni et al., 2010], autism spectrum disorders [Chen et al., 2011] and multiple sclerosis (MS) [Bakshi et al., 2008]. Measures of abnormalities derived from structural MRI are useful in the context of early diagnosis, treatment planning and monitoring of disease progression.

Multiple sclerosis is characterised by the presence of lesions (also called plaques) in the brain or spinal cord of a patient. MS lesions constitute small areas where the MRI intensity differs from the surrounding, normal-appearing tissue. Lesions indicate an acute inflammation or permanent tissue damage leading to the impediment of the transmission of electrical signals along neurons. In the context of MS, MRI data is to a large extent only used in a qualitative way. Typically, a neuroradiologist will study the images, slice by slice, and qualitatively grade the number, size and location of lesions. This information and an examination of the patient is used with clinical guidelines to diagnose and assess the disease. The only quantitative measure that may be used is a lesion count, based on visual inspection. However, there is increasing interest to use image analysis techniques to provide more quantitative measures of lesion count and total lesion volume (also called lesion “load”). Several studies have had moderate success using lesion load for classification and prediction tasks [Aban et al., 2008; Morgan et al., 2010; MacKay Altman et al., 2011]. Other studies have shown that conventional MRI measures have rather

low predictive value and therefore are poor indicators for determining the clinical outcomes in MS [Loevblad et al., 2010].

Existing quantitative methods that are commonly used for the analysis of MS lesions are, to a large extent, unable to exploit the richness of the imaging data. A prominent method is based on the analysis of lesion probability masks that are compared either cross-sectionally or longitudinally [Filli et al., 2012; Holland et al., 2012] which makes it difficult to associate lesion locations with certain covariates of interest. Furthermore, mass-univariate methods such as voxel-based lesion-symptom mapping [Bates et al., 2003] are ill suited for the binary nature of lesion data and cannot account for the underlying spatial structure. Finally, smoothing kernels are commonly used to introduce a degree of spatial correlation in the data. However, the smoothing of lesion masks by means of a Gaussian kernel [Kincses et al., 2011] does not entirely eliminate the non-Gaussian nature of the data and requires an arbitrary choice of smoothing parameter.

Motivated by these challenges, this thesis explores ways of increasing the utilisation of available imaging data by combining quantitative measures derived from MRI with clinical information. A particular focus of this work lies in the use of statistical methods that rely upon or at least are informed by spatial data. We are interested in the analysis of spatial point patterns that are a result of MRI data and propose a Bayesian spatial point process approach that uses locations of individual lesions to model point pattern data. We are drawn to the Bayesian approach due to the very sparse, high-dimensional nature of the neuroimaging data. Bayesian methods also have the advantage of being able to incorporate prior knowledge, reducing the complexity of the analysis and offering greater precision when sample sizes are low.

This work offers several novel contributions. We make a direct comparison of point process and machine learning methods for predicting disease subtype, finding the point process model has superior performance even when based on much less rich data. We extend the hierarchical Poisson/Gamma random field model to allow for covariates and marks. We develop all of this work in the setting of three-dimensional, multiple-class, multiple-realisation point process data, conduct thorough evaluations on simulated data and apply the methods on two real datasets.

The thesis is organised as follows: In Chapter 2 we review the fundamentals of structural magnetic resonance imaging and provide a medical background to multiple sclerosis, including disease pathology, diagnosis, clinical subtypes, treatment and the use of MRI in clinical assessments. The chapter further introduces the theory of spatial point processes, with an emphasis on Cox processes in general and a

formulation based on a Poisson/Gamma random fields, a process derived from the convolution of a kernel with a Gamma random field, in particular. The chapter continues with a brief discussion of classification and prediction procedures with respect to cross-validation and concludes with the presentation of two clinical data sets that form the basis for the application of our models.

Chapter 4 concentrates on four different approaches to classification and prediction of MS lesion data. The four models include a naïve Bayesian classifier, a support vector machine approach based on a large feature set that encompasses clinical as well as imaging data with a focus on geometric measures of individual lesions, a Bayesian spatial generalised linear mixed model that relies on a probit regression framework and, finally, a log-Gaussian Cox process model fitted to the coordinates of the centres-of-mass of individual lesions.

Our main spatial point process model is a Poisson/Gamma random field model that builds upon the work by Wolpert and Ickstadt [1998b] and Kang et al. [2014a] and is presented in Chapter 5. In particular, we outline a hierarchical formulation of the model that allows for the sharing of information between distinct but related point processes, and then propose ways of extending the model. First, we introduce external covariates, which are specific to a given realisation of a point pattern. A second extension combines the spatial point process model with an additional mark process that carries location-specific attributes of each observed point. We further apply and compare different model variants in a simulation study.

Chapter 6 focuses on the application of our spatial point process models to two clinical data sets of patients with different types or stages of MS. Model validation and assessment is done by means of cross-validation using a Bayesian classification/prediction scheme as well as posterior inference on quantities of interest. Finally, Chapter 7 concludes with a brief summary of the work presented and some remarks on future work.

Chapter 2

Background

In this chapter we lay out the background material relevant to the subsequent chapters of the thesis. Section 2.1 covers the technical details of magnetic resonance imaging, including image acquisition and processing methods. Section 2.2 provides medical background on multiple sclerosis. Section 2.3 summarises the statistical methodology underlying the spatial point process models and methods presented later on in Chapter 5. In Section 2.4 we review some aspects of classification including cross-validation and introduce a Bayesian importance sampling algorithm. Finally, Chapter 3 introduces the two MS data sets that are the basis for model applications in Chapter 4 and Chapter 6.

2.1 Neuroimaging: Magnetic Resonance Imaging

In magnetic resonance imaging (MRI) of the brain, a broad distinction is made between functional and structural MRI. Functional MRI (fMRI) is used as a physiological measure that tracks rapidly changing oxygen levels in the brain's blood flow. The technique is widely used to investigate neural activity in response to various tasks. On the other hand, structural MRI (sMRI) is used to visualise details of the anatomy which are unchanging on a short time scale. Due to the necessity of capturing the brain's activity within a limited time frame, the acquisition time of fMRI is significantly shorter resulting in a lack of spatial resolution compared to structural images.

The advance of non-invasive imaging techniques such as MRI has had a large impact on the study of the human brain. Functional MRI in particular has become ubiquitous in the popular press for studies of the brain. To give only a few examples,

researchers have investigated the brain's activity when being in love [Ortigue et al., 2010], while composing music [Lu et al., 2015] or reading a suspense-novel [Lehne et al., 2015] or simply while being at rest [Cole et al., 2010; Lee et al., 2013]. However, structural MRI is the mainstay in clinical settings. It has become an important tool to detect and visualise abnormalities in the physical appearance of the brain and to track changes over time. Structural MRI is highly effective at identifying lesions and therefore is widely used in the diagnosis and assessment of diseases like multiple sclerosis, Alzheimer's, epilepsy and schizophrenia.

MRI is by no account the only imaging technique that is currently used in a clinical context, but for neurological disorders it is often the modality of choice. Other procedures used include (i) Computed Tomography (CT) [Bar-Shalom et al., 2003] which has fine spatial resolution but utilises x-rays and therefore has the inherent disadvantage of exposure to ionising radiation. (ii) Positron Emission Tomography (PET) [Alauddin, 2012], which uses gamma rays emitted by an injected radioactive tracer, can have exquisite biological specificity but has relatively poor spatial resolution and, of course, also entails exposure to ionising radiation. (iii) Electroencephalography (EEG) measures electric currents on the surface of the scalp which are a result of coherent electrical activity in the brain; EEG has good temporal, but poor spatial resolution which limits its usefulness for clinical applications. (iv) Magnetoencephalography (MEG) is similar to EEG but captures magnetic fields emitted by the brain's electrical activity [Sakkalis, 2011]. (v) Less wide-spread techniques include Near-Infrared Spectroscopy (NIRS) which uses light to detect alterations in the concentration of oxygenated and de-oxygenated haemoglobin [Murkin and Arango, 2009], and Nuclear Magnetic Resonance (NMR) spectroscopy [Neema et al., 2007].

Magnetic resonance imaging manipulates the alignment and subsequent relaxation of nuclear spins using a strong pulsed external magnetic field. There are two basic types of imaging sequences that each reflect different relaxation properties of the examined tissue after the nuclear spins have been aligned as a result of the external magnetic field: T_1 -weighted MRI uses the spin-lattice relaxation time after which the longitudinal component of the nuclear spins have fully returned to their equilibrium orientation. The repetition time (TR) between consecutive pulses of the external magnetic field must be short in order to achieve T_1 weighting. Different tissues exhibit different relaxation times. Most importantly, the T_1 relaxation time for water (mainly protons) is about five times larger than for fatty tissue. In contrast, T_2 -weighted scans exploit the spin-spin relaxation time, i.e. the time it takes for the transverse component of the signal to decay to $1/e$ of its initial value.

This is achieved by using a long echo time (TE), that is the time between the radio-frequency pulse and the echo signal, in the scanning protocol. A full review of the physical principles of magnetic resonance imaging is beyond the scope of this thesis. An introductory treatment of the basics of MR physics can be found in Pooley [2005]. For an excellent, detailed review, see for example Currie et al. [2013].

In addition to traditional MRI scans, a standard procedure for MS patients involves the use of a contrast agent, such as Gadolinium, to improve the contrast between lesions and normal brain tissue. Gadolinium is used in combination with T₁-weighted MRI to enhance visibility of acute inflammation which is associated with breaches in the blood-brain-barrier [Filippi et al., 1996].

2.1.1 Preprocessing steps

After acquisition of an MRI scan, the data needs to be processed in order to be used for any statistical analysis, especially when multiple scans and/or multiple subjects are involved in the study. Several steps are commonly used to reduce artefacts, align and standardise the data. Due to the fact that, in most acquisition types, the full MRI is acquired slice-by-slice, time-dependent signals such as BOLD (blood-oxygen-level dependent) contrast imaging, require slice timing corrections. This is a standard preprocessing step for fMRI data, for example.

Spatial realignment is necessary to correct for head motion during a serial image acquisition like fMRI and typically involves a rigid body transformation of the whole image. Furthermore, spatial co-registration to a specific subject or equivalently spatial normalisation to a common brain template is needed in order to align data from multiple subjects who naturally differ in brain size and shape [Friston et al., 2007]. A standard target space for spatial normalisation is provided in form of the atlas of the Montreal Neurological Institute (MNI). Additionally, spatial smoothing of the data with a Gaussian kernel is often used in an attempt to correct still existing misalignments after spatial normalisation. Smoothing decreases the spatial resolution but can increase the signal-to-noise ratio and the Gaussianity of the data.

Orientation and labelling conventions for MRI scans are displayed in the left panel of Figure 2.1, showing the three standard anatomical views: coronal, sagittal and axial. The coronal plane separates front and back (anterior and posterior, labelled A and P in the image). The sagittal plane divides the brain into left and right and the axial plane separates top and bottom (superior and inferior).

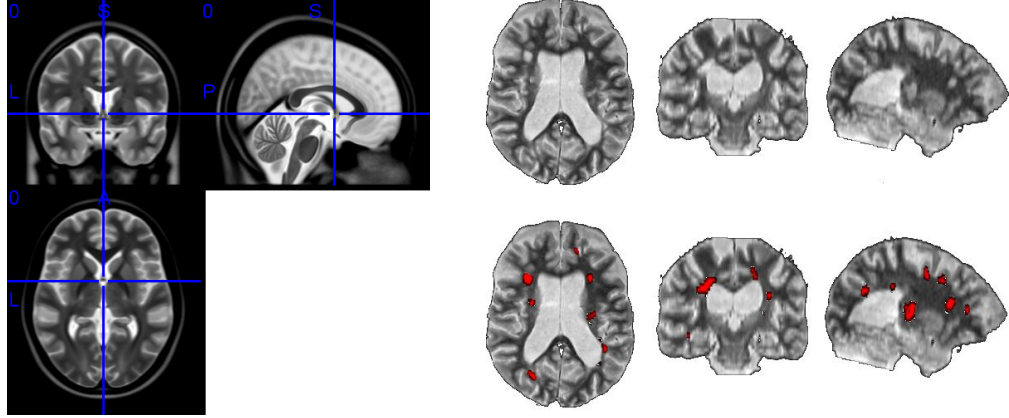


Figure 2.1: *Left panel:* Example MRI of a standard brain template, showing coronal (upper left), sagittal (upper right) and axial (bottom left) views. The crosshairs indicate the origin in MNI space. *Right panel:* Example MRI of an MS patient with and without colouring of lesions on a T₂-weighted scan. The subject shown is part of the GeneMSA data set.

2.2 Medical background: Multiple Sclerosis

Neuroanatomy distinguishes between three main physiological components of the brain: white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF). White matter consists of myelinated axons and forms the connections between various areas of the brain. Its main function is to transmit electrical impulses. On the other hand, GM consists predominantly of neuronal cell bodies and unmyelinated neurons and makes up most of the cortical structure where brain functions such as sensory and motor areas are located. The CSF plays a role in protecting the brain from concussions as well as in regulating the cerebral blood flow. A typical neuron is composed of a cell body (soma), dendrites which act as receiver of electrical signals from other neurons and an axon which transmits impulses from the cell body to other areas of the brain. Axons are also called nerve fibres and are surrounded by a myelin (a kind of lipid) sheath.

Multiple Sclerosis (MS) is a chronic inflammatory and demyelinating disease of the central nervous system that leaves behind scarred tissue in affected regions of the brain and the spinal cord. MS predominantly occurs in the brain's white matter which provides the connections between neurons located in the grey matter. The disease causes damage to the myelin sheath that insulates the nerve fibres and aides the transmission of electric impulses across the brain as well as between the brain and other parts of the body. The neurological damages manifests in form of lesions that can occur across the brain and spinal cord. Lesions are areas of decreased

neuronal conductivity which can take the form of temporal, acute inflammation or chronic tissue damage. Depending on which nerves are affected the symptoms vary considerably from patient to patient, including loss of vision, deprivation of motor and sensory functions, fatigue, cognitive impairment, digestive and sexual dysfunction, mood disorders and chronic pain; for a detailed medical description of MS see for example [Cohen and Rae-Grant, 2010].

2.2.1 Disease pathology

The course of the disease is characterised by episodes of acute relapses where the disease shows high activity and periods of remission during which patients present no or significantly weaker clinical symptoms. In later stages, disease pathology usually takes on a progressive form with increasing severity of symptoms and permanent disability after ten to fifteen years [Cohen and Rae-Grant, 2010].

In general, lesions can be found throughout the brain with a higher disposition in white matter regions around the ventricles. A typical example of lesions found on magnetic resonance imaging are shown in Figure 2.1. Recent studies have also stressed the importance of demyelination and brain atrophy in grey matter [Bendfeldt et al., 2012]. Lesions tend to have an ovoid configuration and in early stages of the disease the lesions are typically thin and elongated.

The pathogenesis of MS is only partially understood. There is no clinical consensus as to what causes the disease. Both T cell and B cell mechanisms¹ have been implicated as well as genetics (e.g. people of Northern European descent appear to be at higher risk [Goldenberg, 2012]) and environmental factors [Compston and Coles, 2008]. However, a recent meta-analysis by Belbasis et al. [2015] has shown that many studies relating MS to environmental risk factors may be inconclusive or even invalid, finding evidence only for the following factors: a biomarker of the Epstein-Barr virus, infectious mononucleosis and smoking. Furthermore, pathogenesis also differs significantly between individuals [Morales et al., 2006]. One of the few common aspects across patients is that early stages of the disease are in most cases characterised by acute inflammatory activity, whereas permanent neurodegeneration is dominant in late and purely progressive MS [Cohen and Rae-Grant, 2010] with symptoms mainly due to axonal loss [Ge, 2006].

Disease burden is often quantified in terms of lesion load, which indicates the total lesion volume or number of lesions. A more general quantitative measure is

¹B lymphocytes and T lymphocytes are types of white blood cells. T cells can be distinguished by the presence of a T-cell receptor.

brain atrophy, which is considered to describe the net accumulative disease burden as the consequence of all types of pathological processes found in the brain. Furthermore, women are more often affected by the disease with a predominance of about 3:2 over male patients [Cohen and Rae-Grant, 2010].

2.2.2 Diagnosis and MS subtypes

First diagnostic criteria for multiple sclerosis were introduced by Schumacher et al. [1965] and two decades later refined by Poser et al. [1983]. The most comprehensive clinical criteria were introduced by McDonald et al. [2001] and are still called the “McDonald criteria”. These guidelines set the standard for the diagnosis of MS. According to the McDonald criteria, patients were grouped into five distinct clinical categories (phenotypes) depending on disease progression, which differs significantly between these subtypes of MS. The subtypes are: relapsing-remitting MS (RRMS), primary-progressive MS (PPMS), secondary-progressive MS (SPMS), progressive-relapsing MS (PRMS) and clinically isolated syndrome (CIS). The McDonald criteria have been updated in 2005 [Polman et al., 2005] and again in 2010 [Polman et al., 2011] to provide better measures of dissemination of lesions in time.

The McDonald criteria are designed for use by a practising physician. The most current revision of the criteria occurred in 2013 and was published by [Lublin et al., 2014]. One of the most important outcomes of the revised guidelines was that PRMS, which affected about 5% of MS patients, was dropped as a distinct clinical category. It was used to describe a disease course that is characterised by regular relapses and a steady progression in symptom severity starting already at disease onset. Additionally, the authors of the revised criteria have put a stronger emphasis on disease activity, measured by clinical relapse rate and MRI findings as well as disease progression, in their recommendations of clinical assessment of MS.

Among the four remaining subtypes, relapsing-remitting MS (RRMS) is the most common type of multiple sclerosis, with about 70–80% of patients belonging to this category when first diagnosed [Cohen and Rae-Grant, 2010]. Over the course of the disease, most RRMS patients will eventually develop into the secondary progressive type. Approximately 15–20% of patients show a gradually progressive course already at disease onset [Cohen and Rae-Grant, 2010; Hurwitz, 2009]. This is the primary-progressive type of MS (PPMS). In contrast to PPMS, where well-defined episodes of acute disease activity are absent, RRMS is characterized by recurrent attacks (so-called relapses), which can vary in frequency and severity. In between relapses, patients display a stable baseline that can be free of any symptoms. How-

ever, a relapse may never completely revert to normal and many patients are left with a residual disability.

Secondary-progressive MS (SPMS) exhibits relapses as well, however, remission occurs only partially and the severity of symptoms increases gradually over time. SPMS is most typically seen 10–15 years after the onset of RRMS [Cohen and Rae-Grant, 2010], with 90% of patients initially diagnosed with RRMS eventually developing into SPMS within 25 years [Ebers, 2001]. There is no distinct transition between the two forms of MS. The relapse rate tends to decrease in SPMS with additional incremental progression between relapses. Furthermore, lesion load tends to be higher in SPMS than in RRMS [Rovaris et al., 2006].

The last disease subtype encompasses clinically isolated syndromes (CIS). In the absence of any prior diagnoses and clinical symptoms, the first acute attack is referred to as a clinically isolated syndrome [Hurwitz, 2009]. The neurological episode is a sign of acute inflammation and must be consistent with demyelination in the central nervous system to be categorised as CIS [Freedman et al., 2014]. Patients with CIS may or may not subsequently develop one of the other courses of MS. In the 2013 revisions of the McDonald criteria, CIS has been recognised as the first clinical presentation of a disease that could be MS. For it to be diagnosed as clinically definite MS (CDMS), dissemination of demyelinating events in time must also occur [Lublin et al., 2014]. A schematic comparison of disease progression for different subtypes is given in Figure 2.2.

In addition, the 2014 guidelines for the first time included the category of radiologically isolated syndrome (RIS). It describes the case where the patient presents no clinical signs or symptoms but incidental imaging findings suggest inflammatory demyelination. RIS is not considered an MS subtype per se since clinical evidence of a demyelinating disease, which is a key criterion in current MS diagnosis, is absent. However, based on location and morphology of lesions as detected on MRI, RIS may suggest the presence of MS [Lublin et al., 2014].

In order to characterise the disease, several clinical outcome measures are used in practice. The Expanded Disability Status Scale (EDSS) [Kurtzke, 1983; Whitaker et al., 1995] was established more than 30 years ago and is widely used but has been criticised for not being responsive or sensitive enough, especially with respect to changes over time. Furthermore, the score is based on a standard – but ultimately subjective – neurological exam which can lead to large variability between raters [Cohen et al., 2012]. The EDSS score is predominantly used in clinical trials to assess neurological disability on a scale of 0–10 in increments of 0.5 and includes

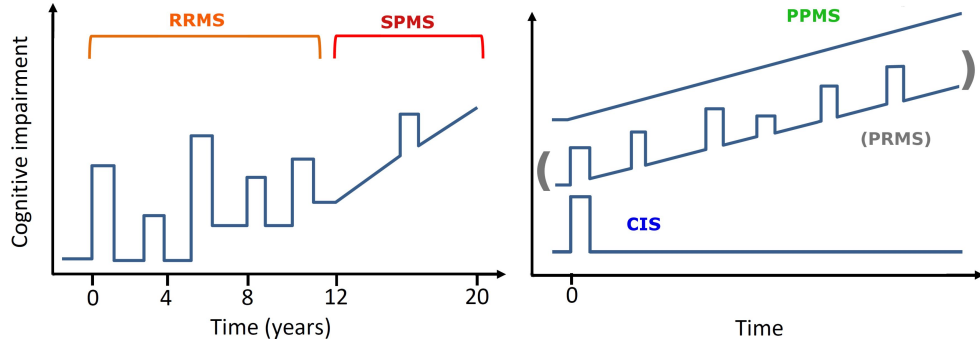


Figure 2.2: Schematic depiction of disease progression for the four clinical subtypes of multiple sclerosis. *Left panel:* Relapsing-remitting MS (RRMS) and secondary-progressive MS (SPMS). Note that the transition from RRMS to SPMS is fluid and occurs typically 10 to 15 years after disease onset. *Right panel:* Clinically isolated syndrome (CIS), primary-progressive MS (PPMS) and progressive-relapsing MS (PRMS). PRMS was dropped in the 2013 revisions of the McDonald criteria. (Figure adapted from Cohen and Rae-Grant [2010]).

several sub-categories that are associated with various functional systems, such as visual (VSLSC), sensory (SENSSC), mental (MNSC), pyramidal (PYRSC, measures the ability to walk), cerebellar (CRBLSC, coordination), brain stem (BRSTMSC), and bowel and bladder (BWLSC) functions.

The paced auditory serial addition task (PASAT) [Gronwall, 1977] is a cognitive test, where a lower score indicates more severe impairment. It measures processing speed, working memory and arithmetic ability. PASAT is reasonably reliable but limited in scope and not specific to MS. New measures have been proposed, including, for example, patient-reported outcomes or biomarker measurements [Cohen et al., 2012].

2.2.3 Therapy

An early and accurate diagnosis is crucial for the efficacy of disease-modifying therapy. There is no curative treatment available for MS, however, early therapy is likely to reduce the frequency of relapses and slow the progression to more severe disability [Loevblad et al., 2010]. In addition to the treatment of symptoms, disease-modifying therapy is aimed at lowering MRI activity (i.e appearance of new lesions or growth of existing lesions) and reducing the risk of permanent disability [Cohen and Rae-Grant, 2010]. Among the four subtypes, disease-modifying therapy predominantly targets the RRMS course. Since the early 1990s several agents have become available: Glatiramer acetate, Interferon beta, Mitoxantrone, Natalizumab,

Fingolimod, Teriflunomide, Dimethyl fumarate, and Alemtuzumab as well as a considerable range of off-label treatment options. All approved agents try to reduce acute inflammation and appear to be more effective in early stages of the disease [Torkildsen et al., 2016]. Therapy, when effective, is able to prolong the time to subsequent attacks, reduce inflammation and formation of new lesions, and can help to slow disability progression and cognitive impairment.

2.2.4 MRI criteria

Neuronal damage develops into lesions which can then be seen on magnetic resonance imaging. Although MRI is the main source of quantitative data for the diagnosis of MS, the correlation between lesions and clinical symptoms is only described as approximate. The reason for this may lie in neural plasticity and repair mechanisms that can compensate for damage. Therefore, MRI scans may not fully reflect functional changes in affected areas [Cohen and Rae-Grant, 2010].

Conventional MRI sequences that show different aspects of disease activity and for which MRI criteria have been developed include T₂-weighted and T₂-weighted fluid-attenuated inversion recovery (FLAIR) images, as well as T₁-weighted scans, both with and without the addition of a contrast agent such as Gadolinium (Gd). T₁-weighted sequences display chronic or persistent hypo-intense lesions. These so-called “black holes” indicate permanent axonal loss and neuronal damage. However, studies about the correlation of black holes and disability have been contradictory and the relationship with relapse rates is similarly inconsistent [Sahraian et al., 2010]. In Gd-enhanced T₁ images, the contrast agent Gadolinium is used to augment the appearance of inflammatory lesions. Gd-enhanced images show breaches of the blood-brain barrier that accompany acute MS attacks, for example during a relapse, revealing regions where the disease is presently active. Finally, T₂-weighted scans are used to assess the cumulative disease burden, i.e. total lesion load. Lesions appear hyper-intense on T₂-weighted images. These sequences are also used to detect the formation of new lesions. However, T₂-weighted sequences lack specificity as several other mechanisms such as edema, inflammation and gliosis (an increase in the number of a certain kind of neuroglial cell) add to the development of T₂ lesions [Sahraian et al., 2010]. A summary of the different types of MS lesions is provided in Table 2.1.

Following the 2001 revisions, the McDonald criteria incorporate data obtained from magnetic resonance imaging to aid clinical diagnosis. However, this

Table 2.1: Summary of MS lesion types.

Type	Appearance on MRI	Characteristics
T ₁	hypo-intense (dark)	chronic, persistent “black holes”, indication of permanent neuronal damage
T ₂	hyper-intense (bright)	used for assessment of total lesion burden
T ₁ -Gd	contrast enhanced (bright)	“active” lesions, indication of acute inflammation

data is to a large extent only used in a qualitative way, assessing existence and general dissemination of lesions in space and time. For example, the diagnostic standard merely states that the presence of lesions in at least two of the typical anatomical regions (periventricular, juxtacortical, infratentorial areas) of the central nervous system or spinal cord is evidence for MS. The most common quantitative measure is lesion load, i.e. the total lesion volume. Previous work on using lesion load for classification have been equivocal [Aban et al., 2008; Morgan et al., 2010; MacKay Altman et al., 2011]. Other studies have shown that conventional MRI measures have rather low predictive value and are therefore poor indicators for determining the clinical outcomes in MS [Loevblad et al., 2010]. There exists a need for quantitative MRI measures, for example to better understand the transition from RRMS to SPMS [Rovaris et al., 2006], and a more quantitative analysis of the data contained in and available through magnetic resonance imaging.

2.2.5 Quantitative analysis

Existing quantitative methods that are commonly used for the analysis of MS lesions are to a large extent unable to reflect the full complexity of the data. These include i) comparing lesion probability masks cross-sectionally or longitudinally [Filli et al., 2012; Holland et al., 2012] which makes it difficult to associate lesion locations with certain covariates of interest, ii) mass-univariate methods such as voxel-based lesion-symptom mapping [Bates et al., 2003] which cannot account for the underlying spatial structure, and iii) smoothing of lesion masks by means of a Gaussian kernel, see e.g. [Kincses et al., 2011], which does not entirely eliminate the non-Gaussian nature of the data and requires an arbitrary choice of smoothing parameter.

These limitations are the core motivation for this work, looking to apply explicit spatial models to MRI-based lesion data.

2.3 Methodological background: Spatial point processes

This section aims to provide some of the methodological background for the models presented in Chapter 5.

Statistical methodology for point pattern data advanced rapidly in the second half of the 20th century. One-dimensional non-homogeneous Poisson processes have been used to model, for example, geomagnetic reversal data, the occurrence of coal-mining disasters and arrivals at an intensive care unit (e.g. Lewis and Shedler [1979] and references therein). Ongoing research and applications can be found across many disciplines, including ecology and forestry [Ickstadt and Wolpert, 1999; Best et al., 2000, 2002; Stoyan and Penttinen, 2000; Niemi and Fernández, 2010; Woodard et al., 2010], geology, astronomy, material sciences and epidemiology [Velázquez et al., 2016]. Research interests focus on diverse topics such as the distributions of plant species [Illian et al., 2013; Renner et al., 2015], the occurrence of earthquakes over a certain period of time [van Lieshout and Stein, 2012], and astronomical data such as the distribution of stars or galaxies [Babu and Feigelson, 1996; Stoica, 2010].

Most of these example applications rely on summary statistics to analyse point patterns. See the extensive literature on summary statistics for point process applications, for example, Baddeley et al. [2000]; Gabriel and Diggle [2009]; Cronie and Van Lieshout [2015]; Cronie and van Lieshout [2016]. However, our focus of interest lies on estimating an underlying intensity function that is able to fully describe the generation of observed point patterns.

In the following, we give a short overview of key concepts and topics relating to the theory of spatial point processes with a particular emphasis on a Bayesian point of view. The main source for this is Møller and Waagepetersen [2004]. Most of the following material is covered in much greater detail in Daley and Vere-Jones [2003]; Daley and Vere-Jones [2008]; Illian et al. [2008]; Møller and Waagepetersen [2007]; Gelfand et al. [2010] and Diggle [2014].

A spatial point process \mathbf{Y} is a random countable set of points $\{y_i\}_{i=1}^{\infty}$ in a space \mathcal{S} . We only consider Euclidean space here, i.e. $\mathcal{S} \subseteq \mathbb{R}^d$. Typically, the region of interest for realisations of the point process will be a bounded subset $\mathcal{B} \subset \mathcal{S}$, for example square kilometre of forest, a geographical region defined by political or natural boundaries, or a section of the night sky. For our purposes, $d=3$ and \mathcal{B} represents the three-dimensional volume of the human brain. Furthermore, if the points in \mathbf{Y} are unique, the process is called simple.

2.3. Methodological background: Spatial point processes

Let A be a bounded Borel set and let $N_{\mathbf{Y}}(A)$ denote the number of points that fall into A . $N_{\mathbf{Y}}(A)$ is the counting measure, i.e. a finite integer-valued measure, on A . Let \mathbf{y} denote a single realisation of \mathbf{Y} consisting of point locations y_i . We require that \mathbf{y} is a locally finite subset of \mathcal{B} , that is the cardinality $n(\cdot)$ of the set $\{\mathbf{y} \cap A\}$ is finite. The space of locally finite point configurations N_{lf} provides the range of \mathbf{Y} , i.e.

$$N_{lf} = \{\mathbf{y} \subseteq \mathcal{B} : n(\mathbf{y} \cap A) < \infty\}, \forall A \subseteq \mathcal{B}. \quad (2.1)$$

There are three distinct but equivalent ways to fully characterise a spatial point process: by its void probabilities (only for simple processes), its finite dimensional distributions and its generating functional.

The finite dimensional distributions are the joint distribution of the number of points of \mathbf{Y} belonging to bounded Borel sets A_1, \dots, A_k , such that

$$\mathbb{P}[N_{\mathbf{Y}}(A_1) = n_1; \dots; N_{\mathbf{Y}}(A_k) = n_k], \quad n_1, \dots, n_k \in \mathbb{N}_0. \quad (2.2)$$

The void probabilities are defined as $v(A) = \mathbb{P}[N_{\mathbf{Y}}(A) = 0]$ and the generating functional for a point process is akin to a probability generating function for a non-negative, integer-valued random variable and is defined as

$$G_{\mathbf{Y}}(u) = \mathbb{E} \left[\exp \left\{ \int_{\mathcal{S}} \ln u(\xi) dN(\xi) \right\} \right] = \mathbb{E} \left[\prod_{\xi \in \mathbf{y}} u(\xi) \right], \quad (2.3)$$

where u is a function from \mathcal{S} to $[0, 1]$ and ξ is a point in \mathcal{S} such that $u(\xi) \leq 1$. Consider as an example $u(\xi) = t^{\mathbb{I}[\xi \in A]}$ with $0 \leq t \leq 1$, then the probability generating functional for $N_{\mathbf{Y}}(A)$ is $G_{\mathbf{Y}}(u) = \mathbb{E}[t^{N_{\mathbf{Y}}(A)}]$.

2.3.1 Spatial Poisson point processes

Poisson point processes are making up the foundations of spatial point process theory. In their simplest variant they describe a random spatial point pattern with no correlation or any kind of interaction between points. Their defining property is that the number of points in any given region is a Poisson distributed random variable. Additionally, given the number of points $N_{\mathbf{Y}}(A)$ that belong to the process \mathbf{Y} and are contained in region $A \subseteq \mathcal{B}$, those points are distributed independently according to a probability density that is proportional to an intensity function λ over that region.

We define the intensity function $\lambda : \mathcal{S} \rightarrow [0, \infty)$ on space $\mathcal{S} \subseteq \mathbb{R}^d$ and

an intensity measure $\Lambda(A) = \int_A \lambda(\xi) d\xi$ that is locally finite, i.e. $\Lambda(A) < \infty$ for all bounded $A \subseteq \mathcal{S}$ and diffuse, i.e. $\Lambda(\xi) = 0$, for all points $\xi \in \mathcal{S}$. A point process \mathbf{Y} on $\mathcal{S} \subseteq \mathbb{R}^d$ is a Poisson point process with intensity function λ and intensity measure Λ if the following two properties hold: (i) For any $A \subseteq \mathcal{S}$ such that $\Lambda(A)$ is finite, $N_{\mathbf{Y}}(A)$ is Poisson distributed with mean $\Lambda(A)$. We write $N_{\mathbf{Y}}(A) \sim \text{Pois}(\Lambda(A))$. (ii) For any disjoint $A_1, \dots, A_k \subseteq \mathcal{S}$, with $k \geq 2$, the random variables $N_{\mathbf{Y}}(A_1), \dots, N_{\mathbf{Y}}(A_k)$ are independent. If for a Poisson process \mathbf{Y} it holds that

$$\Lambda(A) = \int_A \lambda(\xi) d\xi \quad (2.4)$$

with $\Lambda(A) < \infty$ for any bounded $A \subseteq \mathcal{S}$ and λ some non-negative function on \mathcal{S} , then we call \mathbf{Y} a Poisson process with intensity function λ . The intensity is the first-order moment measure and also called the rate function. Analogously, Λ is sometimes called the integrated rate function. We denote $\mathbf{Y} \sim \mathcal{PP}(\mathcal{S}, \lambda)$.

If the intensity function is constant ($\lambda(\xi) = \lambda$, for all $\xi \in \mathcal{S}$), \mathbf{Y} is called a *homogeneous* or *stationary* process with rate λ . The special case of $\mathbf{Y} \sim \mathcal{PP}(\mathcal{S}, 1)$ is called the standard or *unit rate Poisson process*. The void probabilities for a Poisson process are $v(A) = \exp\{-\Lambda(A)\}$ and the probability generating functional is given by $G_{\mathbf{Y}}(u) = \exp\{(1 - u(\xi))\lambda(\xi) d\xi\}$. The density (i.e. the Radon-Nikodym derivative) of the Poisson process with respect to Lebesgue measure does not exist, but it does exist with respect to the unit rate Poisson process $\mathcal{PP}(\mathcal{B}, 1)$ [Møller and Waagepetersen, 2004]. The density is given by

$$\pi(\mathbf{y}|\lambda) = \exp\left\{|\mathcal{B}| - \int_{\mathcal{B}} \lambda(\xi) d\xi\right\} \prod_{y \in \mathbf{y}} \lambda(y), \quad (2.5)$$

which can be interpreted as the density of the sampling distribution of the data.

For most practical applications, a simple Poisson process is not sufficient to describe the observed point patterns. The class of Cox processes is a natural extension to the Poisson process. First introduced by Cox [1955], Cox processes are also called doubly stochastic processes, relating to the way the intensity function is modelled. Whereas the original Poisson process is based on a fixed parametric intensity, in the case of a Cox process, the intensity function is expressed as the realisation of a random variable or a random field.

Hence, let $Z(\xi)$ be a non-negative random field on \mathcal{S} such that $Z(\xi)$ is locally finite and integrable. If $\mathbf{Y}|Z \sim \text{Pois}(\mathcal{S}, Z)$, then \mathbf{Y} is said to be a Cox process driven by Z . The intensity function is simply the expectation over the random field, i.e. $\lambda(\xi) = \mathbb{E}[Z(\xi)]$ and (2.4) still holds for the intensity measure. For simplicity, we

will denote $[\mathbf{Y}|\Lambda] \sim \mathcal{PP}(\mathcal{S}, \Lambda)$ as being a spatial Poisson process driven by random intensity measure Λ [Møller et al., 1998]. Note that for disjoint bounded $A_1, A_2 \subset \mathcal{S}$, $N_{\mathbf{Y}}(A_1)$ and $N_{\mathbf{Y}}(A_2)$ are positively correlated.

Cox processes have the advantage of allowing greater flexibility than, e.g. a parametric inhomogeneous Poisson process, and are better suited to model aggregated point patterns. Additionally, one can incorporate prior knowledge into the intensity function by using a Bayesian framework.

2.3.2 The log-Gaussian Cox process

One of the most commonly used Cox process models is the log-Gaussian Cox process (LGCP) [Møller et al., 1998], where the intensity function is a realisation of a random field Z . Assume that $Z = \ln \lambda$ is a Gaussian random field, that is a random process whose finite dimensional distributions are Gaussian; hence $Z \sim \mathcal{GP}[m(\cdot), c(\cdot, \cdot)]$ [Rasmussen and Williams, 2006]. The distribution of (\mathbf{Y}, Z) is completely determined by the mean and covariance function, which are given by $m(r) = \mathbb{E}[Z(r)]$ and $c(r, s) = \text{Cov}[Z(r), Z(s)]$, respectively. The intensity function of the LGCP is given by

$$\lambda(r) = \exp [m(r) + c(r, r)/2], \quad (2.6)$$

where, for our later applications, we will assume that $c(r, s)$ is isotropic and translation invariant with the form $c(r, s) = \sigma^2 \exp [-\rho \|r - s\|^\delta]$. The exponential term represents the power correlation function and $\delta \in [0, 2]$, which for $\delta=1$ gives an exponential and for $\delta=2$ a Gaussian correlation; σ^2 denotes the variance and $\rho > 0$ is a correlation parameter.

Inference for an LGCP model is typically done under a Bayesian framework and there exist several fast and computationally efficient ways to do posterior inference on an LGCP: Markov chain Monte Carlo techniques [Møller et al., 1998], elliptical slice sampling [Murray et al., 2009], integrated nested Laplace approximations (INLA) [Illian et al., 2012; Simpson et al., 2016], or variational Bayes [Jaakkola and Jordan, 2000]. In Chapter 4, we use an MCMC approach, specifically Hamiltonian Monte Carlo (HMC) [Duane et al., 1987; Neal, 2012]. Several other techniques have been proposed, each of which carries certain advantages and disadvantages. For example, Riemann Manifold Hamiltonian Monte Carlo has been shown in a 2D setting to be a more efficient sampling algorithm [Girolami and Calderhead, 2011]; however, adaptation to 3D problems is prohibitive as it requires the inversion of a large matrix. Taylor and Diggle [2014] have found similar results when comparing the Metropolis-adjusted Langevin algorithm (MALA) [Roberts and Rosenthal,

1998], which uses the gradient of the target distribution to guide a random walk update, to INLA.

2.3.3 Poisson/Gamma random fields

First introduced by Wolpert and Ickstadt [1998b], the Poisson/Gamma random field (PGRF) model was extended by Kang et al. [2014b] to incorporate multiple groups of related point patterns as well as multiple realisations of the underlying stochastic point process.

The fundamental characteristic of the PGRF model is that the intensity function is modelled as a convolution of a spatial kernel and a Gamma random field. In this section we will describe the main building blocks of the model. First, we recount the probability density of the Gamma distribution:

$$h(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu}\right)^\alpha x^{\alpha-1} \exp\left(-\frac{\alpha x}{\mu}\right), \quad (2.7)$$

with $x \geq 0$ and $\alpha, \beta > 0$. Mean and variance of the Gamma distribution are given by

$$\mu = \frac{\alpha}{\beta}, \quad \sigma^2 = \frac{\alpha}{\beta^2} = \frac{\mu^2}{\alpha}. \quad (2.8)$$

A Gamma random field $\Gamma(dx)$ has an inhomogeneous Gamma process distribution, i.e. $\Gamma(dx) \sim \text{Ga}\{\alpha(dx), \beta(x)\}$ with shape measure (also called base measure) $\alpha(dx)$ and inverse scale function (or rate parameter) $\beta(x) > 0$. For $\Gamma(dx)$ to be a Gamma random field the following properties must hold: (i) for some set \mathcal{B} and any partition $\{A_k\}$ such that $\mathcal{B} = \cup_k A_k$, the random variables $\Gamma(A_k)$ follow a Gamma distribution with shape parameter $\alpha(A_k) = \int_{A_k} \alpha(x) dx$ and mean $\alpha(A_k)/\beta(x)$ for any $x \in A_k$; (ii) for disjoint sets $A_k \cap A_l = \emptyset$, $k \neq l$, the Gamma random variables $\Gamma(A_k)$ and $\Gamma(A_l)$ are mutually independent. Note that the shape measure is proportional to the volume and a convenient choice may be to use Lebesgue measure. For the rest of this thesis we take the inverse scale parameter $\beta(x)$ to be constant over space and thus assume $\beta(x) \equiv \beta$.

The Gamma random field can be seen as a set of indicator functions with assigned Gamma distributed random variables. Realisations of Gamma random fields are discrete and consist of a countably infinite number of point masses, so-called jumps. Each jump location θ_m , $m = 1, \dots, \infty$, has an associated magnitude (or jump height) η_m . Then, $\Gamma(A)$ is given by summing up all contributions from jumps belonging to A , i.e. $\Gamma(A) = \sum_{m=1}^{\infty} \{\eta_m : \theta_m \in A\}$. Note that Gamma random

2.3. Methodological background: Spatial point processes

fields are closely related to Dirichlet random fields in so far as Dirichlet random fields are normalised, i.e. given by $\Gamma(A)/\Gamma(\mathcal{B})$.

The Poisson/Gamma random field (PGRF) model relies on a convolution of a Gamma random field with a spatial kernel to describe the underlying Poisson point process \mathbf{Y} . Let $\Lambda(dy)$ denote the intensity measure and let $\lambda(y)$ denote the intensity function associated with \mathbf{Y} . Then $\Lambda(dy)$ is a convolution of a spatial kernel measure $K_{\sigma^2}(dy, x)$ with kernel size σ^2 and a Gamma random field (i.e. a random measure) $\Gamma(dx)$.

In this work we will use a Gaussian kernel with kernel function $k_{\sigma^2}(y, x) = (\frac{1}{\sigma^2})^{d/2} \exp\left\{-\frac{\|y-x\|^2}{2\sigma^2}\right\}$, where $\|y-x\|$ denotes the Euclidean distance in d -dimensional Euclidean space. Suppose a reference measure $\Pi(dy)$, which in our case will be Lebesgue measure, dominates both the kernel measure $K_{\sigma^2}(dy, x)$ as well as the intensity measure $\Lambda(dy)$. We have $K_{\sigma^2}(\mathcal{B}, x) = \int_{\mathcal{B}} k_{\sigma^2}(y, x) dy$ and $\Lambda(\mathcal{B}) = \int_{\mathcal{B}} \lambda(y) dy$.

Conditional on $\Lambda(dy)$, the number of points in a given region A in Euclidean space is a Poisson random variable $N_{\mathbf{Y}}(A)$ with random mean measure $\Lambda(A) = \int_A \lambda(y) \Pi(dy)$, where $\lambda(y)$ is a density with respect to some reference measure $\Pi(dy)$.

An example application in 2D can be found in Best et al. [2000] wherein the authors have used a Poisson/Gamma random field model to analyse the spatial variation of environmental risk factors such as traffic pollution to assess their effect on respiratory disorders in children.

2.3.4 Inverse Lévy measure construction of a Gamma random field

The inverse Lévy measure (ILM) construction provides an efficient method for sampling from independent-increment random fields. For details and a proof, we refer the reader to Theorem 1 and Corollary 2 in Wolpert and Ickstadt [1998b].

The idea of the construction algorithm is based on the Lévy-Khintchine formula for infinitely-divisible distributions. One can use partial sums of independent draws from a standard exponential distribution to generate successive jump heights ν_m of a unit-rate Poisson process. The random measure $\Gamma(dx)$ can be constructed as an infinite sum over latent sources θ_m and associated magnitudes ν_m :

$$\Gamma(dx) = \sum_{m=1}^{\infty} \nu_m \delta_{\theta_m}(dx), \quad (2.9)$$

where $\delta_{\theta_m}(dx)$ stands for Kronecker's delta-function.

The ILM algorithm

The following sampling algorithm is adapted from Wolpert and Ickstadt [1998b]; Wolpert and Ickstadt [1998a]:

1. Fix a large integer M and choose a convenient reference distribution $\Pi(dx)$ on auxiliary space \mathcal{X} such that the shape measure $\alpha(dx)$ has a density $\alpha(x) \equiv \frac{\alpha(dx)}{\Pi(dx)}$.
2. Generate M independent and identically distributed jump locations (latent sources): $\theta_m \stackrel{iid}{\sim} \tilde{\alpha}(dx) = \frac{\alpha(dx)}{\alpha(\mathcal{B})}$.
3. Generate successive jumps of a standard Poisson process on \mathbb{R}^+ : $\zeta_m = \sum_{l=1}^m e_l$, with $e_l \stackrel{iid}{\sim} \text{Exp}(1)$ drawn from a standard exponential distribution.
4. Set jump heights (magnitudes): $\nu_m = \frac{1}{\beta} E_1^{-1}\{\frac{\zeta_m}{\alpha(\mathcal{B})}\}$.
The exponential integral function is given by $E_1(t) = \int_t^\infty e^{-u} u^{-1} du$. Note that $\nu_m = \sup\{u \geq 0 : \zeta(u, \theta_m) \geq \zeta_m\}$.
5. Set $\Gamma(dx) \equiv \sum_{m=1}^M \nu_m \delta_{\theta_m}(dx)$.

The only approximation lies in the fact that M is finite, that is we cannot sample an infinite number of jumps. However, the truncation error can be made arbitrarily small by sampling a large number of latent points. Ideally, $\Pi(dx)$ is chosen in a way such that it is convenient to sample from it. A natural choice, for example, would be Lebesgue measure.

2.3.5 Data augmentation: Auxiliary points

The following data augmentation scheme makes the Gamma random field conjugate and allows for efficient posterior sampling via an MCMC algorithm based on Gibbs sampling [Wolpert and Ickstadt, 1998a,b].

Considering that the distribution of points $N_{\mathbf{Y}}(dy)$ on \mathcal{B} is a finite, integer-valued measure, Wolpert and Ickstadt [1998a] propose to represent $N_{\mathbf{Y}}(\mathcal{B})$ as the sum of a random number of unit point masses. It follows that the locations x_l of these point masses are not necessarily distinct. All data points $\{y_l\}_{l=1}^L \in \mathbf{Y}$ are

2.3. Methodological background: Spatial point processes

mutually independent and follow the same distribution

$$[Y_l = y_l | N_{\mathbf{Y}}(\mathcal{B}), \Gamma(dx), \sigma_j^2] \sim \frac{\Lambda(dy)}{\Lambda(\mathcal{B})} = \frac{\int_{\mathcal{B}} K_{\sigma^2}(dy, x) \Gamma(dx)}{\Lambda(\mathcal{B})}. \quad (2.10)$$

In order to resolve the mixture, for each $y_l \in \mathbf{Y}$ draw an auxiliary random variable $X_l = x_l \in \mathcal{B}$ such that

$$[X_l | Y_l = y_l, N_{\mathbf{Y}}(\mathcal{B}), \Gamma(dx), \sigma^2] \sim \frac{k_{\sigma^2}(y_l, x) \Gamma(dx)}{\lambda(y_l)}. \quad (2.11)$$

The realised values $\{x_l\}$ of conditionally independent random variables $\{X_l\}$ follow discrete distributions

$$\pi(X_l = x_l | N_{\mathbf{Y}}(\mathcal{B}), \Gamma(dx), \sigma^2) = \nu_m k_{\sigma^2}(y_l, \theta_l) / \Lambda(y_l). \quad (2.12)$$

which follows directly from the representation of the intensity measure as $\Lambda(y) = \sum_{m=1}^{\infty} \nu_m k_{\sigma^2} y, \theta_m(dx)$.

The augmented data representation implies that $(\mathbf{X}, \mathbf{Y}) = \{(x_l, y_l), y_l \in \mathbf{Y}\}$ is a Poisson point process on $\mathcal{B} \times \mathcal{B}$ and the joint process is given by

$$[\mathbf{X}, \mathbf{Y} | \Gamma(dx), \sigma^2] \sim \mathcal{PP}\{\mathcal{B} \times \mathcal{B}, K_{\sigma^2}(dy, x) \Gamma(dx)\}. \quad (2.13)$$

2.3.6 The independent PGRF model

Assume there exist a number of J distinct types (or groups) of point patterns, indexed by $j = 1, \dots, J \in \mathbb{N}_0$. Let \mathbf{y}_j denote a single realisation of the point process \mathbf{Y}_j defined on $\mathcal{B} \subset \mathbb{R}^d$, i.e. \mathbf{y}_j is the observed point pattern for type j . We assign a Gamma random field $\Gamma_j(dy)$ to each type j . For now, the $\Gamma_j(dy)$ are independent with common base measure $\Gamma_0(dx)$ and inverse scale parameters β_j . Note that the change in notation is necessary to be consistent with the formulation of the hierarchical model in Chapter 5.

Suppose the intensity measures $\Lambda_j(dy)$ and the finite kernel measures $K_{\sigma_j^2}(dy, x)$ are both dominated by a reference measure $\Pi(dy)$ and take $\Pi(dy)$ to be Lebesgue measure. Hence, the intensity measure of the independent Poisson/Gamma random field (IPGRF) model is given by

$$\Lambda_j(dy) = \int_{\mathcal{B}} K_{\sigma_j^2}(dy, x) \Gamma_j(dx). \quad (2.14)$$

The random mean measure over a subset $A \in \mathcal{B}$

$$\Lambda_j(A) = \int_A \lambda_j(y) \Pi(dy) \quad (2.15)$$

is also defined by the density $\lambda_j(y)$ with respect to some reference measure $\Pi(dy)$. Analogously, the smoothing kernel measure is given by

$$K_{\sigma_j^2}(A, x) = \int_A k_{\sigma_j^2}(y, x) \Pi(dy). \quad (2.16)$$

General model formulation

Each spatial point pattern² is considered as an independent realisation of an individual type of spatial point process. In the case of the IPGRF, the spatial point processes underlying different types are assumed to be completely independent. Hence, let \mathbf{Y}_j be a Poisson point process (a Cox process with realisations $\mathbf{y}_{j,i}$) with intensity measure $\Lambda_j(dy)$ such that

$$\mathbf{Y}_j | \Lambda_j \stackrel{iid}{\sim} \mathcal{PP}\{\mathcal{B}, \Lambda_j(dy)\}, \quad (2.17)$$

$$\Lambda_j(dy) = \int_{\mathcal{B}} K_{\sigma_j^2}(dy, x) \Gamma_j(dx), \quad (2.18)$$

$$\Gamma_j(dx) \sim \mathcal{GRF}\{\Gamma_0(dx), \beta_j\}, \quad (2.19)$$

where \mathcal{GRF} denotes the Gamma random field and $\Gamma_j(dx)$ an independent-increment infinitely-divisible random measure on an auxiliary space \mathcal{X} .

Note that the authors of the original model [Wolpert and Ickstadt, 1998b; Ickstadt and Wolpert, 1999] call this a hierarchical model in the sense that on the first (lowest) stage of the hierarchy the number of points in a given region is modelled as a Poisson random variable with random mean measure as given in (2.15), on the second level the density $\Lambda(dy)$ is modelled as a kernel mixture, and the third stage includes prior distributions on the model parameters.

2.3.7 Multiple realisations

The analysis of neuroimaging data typically involves groups of patients. Taking the example of brain lesions, every individual therefore has their own lesion-based point

²In later chapters, we will consider two real data applications of MS lesions where the locations of individual lesions from different subjects will constitute different spatial point patterns.

pattern. This implies that we have not just one but multiple realisations of the same point process. If we consider different MS subtypes as separate point processes, then each subject's lesion locations constitute one realisation of the point process that is associated with that particular type of MS.

In order to account for multiple independent realisations in our modelling framework, assume that $\{\mathbf{y}_{j,i}\}_{i=1}^{N_j}$ are N_j independent realisations of the same Poisson point process \mathbf{Y}_j . Denote $y_{j,i,l}$ as one observed point (i.e. lesion location) and L_i the number of points in subject i 's data. Thus write $\mathbf{y}_{j,i} = \{y_{j,i,l}\}_{l=1}^{L_i}$ and, for the combined data belonging to group j , $\mathbf{y}_j = \{\{y_{j,i,l}\}_{l=1}^{L_i}\}_{i=1}^{N_j}$.

The data augmentation procedure of Subsection 2.3.5 can easily be modified to incorporate multiple types of point patterns as well as multiple realisations per type. This means that the joint process $(\mathbf{X}_j, \mathbf{Y}_j) = \{(x_{j,i,l}, y_{j,i,l}); y_{j,i,l} \in \mathbf{Y}_j\}$ is a Poisson point process on $\mathcal{B} \times \mathcal{B}$. Therefore, $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_j}$ are N_j independent realisations of $(\mathbf{X}_j, \mathbf{Y}_j)$, each consisting of a set of L_i point-pairs $\{x_{j,i,l}, y_{j,i,l}\}_{l=1}^{L_i}$. The joint process is given by

$$[\mathbf{X}_j, \mathbf{Y}_j | \Gamma_j(dx), \sigma_j^2] \sim \mathcal{PP}\{\mathcal{B} \times \mathcal{B}, K_{\sigma_j^2}(dy, x) \Gamma_j(dx)\}. \quad (2.20)$$

Note that the first marginal of (2.20) recovers the distribution of \mathbf{Y}_j as in (2.17); and similarly, the second marginal gives the distribution of the auxiliary process \mathbf{X}_j .

The full joint distribution

By using the (truncated) inverse Lévy measure construction (presented in Subsection 2.3.4) to establish the Gamma random field $\Gamma_j(dx) = \sum_{m=1}^M \nu_{j,m} \delta_{\theta_{j,m}}(dx)$, the joint distribution for the IPGRF model can then be written as

$$\begin{aligned} & \prod_{j=1}^J \left[\prod_{i=1}^{N_j} \pi(\mathbf{x}_{j,i}, \mathbf{y}_{j,i} | \boldsymbol{\nu}_j, \boldsymbol{\theta}_j, \sigma_j^2) \times \pi(\sigma_j^2) \times \pi(\boldsymbol{\nu}_j | \beta_j) \times \pi(\beta_j) \times \pi(\boldsymbol{\theta}_j) \right] \\ & \propto \prod_{j=1}^J \prod_{i=1}^{N_j} \left[\exp \left\{ - \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_{j,m}) \nu_{j,m} \right\} \prod_{l=1}^{L_{j,i}} \left\{ k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \nu_{j,m} \mathbb{I}_{\theta_{j,m}}(x_{j,i,l}) \right\} \right] \\ & \quad \times \prod_{j=1}^J \left[\exp \{ -E_1(\beta_j \nu_{j,M}) \} \prod_{m=1}^M \frac{1}{\nu_{j,m}} \exp \{ -\nu_{j,m} \beta_j \} \times \pi(\sigma_j^2) \pi(\beta_j) \right]. \quad (2.21) \end{aligned}$$

Here, $\pi(\boldsymbol{\theta}_j)$ is the density of $\boldsymbol{\theta}_j$ with respect to the product of base measures $\prod_m \tilde{\Gamma}_0(dx)$, with $\tilde{\Gamma}_0(dx) = \frac{\Gamma_0(dx)}{\Gamma_0(\mathcal{B})}$. Further, the $\Gamma_0(dx)$ are assumed to be non-

atomic and therefore we have $\pi(\boldsymbol{\theta}_j) = 1$. The densities of all other parameters are with respect to a product of Lebesgue measures. Finally, $\pi(\mathbf{x}_{j,i}, \mathbf{y}_{j,i} | \boldsymbol{\nu}_j, \boldsymbol{\theta}, \sigma_j^2)$ is the density of $(\mathbf{X}_j, \mathbf{Y}_j)$ with respect to the unit rate Poisson process [Møller and Waagepetersen, 2004].

Using the ILM procedure (Subsection 2.3.4) for each subtype separately, we can construct the Gamma random fields as $\Gamma_j(dx) \sim \mathcal{GRF}\{\Gamma_0(dx), \beta_j\}$. Taking the normalised probability measure $\tilde{\Gamma}_0(dx) = \frac{\Gamma_0(dx)}{\Gamma_0(\mathcal{B})}$, let $\Gamma_0(dx) = \frac{\Pi(dx)}{|\mathcal{B}|}$, where $\Pi(dx)$ is Lebesgue measure and $|\mathcal{B}| = \int_{\mathcal{B}} \Pi(dx)$, then $\tilde{\Gamma}_0(dx) = \Gamma_0(dx)$ and $\tilde{\Gamma}_0(\mathcal{B}) = 1$. Therefore, $\Gamma_0(dx)$ are uniformly distributed over \mathcal{B} which implies that a priori the latent source locations $\theta_{j,m}$ are also uniformly distributed over \mathcal{B} . Hence, denote the joint distribution of the jump locations and jump sizes by

$$\{(\theta_{jm}, \nu_{jm})\}_{m=1}^M \sim \text{invLévy}\{\Gamma_0(dx), \beta_j\}. \quad (2.22)$$

2.3.8 Marked point processes

A marked point process is a random sequence of ordered pairs (y, w) with points $y \in \mathbb{R}^d$ (for our purposes $y \in \mathcal{B} \subseteq \mathbb{R}^3$) and marks w in some continuous or discrete mark space \mathcal{W} . The number of points of the process \mathbf{Y} is a random integer-valued measure $N(du)$ on the product space $\mathcal{U} = \mathcal{B} \times \mathcal{W}$ consisting of random ordered pairs (y_n, w_n) . If we treat $N(du)$ as a Poisson random field with mean $\lambda(u)\Pi(du)$, it can be expressed as

$$N(dy \, dw) \sim \text{Po}(\lambda(y)\Pi(dy)\phi(w)\Pi(dw)), \quad (2.23)$$

where $\lambda(y)$ is point process intensity function with respect to the reference measure $\Pi(dy)$ and $\phi(w)$ denotes a mark intensity function with respect to $\Pi(dw)$.

For any Borel set $A \in \mathcal{B}$ and any Borel set $C \in \mathcal{W}$, the expected number of points in A with marks in C satisfies

$$\mathbb{E}[\Lambda(A \times C)] = \Lambda(A)\Phi(C), \quad (2.24)$$

where Φ denotes the mark measure. In case of a stationary (i.e. homogeneous) point process, the intensity measure Λ can be expressed as $\lambda L(A)$ where L denotes Lebesgue measure. In case of spatially independent marks, the marked point process is said to have the *random labelling property* (see e.g. Møller and Waagepetersen [2004] or Cronie and van Lieshout [2016]). The description above gives an outline for a marked point process defined on a product space. Note that not all point

processes that are defined on a product space need to be marked processes. A counter-example would be a spatio-temporal process.

Marks can be either categorical or continuous in nature. In the case of categorical marks, the combined process is also referred to as a multi-type or multivariate Poisson process. The simplest case is a bivariate process where the marks take on labels of $w \in \{1, 2\}$. For multivariate Poisson processes, the random labelling property entails that the marks are spatially independent, i.e. $w_j(y) = w_j$ [Møller and Waagepetersen, 2004]. More interesting is the case of continuous marks which are able to reflect a quantity of interest with some local dependence, for example, the height or diameter of a plant species or the size of a brain lesion.

Applications of marked point process have focused, for instance, on the analysis of long-leaf pine trees in a forest area of 200×200 metres where each tree was marked with its diameter at breast height [Rathbun and Cressie, 1994], and the assessment of the individuality of fingerprints [Lim and Dass, 2011].

2.4 Methodological background: Classification and prediction

2.4.1 Cross-validation

For any classification and prediction task, it is crucial to evaluate the performance of the proposed model. A straightforward way of assessing the accuracy of model predictions is to split the data into a training set and a test set. Ideally, this split occurs before any analysis on the data has been carried out and the test set is not touched until the final evaluation step. Meanwhile, the training set is used by the classifier to learn about the structure of the data. For most applications, however, the supply of data is limited. This is especially true in the case of neuroimaging studies, where the collection of data is very expensive and time-consuming. Holding out a large proportion of the total data set for validation purposes becomes prohibitive as it would not leave enough data to successfully train the classifier. On the other hand, a test set with too few data points would produce a very noisy estimate of the prediction accuracy.

Cross-validation techniques try to maximise the amount of data in the training set whilst simultaneously aiming to minimise the variance of predictive performance. An inherent drawback of cross-validation methods is the need to run multiple instances of the same model. Assume we split the whole data set randomly into k distinct, approximately equally sized groups. For each of the k folds, the data

within that fold constitutes the test set that is used for prediction. The other $k-1$ folds make up the training data. Running the model k times, once for every pair of training and test data, yields an estimate of the model’s classification accuracy. Note that, although the test sets are unique, the training sets for two different folds are highly correlated as they share the same data from $k-2$ folds. If one takes k to be equal to the number of data points, the procedure reduces to leave-one-out cross-validation (LOOCV), meaning that the test set comprises a single data point only.

Models that rely on the optimisation of one or multiple parameter values require a further refinement of k -fold cross-validation: nested cross-validation. In order to determine an optimal parameter value, one can again perform cross-validation within the current training set. Importantly, only data from the training set may be used to optimise a parameter. The parameter optimising cross-validation is nested within the training set of the overall cross-validation routine, hence requiring a new split of the training data into m folds.

Computation of the confusion matrix provides a means to quantify model performance. It lists the proportion of correctly and incorrectly classified data points according to true and predicted group label. The rows i and columns j of the confusion matrix contain the true and predicted number of instances per group, respectively. Entry (i, j) of the matrix therefore contains the number of instances from group i that have been classified as belonging to group j . The overall classification rate is given by the ratio of correctly identified instances to total number of data points in the data set. Whenever sizes of the different classes in a data set are imbalanced, a more informative measure of accuracy is the average classification rate which constitutes the mean over the proportions of successfully predicted instances per group, i.e. the mean of the main diagonal of the confusion matrix.

Cross-validation is a popular way of model evaluation. LOOCV exhibits the smallest bias but also a larger variance than k -fold cross-validation [Molinari et al., 2005]. However, depending on the complexity of the model, other methods of measuring model performance may be more suitable. In particular, permutation tests are often preferable, since they produce an unbiased estimate of predictive performance, especially for small sample data sets [Combrisson and Jerbi, 2015].

2.4.2 A Bayesian classifier for multi-type point patterns

Performing LOOCV on a Bayesian model like the HPGRF model of Chapter 5 is computationally prohibitive as it would require to run a full MCMC simulation for

each subject. Instead, we employ an importance sampling scheme and construct a Bayesian classifier based on the estimated posterior predictive probabilities for different groups in a multi-type point pattern. Similar approaches have been proposed for general Bayesian models, for example in Gelfand et al. [1992]; Alqallaf and Gustafson [2001]; Vehtari and Lampinen [2002]. The classification algorithm presented in this section is adapted from Kang et al. [2014a].

Let \mathcal{D}_n denote the full data set and $\mathcal{D}_{-k} \equiv \{\mathcal{D} \setminus \mathbf{y}_k\}$ the set of data excluding the k^{th} realisation. Assume the data is split into $j = 1, \dots, J$ different groups and denote the individual group indicators by G_k . The posterior predictive probability of \mathbf{y}_k belonging to type j is

$$\begin{aligned} \mathbb{P}(G_k = j | \mathbf{y}_k, \mathcal{D}_{-k}) &= \frac{p_j \int \pi(\mathbf{y}_k | G_k = j, \Theta) \pi(\Theta | \mathcal{D}_{-k}) d\Theta}{\sum_{j'=1}^J p_{j'} \int \pi(\mathbf{y}_k | G_k = j', \Theta) \pi(\Theta | \mathcal{D}_{-k}) d\Theta} \\ &= \int \mathbb{P}(G_k = j | \mathbf{y}_k, \mathcal{D}_{-k}, \Theta) \pi(\Theta | \mathbf{y}_k, \mathcal{D}_{-k}) d\Theta, \end{aligned} \quad (2.25)$$

using Bayes rule and the law of total probability. Prior information about the prevalence of type j is contained in p_j and Θ comprises all model parameters. The density of a single realisation of the point process is provided by (2.5), which becomes

$$\pi(\mathbf{y}_k | G_k = j, \Theta) = \exp \left\{ |\mathcal{B}| - \int_{\mathcal{B}} \lambda_j(y | \Theta) dy \right\} \prod_{y \in \mathbf{y}_k} \lambda_j(y | \Theta). \quad (2.26)$$

Assuming draws from the posterior according to $\Theta^{(t)} \sim \pi(\Theta | \mathcal{D}_{-k})$, for $t = 1, \dots, T$, we denote the density at the t^{th} sample by $\pi_j^{(t)} \equiv \pi(\mathbf{y}_k | G_k = j, \Theta^{(t)})$.

The estimate of the posterior predictive probability can therefore be expressed as

$$\hat{\mathbb{P}}(G_k = j | \mathbf{y}_k, \mathcal{D}_{-k}) = \frac{p_j \sum_{t=1}^T \pi_j^{(t)}}{\sum_{j'=1}^J \left[p_{j'} \sum_{t=1}^T \pi_{j'}^{(t)} \right]}, \quad (2.27)$$

and the predicted group label is given by

$$\hat{G}_k = \arg \max_j \left(p_j \sum_{t=1}^T \pi_j^{(t)} \right), \quad (2.28)$$

which can be obtained by computing LOOCV predictive probabilities.

In order to avoid the need to run a separate Monte Carlo chain for every realisation it is useful to reformulate the predictive probabilities in (2.25) as fol-

lows: Denote with g_k the true group-label of \mathbf{y}_k , then, for $j \in \{1, \dots, J\}$ and $k \in \{1, \dots, n\}$,

$$\mathbb{P}(G_k = j | \mathbf{y}_k, \mathcal{D}_{-k}) = \frac{p_j Q_{jg_k}}{\sum_{j'=1}^J p_{j'} Q_{j'g_k}}, \quad (2.29)$$

with the discounted likelihood given by

$$Q_{jj'} = \int \frac{\pi(\mathbf{y}_k | G_k = j, \Theta)}{\pi(\mathbf{y}_k | G_k = j', \Theta)} \pi(\Theta | \mathcal{D}_n) d\Theta. \quad (2.30)$$

Using T draws from the posterior, the estimate of the discounted likelihood factor can be computed as

$$\hat{Q}_{jg_k} = \frac{1}{T} \sum_{t=1}^T \frac{\pi(\mathbf{y}_k | G_k = j, \Theta^{(t)})}{\pi(\mathbf{y}_k | G_k = g_k, \Theta^{(t)})}. \quad (2.31)$$

Note that with this formulation, $\Theta^{(t)}$ is drawn from the complete data posterior, i.e. $\Theta^{(t)} \sim \pi(\Theta | \mathcal{D}_n)$. The posterior predictive probabilities of (2.27) then become

$$\hat{\mathbb{P}}(G_k = j | \mathbf{y}_k, \mathcal{D}_{-k}) = \frac{p_j \hat{Q}_{jg_k}}{\sum_{j=1}^J p_{j'} \hat{Q}_{j'g_k}} \quad (2.32)$$

and the predicted class is given by

$$\hat{G}_k = \arg \max_j (p_j \hat{Q}_{jg_k}). \quad (2.33)$$

Proofs for the relations presented in this section can be found in the appendix of Kang et al. [2014a].

Chapter 3

Two clinical data sets

There exist two widely used file formats for MRI data: NIfTI, the newer and more versatile format, and DICOM, which is still the standard for older scanners. Details can be found for example in Poldrack et al. [2011]. A considerable range of image analysis and processing software has become available, including the SPM (<http://www.fil.ion.ucl.ac.uk/spm>) and FSL (<https://fsl.fmrib.ox.ac.uk/fsl>) packages which we have used to preprocess the data.

The following sections describe in brief the two clinical data sets of patients with multiple sclerosis that are the basis for model applications in Chapter 4 and Chapter 6.

3.1 The GeneMSA data set

The GeneMSA data set is part of a large multi-centre case-control study initiated in 2003 and conducted by a consortium of leading researchers in MS: the University of California in San Francisco, the Vrije Universiteit Medical Center in Amsterdam, the University Hospital Basel in Switzerland, and the pharmaceutical company Glaxo SmithKline (GSK). The full study analysed 551,642 SNPs in 978 cases and 883 controls. The Medical Image Analysis Center (MIAC) in Basel carried out analysis and processing of the MRI data. The study's initial purpose was aimed at finding gene expressions that can be related to multiple sclerosis. Although the study could identify new candidates for gene associations, the effect sizes were very small [Baranzini et al., 2009].

The data set available for this thesis comprises the Basel cohort of the GeneMSA study and consists of 248 subjects scanned on a Siemens Magnetom Avanto syngo 1.5T scanner at the University Hospital in Basel, Switzerland. 3D-

heavily T_1 -weighted magnetisation-prepared rapid gradient echo (MP-RAGE) images were acquired. The following image parameters were used in the acquisition protocol: echo time 3 ms, time-to-repetition 20.8 ms, flip angle 12, matrix size 240×256 , field of view 24.0×25.6 cm, voxel dimensions $1 \times 1 \times 1 \text{ mm}^3$). Additionally, T_1 -weighted Gadolinium-enhanced sequences were acquired with image parameters: echo time 8 ms, time-to-repetition 467 ms, matrix size $256 \times 256 \times 44$, field of view $240 \times 240 \times 132$ mm. Transaxial contiguous 3 mm dual-echo T_2 -weighted images were acquired using turbo spin echo with image parameters as follows: echo time 14 ms, time-to-repetition 3980 ms, flip angle 180, turbo factor 7, matrix size 192×256 , field of view 18.75×25 cm, voxel dimension $0.9766 \times 0.9766 \times 3.0 \text{ mm}^3$. For more details see Baranzini et al. [2009].

Lesions were independently identified and marked on T_2 -weighted images by two neuro-pathologists using a semi-automatic procedure Kappos et al. [2006a]. Additionally, a third radiologist resolved any disparities. As a result, binary lesion masks with a value of 1 (0) at each voxel indicating the presence (absence) of a lesion were created in native space and then affine registered to MNI standard space at $1 \times 1 \times 1 \text{ mm}^3$ resolution using trilinear interpolation. The transformed lesion masks were thresholded at 0.5 to retain binary values. White matter lesion masks for T_1 -weighted black-hole and T_1 -weighted Gd-enhanced images were created guided by the previously outlined T_2 lesions. Additionally, image segmentations into white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) was carried out. In order to preserve the original within-voxel volumes after non-linear transformation, the segmented images were rescaled using the Jacobian determinants resulting from spatial normalisation. A detailed account of the full image preprocessing pipeline can be found in Bendfeldt et al. [2009] and Bendfeldt et al. [2012].

All patients were categorized into one of the five MS subtypes that had been used until 2013, following to the diagnostic standard defined by McDonald et al. [2001]. The cohort was classified as follows: 10 CIS, 172 RRMS, 13 PPMS, 43 SPMS, 10 PRMS. The data set also includes patient specific covariates such as biological sex, age, disease duration, the seven subscores of the Expanded Disability Status Scale (EDSS) and outcomes of the Paced Auditory Serial Addition Test (PASAT). Figure 3.1 provides a visual overview in form of box plots of the lesion count per subject against the five main covariates in the data set with the EDSS subscores combined into a single clinical measure. The data from the full cohort of 248 subjects and five disease subtypes is used for model applications in Chapter 4.

Following the 2013 revisions [Lublin et al., 2014] of the McDonald criteria and the therein established clinical consensus to drop progressive-relapsing MS (PRMS) as a distinct clinical subtype, we exclude the ten subjects classified as PRMS from our analysis. Due to missing demographic data, we had to further exclude two subjects (one CIS, one RRMS). Hence, our working data set comprises 238 MS patients which have been classified into one of the four clinical disease subtypes as follows: 10 with CIS, 172 with RRMS, 43 with SPMS, and 13 with PPMS.

Table 3.2 provides an overview of the main characteristics of the data set. Computation of the correlation matrix of the five covariates shows high correlation between age and disease duration and weak correlation between the other covariates. A qualitative, visual inspection of point locations across the brain for each subtype reveals no obvious differences in the spatial patterns—for example, large clusters or centres of aggregation—between the groups, apart from the number of points available in the data for each subtype. However, all types show a larger density of points towards the central areas of the brain with fewer lesions occurring in cortical regions (see also the plots of lesion locations in Figure 6.1 and Figure 6.2). This is consistent with the fact that MS is predominantly a white matter disease.

Figure 3.2 shows an exploratory scatter plot of cube-root transformed median lesion volumes per subject against each patient’s total lesion count. The average lesion volume is relatively stable across subtypes. The largest discrepancy can be observed for the CIS subtype, which is associated with early stages of the disease. Subjects in this group on average tend to have fewer and smaller lesions than subjects classified into one of the other subtypes.

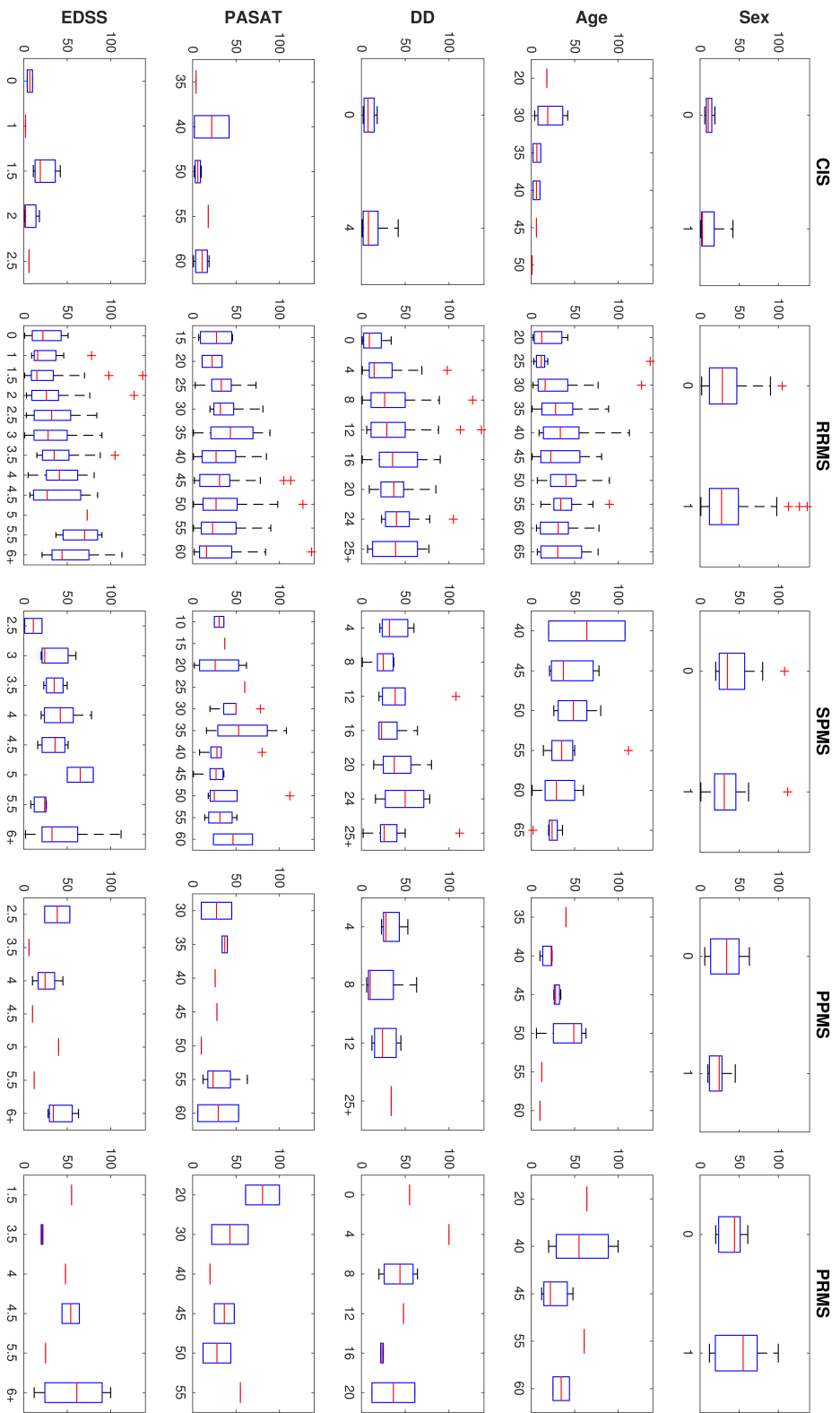


Figure 3.1: Exploratory statistics of the GenEMSA data set: Box plots of lesion count per subject versus demographic and clinical covariates. Sex is codes such that female corresponds to one. The Expanded Disability Status Scale (EDSS) and the Paced Auditory Serial Addition Task (PASAT) are clinical measures of disease severity. DD denotes disease duration (in years). Rows show results for each covariate. Columns indicate one of the five MS subtypes.

Table 3.1: Summary statistics for the GeneMSA data set including the PRMS subtype. The five multiple sclerosis subtypes are CIS (clinically isolated syndrome), RRMS (relapsing-remitting MS), SPMS (secondary-progressive MS), PPMS (primary-progressive MS) and PRMS (progressive-relapsing MS), including standard deviations (SD) and inter-quartile ranges (IQR). The seven subscores of the Expanded Disability Status Scale (EDSS) comprises visual (VSLSC), brain stem (BRSTMSC), pyramidal (PYRSC), cerebellar (CRBLSC), sensory (SENSSC), bowel and bladder (BWLSC) and mental (MNSC) scores. The Paced Auditory Serial Addition Test (PASAT) measures cognitive impairment. Note: This describes the data used for analyses in Chapter 4. The subtype categorisation is based on the McDonald criteria before the 2013 revisions.

	CIS	RRMS	SPMS	PPMS	PRMS
Number of subjects	10	172	43	13	10
Number of T ₁ lesions	41	2519	885	184	251
Number of T ₂ lesions	115	5943	1632	374	451
Number of T ₁ -Gd lesions	0	71	7	1	1
Female / male (% female)	6/4 (60.0%)	134/38 (77.9%)	23/20 (53.5%)	6/7 (46.2%)	5/5 (50%)
Median age in years [IQR]	33.4 [30.5, 42.9]	41.4 [35.0, 48.2]	56.3 [50.9, 61.3]	45.7 [42.2, 51.0]	45.7 [39.3, 56.1]
Median disease duration in months after diagnosis [IQR]	24 [3, 36]	120 [60, 204]	228 [150, 327]	72 [48, 144]	130 [72, 192]
Median total EDSS [IQR]	1.5 [1.1, 2.0]	2.5 [1.5, 3.5]	5.0 [3.5, 6.5]	4.0 [4.0, 6.0]	5.0 [3.5, 6.0]
Median VSLSC [IQR]	0 [0, 0.75]	0 [0, 1]	1 [0, 1]	0.5 [0, 1]	0.5 [0, 1]
Median BRSTMSC [IQR]	0 [0, 1]	1 [0, 1]	1 [0, 2]	1 [1, 1]	1 [0, 2]
Median PYRSC [IQR]	0 [0, 1]	1 [1, 2]	2.5 [2, 3]	3 [2, 3]	2 [2, 3]
Median CRBLSC [IQR]	1 [0, 1]	1 [0, 2]	2 [2, 3]	2.5 [2, 3]	2 [2, 2]
Median SENSSC [IQR]	0 [0, 0.75]	1 [0, 2]	2.5 [2, 3]	2 [1, 3]	2 [0, 3]
Median BWLSC [IQR]	0 [0, 0]	0 [0, 1]	1 [1, 2]	1.5 [1, 2]	1.5 [1, 2]
Median MNSC [IQR]	0 [0, 1]	1 [0, 2]	2 [1, 2]	1 [1, 2]	1.5 [1, 2]
Median PASAT [IQR]	52 [40, 59]	46 [39, 54]	41 [31, 48]	51 [34, 55]	41 [30, 51]

Table 3.2: Summary statistics of the GeneMSA data set for the four remaining multiple sclerosis subtypes, CIS (clinically isolated syndrome), RRMS (relapsing-remitting MS), SPMS (secondary-progressive MS) and PPMS (primary-progressive MS) after the 2013 revisions of the McDonald criteria, excluding progressive-relapsing MS. The table includes standard deviations (SD) and inter-quartile ranges (IQR). Note: This describes the data used for analyses in Chapter 6.

	CIS	RRMS	SPMS	PPMS	Total
Number of subjects	10	172	43	13	238
Total Number of lesions	115	5943	1632	374	8064
Mean number of lesions (SD)	11.50 (12.51)	34.55 (27.69)	37.95 (24.51)	28.77 (17.66)	33.88 (26.60)
Median number of lesions [IQR]	8 [2, 18]	28 [12, 49]	32 [22, 50]	26 [12, 41]	28 [12, 47]
Female / male (% female)	6/4 (60.0%)	134/38 (77.9%)	23/20 (53.5%)	6/7 (46.2%)	169/69 (71.0%)
Median age in years [IQR]	33.4 [30.5, 42.9]	41.4 [35.0, 48.2]	56.3 [50.9, 61.3]	45.7 [42.2, 51.0]	43.8 [37.5, 52.1]
Median disease duration in months after diagnosis [IQR]	24 [3, 36]	120 [60, 204]	228 [150, 327]	72 [48, 144]	120 [56, 192]
Median total EDSS [IQR]	1.5 [1.1, 2.0]	2.5 [1.5, 3.5]	5.0 [3.5, 6.5]	4.0 [4.0, 6.0]	3.0 [2.0, 4.0]
Median PASAT [IQR]	52 [40, 59]	46 [39, 54]	41 [31, 48]	51 [34, 55]	46 [28, 50]
Median lesion volume* [IQR]	2.76 [1.71, 6.20]	2.62 [1.71, 6.91]	2.52 [1.71, 7.09]	2.69 [1.73, 7.06]	2.62 [1.71, 6.99]

* cube-root transformed T₂ lesion volume (in voxels)

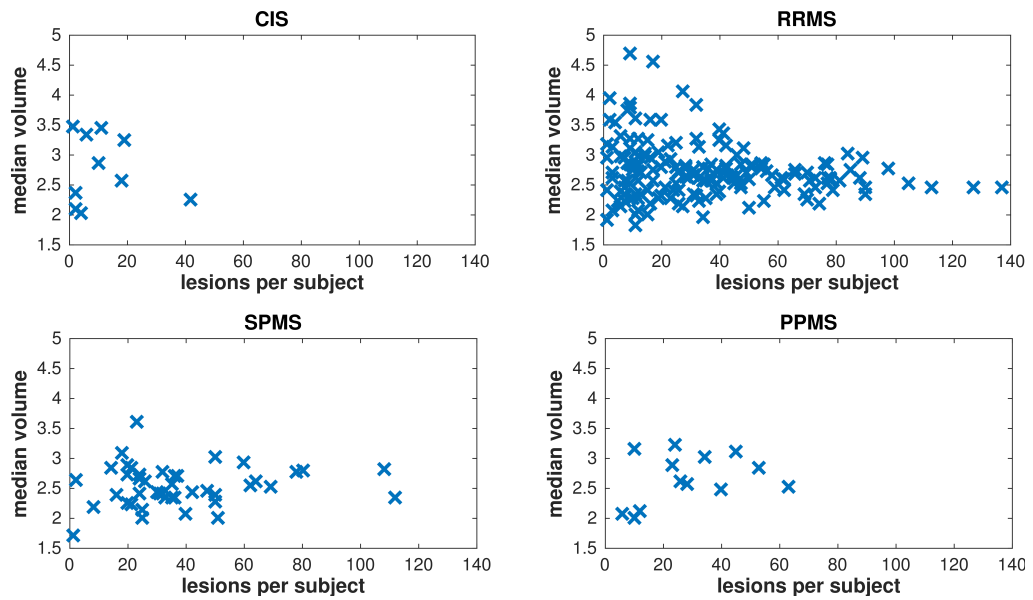


Figure 3.2: Exploratory analysis of the GeneMSA data set: Scatter plots of cube-root transformed median lesion volume versus the number of lesions per subject across the four remaining MS subtypes according to the revised McDonald criteria.

3.2 The BENEFIT data set

The BENEFIT (Betaferon/Betaseron in Newly Emerging multiple sclerosis For Initial Treatment) study [Kappos et al., 2006a; Barkhof et al., 2007] was carried out in 2002–2005. It was a double-blind, placebo-controlled and randomised phase 3 study that involved a total of 98 imaging centres and was financed by Bayer Healthcare. The aim of the study was to evaluate the safety, tolerability and efficacy of the agent interferon beta-1b (IFN β -1b) in patients with initial signs of multiple sclerosis.

Criteria for inclusion in the study were a first demyelinating event suggestive of MS and at least two clinically silent lesions on a T₂-weighted MRI scan. Hence, all subjects were classified into the CIS subtype of MS. The initial cohort comprised 468 subjects. Following a 5:3 ratio, patients were randomly assigned to the treatment or placebo group. Treatment involved a dose of 250 μ g IFN β -1b subcutaneously every other day [Kappos et al., 2006b]. Any subjects from the placebo group who, over the course of the study, were subsequently diagnosed with clinically definite MS (CDMS), were switched to active treatment. The study period extended over 24 months and entailed the collection of clinical scores at 3 to 6 months intervals as well as MRI scans at the end of the study as well as follow up scans 3 years after completion. The baseline MRI scans were taken at field strengths of 0.5–1.5T and included transaxial contiguous 3-mm dual-echo T₂-weighted and T₁-weighted images. Clinical assessment of hyper-intense T₂-lesions, Gadolinium-enhancing lesions and hypo-intense T₁ lesions was carried out at the Image Analysis Centre at the Vrije Universiteit Medical Center in Amsterdam [Barkhof et al., 2007].

Outcomes of the BENEFIT study found an advantage of early treatment on disability progression after three years [Barkhof et al., 2003; Kappos et al., 2007]. However, the difference between treatment and placebo groups did not remain significant at the 5 year follow-up [Kappos et al., 2009]. The study authors mention that crossover from placebo to active treatment may have masked some treatment effects.

For the purposes of this thesis, we use clinical and imaging data at baseline in conjunction with binary information about whether a patient converted to CDMS during the study period. In collaboration with the Medical Image Analysis Center in Basel, we have extracted the lesion masks from the data, transformed them from the original DICOM format to more user friendly NIFTI images and registered them to standard MNI space. Due to poor image quality (scans containing artefacts) and some cases of missing data, about 25% of subjects from the original data set had to be excluded from our analysis. The data used for this work comprises 364 patients,

175 of whom converted to CDMS and 189 remained classified as CIS.

Box plots of how the number of lesions per subject depends on the three demographic and clinical covariates that are available in the data set (sex, age and EDSS score) are shown in Figure 3.4. Scatter plots of median lesion volume versus lesion count are provided in Figure 3.3. On visual inspections, the two groups in the data set, seem to exhibit very similar characteristics. One notable difference is the number of lesions per subject, which tends, on average, to be larger in the CDMS group.

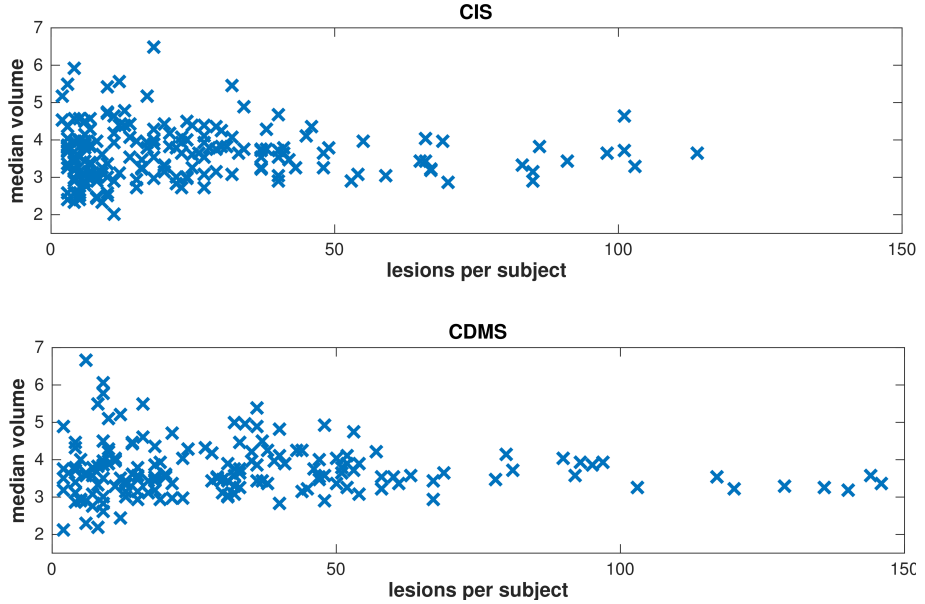


Figure 3.3: Exploratory analysis of the *BENEFIT* data set: Scatter plots of cube-root transformed median lesion volume versus the number of lesions per subject for the two groups, clinically isolated syndrome (CIS) and clinically definite multiple sclerosis (CDMS). The Expanded Disability Status Scale (EDSS) is a clinical measure of disease severity.

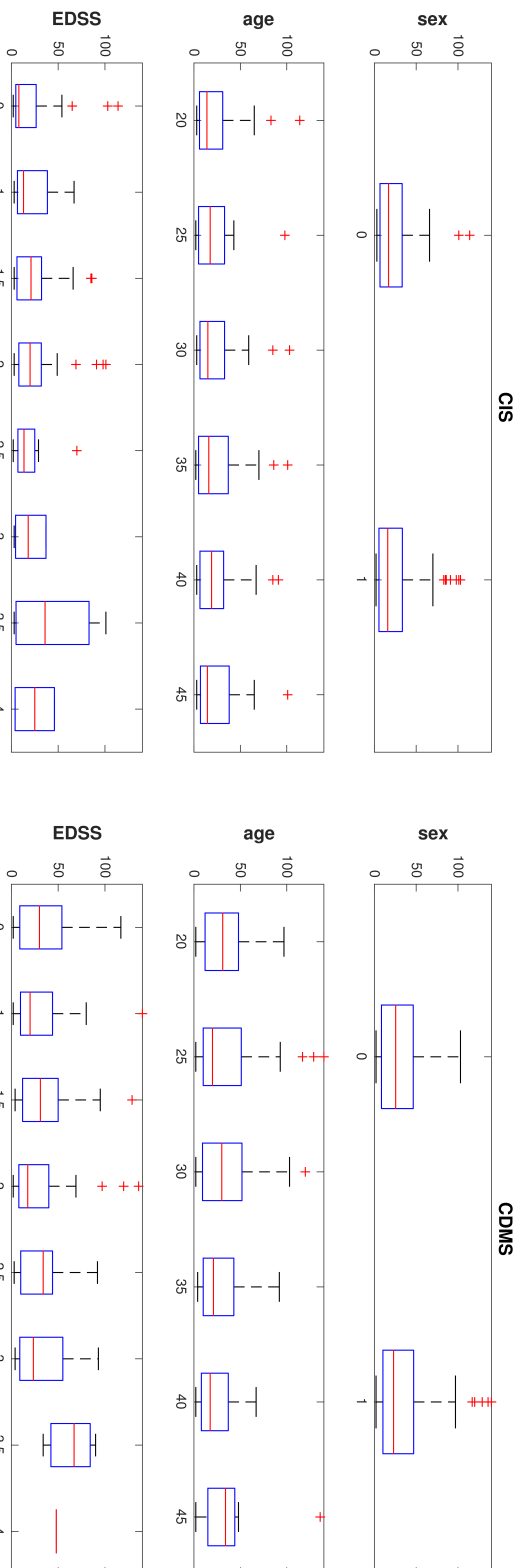


Figure 3.4: Exploratory statistics of the BENEFIT data set: Box plots of lesion count versus demographic and clinical covariates for the two groups, clinically isolated syndrome (CIS) and clinically definite multiple sclerosis (CDMS). Sex is codes such that female corresponds to one. Rows show results for each covariate.

Table 3.3: Summary statistics the two groups, CIS (clinically isolated syndrome) and CDMS (clinically definite multiple sclerosis), of the BENEFIT data set, including standard deviations (SD) and inter-quartile ranges (IQR).

	CIS	CDMS	Total
Number of subjects	189	175	364
Number of lesions	4,813	6,063	10,876
Mean number of lesions (SD)	25.5 (29.9)	34.6 (36.1)	29.9 (32.9)
Median number of lesions [IQR]	16 [3, 101]	24 [2, 140]	19 [3, 124]
Female / male (% female)	128/61 (67.7%)	131/44 (74.9%)	259/105 (71.2%)
Median age in years [IQR]	32 [19, 45]	29 [19, 44]	30 [19, 44]
Median EDSS [IQR]	1.5 [0.0, 3.5]	1.5 [0.0, 3.1]	1.5 [0.0, 3.5]
Median lesion volume* [IQR]	2.25 [1.74, 2.53]	2.20 [1.73, 2.52]	2.23 [1.73, 2.52]

*cube-root transformed T₂ lesion volume (in voxels)

Chapter 4

Comparison of two machine learning approaches and two spatially informed models for MS subtype classification

4.1 Introduction

The aim of this chapter is to investigate and compare several different classification approaches and their application to the prediction of multiple sclerosis subtypes. First, a naïve Bayesian classifier (NBC) (Section 4.2) provides a baseline of minimally expected performance for the other models to compare to. We then propose a machine learning approach based on support vector machines (SVMs) (Section 4.3) that utilises a large number of quantitative features. These features, alongside traditional demographic and clinical measures, include aspects of lesion geometry, measured by Minkowski functionals, and statistics of intra-lesion MRI intensities. Finally, we use two spatially informed models, a Bayesian spatial generalised linear mixed model (BSGLMM) on voxel-wise lesion maps (Section 4.4) and a log-Gaussian Cox process (LGCP) model on lesion location data (Section 4.5).

Each of the main models has its relative merits. The BSGLMM has the interpretability of a traditional generalised linear regression model but can only account for local spatial dependence; the LGCP explicitly accounts for spatial variation in lesion location on a larger scale; and the SVM classifier can revert to using a combination of features from multiple MRI modalities, while the other two only operate on a single type of MR image. Results from all classifiers are summarised in Section 4.6.

4.2 A naïve Bayesian classifier

In general, MR images cannot be directly used for statistical analysis. Several preprocessing steps are usually necessary to prepare the data in a format that is suitable for the model at hand. Due to the large amount of data (typically 10^5 – 10^6 voxels per image) mass-univariate models are commonly used in neuroimaging applications. A mass-univariate approach comprises a separate model fit at each voxel, independently of all the rest. Although other models, for instance multivariate techniques that consider multiple spatial locations jointly, can be entertained, mass-univariate models are often preferred in practice because they are comparatively easy to use and interpret.

A basic variant of a mass-univariate model is a naïve Bayesian classifier (NBC) [Lewis et al., 1998]; for a neuroimaging application, see for example Yarkoni et al. [2011]. The NBC comprises a simple binomial model at each voxel and for each MS subtype. The binary, voxel-by-voxel lesion masks obtained from MRI scans provide the array of input features to the NBC. Crucially, the NBC assumes mutual independence between each element, i.e. each voxel. This assumption, of course, does not reflect the true nature of the data, in which neighbouring voxels are highly correlated.

With respect to the data application, we train the NBC on the full T_1 - and T_2 -weighted lesion masks, respectively. We assign Jeffrey’s prior, $\text{Beta}(0.5, 0.5)$, to the conditional probability of having a lesion at voxel v given a certain subtype. In order to assess classification performance of the NBC we perform leave-one-out cross-validation.

As a means of increasing classification performance, one can consider restricting the space of interest to a global lesion mask where only those voxels with an incidence of two or more lesions over the whole data set are taken into account. Although this reduces the dimensionality by excluding uninformative voxels, it should be noted that the classifier is then unable to pick up on any new lesions outside the mask.

Compared to any other model, the NBC is computationally very efficient, even with extremely large numbers of features. However, ignorance about any spatial dependence between neighbouring voxels constitutes a major drawback. In particular, the classifier has no information about which voxels belong to the same lesion.

4.3 Support vector machines based on lesion-specific features

Our goal for this section is to build a classifier based on support vector machines that exploits a range of geometric features of individual MS lesions, alongside traditional demographic and clinical measures. We use three-dimensional, binary lesion masks from three available MRI modalities (T_1 , T_2 and T_1 -Gd enhanced) as basis for the computation of geometric characteristics.

In the remainder of this section, we first review the fundamentals of SVM theory and then introduce measures of lesion geometry based on Minkowski functionals which will form a large part of the input data to the classifier. The last part summarises the available feature set and discusses feature reduction via principal components analysis.

4.3.1 Support Vector Machines

Support vector machines (SVMs) are an established, supervised machine learning technique and widely used across various disciplines, from neuroimaging [Mourao-Miranda et al., 2005; Mwangi et al., 2012] and cancer research [Huang et al., 2017] to chemistry [Invancic, 2007] and applications in remote sensing [Mountrakis et al., 2011]. What makes SVMs particularly attractive is their generally good performance in comparison to other classifiers as well as their relatively easy extension to non-linear problems using kernels. In addition, careful tuning of regularisation and kernel parameters make them less prone to over-fitting [Cawley and Talbot, 2007]. Linear SVMs are well suited for classification problems where the number of input features is very large, potentially much larger than the number of samples in the data set. However, as the number of features decreases non-linear SVM approaches often perform better due to an increased flexibility in fitting the data. This comes at the cost of having to estimate an additional parameter.

SVMs have been used successfully in diagnostic and prognostic settings based on structural MRI data, including applications in Alzheimer’s disease [Davatzikos et al., 2008], schizophrenia [Koutsouleris et al., 2012] and multiple sclerosis [Weygandt et al., 2011; Bendfeldt et al., 2012]. A recent review of SVMs applied to neurological conditions by Orrù et al. [2012] has found that SVMs have been used successfully in several clinical applications that focused on disease diagnosis, prediction of treatment success and prediction of transition to more severe stages of a disease, respectively.

4.3. Support vector machines based on lesion-specific features

Initially proposed by [Vapnik, 1979], SVM is a family of supervised machine learning techniques that rely on optimising a convex cost function. Because it is a non-parametric method, no assumption is made on the underlying data distribution. The SVM algorithm tries to find a separating hyperplane that splits the data set into two distinct classes. The optimal decision boundary is chosen such that the distance between the data points of the two labelled groups and the hyperplane is maximised. This also makes apparent a drawback of SVM in its original formulation: it can only deal with two-class problems.

In the following, we review the essentials of SVM theory. Numerous textbooks are available which offer a more in-depth description, for example Campbell [2002]; Bishop [2006]; Invanciu [2007] or Vapnik [1999]. The rest of this section is based in parts on Bishop [2006], Burges [1998] and Scholkopf and Smola [2002]. Let $y_i = \pm 1$ denote the class of data point $\mathbf{x}_i \in \mathbb{R}^d$ in feature space and \mathbf{w} be a d -dimensional vector of weights. The canonical representation of the decision hyperplane is given by $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$, where $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the inner product. The geometrical margin of the point \mathbf{x}_i , i.e. the distance to the hyperplane, can be written as

$$m_{(\mathbf{w}, b)}(\mathbf{x}_i) = \frac{y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)}{\|\mathbf{w}\|}. \quad (4.1)$$

Suppose the data is linearly separable, the goal is to find a decision function

$$f(\mathbf{x}) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) \quad (4.2)$$

that satisfies $f(\mathbf{x}_i) = y_i$.

If the training data is not perfectly separable, i.e. a separating hyperplane does not exist, one can allow for soft margins by introducing so-called slack variables ξ_i . Additionally, soft margins take into account (or rather disregard) outliers that may have an undesired impact on the location of the hyperplane. This modified SVM algorithm seeks a linear classifier of the form $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ that minimises the objective function

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i, \quad (4.3)$$

subject to the linear constraints

$$y_i(\langle \mathbf{x}_i, \mathbf{w} \rangle + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i > 0. \quad (4.4)$$

The parameter C is a constant (the so-called box-constraint) that controls the amount of overlap or misclassification. In the limit $C \rightarrow \infty$ no overlap is allowed

and the classifier seeks a perfect separation between the two groups. The support vectors are given by the data points x_i on the boundary, i.e. those for which $y_i(\langle \mathbf{x}_i, \mathbf{w} \rangle + b) = 1 - \xi_i$. This is a quadratic programming problem that is easier to solve by introducing Lagrange multipliers α_i and restating it in terms of the dual Lagrangian of the initial problem

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \sum_{i=1}^n \alpha_i [y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + \xi_i]. \quad (4.5)$$

The Lagrangian must be minimised with respect to \mathbf{w} and b and maximised with respect to all α_i 's. With the help of the Karush-Kuhn-Tucker theorem this leads to the following constraints

$$\alpha_i [y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + \xi_i] = 0 \quad (4.6)$$

and

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i. \quad (4.7)$$

The cases when $\alpha_i > 0$ represent the support vectors. Using the expansion of the weight vector in (4.7), the optimisation problem results in maximising

$$\sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i=1} \sum_{j=1} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \quad (4.8)$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1} \alpha_i y_i = 0$.

Kernel functions

Apart from generic data sets of simple examples, perfect separation of the groups usually cannot be obtained in the native parameter space. There are two aspects of SVM that address this: The box-constraint C allows for a certain amount of misclassification of data points during the optimisation procedure. On the other hand, the kernel function produces a mapping of the inner product of data points into high dimensional feature space. The motivation for using kernels is that, with a suitable choice of kernel function, the data may become separable by finding an optimal separating hyper-surface.

Maximising the target function (4.8) and evaluating the decision function (4.1) requires expensive calculations of inner products in high-dimensional feature

4.3. Support vector machines based on lesion-specific features

space. These computations can be reduced by using a positive definite kernel such that $\langle \mathbf{x}_i, \mathbf{x}_j \rangle \rightarrow K(\mathbf{x}_i, \mathbf{x}_j)$. This transformation is based on Mercer's theorem [Wasserman, 2006] and is known as the *kernel trick* [Scholkopf and Smola, 2002]. It leads to the new decision function

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (4.9)$$

with the following target function to be maximised:

$$\sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i=1} \sum_{j=1} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (4.10)$$

For a detailed review of kernel methods see for instance Campbell [2002]. The idea behind using kernels is based on the computational concept of hyperplanes. All the calculations for the separating plane involve only inner products. Therefore, non-linear kernels can use identical calculations and solution algorithms, and obtain classifiers that are non-linear. The resulting classifiers are hyper-surfaces in some space S , but the space S does not have to be identified or examined. There is no single established kernel function for SVM. Frequently used kernels include radial basis functions (RBF), given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right) \quad (4.11)$$

and sigmoid functions, also called the multilayer perceptron kernel (MLP) due to its connection to neural networks, given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh \left(a \mathbf{x}_i^T \mathbf{x}_j + b \right). \quad (4.12)$$

Besides these two examples there exists a considerable range of more advanced kernel methods, see e.g. Sanchez [2003].

The RBF kernel can be interpreted as smoothing out each data point. The corresponding transformed feature space is an infinite dimensional Hilbert space. The particular advantage of the kernel trick is that the mapping to (potentially infinite) new feature space is never actually carried out, only inner products of the vectors in feature space are needed.

In an initial exploration study on MS lesion data, we found that radial basis functions resulted in consistently higher accuracies than sigmoid functions. Linear or polynomial kernels turned out to be unsuitable to capture the structure of data

in feature space, yielding accuracies barely above chance level.

A majority voting scheme

Our goal is the application of SVM classifiers to a prediction task involving $j=5$ subtypes of MS. This requires an extension of the binary classification scheme of ordinary SVM to a multi-class method [Hsu and Lin, 2002].

The simplest way to incorporate more than two classes is a one-versus-rest approach, that is to train j different classifiers, each distinguishing between one group and the union of all $j-1$ remaining groups. However, with this method the problem arises that the classification becomes inconsistent and ambiguous if more than one of the j classifier estimates returns a positive result.

Computationally more expensive but also more reliable is the one-versus-one approach [Allwein et al., 2001] which involves the training of $j(j-1)/2$ classifiers, one for each pair of classes. To which group a new or unseen data point belongs is then determined by a majority voting scheme. In case of a draw the classifier involving the two classes with the most votes decides.

4.3.2 Minkowski functionals and lesion geometry

First, we use binary lesion masks to identify individual lesions. Each lesion is defined as a cluster, i.e. a collection of contiguous lesion-marked voxels defined by a neighbourhood of order 26, which means that lesion voxels that share at least one face, edge or corner are considered to belong to the same cluster. Lesion clustering was performed using FSL’s cluster algorithm.

We then use Minkowski functionals to extract and quantify aspects of the geometry of each lesion. Minkowski functionals are additive measures that yield local as well as global information with respect to the morphology of an object. In \mathbb{R}^d there exist $d+1$ such measures. In standard three-dimensional Euclidean space, Minkowski functionals are directly related to the geometric quantities volume, surface area, mean breadth and the Euler-Poincaré (EP) characteristic [Arns et al., 2001; Lang et al., 2001].

Due to their completeness property [Hadwiger, 1957; Arns et al., 2001] the functionals can be generalized to configurations with singular edges, such as voxel-based objects consisting of cubic or cuboid lattice grains, which means that they can be directly calculated from MRI data. The volume of a lesion, for instance, is trivially given by the number of voxels it encompasses. The other functionals

are obtained by considering the interface associated with the vertices of each voxel. Each vertex of the lattice is shared by eight neighbouring voxels yielding 2^8 possible configurations. The number of possible configurations, in the case of isotropic voxel size, reduces to 22 due to rotational invariance of the measures. The contribution from different configurations to the (local) Minkowski functionals are then used to compute the global measure for an object.

The measure of mean breadth is related to the integral of mean curvature [Serra, 1988] and represents the mean extension of the lesion taken over all possible rotations in three dimensions, and gives a pose-independent measure of breadth; a more extensive discussion can be found in Arns et al. [2001] and Lang et al. [2001]. Finally, the Euler-Poincaré characteristic is the most basic Minkowski functional and provides a connectivity parameter of the object [Arns et al., 2010]. For a single cluster of voxels and polyhedra in general the EP number is classically defined as $\chi = V - E + F$, where V, E, F denote the number of vertices, edges and faces, respectively. Equivalently, it is given by 1 minus the number of holes plus the number of handles in the geometric object. Furthermore, the whole-brain sum of EP characteristics of single lesions is closely related to the total lesion count. It should be noted that there exist more generalised notions of the Euler-Poincaré characteristic in algebraic topology, see e.g. [Spanier, 1966]. For a general discussion of high dimensional Minkowski functionals we refer to, for example, Legland et al. [2007].

A simplification in the computation of Minkowski functionals can be achieved when considering a restricted level of connectivity between neighbouring voxels. One can reduce the full neighbourhood of 26 adjacent voxels by taking into account only the 6 nearest neighbours on the lattice that share a face. Voxels that only have an edge or a corner in common are in this case considered as disconnected.

4.3.3 The feature set

As an initial preprocessing step, after extracting the geometric characteristics of each lesion from the MRI data, we take the cubic and square root of the volume and area measures respectively [Sormani and Filippi, 2007]. Hence all geometric features are comparable in scale. In order to reduce redundancy and to gain greater sensitivity to (size-independent) lesion shape, we separate volumetric information from the measures for surface area and mean breadth by using adjusted quantities instead. Dividing surface area and mean breadth by their expected value, assuming the lesion was a sphere, produces adjusted measures that reflect the topological information contained in 1D (mean breadth), 2D (surface area) and 3D (volume)

measures and indicate how far the shape of a lesion deviates from a perfect sphere.

The single measures for each lesion do not constitute features that can be used in a classification algorithm, since each patient has a different number of lesions. Based on lesion-specific values for each subject, we compute whole-brain summaries to be used as input features, consisting of the sum total, mean, median, maximum, minimum and standard deviation for each of lesion volume, surface area, mean breadth and the EP characteristic.

The full feature set consists of the following: The non-image-based features for each subject comprise the demographic characteristics sex, age as well as three clinical measures: disease duration, PASAT and EDSS. The image-based features encompass the geometry-derived measures described above as well as the fraction of grey matter (GM) volume to whole brain volume and intra-lesion intensity measures. The grey matter volume can be obtained from MRI scans by using segmentation procedures that are available, for example, within the SPM software package. The intra-lesion intensity measures are computed for each of the three MRI modalities and whole-brain summaries (mean, median and standard deviation) of a subject's MRI intensities within lesions. Since intensities recorded on MR images are relative not absolute, we normalise the intra-lesion intensities by the median intensity of the respective scan. These intensity measures can be viewed as an attempt to capture texture information about the lesions.

The feature set is constructed in a way to intentionally include a range of partially redundant features. Previous studies have mainly focused on lesion count or lesion load, however, with mixed results. For example, Mostert et al. [2010] found that although the total number of T₂ lesions has a small predictive value for progression of disability in RLRM, it appears to be of little relevance in progressive types of MS.

Whole-brain summary measures lose any information about the location of individual lesions. In order to encode a minimum of spatial information into the feature set, we consider splitting the whole-brain summary measures according to 13 regions of interest (ROI) based on white matter (WM) track segmentations derived from the Johns-Hopkins brain atlas (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>) or 11 lobar ROI's based on functional anatomy, respectively. A summary overview of the employed segmentations is provided in Figure 4.1. Lesions that cross the boundary of two or more regions are considered to belong to the region where the majority of their constituent voxels are located.

4.3. Support vector machines based on lesion-specific features

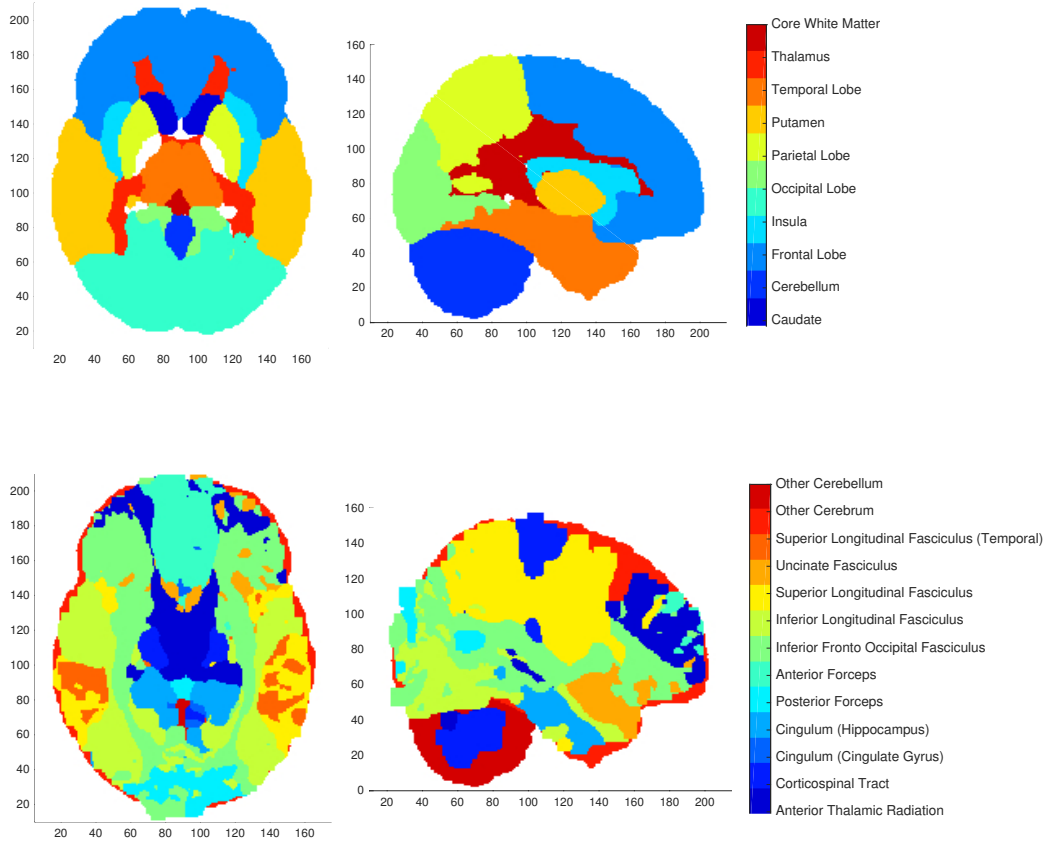


Figure 4.1: Visualisation of two brain parcellations into 11 regions of interest largely based on functionality (*top*) and 13 regions of interest based on white matter track segmentations (*bottom*) based on the Johns-Hopkins brain atlas. Shown are the same axial and sagittal views for both parcellations.

4.3.4 Principal component analysis

In order to reduce the amount of redundant feature information, we use principal component analysis (PCA) to transform the high dimensional feature sets. PCA is a widely used technique for dimensionality reduction and feature extraction. The main principle of PCA is an orthogonal projection of the input data onto a lower dimensional subspace. The projection is carried out in a way such that the variance of the resulting principal components is maximised [Bishop, 2006].

Assume we have a feature set consisting of observations $\{\mathbf{x}_n\}_{n=1}^N$, each containing d measures of interest and therefore spanning a feature space $\mathcal{F} \subset \mathbb{R}^d$. Let \mathbf{X} denote the matrix of size $N \times d$ that contains the data and assume that \mathbf{X} is centred at mean zero. Then the covariance matrix is given by $\mathbf{\Sigma} = \mathbf{X}^\top \mathbf{X} / (N-1)$.

The goal is to project the feature set onto a subspace \mathcal{S} of dimension $p < d$. Define a unit vector \mathbf{u} . In the case of $p=1$, the projection occurs onto a scalar value $\mathbf{u}^\top \mathbf{x}_n$ and the variance of the projected data is given by $\mathbf{u}^\top \mathbf{\Sigma} \mathbf{u}$. In order to maximise this variance with respect to the constraint that $\mathbf{u}^\top \mathbf{u} = 1$, we introduce a Lagrange multiplier λ_1 [Bishop, 2006]. Hence, maximising $\mathbf{u}^\top \mathbf{\Sigma} \mathbf{u} + \lambda_1(1 - \mathbf{u}^\top \mathbf{u})$ yields a stationary point at $\mathbf{\Sigma} \mathbf{u} = \lambda_1 \mathbf{u}$. This means that \mathbf{u} is an eigenvector of $\mathbf{\Sigma}$ and that $\lambda_1 = \mathbf{u}^\top \mathbf{\Sigma} \mathbf{u}$.

The largest value for the variance $\mathbf{u}^\top \mathbf{\Sigma} \mathbf{u}$ of the projected data is therefore obtained by picking the eigenvector \mathbf{u} with the largest eigenvalue λ_1 . The extension from the one-dimensional scenario to the p -dimensional case is straightforward. The projection is given by the p largest eigenvalues and their corresponding eigenvectors of the covariance matrix $\mathbf{\Sigma}$.

For the computation of the principal components it is useful to consider the connection between singular-value decomposition and PCA. Since $\mathbf{\Sigma}$ is symmetric, there exists a diagonal form such that $\mathbf{\Sigma} = \mathbf{V} \mathbf{D} \mathbf{V}^\top$. The columns of \mathbf{V} correspond to the eigenvectors (principal directions) of the covariance matrix and \mathbf{D} is a diagonal matrix with eigenvalues λ_i in decreasing order on the main diagonal. Note that the columns of the matrix product $\mathbf{X} \mathbf{V}$ represent the principal components and the rows of $\mathbf{X} \mathbf{V}$ are the representation of the projected data in the p -dimensional subspace.

If we perform a singular value decomposition on the input data such that $\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{W}^\top$, then the columns of \mathbf{W} correspond to principal directions, the columns of $\mathbf{U} \mathbf{S}$ are the principal components and the singular values s_i are related to the eigenvalues of $\mathbf{\Sigma}$ through $\lambda_i = s_i^2 / (N-1)$. Hence, in order to reduce the dimensionality of the feature space to $p < d$, retain the first p columns of \mathbf{U} and the $p \times p$ upper left square matrix in \mathbf{S} . Finally, the $N \times p$ sized matrix of principal components is given by $\mathbf{U}_{N \times p} \mathbf{S}_{p \times p}$.

4.3.5 Model evaluation

We consider a model space consisting of seven different feature configurations. Table 4.1 summarises the respective feature sets.

Table 4.1: Summary of feature sets for the SVM classifier. The following abbreviations are used: demographic and clinical attributes (DC), grey matter volume ratio (GM), total lesion count (count), lesion volume (load), Euler-Poincaré number (EP), geometric summary measures (GEO), intra-lesion intensity measures (INT), whole-brain summaries (-brain), measures split by regions of interest (-ROI).

Model	Features	Total number of features
M1	DC	12
M2	DC, GM-brain, count, load	19
M3	GEO-brain	60
M4	GM-ROI, EP, load-ROI	27
M5	DC, GM-ROI, INT-brain	32
M6	DC, GM-ROI, GEO-ROI, INT-ROI	920
M7	DC, GM-brain, subset of GEO & INT*	51

*T₂ EP characteristic and median T₂ volume by WM ROI's, standard deviation of T₁ mean breadth, median of T₂ mean breadth, T₁ and T₁-Gd mean intensity

For model evaluation we carry out stratified k -fold cross-validation (see Subsection 2.4.1), where k is determined by the number of elements in the smallest class. In the GeneMSA data set, two groups (CIs and PRMS) consist only of ten subjects, respectively; thus $k=10$ to ensure that each held-out fold contains at least one subject from these groups. The elements of the other groups are then randomly but equally distributed amongst the k sets. Within each cross-validation fold, a three-dimensional parameter optimisation via grid search is performed on the SVM box-constraint, the RBF kernel parameter and the number of principal components. Furthermore, the evaluation of different parameter sets is done by means of nested cross-validation, i.e. in order to ensure independence of the test fold, the training data consisting of $k-1$ folds is once again split into a training ($k-2$ folds) and a test set (1 fold). The overall classification accuracy of the model is estimated by using the so far unseen data from the initially held out fold.

4.4 A Bayesian Spatial Generalised Linear Mixed Model

The Bayesian Spatial Generalised Linear Mixed Model (BSGLMM) has first been proposed by Ge et al. [2014]. In this section we present the cornerstones of the model. For details we refer to the original paper.

As an extension of generalised linear mixed models, a spatial generalised linear mixed model has covariates $\mathbf{x}(s)$ and coefficients $\boldsymbol{\beta}(s)$ that are spatially dependent and comprise the systematic component, $\eta(s) = \mathbf{x}^T(s)\boldsymbol{\beta}(s)$; for our purposes $s \in \mathbb{R}^3$. The form of the systematic component includes any combination of spatially constant or varying covariates and coefficients. For computational reasons [Ge et al., 2014], we use the probit link function, Φ^{-1} . Let s_k denote the k^{th} voxel (or lattice site) and $Y_i(s_k) \in \{0, 1\}$ denote a Bernoulli random variable for subject $i=1, \dots, N$, indicating the presence or absence of a lesion at voxel s_k . For convenience, denote the probability of observing a lesion at s_k by $p_i(s_k) \equiv \mathcal{P}(Y_i(s_k)=1)$. Hence, the random, link and systematic components in the BSGLMM are specified for subject i and voxel k as follows

$$[Y_i(s_k)|p_i(s_k)] \sim \text{Bernoulli}[p_i(s_k)], \quad (4.13)$$

$$\Phi^{-1}\{\mathbb{E}[Y_i(s_k)|p_i(s_k)]\} = \eta_i(s_k), \quad (4.14)$$

$$\eta_i(s_k) = \mathbf{x}_i^T[\boldsymbol{\alpha} + \boldsymbol{\beta}(s_k)]. \quad (4.15)$$

The parameters $\boldsymbol{\alpha}$ represent fixed effects, whereas the elements of $\boldsymbol{\beta}(s_k)$ are spatially varying random effects. The three components, as given above, can be jointly written as a spatial probit regression model with mixed effects:

$$\Phi^{-1}\{\text{Pr}[Y_i(s_k)=1|\eta_i(s_k)]\} = \mathbf{x}_i^T[\boldsymbol{\alpha} + \boldsymbol{\beta}(s_k)]. \quad (4.16)$$

In a fully Bayesian approach we use the following priors: The fixed effect parameters $\boldsymbol{\alpha}$ have flat, improper, uninformative priors. The spatial random effects $\boldsymbol{\beta}(s_k)$ have zero-centred multivariate conditional autoregressive model (MCAR) priors [Besag, 1993; Mardia, 1988], given by

$$[\boldsymbol{\beta}(s_k)|\boldsymbol{\beta}(s_{-k})] \sim \mathcal{MVN}\left(\frac{\sum_{s_r \in \partial s_k} \boldsymbol{\beta}(s_r)}{n(s_k)}, \frac{\boldsymbol{\Sigma}}{n(s_k)}\right), \quad (4.17)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}(s_1), \dots, \boldsymbol{\beta}(s_N))^T$ and $\boldsymbol{\beta}(s_{-k})$ stands for $\boldsymbol{\beta}$ without the element $\boldsymbol{\beta}(s_k)$; ∂s_k denotes the set of neighbours of s_k and $n(s_k)$ the number of neighbouring voxels. The full conditional posterior distribution of $\boldsymbol{\beta}(s)$ is multivariate normal with hyperparameter $\boldsymbol{\Sigma}$ for which we assign an improper, uninformative Wishart prior,

4.5. A Bayesian log-Gaussian Cox process model

such that $\Sigma^{-1} \sim W(\nu, \mathbf{I})$, with $\nu=20$ degrees of freedom and an identity scale matrix. By Brook's lemma, the joint posterior is therefore proportional to

$$\pi[\boldsymbol{\beta}|\Sigma] \propto \exp \left\{ -\frac{1}{2} \sum_{s_l \sim s_k} [\boldsymbol{\beta}(s_l) - \boldsymbol{\beta}(s_k)]^T \Sigma^{-1} [\boldsymbol{\beta}(s_l) - \boldsymbol{\beta}(s_k)] \right\}. \quad (4.18)$$

In the expression above, $s_l \sim s_k$ indicates two neighbouring voxels that share a common face. Note that the degree of spatial regularisation induced by the MCAR prior, as parameterised by Σ , is not fixed, but rather is estimated from the data.

Our model comprises 11 subject-specific covariates plus clinical subtype (coded as five dummy variables), sex, age, disease duration, seven EDSS subscores and the PASAT score. Thus, the covariate vector \mathbf{x}_i has 16 components. Age, disease duration, EDSS and PASAT scores are mean-centred and equal priors ($\pi_j = 1/J$, for $j=1, \dots, J$) are used for each of the five MS subtypes.

The posterior distribution is estimated via Markov chain Monte Carlo. We use Bayes' theorem and an importance sampling approach (cf. Subsection 2.4.2) to obtain leave-one-out-cross-validation (LOOCV) without having to run the sampler N times. An estimate of the predictive probability that subject i is categorised as subtype j is obtained from a sample from the posterior, weighted to discount the (independent) contribution of subject i to the likelihood; algorithmic details can be found in the appendix of Ge et al. [2014].

In order to reduce the overall size of the images, over 2 million voxels, we subsample every other voxel in each of the xyz -directions, resulting in binary images with voxel size $2 \times 2 \times 2 \text{mm}^3$ for a total of 274,596 voxels.

One should note that the BSGLMM almost falls into the class of univariate models in so far as it assumes that all voxels line up across subjects. Consequently, if one would generate new data from the model it would produce lesions only in voxels where lesions have been observed in the known data. Hence, it does not constitute a truly generative model.

4.5 A Bayesian log-Gaussian Cox process model

We have covered spatial point process theory in Section 2.3 and introduced the log-Gaussian Cox process (LGCP) in Subsection 2.3.2.

The log-intensity function of a LGCP is modelled as a linear combination of spatially varying covariates and a Gaussian process (GP). The role of the GP is to account for any random spatial variation and clustering that is not explained

by the covariates. The basis of our input data are lesion locations of MS patients in conjunction with subject-specific covariates that consist of clinical measures for disease severity as well as demographic attributes. For most applications of point process data, the observed point pattern is a single realisation of some underlying process. In our case, the point patterns constitute multiple realisations of a point process, where each realisation reflects the set of lesion locations of a single subject.

Due to the high dimensionality of our application, we use a sampling algorithm based on Hamiltonian Monte Carlo (HMC). In brief, HMC is based on the description of the dynamics of a physical system by a set of differential equations, Hamilton's equations [Duane et al., 1987]. The algorithm emulates the evolution of a particle system which is characterised by time-dependent position (\mathbf{q}) and momentum (\mathbf{p}) vectors. In our case, \mathbf{q} contains the model parameters of interest and \mathbf{p} is introduced via a d -dimensional Normal distribution $\mathcal{N}_d(0, M)$, where d is the dimensionality of the problem and M the mass matrix. Numerical integration for the HMC method is usually based on the leapfrog algorithm, which alternates between updates of the position and momentum vectors. For a general discussion of the HMC algorithm we refer to Neal [2012]. Details of the sampling technique and particular implementation that we have used can be found in Samartsidis et al. [2017].

The intensity function of the LGCP is given in (2.6). Estimation of the inverse scale parameter ρ of the process is challenging. Therefore, before fitting the model, we estimate (and fix) ρ as well as the exponent δ for the distance via a fit of $c(r, s)$ to the pair-correlation-function of all data [Baddeley et al., 2000; Møller and Waagepetersen, 2004]. This is done by means of minimum contrast estimation [Møller and Waagepetersen, 2007]. Results are summarised in Table 4.2. The remaining mean parameter is given a weakly informative prior proportional to $\mathcal{N}(0, 10^8)$. The HMC algorithm is initialised with the mass matrix set to be the identity. The number of leapfrog steps is fixed at 50 and the step size of the integrator is adjusted during burn-in to achieve an acceptance rate of close to 65%, as suggested by Neal [2012].

The calculation of the variance-covariance matrix in the intensity function ((2.6)) is computationally expensive. Møller et al. [1998] have used an algorithm based on circulant embedding to reduce the computational cost. The main idea behind this method the linear algebra result that the eigenvector decomposition of a circulant matrix is equivalent to its discrete Fourier basis. Although the variance-covariance matrix of the LGCP is not circulant, it can be embedded in a matrix that is circulant. Computation of the square root, inversion and multiplication of

the matrix can thereby be accelerated by using the discrete Fourier transform (DFT) and then manipulating the Fourier coefficients. At the end, the desired results can be obtained by computing the inverse DFT and extracting the appropriate sub-matrix. More details are given in Rue and Held [2005].

In order to evaluate the predictive performance of the model, We use LOOCV and importance sampling, as described in Subsection 2.4.1. Note that the LGCP benefits from a non-parametric framework and hence flexibility in modelling spatial data. Furthermore, its relation to multivariate Gaussian distributions make it more computationally tractable than many other point process models.

Table 4.2: Parameter estimates for the LGCP model, for T_1 -weighted and T_2 -weighted lesion data, respectively. Estimates are based on minimum contrast estimation.

T ₁ -weighted lesion data.					
	CIS	RRMS	PPMS	SPMS	PRMS
δ	1.99	0.78	0.96	1.99	1.10
ρ	0.0057	0.0278	0.0013	0.0043	0.0011

T ₂ -weighted lesion data.					
	CIS	RRMS	PPMS	SPMS	PRMS
δ	1.90	0.65	0.72	1.85	1.20
ρ	0.0067	0.0324	0.247	0.015	0.056

4.6 Application: GeneMSA data

In the following we present results from fitting the NBC, SVM, BSGLMM and LGCP models discussed in the previous sections to the GeneMSA data set.

An important notice must be given regarding the data application in this chapter: Most of this work has been carried out before the latest revision of the McDonald criteria [Lublin et al., 2014] became available in 2014 and in which progressive-relapsing MS (PRMS) was abandoned as a distinct clinical subtype of MS. The GeneMSA data set used in this chapter has been classified according to the previous guidelines and our analysis is based on the (now obsolete) assumption that progressive-relapsing multiple sclerosis (PRMS) is a distinct subtype. A discussion of the data set, including summary statistics is given in Section 3.1.

In the remainder of this section, we first discuss the SVM feature set and posterior inference results from the BSGLMM and LGCP models. Towards the end of the section we summarise and compare the predictive performance, based on five-way classification into MS subtypes, for all methods.

4.6.1 SVM feature inference

Ideally, as classification outcome varies depending on the combination of features used as input, one would like to find the most informative subset of features. This is hindered by a very large number of available features (>900 when considering splitting all geometric summary measures by ROI) and the fact that an exhaustive combinatorial search is computationally infeasible. Furthermore, this would likely lead to over-fitting by selecting a feature set that is specifically suited for the available data set but is unlikely to generalise well to new data. On the other hand, simply using all features does impede the classifier’s predictive ability, since a majority of features would be highly correlated and the feature set would contain a large amount of redundant information. Consider, for example, the case of a patient having only one lesion, which, in the case of CIS patients, is not uncommon. Apart from the fact that the measures over 12 out of 13 ROIs will be zero, summary statistics such as mean and maximum lesion volume will be identical. Thus, feature selection or feature reduction becomes a necessity when optimising the classification procedure.

We consider two ways of reducing the number of input features. First, we select a subset of possible feature combinations. The selection process is based on expected clinical relevance of individual features as determined by neurologists at the Medical Image Analysis Center in Basel. Additionally, we take the magnitude of weights of the support vectors in a linear SVM classification into account. Principally, the examination of weights of individual features for the support vectors across different models can help inform which kinds of features are driving the classification procedure. A comparison of the magnitude of SVM weights provides a qualitative assessment of relative importance of different input features.

Inspection of the weights resulting from a SVM classifier provides a way of estimating the importance of different features. Since it is not possible to obtain an analytic expression for the weights and thus separate them in the non-linear case, any visualisation has to rely on weights based on linear SVMs. Even in the linear case the plots are meant purely as a qualitative way of visualising the importance of input variables relative to one another. No quantitative assessments should be

drawn directly from the magnitude of individual weights.

An example of a pairwise classifier discriminating between the PPMS and SPMS subtypes is presented in Figure 4.2. An account of all ten pairwise SVM classifiers and their respective weights for the full feature set based on whole brain summaries is provided in Section B.1 of the appendix. The weights have been standardised to the occurring maximum across all classifiers which is age in RRMS-versus-SPMS. The relevance of different geometry and intensity based features varies depending on which groups are involved in the classification. For instance, median T_2 lesion volume is very significant in RRMS-versus-PPMS, but much less so in other classifiers. Interestingly, for the same classifier, median T_2 lesion volume positively correlates with RRMS whereas mean T_2 lesion volume shows negative correlation. Lesion volume also shows the largest significance of T_1 -Gd features for any of the classifiers. And while GM volume is not of large importance to most of the groups, it dominates the PPMS-versus-SPMS classifier. Also, T_1 maximum lesion area and mean breadth are prominent in three out of four classifiers concerning RLRLM.

The quadratic means (root-mean-square values) of SVM weights summarised in Figure 4.3 allow for a comparison between different sorts of features and their variability. The three demographic attributes sex, age and disease duration carry a considerable amount of information about MS subtypes. Age and disease duration play a strong role in almost all classifiers and are particularly good indicators for CIS and SPMS. Age appears to be highly predictive of SPMS, which reflects the common disease progression from RRMS to SPMS at later stages of MS, whereas CIS commonly marks the beginning of the disease and is, as expected, negatively correlated with age and disease duration. Intra-lesion intensities seem to be useful for the classification of SPMS and PPMS but appear less relevant for the other groups. For the smallest group, PRMS, median lesion volume of both T_1 - and T_2 -weighted scans appear to be informative for classifications against RRMS and PPMS. PRMS is also the only group that consistently positively correlates with being male. Among the seven EDSS subscores, BRSTMSC and MNMSC seem to carry the highest amount of information as indicated by their larger weights in most classifiers.

In summary, with regard to geometric lesion measures, a comparison across different classifiers indicates that the median is in many cases a better measure than the mean, that the maximum lesion volume, area or mean breadth for a single lesion is more meaningful than the respective minimum, and that the Euler-Poincaré characteristic is more significant than a simple lesion count. Data obtained from

T₁-Gd MRI seems of little predictive value, largely because many patients do not present with active lesions that would be visible on a Gadolinium enhanced scan. This suggests that a reduction in the feature set to a single Gd-based characteristic, such as the Euler-Poincaré number, would be a reasonable choice.

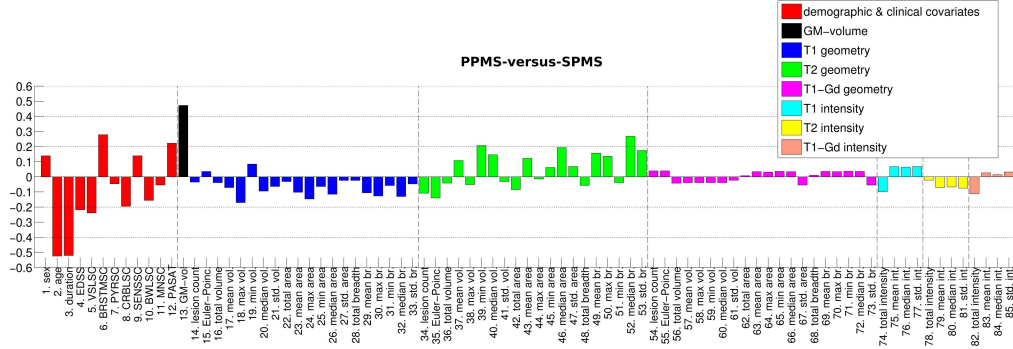


Figure 4.2: Standardised SVM weights for the pairwise classifier involving PPMS and SPMS. All weights have been standardised to the occurring maximum across all classifiers which is age in RRMS-versus-SPMS.

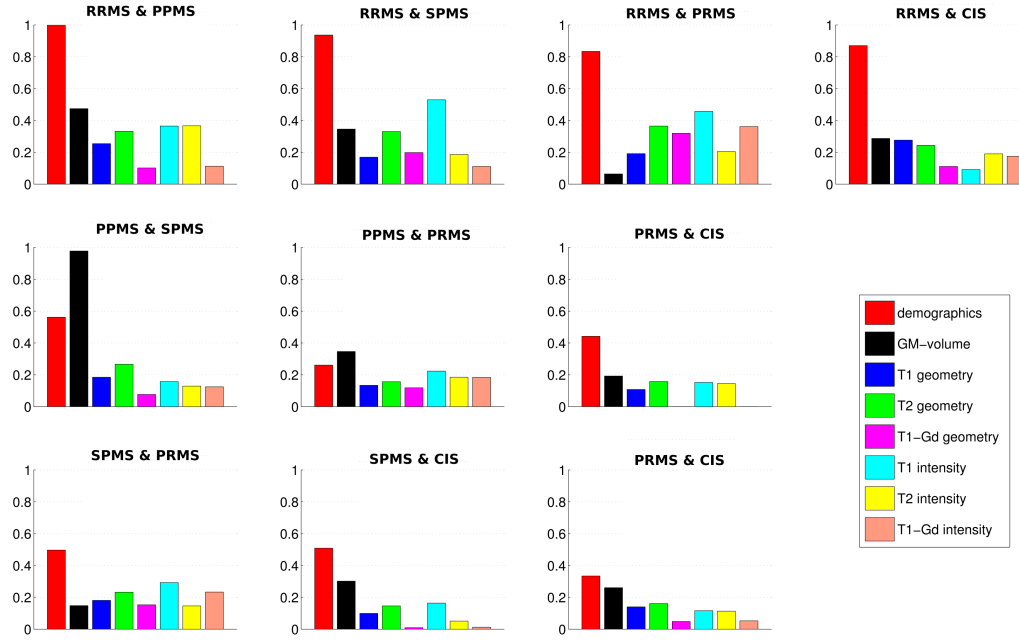


Figure 4.3: Quadratic means of SVM weights for pairwise classifiers, using whole-brain summary measures for all available features. For visualisation purposes, the values are normalised such that the largest is equal to one.

4.6.2 Posterior probability and intensity maps

Figure 4.4 and Figure 4.5 show axial and sagittal slices, respectively, of empirical lesion probabilities based on lesion occurrence and binary lesion masks (top), the estimated mean posterior probabilities from the BSGGLMM model fit (middle) as well as the intensity of the LGCP (bottom). Lesion incidence for CIS differs significantly from the other subtypes. This reflects the fact that CIS patients have the lowest overall lesion load. In contrast, PRMS has the highest per-subject lesion load.

Standardised parameter estimates of subject-specific covariates sex, age, disease duration, PASAT and EDSS measures are given in Figure 4.6. PASAT scores are negatively and EDSS scores positively correlated with lesion occurrence throughout regions of high lesion counts, reflecting higher levels of disability (lower PASAT and higher EDSS scores correspond to more severe MS). Note the strong positive correlation of EDSS with lesion occurrence in the minor and major forceps.

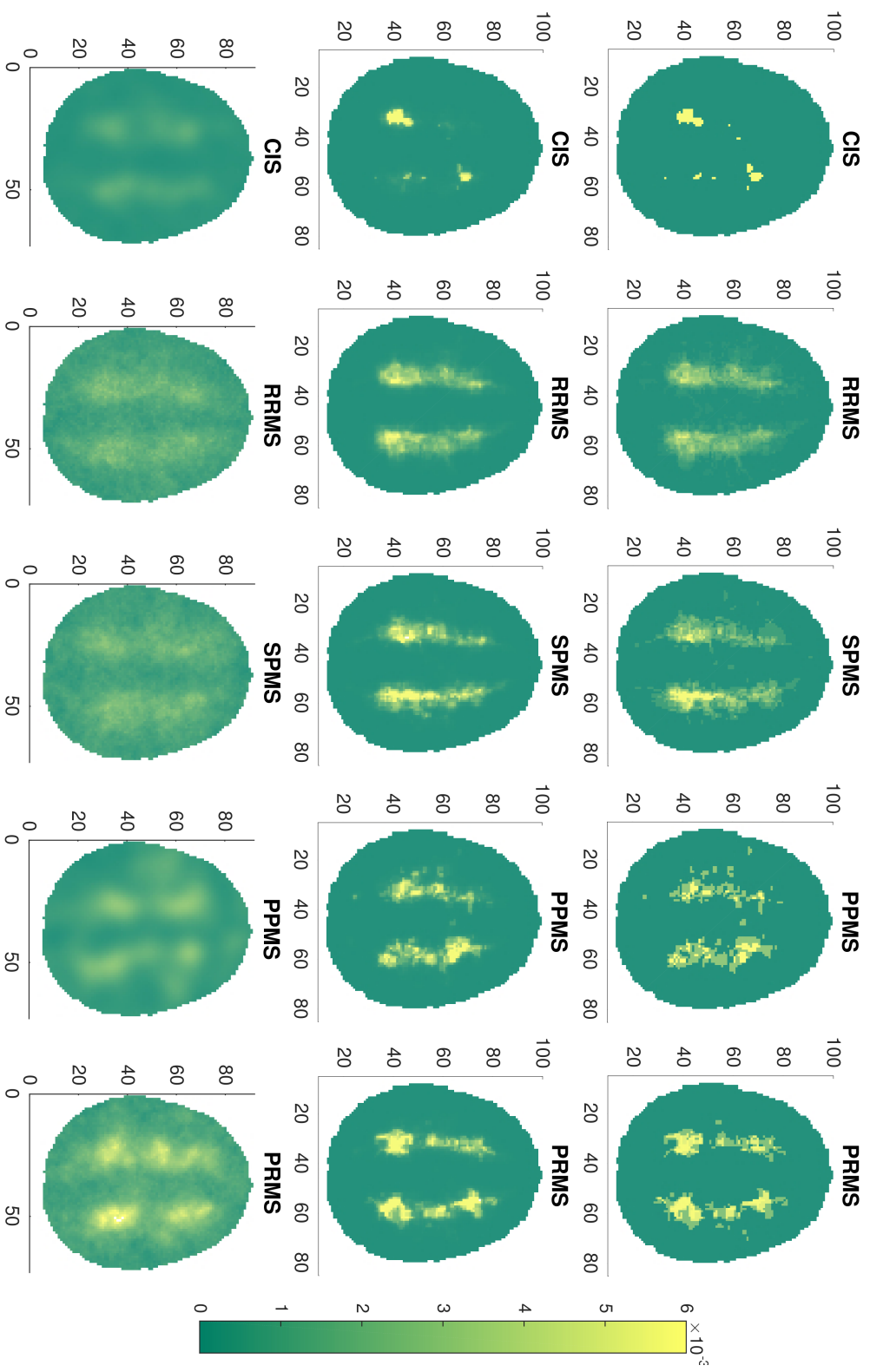


Figure 4.4: Results for the BSGLM and LGCP model fit (sagittal). *Top row:* Empirical lesion prevalence. *Middle Row:* Estimated mean posterior intensity of the BSGLM model. *Bottom row:* Estimated mean posterior intensity of the LGCP model. Axial slice shown at $z=+2$.

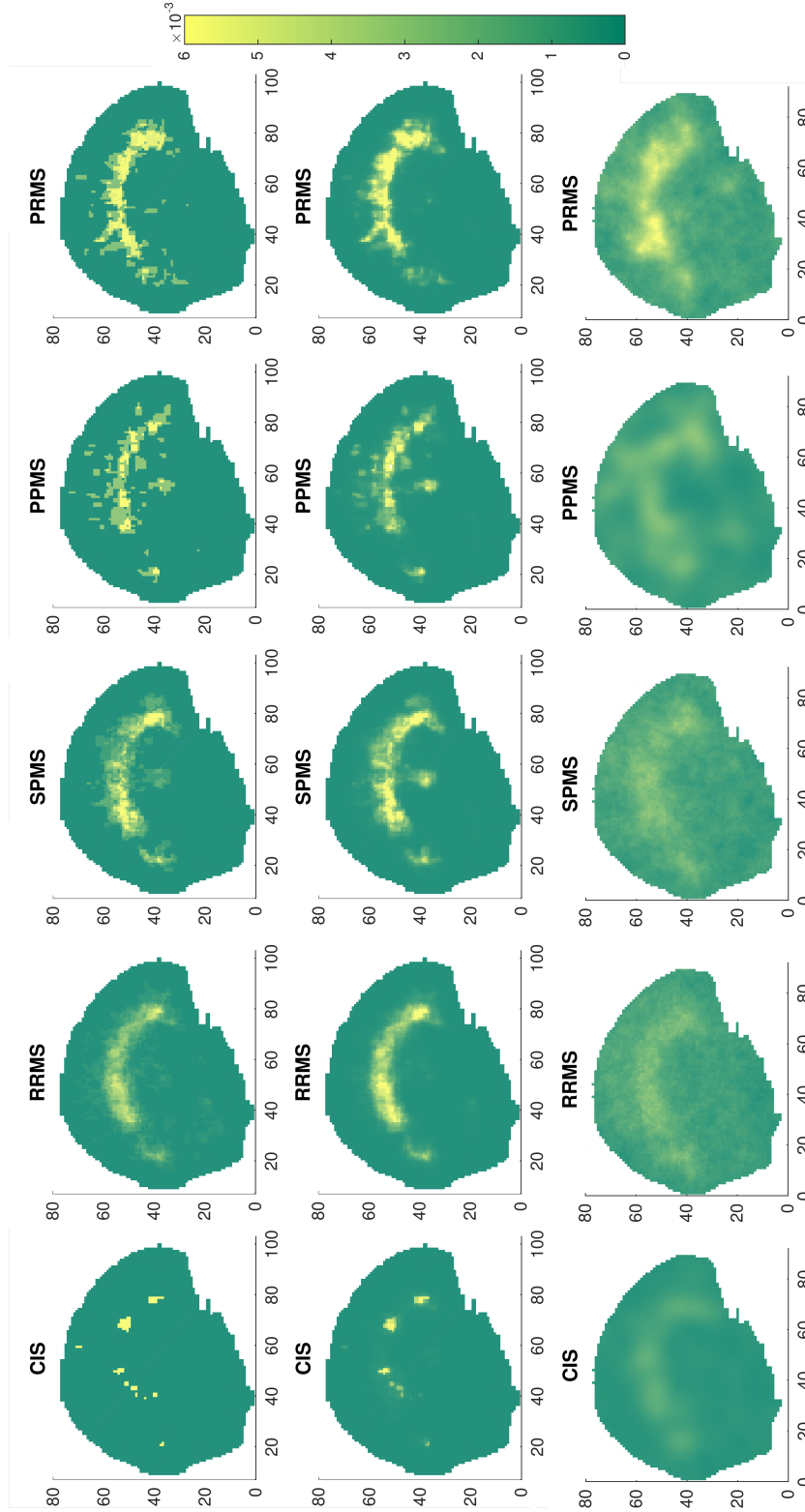


Figure 4.5: Results for the BSLMM and LGCP model fit (sagittal). *Top row:* Empirical lesion prevalence. *Middle Row:* Estimated mean posterior intensity of the LGCP model. *Bottom row:* Estimated mean posterior intensity of the LGCP model. Sagittal slice shown at $x=+20$.

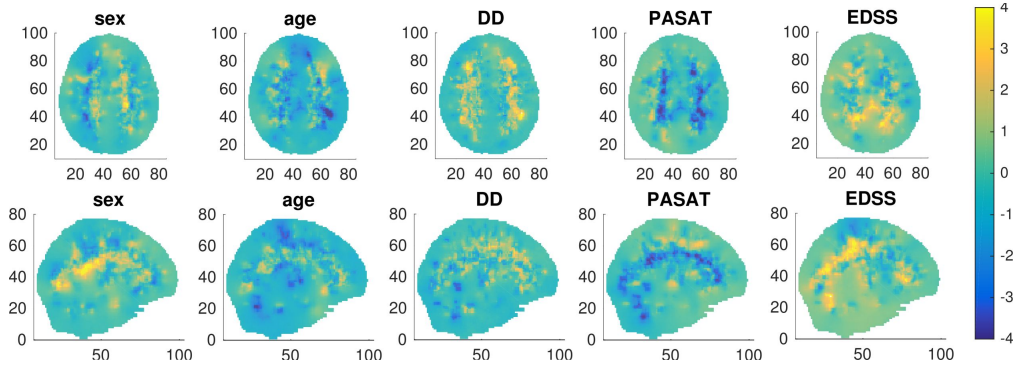


Figure 4.6: Standardised coefficient maps of subject-level covariates as obtained from the BSGLMM model fit. Sex was coded as 0 (male), 1 (female). DD denotes disease duration. The Paced Auditory Serial Addition Test (PASAT) and Expanded Disability Status Scale (EDSS) are clinical scores of disease severity; lower PASAT and higher EDSS scores indicate greater disability. Axial slice at $z=+2$, sagittal slice at $x=+20$.

4.6.3 Prediction Accuracies

We evaluate the predictive performance of all four classification models by computing the confusion matrices. Due to the large discrepancy in membership numbers of the five MS subtypes, the average accuracy is more representative of each method’s performance than the overall accuracy (see Subsection 2.4.1). Also note that chance level of average accuracy lies at 20%.

Table 4.3 and Table 4.4 show confusion matrices for all models based on T_1 and T_2 lesion data, respectively. In case of the SVM classifier, results for two combinations of features from multiple imaging modalities are shown in the tables. A comparison of classification performance reveals the superiority of the spatially informed approaches while the naïve Bayesian approach performs only slightly above chance level. The NBC results are based on employing a lesion mask (only voxels with at least two lesions) on the full MRI data. In contrast, when using all voxels, the NBC yields predictions where every single subject is classified into the largest subtype, RRMS, resulting in an average accuracy exactly equal to chance.

The feature set for the SVM (M4) (see Table 4.1) classifier comprises GM volume by lobar ROI’s, T_1 and T_2 lesion count and lesion volume by WM ROI’s but excludes any demographic or clinical covariates. It reflects the performance of SVM on lesion data when using only traditional measures such as lesion load and count combined with brain atrophy as measured by GM volume-to-whole-brain ratios. The achieved average classification accuracy of 39.4% is well above chance level and indicates that not the covariates but instead the information contained in

MRI data is predominantly driving the predictions.

The feature configuration for SVM (M7) is the one that showed the highest prediction accuracy among our selection of feature sets. The feature set includes GM volume, median T₂ lesion volume and T₂ EP characteristic split into WM ROI's and the following whole brain summaries: standard deviation of T₁ mean breadth, median of T₂ mean breadth, T₁ and T₁-Gd mean intra-lesion intensities; alongside all available demographic and clinical covariates; resulting in an average accuracy of 47.8%.

Although the SVM classifiers are performing better than a naïve mass-univariate approach in terms of average classification accuracy, they struggle particularly with predicting the PPMS and PRMS subtypes. As with the NBC, the majority of misclassified subjects are categorised as belonging to the largest group, RRMS.

The BSGLMM shows strong prediction results with an average accuracy of 78% based on T₁-weighted and 82% based on T₂-weighted data. The BSGLMM confusion matrices show that misclassification predominantly occurs into the CIS subtype. Misclassified patients tend to have fewer and smaller lesions than those that are correctly classified, which is consistent with the clinical presentation of the CIS subtype. Note that classification results improve when using the empirical proportions of group-membership instead of an equal prior, achieving 81.8% (85.5%) average (overall) accuracy on T₁ data and 83.7% (79.9%) on T₂ data.

The LGCP model does not consider any covariates, which, one could expect, might put it at a disadvantage when doing predictions. With respect to the T₁ data, the classifier of the LGCP model performs well on the largest subtype (RRMS) but has difficulty with the two smallest groups (CIS, PPMS). This can at least in part be attributed to the small number of data points available for these subtypes, e.g. there are only eight CIS patients with T₁ lesions in the data set. A further difficulty arises from the fact that only about half as many lesions are visible on T₁-weighted images compared to T₂-weighted scans. The comparatively much higher prediction accuracy in the case of T₂ lesions indicates that additional data would likely increase model performance with respect to the T₁ lesions.

Regarding the T₂ data, the LGCP's predictive accuracy reaches 84.7% overall and 74.7% when averaged across groups. Among the four models considered here, the LGCP is also the closest to a generative model for lesion data, i.e. when simulating new data, it would give much more realistic predictions than the BSGLMM for instance, which assumes independent lesion data conditional on (spatially regularised) coefficients.

Table 4.3: Confusion matrices and prediction accuracies for different classifiers based on T₁ lesion data (except for SVM).

NBC: Overall & average accuracy: 0.580 & 0.219 .					
	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.000	0.400	0.400	0.000	0.200
RRMS	0.018	0.799	0.018	0.128	0.037
PPMS	0.000	0.923	0.077	0.000	0.000
SPMS	0.024	0.756	0.049	0.122	0.049
PRMS	0.200	0.700	0.000	0.000	0.100

SVM (M4): Overall & average accuracy: 0.536 & **0.394**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.454	0.364	0.000	0.182	0.000
RRMS	0.139	0.595	0.064	0.162	0.041
PPMS	0.077	0.385	0.231	0.308	0.000
SPMS	0.070	0.209	0.116	0.488	0.116
PRMS	0.000	0.500	0.000	0.300	0.200

BSGLMM: Overall & average accuracy: 0.654 & **0.783**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	1.000	0.000	0.000	0.000	0.000
RRMS	0.348	0.598	0.030	0.024	0.000
PPMS	0.083	0.000	0.917	0.000	0.000
SPMS	0.216	0.054	0.027	0.703	0.000
PRMS	0.100	0.100	0.100	0.000	0.700

LGCP: Overall & average accuracy: 0.753 & **0.510**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.250	0.375	0.125	0.125	0.125
RRMS	0.056	0.850	0.069	0.019	0.006
PPMS	0.167	0.333	0.333	0.083	0.083
SPMS	0.071	0.119	0.119	0.667	0.024
PRMS	0.111	0.222	0.111	0.111	0.445

Table 4.4: Confusion matrices and prediction accuracies for different classifiers based on T₂ lesion data (except for SVM).

NBC: Overall & average accuracy: 0.592 & **0.280**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.000	0.500	0.200	0.000	0.300
RRMS	0.012	0.781	0.018	0.116	0.073
PPMS	0.000	0.769	0.000	0.154	0.077
SPMS	0.024	0.585	0.024	0.220	0.146
PRMS	0.000	0.600	0.000	0.000	0.400

SVM (M7): Overall & average accuracy: 0.560 & **0.478**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.818	0.182	0.000	0.000	0.000
RRMS	0.162	0.584	0.058	0.081	0.116
PPMS	0.000	0.231	0.308	0.231	0.231
SPMS	0.023	0.093	0.116	0.581	0.186
PRMS	0.000	0.400	0.200	0.300	0.100

BSGLMM: Overall & average accuracy: 0.748 & **0.823**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	1.000	0.000	0.000	0.000	0.000
RRMS	0.238	0.713	0.006	0.043	0.000
PPMS	0.083	0.000	0.917	0.000	0.000
SPMS	0.162	0.000	0.054	0.784	0.000
PRMS	0.200	0.000	0.000	0.100	0.700

LGCP: Overall & average accuracy: 0.847 & **0.747**.

	CIS	RRMS	PPMS	SPMS	PRMS
CIS	0.600	0.100	0.300	0.000	0.000
RRMS	0.035	0.896	0.017	0.029	0.023
PPMS	0.154	0.154	0.692	0.000	0.000
SPMS	0.023	0.116	0.093	0.767	0.000
PRMS	0.111	0.111	0.000	0.000	0.778

4.7 Discussion

In clinical applications, objective classification algorithms have the potential to improve decision making in general and to aid in the prognosis of disease progression or treatment outcomes.

In this chapter, we applied four different classifiers to a data set of multiple sclerosis patients. The results indicate that MRI data of MS lesions contain more information about the disease than currently utilised in clinical assessments. We used several sources of information based on MRI data, geometric and texture measures of individual lesions, demographic and clinical characteristics and spatial information about each lesion’s centre-of-mass to classify each patient into one of five MS subtypes. In general, this is a challenging prediction task, as the differences between MS subtypes in terms of imaging phenotypes are subtle.

We found moderate classification accuracies for the naïve Bayesian and SVM classifiers. In contrast, the Bayesian spatial generalised linear model and the log-Gaussian Cox process model showed significantly better outcomes, an indication that spatial information is crucial to a successful classification scheme.

Whether supervised MRI-based pattern recognition can achieve the level of sensitivity and specificity needed in order to be integrated into clinical applications remains an open question. A shift from single predictive models to ensembles of classifiers may produce more generalisable results by averaging the decisions of numerous predictive models; see for example Koutsouleris et al. [2010]. With regard to MS, incorporating other characteristics and quantitative measures of interest such as biomarkers [Tintoré et al., 2008] and genetic factors [Kelly et al., 1993] into the models could further improve prediction accuracy.

Our application study has shown that widely used machine learning techniques based on support vector machines perform worse than a spatial point process model, even when combining the demographic characteristics and clinical scores of a subject with a large set of lesion specific geometric and intra-lesion intensity features.

In contrast to standard mass-univariate methods, the Bayesian spatial models discussed here exploit the spatial structure of MS lesion maps and take into account the binary nature of lesion data without an arbitrary smoothing parameter. The BSGLMM explicitly includes covariates and spatially varying coefficients. Furthermore, it is able to provide spatial information, e.g. estimates for the spatially varying effects of age, sex, disease duration, EDSS and PASAT, which current empirical methods cannot.

The use of affine registration to align subjects to a common space can have negative consequences with respect to the BSGLMM model, because the rigid body transformation limits the degree of alignment of different brains. A future direction of this work could investigate whether the use of high-dimensional non-linear registration, which can better align brain structures across subjects, will improve upon the current predictive accuracies and covariate maps of the BSGLMM. However, a potential drawback of non-linear registration lies in the fact that lesion volumes may not change proportionally as they do with affine registration. A model for binary lesion data that accounts for local volume changes, as does Voxel Based Morphometry [Ashburner and Friston, 2000] for example, would be able to avoid any effects of potential shrinkage or expansion of individual lesions.

The main challenge of fitting a Cox process model such as the LGCP to spatial point pattern data lies in estimating intractable likelihoods, which is most commonly done via computationally intensive MCMC methods. An interesting extension of the spatial point process approach would be to incorporate subject- as well as lesion-specific attributes into the model, thereby combining the separate kinds of information that the four methods discussed here rely upon. Chapter 5 introduces such an extension based on a Gamma random field approach and applications can be found in Chapter 6.

Our formulation of the LGCP model does not account for any covariates and instead solely relies on point locations as input data. One could expect that utilising the information contained in subject-specific measures such as demographic characteristics and clinical scores would further improve the model fit. Explanatory variables can be incorporated into the log-Gaussian Cox process in a regression framework as proposed, for example, by Samartsidis et al. [2017]. In order to specify a log-linear model over the volume of the brain \mathcal{B} , the authors model the subject-level intensity function at each point $y \in \mathcal{B}$ according to: $\log \lambda_i(y) = \beta_0 + \sum_{p=1}^P \beta_p(y) z_{i,p}$; where β_0 is a baseline parameter and β_p are the regression coefficients. A similar approach could be used to extend our LGCP model.

Chapter 5

Poisson/Gamma random field models for spatial point data

5.1 Introduction

In Chapter 4 we examined how two machine learning techniques, a mass-univariate approach and a Bayesian log-Gaussian Cox process model fared in predicting multiple sclerosis subtypes based on different kinds of information present in the data. In this chapter we consider a more flexible point process model, the Poisson/Gamma random field (PGRF) model. Specifically, we propose and evaluate several extensions of the basic, independent PGRF formulation found in Subsection 2.3.3. Our aim is to combine the spatial information about lesion location that is fundamental to point process models with subject- and lesion-specific attributes that are more common to machine learning models.

This chapter builds on the main statistical theory of spatial Poisson processes as summarised in Section 2.3 and is organised as follows: The hierarchical extension of the independent PGRF model is outlined in Section 5.2. In Section 5.3 we propose a way to incorporate global covariates that contain attributes which are unique to a given realisation of a point pattern, i.e. subject-level characteristics. The focus of Section 5.4 is the inclusion of marks as a means of jointly modelling point-specific attributes together with the traditional spatial point process. Specifics of the implementation of the HPGRF model and its variants (including covariates and/or marks) are discussed in Section 5.5. In Section 5.6 we compare and evaluate different model variants on a simulated set of data. Finally, Section 5.7 contains a discussion of the merits and limitations of the model and concludes with a brief outlook on future work.

5.2 The hierarchical PGRF model

We have described spatial point processes in general and the independent Poisson/Gamma random field model in particular in Section 2.3. Kang et al. [2014a] proposed an extension to the PGRF model of Wolpert and Ickstadt [1998b] to include a hierarchy and therefore interdependence between several types of point patterns. The authors jointly modelled fMRI data of five emotional states (sadness, happiness, anger, fear, disgust). For our purposes, the different types of point patterns correspond to different subtypes, disease categories or pathologies of multiple sclerosis.

As laid out in Section 2.3 we consider a doubly stochastic (Cox) process, such that the spatial intensity measure $\Lambda(dy)$ is described by a density function that is a positive random field. That is, each point process realisation is a realisation of an inhomogeneous Poisson process with intensity function $\lambda(y)$.

The hierarchical Poisson/Gamma random field (HPGRF) model belongs to the class of non-parametric models which entails greater flexibility compared to parametric models and only weak assumptions about the data. The hierarchical structure of the HPGRF model can be interpreted analogous to a random-effects model. The type-specific Gamma random fields are independently drawn realisations of a shared, population level Gamma random field. This allows for the joint analysis of multiple types of point patterns where one assumes a connection between different groups/types and allows for the sharing of information—in the form of positive correlations—across types, i.e. the population-level \mathcal{GRF} provides a link between the random intensities of individual types.

For each type j , the spatial Poisson point processes \mathbf{Y}_j is described by type-specific Gamma random fields $\Gamma_j(dx)$ which themselves are realisations of a common population level Gamma random field $\Gamma_0(dx)$. The general model formulation can be summarised by the following relations

$$\mathbf{Y}_j \sim \mathcal{PP}\{\mathcal{B}, \Lambda_j(dy)\}, \quad (5.1)$$

$$\Gamma_j(dx) \stackrel{iid}{\sim} \mathcal{GRF}\{\Gamma_0(dx), \tau\}, \quad (5.2)$$

$$\Gamma_0(dx) \sim \mathcal{GRF}\{\alpha(dx), \beta\}; \quad (5.3)$$

with the type specific intensity measure expressed as

$$\Lambda_j(dy) = \int_{\mathcal{B}} K_{\sigma_j^2}(dy, x) \Gamma_j(dx). \quad (5.4)$$

Both $\Gamma_0(dx)$ and $\Gamma_j(dx)$ share the same support, i.e. latent sources (jump locations) θ_m . The group specific Gamma random fields have magnitudes (jump heights) $\eta_{j,m}$ associated with these point locations. The $\eta_{j,m}$ are Gamma distributed with individual shape parameters ν_m for each jump location and a common inverse scale parameter τ . We thus write $[\eta_{j,m}|\nu_m, \tau] \stackrel{iid}{\sim} \text{Ga}(\nu_m, \tau)$. The ν_m are themselves part of the population level Gamma random field and represent the jump heights for $\Gamma_0(dx)$ at each θ_m . This implies that the Gamma random fields for both levels of the hierarchy can be expressed as

$$\Gamma_0(dx) = \sum_{m=1}^{\infty} \nu_m \delta_{\theta_m}(dx), \quad (5.5)$$

$$\Gamma_j(dx) = \sum_{m=1}^{\infty} \eta_{j,m} \delta_{\theta_m}(dx). \quad (5.6)$$

The population level \mathcal{GRF} is parameterised by the base measure $\alpha(dx)$, which we take to be Lebesgue measure, and rate parameter β . The construction of $\Gamma_0(dx)$ follows the inverse Lévy measure algorithm as described in Section 2.3. We also use the data augmentation scheme presented in Section 2.3 for estimation of the posterior distribution. In summary, the complete model is formulated as follows:

$$[(\mathbf{Y}_j, \mathbf{X}_j) | \{(\eta_{j,m}, \theta_m)\}_{m=1}^M, \sigma_j^2] \sim \mathcal{PP} \left\{ \mathcal{B} \times \mathcal{B}, K_{\sigma_j^2}(dy, x) \sum_{m=1}^M \eta_{j,m} \delta_{\theta_m}(dx) \right\}, \quad (5.7)$$

$$[\eta_{j,m} | \nu_m, \tau] \stackrel{iid}{\sim} \text{Ga}(\nu_m, \tau), \quad (5.8)$$

$$\{(\theta_m, \nu_m)\}_{m=1}^M \sim \text{invLévy}\{\alpha(dx), \beta\}, \quad (5.9)$$

where $(\mathbf{Y}_j, \mathbf{X}_j)$ denotes the joint process resulting from the data augmentation scheme. Evaluation of the population level intensity in case of multi-type point patterns follows directly from the reweighted mean kernel measure

$$\bar{K}(dy, x) = \frac{1}{J} \sum_{j=1}^J K_{\sigma_j^2}(dy, x) \quad (5.10)$$

and can be written as

$$\Lambda_0(dy) = \sum_{j=1}^J \mathbb{E}[\Lambda_j(dy) | \Gamma_0, \sigma_j^2, \tau] = \frac{1}{\tau} \int_{\mathcal{B}} \bar{K}(dy, x) \Gamma_0(dx). \quad (5.11)$$

5.3. Including covariates

Let $(\mathbf{y}_j, \mathbf{x}_j) = \{(y_{j,i,l}, x_{j,i,l})\}_{l=1}^{L_i}\}_{i=1}^{N_j}$ denote all available realisations of the augmented process $(\mathbf{Y}_j, \mathbf{X}_j)$; in the case of lesion data $(\mathbf{y}_j, \mathbf{x}_j)$ corresponds to the combined point locations of all subjects belonging to subtype j . Denote the vector of jump locations $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^M$, the group-specific magnitudes $\boldsymbol{\eta}_j = \{\eta_{j,m}\}_{m=1}^M$ and the jump heights for the shared random field $\boldsymbol{\nu} = \{\nu_m\}_{m=1}^M$. The target distribution is the joint density of $\{\mathbf{x}_j\}_{j=1}^J, \{\boldsymbol{\eta}_j\}_{j=1}^J, \{\sigma_j^2\}_{j=1}^J, \boldsymbol{\theta}, \boldsymbol{\nu}, \tau$ and β , given $\{\mathbf{y}_j\}_{j=1}^J$. It is proportional to

$$\begin{aligned} & \prod_{j=1}^J \left\{ \pi(\mathbf{y}_j, \mathbf{x}_j | \boldsymbol{\eta}_j, \boldsymbol{\theta}, \sigma_j^2) \times \pi(\sigma_j^2) \times \pi(\boldsymbol{\eta}_j | \boldsymbol{\nu}, \tau) \right\} \times \pi(\boldsymbol{\theta}) \times \pi(\tau) \times \pi(\boldsymbol{\nu} | \beta) \times \pi(\beta) \\ & \propto \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[- \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} \right] \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \right\} \\ & \quad \times \prod_{j=1}^J \left\{ \pi(\sigma_j^2) \prod_{m=1}^M \left[\frac{\tau^{\nu_m}}{\Gamma(\nu_m)} \eta_{j,m}^{\nu_m-1} \exp(-\tau \eta_{j,m}) \right] \right\} \\ & \quad \times \pi(\tau) \times \exp[-E_1(\beta \nu_M)] \prod_{m=1}^M \left\{ \frac{1}{\nu_m} \exp[-\nu_m \beta] \right\} \times \pi(\beta). \end{aligned} \quad (5.12)$$

Note that $\pi(\mathbf{y}_j, \mathbf{x}_j | \boldsymbol{\eta}_j, \boldsymbol{\theta}, \sigma_j^2)$ is the density of $(\mathbf{Y}_j, \mathbf{X}_j)$ with respect to a unit rate Poisson process [Møller and Waagepetersen, 2004; Kang et al., 2014a]. $\pi(\boldsymbol{\theta})$ is a density with respect to the product measure $\prod_{m=1}^M \tilde{\alpha}(dx)$ and we take $\pi(\boldsymbol{\theta}) \propto 1$. We further assume the base measure $\alpha(dx)$ to be Lebesgue measure. This means that $\alpha(dx)$ is non-atomic and therefore the θ_m are distinct. The densities of the other parameters are with respect to Lebesgue measure and $\mathbb{I}_{\theta_m}(x_{j,i,l})$ denotes the indicator function with $\mathbb{I}_{\theta_m}(x_{j,i,l}) = 1$ if $\theta_m = x_{j,i,l}$ and 0 otherwise.

5.3 Including covariates

Some of the benefits of model-based approaches, such as the independent and inter-dependent variants of the Poisson/Gamma random field model, lie in their flexibility and lack of strong prior assumptions about the structure of the data. However, with respect to multiple realisations these models implicitly assume that the underlying intensity function is exactly the same across all realisations. In the case of a cohort of patients with MS, each patient's lesion pattern represents an independent realisation of a PGRF. The HPGRF model is able to capture differences between subgroups of these point patterns, but systematic differences between separate realisations (i.e. subjects) within a type or subgroup are not registered. With regard to lesion data,

such differences can arise due to, for instance, demographic or clinical particularities of individual patients.

Apart from the exact spatial location of objects or events, it is often desirable to include further information. This additional information could, for example in ecological studies, relate environmental factors such as soil properties or pollution levels to the observed spatial distribution of certain plants. Diggle [1990] and Diggle and Rowlingson [1994] proposed a multiplicative model for the intensity function of the spatial point process with the goal to estimate elevated risk of disease due to environmental factors. Spatial covariates have also been used for the analysis of tropical rainforests [Burslem et al., 2001] or the habitats of koalas [Moore et al., 2010; Illian et al., 2012]. All of these applications consider a single realisation of a two-dimensional point process. In our case and with regard to neuroimaging data in general, we are dealing with multiple realisations (i.e. subjects) and thus would like to consider external covariates that are specific to each realisation.

We adopt a multiplicative approach similar to Diggle and Rowlingson [1994]. By introducing a scaling parameter to the intensity function the model can account for deviations from the population mean when assuming standardised covariates. We model contributions from subject-specific attributes to the intensity function with a log-link dependence as in univariate Poisson regression models.

5.3.1 Poisson regression

Following the theory of Generalised Linear Models (GLM), see for example McCullagh and Nelder [1989] for a detailed textbook reference, we assume the response distribution to be in the exponential family. The canonical link function for a Poisson model is an exponential function. Let $z_{j,i,p}$, for $p = 1, \dots, P$, denote the set of covariates belonging to subject i of subtype j . Regression coefficients are denoted as γ_p and shared across subjects as well as subtypes. Hence, the intensity function for a single realisation can be written as

$$\lambda_{j,i}(y) = \exp \left[\sum_{p=1}^P z_{j,i,p} \gamma_p \right] \int_{\mathcal{B}} k_{\sigma_j^2}(y, x) \Gamma_j(dx). \quad (5.13)$$

Analogously, the intensity measure for a single realisation becomes

$$\Lambda_{j,i}(dy) = \exp \left[\sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right] \Lambda_j^*(dy), \quad (5.14)$$

5.4. Marked hierarchical PGRF models

where the intensity measure $\Lambda_j^*(dy)$ is the same as in (5.4) of the original model, i.e.

$$\Lambda_j^*(dy) = \int_{\mathcal{B}} K_{\sigma_j^2}(dy, x) \Gamma_j(dx) = \int_{\mathcal{B}} K_{\sigma_j^2}(dy, x) \sum_{m=1}^M \nu_{j,m} \delta_{\theta_{j,m}}(dx). \quad (5.15)$$

For multiple realisations, the intensity measure again is given by

$$\Lambda_j(dy) = \prod_{i=1}^{N_j} \Lambda_{j,i}(dy). \quad (5.16)$$

Finally, the full joint density can be written as

$$\begin{aligned} \pi(\cdot) \propto & \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[- \exp \left(\sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right) \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} + \sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right] \right. \\ & \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \Big\} \\ & \times \prod_{j=1}^J \left\{ \pi(\sigma_j^2) \prod_{m=1}^M \left[\frac{\tau^{\nu_m}}{\Gamma(\nu_m)} \eta_{j,m}^{\nu_m-1} \exp(-\tau \eta_{j,m}) \right] \right\} \\ & \times \pi(\tau) \times \exp[-E_1(\beta \nu_M)] \prod_{m=1}^M \left\{ \frac{1}{\nu_m} \exp[-\nu_m \beta] \right\} \times \pi(\beta). \end{aligned} \quad (5.17)$$

5.4 Marked hierarchical PGRF models

We have briefly introduced the concept of marked point processes at the end of Section 2.3. In the present section, we first add a constant and spatially independent mark process to the model described in Section 5.2 and then introduce the fully spatially varying process where the mark distribution also depends on the underlying spatial intensity.

We carefully draw the distinction between covariates and marks. Covariates are general properties of a whole realisation, like a demographic variable of an MS patient (e.g. age, gender), or large-scale non-local environmental properties (e.g. soil properties, pollution levels). Marks are small-scale, local attributes of individual objects or events, and in our setting are directly attached to a point location. Marks may or may not be independent of the point process. For example, a simple marked Cox process comprises a set of independent and identically distributed marks w according to some density $\pi(w)$ that are also independent from a Cox process with

intensity function $\lambda(y)$. Then the density of the joint process is simply given by $\pi(w)\lambda(y)$ [Møller and Waagepetersen, 2004]. Note that the HPGRF model can already be regarded as a marked point process model in the sense of having attached a categorical (type-specific) mark to each point location. In the following, we always mean continuous, spatially dependent marks when referring to the “marked” process.

Marks that depend on the intensity function of the point process can be used to model quantities that are correlated (positively or negatively) with the probability of observing points in any given region. Diggle et al. [2010] have used intensity-dependent marks in conjunction with preferential sampling in a geostatistics setting. Other recent examples in the literature of models for point patterns that include spatially dependent marks are Ho and Stoyan [2008]; Myllymäki and Penttinen [2009]. In terms of MS lesion data, high rates of lesion incidence may be indicative of areas of the brain where lesions are likely to persist and grow over time. Additionally, several small lesions in early stages of MS may merge into fewer bigger lesions as the disease progresses. Therefore, when modelling lesional point patterns, marks that carry information about the size of individual lesions may be able to explain some of the additional variance in the data.

5.4.1 The marked HPGRF model with additional covariates

In the previous section we have related point process intensities to subject-specific attributes by introducing spatially non-varying covariates such as demographic data or clinical scores. By associating a mark with each point we can also model point-specific attributes.

For all of the following, we keep these non-spatial regressors as part of the model. However, the model without covariates can easily be obtained by setting all predictors $z_p \equiv 0$. We call the non-marked, spatial point process part of the model the *ground process*, as opposed to the combined or *marked process*, which includes a mark distribution placed at each point $y_{j,i,l} \in \mathbf{Y}_j$. We consider only marks that are *independently marked*, following definition (3) in Cronie and van Lieshout [2016] which defines a marked process as being independently marked if, given the locations of the corresponding points of the ground process, the marks are independent random variables with a distribution that depends only on the point location (cf. definition 6.4III in Daley and Vere-Jones [2003]). Therefore, any interaction in the marked process is due to interactions in the ground process. Note that we are not assuming any interaction between marks themselves. From before, the density for the ground process is given by (5.17).

5.4.2 Assuming an independent mark distribution

If, in addition to the marks being independent of one another, the mark process with marks $w \in \mathcal{W}$ does not depend on the spatial location, the combined process \mathbf{Y} is said to have the random labelling property [Cronie and van Lieshout, 2016]. If we further assume that the marks are normally distributed, then $w_{j,i,l}$ are independent Gaussian marks with a simple, spatially constant mark distribution with a type-specific baseline a_j and random fluctuations $\epsilon_{j,i,l}$ such that $w(y_{j,i,l}) \equiv w_{j,i,l} = a_j + \epsilon_{j,i,l}$. For lesion l of subject i in type j , the random error $\epsilon(y_{j,i,l}) \equiv \epsilon_{j,i,l}$ has a Gaussian distribution with mean 0 and variance σ_w^2 and we write

$$w_{j,i,l} \stackrel{iid}{\sim} \mathcal{N}(a_j, \sigma_w^2). \quad (5.18)$$

In our application of the model to MS lesion data, we take the mark on each point to reflect individual, volume-adjusted lesion size $V_l^* = V_l^{1/3}$ for lesion l of a given subject. Assume that the mark distribution differs for different MS subtypes j and denote prior distributions on the parameters a_j by $\pi(a_j)$. The mark space \mathcal{W} comprises the non-negative real line up to the (adjusted) total brain volume, i.e. $w(y) \in [0, |\mathcal{B}|^{1/3}]$. Note that realistic values of lesion volume are of course much smaller than the total brain volume. To enforce positive mark values one can model the marks instead as log-Gaussian random variables, resulting in $w_{j,i,l} \stackrel{iid}{\sim} \log\text{-}\mathcal{N}(a_j, \sigma_w^2)$. Denote the set of marks for type j as $\mathbf{w}_j = \{\{w_{j,i,l}\}_{l=1}^{L_i}\}_{i=1}^{N_j}$. The joint likelihood is schematically given by $\pi(\mathbf{y}_j, \mathbf{w}_j) = \pi(\mathbf{y}_j) \times \pi(\mathbf{w}_j)$, leading to the following expression for the joint density:

$$\begin{aligned} \pi[\cdot] \propto & \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[- \exp \left(\sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right) \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} + \sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right] \right. \\ & \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \Big\} \\ & \times \prod_{j=1}^J \left\{ \pi(\sigma_j^2) \prod_{m=1}^M \left[\frac{\tau^{\nu_m}}{\Gamma(\nu_m)} \eta_{j,m}^{\nu_m-1} \exp(-\tau \eta_{j,m}) \right] \right\} \times \pi(\tau) \\ & \times \exp[-E_1(\beta \nu_M)] \prod_{m=1}^M \left\{ \frac{1}{\nu_m} \exp[-\nu_m \beta] \right\} \times \pi(\beta) \\ & \times \prod_{j=1}^J \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \frac{1}{\sigma_w} \exp \left[-\frac{(\ln w_{j,i,l} - a_j)^2}{2\sigma_w^2} \right] \times \pi(a_j) \right\} \times \pi(\sigma_w^2), \end{aligned} \quad (5.19)$$

where the last line captures the contribution from marks and the others represent

the ground process.

5.4.3 The intensity-marked HPGRF model

Ho and Stoyan [2008] introduced a density-dependent marked Cox process with marks w on points y of a log-Gaussian Cox process as $w(y) = a + b\lambda(y) + \epsilon(y)$. The mark distribution therefore is spatially dependent and driven by the intensity of the point process. The case of $b > 0$ ($b < 0$) results in the marks being large (small) in regions of high point density.

Adapting the intensity-marked process to the HPGRF model, we make the baseline parameter a and the interaction coefficient b type dependent. The log-marks are modelled as $\ln w_{j,i,l} = a_j + b_j\lambda(y_{j,i,l}) + \epsilon(y_{j,i,l})$ with the mark distribution thus given as

$$w_{j,i,l} | \lambda(y_{j,i,l}) \sim \log\mathcal{N}(a_j + b_j\lambda(y_{j,i,l}), \sigma_w^2). \quad (5.20)$$

Finally, the full joint density of the intensity marked hierarchical Poisson/Gamma random field (imHPGRF) model is expressed as

$$\begin{aligned} \pi[\cdot] \propto & \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[- \exp \left(\sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right) \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} + \sum_{p=1}^P z_{j,i,p} \gamma_{j,p} \right] \right. \\ & \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \Big\} \\ & \times \prod_{j=1}^J \left\{ \pi(\sigma_j^2) \prod_{m=1}^M \left[\frac{\tau^{\nu_m}}{\Gamma(\nu_m)} \eta_{j,m}^{\nu_m-1} \exp(-\tau \eta_{j,m}) \right] \right\} \times \pi(\tau) \\ & \times \exp[-E_1(\beta \nu_M)] \prod_{m=1}^M \left\{ \frac{1}{\nu_m} \exp[-\nu_m \beta] \right\} \times \pi(\beta) \\ & \times \prod_{j=1}^J \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \frac{1}{\sigma_w} \exp \left[-\frac{(\ln w_{j,i,l} - c_{j,i,l})^2}{2\sigma_w^2} \right] \times \pi(a_j) \times \pi(b_j) \right\} \times \pi(\sigma_w^2), \end{aligned} \quad (5.21)$$

where

$$c_{j,i,l} \equiv a_j + b_j \lambda_j(y_{j,i,l}) = a_j + b_j \left\{ \exp \left[\sum_{p=1}^P z_{j,i,p} \gamma_p \right] \sum_{m=1}^M k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \eta_{j,m} \right\}. \quad (5.22)$$

5.4.4 Variance-stabilised marks

The posterior estimation of mark parameters depends on a smooth estimate of the underlying spatial field. A noisy (i.e. “bumpy”) estimate of the spatial intensity makes it harder for the model to correctly estimate the mark parameters. An increased variability in the spatial field drives the mark parameters closer to the grand mean. Hence, the estimates of the mark parameters will show a bias. The estimate of the baseline parameter a will become inflated and the coefficient b linking the marks to the spatial intensity will be underestimated. In short, in almost all cases we will observe that $\text{var}(w|\lambda) \leq \text{var}(w|\hat{\lambda})$.

A way to reduce this bias would be to use a square-root- or log-transform to link the spatial intensity with the mark process; for example, by assuming a mark distribution of the form

$$w_{j,i,l} \sim \mathcal{N}(a_j + b_j \log \lambda_j(y_{j,i,l})). \quad (5.23)$$

or

$$w_{j,i,l} \sim \log\text{-}\mathcal{N}(a_j + b_j \log \lambda_j(y_{j,i,l})). \quad (5.24)$$

Note that such a model formulation would introduce a sign change in the interpretation of the mark parameter b as the log-intensity function is always negative.

5.5 Posterior approximation and sampling algorithm

In this section, we briefly discuss the MCMC scheme to sample the model parameters of the intensity-marked HPGRF model with additional covariates. Algorithmic details of all update routines, including the non-marked HPGRF variants, are relegated to Appendix A.

The MCMC algorithm that we use is essentially a Gibbs sampler. Most updates are not conjugate which means that additional MCMC techniques such as Metropolis-Hastings-within-Gibbs schemes or adaptive-rejection sampling are necessary. A simple proposal distribution for the Metropolis-Hastings algorithm that satisfies the detailed balance condition follows a multi-dimensional random walk. By adjusting the width of the proposal distribution during burn-in, one can modify the step size, i.e. the distance between subsequent proposals, in order to more effectively traverse the space of the posterior distribution and improve mixing. Random walk Metropolis-Hastings updates carry a number of disadvantages with them, most notably a high correlation between subsequent samples, potentially slow mix-

ing and slow convergence to the target distribution. Still, for many high dimensional Bayesian settings, this is the most viable option. We use Metropolis-Hastings-within-Gibbs sampling paired with symmetric random walk proposals to draw from the full conditional distributions for regression coefficients γ_p , latent sources θ_m , shared jump heights ν_m and shared inverse scale parameter β of the Gamma random field, and kernel variance σ_j^2 . Both inverse scale parameters β and τ are assigned Gamma priors, making the update of the group-specific inverse scale parameter τ conjugate. Conjugate priors are also used for the mark parameters, i.e a Normal prior on a_j and b_j , and an inverse Gamma prior on σ_w^2 .

A natural (and simple) choice for base measure $\alpha(dx)$ of the Gamma random field is to use Lebesgue measure. We normalise by the Lebesgue measure of the subset \mathcal{B} of \mathbb{R}^n on which the point process is defined. In our case $\mathcal{B} \subset \mathbb{R}^3$ and \mathcal{B} is equivalent to the human brain. We write $\alpha(dx) = \Pi(dx)/|\mathcal{B}|$, where $\Pi(dx)$ denotes Lebesgue measure and the Euclidean volume is given by $|\mathcal{B}| = \int_{\mathcal{B}} \Pi(dx)$. For the inverse Lévy construction of the Gamma random field, this implies that $\tilde{\alpha}(\mathcal{B}) = 1$ and $\tilde{\alpha}(dx) = \alpha(dx)$. The update for auxiliary points $x_{j,i,l}$ of the data augmentation scheme is directly given by the conditional distribution which assigns individual points to a jump location θ_m . Conditional on the presence of one or more auxiliary points at a given jump location, the θ_m are updated using a symmetric random walk and a Metropolis-Hastings acceptance step. For jumps that do not share a location with any auxiliary points, the update consists of a uniform draw over the whole domain. Group-specific jump heights $\eta_{j,m}$ follow a Gamma distribution combined with a term in non-closed form, hence a proposal is drawn from the Gamma distribution and accepted according to the corresponding Metropolis-Hastings ratio.

All parameters are assigned vague or weakly informative priors. Specific settings are given for the simulation study in Section 5.6 and for the real data applications in Chapter 6. Conditional posterior distributions, updates and sampling algorithms for all model parameters are provided in Appendix A.

5.6 Simulation study

In this section, we create a simulated data set to demonstrate how the HPGRF model can incorporate additional information, both spatially dependent and independent, alongside traditional point data to arrive at more accurate estimates for the intensity.

5.6.1 Simulated spatial data

The synthetic data set is based on three-dimensional point patterns on a volume $\mathcal{V} = [0, 100] \times [0, 100] \times [0, 20]$. We simulate three groups (types) A, B and C of points that share some areas of increased intensity. The true intensity functions for the three types are mixtures of four three-dimensional Gaussian densities $\phi_3(x; \mu, \Sigma)$, which are characterised by means μ_n and covariance structures Σ_n , for $n=1, 2, 3, 4$. For each group $j=A, B, C$ we draw $N_j=30$ realisations according to

$$\mathbf{Y}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{PP} \{ \mathcal{V}, \lambda_j(x) dx \}, \quad (5.25)$$

where $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ indicate the sets of parameters characterising the Gaussian densities. The true intensities for the three groups are composed as follows:

$$\begin{aligned} \lambda_A(x) &= \epsilon + \alpha_2 \phi_3(x; \mu_2, \Sigma_2) + \alpha_3 \phi_3(x; \mu_3, \Sigma_3), \\ \lambda_B(x) &= \epsilon + \alpha_2 \phi_3(x; \mu_2, \Sigma_2) + \alpha_4 \phi_3(x; \mu_4, \Sigma_4), \\ \lambda_C(x) &= \epsilon + \alpha_1 \phi_3(x; \mu_1, \Sigma_1) + \alpha_2 \phi_3(x; \mu_2, \Sigma_2) + \alpha_3 \phi_3(x; \mu_3, \Sigma_3). \end{aligned} \quad (5.26)$$

A homogeneous, non-zero background intensity is denoted by $\epsilon=20/|\mathcal{V}|$. The re-

Table 5.1: Parameter specification for simulated data sets.

n	1	2	3	4
α_n	4	5	1	3
μ_n	(60, 75, 10)	(70, 30, 10)	(40, 50, 10)	(10, 20, 10)
Σ_n	$\begin{bmatrix} 30 & 15 & 10 \\ 15 & 15 & 5 \\ 10 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} 30 & -10 & -10 \\ -10 & 40 & 5 \\ -10 & 5 & 10 \end{bmatrix}$	$\begin{bmatrix} 20 & -5 & 5 \\ -5 & 10 & -5 \\ 5 & -5 & 5 \end{bmatrix}$	$\begin{bmatrix} 10 & 5 & 0 \\ 5 & 20 & 5 \\ 0 & 5 & 10 \end{bmatrix}$

maining intensity parameters are summarised in Table 5.1. The ground truth exhibits accumulation of points in four regions (see top row in Figure 5.3). All three simulated types share clustering in region IV. Groups A and C have common clustering in region II. Regions I and III show an increase in intensity only for group B and group C, respectively. The event locations of the simulated data set are shown in the second row of Figure 5.3.

5.6.2 Simulated covariates

We simulate a set of three covariates for each realisation $\mathbf{x}_{j,i}$ of the point process. These covariates consist of one completely random covariate (z_1), one highly informative covariate (z_2) that is directly proportional to the log-count of points for that realisation and one covariate (z_3) that is correlated with the group to which the point pattern belongs. Summary statistics of the simulated covariates are listed in Table 5.2. The covariates are standardised to mean zero and standard deviation one before running the model. Note that the spatial process provides the intercept in the regression framework, thus there is no intercept needed in the design matrix.

Table 5.2: Summary statistics of simulated points and covariates.

	<i>A</i>	<i>B</i>	<i>C</i>	Total
Total # of points	621	824	935	2380
Mean # of points (SD)	20.70 (4.44)	27.47 (5.78)	31.17 (5.04)	26.44 (6.68)
Median # of points	21	28.5	32	26
IQR # of points	[18, 23]	[24, 30]	[27, 34]	[22, 32]
Mean z_1 (SD)	1.59 (0.96)	1.69 (0.86)	1.60 (0.77)	1.63 (0.86)
Median z_1	1.62	1.64	1.54	1.59
IQR z_1	[0.85, 2.59]	[0.85, 2.39]	[0.95, 2.15]	[0.85, 2.40]
Mean z_2 (SD)	0.16 (0.09)	5.12 (0.09)	10.15 (0.09)	5.14 (4.10)
Median z_2	0.17	5.10	10.15	5.10
IQR z_2	[0.06, 0.24]	[5.04, 5.21]	[10.07, 10.23]	[0.24, 10.07]
Mean z_3 (SD)	0.75 (0.06)	0.82 (0.06)	0.86 (0.04)	0.81 (0.07)
Median z_3	0.76	0.84	0.87	0.81
IQR z_3	[0.72, 0.78]	[0.79, 0.85]	[0.82, 0.88]	[0.77, 0.87]

5.6.3 Simulated mark process

We use intensity-dependent Gaussian marks $w_{j,i,l} \equiv f(\lambda(x_{j,i,l}))$. The individual mark values are sampled from a Normal distribution whose mean is characterised by a spatially constant baseline parameter a_j and an intensity-dependent term with interaction strength b_j such that

$$w_{j,i,l} \sim \mathcal{N}(a_j + b_j \lambda_j(x_{j,i,l}), \sigma_w). \quad (5.27)$$

The variance parameter σ_w is shared across the three types. Chosen parameter values for the simulation of the marks are listed in Table 5.6 and summary statistics

5.6. Simulation study

are provided in Table 5.3. For illustration, an intensity map of the expected mark value for a slice at $z=10$ is provided in Figure 5.4.

Table 5.3: Mean and standard deviation of simulated mark values per group. Regions I–IV are defined in Figure 5.3.

	Total	I	II	III	IV	Rest
<i>A</i>	6.01 (2.80)	3.92 (0.62)	6.42 (2.64)	4.14 (0.95)	8.19 (2.89)	4.31 (1.04)
<i>B</i>	7.25 (2.73)	9.00 (2.87)	4.70 (1.23)	4.744 (0.78)	8.55 (2.42)	5.20 (1.01)
<i>C</i>	2.88 (1.01)	4.31 (1.62)	9.96 (5.36)	5.32 (1.66)	3.06 (1.01)	5.64 (4.10)

5.6.4 Simulation setup

Details of the MCMC updates are provided in Appendix A. We set the number of jump locations for the approximation of the Gamma random field to $M=5,000$ and keep the size of the smoothing kernel fixed, setting $\sigma_j^2=4.0$ for all three types. The MCMC algorithm is run for a total number of 15,000 iterations with a burn-in of 5,000 iterations. After burn-in the chain is thinned to retain every 10th draw from the posterior, resulting in 1,000 posterior samples.

The prior specifications are as follows: For inverse scale parameters of the Gamma random fields we use vague priors $\beta \sim \mathcal{G}(0.001, 0.001)$ and $\tau \sim \mathcal{G}(0.001, 0.001)$. In runs where the kernel variance is being estimated we assign an uninformative uniform prior, $\sigma_j^{-2} \sim \mathcal{U}[0.01, 50]$, otherwise the parameter is held fixed. The mark parameters are given broad Normal priors, $a_j \sim \mathcal{N}(0.1, 100)$, $b_j \sim \mathcal{N}(1.0, 2000)$. Due to conjugacy the mark variance is assigned an inverse-Gamma prior, chosen as $\sigma_w^2 \sim \mathcal{IG}(5.0, 10.0)$. The variance parameters for any random walk Metropolis-Hastings updates are initialised to $\zeta_*^2=0.01$ and adjusted every 25 iterations during burn-in, depending on the acceptance rate over the previous 100 iterations and with optimal target acceptance rate of 44% for one-dimensional parameters [Gelfand et al., 2010].

To run the imHPGRF model with this configuration on an Intel Core i7-3770 CPU with 3.40GHz takes over 100 hours. By employing OpenMP to parallelise parts of the code and utilise up to six cores simultaneously, we could achieve a 3–4 fold reduction, resulting in a total computation time of approximately 30 hours.

5.6.5 Simulation results and model assessment

We carry out three separate runs to ensure convergence and stability of results. The different starting positions are drawn from the prior configurations and include different seed numbers. We find that the same realisations are misclassified in each run and that the 95% credible intervals of the posterior parameter estimates overlap. The model parameters show good mixing behaviour (Figure 5.1) and the autocorrelation function declines rapidly for most parameters (Figure 5.2). However, the interaction parameters b of the mark process as well as shared inverse scale parameter β of the population-level Gamma random field exhibit some long-range autocorrelations.

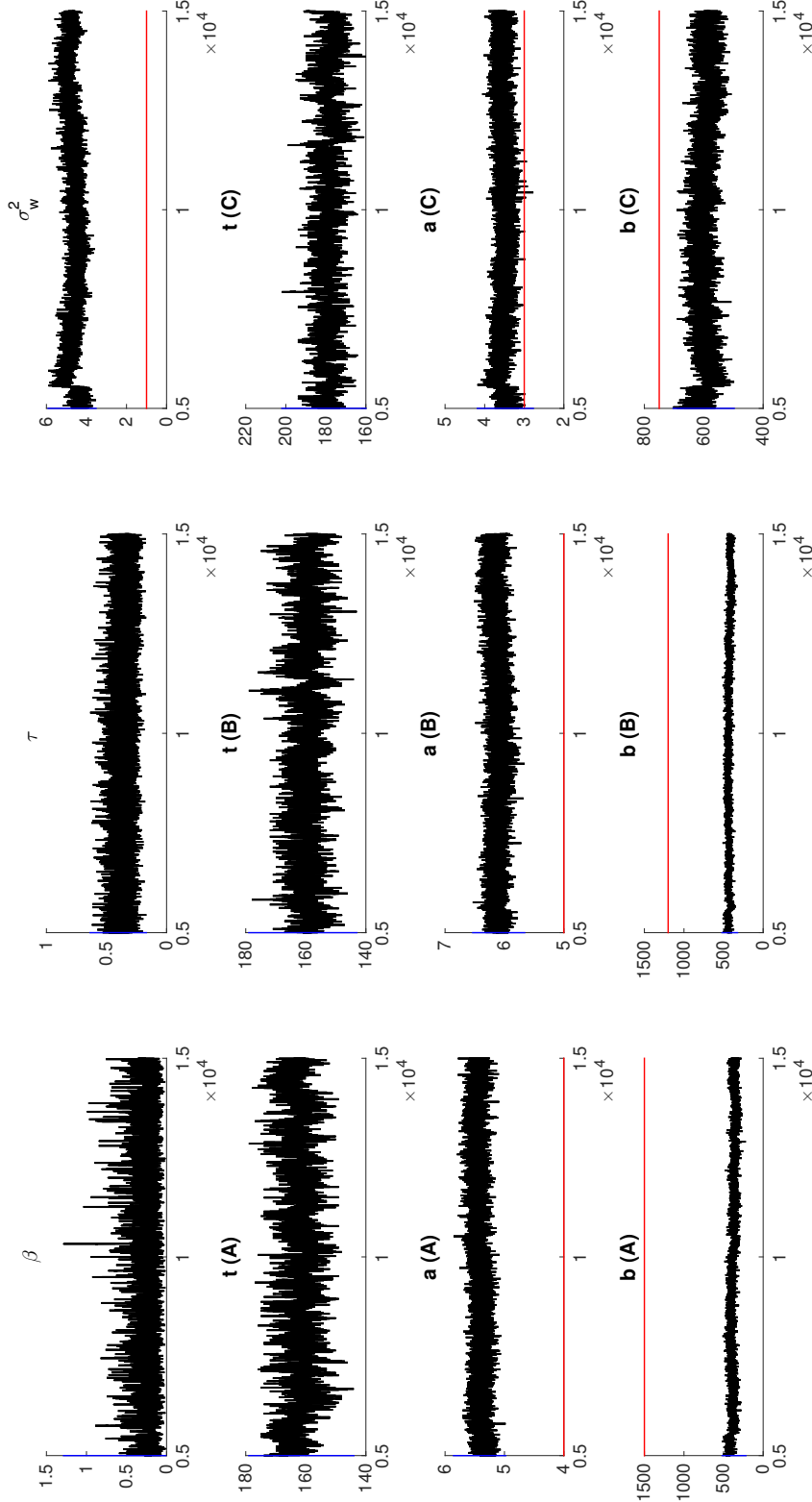


Figure 5.1: Posterior traceplots of scalar model parameters when applying the imHPGRF model to simulated data. Traceplots are based on the non-thinned chain post burn-in. Red lines indicate true parameter values. β and τ are the shared and group-specific inverse scale parameters of the Gamma random fields, respectively. a , b and σ_w^2 are parameters of the mark process and t indicates the number of unique point locations around which auxiliary points from the data augmentation scheme cluster.

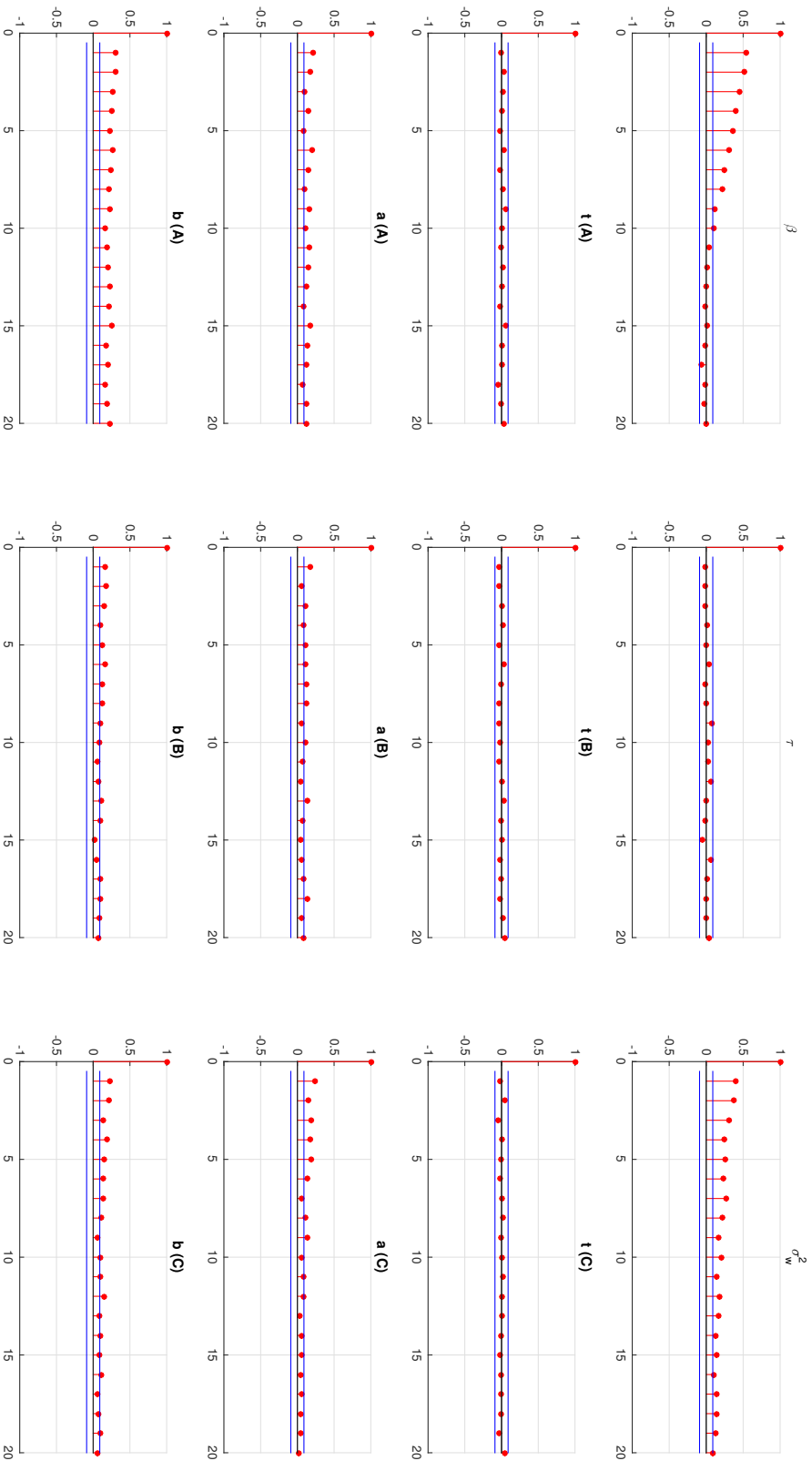


Figure 5.2: Autocorrelation of scalar model parameters when applying the imHPGRF model to simulated data. Autocorrelation functions are based on the thinned chain post burn-in. β and τ are the shared and group-specific inverse scale parameters of the Gamma random fields, respectively. a , b and σ_w^2 are parameters of the mark process and t indicates the number of unique point locations around which auxiliary points from the data augmentation scheme cluster.

In Figure 5.3 two-dimensional slices of the estimated posterior mean intensity for the plain HPGRF model and the intensity-marked HPGRF are shown, respectively. A two-dimensional map of estimated mark intensity is shown in Figure 5.4.

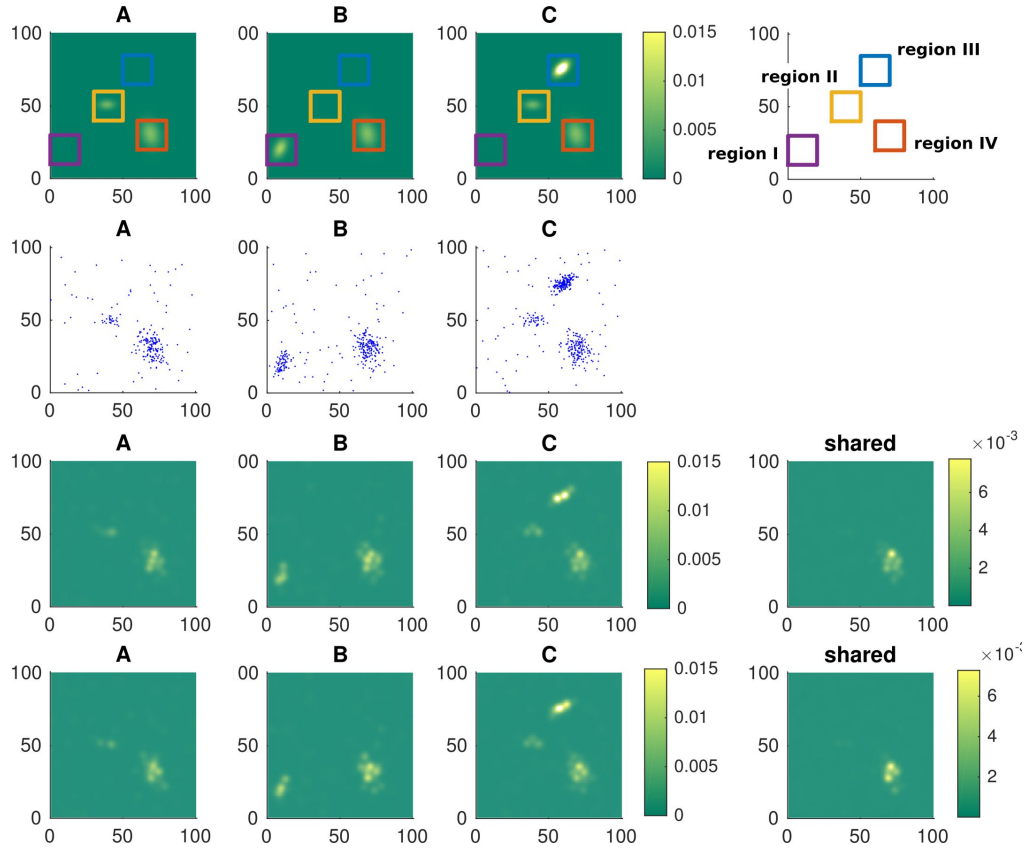


Figure 5.3: *Top row:* True spatial intensity maps for three simulated types A , B and C of point pattern data. Regions I–IV indicate areas of high intensity. *Second row:* Locations of simulated points for the three groups. The patterns show the full data set (30 realisations for each of the three point processes), projected onto the xy -plane. The total number of points is 2,380 (A : 621, B : 824, C : 935). *Third row:* Estimated posterior mean spatial intensity for the non-marked HPGRF model. *Bottom row:* Estimated posterior mean spatial intensity for the marked imHPGRF model. Slice shown at $z=10$.

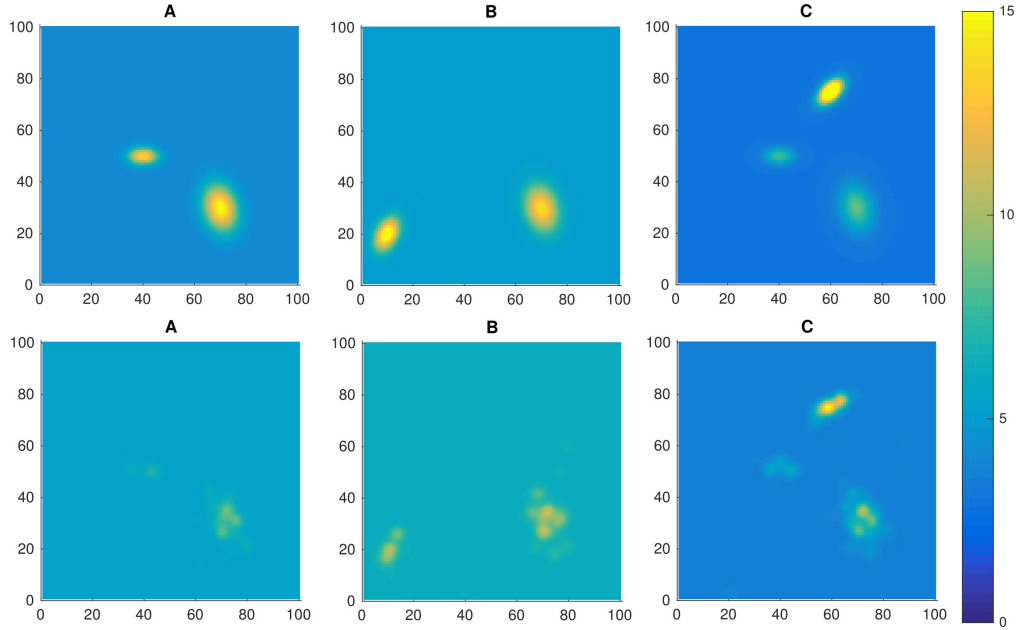


Figure 5.4: *Top row:* True expected intensity of the simulated mark process. *Bottom row:* Estimated posterior mean intensity as a result of the imHPGRF model fit. Slice shown at $z=10$.

The integrated mean posterior intensities provide an estimate for the expected number of points per type (see Table 5.4). The total number of points constitutes a first order property of the process and comparing the posterior estimate with the empirical count as provided by the data is a standard model check. In the case of the imHPGRF model variant with covariates, 95% credible intervals of mean posterior count estimates for each of the three groups A, B and C cover 53% (16/30), 37% (11/30) and 43% (13/30) of the simulated data, respectively.

We consider a second order posterior predictive test [Diggle, 2014; Baddeley et al., 2005] that involves computation of the L -function [Møller and Waagepetersen, 2004, 2007; Illian et al., 2009]. Given an intensity λ_j and a realisation \mathbf{y}_j , the L -function for distance r is computed as

$$L(r|\mathbf{y}_j, \lambda_j(y)) = \left[\frac{3}{4\pi|\mathcal{B}|} \sum_{y_1^*, y_2^* \in \mathbf{y}_j} \frac{\mathbf{1}_{[||y_1, y_2|| \leq r]}}{\lambda_j(y_1^*) \cdot \lambda_j(y_2^*)} \right]^{\frac{1}{3}}, \quad (5.28)$$

where y_1^* and y_2^* are individual points belonging to the same posterior realisation $\mathbf{y}_j^{(t)}$ or initial data point $\mathbf{y}_{j,i}$. Illian et al. [2009] suggest to look at the differences

5.6. Simulation study

Table 5.4: Data and mean posterior estimate of the average number of points in each group. Posterior estimates are based on 1,000 samples of the integrated spatial intensity per type.

HPGRF			
	Empirical mean	Posterior mean	95% CI
<i>A</i>	20.70	20.90	(19.17, 22.68)
<i>B</i>	27.47	27.64	(25.23, 30.03)
<i>C</i>	31.17	31.23	(29.21, 33.94)
Total	26.44	25.28	(21.47, 30.20)

imHPGRF			
	Empirical mean	Posterior mean	95% CI
<i>A</i>	20.70	20.95	(19.41, 22.62)
<i>B</i>	27.47	27.62	(25.71, 29.44)
<i>C</i>	31.17	31.25	(29.42, 33.22)
Total	26.44	24.91	(21.31, 29.65)

imHPGRF with covariates			
	Empirical mean	Posterior mean	95% CI
<i>A</i>	20.70	20.96	(19.25, 23.31)
<i>B</i>	27.47	27.59	(25.54, 29.78)
<i>C</i>	31.17	31.25	(29.47, 33.41)
Total	26.44	25.29	(21.98, 29.72)

$\Delta_{j,i}(r) \equiv L(r|\mathbf{y}_{j,i}, \lambda_j^{(t)}(y)) - L(r|\mathbf{y}_j^{(t)}, \lambda_j^{(t)}(y))$, where $\lambda_j^{(t)}$ denotes the t^{th} posterior estimate of the intensity. Hence, $\Delta_{j,i}(r)$ compares the L -functions of the input data with the L -function of a posterior realisation, given the posterior estimate of the intensity function at that sample draw. By calculating this difference between input data and all posterior predictive samples $t=1, \dots, T$ one obtains 95% credible intervals for $\Delta_{j,i}(r)$ at various distances. Ideally, the credible interval for each subject (or simulated realisation) i contains zero, which would indicate that the model performs well in explaining the data. We use a range of $[0, 100]$ in steps of $\Delta r=1$. For $r \leq 20$, more than 85% of the simulated data have a credible interval that contains zero. For distances greater than 20, 70% of all data points satisfy this condition. Overall, the posterior predictive checks indicate that the model is able to reproduce the general structure of the input data; see Figure 5.5.

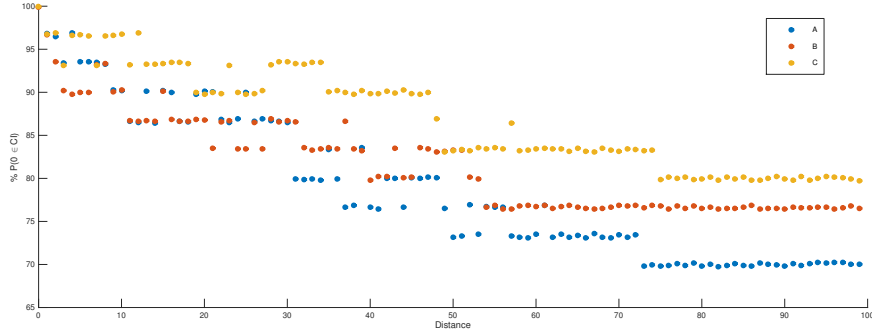


Figure 5.5: Results of the posterior predictive check for second order properties of the point process. Shown are the proportion of realisations which have credible intervals $\Delta_{j,i}(r)$ that contain zero. The posterior predictive samples were generated from 1,000 posterior draws of the intensity. Note: A small jitter was added to make overlapping points distinguishable.

We further assess the performance of the model variants by computing the integrated mean squared error (IMSE) and integrated weighted mean squared error (IWMSE) given by the relations

$$\begin{aligned} IMSE &= \frac{1}{J} \sum_{j=1}^J \int_Z \left[\hat{\lambda}_j(y) - \lambda_j(y) \right]^2 dy, \\ IWMSE &= \frac{1}{J} \sum_{j=1}^J \int_Z \lambda_j(y) \left[\hat{\lambda}_j(y) - \lambda_j(y) \right]^2 dy. \end{aligned} \quad (5.29)$$

In (5.29) above, $\hat{\lambda}_j(y)$ denotes the estimated intensity function for type j while $\lambda_j(y)$ represents the true intensity. Note that the IWMSE increases the relative contribution of regions with larger true intensity. The results are summarised in Table 5.5. The imHPGRF model with and without additional covariates exhibits an lower overall IMSE and IWMSE than the un-marked HPGRF variants. The addition of covariates has mixed effects as it slightly increases the discrepancy between fitted and true intensity for group C in case of the HPGRF and for groups A and B in case of the imHPGRF.

Posterior estimates of the mark parameters are summarised in Table 5.6. The mark residuals are given by the deviation of estimated values \hat{w}_i from the truth as given by the underlying process that generates the simulated data such that $r_{j,i,l} = \mathbb{E}[w_{j,i,l}] - \hat{w}_{j,i,l}$, for each mark value $w_{j,i,l}$. The index l ranges from one to L_i , with L_i being the number of points per realisation i in group j . The expectation $\mathbb{E}[w_{j,i,l}]$ is given by the mean of the true mark distribution in (5.27). The estimated

values $\hat{w}_{j,i,l}$ are mean posterior estimates of the mark process.

Standardised root-mean-squared errors (RMSE) are computed as $(\sum(w_{j,i,l} - \hat{w}_{j,i,l})^2)^{1/2} / (\sum(w_{j,i,l} - \bar{w}_{j,i,l})^2)^{1/2}$, where the sum runs over i and l for group-specific estimates of the RMSE, and over all three indices for the total RMSE score. Values smaller than one indicate that the addition of marks can explain some proportion of the variance in the data (see Table 5.5). The marks appear to improve the model fit for group C , which contains the highest intensity levels and therefore the largest effect on the intensity-dependent mark process. Residuals of the estimated mark process are shown in Figure 5.6. Histograms of mark residuals are shown in Figure 5.7. The two rows in the lower half of the figure show residuals with respect to mark values that are based either on posterior estimates of the spatial intensity (third row) or posterior estimates of the mark parameters (last row). This can help to determine as to what extent a deviation in the estimated intensity from the true intensity influences the mark process.

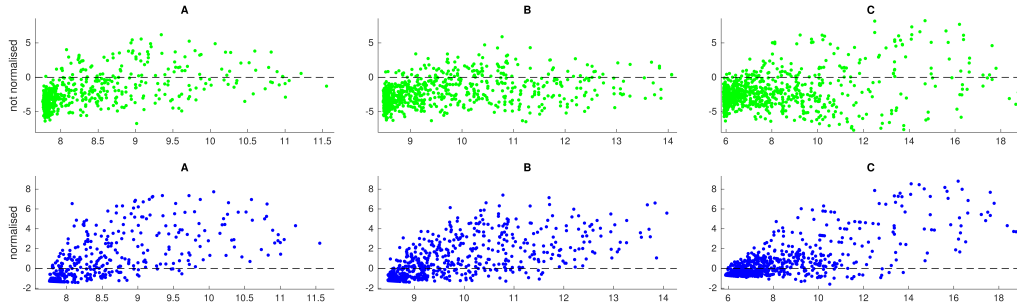


Figure 5.6: Mark residuals versus predicted mark values. The top row shows the plain residuals $r_{j,i,l}$. The bottom row shows standardised residuals based on the estimated MCMC standard error.

A summary of posterior results for regression coefficients is listed in Table 5.7. The coefficient γ_1 of the random covariate is estimated close to zero. As expected, the second covariate which was modelled as being directly dependent on the number of points per realisation is found to be the most informative predictor. The third simulated covariate is proportional to the group label and shows a weak positive correlation with overall count, which is consistent with the average number of points within the three different types of point pattern. Finally, confusion matrices for all model variants are summarised in Table 5.8. The plain HPGRF already achieves a

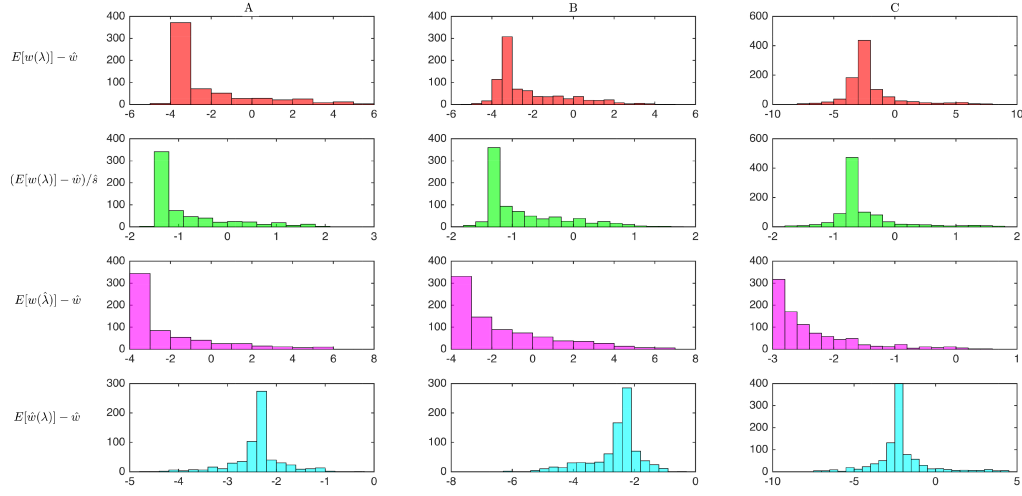


Figure 5.7: Histograms of mark residuals when applying the imHPGRF model to the simulated data set. *First row:* Residuals with respect to the true spatial intensity λ . *Second row:* Standardised residuals based on the estimated standard error \hat{s} . *Third row:* Residuals with respect to the mean posterior estimate of the spatial intensity $\hat{\lambda}$. *Bottom row:* Residuals with respect to marks $\hat{w}(y)$ that are computed using the true spatial intensity but posterior mean estimates for the mark parameters.

high classification accuracy of 93.3%. Inclusion of covariates and marks increases the prediction accuracy to 95.6% and 98.9%, respectively.

Table 5.5: Summary of integrated mean squared errors ($\times 10^{-4}$), integrated weighted mean squared errors ($\times 10^{-6}$) and standardised mark residuals based on root-mean-squared errors, for different model variants.

	HPGRF	HPGRF + cov.	imHPGRF	imHPGRF + cov.
$IMSE - \text{total}$	713	717	673	686
$IMSE - A$	142	137	136	138
$IMSE - B$	222	218	203	216
$IMSE - C$	348	362	335	332
$IWMSE - \text{total}$	314	309	280	293
$IWMSE - A$	37	43	35	40
$IWMSE - B$	63	65	63	64
$IWMSE - C$	214	201	192	189
Marks: std. $\overline{RMSE} - \text{total}$	/	/	0.971	1.039
Marks: std. $RMSE - A$	/	/	1.140	1.279
Marks: std. $RMSE - B$	/	/	1.051	1.035
Marks: std. $RMSE - C$	/	/	0.749	0.863

Table 5.6: Summary of ground truth and estimated posterior mean mark parameters, including 95% credible intervals.

Truth				
	A	B	C	Shared
a_j	4.0	5.0	3.0	/
b_j	1500	1200	750	/
σ_w	/	/	/	1.0

imHPGRF				
	A	B	C	Shared
a_j	5.42 [5.18, 5.61]	6.13 [5.89, 6.35]	3.54 [3.20, 3.84]	/
b_j	383 [308, 449]	430 [372, 478]	597 [546, 654]	/
σ_w	/	/	/	4.65 [4.05, 5.33]

imHPGRF with covariates				
	A	B	C	Shared
a_j	5.55 [5.37, 5.84]	6.35 [6.15, 6.57]	3.80 [3.46, 3.99]	/
b_j	317 [255, 390]	386 [354, 417]	625 [582, 675]	/
σ_w	/	/	/	4.42 [3.96, 4.81]

Table 5.7: Summary of ground truth and mean posterior estimates of regression coefficients, including 95% credible intervals.

	Truth	HPGRF + cov.	imHPGRF + cov.
γ_1	0.0	-0.016 [-0.050, 0.012]	-0.017 [-0.035, 0.008]
γ_2	2.0	1.283 [0.985, 1.633]	2.748 [1.879, 3.676]
γ_3	0.2	0.182 [0.132, 0.226]	0.197 [0.157, 0.247]

Table 5.8: Summary of confusion matrices and prediction accuracies for the simulation study. Rows correspond to truth.

HPGRF: Prediction accuracy: 0.933 .			
	<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	0.97 (29)	0.03 (1)	0.0 (0)
<i>B</i>	0.17 (5)	0.83 (25)	0.0 (0)
<i>C</i>	0.0 (0)	0.0 (0)	1.0 (30)

HPGRF with covariates: Prediction accuracy: **0.956**.

	<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	0.93 (28)	0.07 (2)	0.0 (0)
<i>B</i>	0.07 (2)	0.93 (28)	0.0 (0)
<i>C</i>	0.0 (0)	0.0 (0)	1.0 (30)

imHPGRF (with and without covariates): Prediction accuracy: **0.989**.

	<i>A</i>	<i>B</i>	<i>C</i>
<i>A</i>	1.0 (30)	0.0 (0)	0.0 (0)
<i>B</i>	0.03 (1)	0.97 (29)	0.0 (0)
<i>C</i>	0.0 (0)	0.0 (0)	1.0 (30)

5.7 Discussion

In this chapter we focused on a particular class of Cox process models, namely the Poisson/Gamma random field model, for multi-type spatial point patterns. We presented an intensity-marked hierarchical point process model that is also able to incorporate covariates unique to a given point pattern. Our imHPGRF model is based on and extends the approaches of Wolpert and Ickstadt [1998b] and Kang et al. [2014a]. The HPGRF model is able to share information between different point patterns and can model more complicated multivariate spatial structures than the PGRF of Wolpert and Ickstadt [1998b]. While the HPGRF model of Kang et al. [2014a] accounts for positive correlations between different types of spatial point patterns, it does not include any local or global information beyond the xyz -coordinates of individual events. The imHPGRF model on the other hand combines the underlying spatial stochastic process with a potentially intensity-dependent mark process that contains additional information about individual point locations.

We simulated three-dimensional spatial point patterns from three modified Thomas processes [Van Lieshout and Baddeley, 2002]. The Thomas process belongs to the family of Cox cluster processes and has the advantage that it is easy to sample from it. Our real datasets based on MS lesion locations show accumulation of points in brain areas with high white matter density. Choosing a process that exhibits clustering for the simulation study therefore seems justified. Note that a direct “simulation under the model” is not possible in our case, due to the non-parametric nature and large number of latent variables of the model.

The simulation study carried out in this chapter demonstrated that modelling the intensity function of a Cox process via the convolution of a Gamma random field with a Gaussian kernel can successfully reproduce the true intensity in a sparse, three-dimensional setting. Furthermore, adding covariates and marks can improve posterior intensity estimates and aid in classification and prediction tasks. However, in real data applications, the choice of how to model the mark structure and its dependence on the spatial intensity function needs to take into account the particular distribution of marks in the observed data. The Gaussian and log-Gaussian mark dependencies that we have explored here, are likely to be ill suited in cases where mark values show a lot of variation. Additionally, the intensity-dependent mark model necessarily requires a good fit of the spatial intensity in order to be able to estimate the mark parameters.

The implementation of the model is not straightforward and assessment of convergence and fine-tuning of the MCMC algorithm needs to be done separately

for every application. Due to the large number of latent variables, convergence diagnostics for the imHPGRF model have to be selective. Alongside model parameters, convergence of the intensity function can only be monitored at a finite number of points. Even that is not a guarantee that the MCMC algorithm has fully converged. At the same time, model-based methods like the imHPGRF allow for the quantification of uncertainty through standard errors obtained from the posterior distribution of any parameter of interest.

5.7.1 Limitations

An inherent limitation of any variant of the Poisson/Gamma random field model lies in the truncation error when approximating the Gamma random field with a finite number of jumps. Theoretically, this truncation error can be made arbitrarily small by increasing the number of latent sources M . In practice, M usually needs to be significantly larger than the total number of observed points in the data to produce a good model fit. For constant base measure $\alpha(dx)$ and constant inverse scale parameter β , the inverse Lévy measure algorithm generates jump heights θ_m in strictly decreasing order [Wolpert and Ickstadt, 1998b]. In order to estimate the truncation error, one can sample a random number M_ϵ that represents the number of jumps necessary until a previously chosen $\epsilon > 0$ would exceed the next jump height. Then the expected truncation error obeys the bound $\mathbb{E}[\Gamma(\mathcal{B}) - \Gamma_{M_\epsilon}(\mathcal{B})] \leq \epsilon \alpha(\mathcal{B})$ [Wolpert and Ickstadt, 1998b].

The parameter estimates given in Table 5.6 show that, regardless of including covariates or not, the imHPGRF model struggled to predict the true mark parameter values. The intensity-dependent mark process falls short when trying to model individual mark values that are far from the mean. These cases are not captured well by the model as it tries to estimate a smooth spatial intensity function. Large mark values would have to be the result of either a high baseline parameter, thereby increasing all mark values or a very large interaction coefficient on the side of the mark process; or, with respect to the Poisson point process, large local values in the intensity, similar to spikes. Another reason for why the parameter estimates are not very good may be attributed to the relatively small number of 30 point pattern realisations per group that constitutes the simulated dataset. This choice was based on the intent to mimic real data sets where usually only small sample sizes are available for different disease subtypes. We would expect that a larger number of training data would increase the accuracy of the parameter estimates.

The suboptimal results for the mark parameters have implications for the

application of the model to real data sets. The simulation study indicates that the model is better suited for applications with marks that are close to the overall mean, especially when sample sizes are small. In cases with large discrepancies between individual mark values or a large number of outliers, the estimation of parameters of the mark process is likely to be very challenging. With respect to our applications to lesion data (see the following chapter), we would expect that the use of cube-root transformed lesion volumes as marks lies within the regime for which the model is adequate.

A further drawback arises from the computational cost of large data sets. As the number of point locations in the input data increases, the required amount of jumps as well as the number of auxiliary points in the data augmentation scheme grows. Both increases lead to a larger-than-linear inflation of computational burden for the MCMC sampling algorithm. Additionally, a large domain over which the point process is defined or, equivalently, a high spatial resolution of the estimated intensity come at an exponentially increasing computational cost, particularly in three-dimensional applications.

5.7.2 Further extensions

The role of covariates in our log-linear approach is to model a homogeneous effect that shifts the intensity function away from the group mean. The magnitude of the effect depends on subject-specific attributes which function as explanatory variables. A different way to interpret this is to look at the Poisson/Gamma random field as representing unobserved, spatially correlated covariates.

An interesting area for future work would be to introduce spatially varying covariates. One way of doing so would be to model covariates additively. Instead of subject-specific covariates with a log-link function, Ickstadt and Wolpert [1999] have noted that the identity link would permit to add an underlying random field to the intensity, which could be allowed to vary spatially. With respect to applications to MS lesion data, a candidate for a location-specific covariate (in the sense of Ickstadt and Wolpert [1999]) would be a white matter density map. Such an attribute is defined over the whole space and is independent of any observed point location. Lesions predominantly occur in white matter regions and therefore, considering white matter density as a baseline intensity could improve local estimation of the intensity.

Another approach could focus on covariates with a local effect, i.e. by making the regression coefficients dependent on spatial location, thus assuming $\gamma = \gamma(y), y \in \mathcal{B}$. While non-spatial regression coefficients γ may be assumed to follow a Normal

distribution, one could model the $\gamma(y)$ as realisations of Gaussian random fields.

There are several ways of extending the modelling of marks. In addition to the intensity dependence, the mark process of the imHPGRF model could also depend on another spatially varying function:

$$w_{j,i,l} = a_j + b_j \lambda_j(y_{j,i,l}) + c \kappa(y_{j,i,l}) + \epsilon_{j,i,l}, \quad (5.30)$$

where $\kappa(y)$ denotes a location-dependent mark function defined on \mathcal{B} . Even an interacting intensity-marked HPGRF model with an additional, spatially informed mark could be considered; for example, in the form of

$$w_{j,i,l} = a_j + b_j \lambda_j(y_{j,i,l}) + c_j \kappa(y_{j,i,l}) + d_j \lambda_j(y_{j,i,l}) \kappa(y_{j,i,l}) + \epsilon_{j,i,l}, \quad (5.31)$$

where an interaction term between the intensity function $\lambda(y)$ and the spatially dependent mark function $\kappa(y)$ is included in the mark process. Increasingly complex mark structures are more challenging to estimate and may require a large amount of data.

Alternatively, one could consider the inclusion of multivariate marks or spatially varying regression coefficients for global covariates. Another, potentially very challenging extension would involve the incorporation of a spatio-temporal framework into the HPGRF model. Spatio-temporal formulations of spatial point process models involving Gaussian processes have been proposed in the literature [Gelfand et al., 2010]. For the log-Gaussian Cox process, for example, the incorporation of a temporal dimension leads to a spatio-temporal stochastic intensity function of the form $\lambda(y, t) = \exp\{G(y, t)\}$ where G is a Gaussian process. Applications of such models can be found in geostatistical and epidemiological settings, for example, the monitoring of disease outbreaks [Diggle et al., 2013].

Chapter 6

Application of spatial point process models to MS lesion data

This chapter explores the intensity-marked hierarchical Poisson/Gamma random field model introduced in Chapter 5 as applied to two clinical data sets. The two data sets, GeneMSA (Section 6.1) and BENEFIT (Section 6.2), comprise structural MRI data on patients with different forms of MS. The GeneMSA data considers different MS subtypes, while the BENEFIT data looks at conversion from a clinically isolated syndrome to clinically definite MS. Details of the two data sets were given in Chapter 3.

Data preprocessing

For both data sets, we base our analysis on lesion locations, precisely, the spatial locations of the geometric centres-of-mass of individual lesions found in T_2 -weighted MR images. After co-registration and normalisation of the images to standard MNI space as described in Subsection 2.1.1, the same affine transformation was applied to the lesion masks. The aligned lesion masks were thresholded at 0.5 to create binary images of lesion occurrence for each patient. Individual lesions were defined on the basis of a 26-order neighbourhood as computed with the cluster command in the FSL software package.

6.1 Application: GeneMSA data

The goal of the analyses in this section is to perform inference on the differences between MS subtypes and to assess the performance of our spatial point process models when predicting the subtype of new or previously held-out data. Taking into account the correlation between groups, induced by the hierarchical structure of the model, we can identify regions of lesion occurrence that are common across different subtypes. The shared, population-level intensity function on the top of the hierarchy can provide insights as to the degree of commonality between groups. We make use of the various types of data available for each MS patient by including demographic attributes and clinical measures as covariates, as well as considering individual lesion volume as intensity-dependent marks attached to the location of their centre-of-mass.

Our main focus lies on the application of the intensity-marked hierarchical Poisson/Gamma random field (imHPGRF) model that we have introduced in Section 5.4. Hence, the results presented in the following sections are with respect to the imHPGRF. In order to evaluate model performance we also use other PGRF model variants.

6.1.1 Algorithmic details and posterior computation

We run the MCMC algorithm for a total of 16,000 iterations, discarding the first 6,000 as burn-in. Estimation of posterior quantities is based on 1,000 samples, using every 10th iteration after burn-in and discarding the rest to decrease autocorrelations between samples. We assume a Gaussian representation for the mark process. The prior specifications are as follows: For inverse scale parameters of the Gamma random fields we use vague priors $\beta \sim \mathcal{G}(0.001, 0.001)$ and $\tau \sim \mathcal{G}(0.001, 0.001)$. In runs where the kernel variance is being estimated we assign an uninformative uniform prior, $\sigma_j^{-2} \sim \mathcal{U}[0.01, 50]$, otherwise the parameter is held fixed. The mark parameters are given broad Normal priors, $a_j \sim \mathcal{N}(0.1, 100)$, $b_j \sim \mathcal{N}(1.0, 1000)$. Due to conjugacy the mark variance is assigned an inverse-Gamma prior; we choose $\sigma_w^2 \sim \mathcal{IG}(5.0, 10.0)$ based on an empirical analysis of expected cube-root transformed lesion volume. The variance parameters for any random walk Metropolis-Hastings updates are initialised to $\zeta_*^2=0.01$ and adjusted every 25 iterations during burn-in, depending on the acceptance rate over the previous 100 iterations and with optimal target acceptance rate of 44% for one-dimensional parameters [Gelfand et al., 2010].

For the Bayesian classifier in the prediction part of the model, we assign equal

prior probability to each MS subtype. Alternatively, one could use empirical priors and thereby use prior information about discrepancies in expected group sizes.

The truncation parameter M for the evaluation of all Gamma random fields is set to $M=15,000$. This is sufficiently larger than the total number of data points as recommended by Wolpert and Ickstadt [1998b]. In our experience, setting M to roughly twice the number of points in the data keeps both the computational cost and the truncation error reasonably small.

We performed three separate runs of the same imHPGRF model, initialised with different seed numbers and different values for the prior hyper-parameters, to ensure that the algorithm converges to the same posterior distribution. None of the parameters have shown sensitive behaviour with regard to prior specifications. The mean posterior estimates of the three separate runs overlap for all quantities of interest and fall within the range of each other's standard error. Finally, results and posterior diagnostics indicate convergence of the MCMC algorithm for all model variants.

6.1.2 Posterior results and prediction

Estimated mean posterior intensity maps per MS subtype are displayed in Figure 6.1 and Figure 6.2. The figures show an axial and a sagittal view of the brain, respectively. For comparison, the empirical intensity based on binary lesion masks is provided in the second row of each figure. The imHPGRF is able to capture accurately regions of high lesion incidence and distinguish between the four subtypes. The shared intensity estimate indicates regions of common clustering.

Some discrepancies between the empirical incidence rate and the posterior estimates of the intensity function can be observed. For instance, the SPMS and PPMS subtypes show higher intensities in the area around voxel [60,40] on the empirical plots (see middle row of Figure 6.1) compared to the estimated intensity functions. This difference could be a result of several patients in each of the two subtypes having large lesions in this particular region. The empirical intensity is based on the collation of the binary lesion masks of all subjects within a given subtype. Many lesions overlapping at a particular location results in a high incidence rate. On the other hand, the spatial intensity function of the imHPGRF model is based on the coordinates of the centres-of-mass of lesions and takes lesion size only into account via the mark process.

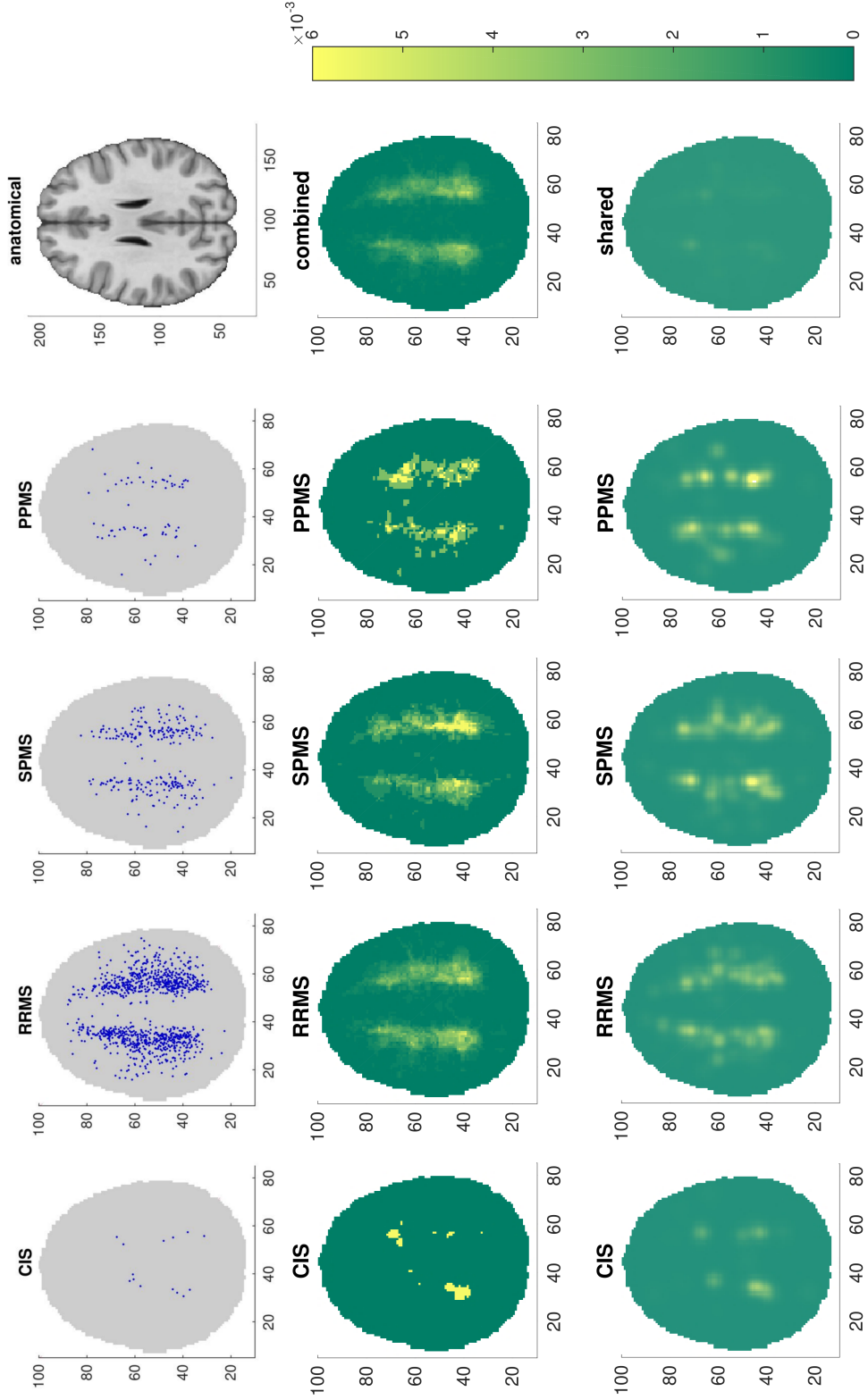


Figure 6.1: Results from fitting the imHPGRF model to the GeneMSA data, showing empirical and estimated intensity functions per MS subtype. *Middle row:* Lesion locations that fall within $\pm 4mm$ of the current slice. *Second row:* Empirical intensity maps based on binary lesion masks. *Bottom row:* Estimated posterior mean intensities at each voxel. The intensity estimates are based on 1,000 posterior samples. Shared intensities in the column on the right are based on the population level Gamma random field that is shared across all subtypes. All panels show an axial slice at $z=+2$.

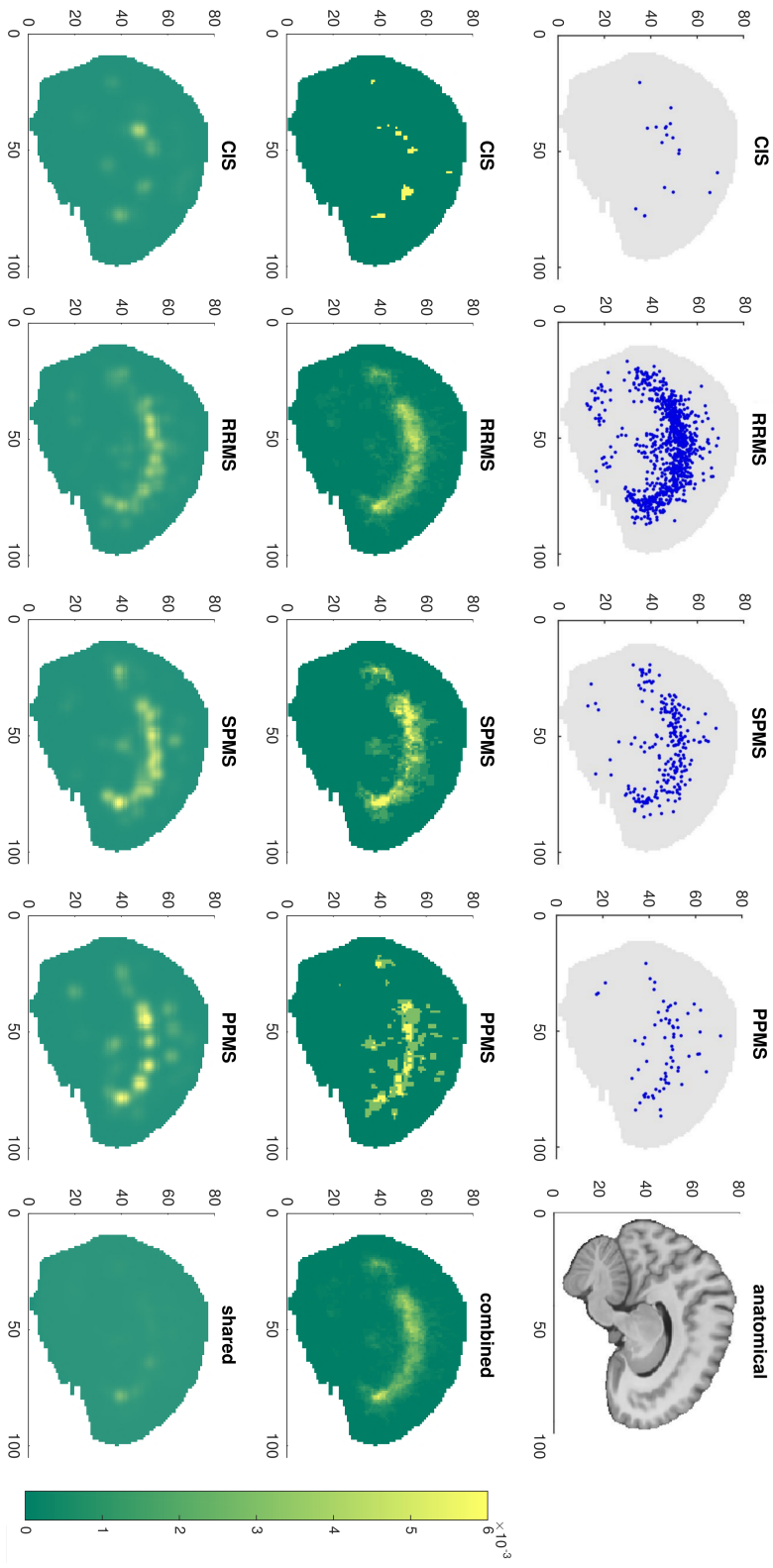


Figure 6.2: Empirical and estimated intensity functions per MS subtype as described in Figure 6.1. All panels show a sagittal slice at $x=+10$.

Median posterior estimates including 95% credible intervals for model parameters of interest are listed in Table 6.1 and Table 6.2. The baseline parameters a_j of the mark process reflect the mean of the mark distribution. The estimated parameter values are similar to the average mark value as given by the data. At the same time, estimates for the link parameters b_j are relatively small and the credible intervals for the CIS and PPMS types even include zero. This indicates a weak dependence of the marks on the underlying spatial intensity. The model favours estimation of the grand mean rather than local fluctuations in the mark process. Consequently, most of the variation in mark values is transferred into the variance parameter σ_w^2 .

Furthermore, the parameter estimation becomes more difficult for groups with few data points. The estimates of mark parameters for the CIS and PPMS subtypes exhibit larger posterior variances than their counterparts in the RRMS and SPMS types. This indicates that a more accurate estimate of mark parameters and thus a more pronounced distinction between different groups may be achievable given more data points.

In Figure 6.3, residuals are shown for the estimated marks. The panels show an overestimation of small mark values for all four groups. An clear underestimation of the volume of large lesions that differ substantially from the average also occurs. Overall, the model seems to favour the grand mean and is not able to capture some of the variation in the mark distribution.

Table 6.1: Posterior parameter estimates for the imHPGRF model including covariates: Type-specific parameters. The mark parameters a and b represent the mean and interaction strength with the spatial intensity of the mark distribution, respectively.

	CIS	RRMS	SPMS	PPMS
a_j (median)	2.93	2.82	2.85	3.07
a_j (95% CI)	[2.66, 3.22]	[2.77, 2.87]	[2.75, 2.94]	[2.90, 3.24]
b_j (median)	15.0	197.3	133.2	29.1
b_j (95% CI)	[-43.1, 71.8]	[160.8, 231.7]	[85.6, 174.2]	[-22.5, 79.4]

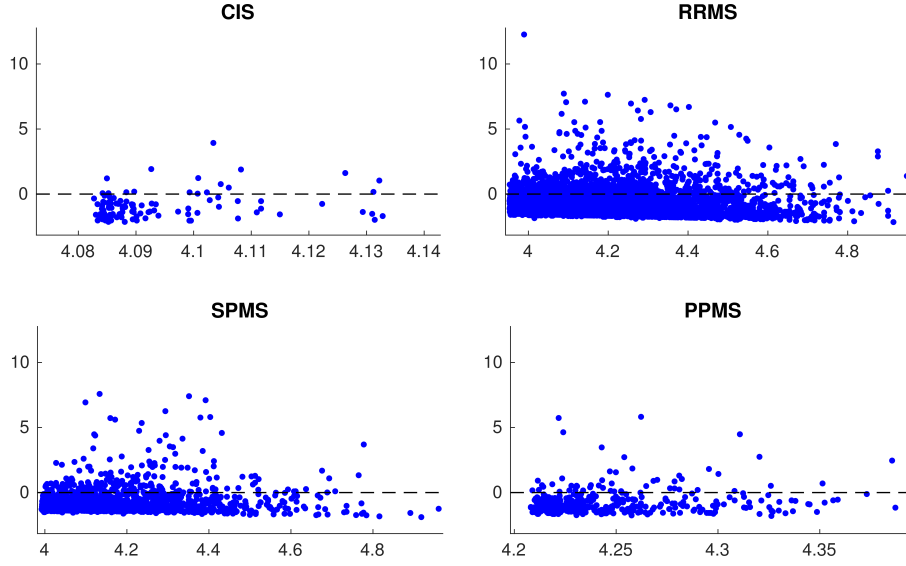


Figure 6.3: Mark residuals of posterior estimates of individual lesion volume with respect to empirical data versus predicted values.

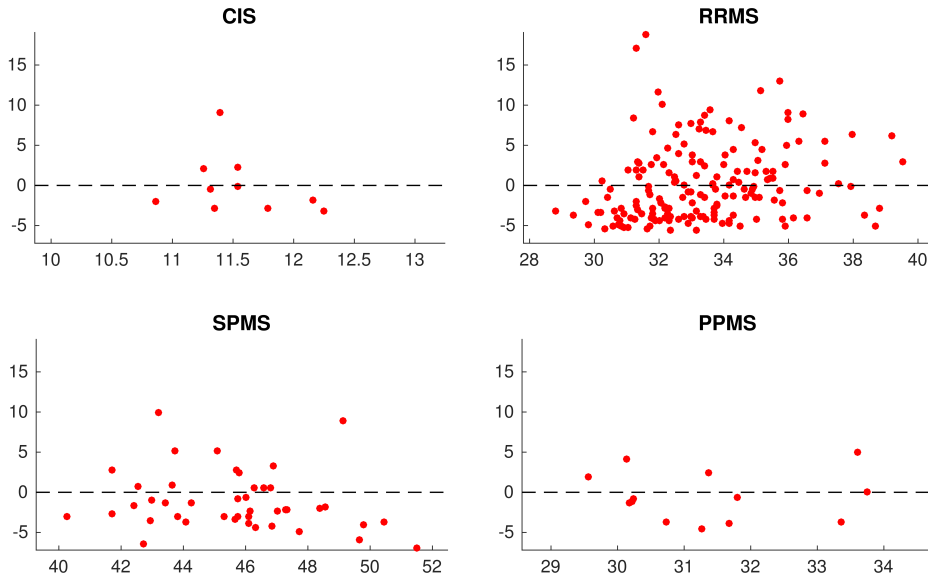


Figure 6.4: Covariate residuals from Poisson regression of posterior estimates of lesion count per subject with respect to empirical data versus predicted values.

Table 6.2: Posterior parameter estimates for the imHPGRF model including covariates: Shared parameters common to all types. Regression coefficients for subject-level covariates are denoted as γ_p , the corresponding covariate is indicated in brackets. The mark variance is denoted by σ_w^2 .

	Median	95% CI
σ_w^2	2.28	[2.21, 2.35]
γ_1 (sex)	0.001	[-0.021, 0.024]
γ_2 (age)	0.033	[0.001, 0.062]
γ_3 (disease duration)	0.028	[-0.004, 0.055]
γ_4 (PASAT)	0.009	[-0.017, 0.033]
γ_5 (EDSS)	0.036	[0.005, 0.071]

Posterior estimates of the regression coefficients γ_p for demographic and clinical covariates (Table 6.2) indicate that EDSS score, followed by age and disease duration, are most informative. Residuals based on the Poisson regression framework and with respect to the predicted number of lesions per subject are shown in Figure 6.4.

Confusion matrices and LOOCV prediction rates for the imHPGRF model with covariates as well as other model variants are summarised in Table 6.3. The independent PGRF model (IPGRF) estimates a separate intensity function for each group of point pattern data, which, in principle, amounts to running the PGRF model of Wolpert and Ickstadt [1998b] multiple times, once for each group. The IPGRF lacks the hierarchical structure of the HPGRF models and no information about clustering is shared across different classes.

The inclusion of lesion volume in the form of intensity-dependent marks does not have a large effect on classification performance. The plain model (HPGRF) achieves an average accuracy of 80.1%, while the marked model (imHPGRF) reaches 82.8%. Adding covariates again improves prediction performance. The imHPGRF model plus covariates achieves an average classification accuracy of 87.6% and a total accuracy of 85.7%.

Although one group (RRMS) is dominating in terms of subject numbers, the model does a good job in predicting the subtype label for the two smallest classes (CIS and PPMS). Clinically, RRMS and SPMS are closely related. Most patients who are initially diagnosed with relapsing-remitting MS eventually develop into the secondary-progressive course of the disease. This may be reflected in the confusion of SPMS predominantly with RRMS in cases of wrongly predicted subtype. The

reverse (RRMS being classified as SPMS) appears not to happen as often, as most misclassified subjects with RRMS are categorised as CIS. A closer examination of misclassified data shows that patients with few lesions (five or less) tend to be labelled as CIS. On the other hand, patients wrongly predicted as having the SPMS subtype have a disproportionally large number of lesions. This is consistent with clinical findings and expectations of subtype characteristics.

Finally, when comparing individual posterior predictive probabilities for different subjects, the model displays much greater certainty about the ones that classified correctly. In most of these cases the predicted probability for the correct class exceeds 90%, whereas for misclassified subjects the predictive probabilities are more evenly split among all groups.

Table 6.3: Confusion matrices and prediction accuracies (see Subsection 2.4.1 for a definition of overall and average accuracy) for different model variants. Note that chance level is at 25% average accuracy and, at best, 72.3% overall accuracy which would be obtained by classifying everything into the largest group, RRMS.

IPGRF: Overall & average acc.: 0.618 & 0.459 .				
	CIS	RRMS	SPMS	PPMS
CIS	0.20 (2)	0.10 (1)	0.60 (6)	0.10 (1)
RRMS	0.19 (33)	0.69 (118)	0.08 (13)	0.05 (8)
SPMS	0.09 (4)	0.40 (17)	0.49 (21)	0.02 (1)
PPMS	0.08 (1)	0.31 (4)	0.31 (2)	0.46 (6)

HPGRF: Overall & average acc.: 0.798 & 0.801 .				
	CIS	RRMS	SPMS	PPMS
CIS	0.90 (9)	0.10 (1)	0.0 (0)	0.0 (0)
RRMS	0.12 (21)	0.81 (140)	0.03 (5)	0.03 (6)
SPMS	0.05 (2)	0.23 (10)	0.72 (31)	0.0 (0)
PPMS	0.08 (1)	0.15 (2)	0.0 (0)	0.77 (10)

imHPGRF: Overall & average acc.: 0.811 & 0.828 .				
	CIS	RRMS	SPMS	PPMS
CIS	0.90 (9)	0.10 (1)	0.0 (0)	0.0 (0)
RRMS	0.13 (23)	0.82 (141)	0.02 (4)	0.02 (4)
SPMS	0.05 (2)	0.21 (9)	0.74 (32)	0.0 (0)
PPMS	0.0 (0)	0.15 (2)	0.0 (0)	0.85 (11)

imHPGRF with covariates: Overall & average acc.: 0.857 & **0.876**.

	CIS	RRMS	SPMS	PPMS
CIS	1.0 (10)	0.0 (0)	0.0 (0)	0.0 (0)
RRMS	0.10 (17)	0.87 (149)	0.01 (1)	0.03 (5)
SPMS	0.05 (2)	0.16 (7)	0.79 (34)	0.0 (0)
PPMS	0.0 (0)	0.15 (2)	0.0 (0)	0.85 (11)

6.1.3 Model assessment

Similar to what we did for the simulation study in the last chapter, we perform three separate runs of the model with different starting positions randomly drawn from priors within the parameter space and with different seed numbers. All three MCMC chains display the same convergence behaviour, giving similar overall results.

Convergence of the MCMC algorithm is assessed via visual inspection of traceplots of the model parameters. Traceplots, together with plots of the autocorrelation function, for one run are shown in Section B.2 of the appendix. The figures show very good mixing behaviour for all parameters when the width of the smoothing kernel is held fixed. The autocorrelation drops rapidly within ten iterations. Note that estimating the smoothing kernel variance σ_j^2 induces higher correlation between samples and requires a longer burn-in as the σ_j^2 exhibit significantly slower convergence behaviour. Type-specific parameters such as the mark parameters a_j and b_j mix more slowly and show longer auto-correlations for the subtypes with the largest number of data points, RRMS and SPMS.

As in Subsection 5.6.5, we again look at posterior predictive checks and compute first and second order quantities that provide an indication of the validity of the model. The integrated mean posterior intensities yield an estimate of the expected number of points in each group. The results for the imHPGRF model are summarised in Table 6.4. For all four subtypes, the posterior estimates are close to the empirical mean number of points in the data set. In fact, the 95% credible intervals cover all empirical means, indicating a good fit of the subtype-specific spatial intensities.

As a second order statistic, we examine the L -function (cf. the expression in (5.28)); in particular, the difference $\Delta_{j,i}(r)$ between the L -function of a subject's point pattern and the posterior sample of a point pattern belonging to the same subtype j . The distance parameter r ranges from 0 to 100 voxels in steps of 1 voxel, corresponding to steps of $2mm$ and a maximum of $200mm$. For each distance value r one can compute the proportion of subjects whose 95% credible interval for $\Delta_{j,i}(r)$ includes zero, using 1,000 posterior realisations. The credible intervals from over 80% of subjects cover zero for distances $r \leq 25$. For distances greater than that the proportion remains above 60%, indicating a good overall fit of the model.

Table 6.4: Empirical data and posterior estimates of the average number of lesions per subject for each MS subtype based on integrated spatial intensities.

	CIS	RRMS	SPMS	PPMS
Empirical mean	11.5	34.6	38.0	28.8
Posterior mean	12.6	33.9	41.6	30.6
95% CI	[10.3, 14.7]	[33.0, 34.9]	[39.2, 44.4]	[27.7, 33.9]

6.1.4 A GLM analysis

As a sanity check, we ignore the spatial locations and use a Generalised Linear Model (GLM) [McCullagh and Nelder, 1989] to assess subject-specific covariates independently from the fit of the spatial point process model. The response is simply the number of lesions per subject. The covariates include biological sex, age, disease duration, PASAT and EDSS scores. We fit a Poisson regression model using the canonical exponential link function. The set of regressors include nuisance parameters that indicate to which group any given subject belongs.

Results from the GLM analysis are listed in Table 6.5 and displayed in Figure 6.5. Overall, the number of lesions per subject shows a weak dependence on the predictors. Among all covariates, EDSS appears to be the most indicative, followed by disease duration and age. Biological sex and PASAT do not seem to be good predictors with p-values larger than 0.01 in both cases. These results are consistent with the coefficient estimates obtained from the imHPGRF model (see Table 6.2), where EDSS and age were found to be the most informative characteristics.

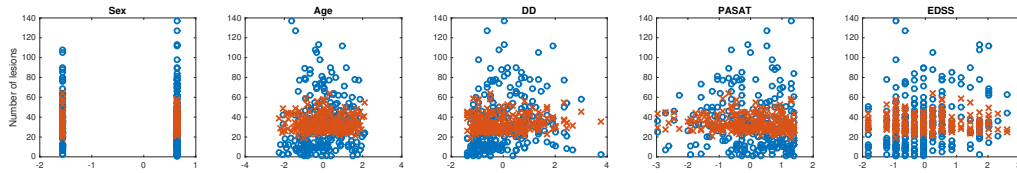


Figure 6.5: Results from fitting a GLM to the GeneMSA data set. The x-axes show the (standardised) explanatory variables sex (female is coded positive), age, disease duration (DD), Paced Auditory Serial Addition Task (PASAT), Expanded Disability Status Scale (EDSS), respectively. The dependent variable is the number of lesions per subject. Red crosses indicate the fitted values; blue circles indicate true values.

A caveat needs to be pointed out here. The GLM regression model is based on aggregate measures of the whole population/dataset whereas the Poisson Gamma random field models are specified on the level of individual subjects. A direct comparison of effect sizes and even the direction (i.e. signs of coefficients) of effects between the two models is not possible; a problem that is generally known as the ecological fallacy [Robinson, 1950]. One could therefore expect differences between effects for groups based on the GLM analysis and effects for individuals based on the point process models.

Table 6.5: Results from fitting a GLM to subject-specific covariates, including standard errors on the regression coefficients. The number of lesions per subject constitutes the dependent variable in a log-linear regression framework. PASAT and EDSS are clinical scores measuring disease severity. DD stands for disease duration, i.e. the time between the initial diagnosis and the MRI scan on which our analysis is based. A negative coefficient for the covariate sex indicates a positive correlation with being male.

	Intercept	Sex	Age
Coefficient (s.e.)	3.496 (0.011)	-0.026 (0.011)	-0.099 (0.014)
p-value	/	$2.25 \cdot 10^{-02}$	$2.86 \cdot 10^{-12}$
t-statistic	/	-2.28	-6.98

	DD	PASAT	EDSS
Coefficient (s.e.)	0.157 (0.013)	0.017 (0.015)	0.167 (0.012)
p-value	$5.08 \cdot 10^{-32}$	$1.42 \cdot 10^{-01}$	$8.61 \cdot 10^{-41}$
t-statistic	11.78	1.47	13.37

6.2 Application: BENEFIT data

The BENEFIT data set consists of 364 patients with an initial diagnosis of a clinically isolated syndrome (CIS), 175 (48%) of which developed clinically definite multiple sclerosis (CDMS) within the study period of 24 months. To date, no reliable method exists to predict who will, and will not, develop CDMS. A recent study by Wottschel et al. [2015] showed that a linear SVM classifier correctly predicted the conversion (or non-conversion) to CDMS in 71% of patients after 12 months, and in 68% after 36 months, on the basis of different combinations of lesional and clinical features.

The aim of this section is to apply our intensity-marked Poisson/Gamma random field model to the data and try to predict conversion to CDMS within 24 months of the first clinical signs of the disease. Similar to the prediction of MS subtype in the GeneMSA data set in Section 6.1, we use the coordinates of centres-of-mass of individual lesions to estimate spatial intensity functions for converters and non-converters. We compare the performance of the plain HPGRF model with the imHPGRF model which includes lesion volume as location-specific marks as well as both model variants with and without additional subject-specific covariates. An overview of the data set is provided in Section 3.2.

6.2.1 Algorithmic details and posterior computation

We run the MCMC algorithm for a total of 16,000 iterations, discarding the first 6,000 as burn-in. Estimation of posterior quantities is based on 1,000 samples, using every 10th iteration after burn-in and discarding the rest. As in Section 6.1 we assume a Gaussian representation for the mark process. The prior specifications are as follows: We assign vague priors to the Gamma random field parameters $\beta \sim \mathcal{G}(0.001, 0.001)$ and $\tau \sim \mathcal{G}(0.001, 0.001)$. The mark parameters have broad Gaussian priors $a_j \sim \mathcal{N}(0.1, 100)$, $b_j \sim \mathcal{N}(1.0, 1000)$ and $\sigma_w^2 \sim \mathcal{IG}(5.0, 10.0)$ based on an empirical analysis of expected cube-root transformed lesion volume. The variance for any random walk Metropolis-Hastings update is initialised to $\zeta_*=0.1$ and adjusted every 25 iterations during burn-in. The adjustment depends on the acceptance rate of update proposals as measured over the previous 100 iterations. The kernel variance in the convolution with the Gamma random field is held fixed for both groups at $\sigma_j^2=5.0$. The Bayesian classifier charged with predicting conversion/non-conversion uses equal prior probabilities for each of the two groups. The truncation parameter M for the evaluation of all Gamma random fields is set to $M=18,000$.

6.2.2 Posterior results and prediction

Figure 6.6 and Figure 6.7 summarise results from fitting the imHPGRF model to the BENEFIT data set. The two figures show an axial and a sagittal slice, respectively, of the estimated mean posterior intensity function in real space and on the log-scale, as well as lesion locations of the input data. As can be seen from the intensity maps for CIS and CDMS, the point patterns of both groups appear very similar, a fact that is also reflected in the high values of the shared intensity map.

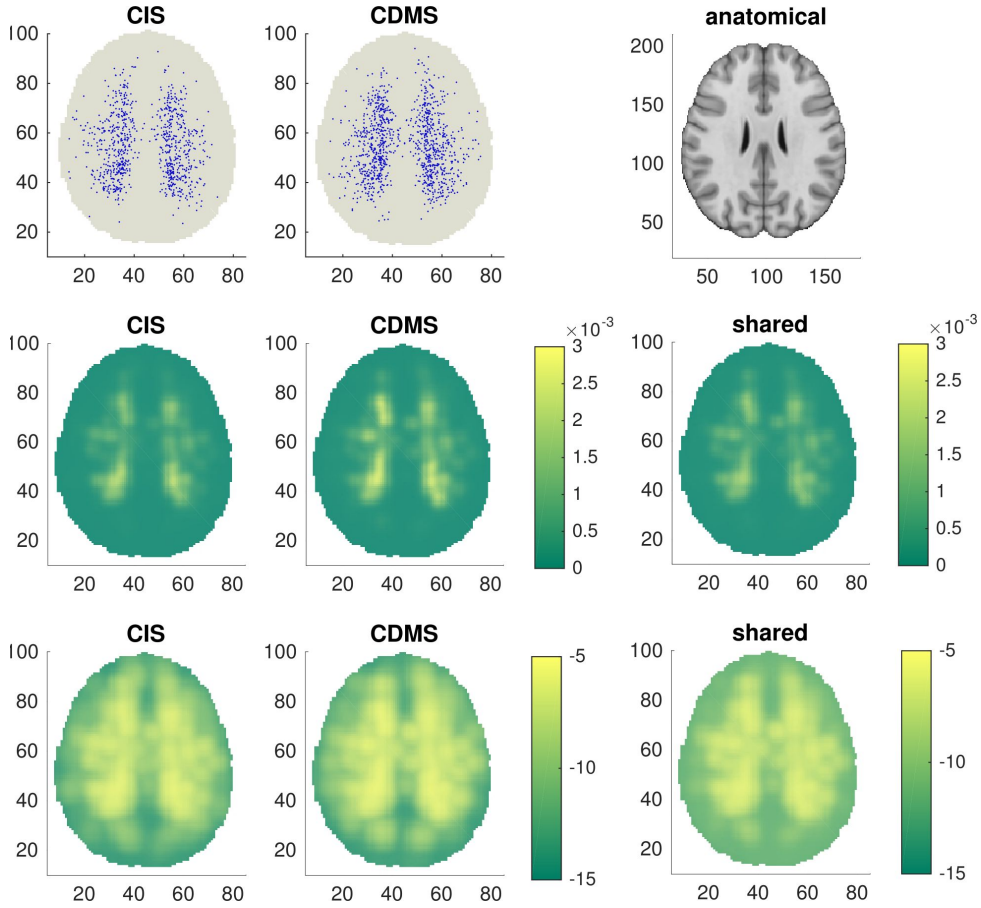


Figure 6.6: Results from fitting the imHPGRF model to the BENEFIT data set. The plots show an axial slice at $z=+2$. Intensity estimates are based on 1,000 posterior samples. *Top row:* Lesion locations that fall within $\pm 4mm$ of the current slice and anatomical reference. *Middle row:* Estimated posterior mean intensities for the two groups in the left and centre panel, estimated shared (population-level) intensity in the panel on the right. *Bottom row:* Log-transform of the estimated posterior mean intensities.

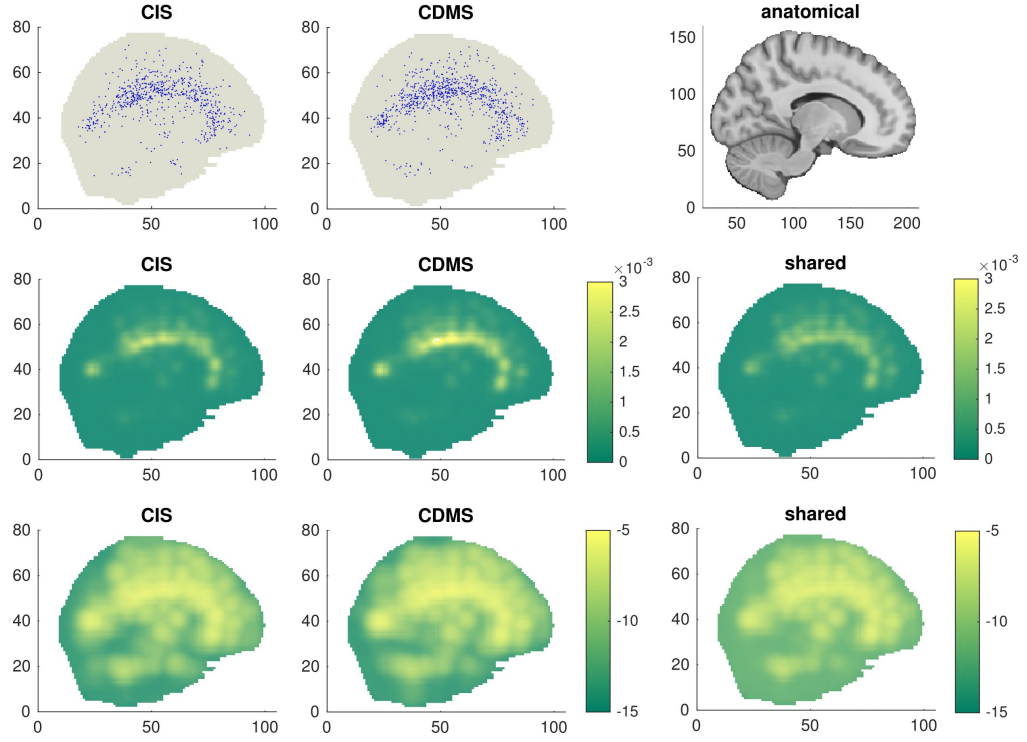


Figure 6.7: Results from fitting the imHPGRF model to the BENEFIT data set. The plots show a sagittal slice at $x=+10$. Intensity estimates are based on 1,000 posterior samples. *Top row:* Lesion locations that fall within $\pm 4mm$ of the current slice and anatomical reference. *Middle row:* Estimated posterior mean intensities for the two groups in the left and centre panel, estimated shared (population-level) intensity in the panel on the right. *Bottom row:* Log-transform of the estimated posterior mean intensities.

Median posterior estimates of the main model parameters, including 95% credible intervals, are presented in Table 6.6 and Table 6.7. Large values for the baseline parameter a of the mark process for both groups and moderate values of the coefficient b for the intensity-dependence indicate a weak correlation between spatial intensity and the mark process. The credible intervals of the regression coefficients cover zero for two out of the three subject-specific covariates. Only age appears to be somewhat informative with respect to the number of lesions per patient.

Residuals from estimated mark values and the Poisson regression are shown in Figure 6.8 and Figure 6.9, respectively. Similar to what we found for the GeneMSA data set, the mark process tends to favour the grand mean, leading to a modest overestimation of a majority of marks and a stark underestimation of few very large mark values. The plots of Poisson regression residuals show on average a slight overestimation of the number of lesions per subject. The mark residuals in

Figure 6.8 show an overestimation of mark values for the majority of lesions. The skewed distribution may indicate that, instead of using the lesion volume directly, modelling the log lesion volume in the mark process may be a better choice. However, with respect to the classification outcomes, we did not find an increase in prediction accuracy when applying the log-Normal mark model.

Confusion matrices alongside classification rates as determined via LOOCV are summarised in Table 6.8. We have applied independent, hierarchical and marked variants of the Poisson/Gamma random field model to the data set. The hierarchical formulation is superior to the separate modelling of groups, since it can share information across the groups. The inclusion of marks and covariates, however, appears to have little effect on the predictive accuracy of the classifier. The highest achieved classification accuracy is 68.7%.

Table 6.6: Posterior parameter estimates for the imHPGRF model including covariates: Type-specific parameters. The mark parameters a and b represent the mean and interaction strength with the spatial intensity of the mark distribution, respectively.

	CIS	CDMS
a_j (median)	3.97	3.86
a_j (95% CI)	[3.90, 4.08]	[3.80, 3.94]
b_j (median)	149.5	[92.4, 207.5]
b_j (95% CI)	230.3	[182.1, 276.4]

Table 6.7: Posterior parameter estimates for the imHPGRF model including covariates: Shared parameters common to all types. Regression coefficients for subject-level covariates are denoted as γ_p , the corresponding covariate is indicated in brackets. The mark variance is denoted by σ_w^2 .

	Median	95% CI
σ_w^2	4.34	[4.22, 4.47]
γ_1 (sex)	0.016	[-0.004, 0.035]
γ_2 (age)	-0.028	[-0.048, -0.009]
γ_3 (EDSS)	0.006	[-0.013, 0.026]

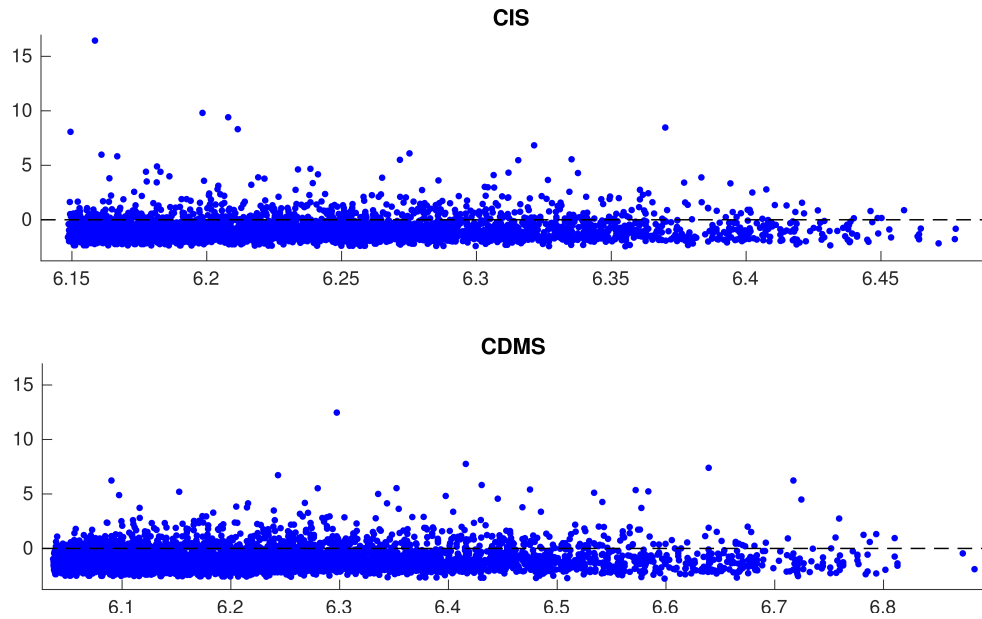


Figure 6.8: Mark residuals of posterior estimates of individual lesion volume versus predicted values.

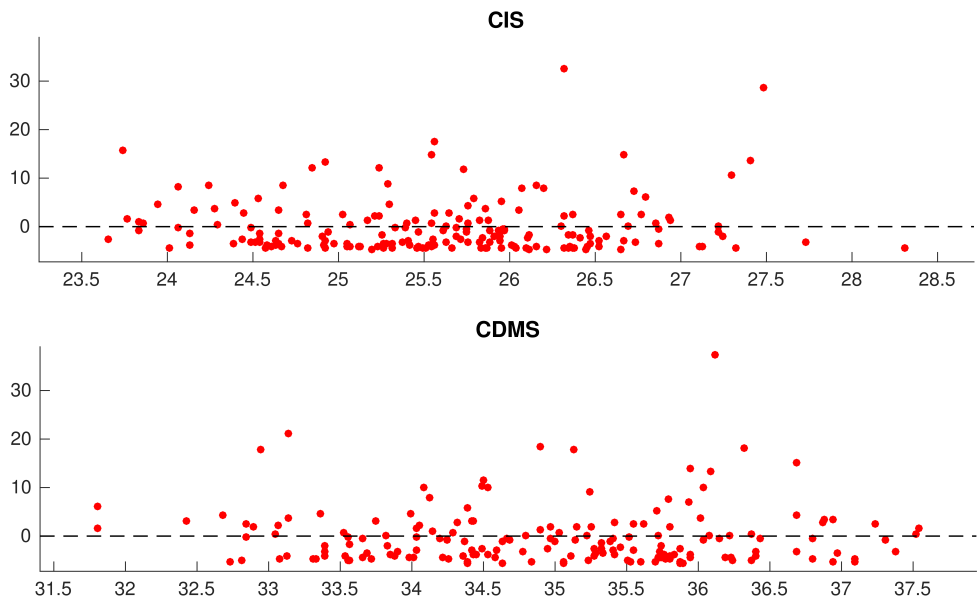


Figure 6.9: Covariate residuals of posterior estimates of lesion count per subject versus predicted values.

Table 6.8: Confusion matrices for different model variants, listing average and total classification accuracy for each variant.

IPGRF: Avg. acc. 56.4%, total acc. 56.9%.

	CIS	CDMS
CIS	0.68 (128)	0.32 (61)
CDMS	0.55 (96)	0.45 (79)

HPGRF: Avg. acc. 66.0%, total acc. 66.5%.

	CIS	CDMS
CIS	0.78 (147)	0.22 (42)
CDMS	0.46 (80)	0.54 (95)

imHPGRF: Avg. acc. 66.1%, total acc. 66.5%.

	CIS	CDMS
CIS	0.77 (145)	0.23 (44)
CDMS	0.45 (78)	0.55 (97)

imHPGRF with covariates: Avg. acc. 68.3%, total acc. 68.7%.

	CIS	CDMS
CIS	0.78 (148)	0.22 (41)
CDMS	0.42 (73)	0.58 (102)

6.2.3 Model assessment

Convergence of the MCMC algorithm is assessed via visual inspection of traceplots of the model parameters. Traceplots, together with plots of the autocorrelation function, are shown in Figure B.9 and Figure B.10, respectively. The figures show good mixing behaviour for all parameters. The longest autocorrelation can be observed for the group-specific inverse scale parameter τ of the Gamma random fields. Based on the thinned MCMC chain, the autocorrelation for τ falls below the confidence bounds after five to ten samples. For all other parameters, subsequent samples are already uncorrelated.

We performed three separate runs of the same imHPGRF model with different starting positions drawn from priors to evaluate convergence of the MCMC

algorithm. None of the parameters have shown sensitive behaviour with regard to prior specifications. The mean posterior estimates of the three separate runs overlap for all quantities of interest and fall within the range of each other's standard error which indicates convergence of the Monte Carlo chain.

We compute first and second order posterior predictive checks using 1,000 samples from the posterior. A comparison of the estimated number of points per group, based on the integrated mean posterior intensity function, with the data is given in Table 6.9. The model fit is very close to the empirical mean for both groups. The second order statistic based on differences in L -functions also indicates a good model fit. We compute the 95% credible intervals of $\Delta_{j,k}(r)$ for distances r ranging from 0 to 100 voxels in steps of 1, corresponding to a maximum distance of 200mm. The proportion of subject whose credible interval covers zero lies above 80% for distances $r \leq 40mm$ and remains above 65% for $r > 40mm$.

Table 6.9: Empirical data and posterior estimates of the average number of lesions per subject for CIS and CDMS groups based on integrated spatial intensities.

	CIS	CDMS
Empirical mean	25.5	34.7
Posterior mean	25.9	34.6
95% CI	[25.2, 26.7]	[33.7, 35.4]

6.2.4 A GLM analysis

Finally, results from fitting a GLM to three subject-level covariates, biological sex, age and EDSS, are summarised in Table 6.10 and visualised in Figure 6.10. The set of regressors for the log-linear model include a nuisance parameter indicating group membership. The signs of regression coefficients in the GLM are consistent with our estimates using the imHPGRF model. The most informative covariate in both cases appears to be age. However, note that this comparison is only valid if inferences on the group level and the individual show the same effects, as pointed out in Subsection 6.1.4 based on the concept of the ecological fallacy.

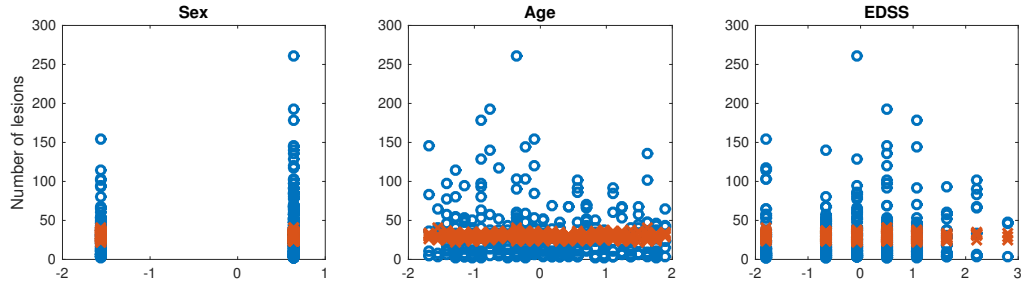


Figure 6.10: Results from fitting a GLM to the BENEFIT data set. The x-axes show the standardised explanatory variables sex (female is coded positive), age and Expanded Disability Status Scale (EDSS). The dependent variable is the number of lesions per subject. Red crosses indicate the fitted values; blue circles indicate true values.

Table 6.10: Summary of results from fitting a GLM to subject-specific covariates, including standard errors (s.e.) for coefficient estimates.

	Intercept	Sex	Age	EDSS
Coefficient (s.e.)	3.391 (0.010)	0.043 (0.011)	-0.094 (0.010)	0.062 (0.010)
p-values	/	$1.52 \cdot 10^{-05}$	$1.07 \cdot 10^{-21}$	$1.30 \cdot 10^{-10}$
t-statistic	/	4.33	-9.57	6.42

6.3 Discussion

In this chapter we have applied our hierarchical Poisson/Gamma random field models to two clinical data sets of patients with multiple sclerosis. To our knowledge, this is the first application of spatial point process models to MRI lesion data. Common ways of analysing neuroimaging data rely on mass-univariate approaches on usually do not take into account any spatial structure of the data. Furthermore, existing methods do not allow for the rich inference that is possible with a generative model like the HPGRF. The estimated intensity function in the imHPGRF model first and foremost is informed by the point locations in the data, but also depends on external covariates and marks attached to each point. We have used the intensity function to predict and classify new or previously unseen lesion patterns into one of several disease subtypes. Apart from classification tasks, the model allows for the interpretation of the intensity function from a clinical point of view. Differences in intensity between disease subtypes or otherwise distinct groups of patients have the potential to better understand, for example, which areas of the brain show an accumulation of lesions; and, furthermore, which groups share a higher rate of lesions in a particular region.

Despite our focus on neuroimaging data, note that the methods presented in this thesis are directly generalisable to any other multi-type point pattern problem, with the exception of repulsive point processes. The framework of the family of Poisson/Gamma random field models only allows for the modelling of positive correlations between the number of points. Repulsive interactions between neighbouring points would have to be incorporated separately, for example, in the marks. This implies also a shortcoming of the model with respect to lesion data, or generally, any point pattern data where the points resemble finite dimensional objects. In its current formulation, the model assumes that two points belonging to the same realisation can be located arbitrarily close to one another. Modelling the mark process in a way that would accommodate something like an exclusion radius could therefore be able to better reflect the true nature of the data.

There are several ways of changing and extending the analysis of these multiple sclerosis data sets. In its present form the HPGRF model cannot accommodate point pattern data belonging to the same subject but resulting from different imaging modalities. As discussed in Subsection 2.2.4, there are three principal types of MRI acquisition for imaging MS lesions, “black holes” on T_1 -weighted MRI, T_2 -weighted hyper-intense lesions and T_1 -weighted Gadolinium-enhanced lesions. These lesions vary both in their appearance (including shape, size, count and location) and clinical interpretation. Integration of multi-modality data into the model, for example by adding another level to the hierarchy, has the potential to produce more accurate estimates of the intensity function.

The application of our point process model to the GeneMSA data set has shown that predictive performance can be high even with very sparse point patterns and small group sizes for some of the target classes. In the application, we have used individual, cube-root transformed lesion volume as marks and modelled them with a Normal distribution which is straightforward to interpret. However, other distributions may be more suitable. For example, a log-Gaussian distribution would account for the fact that lesion volume is strictly positive. Yet, when applying a log-Gaussian mark structure to modelling of the GeneMSA data set, we have not found any substantial differences in the intensity estimates or prediction accuracy compared to the model with Gaussian marks. Similarly, changing the mark structure in the analysis of the BENEFIT data set to account for a log-dependency did not produce different results.

One could also utilise other attributes about individual lesions as marks. For example, various measures of lesion geometry, similar to the ones we have used to train an SVM classifier in Chapter 4, can be computed from MRI lesion masks.

Other measures of interest could include texture features derived from MR intensities. For a comprehensive account of these attributes, one could further extend the model to include multi-variate marks.

We have also looked into using variance-stabilised marks as described in Subsection 5.4.4 where the marks are modelled via a log-dependence on the spatial intensity. The log-transform has the benefit of smoothing the posterior estimate of the intensity. However, we found that it only modestly affected the estimation of the mark process.

A general benefit of non-parametric models is that they make only weak assumptions about the structure of the data. In our applications of the imHPGRF model, we fixed the kernel parameters σ_j^2 rather than estimate them alongside the other model parameters. There are several reasons for doing this. First, holding the kernel variance fixed significantly reduces the computational cost of each iteration by allowing to pre-compute and re-use the values for the integrated kernel function over the whole brain. Second, a fixed kernel width prevents the model from estimating “grainy” intensity functions, which may be the natural tendency in cases where few data points are available; while a good fit, these intensity functions may be difficult to interpret. However, and depending on the application, if one can assume that there are stark discrepancies on different spatial scales between the point processes of different types, then estimating the kernel variance parameter is likely to produce better results. During our exploration of various settings for the HPGRF and imHPGRF models, we have found that estimating the smoothing kernel separately for each subtype generally does not improve model performance.

The imHPGRF model with covariates performs similarly to the log-Gaussian Cox process model in Chapter 4 (87.6% average classification accuracy for the imHPGRF model versus 85.1% for the LGCP model). Although, it must be noted, in the application of the LGCP, the data comprised an additional subtype which the model had to estimate.

As already noted, one drawback of our model, and, in fact, three-dimensional point process models in general, is the computational cost which scales non-linearly with the size of input data. Employing tools for parallel programming such as OpenMP is a way of reducing the time necessary to run the MCMC algorithm. Additionally, implementation of the code for use on one or even multiple GPUs has a large potential to reduce execution time as many parameter updates as well as voxel-by-voxel intensity computations can principally be carried out in parallel. However, transferring the current CPU code would be a non-trivial task.

Finally, methods for preliminary analysis of point pattern data could include

basic clustering techniques to determine, for instance, how many classes the clustering algorithm can detect using simple distance measures. One could also use kernel density estimation [van Lieshout and Stein, 2012] as a possibly easier way of modelling the underlying spatial intensity.

The application studies in this chapter have shown that the HPGRF model is able to accurately estimate an underlying intensity function that can generate point patterns like the ones observed in the data. The addition of a mark process had only limited effects in the two applications of MS lesion data. The cube-root transformed lesion volume that we used here as marks may not be as informative as we had expected. The inclusion of demographic and clinical covariates per subject slightly improved predictive performance in the BENEFIT data set and more so in the GeneMSA application. In both cases, age appeared to be among the most informative predictors for conversion to CDMS and MS subtype, respectively.

There is certainly potential to further improve these methods. An interesting extension with respect to lesion data would be to include a white matter density map as a spatially varying covariate. Furthermore, a spatio-temporal framework that accounts for follow-up scans and changes in a subject's lesion pattern over time may be able to aid decision making in a clinical setting, for example, in the prediction of disease progression or treatment outcomes.

Chapter 7

Conclusions

7.1 Contributions

In this thesis we have explored various approaches to modelling MRI lesion data in patients with multiple sclerosis, with a particular emphasis on Bayesian spatial point process models.

In Chapter 4 we have contrasted two machine learning methods with a probit regression model and a spatial point process model. The machine learning approaches included a simple naïve Bayes classifier and support vector machines that utilised a rich feature set of MRI-based measures as well as non-imaging attributes. The mass-univariate regression model relied on voxel-by-voxel binary lesion masks and the spatial point process model used locations of lesion centres to estimate the intensity function of a doubly-stochastic Poisson process. The success of the point process approach in the prediction task may be unexpected as it does not take into account any patient-specific information other than the coordinates of individual lesions. Traditionally, demographic attributes and clinical scores are heavily relied upon in the clinical management of MS. Although less successful, the machine learning approach showed that a quantitative assessment of lesions on MRI data can be improved by considering geometry-based measures.

In Chapter 5 we have presented novel extensions to the Poisson/Gamma random field model [Wolpert and Ickstadt, 1998b; Kang et al., 2014a]. The results of a simulation study suggest that the method can successfully model three-dimensional sparse multi-type point patterns, including external covariates and location-dependent marks. Additionally, by using a Bayesian importance sampling approach, the model can be used to accurately predict different types of point pattern with only a moderate increase in computational cost.

Motivated by the limited use of spatial data resulting from magnetic resonance imaging in the assessment of multiple sclerosis, in Chapter 6 we apply our spatial point process model to two clinical data sets of MS patients. The model performs well, even with small sample sizes. However, there is room for improvement with regard to which covariates and lesion-specific attributes are most informative when used in combination with the point pattern data. Only few models have been proposed that fit both an underlying point process and a dependent mark process to point pattern data [Ho and Stoyan, 2008; Myllymäki and Penttinen, 2009]. To our knowledge, none of these have been applied to three-dimensional point patterns and joint models of global covariates, continuous marks and point locations have not been considered.

An advantage of our Poisson/Gamma random field model over kernel-based methods lies in the fact that the PGRF does not require the fixing of an (arbitrary) kernel size as smoothing parameter. Our model formulation explicitly allows for the estimation of the kernel variance that is used in the convolution with the Gamma random field. Additionally, the Bayesian framework allows for a wide range of inferences on quantities of interest and model assessment based on posterior predictive probabilities. With regard to MS lesion data, the probabilistic formulation of the PGRF model inherently accounts for the random occurrence of lesions across the brain and therefore is likely to be a more accurate representation of the unobserved data-generating process than parametric methods.

For some applications it may be useful to include a general offset or baseline for the intensity function. For instance, by splitting the intensity function into a smoothly varying baseline intensity and a residual stochastic process, the baseline intensity can be used to model the overall expected incidence rate in the population. The remaining spatial variation would be explained by the stochastic process. The intensity offset could be estimated from the data, for example, with the help of kernel smoothing techniques. [Diggle et al., 2005] have used an approach like this to create a spatio-temporal model of surveillance data on gastroenteric disease. The baseline intensity is used to model the normal incidence rate, whereas the stochastic process models deviations from the expected number of cases of the disease. By employing a critical threshold, this model can be used to detect abnormal fluctuations in the stochastic part of the intensity function, which in turn may warrant a public health intervention.

We did not include an offset for the intensity for several reasons. First, any choice of kernel width is to some extent arbitrary and it is unclear which values

would be appropriate since MRI data already contain an implicit smoothing as a result of the acquisition process. Furthermore, for certain areas in the brain, e.g. in the deep grey matter, we would not expect to find many lesions due to the general disease pathology of multiple sclerosis. On the other hand, incorporating a spatial mask that would exclude such areas could result in losing valuable information from any discarded points. Additionally, a large kernel smoother could bias the model, especially for groups, i.e. disease subtypes, for which sample size is small. Finally, smoothing could lead to cases where local differences in the incidence rate may not be picked up by the model.

7.2 Future work

The analysis of neuroimaging data remains an interesting and challenging area of statistical research. The development of new methodology and the advancement of existing approaches has important implications for the evaluation and treatment of neurological diseases such as MS. The methods presented in this work can be a basis for ways of extending the use of quantitative measurements based on MRI data in clinical practice. The use of spatial information in particular has the potential to inform clinical decision making in the management of MS. In addition, our spatial point process models are not limited to a specific type of application but are directly applicable to multi-type point pattern data in general.

The hierarchical PGRF model considers multiple types of point pattern and multiple realisations of each point process. In future work we would like to extend the model structure by introducing additional levels to the hierarchy. This could be particularly useful in neuroimaging applications for which multiple sources of point pattern data exist, for example, different MRI modalities.

An interesting idea that we have not pursued in this work but one that may have considerable merit in neuroimaging applications in which the objects that give rise to individual points have a non-trivial geometry is to model each object (lesion) with multiple points instead of only one, each point carrying a radius. In the simplest formulation, each lesion would be represented by a collection of spheres; in a more elaborate version, spherical harmonics could be used to account for more complex shapes. A categorical mark on each point could carry the information which points belong to the same lesion.

Other lines of research could focus on the incorporation of a spatio-temporal framework, making our current point process models available to applications in which changes over time are of particular interest. Clinical assessment, prognosis

and the monitoring of disease progression and treatment outcomes are only one area where the inclusion of time would naturally be of importance.

A direct extension that could be considered is the addition of other marks to the imHPGRF model in order to account for a multivariate mark structure, for example, marks based on intra-lesion intensity measures, hence including more and different sources of information about individual point locations. A direction of work concerning computational cost and the problem of poor scaling of the algorithm with very large data sets would be the implementation on a GPU. Currently, our models use some degree of parallelisation through the use of OpenMP. However, we believe the potential of GPU's to carry out parallel computations on a much larger scale would lead to a substantial further reduction in running time.

Appendix A

imHPGRF supplements

A.1 Posterior computation

In the following we denote with $\pi(\theta|\cdot)$ the full conditional posterior distribution of any parameter θ given all other parameters of the model. We employ a Gibbs sampling scheme, updating the model parameters consecutively for each iteration of the Monte Carlo algorithm. Details of each parameter update are given below and include Metropolis-within-Gibbs sampling, random walk Metropolis-Hastings updates as well as direct draws from standard distributions in cases where conjugate priors can be used.

The sampling algorithms presented in this section are with respect to the intensity-marked HPGRF model with global covariates as described in Section 5.4.

The joint posterior density is given in (5.21) and, for completeness, repeated here:

$$\begin{aligned}
\pi[\cdot] \propto & \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[- \exp \left(\sum_{p=1}^P z_{j,i,p} \gamma_p \right) \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} + \sum_{p=1}^P z_{j,i,p} \gamma_p \right] \right. \\
& \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \Big\} \\
& \times \prod_{j=1}^J \left\{ \pi(\sigma_j^2) \prod_{m=1}^M \left[\frac{\tau^{\nu_m}}{\Gamma(\nu_m)} \eta_{j,m}^{\nu_m-1} \exp(-\tau \eta_{j,m}) \right] \right\} \times \pi(\tau) \\
& \times \exp[-E_1(\beta \nu_M)] \prod_{m=1}^M \left\{ \frac{1}{\nu_m} \exp[-\nu_m \beta] \right\} \times \pi(\beta) \\
& \times \prod_{j=1}^J \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \frac{1}{\sigma_w} \exp \left[- \frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2\sigma_w^2} \right] \times \pi(a_j) \times \pi(b_j) \right\} \\
& \times \pi(\sigma_w^2). \tag{A.1}
\end{aligned}$$

Modifications of update routines for different variants of the HPGRF model are indicated for each set of parameters. In order to simplify some of the expressions below, we define the following quantities: the subject-specific covariate term $Z_{j,i} \equiv \exp \left[\sum_{p=1}^P z_{j,i,p} \gamma_p \right]$, the current estimate of the intensity function at a given point $\lambda_{j,i,l} \equiv \sum_{m=1}^M k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \eta_{j,m}$ and the integrated kernel over the whole domain for a given jump location $K_j \equiv \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m}$. Further, denote the total number of points in type j as $L_j = \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} 1$. In the case of Metropolis-Hastings updates, the acceptance ratio will be denoted as $r(\tilde{\theta}, \theta)$ with $\tilde{\theta}$ representing the new proposal and θ being the current parameter value.

A.1.1 Sampling of mark parameters

Note that we assume a log-Gaussian mark process throughout the following. In fact, the updates remain also valid in the case of a Gaussian mark distribution when replacing $\ln w_{j,i,l}$ with $w_{j,i,l}$. However, the interpretation of mark parameters a_j and b_j changes.

Updating a_j

Use a Normal prior for the type-specific baseline parameter, i.e. $a_j \sim \mathcal{N}(\mu_a, \zeta_a^2)$. Hence, the full conditional is proportional to

$$\pi(a_j|\cdot) \propto \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \exp \left[-\frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2\sigma_w^2} \right] \right\} \times \exp \left[-\frac{(a_j - \mu_a)^2}{2\zeta_a^2} \right]. \quad (\text{A.2})$$

Due to conjugacy, one can draw directly from

$$a_j|\cdot \sim \mathcal{N} \left(\frac{\sigma_w^2 \mu_a + \zeta_a^2 L_j \overline{m_j}}{L_j \zeta_a^2 + \sigma_w^2}, \frac{\sigma_w^2 \zeta_a^2}{L_j \zeta_a^2 + \sigma_w^2} \right). \quad (\text{A.3})$$

Here, $\overline{m_j} \equiv \frac{1}{L_j} \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} (\ln w_{j,i,l} - b_j Z_{j,i} \lambda_{j,i,l})$ stands for the re-parameterised mean log-mark per type.

Updating b_j

Use a Normal prior on the type-specific interaction parameter such that $b_j \sim \mathcal{N}(\mu_b, \zeta_b^2)$. The full conditional distribution is given by

$$\pi(b_j|\cdot) \propto \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \exp \left[-\frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2\sigma_w^2} \right] \right\} \times \exp \left[-\frac{(b_j - \mu_b)^2}{2\zeta_b^2} \right], \quad (\text{A.4})$$

which can be rewritten as

$$\pi(b_j|\cdot) \propto \exp \left\{ -\sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \frac{(b_j - m_{j,i,l})^2}{2s_{j,i,l}} \right\} \times \exp \left\{ -\frac{(b_j - \mu_b)^2}{2\zeta_b^2} \right\}, \quad (\text{A.5})$$

where $m_{j,i,l} \equiv (\ln w_{j,i,l} - a_j)/(Z_{j,i} \lambda_{j,i,l})$ and $2s_{j,i,l} \equiv 2\sigma_w^2/(Z_{j,i} \lambda_{j,i,l})^2$.

Hence, the update is conjugate and follows

$$b_j|\cdot \sim \mathcal{N} \left[\left(\sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \frac{m_{j,i,l}}{s_{j,i,l}} + \frac{\mu_b}{\sigma_b^2} \right) \left(\sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \frac{1}{s_{j,i,l}} + \frac{1}{\sigma_b^2} \right)^{-1}, \left(\sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \frac{1}{s_{j,i,l}} + \frac{1}{\sigma_b^2} \right)^{-1} \right]. \quad (\text{A.6})$$

Updating σ_w^2

Due to conjugacy we place a gamma-prior on the mark precision which is equivalent to an inverse gamma-prior on the mark variance parameter, i.e. $\frac{1}{\sigma_w^2} \sim \Gamma(\alpha_{\sigma_w}, \beta_{\sigma_w})$ or $\sigma_w^2 \sim \mathcal{IG}(\alpha_{\sigma_w}, \beta_{\sigma_w})$. The full conditional is proportional to

$$\begin{aligned} \pi(\sigma_w^2 | \cdot) &\propto \left[\prod_{j=1}^J \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \frac{1}{\sigma_w} \exp \left\{ -\frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2\sigma_w^2} \right\} \right] \\ &\quad \times \sigma_w^{-2(1+\alpha_{\sigma_w})} \exp \left\{ -\frac{\beta_{\sigma_w}}{\sigma_w^2} \right\} \\ &\propto \left(\frac{1}{\sigma_w^2} \right)^{\alpha_{\sigma_w} + \frac{L}{2} - 1} \exp \left\{ -\frac{1}{\sigma_w^2} \left[\sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2} + \beta_{\sigma_w} \right] \right\}. \end{aligned} \quad (\text{A.7})$$

Therefore, the updating step results in a draw from

$$\sigma_w^2 | \cdot \sim \mathcal{IG} \left(\alpha_{\sigma_w} + \frac{L}{2}, \beta_{\sigma_w} + \frac{1}{2} \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} (\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2 \right), \quad (\text{A.8})$$

where L denotes the total number of points, i.e. $L = \sum_{j=1}^J L_j$, and $\mathcal{IG}(\alpha, \beta)$ is given as $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}$. Note that the mean of the inverse gamma distribution is given by $\frac{\beta}{\alpha-1}$ and the variance by $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, for $\alpha > 2$.

A.1.2 Sampling of regression coefficients

Updating γ_p

The full log-conditional is proportional to

$$\begin{aligned} \log \pi(\gamma_p | \cdot) &\propto \sum_{j=1}^J \sum_{i=1}^{N_j} \left[-\exp \left\{ \sum_{p'=1}^P z_{j,i,p'} \gamma_{p'} \right\} K_j + z_{j,i,p} \gamma_p \right] - \\ &\quad - \frac{1}{2\sigma_w^2} \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left(\ln w_{j,i,l} - a_j - b_j \lambda_{j,i,l} \exp \left\{ \sum_{p'=1}^P z_{j,i,p'} \gamma_{p'} \right\} \right)^2, \end{aligned} \quad (\text{A.9})$$

Assume an improper, uninformative prior to initialise the regression coefficients: $\pi(\gamma_p) \sim 1$. Use a random-walk Metropolis-Hastings update and draw a new value for $\tilde{\gamma}_p$ by sampling from $\tilde{\gamma}_p \sim \mathcal{N}(\gamma_p, \zeta_\gamma^2)$. Set $\gamma_p \equiv \tilde{\gamma}_p$ with probability $\min\{r(\tilde{\gamma}_p, \gamma_p), 1\}$ or equivalently, $\min\{\log r(\tilde{\gamma}_p, \gamma_p), 0\}$. The log-acceptance ratio can be expressed as

$$\begin{aligned} \log r(\tilde{\gamma}_p, \gamma_p) &= \sum_{j=1}^J \left[K_j \sum_{i=1}^{N_j} (Z_{j,i} - \tilde{Z}_{j,i}) \right] + (\tilde{\gamma}_p - \gamma_p) \sum_{j=1}^J \sum_{i=1}^{N_j} z_{j,i,p} \\ &+ \frac{1}{2\sigma_w^2} \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left[(\ln w_{j,i,l} - a_j - b_j \lambda_{j,i,l} Z_{j,i})^2 - (\ln w_{j,i,l} - a_j - b_j \lambda_{j,i,l} \tilde{Z}_{j,i})^2 \right], \end{aligned} \quad (\text{A.10})$$

where $\tilde{Z}_{j,i} \equiv \exp \left\{ \sum_{p' \neq p}^P z_{j,i,p'} \gamma_{p'} + z_{j,i,p} \tilde{\gamma}_p \right\}$. The expression above can be written as

$$\begin{aligned} \log r(\tilde{\gamma}_p, \gamma_p) &= \sum_{j=1}^J \left[K_j \sum_{i=1}^{N_j} (Z_{j,i} - \tilde{Z}_{j,i}) \right] + (\tilde{\gamma}_p - \gamma_p) \sum_{j=1}^J \sum_{i=1}^{N_j} z_{j,i,p} \\ &+ \frac{1}{2\sigma_w^2} \sum_{j=1}^J \sum_{i=1}^{N_j} \left\{ (\tilde{Z}_{j,i} - Z_{j,i}) b_j \sum_{l=1}^{L_i} \left[\lambda_{j,i,l} \left(2(\ln w_{j,i,l} - a_j) - b_j \lambda_{j,i,l} (\tilde{Z}_{j,i} + Z_{j,i}) \right) \right] \right\} \end{aligned} \quad (\text{A.11})$$

The update is easily modified to model variants without marks by removing the second line in (A.11). The update is the same for the IPGRF model.

A.1.3 Sampling of Gamma random fields

Updating auxiliary points $x_{j,i,l}$

As can be seen directly from the full target distribution, the conditional distribution of $x_{j,i,l}$ given all other parameters is

$$\begin{aligned} \pi(x_{j,i,l} | y_{j,i,l}, \Gamma_j, \sigma_j^2) &\propto \sum_{m=1}^M \eta_{j,m} k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \mathbb{I}_{\theta_m}(x_{j,i,l}) \\ \pi(x_{j,i,l} = \theta_m | \cdot) &\propto \eta_{j,m} k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \end{aligned} \quad (\text{A.12})$$

A way to sample from this distribution is to draw an index m^* from a multinomial distribution with probabilities given by the normalised cumulative probabilities of

(A.12). Then assign the auxiliary point $x_{j,i,l}$ to the jump location θ_{m^*} . There are no modifications of the update routine necessary for HPGRF or IPGRF model variants with or without marks and with or without covariates.

Updating jump locations θ_m

The full conditional for the set of jump locations $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^M$ is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}|\cdot) \propto & \prod_{j=1}^J \prod_{i=1}^{N_j} \left\{ \exp \left[-Z_{j,i} \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} \right] \right. \\ & \times \prod_{l=1}^{L_i} \left[k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \sum_{m=1}^M \eta_{j,m} \mathbb{I}_{\theta_m}(x_{j,i,l}) \right] \Big\} \\ & \times \prod_{j=1}^J \left\{ \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \exp \left[-\frac{(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2}{2\sigma_w^2} \right] \right\}. \end{aligned} \quad (\text{A.13})$$

The total number M of jump locations must be pre-specified in the algorithm and should be sufficiently larger than the number of points in the data set. To initialise the jump sources of the Gamma random field, draw M points uniformly distributed over \mathcal{B} . The auxiliary points share the same locations as the jump sources, i.e. each $x_{j,i,l}$ equals some θ_m . These assignments are not unique which means there exists a set of distinct points $\{\theta_t\}_{t=1}^T$, with $T \leq L < M$, that is a subset of $\{\theta_m\}_{m=1}^M$. Each θ_t has at least one $x_{j,i,l}$ assigned to it. In order to sample the first T latent sources, that is for all $m \leq T$ when one or more auxiliary points are clustered at the jump location θ_m , use a random-walk Metropolis-Hastings update with proposal distribution $\tilde{\theta}_m \sim \mathcal{N}(\theta_m, \zeta_\theta^2)$. The hyperparameter ζ_θ^2 can be adjusted during burn-in to optimise the acceptance rate. The log-acceptance ratio is given by

$$\begin{aligned} \log r(\tilde{\theta}_m, \theta_m) = & \sum_{j=1}^J \sum_{i=1}^{N_j} \left[K_{\sigma_j^2}(\mathcal{B}, \theta_m) - K_{\sigma_j^2}(\mathcal{B}, \tilde{\theta}_m) \right] \eta_{jm} Z_{j,i} \\ & + \frac{1}{2\sigma_w^2} \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left[(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2 \right. \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} & \left. - (\ln w_{j,i,l} - a_j - b_j Z_{j,i} \tilde{\lambda}_{j,i,l})^2 \right] \\ & - \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \ln k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \mathbb{I}_{x_{j,i,l}}(\theta_m), \end{aligned} \quad (\text{A.15})$$

where $\tilde{\lambda}_{j,i,l} \equiv \sum_{m' \neq m}^M k_{\sigma_j^2}(y_{j,i,l}, \theta_{m'}) \eta_{jm'} + k_{\sigma_j^2}(y_{j,i,l}, \tilde{\theta}_m) \eta_{jm}$. The expression above can be written as

$$\begin{aligned} \log r(\tilde{\theta}_m, \theta_m) &= \sum_{j=1}^J \sum_{i=1}^{N_j} \left[K_{\sigma_j^2}(\mathcal{B}, \theta_m) - K_{\sigma_j^2}(\mathcal{B}, \tilde{\theta}_m) \right] \eta_{j,m} Z_{j,i} \\ &\quad + \frac{1}{2\sigma_w^2} \sum_{j=1}^J \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left[\left(k_{\sigma_j^2}(y_{j,i,l}, \tilde{\theta}_m) - k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \right) \right. \\ &\quad \times \eta_{jm} b_j Z_{j,i} \left\{ 2(\ln w_{j,i,l} - a_j) - b_j Z_{j,i} \eta_{jm} \left(k_{\sigma_j^2}(y_{j,i,l}, \tilde{\theta}_m) - k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \right) \right\} \\ &\quad \left. - \ln k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \mathbb{I}_{x_{j,i,l}}(\theta_m) \right]. \end{aligned} \quad (\text{A.16})$$

For the remaining $M - T$ jump locations with indices $T < m \leq M$, the new proposal is drawn from a uniform distribution over \mathcal{B} , i.e. $\tilde{\theta}_m \sim \mathcal{U}(\mathcal{B})$. The acceptance ratio remains the same as before except that now $\mathbb{I}_{x_{j,i,l}}(\theta_m) = 0$.

The modification for model variants without covariates entails setting all $Z_{j,i} = 1$. For the plain HPGRF without marks and covariates, the log-acceptance ratio simplifies to

$$\begin{aligned} \log r(\tilde{\theta}_m, \theta_m) &= \left[K_{\sigma_j^2}(\mathcal{B}, \theta_m) - K_{\sigma_j^2}(\mathcal{B}, \tilde{\theta}_m) \right] \eta_{j,m} N_j \\ &\quad - \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \ln k_{\sigma_j^2}(y_{j,i,l}, x_{j,i,l}) \mathbb{I}_{x_{j,i,l}}(\theta_m). \end{aligned} \quad (\text{A.17})$$

Updating type level jump heights $\eta_{j,m}$

The full conditional distribution for type-specific jump magnitudes $\boldsymbol{\eta} = \{\{\eta\}_{m=1}^M\}_{j=1}^J$ is given by

$$\begin{aligned} \pi(\eta_{j,m} | \cdot) &\propto \prod_{i=1}^{N_j} \exp \left\{ -Z_{j,i} K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} \right\} \eta_{j,m}^{t_{j,m} + \nu_m - 1} \exp \{ -\tau \eta_{j,m} \} \\ &\quad \times \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \left[\exp \left\{ -\frac{1}{2\sigma_w^2} \left(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \sum_{m=1}^M k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \eta_{j,m} \right)^2 \right\} \right], \end{aligned} \quad (\text{A.18})$$

A.1. Posterior computation

where with $t_{j,m} \equiv \sum_{i=1}^{N_j} \sum_{l=1}^{L_{j,i}} \mathbb{I}_{\theta_m}(x_{j,i,l})$. The first product term can be expressed as a Gamma distribution, returning

$$\begin{aligned} \pi(\eta_{j,m}|\cdot) \propto & \Gamma \left[\nu_m + t_{j,m}, \sum_{i=1}^{N_j} Z_{j,i} K_{\sigma_j^2}(\mathcal{B}, \theta_m) + \tau \right] \\ & \times \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \left[\exp \left\{ -\frac{1}{2\sigma_w^2} (\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2 \right\} \right]. \end{aligned} \quad (\text{A.19})$$

To update $\eta_{j,m}$, draw a proposal according to

$$\tilde{\eta}_{j,m}|\cdot \sim \Gamma \left[\nu_m + t_{j,m}, K_{\sigma_j^2}(\mathcal{B}, \theta_m) \sum_{i=1}^{N_j} Z_{j,i} + \tau \right] \quad (\text{A.20})$$

and accept with probability $\min \{0, \log r(\tilde{\eta}_{j,m}, \eta_{j,m})\}$, with the log-acceptance ratio given by

$$\begin{aligned} \log r(\tilde{\eta}_{j,m}, \eta_{j,m}) = & \frac{1}{2\sigma_w^2} \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left[(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2 \right. \\ & \left. - \left(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \tilde{\lambda}_{j,i,l} \right)^2 \right] \end{aligned} \quad (\text{A.21})$$

and $\tilde{\lambda}_{j,i,l} \equiv \sum_{m' \neq m}^M k_{\sigma_j^2}(y_{j,i,l}, \theta_{m'}) \eta_{j,m'} + k_{\sigma_j^2}(y_{j,i,l}, \theta_m) \tilde{\eta}_{j,m}$.

For the un-marked HPGRF model, the update is conjugate and reduces to drawing from

$$\eta_{j,m}|\cdot \sim \Gamma \left\{ \nu_m + \sum_{i=1}^{N_j} \sum_{l=1}^{L_{j,i}} \mathbb{I}_{\theta_m}(x_{j,i,l}), N_j K_{\sigma_j^2}(\mathcal{B}, \theta_m) + \tau \right\}. \quad (\text{A.22})$$

Updating population level jump heights ν_m

Define $c_m \equiv \prod_{j=1}^J \tau e^{-\beta} \eta_{j,m}$. The full conditional for all but the last population level jump heights is given by

$$\pi(\nu_{m < M}|\cdot) \propto \frac{c_m^{\nu_m}}{\Gamma[\nu_m]} \nu_m^{-1}. \quad (\text{A.23})$$

Due to the inverse Lévy measure construction of the Gamma random field, the analogous expression for the last jump height is

$$\pi(\nu_M|\cdot) \propto \exp\{-\mathbb{E}_1[\beta\nu_M]\} \frac{c_M^{\nu_M}}{\Gamma[\nu_M]} \nu_M^{-1}. \quad (\text{A.24})$$

We use a random-walk Metropolis-Hastings update with a symmetric proposal distribution, thus $\tilde{\nu}_m \sim \mathcal{N}(\nu_m, \zeta_\nu^2)$. The log-acceptance rate can be expressed as

$$\begin{aligned} \log r(\tilde{\nu}_m, \nu_m) &= \ln c_m(\tilde{\nu}_m - \nu_m) - (\ln \Gamma[\tilde{\nu}_m + 1] - \ln \Gamma[\nu_m + 1]) \\ &\quad - (\mathbb{E}_1[\beta\tilde{\nu}_M] - \mathbb{E}_1[\beta\nu_M]) \delta_{m,M}, \end{aligned} \quad (\text{A.25})$$

where we used the relation $\ln \Gamma[x] + \ln x = \ln \Gamma[x + 1]$.

Alternatively, by using a Taylor expansion around $\Gamma[x]$, one can show that $\pi(\nu_m|\cdot)$ are log-concave functions and hence ν_m can also be updated via Adaptive Rejection Sampling.

Updating type level inverse scale parameter τ

We use a Gamma prior for the group level inverse scale parameter: $\tau \sim \Gamma(a_\tau, b_\tau)$. The full conditional is given by

$$\pi(\tau|\cdot) \propto \tau^{J \sum_{m=1}^M \nu_m + a_\tau - 1} \exp \left\{ - \left(b_\tau + \sum_{j=1}^J \sum_{m=1}^M \eta_{j,m} \right) \tau \right\}, \quad (\text{A.26})$$

which implies that the update is conjugate and that τ can be drawn according to

$$\tau|\cdot \sim \Gamma \left(J \sum_{m=1}^M \nu_m + a_\tau, \sum_{j=1}^J \sum_{m=1}^M \eta_{j,m} + b_\tau \right) \quad (\text{A.27})$$

Updating population level inverse scale parameter β

We again use a Gamma prior such that $\beta \sim \Gamma(a_\beta, b_\beta)$. The full conditional is given by

$$\pi(\beta|\cdot) \propto \beta^{a_\beta - 1} \exp \left\{ - \left(b_\beta + \sum_{m=1}^M \nu_m \right) \beta - \mathbb{E}_1(\beta\nu_M) \right\}. \quad (\text{A.28})$$

A.1. Posterior computation

For the update we employ a symmetric random-walk Metropolis-Hastings step by drawing a proposal value from $\tilde{\beta} \sim \mathcal{N}(\beta, \zeta_\beta^2)$. The full conditional in the log-domain is

$$\log \pi(\beta|\cdot) \propto (a_\beta - 1) \log \beta - (b_\beta + \sum_{m=1}^M \nu_m) \beta - \mathbb{E}_1[\beta \nu_M] \quad (\text{A.29})$$

and thus the log-acceptance rate becomes

$$\log r(\tilde{\beta}, \beta) = (a_\beta - 1)(\log \tilde{\beta} - \log \beta) - (b_\beta + \sum_{m=1}^M \nu_m)(\tilde{\beta} - \beta) - \mathbb{E}_1(\tilde{\beta} \nu_M) + \mathbb{E}_1[\beta \nu_M]. \quad (\text{A.30})$$

A.1.4 Sampling of kernel parameters

Updating σ_j^2

We use a uniform prior: $\pi(\sigma_j^{-2}) \sim \mathcal{U}[a_\sigma, b_\sigma]$. The full conditional of the type-specific kernel variance σ_j^2 is given by

$$\begin{aligned} \pi(\sigma_j^2|\cdot) \propto \exp \left\{ -K_j \sum_{i=1}^{N_j} Z_{j,i} - \frac{S_j^2}{2\sigma_j^2} - \frac{d}{2} L_j \ln(\sigma_j^2) \right\} \\ \times \prod_{i=1}^{N_j} \prod_{l=1}^{L_i} \exp \left\{ -\frac{1}{2\sigma_w^2} \left(\ln w_{j,i,l} - a_j \right. \right. \\ \left. \left. - b_j Z_{j,i} \left[\sum_{m=1}^M \left(\frac{1}{\sigma_j^2} \right)^{d/2} \exp \left\{ -\frac{\|y_{j,i,l} - \theta_m\|^2}{2\sigma_j^2} \right\} \eta_{j,m} \right] \right)^2 \right\} \end{aligned} \quad (\text{A.31})$$

with $Z_{j,i} \equiv \exp \left\{ \sum_{p=1}^P z_{j,i,p} \gamma_p \right\}$, $S_j^2 \equiv \frac{1}{2} \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \|x_{j,i,l} - y_{j,i,l}\|^2$ and $L_j \equiv \sum_{i=1}^{N_j} L_i$.

We use a symmetric random-walk update to draw $\tilde{\sigma}_j^2 \sim \mathcal{N}(\sigma_j^2, \zeta_\sigma^2)$. Assuming the draw falls within $\mathcal{U}[a_\sigma, b_\sigma]$, the Metropolis-Hastings log-acceptance rate is given by

$$\begin{aligned} \log r(\tilde{\sigma}_j^2, \sigma_j^2) = \left(K_j - \tilde{K}_j \right) \sum_{i=1}^{N_j} Z_{j,i} + \frac{S_j^2}{2} \left(\frac{1}{\sigma_j^2} - \frac{1}{\tilde{\sigma}_j^2} \right) + \frac{d}{2} L_j (\ln \sigma_j^2 - \ln \tilde{\sigma}_j^2) \\ + \frac{1}{2\sigma_w^2} \sum_{i=1}^{N_j} \sum_{l=1}^{L_i} \left[(\ln w_{j,i,l} - a_j - b_j Z_{j,i} \lambda_{j,i,l})^2 - (\ln w_{j,i,l} - a_j - b_j Z_{j,i} \tilde{\lambda}_{j,i,l})^2 \right], \end{aligned} \quad (\text{A.32})$$

where $K_j \equiv \sum_{m=1}^M K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m}$, $\tilde{K}_j \equiv \sum_{m=1}^M K_{\tilde{\sigma}_j^2}(\mathcal{B}, \theta_m) \eta_{j,m}$ and $\tilde{\lambda}_{j,i,l} \equiv \sum_{m=1}^M k_{\tilde{\sigma}_j^2}(y_{j,i,l}, \theta_m) \eta_{j,m}$. The log-acceptance rate can also be written as

$$\begin{aligned} \log r(\tilde{\sigma}_j^2, \sigma_j^2) &= (K_j - \tilde{K}_j) \sum_{i=1}^{N_j} Z_{j,i} + \frac{S_j^2}{2} \left(\frac{1}{\sigma_j^2} - \frac{1}{\tilde{\sigma}_j^2} \right) + \frac{d}{2} L_j (\ln \sigma_j^2 - \ln \tilde{\sigma}_j^2) \\ &+ \frac{1}{2\sigma_w^2} \sum_{i=1}^{N_j} b_j Z_{j,i} \sum_{l=1}^{L_i} \left\{ \left(\tilde{\lambda}_{j,i,l} - \lambda_{j,i,l} \right) \left[2(\ln w_{j,i,l} - a_j) - b_j Z_{j,i} \left(\tilde{\lambda}_{j,i,l} - \lambda_{j,i,l} \right) \right] \right\} \end{aligned} \quad (\text{A.33})$$

In case of the un-marked HPGRF model without covariates, the full conditional reduces to

$$\pi(\sigma_j^2 | \cdot) \propto \exp \left\{ - \sum_{m=1}^M \left[N_j K_{\sigma_j^2}(\mathcal{B}, \theta_m) \eta_{j,m} \right] - \frac{S_j^2}{2\sigma_j^2} - \frac{d}{2} L_j \log(\sigma_j^2) \right\} \text{I}_{[a_\sigma, b_\sigma]}(\sigma_j^{-2}) \quad (\text{A.34})$$

and the log-acceptance ratio is given by

$$\log r(\tilde{\sigma}_j, \sigma_j) = N_j(K_j - \tilde{K}_j) + \frac{S_j^2}{2} \left(\frac{1}{\sigma_j^2} - \frac{1}{\tilde{\sigma}_j^2} \right) + \frac{d}{2} L_j [\ln(\sigma_j^2) - \ln(\tilde{\sigma}_j^2)]. \quad (\text{A.35})$$

A.1.5 Intensity computation

The group level intensity measure at each MCMC draw for voxel v_r is given by

$$\Lambda_j(v_r) = \sum_{m=1}^M k_{\sigma_j^2}(v_r, \theta_m) \eta_{j,m}. \quad (\text{A.36})$$

After N draws from the posterior, the mean estimated posterior intensity is simply given by

$$\hat{\Lambda}_j(v_r) = \frac{1}{N} \sum_{n=1}^N \Lambda_j^{(n)}(v_r). \quad (\text{A.37})$$

Analogously, the population level intensity is given by

$$\Lambda_{pop}(v_r) = \frac{1}{J} \frac{1}{\tau} \sum_{j=1}^J \sum_{m=1}^M k_{\sigma_j^2}(v_r, \theta_m) \nu_m \quad (\text{A.38})$$

and

$$\hat{\Lambda}_{pop}(v_r) = \frac{1}{N} \sum_{n=1}^N \Lambda_{pop}^{(n)}(v_r). \quad (\text{A.39})$$

The standard deviation of mean posterior intensities is given by

$$\hat{\sigma}_{\Lambda}(v_r) = \left[\frac{1}{N-1} \left(\sum_{n=1}^N \left[\Lambda^{(n)}(v_r) \right]^2 - N \left[\hat{\Lambda}(v_r) \right]^2 \right) \right]^{\frac{1}{2}} \quad (\text{A.40})$$

where $\Lambda \in \{\Lambda_j, \Lambda_{pop}\}$.

Appendix B

Application supplements

B.1 Supplement to the application of the SVM classifier to the GeneMSA data set

B.1. Supplement to the SVM classifier

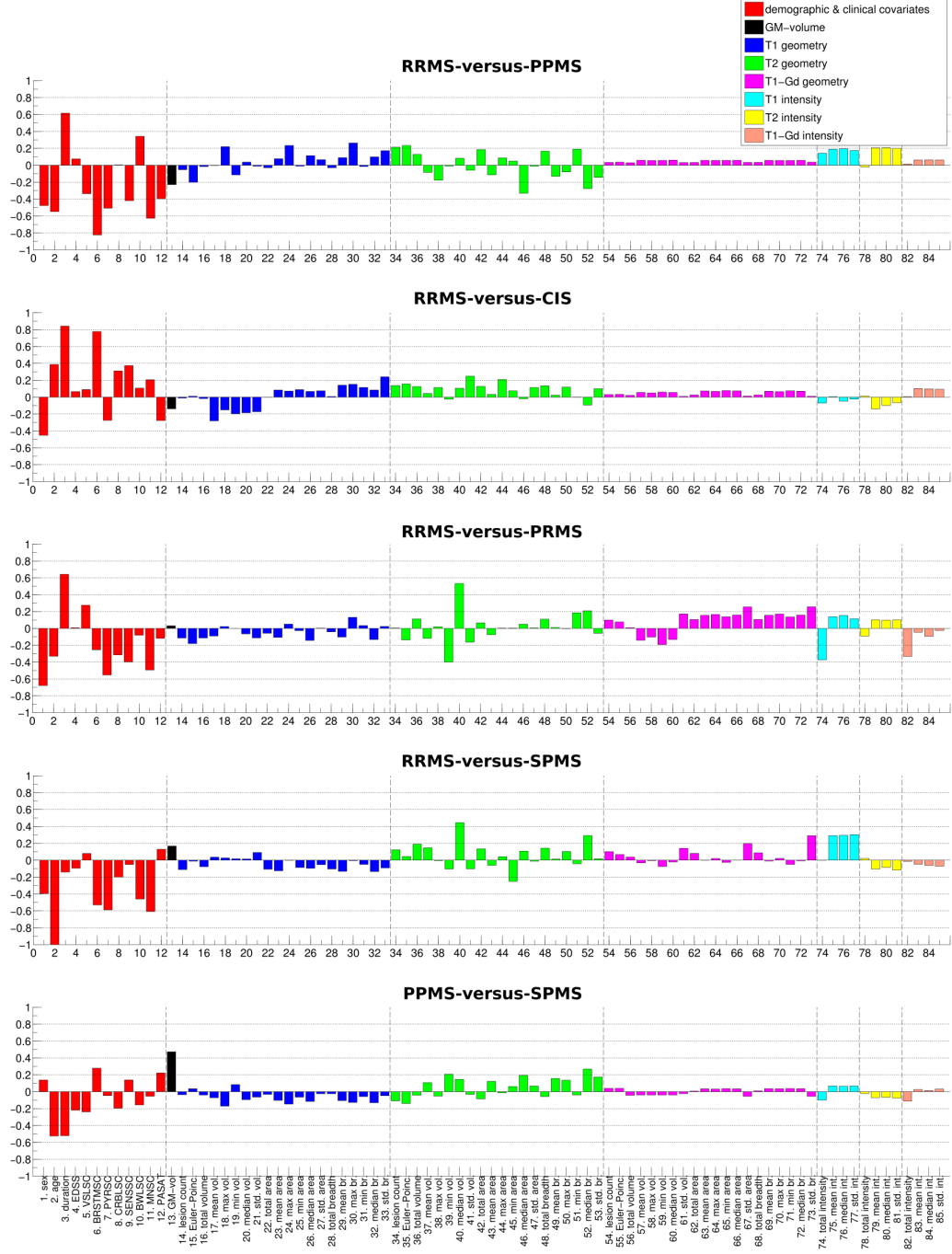


Figure B.1: Standardised SVM weights for five pairwise classifiers (I). All weights have been standardised to the occurring maximum across all classifiers which is age in RRMS-versus-SPMS.

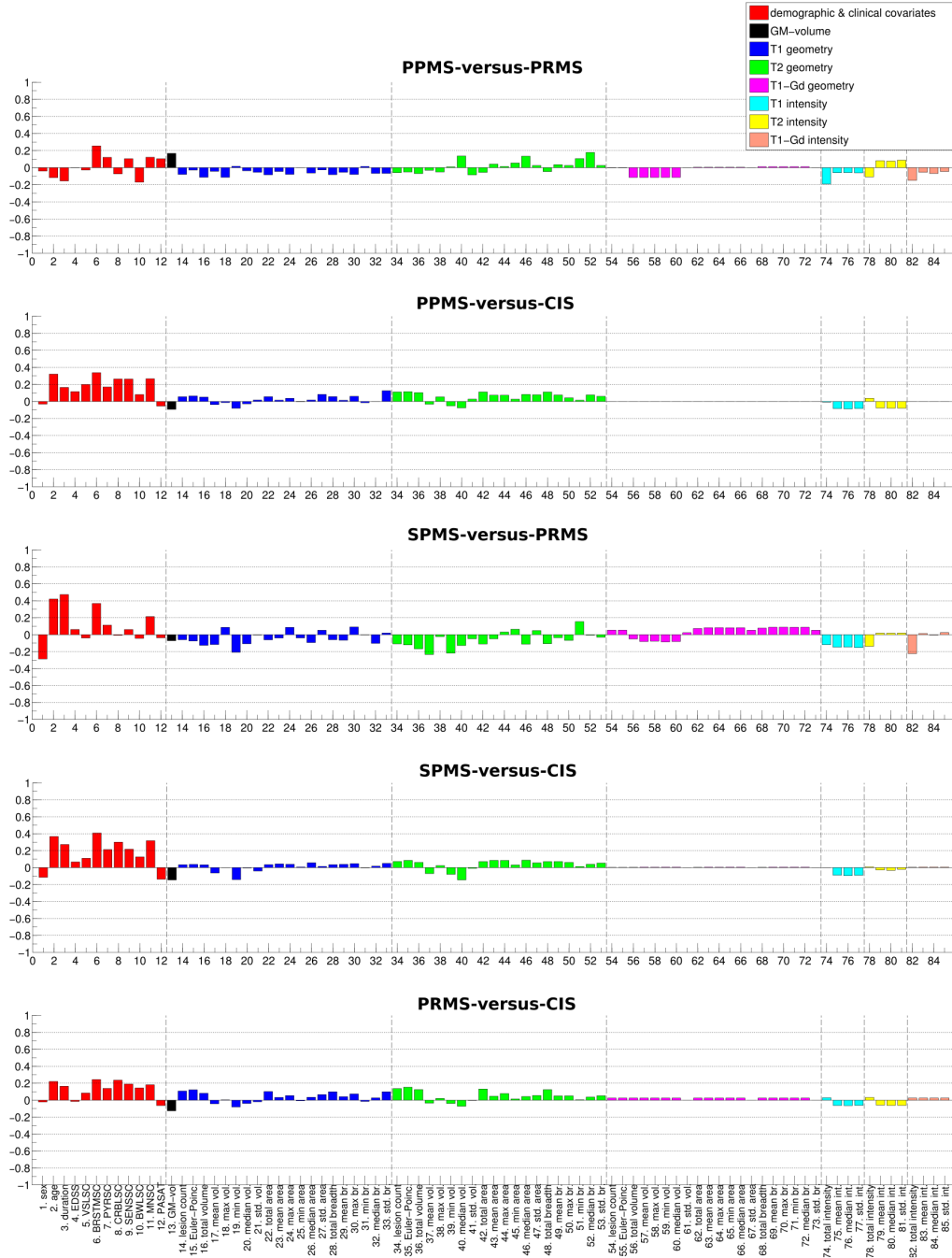


Figure B.2: Standardised SVM weights for five pairwise classifiers (II). All weights have been standardised to the occurring maximum across all classifiers which is age in RRMS-versus-SPMS.

B.2 Supplement to the application of the imHPGRF model to the GeneMSA data set

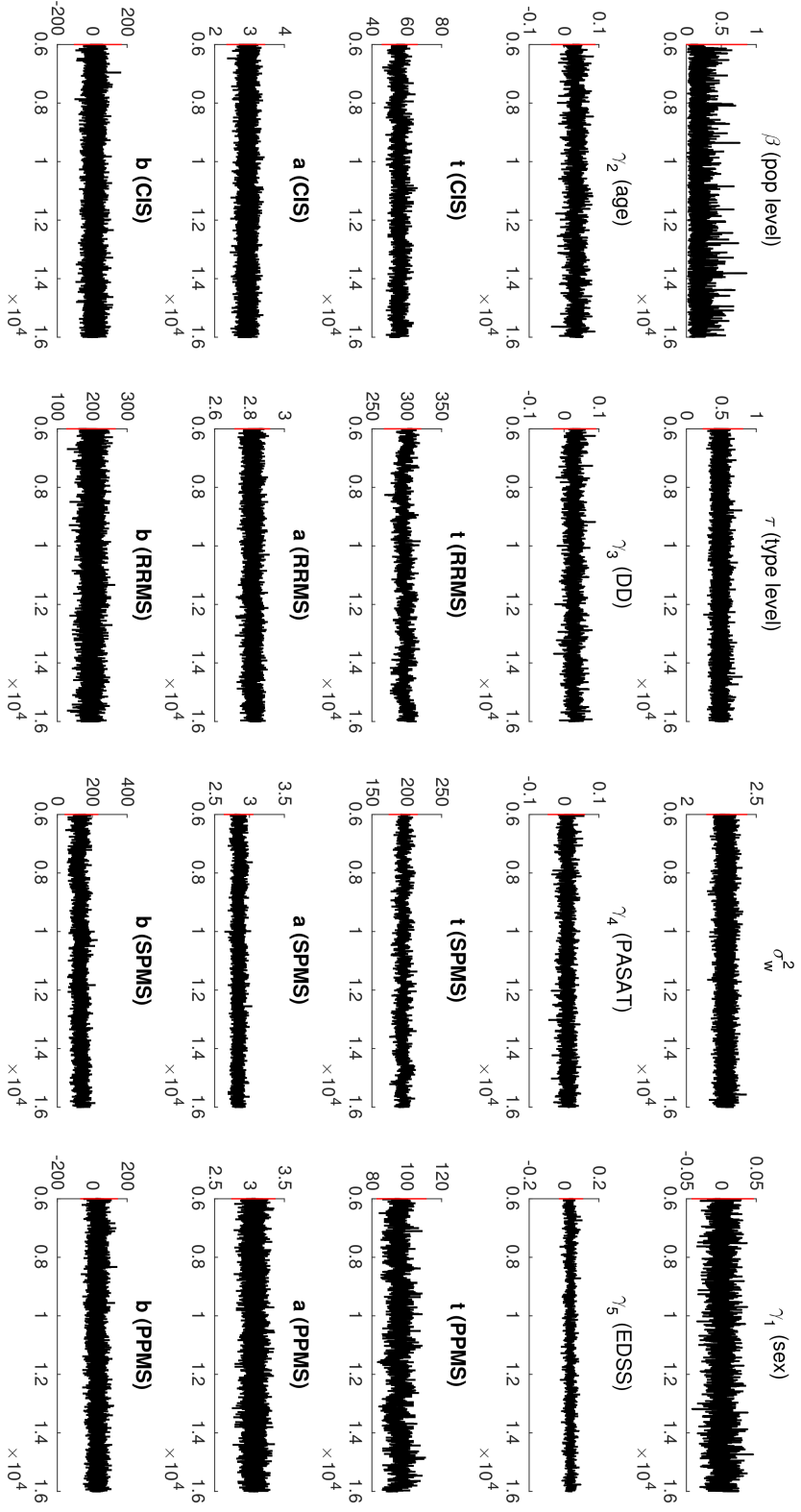


Figure B.3: Traceplots for parameters and quantities of interest of the intensity-marked hierarchical Poisson/Gamma random field model. β and τ are the shared and group-specific inverse scale parameters of the Gamma random fields, respectively. Regression coefficients for subject-level covariates are denoted as γ_p , the corresponding covariate is indicated in brackets. a , b and σ_w^2 are parameters of the mark process and t indicates the number of unique point locations around which auxiliary points from the data augmentation scheme cluster.

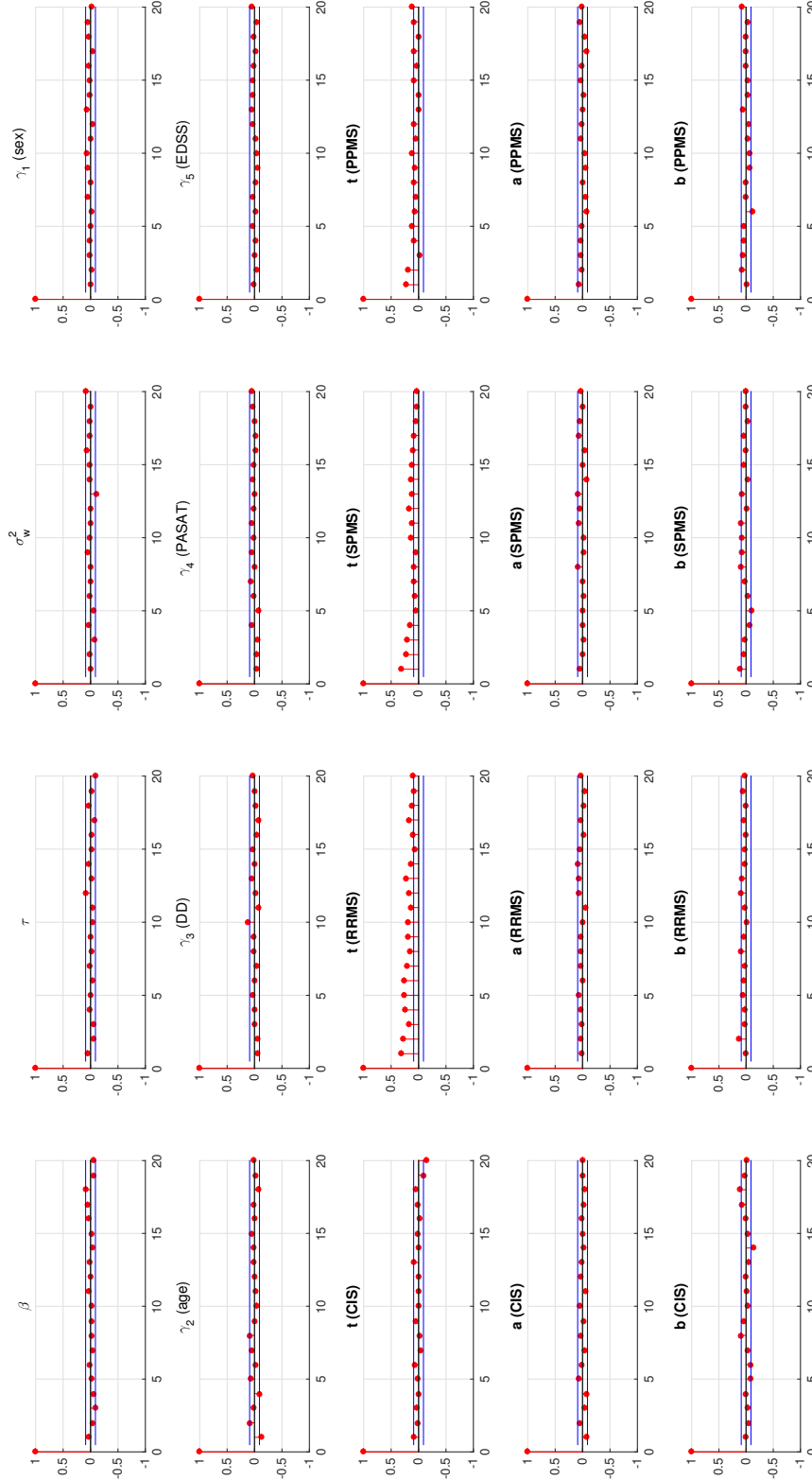


Figure B.4: Autocorrelation plots for parameters of the imHPGRF model based on the thinned MCMC chain. β and τ are the shared and group-specific inverse scale parameters of the Gamma random fields, respectively. Regression coefficients for subject-level covariates are denoted as γ_p , the corresponding covariate is indicated in brackets. a , b and σ_w^2 are parameters of the mark process and t indicates the number of unique point locations around which auxiliary points from the data augmentation scheme cluster.

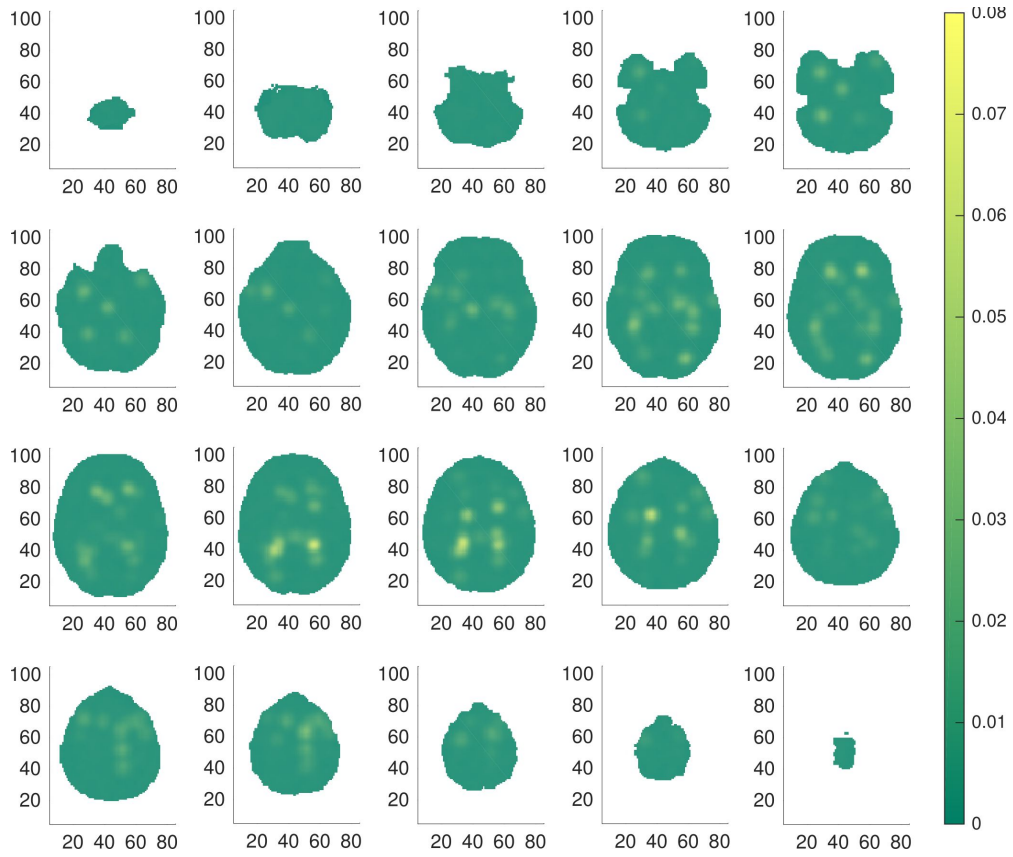


Figure B.5: Square-root transformed posterior mean intensity for MS subtype CIS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

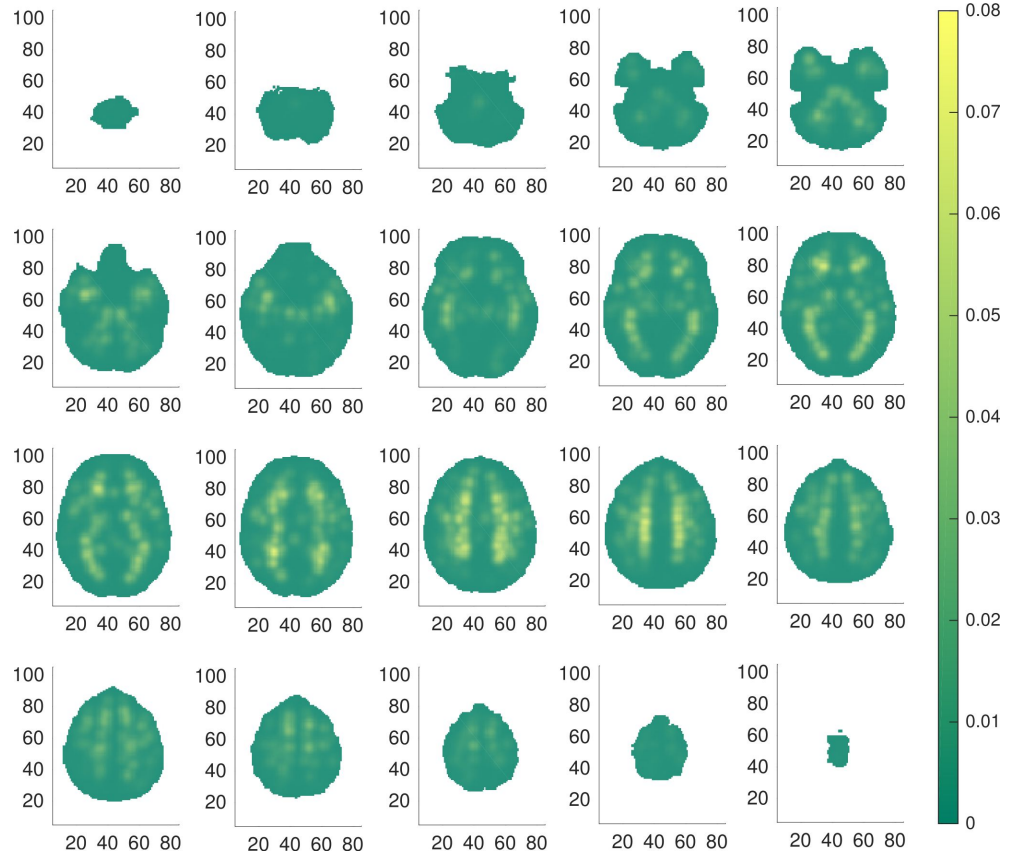


Figure B.6: Square-root transformed posterior mean intensity for MS subtype RRMS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

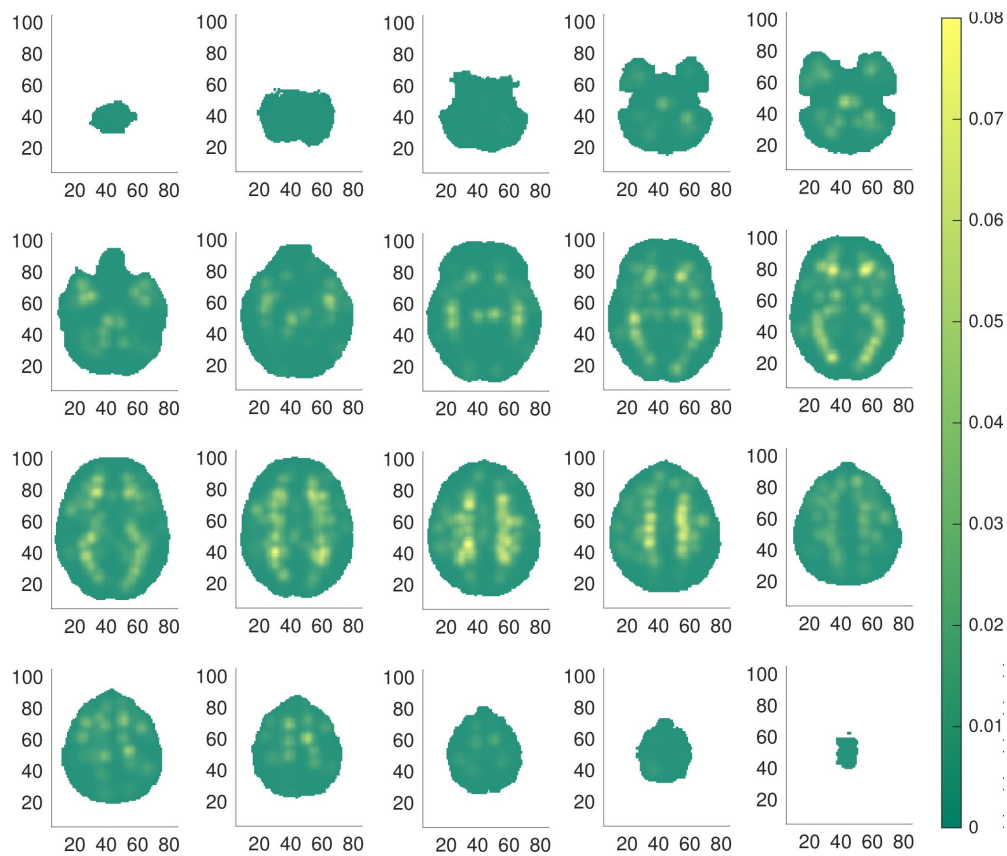


Figure B.7: Square-root transformed posterior mean intensity for MS subtype SPMS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

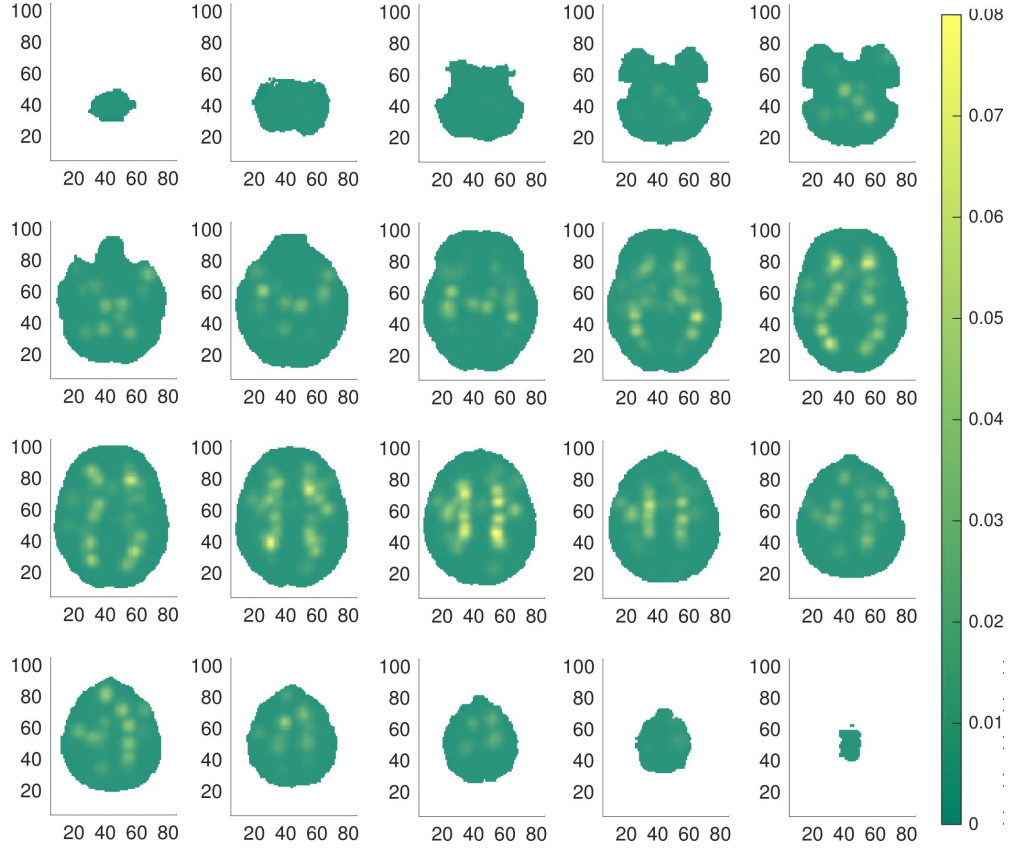


Figure B.8: Square-root transformed posterior mean intensity for MS subtype PPMS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

B.3 Supplement to the application of the imHPGRF model to the BENEFIT data set

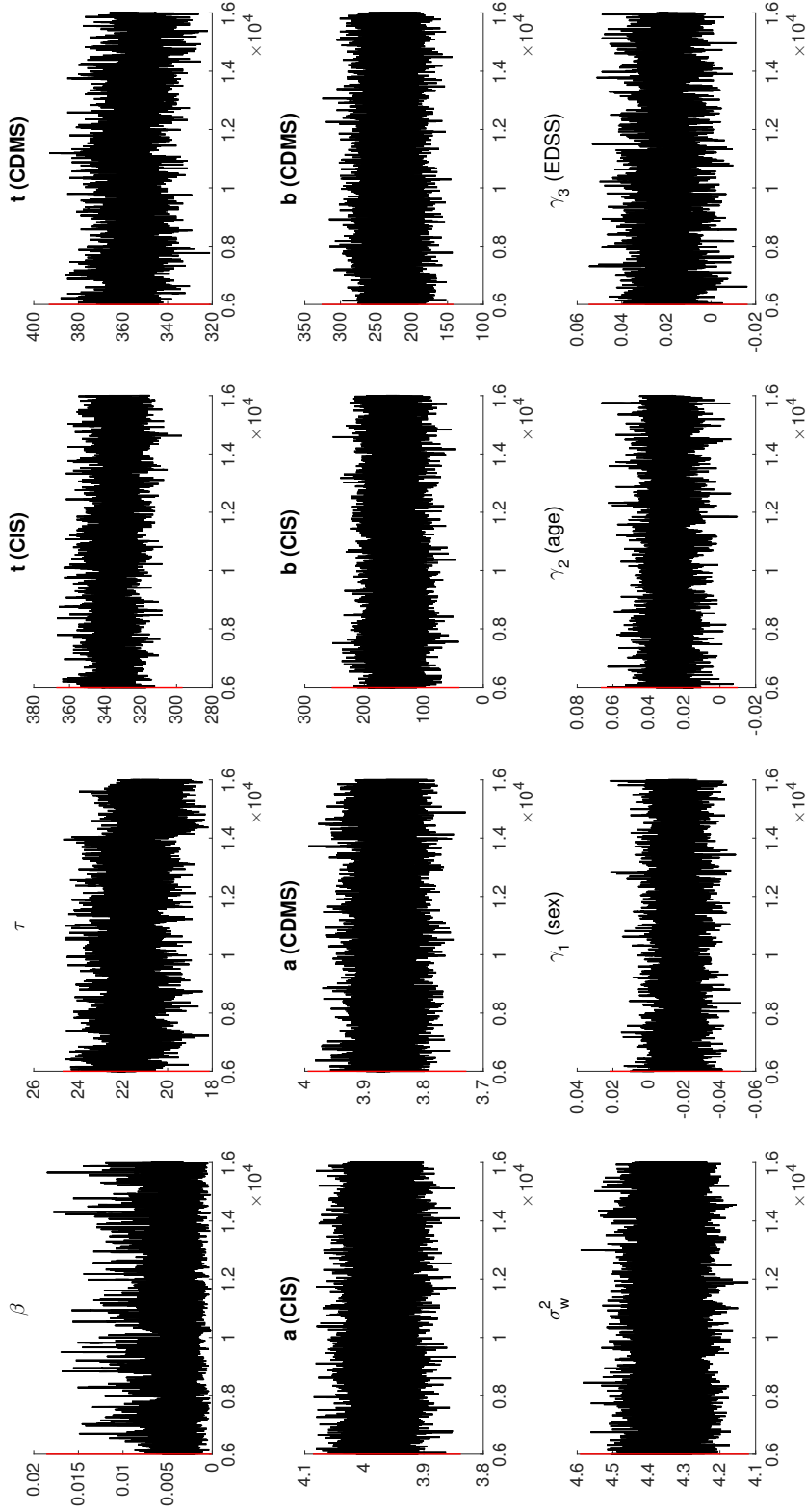


Figure B.9: Traceplots (non-thinned) for parameters and quantities of interest of the imHPGRF model. β and τ denote the population-level and group-specific inverse scale parameters of the Gamma random fields, respectively. The mark process is determined by the baseline parameters a , the interaction strength b with the spatial field and the mark variance σ_w^2 . The number of unique point locations around which auxiliary points of the augmented data cluster are denoted by t .

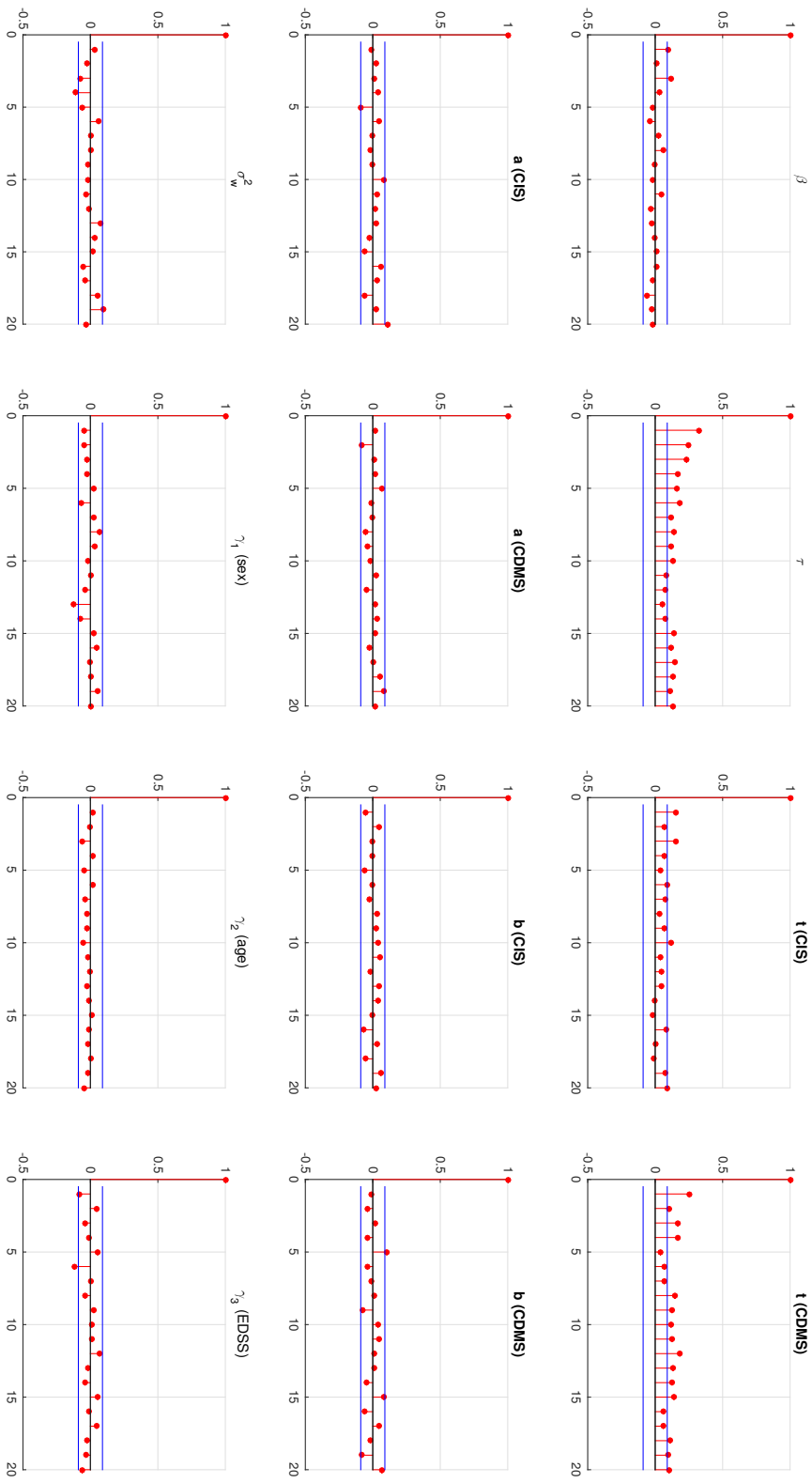


Figure B.10: Autocorrelation plots (thinned) for parameters of the imHPGRF model fit. β and τ are the shared and group-specific inverse scale parameters of the Gamma random fields, respectively. Regression coefficients for subject-level covariates are denoted as γ_j , the corresponding covariate is indicated in brackets. a , b and σ_w^2 are parameters of the mark process and t indicates the number of unique point locations around which auxiliary points from the data augmentation scheme cluster.

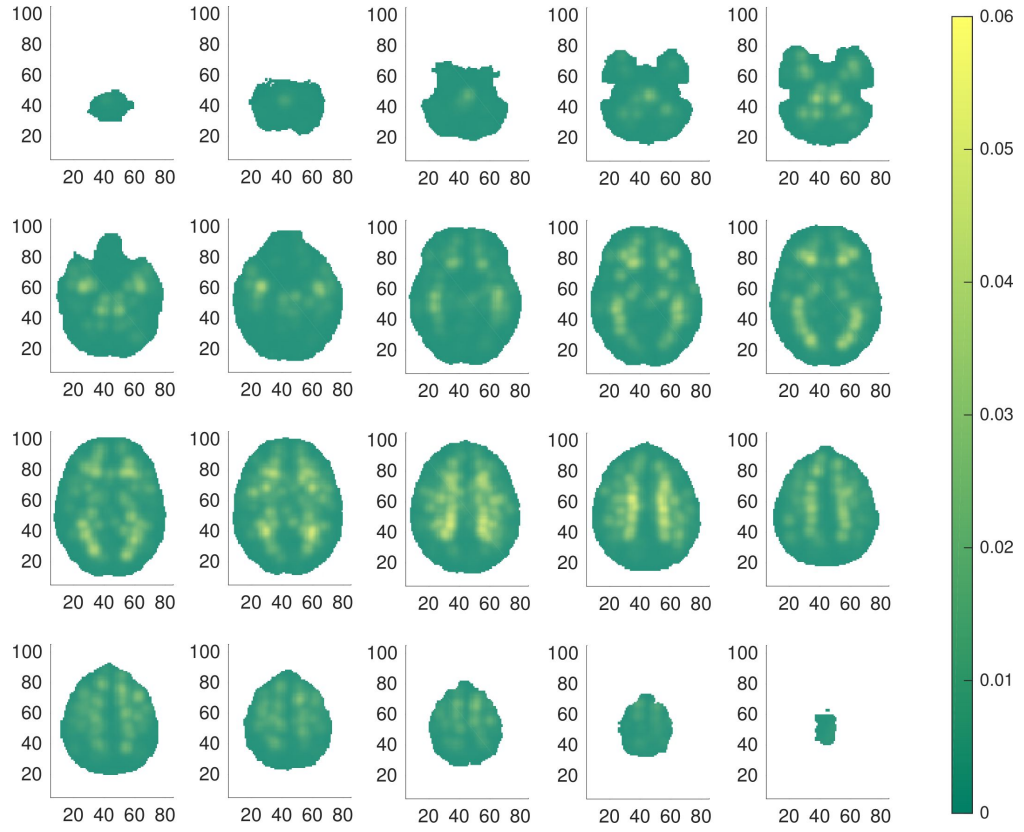


Figure B.11: Square-root transformed posterior mean intensity for CIS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

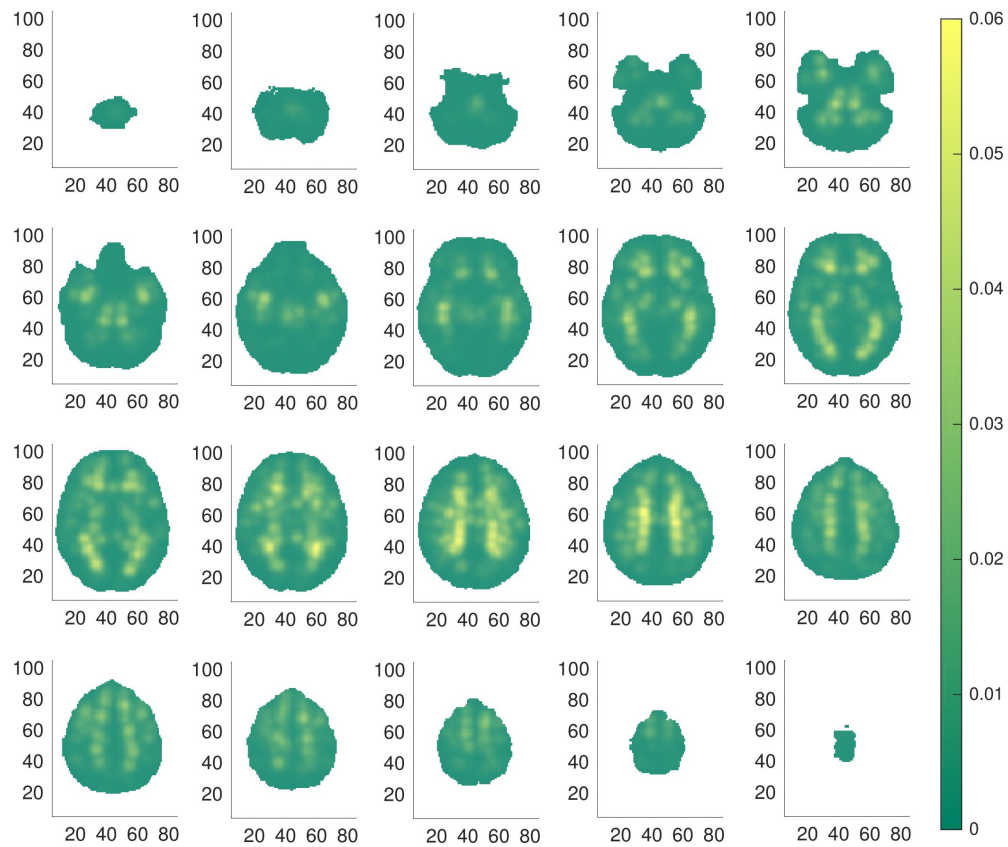


Figure B.12: Square-root transformed posterior mean intensity for CIS. The estimated intensity is based on 1,000 posterior samples. The panels correspond (in increasing order from left to right and top to bottom) to axial slices at $z = -64 : 6 : +50$.

Bibliography

- Aban, I.B., Cutter, G.R., Mavinga, N. (2008). Inferences and power analysis concerning two negative binomial distributions with an application to MRI lesion counts data. *Computational Statistics & Data Analysis* **53**(3), 820–833.
- Alauddin, M.M. (2012). Positron emission tomography (PET) imaging with (18)F-based radiotracers. *American Journal of Nuclear Medicine and Molecular Imaging* **2**(1), 55–76.
- Allwein, E.L., Schapire, R.E., Singer, Y. (2001). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research* **1**, 113–141.
- Alqallaf, F., Gustafson, P. (2001). On cross-validation of Bayesian models. *Canadian Journal of Statistics* **29**(2), 333–340.
- Arns, C.H., Knackstedt, M.A., Mecke, K. (2010). 3D structural analysis: Sensitivity of Minkowski functionals. *Journal of Microscopy* **240**(3), 181–196.
- Arns, C.H., Knackstedt, M.A., Pinczewski, W.V., Mecke, K.R. (2001). Euler-Poincaré characteristics of classes of disordered media. *Phys. Rev. E* **63**(3), 31112.
- Ashburner, J., Friston, K.J. (2000). Voxel-based morphometry - The methods. *NeuroImage* **11**(6 Pt 1), 805–21.
- Babu, G.J., Feigelson, E.D. (1996). Spatial point processes in astronomy. *Journal of Statistical Planning and Inference* **50**(3), 311–326.
- Baddeley, A., Turner, R., Møller, J., Hazelton, M. (2005). Residual analysis for spatial point processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **67**(5), 617–666.
- Baddeley, A.J., Møller, J., Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**(3), 329–350.

- Bakshi, R., Thompson, A.J., Rocca, M.A., Pelletier, D., Dousset, V., Barkhof, F., Inglese, M., Guttman, C.R., Horsfield, M.A., Filippi, M. (2008). MRI in multiple sclerosis: Current status and future prospects. *The Lancet Neurology* **7**(7), 615–625.
- Bar-Shalom, R., Yefremov, N., Guralnik, L., Gaitini, D., Frenkel, A., Kuten, A., Altman, H., Keidar, Z., Israel, O. (2003). Clinical performance of PET/CT in evaluation of cancer: Additional value for diagnostic imaging and patient management. *Journal of Nuclear Medicine* **44**(8), 1200–9.
- Baranzini, S.E., Wang, J., Gibson, R.A., Galwey, N., Naegelin, Y., Barkhof, F., Radue, E.W., Lindberg, R.L.P., Uitdehaag, B.M.G., Johnson, M.R., Angelakopoulou, A., Hall, L., Richardson, J.C., Prinjha, R.K., Gass, A., Geurts, J.J.G., Kragt, J., Sombekke, M., Vrenken, H., Qualley, P., Lincoln, R.R., Gomez, R., Caillier, S.J., George, M.F., Mousavi, H., Guerrero, R., Okuda, D.T., Cree, B.A.C., Green, A.J., Waubant, E., Goodin, D.S., Pelletier, D., Matthews, P.M., Hauser, S.L., Kappos, L., Polman, C.H., Oksenberg, J.R. (2009). Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Human Molecular Genetics* **18**(4), 767–778.
- Barkhof, F., Polman, C.H., Radue, E.W., Kappos, L., Freedman, M.S., Edan, G., Hartung, H.P., Miller, D.H., Montalbán, X., Poppe, P., de Vos, M., Lasri, F., Bauer, L., Dahms, S., Wagner, K., Pohl, C., Sandbrink, R. (2007). Magnetic resonance imaging effects of interferon beta-1b in the BENEFIT study: Integrated 2-year results. *Annals of Neurology* **64**(9), 1292–1298.
- Barkhof, F., Rocca, M., Francis, G., Van Waesberghe, J.H.T.M., Uitdehaag, B.M.J., Hommes, O.R., Hartung, H.P., Durelli, L., Edan, G., Fernández, O., Seeldrayers, P., Sørensen, P., Margrie, S., Rovaris, M., Comi, G., Filippi, M. (2003). Validation of diagnostic magnetic resonance imaging criteria for multiple sclerosis and response to interferon β 1a. *Annals of Neurology* **53**(6), 718–724.
- Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience* **6**(5), 448–450.
- Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J.P.A., Tzoulaki, I. (2015). Environmental risk factors and multiple sclerosis: An umbrella review of systematic reviews and meta-analyses. *The Lancet Neurology* **14**(3), 263–273.

- Bendfeldt, K., Klöppel, S., Nichols, T.E., Smieskova, R., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., Kappos, L., Radü, E.W., Borgwardt, S.J. (2012). Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage* **60**(1), 400–408.
- Bendfeldt, K., Kuster, P., Traud, S., Egger, H., Winklhofer, S., Müller-Lenke, N., Naegelin, Y., Gass, A., Kappos, L., Matthews, P.M., Nichols, T.E., Radü, E.W., Borgwardt, S.J. (2009). Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis - A longitudinal voxel-based morphometry study. *NeuroImage* **45**(1), 60–67.
- Besag, J. (1993). Towards Bayesian image analysis. *Journal of Applied Statistics* **20**(5-6), 107–119.
- Best, N., Ickstadt, K., Wolpert, R.L., Cockings, S., Elliott, P., Bennett, J., Bottle, A., Reed, S. (2002). Modelling the impact of traffic-related air pollution on childhood respiratory illness, *in: Gatsonis, C., Kass, R.E., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., West, M. (Eds.), Case Studies in Bayesian Statistics, Volume V*. Springer New York, pp. 183–259.
- Best, N.G., Ickstadt, K., Wolpert, R.L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association* **95**, 1076–1088.
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business Media, New York, Singapore.
- Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**, 121–167.
- Burslem, D.F.R.P., Garwood, N.C., Thomas, S.C. (2001). Tropical forest diversity: The plot thickens. *Science* **291**(5504), 607.
- Campbell, C. (2002). Kernel methods: A survey of current techniques. *Neurocomputing* **48**(1), 63–84.
- Cawley, G.C., Talbot, N.L.C. (2007). Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters. *Journal of Machine Learning Research* **8**, 841–861.
- Chen, R., Jiao, Y., Herskovits, E.H. (2011). Structural MRI in autism spectrum disorder. *Pediatric Research* **69**(5-2).

- Cohen, J.A., Rae-Grant, A. (2010). *Handbook of Multiple Sclerosis*. Springer Healthcare, London.
- Cohen, J.A., Reingold, S.C., Polman, C.H., Wolinsky, J.S. (2012). Disability outcome measures in multiple sclerosis clinical trials: Current status and future prospects. *The Lancet Neurology* **11**(5), 467–476.
- Cole, D.M., Smith, S.M., Beckmann, C.F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Frontiers in Systems Neuroscience* **4**(April), 8.
- Combrisson, E., Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods* **250**(0), 126–136.
- Compston, A., Coles, A. (2008). Multiple sclerosis. *The Lancet* **372**(9648), 1502–1517.
- Cox, D.R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B* **17**(2), 129–164.
- Cronie, O., van Lieshout, M.N.M. (2016). Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics* **68**(4), 905–928.
- Cronie, O., Van Lieshout, M.N.M. (2015). A J -function for inhomogeneous spatio-temporal point processes. *Scandinavian Journal of Statistics* **42**(1), 562–579.
- Currie, S., Hoggard, N., Craven, I.J., Hadjivassiliou, M., Wilkinson, I.D. (2013). Understanding MRI: Basic MR physics for physicians. *Postgraduate Medical Journal* **89**(1050), 209–23.
- Daley, D.J., Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer. 2nd edition.
- Daley, D.J., Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Springer. 2nd edition.
- Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M. (2008). Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging* **29**(4), 514–523.

- Diggle, P. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman Hall/CRC Press. 3 edition.
- Diggle, P., Rowlingson, B., Su, T.L. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16**(5), 423–434.
- Diggle, P.J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A* **153**(3), 349–362.
- Diggle, P.J., Menezes, R., Su, T.L. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **59**(2), 191–232.
- Diggle, P.J., Moraga, P., Rowlingson, B., 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., Rowlingson, B.S. (1994). A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society. Series A* **157**(3), 433–440.
- Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195**(2), 216–222.
- Ebers, G.C. (2001). Natural history of multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry* **71**, 16–19.
- Filippi, M., Yousry, T., Campi, A., Kandziora, C., Colombo, B., Voltz, R., Martinelli, V., Spuler, S., Bressi, S., Scotti, G., Comi, G. (1996). Comparison of triple dose versus standard dose gadolinium-DTPA for detection of MRI enhancing lesions in patients with MS. *Neurology* **46**(2), 379–84.
- Filli, L., Hofstetter, L., Kuster, P., Traud, S., Müller-Lenke, N., Naegelin, Y., Kappos, L., Gass, A., Sprenger, T., Nichols, T.E., Vrenken, H., Barkhof, F., Polman, C., Radue, E.W., Borgwardt, S.J., Bendfeldt, K. (2012). Spatiotemporal distribution of white matter lesions in relapsing-remitting and secondary progressive multiple sclerosis. *Multiple Sclerosis Journal* **18**(11), 1577–1584.
- Freedman, M.S., Comi, G., De Stefano, N., Barkhof, F., Polman, C.H., Uitdehaag, B.M.J., Lehr, L., Stubinski, B., Kappos, L. (2014). Moving toward earlier treatment of multiple sclerosis: Findings from a decade of clinical trials and implications for clinical practice. *Multiple Sclerosis and Related Disorders* **3**(2), 147–155.

- Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews. Neurology* **6**(2), 67–77.
- Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.) (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.
- Gabriel, E., Diggle, P.J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica* **63**(1), 43–51.
- Ge, T., Müller-Lenke, N., Bendfeldt, K., Nichols, T.E., Johnson, T.D. (2014). Analysis of multiple sclerosis lesions via spatially varying coefficients. *The Annals of Applied Statistics* **8**(2), 1095–1118.
- Ge, Y. (2006). Multiple sclerosis: The role of MR imaging. *American Journal of Neuroradiology* **27**(6), 1165–1176.
- Gelfand, A., Diggle, P.J., Fuentes, M., Guttorp, P. (2010). *Handbook of Spatial Statistics*. Chapman and Hall/CRC, Boca Raton.
- Gelfand, A.E., Dey, D., Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion), in: *Bernardo, J. (Ed.), Bayesian Statistics*. Oxford University Press. Vol. 4, pp. 147–167.
- Girolami, M., Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **73**(2), 123–214.
- Goldenberg, M.M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics* **37**(3), 175–84.
- Gronwall, D.M.A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills* **44**(2), 367–373.
- Hadwiger, H. (1957). *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*. Springer, Berlin.
- Ho, L.P., Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters* **78**(10), 1194–1199.

- Holland, C.M., Charil, A., Csapo, I., Liptak, Z., Ichise, M., Khoury, S.J., Bakshi, R., Weiner, H.L., Guttmann, C.R.G. (2012). The relationship between normal cerebral perfusion patterns and white matter lesion distribution in 1,249 patients with multiple sclerosis. *Journal of Neuroimaging* **22**(2), 129–136.
- Hsu, C.W., Lin, C.J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* **13**(2), 415–425.
- Huang, M.W., Chen, C.W., Lin, W.C., Ke, S.W., Tsai, C.F., Saunders, C., Jassal, S., Cruz, J., Wishart, D., Gayathri, B., Sumathi, C., Santhanam, T., Kharya, S., Kourou, K., Exarchos, T., Exarchos, K., Karamouzis, M., Fotiadis, D., Shajahan, S., Shanthi, S., Manochitra, V., Shrivastava, S., Sant, A., Aharwal, R., Ahmad, L., Eshlaghy, A., Poorebrahimi, A., Ebrahimi, M., Razavi, A., Huang, C.L., Liao, H.C., Chen, M.C., Senturk, Z., Kara, R., You, H., Rumble, G., Kittler, J., Hatef, M., Duin, R., Matas, J., Lavanya, D., Rani, K., He, H., Garcia, E., Fawcett, T., Vapnik, V., Byun, H., Lee, S.W., Das, S., Das, K., Mishra, D., Shaw, K., Mishra, S., Buck, T., Zhang, B., Mezghani, D., Boujelbene, S., Ellouze, N., Baten, A., Chang, B., Halgamge, S., Li, J., Frosyniotis, D., Stafylopatis, A., Likas, A., Rokach, L., Wozniak, M., Grana, M., Corchado, E., Happel, B., Murre, J., Breiman, L., Kim, H.C., Pang, S., Je, H.M., Kim, D., Bang, S., Schapire, R., Joshi, J., Doshi, R., Patel, J., Powers, D., Oh, I., Lee, J., Moon, B. (2017). SVM and SVM ensembles in breast cancer prediction. *PLoS ONE* **12**(1), e0161501.
- Hurwitz, B. (2009). The diagnosis of multiple sclerosis and the clinical subtypes. *Annals of Indian Academy of Neurology* **12**(4), 226–230.
- Ickstadt, K., Wolpert, R. (1999). Spatial regression for marked point processes. *Bayesian Statistics* **6**, 323–341.
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons, Ltd, Chichester, UK.
- Illian, J.B., Martino, S., Sørbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Esquivias, M.P., Travis, J.M.J. (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution* **4**(4), 305–315.
- Illian, J.B., Møller, J., Waagepetersen, R.P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics* **16**(3), 389–405.

- Illian, J.B., Sørbye, S.H., 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.
- Invanciu, O. (2007). Applications of Support Vector Machines in Chemistry, *in: Lipkowitz, K.B., Cundari, T.R. (Eds.), Reviews in Computational Chemistry*. Wiley-VCH, Weinheim. Vol. 23, pp. 291–400.
- Jaakkola, T.S., Jordan, M.I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing* **10**(1), 25–37.
- Kang, J., Johnson, T.D., Wager, T.D., Nichols, T.E. (2014)a. A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. *Annals of Applied Statistics* **8**(3), 1800–1824.
- Kang, S.Y., McGree, J., Mengersen, K. (2014)b. Bayesian hierarchical models for analysing spatial point-based data at a grid level: a comparison of approaches. *Environmental and Ecological Statistics* **22**(2), 297–327.
- Kappos, L., Antel, J., Comi, G., Montalban, X., O’Connor, P., Polman, C.H., Haas, T., Korn, A.A., Karlsson, G., Radü, E.W. (2006)a. Oral fingolimod (FTY720) for relapsing multiple sclerosis. *New England Journal of Medicine* **355**(11), 1124–1140.
- Kappos, L., Freedman, M.S., Polman, C.H., Edan, G., Hartung, H.P., Miller, D.H., Montalbán, X., Barkhof, F., Radü, E.W., Bauer, L., Dahms, S., Lanius, V., Pohl, C., Sandbrink, R. (2007). Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: A 3-year follow-up analysis of the BENEFIT study. *Lancet* **370**(9585), 389–397.
- Kappos, L., Freedman, M.S., Polman, C.H., Edan, G., Hartung, H.P., Miller, D.H., Montalbán, X., Barkhof, F., Radü, E.W., Metzger, C., Bauer, L., Lanius, V., Sandbrink, R., Pohl, C. (2009). Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. *The Lancet Neurology* **8**(11), 987–997.
- Kappos, L., Polman, C.H., Freedman, M.S., Edan, G., Hartung, H.P., Miller, D.H., Montalban, X., Barkhof, F., Bauer, L., Jakobs, P., Pohl, C., Sandbrink, R. (2006)b. Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. *Neurology* **67**(7), 1242–9.

- Kelly, M.A., Cavan, D.A., Penny, M.A., Mijovic, C.H., Jenkins, D., Morrissey, S., Miller, D.H., Barnett, A.H., Francis, D.A. (1993). The influence of HLA-DR and -DQ alleles on progression to multiple sclerosis following a clinically isolated syndrome. *Human Immunology* **37**(3), 185–191.
- Kincses, Z.T., Ropele, S., Jenkinson, M., Khalil, M., Petrovic, K., Loitfelder, M., Langkammer, C., Aspeck, E., Wallner-Blazek, M., Fuchs, S., Jehna, M., Schmidt, R., Vécsei, L., Fazekas, F., Enzinger, C. (2011). Lesion probability mapping to explain clinical deficits and cognitive performance in multiple sclerosis. *Multiple Sclerosis Journal* **17**(6), 681–689.
- Koutsouleris, N., Borgwardt, S., Meisenzahl, E.M., Bottlender, R., Möller, H.J., Riecher-Rössler, A. (2012). Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: Results from the fepsy study. *Schizophrenia Bulletin* **38**(6), 1234–1246.
- Koutsouleris, N., Patschurek-Kliche, K., Scheuerecker, J., Decker, P., Bottlender, R., Schmitt, G., Rujescu, D., Giegling, I., Gaser, C., Reiser, M., Möller, H.J., Meisenzahl, E.M. (2010). Neuroanatomical correlates of executive dysfunction in the at-risk mental state for psychosis. *Schizophrenia Research* **123**(2-3), 160–174.
- Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* **33**(11), 1444–1452.
- Lang, C., Ohser, J., Hilfer, R. (2001). On the analysis of spatial binary images. *Journal of Microscopy* **203**(3), 303–313.
- Lee, M.H., Smyser, C.D., Shimony, J.S. (2013). Resting-state fMRI: A review of methods and clinical applications. *American Journal of Neuroradiology* **34**(10), 1866–1872.
- Legland, D., Kiêu, K., Devaux, M.F. (2007). Computation of Minkowski measures on 2D and 3D binary images. *Image Analysis and Stereology* **26**(2).
- Lehne, M., Engel, P., Rohrmeier, M., Menninghaus, W., Jacobs, A.M., Koelsch, S. (2015). Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLoS ONE* **10**(5), e0124550.
- Lewis, D.D., Nedellec, C., Rouveirol, C. (1998). Naïve (Bayes) at forty: The independence assumption in information retrieval. *Machine Learning: ECML-98*, 4–15.

- Lewis, P.A.W., Shedler, G.S. (1979). Simulation of nonhomogenous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26**(42), 403–414.
- van Lieshout, M.N.M., Stein, A. (2012). Earthquake modelling at the country level using aggregated spatio-temporal point processes. *Mathematical Geosciences* **44**(3), 309–326.
- Lim, C.Y., Dass, S.C. (2011). Assessing fingerprint individuality using EPIC: A case study in the analysis of spatially dependent marked processes. *Technometrics* **53**(2), 112–124.
- Loevblad, K.O., Anzalone, N., Doerfler, A., Essig, M., Hurwitz, B., Kappos, L., Lee, S.K., Filippi, M. (2010). MR imaging in multiple sclerosis: Review and recommendations for current practice. *American Journal of Neuroradiology* **31**(6), 983–989.
- Lu, J., Yang, H., Zhang, X., He, H., Luo, C., Yao, D. (2015). The brain functional state of music creation: An fMRI study of composers. *Scientific Reports* **5**, 12277.
- Lublin, F.D., Reingold, S.C., Cohen, J.A., Cutter, G.R., Sørensen, P.S., Thompson, A.J., Wolinsky, J.S., Balcer, L.J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P.A., Clanet, M., Comi, G., Fox, R.J., Freedman, M.S., Goodman, A.D., Inglese, M., Kappos, L., Kieseier, B.C., Lincoln, J.A., Lubetzki, C., Miller, A.E., Montalban, X., O'Connor, P.W., Petkau, J., Pozzilli, C., Rudick, R.A., Sormani, M.P., Stüve, O., Waubant, E., Polman, C.H. (2014). Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology* **83**(3), 278–286.
- MacKay Altman, R., Petkau, A.J., Vrecko, D., Smith, A. (2011). A longitudinal model for magnetic resonance imaging lesion count data in multiple sclerosis patients. *Statistics in Medicine* **31**(5), 449–469.
- Mardia, K.V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**(2), 265–284.
- McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, London, New York. 2nd edition.
- McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., Van Den Noort, S., Weinshenker, B.Y., Wolinsky, J.S. (2001). Recommended diagnostic criteria for multiple sclerosis:

- Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology* **50**(1), 121–127.
- Molinaro, A.M., Simon, R., Pfeiffer, R.M. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics* **21**(15), 3301–3307.
- Møller, J., Syversveen, A.R., Waagepetersen, R.P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**(3), 451–482.
- Møller, J., Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Vol. 23. Chapman and Hall CRC, Boca Raton, London, New York.
- Møller, J., Waagepetersen, R.P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics* **34**(4), 643–684.
- Moore, B.D., Lawler, I.R., Wallis, I.R., Beale, C.M., Foley, W.J. (2010). Palatability mapping: A koala’s eye view of spatial variation in habitat quality. *Ecology* **91**(11), 3165–3176.
- Morales, Y., Parisi, J.E., Lucchinetti, C.F. (2006). The pathology of multiple sclerosis: Evidence for heterogeneity. *Advances in Neurology* **98**, 27–45.
- Morgan, C.J., Aban, I.B., Katholi, C.R., Cutter, G.R. (2010). Modeling lesion counts in multiple sclerosis when patients have been selected for baseline activity. *Multiple Sclerosis* **16**(8), 926–934.
- Mostert, J.P., Koch, M.W., Steen, C., Heersema, D.J., De Groot, J.C., De Keyser, J. (2010). T2 lesions and rate of progression of disability in multiple sclerosis. *European Journal of Neurology* **17**(12), 1471–1475.
- Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review. *Journal of Photogrammetry and Remote Sensing* **66**(3), 247–259.
- Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *NeuroImage* **28**(4), 980–995.
- Murkin, J.M., Arango, M. (2009). Near-infrared spectroscopy as an index of brain and tissue oxygenation. *British Journal of Anaesthesia* **103**(Supplement 1), i3–i13.

- Murray, I., Adams, R.P.R., MacKay, D.D.J.C. (2009). Elliptical slice sampling. *Journal of Machine Learning Research: Workshop and Conference Proceeding* **9**(6), 541–548.
- Mwangi, B., Ebmeier, K.P., Matthews, K., Douglas Steele, J. (2012). Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain* **135**(5), 1508–1521.
- Myllymäki, M., Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica* **63**(4), 450–473.
- Neal, R.M. (2012). MCMC using Hamiltonian dynamics, *in: Brooks, S., Gelman, A., Jones, G., Meng, X.L. (Eds.), Handbook of Markov Chain Monte Carlo*. Chapman & Hall / CRC Press.
- Neema, M., Stankiewicz, J., Arora, A., Guss, Z., Bakshi, R. (2007). MRI in multiple sclerosis: What’s inside the toolbox? *Neurotherapeutics* **4**(4), 602–617.
- Niemi, A., Fernández, C. (2010). Bayesian spatial point process modeling of line transect data. *Journal of Agricultural, Biological, and Environmental Statistics* **15**(3), 327–345.
- Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience & Biobehavioral Reviews* **36**(4), 1140–1152.
- Ortigue, S., Bianchi-Demicheli, F., Patel, N., Frum, C., Lewis, J.W. (2010). Neuroimaging of love: fMRI meta-analysis evidence toward new perspectives in sexual medicine. *Journal of Sexual Medicine* **7**(11), 3541–3552.
- Poldrack, R., Mumford, J., Nichols, T.E. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge University Press, New York.
- Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O’Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology* **69**(2), 292–302.

- Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W., Sandberg-Wollheim, M., Thompson, A.J., Weinshenker, B.G., Wolinsky, J.S. (2005). Diagnostic criteria for multiple sclerosis: 2005 Revisions to the "McDonald Criteria". *Annals of Neurology* **58**(6), 840–846.
- Pooley, R.A. (2005). Fundamental physics of MR imaging. *Radiographics* **25**(4), 1087–1099.
- Poser, C.M., Paty, D.W., Scheinberg, L., McDonald, W.I., Davis, F.A., Ebers, G.C., Johnson, K.P., Sibley, W.A., Silberberg, D.H., Tourtellotte, W.W. (1983). New diagnostic criteria for multiple sclerosis: Guidelines for research protocols. *Annals of Neurology* **13**(3), 227–31.
- Rasmussen, C.E., Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- Rathbun, S.L., Cressie, N.a. (1994). A space-time survival point process for a longleaf pine forest in southern Georgia. *Journal of the American Statistical Association* **89**(428), 1164–1174.
- Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution* **6**(4), 366–379.
- Roberts, G.O., Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **60**(1), 255–268.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* **15**(3), 351.
- Rovaris, M., Confavreux, C., Furlan, R., Kappos, L., Comi, G., Filippi, M. (2006). Secondary progressive multiple sclerosis: Current knowledge and future challenges. *The Lancet Neurology* **5**(4), 343–54.
- Rue, H., Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC.
- Sahraian, M.A., Radue, E.W., Haller, S., Kappos, L. (2010). Black holes in multiple sclerosis: Definition, evolution, and clinical correlations. *Acta Neurologica Scandinavica* **122**(1), 1–8.

- Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine* **41**(12), 1110–1117.
- Samartsidis, P., Wager, T.D., Barrett, L.F., Atzil, S., Eickhoff, S.B., Nichols, T.E., Johnson, T.D. (2017). Bayesian log-Gaussian Cox process regression with applications to fMRI meta-analysis. *Eprint arXiv:1701.02643* .
- Sanchez, V. (2003). Advanced support vector machines and kernel methods. *Neurocomputing* **55**(1-2), 5–20.
- Scholkopf, B., Smola, A.J. (2002). *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, Massachusetts.
- Schumacher, G.A., Beebe, G., Kibler, R.F., Kurland, L.T., Kurtzke, J.F., McDowell, F., Nagler, B., Sibley, W.A., Tourtellotte, W.W., Willmon, T.L. (1965). Problems of experimental trials of therapy in multiple sclerosis: Report by the panel on the evaluation of experimental trials of therapy in multiple sclerosis. *Annals of the New York Academy of Sciences* **122**, 552–68.
- Serra, J. (1988). *Image Analysis and Morphology: Theoretical Advances*. Vol. 2. Academic Press, London.
- Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika* **103**(1), 49–70.
- Sormani, M.P., Filippi, M. (2007). Statistical issues related to the use of MRI data in multiple sclerosis. *Journal of Neuroimaging* **17**, 56–59.
- Spanier, E. (1966). *Algebraic Topology*. Springer, New York.
- Stoica, R.S. (2010). Marked point processes for statistical and morphological analysis of astronomical data. *European Physical Journal: Special Topics* **186**(1), 123–165.
- Stoyan, D., Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15**(1), 61–78.
- Taschler, B., Ge, T., Bendfeldt, K., Mueller-Lenke, N., Johnson, T.D., Nichols, T.E. (2014). Spatial modeling of multiple sclerosis for disease subtype prediction. *Lecture Notes in Computer Science* **8674**(2), 797–804.

- Taylor, B.M., Diggle, P.J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation* **84**(10), 2266 – 2284.
- Tintoré, M., Rovira, A., Río, J., Tur, C., Pelayo, R., Nos, C., Téllez, N., Perkal, H., Comabella, M., Sastre-Garriga, J., Montalban, X. (2008). Do oligoclonal bands add information to MRI in first attacks of multiple sclerosis? *Neurology* **70**(13-2), 1079–1083.
- Torkildsen, O., Myhr, K.M., Bø, L. (2016). Disease-modifying treatments for multiple sclerosis: A review of approved medications. *European Journal of Neurology* **23**(S1), 18–27.
- Van Lieshout, M.N.M., Baddeley, A.J. (2002). Extrapolating and interpolating spatial patterns, *in: Lawson, A.B., Denison, D.G.T. (Eds.), Spatial Cluster Modelling*. Chapman & Hall/CRC, Boca Raton, FL, pp. 61–86.
- Vapnik, V. (1979). *Estimation of Dependences Based on Empirical Data*. Vol. 27 of *Nauka*. English translation: Springer Verlag, New York, 1982.
- Vapnik, V.N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks* **10**(5), 988–99.
- Vehtari, A., Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation* **14**(10), 2439–2468.
- Velázquez, E., Martínez, I., Getzin, S., Moloney, K.A., Wiegand, T. (2016). An evaluation of the state of spatial point pattern analysis in ecology. *Ecography* **39**(11), 1042–1055.
- Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics, Springer, New York.
- Weygandt, M., Hackmack, K., Pfüller, C., Bellmann-Strobl, J., Paul, F., Zipp, F., Haynes, J.D. (2011). MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS ONE* **6**(6), e21138.
- Whitaker, J.N., McFarland, H.F., Rudge, P., Reingold, S.C. (1995). Outcomes assessment in multiple sclerosis clinical trials: A critical analysis. *Multiple Sclerosis Journal* **1**(1), 37–47.
- Wolpert, R., Ickstadt, K. (1998)a. Simulation of Lévy random fields, *in: Dey, D., Müller, P., Sinha, D. (Eds.), Practical Nonparametric and Semiparametric*

- Bayesian Statistics*. Springer New York. Vol. 133 of *Lecture Notes in Statistics*, pp. 227–242.
- Wolpert, R.L., Ickstadt, K. (1998)b. Poisson/Gamma random field models for spatial statistics. *Biometrika* **85**(2), 251–267.
- Woodard, D.B., Wolpert, R.L., O’Connell, M.A. (2010). Spatial inference of nitrate concentrations in groundwater. *Journal of Agricultural, Biological, and Environmental Statistics* **15**(2), 209–227.
- Wottschel, V., Alexander, D., Kwok, P., Chard, D., Stromillo, M., De Stefano, N., Thompson, A., Miller, D., Ciccarelli, O. (2015). Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage: Clinical* **7**, 281–287.
- Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* **8**(8), 665–70.