UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 120

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 120

STATISTICAL MODELS AND INFERENCE FOR SPATIAL POINT PATTERNS WITH INTENSITY-DEPENDENT MARKS

MARI MYLLYMÄKI

JYVÄSKYLÄ
2009

UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 120

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 120

STATISTICAL MODELS AND INFERENCE FOR SPATIAL POINT PATTERNS WITH INTENSITY-DEPENDENT MARKS

MARI MYLLYMÄKI

To be presented, with the permission of the Faculty of Mathematics and Science of the University of Jyväskylä, for public criticism in Auditorium Paulaharju, Villa Rana, on November 7th, 2009, at 12 o'clock noon.

JYVÄSKYLÄ
2009

Editor: Pekka Koskela
Department of Mathematics and Statistics
P.O. Box 35 (MaD)
FI–40014 University of Jyväskylä
Finland

ISBN 978-951-39-3683-9
ISSN 1457-8905

University Printing House
Jyväskylä 2009

Abstract

This work belongs to the field of statistics for marked spatial point patterns. Such data contain locations of objects in a 2- or 3-dimensional space and their properties. These elements are called points and marks, respectively. An example is trees in a forest area with the diameter at breast height as a measurement for each tree. Point processes are models for points. Their theory is well-developed and useful model constructions are available. Adding the marks to point pattern data leads to new mathematical and statistical problems.

A practically and theoretically interesting problem is marked point patterns where the statistical properties of marks depend locally on point intensity. Such dependence can be observed e.g. in plant ecology and materials science: the plants can be systematically smaller in areas with high plant density than with low plant density, for example, because of competition of resources, or the size of particles may vary along the density of particles in some material. The present knowledge of intensity-dependent constructions does not cover the needs of applications and, hence, new marked point process models and model fitting techniques are needed.

This work concentrates on statistical models and methods for marked point patterns with intensity-dependent real-valued or qualitative marks. In the first group of models, the mark distribution is defined conditional on the local intensity. The main novelty of the new models is that they allow both the mean and the variance of the mark distribution to depend on the local intensity. These models are used for analysing a tropical rainforest data set. Secondly, a procedure that uses independent mark-dependent thinning is suggested to create models with dependence between the marks and the intensity. Theoretical mark characteristics are derived for both models. Thirdly, a modelling approach is proposed for set-marked data where two different point densities are expected in a set and in its complement and the marks are indicators of the set. These marks are intensity-dependent by construction. This model is used to analyse pine saplings growing in an area which is heterogeneous due to soil treatment. In addition, as a fourth example of intensity-dependently marked point processes, a mathematical characterization of the Bitterlich plot is given. This plot is a point pattern of trees born through sampling which is proportional to the size of a tree, widely used in forestry.

The new models allow detailed modelling of intensity-dependence and are applicable to other applications as well. Importantly, methods for statistical inference are suggested for conditionally defined and set-marked models. The main inferential approach is Bayesian, where Markov chain Monte Carlo (MCMC) methods are used in posterior computation. The methods are demonstrated through simulation studies before applying them to real data.

The work has been conducted at the Department of Mathematics and Statistics, University of Jyväskylä, and as a part of the project "Dependent geostatistical marking of spatial Cox point processes" of the Academy of Finland. The work has connection to an international research project "Spatial Analysis of Tropical Forest Biodiversity" coordinated by the University of York in Great Britain.

Keywords: Bayesian modelling, Bitterlich sampling, density-dependence, Gaussian excursion set, log Gaussian Cox process, mark-dependent thinning, marked point process, MCMC, pine saplings, random set marked Cox process, tropical rainforest.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Antti Penttinen, for introducing me to the field of point process statistics and to other aspects of academic life. I am most indebted to him for his guidance and support during this work.

I have had the privilege to participate in the Working Group on Spatial Analysis of Tropical Forest Biodiversity funded by UKPopNet and the Centre for Population Biology. Discussions with other participants in York stimulated this work greatly. I hereby thank all the members of the Working Group. The rainforest data set used in this thesis origins from the Forest Dynamics Plot of Barro Colorado Island, which is made possible through the generous support of the U.S. National Science Foundation, the John D. and Catherine T. MacArthur Foundation, the Smithsonian Tropical Research Institute and through the hard work of over 100 people from 10 countries over the past two decades. The BCI Forest Dynamics Plot is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

Regarding the sapling data set used in this work, my thanks are due to Dr. Timo Saksa (Finnish Forest Research Institute, Suonenjoki Research Unit). He has pointed out the forest research problem concerning the pine saplings and provided us with the data.

I would like to thank Professor Stefan Geiss (University of Jyväskylä) for advice on Gaussian random processes. I am also very grateful to Professor Dietrich Stoyan (TU Bergakademie Freiberg) for valuable discussions and comments on marked point processes. In addition, Professor Erkki Tomppo (Finnish Forest Research Institute, Vantaa Research Unit) deserves special thanks for his useful remarks on my thesis. Further, my thanks go to the reviewers Dr. Marie-Colette van Lieshout (CWI, Amsterdam) and Professor Joël Chadœuf (INRA, Avignon) for their constructive comments.

This research has been done at the Department of Mathematics and Statistics, University of Jyväskylä. The research has been mainly funded by the Academy of Finland (Project number 111156). I particularly appreciate the project of the Academy of Finland for funding conferences, meetings in York, and visits to universities and research institutes. These experiences have contributed substantially to completion of this thesis. In addition, personal grants have been awarded by Ellen ja Artturi Nyyssönen Foundation. I acknowledge all these sources.

I especially would like to thank all the colleagues at the Department of Mathematics and Statistics for providing a pleasant working environment. I am thankful to all the comments and advice I have got.

My thanks also belong to my parents for their encouragement and support. Moreover, I am very grateful to my brother and friends for being present in my life. Nevertheless, most grateful I am to Juha for making my everyday life so enjoyable.

Jyväskylä, October 2009

Mari Myllymäki

Contents

1 Introduction **1**

1.1 Constructions of intensity-dependent marks 2

1.2 Inference for the intensity-dependent marks 6

1.3 Applications 8

1.4 Articles 9

1.5 Organization of the thesis 9

2 Preliminaries for marked point processes **11**

2.1 Characteristics of marks and points 11

2.2 Point process models 14

3 Example data and scientific problems **17**

3.1 A Norway spruce forest 17

3.2 Data from a tropical rainforest 17

3.3 Pine saplings growing in treated soil 20

4 Conditional intensity-dependent regression marking of log Gaussian Cox processes **26**

4.1 Log-intensity marked Cox process 27

4.2 Intensity-marked Cox process 31

4.3 Geostatistical model for preferential sampling 32

5 Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes **34**

5.1 Gaussian intensity-marked Cox process 35

5.2 Gamma intensity-marked Cox process 37

6 Statistical inference for conditionally marked Cox point processes 42

6.1 Empirical data analysis on marking . . . 43

6.1.1 Kernel estimation of the intensity function . . . 43

6.1.2 Moment-based estimators and conditional likelihood for the marks . . . 44

6.1.3 Application to the log-intensity marked Cox process . . . 45

6.2 Minimum contrast estimation . . . 48

6.2.1 Estimation of log Gaussian Cox processes . . . 49

6.2.2 Minimum contrast estimation applied to the mark variogram . . . 50

6.2.3 A simulation experiment . . . 50

6.3 Bayesian modelling . . . 55

6.4 A simulation study . . . 58

6.5 Modelling the dbh of trees in a tropical rainforest data set . . . 63

6.6 Discussion on intensity-dependent markings . . . 66

7 Independent mark-dependent thinning of geostatistically marked point processes 68

7.1 Mark-dependently thinned process . . . 68

7.2 First- and second-order characteristics for the thinned model . . . 69

7.2.1 General retention probability function . . . 69

7.2.2 An exponential retention probability function . . . 73

7.3 A simulation experiment . . . 74

7.4 Discussion on the thinned process . . . 74

8 Mathematical characterization of Bitterlich forest: what is seen through a relascope 77

8.1 Point-centred Bitterlich point process . . . 79

8.2 Stationary Bitterlich point process . . . 81

8.2.1 First-order characteristics . . . 82

8.2.2 Second-order characteristics . . . 83

8.3 A simulation experiment . . . 85

8.4 Bitterlich sampling applied to a primeval forest . . . 86

8.5 Discussion on the Bitterlich processes . . . 87

9 Bayesian inference for Gaussian excursion set generated Cox processes with set-marking **90**

9.1 Gaussian excursion set generated Cox process with set-marking 91

9.1.1 Gaussian excursion set generated Cox process 91

9.1.2 Set-marking 92

9.2 Bayesian modelling 92

9.2.1 The likelihood for the set-marked process 93

9.2.2 Prior specification 93

9.2.3 Implementation 94

9.2.4 Acceptance probabilities 96

9.3 A simulation study 96

9.4 Modelling pine saplings growing in treated soil 98

9.5 Discussion on Bayesian inference for set-marked data 102

Bibliography **105**

A Derivation of mark characteristics **109**

Chapter 1

Introduction

This work develops models and methods for statistical analysis of marked point patterns where the marks are allowed to depend on the local point density. Besides constructing new models, the emphasis is in the theoretical characteristics of the new models and in Bayesian inference. Algorithms, that use Markov chain Monte Carlo simulation, are developed for the new density-dependently marked models in order to be able to simulate the posterior distribution of the unknowns. The new methods are applied both to simulated and real data. Two applications in forest science have been a driving force behind the work.

Spatial point patterns are collections of objects distributed randomly in a (subset of) multidimensional real space $\mathbb{R}^d$, $d \geq 2$. Point process statistics aims to analyse the random structure of such patterns, see e.g. Cressie (1993), Diggle (2003), Illian et al. (2008), Møller and Waagepetersen (2004), Ripley (1981), Ripley (1988), Stoyan et al. (1995) and Stoyan and Stoyan (1994). Often the points are provided with measured (random) quantities, called marks, that give extra information on the phenomenon. The marks can be very general, but often the observed marks are real-valued or qualitative measurements made at the point locations. Such marks can either be a quality of an object, for example its size, or they can represent some property of the environment such as soil type. These kind of marks are of interest in this work. In the mathematical model, a mark can be considered as an extra coordinate to the point location, but because the marks are observed only at point locations, the role of points and marks is different. Therefore, the marks are separated from points in the statistical model.

There are many possible causalities and spatial structures in marked point patterns. The location information can even be a nuisance the interest being only in phenomena seen through the marks, or the point locations can be of main interest for which the marks provide extra information, for example in the form of the point type. In some applications, the points and marks are closely coupled, in which case their mutual interactions are of interest. An important example is the density-dependence of marks. The theory of point processes is well-developed, but advances in modelling marked point patterns are still needed for understanding those kind of patterns more accurately, see e.g. Illian et al. (2008) and Stoyan et al. (1995).

A marked point process $N_m = \{[x_i; m(x_i)]\}$, where x_i stands for a point location and the mark $m(x_i)$ is a random variable attached to x_i, describes how the marks and points are distributed in a space. The corresponding unmarked point process $N = \{x_i\}$ is often of interest of its own, but in this study, the focus is in marks and in their relationship to points. A marking model may describe how the marks are born given the points or how the marked points have been

generated. The simplest marked point process model is the *independently marked* (or *randomly labelled*) point process (see e.g. Illian et al., 2008). In independent marking, the mark for each point $x_i \in N$ is drawn from a probability distribution independently of the other marks and independently of N. Another simple model for qualitative marks is the *random superposition model*: the point process is assumed to be a superposition of independent subprocesses $N_{\{j\}}$ consisting of points with mark j. The analysis of the marks of a marked point pattern often starts by testing whether the marks can be considered independent of each other or not, see e.g. Illian et al. (2008, p. 460). If the marks are independent, then methods developed for independent data can be used in the statistical analysis of the marks.

More general models are needed, if the marks are spatially correlated. If the marks are independent of the points, they may be described through *geostatistical marking* (Mase, 1996; Schlather et al., 2004; Illian et al., 2008) where the marks are drawn from a random field $\{U(s)\}$ being independent of the point process N: the marks are

$$m(x_i) = U(x_i) \quad \text{for } x_i \in N.$$

These marks are allowed to be correlated, but they are independent of N by construction. The independence of marks from the points can be tested as suggested by Schlather et al. (2004), Guan (2006) and Illian et al. (2008, p. 460). Following Mase (1996), the point process can be interpreted as a sampling design for the random field $\{U(s)\}$ in geostatistical marking: the marks are a representative sample from $\{U(s)\}$. Thus, properties of $\{U(s)\}$ can be deduced from the marks using standard techniques for geostatistical data, see e.g. Chilès and Delfiner (1999), Cressie (1993), Lantuéjoul (2002) and Wackernagel (1998). Likewise, the points can be analysed separately from marks. Therefore, the assumption of the independence between points and marks simplifies the statistical analysis.

The independence of points and marks may not hold in an application under investigation. For example, plants can tend to be small in areas of high plant density because of competition of resources, or they can be large because of good soil properties. Thus, more flexible marked point process models are needed.

1.1 Constructions of intensity-dependent marks

Conditional marking

One of the objectives of this work is to construct flexible marked point process models $N_m = \{[x_i; m(x_i)]\}$. The main construction starts from an unmarked point process $N = \{x_i\}$ and provides each point $x_i \in N$ by a real-valued mark $m(x_i)$. This procedure is called marking. This kind of marking models are conditional: they define the distribution of marks given the points. Modelling the marks conditional on points is natural particularly for data where the birth of the points precedes the marks to be investigated, and the point configuration affects these marks. The focus is in intensity-dependent markings which means that the distribution of marks is affected by the local point density.

A step forward from independent and geostatistical marking is *intensity-dependent* marking suggested by Ho and Stoyan (2008), Menezes (2005) and Myllymäki (2006) for the stationary log Gaussian Cox process. The leading measure of the log Gaussian Cox process is a random intensity $\Lambda(s) = \exp(Z(s))$, $s \in \mathbb{R}^d$, where $\{Z(s)\}$ is a (stationary) Gaussian random field

(Møller et al., 1998; Møller and Waagepetersen, 2004). In these markings, the mean of the conditional mark distribution of $m(x_i)$ given $\Lambda(x_i)$ is a function of $\Lambda(x_i)$ and the marks are conditionally independent given the intensity $\{\Lambda(s)\}$. In particular, the marking suggested by Ho and Stoyan (2008) is

$$m(x_i) = a + b\Lambda(x_i) + \epsilon(x_i) \quad \text{for } x_i \in N, \tag{1.1}$$

where $\epsilon(x_i)$ is a Gaussian random error. The marking suggested in Myllymäki (2006) is similar regression type marking where the intensity-dependence of marks is modelled through $\{Z(s)\}$, that is $\{\log(\Lambda(s))\}$: the marks are

$$m(x_i) = a + bZ(x_i) + \epsilon(x_i) \quad \text{for } x_i \in N. \tag{1.2}$$

The model introduced by Menezes (2005) differs from the above two markings in the sense that it has been developed for preferential sampling in geostatistics: In geostatistics, it is commonly believed that the point process for sample points does not depend on the data process. However, a prior scientific knowledge of the spatial variable of interest or reasons of cost can lead to the gathering of samples in areas that are believed critical, or where the spatial variable of interest obtains its largest values, for example. Menezes (2005) introduces a parametric stochastic model for such situations. The objective is to use the (fitted) model in spatial prediction of a Gaussian random field $\{Z(s)\}$. The sample points of the model are assumed to be generated by the intensity $\Lambda(s) = \exp(c + dZ(s))$ and the marks are a noisy version of $\{Z(s)\}$:

$$m(x_i) = Z(x_i) + \epsilon(x_i) \quad \text{for } x_i \in N. \tag{1.3}$$

This model is close to (1.2), but the interpretation is different.

The markings of Ho and Stoyan (2008), Menezes (2005) and Myllymäki (2006) allow the marks to be correlated and dependent on N. The log Gaussian Cox process as a point process model is a natural choice for two reasons. First, intensity-dependent marking presumes the existence of local variation in the point density, and thus, only clustered or heterogeneous point process models are relevant. Second, the log Gaussian Cox process is a flexible model with nice theoretical properties (Møller et al., 1998; Møller and Waagepetersen, 2004). The suggested markings are useful models that allow the marks to be small (large) in areas of high (low) point density, but they are not able to describe patterns where the variation of marks depends on the local intensity as well: in areas with high point density, in addition to being smaller, the marks may also tend to vary less than in areas with low point density, or vice versa. Consequently, there is a need for more flexible intensity-dependent markings, that allow also conditional heteroscedasticity of marks.

We extend the family of intensity-dependently marked processes by *conditionally heteroscedastic* intensity-dependent markings. Our new markings are for the log Gaussian Cox process as the previous intensity-dependent marking models by Ho and Stoyan (2008) and Menezes (2005). The first two models are heteroscedastic extensions to the model of Ho and Stoyan (2008) whose regression type marking (1.1) can be rewritten as the conditional mark distribution

$$m(x_i)|\,\Lambda(x_i) \sim N(a + b\Lambda(x_i), d^2) \quad \text{for } x_i \in N$$

with constant variance d^2. We further introduce conditional exponential and conditional gamma distribution for marks. Generally, the new conditionally independent markings are of the form

$$m(x_i)|\,\Lambda(x_i) \sim F_{m|\Lambda}(\cdot|\,\Lambda(x_i)) \quad \text{for } x_i \in N, \tag{1.4}$$

where $F_{m|\Lambda}$ is a parametric conditional distribution of the mark $m(x_i)$ at (fixed) $x_i \in N$ given the intensity $\Lambda(x_i)$. In the models specifically discussed in this work, the mean and variance of the conditional mark distribution of $m(x_i)$ given $\Lambda(x_i)$ are defined as a function of $\Lambda(x_i)$. These models allow detailed modelling of intensity-dependence including heteroscedasticity.

The construction which relies on conditional independence of marks given the intensity creates a correlation structure to marks. In this work, theoretical first- and second-order mark characteristics, such as mark correlation and mark variogram function (see e.g. Schlather, 2001a; Illian et al., 2008), are derived for the new conditional markings. The first-order characteristics, mean mark and mark variance, describe the marks irrespective of point locations. The second-order characteristics are functional measures of dependence between the marks, or of dependence between the marks and points. These help in model identification, interpretation, fitting and also in evaluating the goodness-of-fit.

After completing this thesis, we have been informed about a forthcoming paper Diggle et al. (2009).

Thinning

As an alternative to conditional marking, a possibility for creating intensity-dependent marks is to use point process transformations such as thinning (see e.g. Illian et al., 2008). Many point processes can be obtained by applying thinning to a simpler model. For example, if location-dependent $p(x)$-thinning, where a point at x is deleted with probability $1-p(x)$ independently of deletion of the other points (Illian et al., 2008, p. 365), is applied to a homogeneous Poisson process, it results in an inhomogeneous Poisson process. If the thinning function is random and thinning is independent, then the obtained process is a Cox process. Similar strategy is here used to create new marked point processes from a simpler one. We suggest the use of mark-dependent thinning to simpler marked point processes in order to create dependence between the intensity and marks. Our first thinned model applies independent mark-dependent thinning to a geostatistically marked point process. Interestingly, the two simple procedures, geostatistical marking and independent thinning, when applied sequentially to a base point process, are able to create dependence between the marks and points. Theoretical formulas are here derived for the first- and second-order characteristics of the thinned model with a general mark-dependent thinning function.

Thinning procedures can also be dependent such as in Matérn inhibition point processes (Matérn, 1960, 1986; Illian et al., 2008). Dependent thinnings may be more realistic for some applications, but independent thinnings are useful because of their simplicity and because they may allow analytical calculations of process characteristics to be tractable. Therefore, independent mark-dependent thinning is regarded as a reasonable starting point for creating new marked point process models through thinning.

Bitterlich sampling

In some practical problems, thinning may be specifically defined. Such a forestry-specific problem is Bitterlich sampling, which is a widely used technique in practical forestry. In Bitterlich sampling, an investigator selects trees to a sample plot (called Bitterlich plot or angle count plot) using a simple optical instrument called relascope. The relascope allows the observer, standing at a sample point, to deduce whether an angle subtended by a tree is larger than 2α, where α is an angle fixed in advance. This selection is based on a simple geometric rule: if the distance from the observer to a tree is smaller than $m/(2\sin\alpha)$, where m is the diameter of

the tree at breast height (dbh), then the tree belongs to the Bitterlich plot (see Figure 8.1). The popularity of the Bitterlich sampling is due to the fact that, assuming the cross-sections of the trees to be circular, the number of trees in a Bitterlich plot multiplied by $\sin^2 \alpha$ is an unbiased estimator of the basal area per unit area of forest, see e.g. Ripley (1981) and Illian et al. (2008). Most of the theoretical work concerning Bitterlich sampling studies the properties of this estimator such as its efficiency compared to plot sampling, see e.g. Holgate (1967) and Penttinen (1988). In this work, the Bitterlich sampling is viewed as mark-dependent thinning of a (stationary) marked point process $N_{b,m}$: A forest is described by $N_{b,m} = \{[y_j; m(y_j)]\}$, where y_j stands for a location of a tree and $m(y_j)$ for its dbh. The Bitterlich plot is considered to be a realisation of a marked point process that is obtained by applying the relascope thinning to $N_{b,m}$ at a random location. We call this new marked point process point-centred Bitterlich point process. Obviously, the relascope thinning is mark-dependent, because a large tree of $N_{b,m}$ is retained at a larger distance from the sample point than a small tree. One reason, why this thinning is interesting from the point of view marked point process models, is that the obtained process is an example of a centred marked point process defined relative to the sample point. Further, in the created process the marks are intensity-dependent. This can be seen from the theoretical characteristics that are derived for the point-centred Bitterlich point process. The contribution of our approach to the Bitterlich sampling is that it provides mathematical characterization of the Bitterlich plot: it describes what the Bitterlich plot looks like.

Applying relascope thinning to $N_{b,m}$ at a set of random locations allows us to construct a stationary marked point process, which we call (stationary) Bitterlich point process. In this constuction, a point $[y_j; m_j]$ of $N_{b,m}$ is retained if its distance at least from one of the sample points is less than $m_j/(2 \sin \alpha)$. This process is stationary if the sample points are distributed according to a stationary point process, but there is variation in the local intensity function and mark distribution defined relative to the sample points. We consider the case where the sample points are distributed according to a homogeneous Poisson process. This corresponds to uniform random sampling. The resulting model is a further example of a marked point process where the marks depend locally on the intensity.

Set-marking

Besides the log Gaussian Cox process, another Cox point process is employed in this work. It is the random set generated Cox process with random intensity

$$\Lambda(s) = \lambda_1 \mathbf{1}(s \in \Theta) + \lambda_2 \mathbf{1}(s \notin \Theta), \quad s \in \mathbb{R}^d,$$

where Θ is a random closed set in $\mathbb{R}^d$ and $\mathbf{1}(\cdot)$ stands for the indicator function. That is, the random intensity of this process is defined by Θ such that different point intensities, λ_1 and λ_2, appear in the two phases formed by Θ and its complement, Θ^c. The point pattern is observed as a set of point coordinates in a bounded region $W \subset \mathbb{R}^d$. For this particular model, it is useful to observe additionally the information on the phase of the location of each point. In practice, this kind of information may often be economical to collect. A marked point process, called random set marked Cox process, is obtained with so-called set-marks (Penttinen and Niemi, 2007): the mark $m(x_i)$ obtains the value 1 if $x_i \in \Theta$, and $m(x_i) = 2$ if $x_i \in \Theta^c$. The interest in the random set marked Cox process lies in the two intensities, λ_1 and λ_2, and in the properties of Θ. If $\lambda_1 \neq \lambda_2$, the marks are not a representative sample from the random set, and hence they cannot be directly used for deducing properties of the random set. These marks are an example of intensity-dependent qualitative marks.

Penttinen and Niemi (2007) mentions the germ-grain model (see e.g. Stoyan et al., 1995) as a possible model for the random set, but their modelling approach does not demand specification of the random set model. We suggest the use of Gaussian excursion sets as a parametric model for the random set. More precisely, a stationary Gaussian random field $\{Z(s) : s \in \mathbb{R}^d\}$ with continuous sample paths is generated, and the random closed set is defined as

$$\Theta = \{s \in \mathbb{R}^d : Z(s) \geq l\}, \tag{1.5}$$

where l is a (given) threshold level, see e.g. Lantuéjoul (2002, p. 205). The mean and covariance function of $\{Z(s)\}$ together with the threshold level l define properties of the random set. The resulting set-marked model is called *Gaussian excursion set generated Cox process with set-marking*. This model is supposed to be a flexible and parsimonious model to be applied to set-marked data. The flexibility of the Gaussian excursion set is supposed to be an advantage over the germ-grain model.

Also the set-marked process can be obtained through independent mark-dependent thinning: Assume $\lambda_1 > \lambda_2$ without loss of generality, and let N_1 be a homogeneous Poisson process in $\mathbb{R}^d$ with intensity λ_1. A realisation of the random set marked Cox process is a result of independent thinning applied to N_1 such that the thinning probability is 0 if $m(x) = 1$, and λ_2/λ_1 if $m(x) = 2$. Moreover, the set-marks could be specified conditional on the intensity, but this characterization is not especially important for this model. It is more reasonable to think that the random set defines both the intensity and the marks.

1.2 Inference for the intensity-dependent marks

The use of intensity-dependently marked point processes in data analysis is a recent topic. Model fitting has not been considered in the seminal paper Ho and Stoyan (2008) and previous statistical modelling of marked point patterns concentrates on the analysis of points and marks separately. Some suggestions for model fitting for intensity-dependently marked Cox processes are made by Ho (2006) and Menezes (2005). Ho (2006) fits the log Gaussian Cox process with marking (1.1) or (1.3) to a data set using least squares method applied to second-order characteristics of points and marks, and Menezes (2005) considers estimation allowing so-called *length-bias*, see Menezes (2005, p. 182). A further example of modelling intensity-dependent marks is Penttinen and Niemi (2007) where an estimation method is suggested for random set marked Cox processes. This method is based on second-order characteristics of point processes.

These existing model fitting techniques are useful at an early stage of data analysis. However, the minimum contrast estimation or least squares method, although being a widely applicable and computationally easy approach for parameter estimation (see e.g. Diggle, 1979; Diggle, 2003, p. 86; Illian et al., 2008, p. 451), is criticized because it is based on some user-specified choices, see e.g. Cressie (1993, p. 666) and Guan (2006). Moreover, as regards the random set marked Cox process, the method of Penttinen and Niemi (2007) can be sensitive to the estimation of the Ripley's K-function. It would also be preferable to obtain some idea about the uncertainties of the parameter estimates, which these methods do not yield as such; an application of parametric bootstrap would help. In this work, we suggest the use of Bayesian methods for statistical inference for the intensity-dependently marked log Gaussian Cox processes and for the random set marked Cox process. A great advantage of the Bayesian method is that it

allows to estimate the parameters of a model simultaneously. It also provides uncertainties for the parameters in terms of the posterior interval or standard deviation of the marginal posterior distributions. In addition, the latent structures of the models are naturally handled in the hierarchical Bayesian setting, and the information in data is utilized more efficiently than in those estimation methods based only on second-order characteristics. An overview on the use of Bayesian computing in point process statistics can be found in Møller and Waagepetersen (2004).

For the intensity-dependently marked log Gaussian Cox processes, we propose an empirical Bayesian approach where the Markov chain Monte Carlo (MCMC) method is used in the simulation of the posterior distribution of the intensity and the parameters of the marking equation (1.4). In this hierarchical Bayesian model, the intensity and marking form the top level; at the bottom level is the (parametric) model for the intensity. The priors for the marking parameters are set according to the marking model, and according to our model choice, the prior distribution of the intensity is log Gaussian. The model parameters of this intensity are hyperparameters of the intensity-dependently marked Cox process model, and they are to be estimated from the point data in our modelling approach. Assuming the log Gaussian Cox process model, those hyperparameters are the mean and parameters of the chosen covariance function of the Gaussian random field $\{Z(s)\}$ that generates the intensity through $\Lambda(s) = \exp(Z(s))$. For estimation of the mean and covariance parameters, one can employ the minimum contrast method (Møller et al., 1998; Møller and Waagepetersen, 2004; Illian et al., 2008; Guan and Sherman, 2007), approximative likelihood method (Tanaka et al., 2008; see also Illian et al., 2008), composite likelihood method (Guan, 2006) or Bayesian method (Møller and Waagepetersen, 2004).

The parameters of the prior distribution of the intensity affect the values of the intensity and the mark parameters, of course, but since the intensity itself contains a lot of variation, the effect of fixing those hyperpriors is supposed to be quite small. The estimation of the hyperparameters from the point data allows stable estimation of the intensity and marking parameters.

In the MCMC simulation, the conditional likelihood of the intensity-dependently marked log Gaussian Cox process given $\{Z(s)\}$ (or $\{\Lambda(s)\}$) is employed. This likelihood depends on the whole process $\{Z(s)\}$, not only on the process at the points $x_i \in N$ where the conditional mark distribution (1.4) is defined. The points x_i depend on $\{Z(s)\}$, and thus, $\{Z(x_i) : x_i \in N\}$ is not a representative sample from $\{Z(s)\}$ and cannot be used for deducing properties of $\{Z(s)\}$ directly. Therefore, it is not enough to have $\{Z(x_i) : x_i \in N\}$ in the set of unknowns in the MCMC simulation: In the implementation, the Gaussian random field $\{Z(s)\}$ is approximated by a discrete Gaussian random field defined on a partition of the observation window W into a disjoint rectangular sets of equal size. The values of $\{Z(s)\}$ are simulated both at the centre points of those sets, being independent of $\{Z(s)\}$, and at the points $x_i \in N$. Therefore, the algorithm produces also a prediction for the intensity surface in W as a posterior mean of the simulated values of the intensity $\Lambda(s) = \exp(Z(s))$ at the centre points which form a grid in W. For this prediction the mark information is utilized, too, which is an extension to the method presented by Møller et al. (1998) for intensity estimation. In our MCMC simulation, Metropolis-Hastings algorithm is utilized in updating variables.

For our second case, Gaussian excursion set generated Cox processes with set-marking, we adopt a fully Bayesian method for inferring the parameters of the model and estimating the random set. The Gaussian excursion set is chosen as a parametric model for the random set since it is supposed to be a flexible model, suitable to be used as a prior model in the Bayesian approach.

Thus, also this model includes a latent Gaussian random field $\{Z(s)\}$. Here it generates the Gaussian excursion set, our random set, see (1.5). Gaussian excursion sets have also earlier been used as a hidden structure in spatial statistics in De Oliveira (2000) where prediction of the random set from sampled points is considered. The sampled points are independent of the random closed set (or Gaussian random field) and thus, the obtained marked point process $\{[x_i; m(x_i)]\}$ is geostatistically marked in the terminology of Schlather et al. (2004), see also Mase (1996). The situation in the random set marked Cox process is different. The observations are points of the process and hence not independent of the underlying random closed set; the marking is intensity-dependent. This is an additional complexity in model fitting, and hence, the Bayesian modelling considered here is a contribution to the statistical methodology for marked point patterns.

As in the inference for the intensity-dependently marked log Gaussian Cox processes, the Gaussian random field must be discretized in implementation. The random field $\{Z(s)\}$ is considered at the points x_i of the random set generated Cox process and at $s_1, \ldots, s_k$ forming a grid $\mathcal{G}$ covering the observation window W. The values of $\{Z(s)\}$ on $\mathcal{G}$ are again needed in the approximation of the conditional likelihood given $\{Z(s)\}$, which also in the case of this marked point process depends on the whole process $\{Z(s)\}$. Therefore, the unknowns of the set-marked model to be simulated in the MCMC algorithm are $Z = (Z(x_1), \ldots, Z(x_n), Z(s_1), \ldots, Z(s_k))'$, where n is the number of observed points, the intensities λ_1 and λ_2, and the parameters of $\{Z(s)\}$ (or Z). Because of confounding of parameters, the threshold level is fixed at $l = 0$.

The stationary Gaussian random field $\{Z(s)\}$ is completely characterized by its mean μ_Z and covariance function $C_Z(r)$. As previous studies discuss, see De Oliveira (2000) and Diggle et al. (1998), it is probably not possible to deduce the form of the covariance function from the point pattern data. Therefore, a suitable parametrized covariance function is chosen. We use the Matérn family of covariance functions. The chosen covariance function defines the covariance matrix of the prior distribution of Z, which is a $(n + k)$-dimensional normal distribution (with mean μ_Z). We suggest prior distributions for the other parameters and hyperparameters, and use Metropolis-Hastings steps then to update the variables in the MCMC algorithm. The algorithm as such is slow mainly since the updating of a parameter of the covariance function $C_Z(r)$ demands an inversion of the covariance matrix of Z. Discretization of the covariance parameter is proposed, which leads to great improvement in computation time. The Bayesian estimation results in posterior distribution for all the unknowns. A prediction for the random set can be obtained from the simulated values of Z at $\mathcal{G}$.

Bayesian algorithms are suggested for statistical reasoning on two Cox processes with intensity-dependent marks or set-marks. For both models quite extensive simulation experiments are performed in order to demonstrate the models and the functionality of the posterior computation, and further, the methods are applied to real data.

1.3 Applications

This work is motivated by two data sets. In the first one, the locations of trees of the species *Trichilia tuberculata* are observed in an area of a tropical rainforest together with the measured diameters at breast height, see Figure 3.6. In this data set, the trees tend to be small in areas with high tree density, whilst in areas with low density both large and small trees are abundant. The conditionally heteroscedastic intensity-dependent markings are developed keeping

this application in mind, and the markings are used for analysing this data set. The modelling gives information on the interactions between the conspecific trees, which is one aspect of the forest structure.

The second data consist of locations of pine saplings growing in a forest area that has been recently clear-cut, see Figure 3.8. In this area, before planting and natural seeding, the soil has been mounded removing mechanically the vegetation and a part of turf in patches. The treatment has formed a blotchy soil structure. The density of saplings is expected to be higher in treated patches than in the area outside patches because of better germination of naturally distributed seeds or because of cultivation. The Gaussian excursion set generated Cox process is a candidate for a statistical model for the saplings: It is reasonable to assume that the treated area is random, and thus it may be modelled by a random set. Because early-stage competition of trees is often negligible, it is reasonable to model the sapling locations in the treated and untreated areas by two Poisson processes with different intensities. Therefore, the random set generated Cox process is supposed to be especially suitable model for these pine saplings. Further, in addition to the locations of saplings, the soil type, treated or not treated, is observed at the location of each sapling. This information is economically collected in the field inventory. Data on pine saplings are modelled in this work using the Gaussian excursion set generated Cox process model with set-marking. The modelling allows to evaluate the forest regeneration intensities and it also gives other information on the data.

1.4 Articles

Two articles origin from this thesis. Chapter 9 is based on the article

> Myllymäki, M. and Penttinen, A. (2009). Bayesian inference for Gaussian excursion set generated Cox processes with set-marking. Accepted for publication in Statistics & Computing. DOI: 10.1007/s11222-009-9123-1.

Parts of the text of this article and Figures 3.8, 9.4, 9.5, 9.8 and 9.10 are used here with kind permission from Springer Science + Business Media. The conditionally heteroscedastic marking models and the Bayesian inference for these models will be published in

> Myllymäki, M. and Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. Accepted for publication in Statistica Neerlandica.

Articles from the works of Chapters 7 and 8 are under construction.

1.5 Organization of the thesis

The thesis is arranged as follows. Preliminaries of marked point processes and mark characteristics are given in Chapter 2. Cox processes as well as independent and geostatistical marking are also recalled. Example data are explained and illustrated in Chapter 3. In Chapter 4, the regression type intensity-dependent marking strategies are discussed. New conditionally heteroscedastic markings are introduced in Chapter 5, and in Chapter 6, statistical inference is considered for conditionally marked Cox processes. Two marking strategies based on thinning

are presented in Chapters 7 and 8. Chapter 9 is for Bayesian inference for Gaussian excursion set generated Cox processes with set-marking. The models, their inference and possible extensions are discussed in the corresponding chapters.

Chapter 2

Preliminaries for marked point processes

2.1 Characteristics of marks and points

Spatial marked point processes $N_m = \{[x_i; m(x_i)]\}$ are models for collections of randomly distributed points in (a subset W of) $\mathbb{R}^d$ provided with measured quantities called marks. The corresponding (unmarked) point process is denoted by $N = \{x_i\}$ and it describes the locations of objects, whereas the mark $m(x_i)$, denoted by m_i for short, is a random variable attached to the point x_i of N. The marks can be very general: the marks are assumed to be elements of a so-called Polish space, see e.g. Stoyan et al. (1995) and Daley and Vere-Jones (2003). Often observed marks are real-valued or qualitative measurements, which kind of marks are considered in this work.

In this work, we assume that N is a stationary and isotropic simple locally finite point process. Stationarity and isotropy mean that $N + s = \{x_i + s\}$ and $\mathbf{r}N = \{\mathbf{r}x_i\}$ have the same distribution as N for any $s \in \mathbb{R}^d$ and for any (Euclidean) rotation $\mathbf{r}$ around the origin, respectively. Simplicity means that multiple points are not allowed, and local finiteness that each bounded subset of $\mathbb{R}^d$ contains only a finite number of points of N. Further, $N_m = \{[x_i; m(x_i)]\}$ is assumed to be a stationary and isotropic simple locally finite marked point process meaning that the translated process $\{[x_i + s; m(x_i)]\}$ and the rotated process $\{[\mathbf{r}x_i; m(x_i)]\}$ have the same distribution as N_m. Note that the marks are kept untouched in the translation and rotation. Simplicity and local finiteness also concern the points.

One type of construction of marked point processes takes the distribution of points and then, conditional on points, specifies the marks. The strategy that transforms an unmarked point process N to a marked point process N_m by providing each point x_i of N with a random variable $m(x_i)$ is called marking. Two well-known examples are *independent marking* and *geostatistical marking*. In independent marking, the mark for each point x_i is drawn from a probability distribution independently of other marks and N. This model is often used as a reference model. In geostatistical marking (Mase, 1996; Schlather et al., 2004; Illian et al., 2008), the marks are drawn from a (stationary) random field $\{U(s)\}$ which is independent of N:

$$m(x_i) = U(x_i) \quad \text{for } x_i \in N.$$

This marking is able to generate correlated marks, but the marks are independent of N by construction.

There are various first- and second-order characteristics that describe the properties of marked point processes, see e.g. Stoyan and Stoyan (1994), Stoyan et al. (1995), Schlather (2001a), Schlather et al. (2004) and Illian et al. (2008). Their empirical counterparts are applied in empirical data analysis, model identification, model fitting, evaluation of goodness-of-fit and in model interpretation. The mark characteristics have a special nature since the marks are only observed at $x_i \in N$. They are conditional quantities (in the Palm sense, see e.g. Stoyan et al., 1995): Let $\mathbb{E}_x$ and var_x stand for the conditional expectation and variance, respectively, given that there is a point of N at the location x. Further let $\mathbb{E}_{xy}$ refer to the conditional expectation given there are two points of N at the locations x and y. Because of stationarity and isotropy, it suffices to consider expectations $\mathbb{E}_o$ and $\mathbb{E}_{o\mathbf{r}}$ with $\|\mathbf{r}\| = r$.

The first-order characteristics are the intensity λ, which tells the mean number of points of N per unit volume, and the mark distribution that describes the marks irrespective of point locations. Often the mean and variance of quantitative marks are of interest. Definitions are recalled in the following for real-valued marks: For any Borel sets B in $\mathbb{R}^d$ and L in $\mathbb{R}$, let $N_m(B \times L)$ stand for the number of points in B with mark in L. The corresponding mean number of points satisfies

$$\mathbb{E}[N_m(B \times L)] = \lambda \nu_d(B) M(L), \tag{2.1}$$

where ν_d denotes the d-dimensional Lebesgue measure and M is called mark distribution, see e.g. Stoyan et al. (1995). For real-valued marks, M is described by the mark distribution function F_M, $F_M(m) = M((-\infty, m])$ for $-\infty < m < \infty$, and assuming a mark probability density function $f_m(m)$ exist, it holds

$$F_M(m) = \int_{-\infty}^{m} f_m(m) \mathrm{d}m. \tag{2.2}$$

The mean mark is

$$\mu_m = \mathbb{E}_o(m(o)) = \int_{-\infty}^{\infty} m M(\mathrm{d}m) = \int_{-\infty}^{\infty} m f_m(m) \mathrm{d}m, \tag{2.3}$$

and the mark variance

$$\sigma_m^2 = \mathrm{var}_o(m(o)) = \int_{-\infty}^{\infty} (m - \mu_m)^2 M(\mathrm{d}m) = \int_{-\infty}^{\infty} (m - \mu_m)^2 f_m(m) \mathrm{d}m.$$

The second-order characteristics can be classified to those concerning only points and to mark characteristics. The second-order characteristics for points that is used here is the pair-correlation, which is defined through

$$g(r) = \frac{\rho^{(2)}(r)}{\lambda^2} \quad \text{for } r > 0, \tag{2.4}$$

where $\rho^{(2)}(r)$ is the second-order product density of a stationary and isotropic process. For a stationary and isotropic process the product density $\rho^{(2)}(o, \mathbf{r})$ depends only on $r = \|o - \mathbf{r}\|$, and it is written $\rho^{(2)}(r)$. In other words, $\rho^{(2)}$ is the density of the factorial moment measure $\alpha^{(2)}$: Assume that $\alpha^{(2)}$ is locally finite and absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^{2d}$. Then for Borel sets B_1 and B_2 in $\mathbb{R}^d$

$$\alpha^{(2)}(B_1 \times B_2) = \mathbb{E}\left(\sum_{x_1, x_2 \in N}^{\neq} \mathbf{1}_{B_1}(x_1) \mathbf{1}_{B_2}(x_2) \right) = \int_{B_1} \int_{B_2} \rho^{(2)}(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2, \tag{2.5}$$

where $\sum^{\neq}$ denotes that the summation is only over distinct points $x_1, x_2 \in N$, see e.g. Stoyan et al. (1995), Stoyan and Stoyan (1994) or Illian et al. (2008). Intuitively, $\rho^{(2)}(o, \mathbf{r})\,\mathrm{d}o\,\mathrm{d}\mathbf{r}$ is the probability that two infinitesimal disjoint regions of volumes $\mathrm{d}o$ and $\mathrm{d}\mathbf{r}$ both contain exactly one point of N.

In this work the following second-order mark characteristics, summarized in Table 2.1, are exploited. For comparison, the characteristics for an independently and geostatistically marked point process are recalled. In the latter marking, the marks are determined by a random field $\{U(s)\}$ with mean μ_U, variance σ_U^2, covariance function $C_U(r)$ and variogram $\gamma_U(r)$.

Table 2.1: Mark characteristics.

mark characteristic	definition for $r > 0$	independent marking	geostatistical marking
Stoyan's κ_{mm}:	$\kappa_{mm}(r) = \mathbb{E}_{o\mathbf{r}}\{m(o)m(\mathbf{r})\}$	μ_m^2	$\mu_U^2 + C_U(r)$
Stoyan's k_{mm}:	$k_{mm}(r) = \kappa_{mm}(r)/\mu_m^2$	1	$1 + C_U(r)/\mu_U^2$
$E(r)$-function:	$E(r) = \mathbb{E}_{o\mathbf{r}}\{m(o)\}$	μ_m	μ_U
$V(r)$-function:	$V(r) = \mathbb{E}_{o\mathbf{r}}\{[m(o) - E(r)]^2\}$	σ_m^2	σ_U^2
mark covariance function:	$C_m(r) = \kappa_{mm}(r) - (E(r))^2$	0	$C_U(r)$
mark correlation function:	$\rho_m(r) = [\kappa_{mm}(r) - (E(r))^2]/V(r)$	0	$C_U(r)/\sigma_U^2$
mark variogram:	$\gamma_m(r) = \frac{1}{2}\mathbb{E}_{o\mathbf{r}}\{[m(o) - m(\mathbf{r})]^2\}$	σ_m^2	$\gamma_U(r)$

The characteristics of Table 2.1 are of the form $\kappa_f(r) = \mathbb{E}_{o\mathbf{r}}\{f(m(o), m(\mathbf{r}))\}$, $\|\mathbf{r}\| = r > 0$, with or without scaling, see e.g. Illian et al. (2008) or Schlather (2001a). The expectations are with respect to the two-point mark distribution $M_{x_1,x_2}(\mathrm{d}m_1, \mathrm{d}m_2)$, which is the joint distribution of the marks $m_1 = m(x_1)$ and $m_2 = m(x_2)$ on condition (in the Palm sense) that there exist points of N at x_1 and x_2. Mathematical arguments for the existence of this distribution can be found in Stoyan et al. (1995, p. 114). Because the process is assumed to be stationary and isotropic, the mark distribution M_{x_1,x_2} depends on the points x_1 and x_2 only through their distance $r = \|x_1 - x_2\|$, and it can be written M_r. More technically,

$$\kappa_f(r) = \frac{\rho_f^{(2)}(r)}{\rho^{(2)}(r)}, \quad \text{for } r > 0,\ \rho^{(2)}(r) > 0, \tag{2.6}$$

where $\rho_f^{(2)}(r)$ is the density (assuming it exists) of the factorial moment measure $\alpha_f^{(2)}$ defined by

$$\alpha_f^{(2)}(B_1 \times B_2) = \mathbb{E}\left(\sum_{[x_1;\,m_1],\,[x_2;\,m_2] \in N_m}^{\neq} f(m_1, m_2)\mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_n)\right). \tag{2.7}$$

The function $\kappa_f(r)$ is often scaled by dividing it by a factor

$$c_f = \int\int f(m_1, m_2)M(\mathrm{d}m_1)M(\mathrm{d}m_2), \tag{2.8}$$

which leads to the f-mark correlation function

$$k_f(r) = \frac{\kappa_f(r)}{c_f} \quad \text{for } r > 0. \tag{2.9}$$

The functions $k_{mm}(r)$ and $\gamma_m(r)$ are functional measures of dependence (or association) between marks, and further, the functions $C_m(r)$ and $\rho_m(r)$ correspond to the classical definitions of covariance and correlation, respectively. Instead, $E(r)$ and $V(r)$ are characteristics for dependence between marks and points, see Schlather (2001a) and Schlather et al. (2004). They are the conditional expectation and variance of a mark given that there is a further point at the distance r apart. Note that for a geostatistically marked point process these quantities are constant. This can be used to test the hypothesis of geostatistical marking, see Schlather et al. (2004). Guan (2006) gives further tools for studying independence of marks and points, see also Illian et al. (2008). Further, note that the mark variogram $\gamma_m(r)$ coincides with the geostatistical variogram only in specific situations such as independent and geostatistical marking, which is considered by Wälder and Stoyan (1996) and Stoyan and Wälder (2000), see also Illian et al. (2008). Schlather (2001a) extends the characteristics of Table 2.1 to include the case $r = 0$ by defining $k_{mm}(0) = \mathbb{E}_o[(m(o))^2]/\mu_m^2$, $E(0) = \mu_m$, $V(0) = \sigma_m^2$ and $\gamma_m(0) = 0$.

The empirical counterparts of the characteristics of Table 2.1 are called mark summaries. Let $W \subset \mathbb{R}^d$ be an observation window. An estimator of $\kappa_f(r)$ is (Schlather, 2001a)

$$\hat{\kappa}_f(r) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{x\in N} f(m(x), m(x)), & \text{if } r = 0, \\ \dfrac{\sum^{\neq}_{[x;m(x)],[y;m(y)]\in N_m} f(m(x), m(y))\mathbf{1}_W(x)\mathbf{1}_W(y)\mathbf{1}_{N_r}(x,y)/\nu_d(W_x \cap W_y)}{\sum^{\neq}_{x,y\in N} \mathbf{1}_W(x)\mathbf{1}_W(y)\mathbf{1}_{N_r}(x,y)/\nu_d(W_x \cap W_y)}, & \text{if } r > 0. \end{cases} \tag{2.10}$$

where $\delta > 0$ is a suitably small number, $W_x = W + x = \{s + x : s \in W\}$ and

$$N_r = \{(x_i, x_j) : r - \delta < \|x_i - x_j\| \leq r + \delta, i, j = 1., \ldots, n, i \neq j\}.$$

That is, N_r consists of all pairs of points being the distance $r \pm \delta$ apart. In (2.10) $\mathbf{1}_{N_r}(x, y)$ can be replaced by a kernel function $K(r - \|x - y\|)$. The estimator (2.10) is used in this work.

For independent and geostatistical marking, the mark characteristics of Table 2.1 do not depend on the characteristics of N. Instead, for the intensity-dependently marked models considered in this work, the points and marks are closely coupled, and therefore, the mark characteristics also depend on the distributional properties of points.

2.2 Point process models

Many of the marking models presented in the following chapters apply the class of Cox processes as point process models. It is a widely used point process class for aggregated or clustered point patterns, see e.g. Møller and Waagepetersen (2004) and Illian et al. (2008). A Cox process is defined by a random intensity function $\{\Lambda(s) : s \in \mathbb{R}^d\}$: given a realisation $\Lambda(s) = \lambda(s)$, $s \in \mathbb{R}^d$, the points follow an inhomogeneous Poisson process having $\lambda(s)$ as the intensity

function. Stationarity implies that $\lambda = \mathrm{E}(\Lambda(s))$ is constant, but the point intensity in its realisations is defined by $\Lambda(s) = \lambda(s)$.

In this work, two different Cox processes are applied. These are the log Gaussian Cox process (Møller et al., 1998; Møller and Waagepetersen, 2004) and the random set generated Cox process (Penttinen and Niemi, 2007; Illian et al., 2008). Particularly, the new approach of this work for the random set generated Cox process suggests that the random set is a Gaussian excursion set generated by a stationary Gaussian random field. Therefore, both of the Cox process models employ a real-valued (stationary) Gaussian random field.

It is assumed that $\{Z(s) : s \in \mathbb{R}^d\}$ is a real-valued Gaussian random field which is characterized by the mean function

$$\mu_Z(s) = \mathbb{E}[Z(s)],$$

$s \in \mathbb{R}^d$, and the (positive definite) covariance function

$$C_Z(s_1, s_2) = \mathbb{E}[Z(s_1) - \mu_Z(s_1)][Z(s_2) - \mu_Z(s_2)],$$

$s_1, s_2 \in \mathbb{R}^d$. Its finite-dimensional distributions $\mathbf{P}(Z(s_1) \in F_1, \ldots, Z(s_n) \in F_n)$ are Gaussian for all $n \geq 1$, all $s_1, \ldots, s_n \in \mathbb{R}^d$ and all Borel sets $F_1, \ldots, F_n \subseteq \mathbb{R}^d$, see e.g. Adler (1981). A Gaussian random field is stationary if and only if $\mu_Z(s) \equiv \mu_Z$ and $C_Z(s_1, s_2)$ depends only on $s_1 - s_2$ for all s, s_1 and $s_2 \in \mathbb{R}^d$. Then it is written $C_Z(s_1, s_2) = C_Z(s_1 - s_2)$. Moreover, the random field $\{Z(s)\}$ is stationary and isotropic if, in addition, $C_Z(s_1 - s_2)$ depends on $r = \|s_1 - s_2\|$ only. In what follows, the isotropic covariance function is denoted by $C_Z(r)$. The mean of the stationary Gaussian random field $\{Z(s)\}$ is denoted by μ_Z, and the variance by σ_Z^2. It is assumed $\mathbb{E}|Z(s)|^2 < \infty$ for all $s \in \mathbb{R}^d$, see e.g. Adler (1981, p. 23).

Log Gaussian Cox process

Most of the marking strategies of this work are for the log Gaussian Cox process, which is generated by a real-valued Gaussian random field $\{Z(s) : s \in \mathbb{R}^d\}$ through

$$\Lambda(s) = \exp(Z(s)) \quad \text{for } s \in \mathbb{R}^d,$$

see Møller et al. (1998), Møller and Waagepetersen (2004) and Illian et al. (2008). The log Gaussian Cox process is completely characterized by $\{Z(s)\}$, which is here assumed to be a stationary and isotropic Gaussian random field with mean μ_Z, variance σ_Z^2 and (valid) covariance function $C_Z(r)$. Assuming the stationarity and isotropy, the intensity of the log Gaussian Cox process is

$$\lambda = \exp\{\mu_Z + \sigma_Z^2/2\} \tag{2.11}$$

and the pair-correlation

$$g(r) = \exp\{C_Z(r)\} \quad \text{for } r > 0.$$

These formulas appear in the mark characteristics of intensity-dependently marked log Gaussian Cox processes.

Random set generated Cox process

A stationary and isotropic random closed set in $\mathbb{R}^d$ is a stochastic process whose realisations are closed subsets of $\mathbb{R}^d$, see e.g. Stoyan et al. (1995) and Lantuéjoul (2002). The random set divides the space $\mathbb{R}^d$ into two phases, Θ and Θ^c, the complement of Θ. The random set generated Cox process is, conditional on Θ, a union of two independent Poisson processes, one

in Θ and another in Θ^c with intensities $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, respectively. It is a Cox process with the random intensity function

$$\Lambda(s) = \lambda_1 \mathbf{1}(s \in \Theta) + \lambda_2 \mathbf{1}(s \notin \Theta), \quad s \in \mathbb{R}^d, \tag{2.12}$$

where $\mathbf{1}(\cdot)$ stands for the indicator function. All properties of the random set generated Cox process are inherited from the random set, the intensities λ_1 and λ_2 and conditionally independent and uniform scattering of points. If $\lambda_1 = \lambda_2 = \lambda$, then the Poisson process with intensity λ is obtained. Hence of interest is in general the case $\lambda_1 \neq \lambda_2$. The process is introduced in Penttinen and Niemi (2007) and a special case (with $\lambda_2 = 0$), called interrupted point process, in Stoyan (1979).

The random set generated Cox process is considered in Chapter 9 with a Gaussian excursion set as a model for the random set, and with set-marking. The Gaussian excursion set is supposed to be a flexible and parsimonious model for the random set: all the distributional properties of the set are inherited from a Gaussian random field and a fixed threshold level. The set-marks are defined by

$$m(x_i) = 2 - \mathbf{1}(x_i \in \Theta).$$

This marking leads to a marked point process $N = \{x_i; m(x_i)\}$ called random set marked Cox process by Penttinen and Niemi (2007).

Chapter 3

Example data and scientific problems

3.1 A Norway spruce forest

A data set from a Norway spruce (sometimes called Norwegian spruce) forest in a natural forest stand in Saxonia, Germany, is shown in Figure 3.1. The data set is available in R library `spatstat` (Baddeley and Turner, 2005), see also e.g. Illian et al. (2008, p. 405). The data set consist of locations of trees in a 56 m $\times$ 38 m -plot and diameters of the trees at breast height. The diameter at breast height (dbh) is the diameter of a tree at the height of 1.3 m above the ground level.

Mark summaries (see Table 2.1) calculated from the marked point pattern of the spruce trees are shown in Figure 3.2. These summaries do not give clear evidence against independent (or geostatistical) marking. Further tests can be performed for testing independent or geostatistical marking, see Illian et al. (2008, p. 460), who conclude that the marking may be considered independent.

The Norway spruce forest data set is shown here as an example of data where independent (or geostatistical) marking may be a valid marking strategy. When the marks and points can be considered independent, they can be analysed separately using standard techniques for point patterns and for independent or geostatistical data, see e.g. Chilès and Delfiner (1999), Lantuéjoul (2002) and Cressie (1993). The independence of marks and points makes the analysis easier, but it may not hold in practice as the examples of Sections 3.2 and 3.3 show.

3.2 Data from a tropical rainforest

Figure 3.3 presents a marked point pattern of the species *Trichilia tuberculata* in a tropical rainforest. The data origins from a 50 ha Forest Dynamics Plot in 1990 at Barro Colorado Island, Panama, see Hubbell and Foster (1983), Condit et al. (1996) and Condit (1998). The rainforest plot contains tens of species and thousands of trees. All trees with dbh $\geq$ 10 mm have been measured. Precisely, there are 316 species and 244059 trees in the plot measured in 1990. One has to start with a single species analysis, which gives important information on one aspect of the forest structure. In the data set of Figure 3.3, the points are locations of trees and the marks are diameters at breast height. This pattern consisting only of conspecific trees

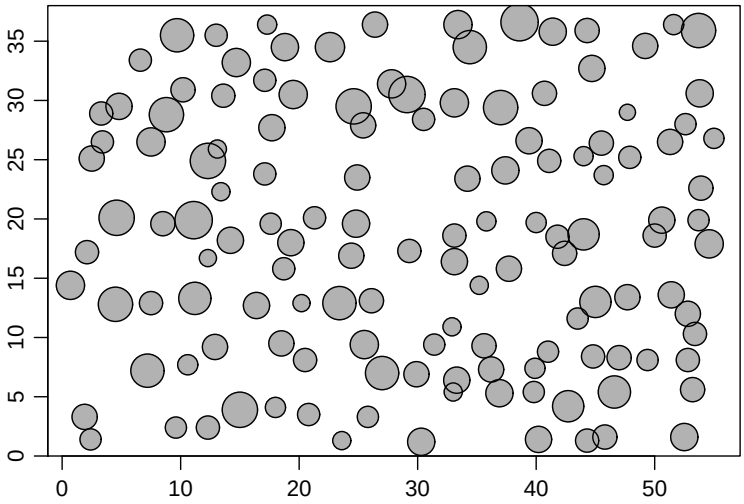

Figure 3.1: A Norway spruce forest. Each point corresponds to a location of a tree and the diameter of a circle is proportional to the diameter of the tree at breast height.

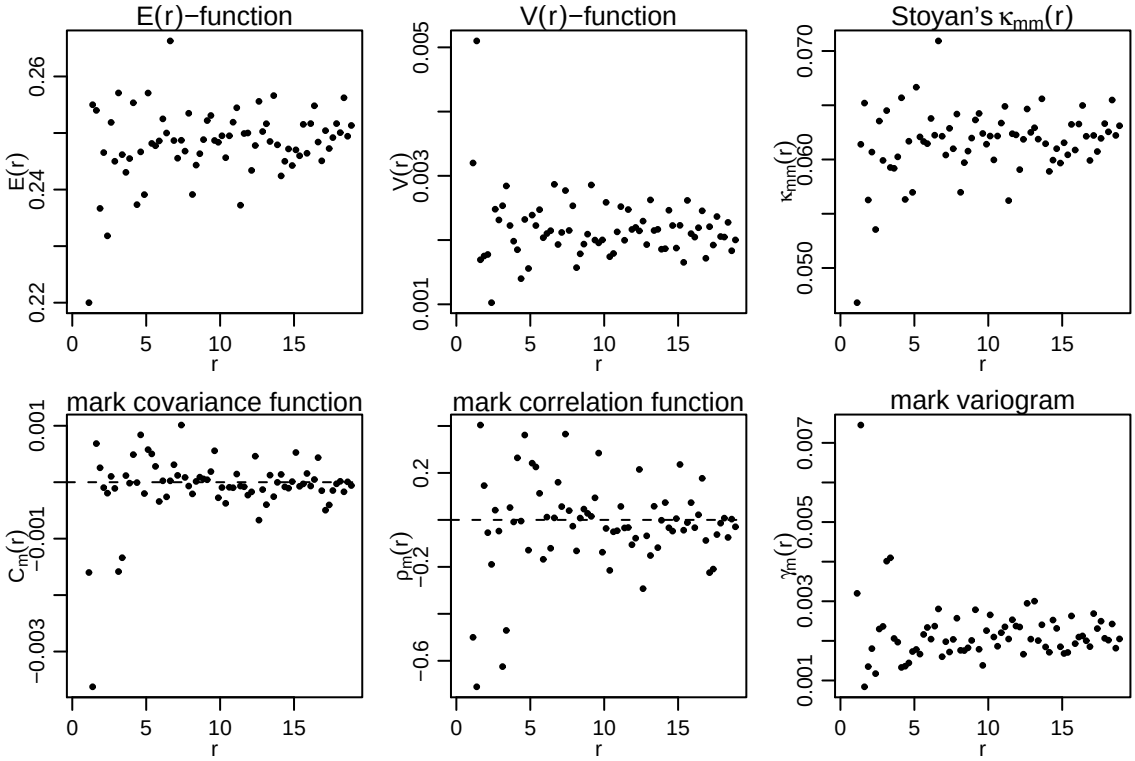

Figure 3.2: The mark summaries (*dots*) calculated from the marked point pattern of Norway spruce trees of Figure 3.1.

is extensive as its own and can not be modelled by means of stationary processes. There is variation in the intensity of points over the region, and elevation and slope maps are available, too, see Figure 3.5.

For analysing the data set one can proceed by investigating large-scale or small-scale properties.

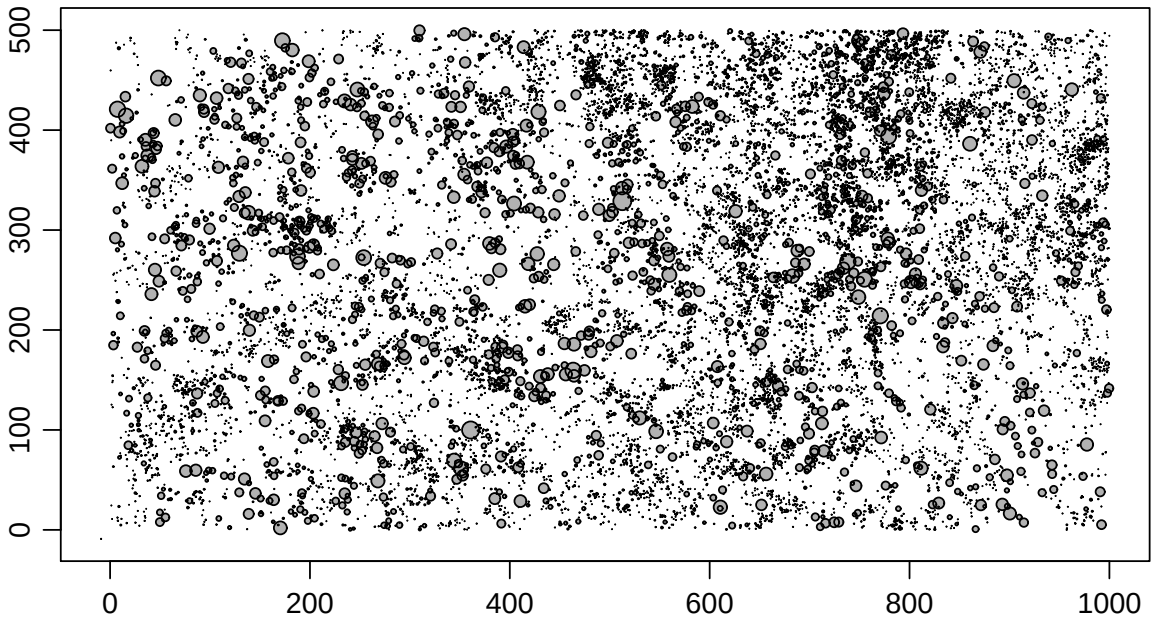

Figure 3.3: A marked point pattern of *T. tuberculata*. The diameter of a circle is proportional to the diameter of the tree at breast height.

If the interest is in the large-scale behaviour, then only long-range point or mark fluctuations and their dependencies on covariates are relevant. The small-scale variation, including dependencies between (individual) points and marks, can be studied by marked point process methods applied to approximately homogeneous sampled parts of the data. The short-range interactions of points might also be studied by means of summary characteristics for inhomogeneous patterns, see e.g. Illian et al. (2008, p. 279) and references there.

The large scale variation of points and marks may be illustrated by the means of regionalised fields. The point-sum measure

$$I_R(s) = \sum_{[x_i;\, m(x_i)] \in N_m} \mathbf{1}_{b(s,R)}(x_i), \quad s \in W \subseteq \mathbb{R}^d,$$

calculates the number of trees in a circle with fixed radius R. If the point-sum measure is scaled by the the volume of the intersection of $b(s, R)$ and the observation window W, it is a simple estimator of the intensity of the point pattern. The mark-sum measure (Stoyan, 1984)

$$S_R(s) = \sum_{[x_i;\, m(x_i)] \in N_m} m(x_i)\mathbf{1}_{b(s,R)}(x_i), \quad s \in W \subseteq \mathbb{R}^d,$$

counts the sum of marks in a circle $b(s, R)$. The mark-sum measure may also be scaled by the volume of the intersection of $b(s, R)$ and W, but different scaling may be applied as well: the so-called normalized mark-sum measure is

$$S_R^*(s) = \frac{S_R(s)}{I_R(s)} \quad \text{for } I_R(s) > 0, \quad s \in W \subseteq \mathbb{R}^d.$$

It describes the spatial variation of marks.

These regionalised fields have been calculated for the *T. tuberculata* data of Figure 3.3 with $R = 50$ m using the edge correction, where the measures are divided by the area of the intersection of $b(s, R)$ and the observation window W, see Figure 3.4. The point-sum measure indicates heavy concentration of trees on the top right of the image, which is the area with the highest elevation. The mark-sum measure is more uniformly distributed over the area, but still there is a noticeable concentration in the region with high elevation. The normalized mark-sum field calculated for *T. tuberculata* indicates that the large trees are on the top left and the bottom right parts of the study area. These are regions where the tree density is quite low. This could be considered as a kind of density-dependence of marks at large scale.

The large-scale properties and dependence on covariates may further be studied by geostatistical methods or one can also proceed in terms of inhomogeneous spatial point processes (ignoring the marks), see e.g. Waagepetersen (2008) and Waagepetersen and Guan (2007). In order to study the short-range interaction of points and marks, a number of (nearly) homogeneous subplots may be chosen and methods for stationary patterns may then be applied to these. Here a subplot is chosen such that it is relatively homogeneous with respect to elevation and slope, see Figure 3.5. In this subregion, $[200, 400] \times [100, 300]$, the elevation varies between 140 m and 144 m (above the sea level) and the slope is small. The marked point pattern is shown in Figure 3.6. In addition, the dbh-marks are plotted against the estimated local tree density at the tree locations, which reveals an interesting feature in the data: there is dependence between the tree density and dbh. Further, $E(r)$- and $V(r)$-functions estimated from the marked point pattern are shown in Figure 3.7 together with the histogram of marks. Since the summaries are definitely not constants in r, independent or geostatistical marking is not an option here.

This work aims at constructing stationary marked point process models that are able to explain the small-scale behaviour of the data set of Figure 3.6. Indeed, the data set is modelled by the new intensity-dependent marking models in Section 6.5. Similar modelling could be applied to other subregions of the *T. tuberculata* data and the obtained estimates considered in relation to covariates.

3.3 Pine saplings growing in treated soil

The two marked point patterns of pine saplings, shown in Figure 3.8, origin from a study conducted by the Finnish Forest Research Institute, Suonenjoki Research Unit. The saplings grow in a forest area where, after clear-cutting, the soil surface has been treated mechanically by removing the vegetation and part of turf in patches. This treatment, that has been done before planting and natural seeding, has created a blotchy soil structure that divides the soil into two parts: treated and untreated. The coordinates of each sapling have been measured and, in addition, for each sapling it is known whether its location is in treated soil or untreated. In other words, the observations consist of the sapling locations x_i and set-marks $m(x_i)$ where the set is defined by the treated patches. The marks can be coded such that $m(x_i)$ obtains the value 1 if the sapling at x_i is in treated soil, and 2 if x_i belongs to untreated soil.

The interest is in the effect of the pre-regeneration soil treatment to the sapling density. The density is expected to be higher in patches because the treated soil is subject to planting, and because it is supposed to provide better conditions for germination of naturally distributed

seeds and for growth of saplings. The origin of saplings, planted or naturally born, is omitted here: the total intensities, in treated and untreated parts, are of interest. If the area fraction p of treated soil in an observation window W was known, the intensities could be estimated by $n_1/(p|W|)$ and $n_2/((1-p)|W|)$, where $|W|$ is the area of the observation window W, n_1 is the number of saplings with $m(x_i) = 1$ and n_2 the number of saplings with $m(x_i) = 2$. However, only information on the soil structure is at the sapling locations x_i. Since the observations $m(x_i)$ depend on the soil structure, properties of the soil structure cannot be deduced directly from these observations. This is due to expecting two different intensities, one in treated and another in untreated part of the soil. Consequently, more advanced methods are needed for analysis of these data sets.

The data are modelled in Chapter 9 by the means of Gaussian excursion set generated Cox processes. The fully Bayesian analysis gives estimates (with uncertainties) for the intensities and also describes the treated area in the form of the area fraction and covariance properties. In addition, a prediction for the treated area is given.

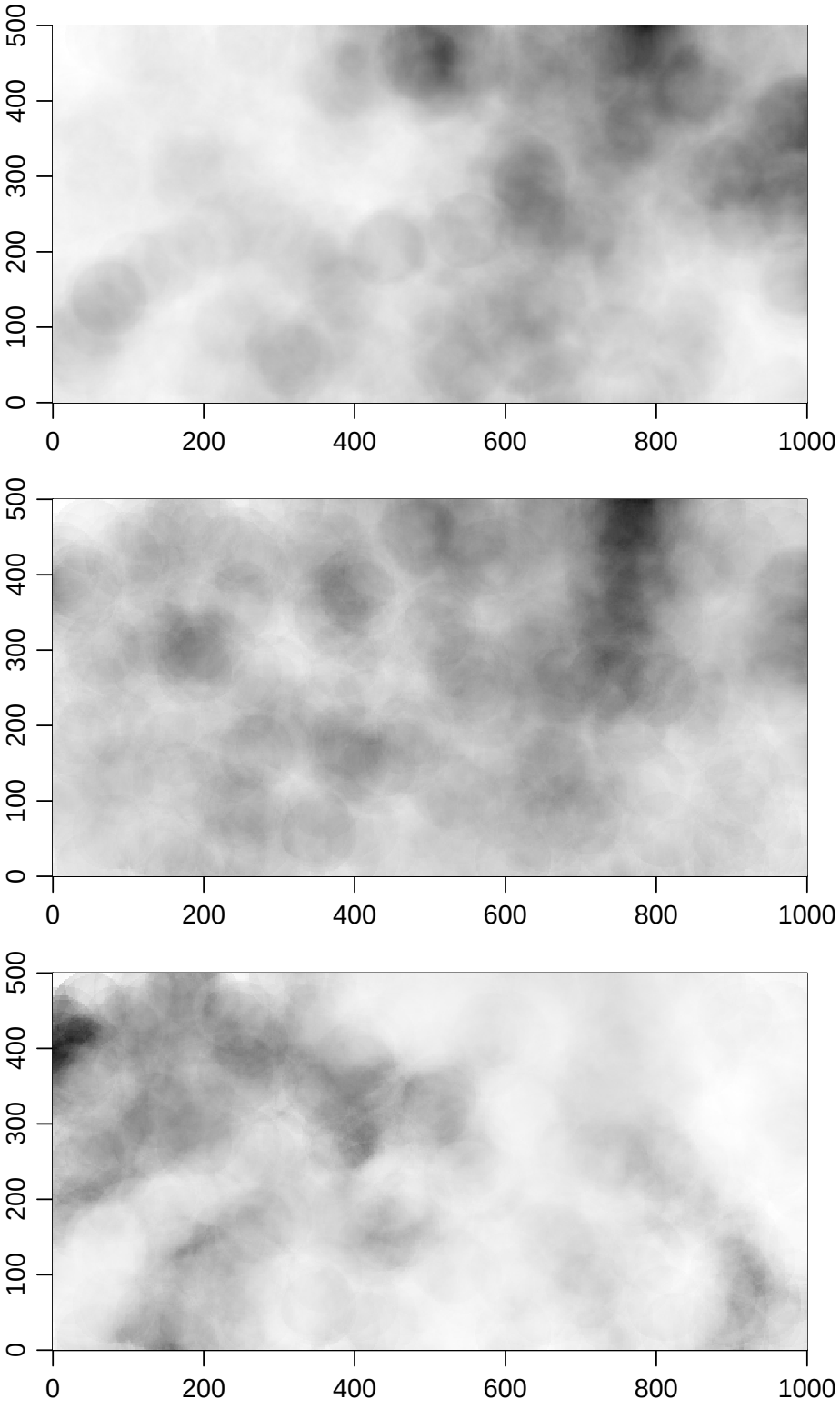

Figure 3.4: The point-sum measure $\{I_{50}(s)\}$ (*at the top*), mark-sum measure $\{S_{50}(s)\}$ (*in the middle*) and normalized mark-sum measure $\{S^*_{50}(s)\}$ (*at the bottom*) for the *T. tuberculata* data set of Figure 3.3. Dark colour corresponds to a high value of a random field.

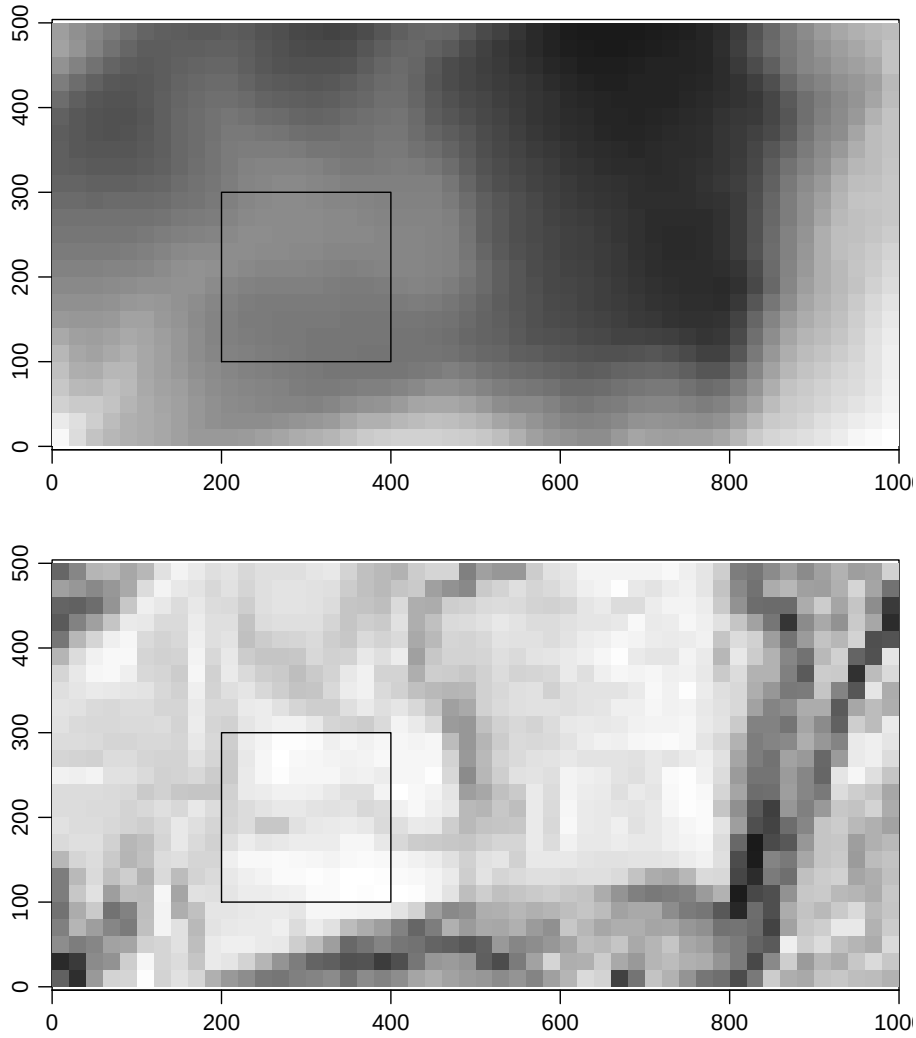

Figure 3.5: Elevation (*upper*) and slope (*lower*) maps with the box indicating the location of the plot of Figure 3.6. Dark color corresponds to a high value of elevation/slope.

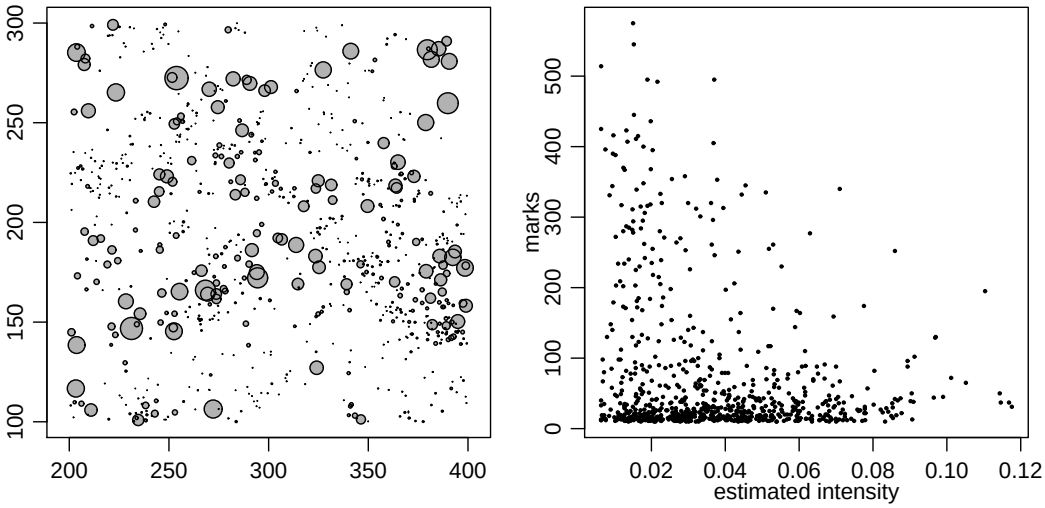

Figure 3.6: *On the left*: The marked point pattern of *T. tuberculata* in the subregion $[200, 400] \times [100, 300]$. The diameter of a circle is proportional to the diameter of the tree at breast height. *On the right*: The diameters of trees (in mm) plotted agains the estimated tree densities at the tree locations.

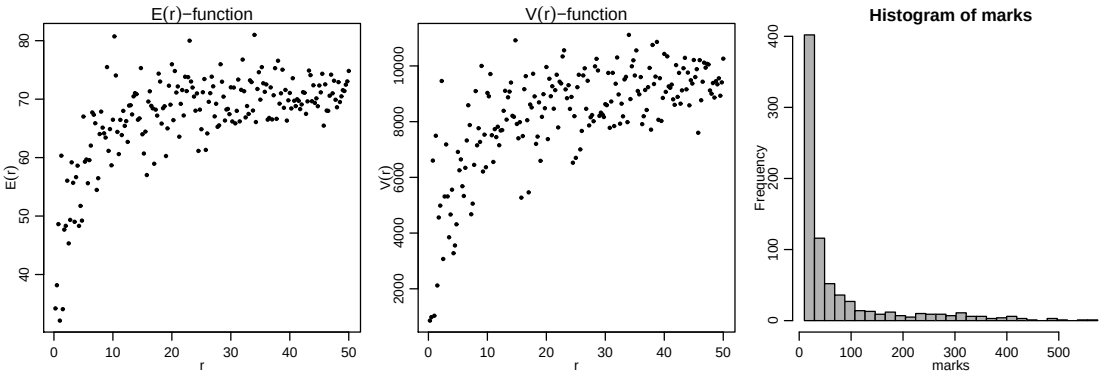

Figure 3.7: The $E(r)$ and $V(r)$ summaries calculated from the marked point pattern of *T. tuberculata* of Figure 3.6 and the histogram of the marks (diameters of trees at breast height in mm).

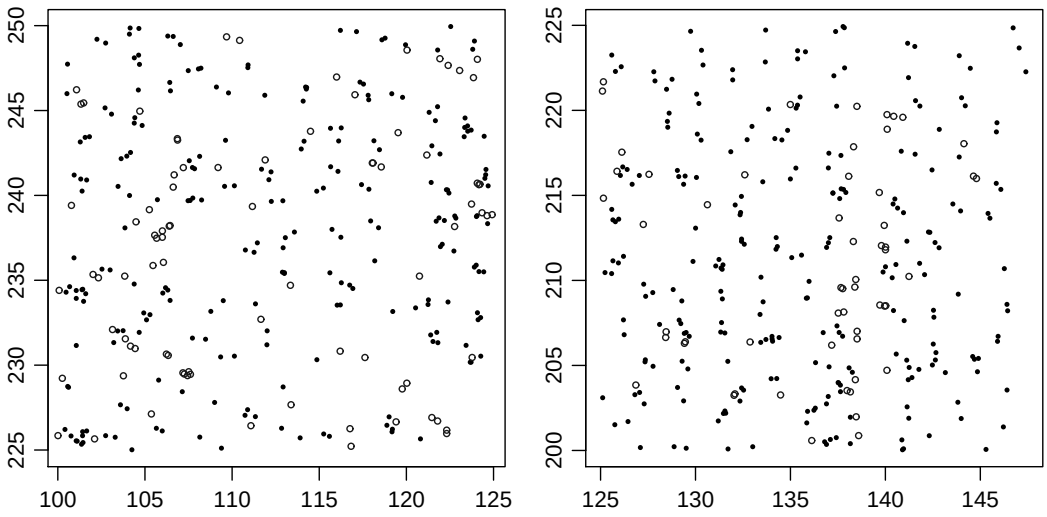

Figure 3.8: The first (*on the left*) and the second (*on the right*) sapling data set. Black point means that the sapling is in a treated patch, white point that the sapling is not in a patch.

Chapter 4

Conditional intensity-dependent regression marking of log Gaussian Cox processes

In the simplest marking strategy, independent marking, which is often used as a reference model, and in more general geostatistical marking (Mase, 1996; Schlather et al., 2004; Illian et al., 2008) the point process N does not affect the distributional properties of the marks. An important situation where these markings are not valid strategies is the case where the point density affects the marks, called density-dependence in plant ecology. For example, in a forest, trees may be systematically small in regions of high tree density because of competition of nutrients and light, or systematically large because of good soil properties.

A step forward is *intensity-dependent* marking suggested by Ho and Stoyan (2008), Menezes (2005) and Myllymäki (2006) for the stationary log Gaussian Cox process generated by a random intensity $\{\Lambda(s)\}$. In these markings, the mean of the conditional mark distribution of $m(x_i)$ given $\Lambda(x_i)$ is a function of $\Lambda(x_i)$ and the marks are conditionally independent given the intensity $\{\Lambda(s)\}$. In this construction, the marks are allowed to be marginally correlated and the marking depends on N. The log Gaussian Cox process as a point process model is a natural choice for two reasons. First, intensity-dependent marking presumes the existence of local variation in the point intensity, and thus, only clustered or heterogeneous point process models are relevant. Second, the log Gaussian Cox process is a flexible model with nice theoretical properties, see Møller et al. (1998) and Møller and Waagepetersen (2004).

In the markings by Ho and Stoyan (2008), Menezes (2005) and Myllymäki (2006), the correlation structure of marks is inherited from the Gaussian random field that generates the Cox process. Here an extension is presented to the model considered already in Myllymäki (2006). The extension model uses a further (independent) Gaussian random field to model the additional correlation of marks that is not due to the intensity. This construction allows the range of correlation of marks to differ from the range of correlation of points. It could be applied also for the other models. Unfortunately, in statistical modelling, it is probably difficult to separate the effect of the intensity and of the external Gaussian random field.

The extended model and its properties are presented in Section 4.1. Thereafter, in Sections 4.2 and 4.3, the other two models are recalled and their properties are given based on Ho and Stoyan (2008) and Menezes (2005).

4.1 Log-intensity marked Cox process

Let $N = \{x_i\}$ be a log Gaussian Cox process in $W \subseteq \mathbb{R}^d$ (see Section 2.2). The marks of the *log-intensity marked Cox process* are constructed by

$$m(x_i) = \alpha + \beta Z(x_i) + U(x_i) \quad \text{for } x_i \in N, \tag{4.1}$$

where α and β are model parameters and $U(x_i)$ is a Gaussian random error with mean 0 and variance σ_U^2 being independent of $\{Z(s)\}$. This construction is conditional on $\{Z(x_i) : x_i \in N\}$.

An extension is obtained if $U(x_i)$s origin from $\{U(s)\}$, which is a zero-mean motion-invariant Gaussian random field independent of $\{Z(s)\}$. The random field $\{U(s)\}$ is allowed to be correlated. Its covariance function is $C_U(r)$ and the variance σ_U^2. An interesting feature of this extension is that it combines two sources of variation into the marking: the spatial correlation structure of the marks is inherited from the random fields $\{Z(s)\}$ and $\{U(s)\}$. The intensity-dependence of marks is modelled through $\{Z(s)\}$, and $\{U(s)\}$ is a geostatistical component. Heuristically, if the marks are explained by the intensity-generated process ($\beta \neq 0$), then the residuals $U(x_i) = m(x_i) - \alpha - \beta Z(x_i)$ may still contain autocorrelation. This is modelled using an external (and independent) random field. This random field $\{U(s)\}$ can be interpreted as the influence of unobserved variables affecting the marks but not the intensity. In the following, the extension, where $U(x_i)s$ in (4.1) origin from $\{U(s)\}$, is considered.

Note that if $\beta = 0$, then (4.1) leads to a geostatistically marked point process $\{[x_i; U(x_i)]\}$. If $\beta < 0$, then the marks are small in regions of high point density, while positive β yields large marks in regions with high intensity.

The formal calculation of mean of $\alpha + \beta Z(x_i) + U(x_i)$ would yield $\alpha + \beta\mu_Z$, but this is not the true mean mark while the marks are only observed at $x_i \in N$ which locations depend on $\{Z(s)\}$. The mean mark is

$$\mu_m = \alpha + \beta\mu_Z + \beta\sigma_Z^2 \,.$$

In fact, for this marking it is possible to derive the whole mark distribution: the (Palm) distribution of a mark $m(x_i)$ on condition $x_i \in N$ is the normal distribution

$$m(x_i) \sim N(\alpha + \beta\mu_Z + \beta\sigma_Z^2, \beta^2\sigma_Z^2 + \sigma_U^2) \quad \text{for } x_i \in N. \tag{4.2}$$

The proof of (4.2) can be found in Appendix A. Therefore, the mark variance is $\sigma_m^2 = \beta^2\sigma_Z^2 + \sigma_U^2$. That is, the dependence of marks on N affects the mean mark but not the mark variance.

The n-point mark distribution of the log-intensity marked Cox process is also analytically tractable for $n \geq 2$. The most important distribution is probably the two-point mark distribution. Recall that it is the distribution of marks $m(x_1)$ and $m(x_2)$ on condition that there are points of N at x_1 and x_2 and these are a distance r apart. Some of the second-order characteristics, that are derived below, can be obtained directly from the two-point mark distribution. Similarly n-point mark distribution of n marks $m(x_1), \ldots, m(x_n)$ is conditional on n points $x_1, \ldots, x_n$ being particular distances $\|x_i - x_j\|$, $i, j = 1, \ldots, n$, apart. For example, the three-point mark distribution is conditional on the triangular configuration of the locations of three points of N. The n-point distribution is here given in the general form for n marks. The two-point mark distribution is obtained as a particular case with $n = 2$.

Let $\{V(s)\}$ stand for a random field that is obtained by $V(s) = \alpha + \beta Z(s) + U(s)$ for $s \in W$. Its mean is

$$\mu_V = \alpha + \beta\mu_Z$$

and covariance function

$$C_V(r) = \beta^2 C_Z(r) + C_U(r)\,.$$

The (Palm) distribution of n marks $m(x) := (m(x_1), \ldots, m(x_n))'$ on condition $x_1, \ldots, x_n \in N$, called n-point mark distribution, is the n-dimensional normal distribution

$$m(x) \sim N_n(\mathbf{1}\mu_V + \beta \mathrm{diag}(\Sigma_{Z(x)})\mathbf{1}, \Sigma_{V(x)}), \tag{4.3}$$

where $\mathbf{1} = (1, 1, \ldots, 1)'$ is $n \times 1$ -vector, $\mathrm{diag}(\Sigma_{Z(x)})$ is the diagonal matrix of the variances of $(Z(x_1), \ldots, Z(x_n))'$ and $\Sigma_{V(x)}$ the $n \times n$ -covariance matrix of $(V(x_1), \ldots, V(x_n))'$. The proof of (4.3) is given in Appendix A. Note that the expected value of each mark $m(x_i)$, $x_i \in N$, is $\mu_V + \beta\sigma_Z^2$, but the values of marks at points $x_1, \ldots, x_n$ are also affected by the point configuration and the correlation structure determined by $C_V(r)$. Recall that the covariance matrix of $V(x_1), \ldots, V(x_n)$ (or $Z(x_1), \ldots, Z(x_n)$) is determined by the covariance function of the corresponding random field and the distances $\|x_i - x_j\|$, $i, j = 1, \ldots, n$: the element (i, j) of the covariance matrix $\Sigma_{V(x)}$, where i denotes the row and j the column of the matrix, is equal to $C_V(\|x_i - x_j\|)$ for all $i, j = 1, \ldots, n$.

The characteristics $E(r)$ and $V(r)$, suggested by Schlather et al. (2004) to test geostatistical marking, can also be determined for the log-intensity marked Cox process. They are

$$E(r) = \alpha + \beta(\mu_Z + \sigma_Z^2 + C_Z(r)) = \mu_m + \beta C_Z(r) \tag{4.4}$$

and

$$V(r) = C_V(0) = \beta^2\sigma_Z^2 + \sigma_U^2, \tag{4.5}$$

see Appendix A. Thus, $E(r)$ depends on the distance r, but $V(r)$ does not. Recall, that these are characteristics of a mark given that there is a further point of the process at the distance r apart. The result (4.3) gives that the covariance between $m(x_1)$ and $m(x_2)$ at points x_1 and x_2 being a distance r apart is $C_V(r)$. Thus, Stoyan's $\kappa_{mm}(r)$-function is

$$\begin{aligned}
\kappa_{mm}(r) &= \mathbb{E}_{o\mathbf{r}}[m(o)m(\mathbf{r})] \\
&= \mathbb{E}_{o\mathbf{r}}\{[m(o) - E(r)][m(\mathbf{r}) - E(r)]\} + [E(r)]^2 \\
&= C_V(r) + (\mu_m + \beta C_Z(r))^2 \quad \text{for } r > 0,
\end{aligned}$$

and Stoyan's $k_{mm}(r)$-function

$$k_{mm}(r) = \frac{C_V(r) + (\mu_m + \beta C_Z(r))^2}{\mu_m^2} \quad \text{for } r > 0,$$

and

$$k_{mm}(0) = \frac{\mathbb{E}_o[(m(o))^2]}{\mu_m^2} = \frac{\sigma_m^2 + \mu_m^2}{\mu_m^2} = 1 + \frac{\beta^2\sigma_Z^2 + \sigma_U^2}{(\alpha + \beta\mu_Z + \beta\sigma_Z^2)^2}\,.$$

The mark covariance and mark correlation are $C_m(r) = C_V(r)$ and $\rho_m(r) = C_V(r)/\sigma_V^2$ and, from the formulas in Table 2.1, it can be deduced that the mark variogram is

$$\gamma_m(r) = V(r) - C_m(r) \quad \text{for } r > 0,$$

that is

$$\gamma_m(r) = \beta^2\sigma_Z^2 + \sigma_U^2 - \beta^2 C_Z(r) - C_U(r) \quad \text{for } r > 0. \tag{4.6}$$

Since the mark characteristics of Table 2.1 are known for the log-intensity marked Cox process, they can be used in model interpretation, model fitting and testing the goodness-of-fit.

Example 4.1. One realisation of a log Gaussian Cox process is simulated in a window of size 200×200 using the following Gaussian random field parameters: the mean is $\mu_Z = 4.0$ and the exponential covariance function

$$C_Z(r;\theta) = \sigma_Z^2 e^{-r/\phi_Z}$$

with $\sigma_Z^2 = 1.5$ and $\phi_Z = 6.0$. The realisation is shown in Figure 4.1. The simulation of the Gaussian random field is performed using the function `GaussRF` in the R library `RandomFields`, see Schlather (2001b), and the random field is simulated on a grid with 200×200 cells.

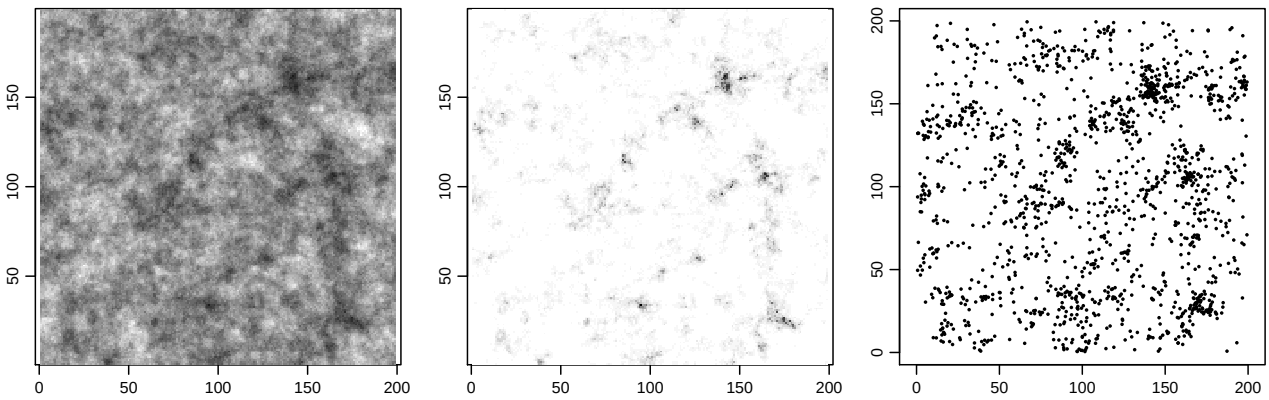

Figure 4.1: The point pattern with 1523 points *on the right* is a realisation of the log Gaussian Cox process with intensity $\{\Lambda(s)\}$ (*in the middle*), which has been generated by the realisation of the Gaussian random field $\{Z(s)\}$ (*on the left*). Dark color corresponds to a high value of a random field.

The log Gaussian Cox process $N = \{x_i\}$ is marked by the marking model

$$m(x_i) = 20 - 10Z(x_i) + U(x_i) \quad \text{for } x_i \in N,$$

where $U(x_i) \sim N(0, 10)$ are independent. A realisation is shown in Figure 4.2. Further, mark summaries calculated from the realisation are shown in Figure 4.3 with their theoretical counterparts. The function `mpp.characteristics` in R library `MarkedPointProcess` has been used to calculate the summaries, see Schlather et al. (2004). Here no edge correction is used. For comparison, the mark variogram has been calculated also with translation edge correction, but the estimates do not differ much from the ones obtained with no edge correction. Several simulation experiments (not presented here) show that there is great variation in mark summaries, especially in $V(r)$, calculated from realisations of the model with given parameters.

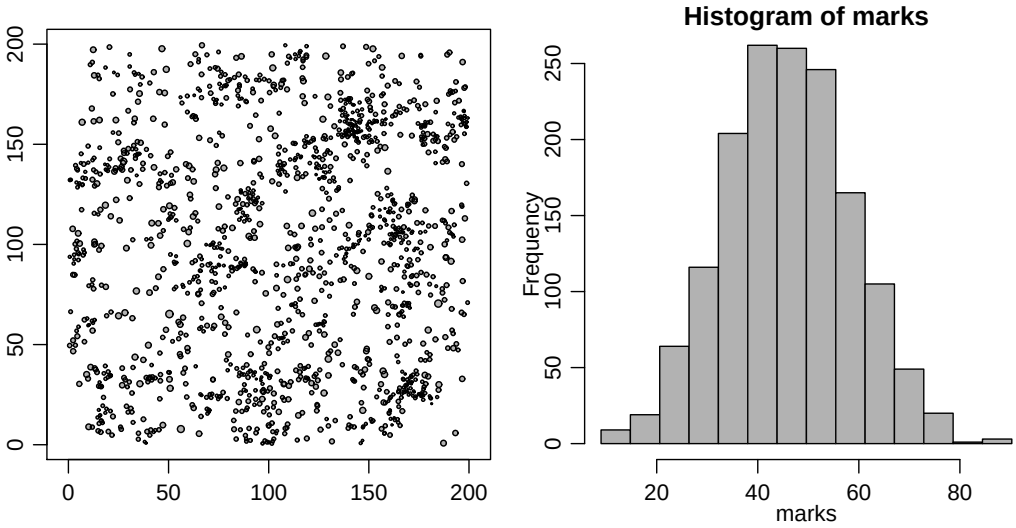

Figure 4.2: A realisation of the log-intensity marked Cox process of Example 4.1. *On the left*: The simulated marked point pattern. The diameter of a circle is proportional to the size of the mark. *On the right*: The histogram of the marks.

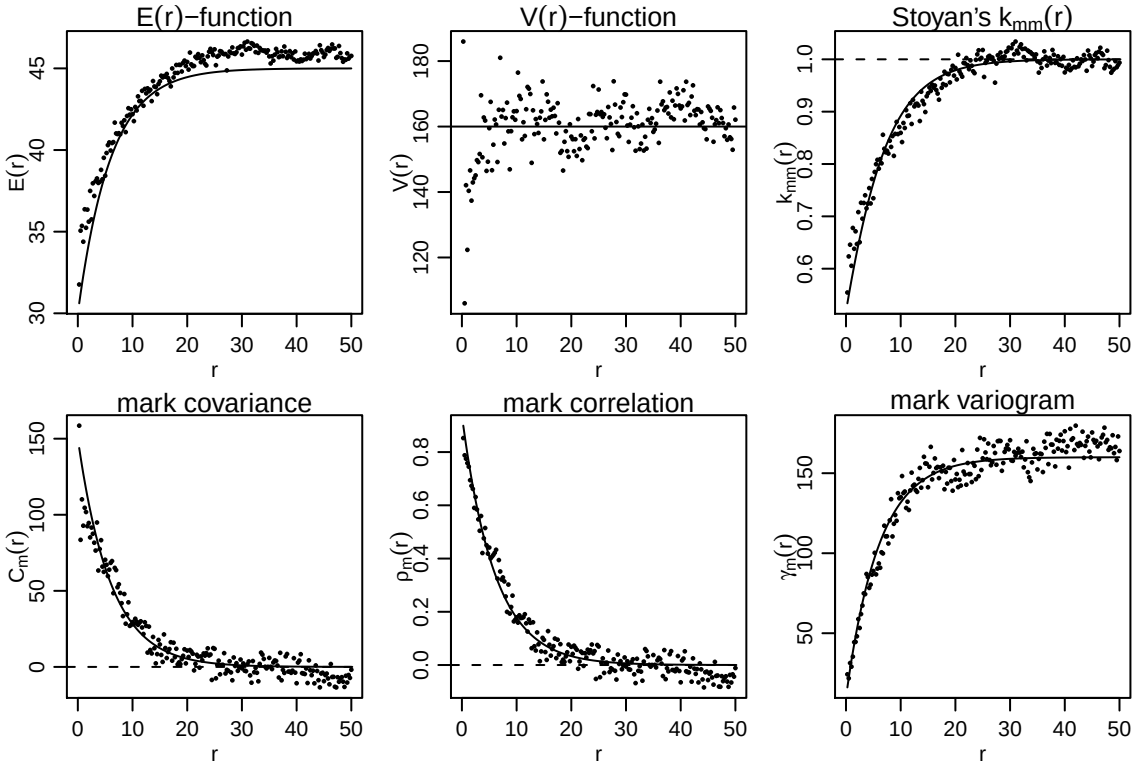

Figure 4.3: Mark summaries (*dots*) calculated from the realisation of the log-intensity marked Cox process of Figure 4.2 and their theoretical counterparts (*solid lines*) with parameters used in simulation. *Dashed* (horizontal) *lines* in the figures of $k_{mm}(r)$, $C_m(r)$ and $\rho_m(r)$ correspond to the case with no dependence.

4.2 Intensity-marked Cox process

Ho and Stoyan (2008) considers an *intensity-marked Cox process*, in which point density and mark sizes are closely coupled. The log Gaussian Cox process is considered as a model for the point locations and the marks are provided by

$$m(x_i) = a + b\Lambda(x_i) + \epsilon(x_i) \quad \text{for } x_i \in N, \tag{4.7}$$

where a and b are model parameters and $\epsilon(x_i)$ is a Gaussian random error with mean 0 and variance τ^2 being independent of $\{\Lambda(s)\}$. The dependence between marks and points affects mark characteristics. The following characteristics are derived for the intensity-marked Cox process in Ho and Stoyan (2008):

- the mean mark $\mu_m = a + b\lambda \exp(\sigma_Z^2)$
- the second moment of marks $\mu_{m2} = a^2 + b^2\lambda^2 \exp(3\sigma_Z^2) + \tau^2 + 2ab\lambda \exp(\sigma_Z^2)$ and, consequently, the mark variance $\sigma_m^2 = \tau^2 + b^2\lambda^2 e^{2\sigma_Z^2}\left(e^{\sigma_Z^2} - 1\right)$
- Stoyan's $k_{mm}(r)$-function

$$k_{mm}(r) = \begin{cases} \dfrac{a^2 + 2ab\lambda e^{\sigma_Z^2 + C_Z(r)} + b^2\lambda^2 e^{2\sigma_Z^2 + 3C_Z(r)}}{(a + b\lambda e^{\sigma_Z^2})^2}, & \text{for } r > 0, \\ \dfrac{a^2 + 2ab\lambda e^{\sigma_Z^2} + b^2\lambda^2 e^{3\sigma_Z^2} + \tau^2}{(a + b\lambda e^{\sigma_Z^2})^2}, & \text{for } r = 0. \end{cases}$$

- $E(r)$-function

$$E(r) = \begin{cases} a + b\lambda e^{\sigma_Z^2 + C_Z(r)}, & \text{for } r > 0, \\ a + b\lambda e^{\sigma_Z^2}, & \text{for } r = 0. \end{cases}$$

- mark variogram

$$\gamma(r) = \begin{cases} b^2\lambda^2 e^{2\sigma_Z^2 + 2C_Z(r)}\left(e^{\sigma_Z^2} - e^{C_Z(r)}\right) + \tau^2, & \text{for } r > 0, \\ 0, & \text{for } r = 0. \end{cases}$$

The intensity-marked Cox process model is very similar to the log-intensity marked Cox process. In intensity-marked Cox process, the point process is directly marked by the intensity $\{\Lambda(s)\}$ (of a log Gaussian Cox process), whereas $\{\log(\Lambda(s))\}$ is used to generate the marks in the log-intensity marked Cox process.

The variance of the random field $\{Z(s)\}$ affects the shape of the mark distribution of the intensity-marked Cox process. If the variance is small, the distribution is almost symmetric, whereas large variance yields a distribution that is positively skew (longer right tail) with $b > 0$ and negatively skew with $b < 0$.

4.3 Geostatistical model for preferential sampling

Menezes (2005) develops a model for geostatistical sampling where the sampling design depends on an observed random field. The sampling points and the values of the observed random field can be considered as a marked point pattern, but the interest is in fact in the random field, which can be observed wherever the investigator's objective is to study its properties.

The model for point locations is a log Gaussian Cox process $N = \{x_i\}$ with the intensity

$$\Lambda(s) = \exp\{c + dZ(s)\},$$

where c and d are real-valued parameters and $\{Z(s)\}$ is a Gaussian random field with mean μ_Z, variance σ_Z^2 and covariance function $C_Z(r)$ as before. The intensity and the pair-correlation of this log Gaussian Cox process are $\lambda = \exp\{c + d\mu_Z + d^2\sigma_Z^2/2\}$ and $g(r) = \exp\{d^2 C_Z(r)\}$ for $r > 0$, respectively (Ho and Stoyan, 2008). The case $d = 0$ corresponds to a homogeneous Poisson process with intensity $\exp(c)$. The marks are defined by

$$m(x_i) = Z(x_i) + \epsilon(x_i) \quad \text{for } x_i \in N, \tag{4.8}$$

where $\epsilon(x_i)$ is a Gaussian random error with mean 0 and variance τ^2 being independent of $\{Z(s)\}$. The marks (4.8) can be considered as a noisy version of the underlying random field $\{Z(s)\}$ that is of interest. The model is called *geostatistical model for preferential sampling* (GMPF).

Clearly, the marks are not in general independent of the underlying point process and this is seen through the mark characteristics, which are studied in Ho and Stoyan (2008). The model has the following mark characteristics:

- the mean mark $\mu_m = \mu_Z + d\sigma_Z^2$
- the second moment of marks $\mu_{m2} = \sigma_Z^2 + (\mu_Z + d\sigma_Z^2)^2 + \tau^2$ and, consequently, the mark variance $\sigma_m^2 = \sigma_Z^2 + \tau^2$
- Stoyan's $k_{mm}(r)$-function

$$k_{mm}(r) = \begin{cases} \dfrac{C_Z(r) + [\mu_Z + d\sigma_Z^2 + dC_Z(r)]^2}{(\mu_Z + d\sigma_Z^2)^2}, & \text{for } r > 0, \\ \dfrac{\sigma_Z^2 + (\mu_Z + d\sigma_Z^2)^2 + \tau^2}{(\mu_Z + d\sigma_Z^2)^2}, & \text{for } r = 0. \end{cases}$$

- $E(r)$-function

$$E(r) = \begin{cases} \mu_Z + d\sigma_Z^2 + dC_Z(r), & \text{for } r > 0, \\ \mu_Z + d\sigma_Z^2, & \text{for } r = 0. \end{cases}$$

- mark variogram

$$\gamma(r) = \begin{cases} \sigma_Z^2 + \tau^2 - C_Z(r), & \text{for } r > 0, \\ 0, & \text{for } r = 0. \end{cases}$$

The marking (4.8) equals the marking of the log-intensity marked Cox process up to parametrization with respect to the intensity of the point process. Indeed, the marking (4.8) can be rewritten as

$$m(x_i) = -\frac{c}{d} + \frac{1}{d}\log(\Lambda(x_i)) + \epsilon(x_i) \quad \text{for } x_i \in N.$$

Thus c and d correspond to $-\alpha/\beta$ and $1/\beta$ in the marking (4.1), respectively. The intensity is provided by a Gaussian random field in both processes: since $\{Z(s)\}$ is Gaussian random field, so is $\{c + dZ(s)\}$. However, the fact that in the geostatistical model the interest is in the random field $\{Z(s)\}$ distinguishes it from the log-intensity marked Cox process.

The parametrization is a matter of interpretation of the model. The GMPF has been developed to investigate the random field $\{Z(s)\}$ that corresponds to some geostatistical variable, such as soil fertility, which can be measured over the whole investigated area. The fact, that the value of the variable can be obtained wherever the investigator likes to have it, distinguishes a regionalised variable from an usual marked point pattern, where the value of a mark can be measured only at predefined locations. However, the soil fertility can be considered to be measured through the height of specific trees, for example, and is then only observed at tree locations.

The parameter d in (4.8) identifies the degree of preferability in sampling. The sampling locations, if taken from areas where $\{Z(s)\}$ is expected to present larger values, can be considered as a realisation of a point process with a positive value of d. On the contrary, negative d means negative association between $\{Z(s)\}$ and the intensity of sampling locations (points). In geostatistical way of thinking, this means that more sampling locations are collected in areas where $\{Z(s)\}$ is expected to get smaller values. In analysing marked point patterns, $d > 0$ means that marks are larger in areas with high point density, whereas negative d yields negative association between mark size and intensity. Thus, d in GMPF can be interpreted similarly as β in the log-intensity marked Cox process or b in the intensity-marked Cox process.

Chapter 5

Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes

The intensity-marked point processes of Ho and Stoyan (2008), Menezes (2005) and Myllymäki (2006), recalled in Chapter 4, are models for marked point patterns where the mark size depends locally on the point density. However, these models assume that the variance of the conditional mark distribution does not depend on the point intensity, which can be a severe restriction as in the following example.

Figure 3.6 presents a marked point pattern of the species *T. tuberculata* in a tropical rainforest. The marks, which are the diameters of trees at breast height (dbh), are plotted against the estimated local tree density at the tree locations: the trees are small when the intensity is high whilst in low intensities both large and small trees are abundant. This means that not only the mean but also the variance of a mark depends on the local intensity, which calls for new *conditionally heteroscedastic* marking models.

The new heteroscedastic marking schemes introduced in this chapter use the (stationary) log Gaussian Cox process as an unmarked point process model as the markings in Chapter 4. The new markings are extensions to the ones in Ho and Stoyan (2008) whose regression type marking model (4.7) can be rewritten as the conditional mark distribution

$$m(x_i)|\,\Lambda(x_i) \sim N(a + b\Lambda(x_i), d^2) \tag{5.1}$$

with constant variance d^2. Our first two generalisations are heteroscedastic modifications of this model. As a third model the gamma family of distributions is suggested, including the exponential model

$$m(x_i)|\,\Lambda(x_i) \sim \mathrm{Exp}\left(\frac{1}{a + b/\Lambda(x_i)}\right).$$

Generally, the new markings are of the form

$$m(x_i)|\,\Lambda(x_i) \sim F_{m|\Lambda}(\cdot|\,\Lambda(x_i)) \quad \text{for } x_i \in N, \tag{5.2}$$

where $F_{m|\Lambda}$ is a parametric conditional distribution of a mark $m(x_i)$ at (fixed) $x_i \in N$ given the intensity $\Lambda(x_i)$. These models allow detailed modelling of intensity-dependence including heteroscedasticity.

Mean mark, mark variance and various mark correlations, see Table 2.1, are derived for the new models. These characteristics can be applied in model selection, fitting, interpretation and in assessing the goodness-of-fit, see Chapter 6. The models are here further illustrated by simulated examples showing the curves for the mark characteristics as well. The obtained marked point patterns are used as synthetic data sets in Section 6.2.3.

5.1 Gaussian intensity-marked Cox process

Let $N = \{x_i\}$ be a log Gaussian Cox process in $W \subseteq \mathbb{R}^d$ and construct the marks by

$$m(x_i)|\,\Lambda(x_i) \sim N\left(a + b\Lambda(x_i), c^2\Lambda(x_i) + d^2\right) \quad \text{for } x_i \in N. \tag{5.3}$$

The marking (5.3) with positive parameters a, b, c^2 and d^2 creates a (conditional) mark distribution with increasing mean and variance along the intensity. The model is a modification of the intensity-marked Cox process considered by Ho and Stoyan (2008), the marking of which equals (5.3) with $c = 0$.

A model (distribution) similar to (5.3) is called *Generalized Normal Variance-Mean Model* by Tjetjep and Seneta (2006). Note that $\Lambda(x_i)$ is a *positive* random variable.

The parameter c does not affect the mean mark μ_m, $E(r)$ and $k_{mm}(r)$, for $r > 0$, and these are the same for the markings (4.7) and (5.3), see Section 4.2, Ho and Stoyan (2008) and Illian et al. (2008). Instead, there is an increase in the mark variance due to the term $c^2\Lambda(x_i)$ in (5.3):

$$\sigma_m^2 = b^2\lambda^2 e^{2\sigma_Z^2}\left(e^{\sigma_Z^2} - 1\right) + c^2\lambda e^{\sigma_Z^2} + d^2$$

and

$$\kappa_{mm}(0) = a^2 + 2ab\lambda e^{\sigma_Z^2} + b^2\lambda^2 e^{3\sigma_Z^2} + c^2\lambda e^{\sigma_Z^2} + d^2,$$

where λ and σ_Z^2 are parameters of the log Gaussian Cox process, see Appendix A. The second-order characteristics $V(r)$ and $\gamma_m(r)$ also change. For the marking (5.3), these are

$$V(r) = d^2 + c^2\lambda e^{\sigma_Z^2 + C_Z(r)} + b^2\lambda^2 e^{2\sigma_Z^2 + 2C_Z(r)}\left(e^{\sigma_Z^2} - 1\right) \quad \text{for } r > 0$$

and

$$\gamma_m(r) = d^2 + c^2\lambda e^{2\sigma_Z^2 + C_Z(r)} + b^2\lambda^2 e^{2\sigma_Z^2 + 2C_Z(r)}\left(e^{\sigma_Z^2} - e^{C_Z(r)}\right) \quad \text{for } r > 0,$$

where $C_Z(r)$ is the covariance function of the Gaussian random field $\{Z(s)\}$. The mark characteristics $E(r)$, $V(r)$, $\kappa_{mm}(r)$ and $\gamma_m(r)$ are derived using the equation

$$\begin{aligned} \mathbb{E}_{o\mathbf{r}}\{f(m(o), m(\mathbf{r}))\} &= \frac{\mathbb{E}[f(m(o), m(\mathbf{r}))\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} \\ &= \frac{\mathbb{E}\{\Lambda(o)\Lambda(\mathbf{r})\mathbb{E}[f(m(o), m(\mathbf{r}))|\Lambda(o), \Lambda(\mathbf{r})]\}}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]}, \end{aligned} \tag{5.4}$$

see Ho and Stoyan (2008) and Appendix A. In fact, for the calculation of these mark characteristics only the conditional mean $\mathbb{E}_o[m(o)|\,\Lambda(o)]$ and variance $\mathrm{var}_o[m(o)|\,\Lambda(o)]$ need to be specified for the model, not the whole conditional mark distribution. For calculation of (5.4) expectations of the form $\mathbb{E}[e^X]$ and $\mathbb{E}[e^{t_1X + t_2Y}]$, where X and Y are Gaussian random variables

and t_1, t_2 constants, are needed. These are easily obtained by the formulas of the moment generating functions of the 1- and 2-dimensional normal distributions, respectively.

Negative dependence between the marks and the (local) intensity is also of interest. The first chance is to set $b < 0$ in (5.1) or in (5.3). For the model with marking (5.3), the conditional mark variance is larger in areas with high intensity and with 'small' mean mark (if $c \neq 0$) and, if $C_Z(r)$ is a decreasing function, then $V(r)$ is decreasing regardless of the sign of b. Having the log Gaussian Cox process as a model for point locations, the resulting mark distribution tends to be negatively skew (longer left tail) if $b < 0$.

An example for which the markings (5.1) and (5.3) are not appropriate is the marked point pattern of *T. tuberculata* in Figure 3.6: the empirical $E(r)$ and $V(r)$ both increase along r and the mark distribution is positively skew, see Figure 3.7. A possible marking which leads to such characteristics has the conditional mean mark and mark variance linearly dependent on $1/\Lambda(x_i)$. The marking

$$m(x_i)|\,\Lambda(x_i) \sim N\left(a + b\frac{1}{\Lambda(x_i)}, c^2\frac{1}{\Lambda(x_i)} + d^2\right) \quad \text{for } x_i \in N \tag{5.5}$$

with $a, b, c^2, d^2 > 0$ produces a (conditional) mark distribution with decreasing mean and variance along the increase in the intensity. If $c = 0$, only the conditional mean of marks depends on the intensity, and if $b = 0$, only the variance.

The mean mark and mark variance are now

$$\mu_m = a + \frac{b}{\lambda} \quad \text{and} \quad \sigma_m^2 = b^2 e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right) + \frac{c^2}{\lambda} + d^2, \tag{5.6}$$

see Appendix A. Assuming that $C_Z(r)$ is a decreasing function, the characteristics

$$E(r) = a + \frac{b}{\lambda}e^{-C_Z(r)} \quad \text{for } r > 0$$

and

$$V(r) = d^2 + \frac{c^2}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}\left(e^{\sigma_Z^2} - 1\right)e^{-2C_Z(r)} \quad \text{for } r > 0$$

increase along r. Stoyan's $k_{mm}(r)$-function is

$$k_{mm}(r) = \begin{cases} \dfrac{a^2 + \left(\frac{2ab}{\lambda} + \frac{b^2}{\lambda^2}\right)e^{-C_Z(r)}}{\left(a + \frac{b}{\lambda}\right)^2}, & \text{for } r > 0, \\ 1 + \dfrac{b^2 e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right) + \frac{c^2}{\lambda} + d^2}{\left(a + \frac{b}{\lambda}\right)^2}, & \text{for } r = 0. \end{cases}$$

It indicates that at short inter-point distances the marks tend to be small. The mark variogram

$$\gamma_m(r) = d^2 + \left(\frac{c^2}{\lambda} - \frac{b^2}{\lambda^2}\right)e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{\sigma_Z^2 - 2C_Z(r)} \quad \text{for } r > 0 \tag{5.7}$$

also points out positive association of the marks at short interpoint distances. These formulas are derived in Appendix A.

The marking (5.5) allows one to obtain second-order characteristics, especially the $E(r)$- and $V(r)$-functions, which are in agreement with their empirical counterparts of *T. tuberculata*.

Example 5.1. The log Gaussian Cox process $N = \{x_i\}$ of the Example 4.1 is marked by the marking model

$$m(x_i)|\,\Lambda(x_i) \sim N(20 + 0.2/\Lambda(x_i), 0.5^2/\Lambda(x_i) + 2^2).$$

A realisation is shown in Figure 5.1, and in Figure 5.2, the mark summaries estimated from the realisation are plotted with their theoretical counterparts.

There is much variation in realisations of the conditionally heteroscedastic intensity-dependent models, and therefore, the empirical summaries calculated from realisations may differ even quite a lot from the theoretical characteristics. The shape of the mark summaries stays the same, but the level they reach varies among realisations. Note that the underlying Cox process realisation affects the marks.

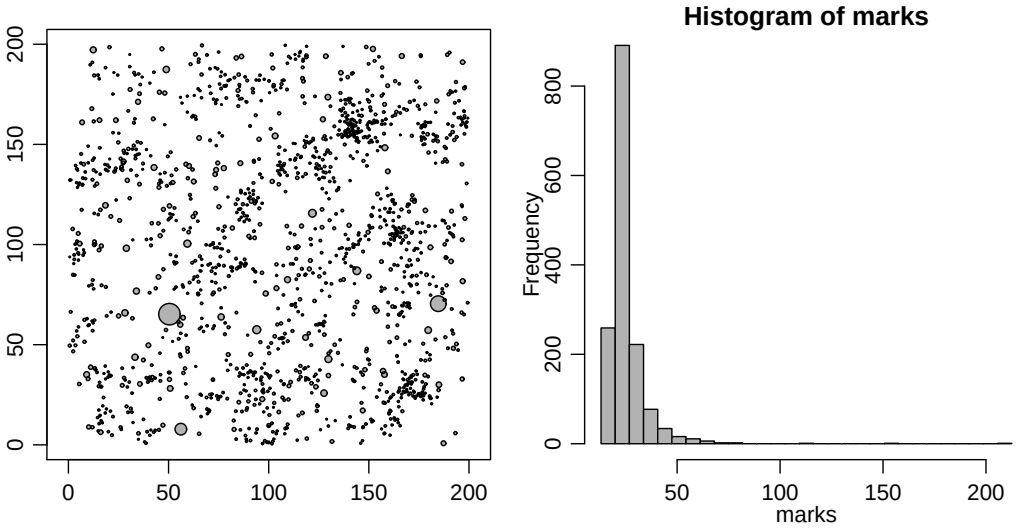

Figure 5.1: A realisation of the Gaussian intensity-marked Cox processes of Example 5.1. *On the left*: The simulated marked point pattern. The diameter of a circle is proportional to the size of the mark. *On the right*: The histogram of the marks.

5.2 Gamma intensity-marked Cox process

In our applications, the model assumption

$$\mathbb{E}_{x_i}[m(x_i)|\Lambda(x_i)] = a + b/\Lambda(x_i)$$

seems to be reasonable. A further idea is to construct the marks according to

$$m(x_i)|\,\Lambda(x_i) \sim \text{Exp}\left(\frac{1}{a + b/\Lambda(x_i)}\right) \quad \text{for } x_i \in N \tag{5.8}$$

resulting in the conditional mark variance

$$\text{var}_{x_i}[m(x_i)|\,\Lambda(x_i)] = (a + b/\Lambda(x_i))^2.$$

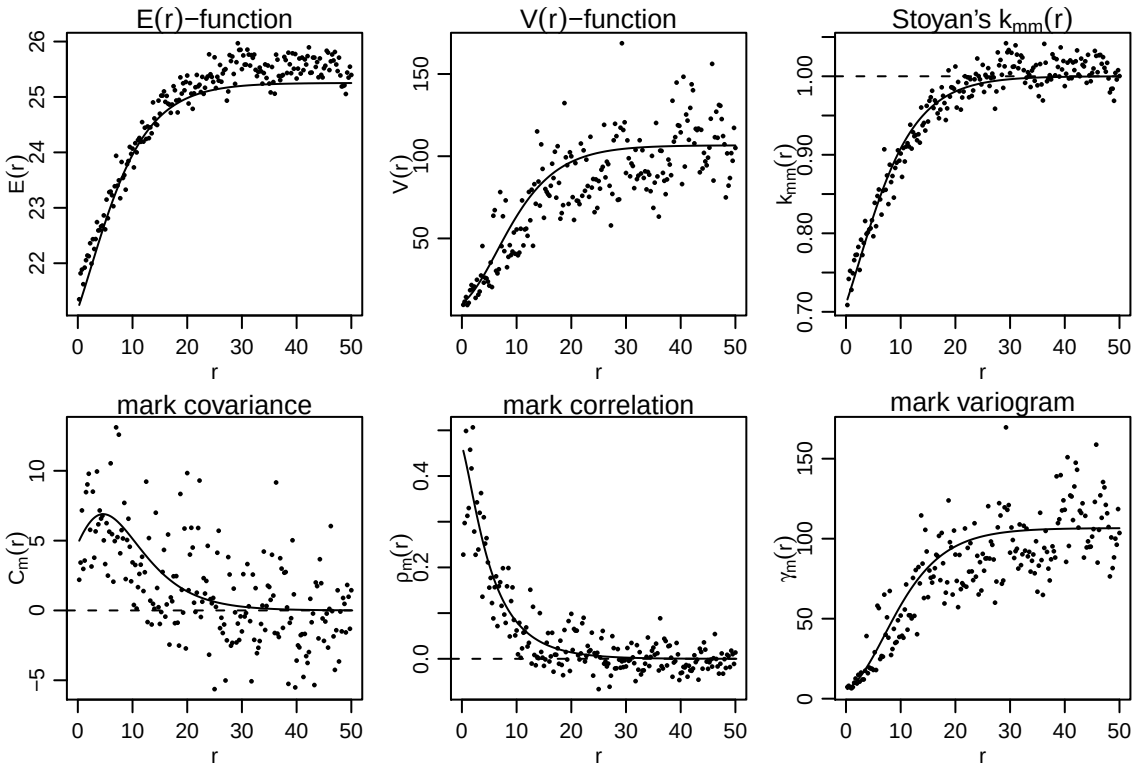

Figure 5.2: Mark summaries (*dots*) calculated from the realisation of the Gaussian intensity-marked Cox process of Figure 5.1 and their theoretical counterparts (*solid lines*) with parameters used in simulation. *Dashed* (horizontal) *lines* in the figures of $k_{mm}(r)$, $C_m(r)$ and $\rho_m(r)$ correspond to the case with no dependence.

Since the exponential distribution is a special case of the gamma distribution with the shape parameter $\alpha = 1$, the gamma model with (fixed) shape parameter α is considered in the following.

Let $N = \{x_i\}$ be a log Gaussian Cox process in $W \subseteq \mathbb{R}^d$ and

$$m(x_i)|\,\Lambda(x_i) \sim \text{Gamma}\,(\alpha, \beta) \quad \text{for } x_i \in N \tag{5.9}$$

with the shape parameter $\alpha > 0$ and the rate parameter $\beta = 1/(a + b/\Lambda(x_i))$. Consequently, the conditional mean mark and mark variance of a mark $m(x_i)$ at $x_i \in N$ given $\Lambda(x_i)$ are α/β and α/β^2, respectively. Thus, both of these depend on the intensity if $b \neq 0$. Of interest here is the case $b > 0$. (Positive dependence between the marks and intensity can be studied by the model with $\beta = 1/(a + b\Lambda(x_i))$, for example.) Note also that β needs to be positive, which is definitely satisfied if $a > 0$ and $b > 0$.

The intensity-dependence of the marks can be deduced from the mark characteristics. The mean mark and the mark variance of the mark (5.9) are

$$\mu_m = \alpha\left(a + \frac{b}{\lambda}\right) \quad \text{and} \quad \sigma_m^2 = \alpha\left(a^2 + \frac{2ab}{\lambda} + b^2 e^{-2\mu_Z}\right) + \alpha^2 b^2 e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right).$$

Further,

$$E(r) = \alpha\left(a + \frac{b}{\lambda}e^{-C_Z(r)}\right) \quad \text{for } r > 0,$$

$$V(r) = \alpha\left(a^2 + \frac{2ab}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{\sigma_Z^2 - 2C_Z(r)}\right) + \alpha^2\frac{b^2}{\lambda^2}e^{-2C_Z(r)}\left(e^{\sigma_Z^2} - 1\right) \quad \text{for } r > 0,$$

$$k_{mm}(r) = \begin{cases} \dfrac{a^2 + \left(\frac{2ab}{\lambda} + \frac{b^2}{\lambda^2}\right)e^{-C_Z(r)}}{\left(a + \frac{b}{\lambda}\right)^2}, & \text{for } r > 0, \\ 1 + \dfrac{\alpha\left(a^2 + \frac{2ab}{\lambda} + b^2e^{-2\mu_Z}\right) + \alpha^2b^2e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right)}{\alpha^2\left(a + \frac{b}{\lambda}\right)^2}, & \text{for } r = 0 \end{cases}$$

and

$$\gamma_m(r) = \alpha\left(a^2 + \frac{2ab}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{\sigma_Z^2 - 2C_Z(r)}\right) + \alpha^2\frac{b^2}{\lambda^2}e^{-C_Z(r)}\left(e^{\sigma_Z^2 - C_Z(r)} - 1\right) \quad \text{for } r > 0,$$

see Appendix A. The characteristics of the marking (5.8) are obtained by selecting $\alpha = 1$.

The above theoretical characteristics show that the gamma intensity-marking, together with the log Gaussian Cox process, is able to generate marked point patterns with clusters of points having small marks, but simultaneously allow large variation for the marks. These characteristics are particularly useful tools in the preliminary analysis of data. They help in model selection for the analysis of real data.

Example 5.2. The log Gaussian Cox process $N = \{x_i\}$ of the Example 4.1 is marked by the marking model

$$m(x_i)|\,\Lambda(x_i) \sim \text{Exp}(1/(30 + 0.8/\Lambda(x_i)),$$

see Figures 5.3 and 5.4. The correlation between marks is quite weak in this realisation when measured by the correlation function $\rho_m(r)$, whereas $E(r)$- and $V(r)$-functions indicate clear dependence between marks and points. For comparison, a realisation of the gamma intensity-marked Cox process is simulated with the same $\beta = 1/(30 + 0.8/\Lambda(x_i))$, and with shape parameter $\alpha = 0.5$, see Figures 5.5 and 5.6. There is much variation in the mark summaries of realisations of exponential and gamma intensity-marked Cox processes.

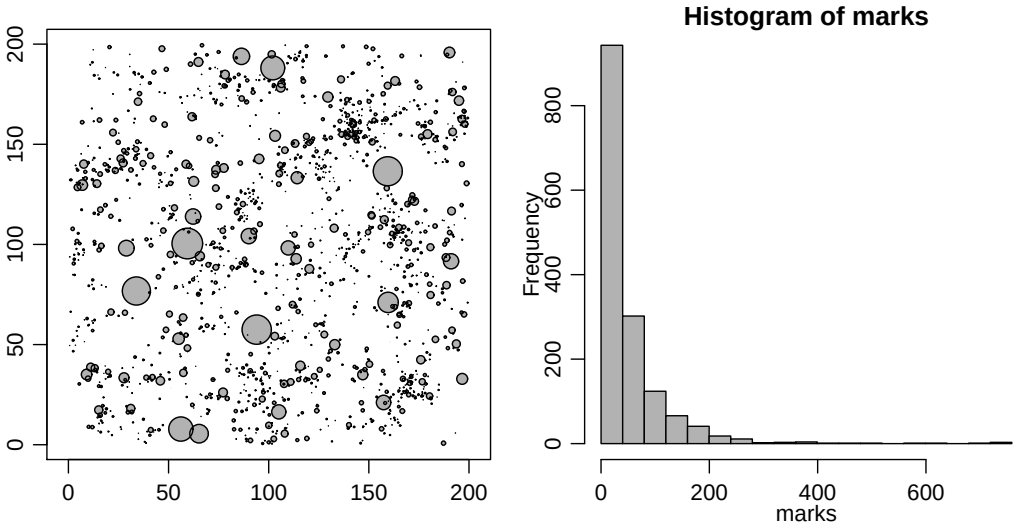

Figure 5.3: A realisation of the exponential intensity-marked Cox process of Example 5.2. *On the left*: The simulated marked point pattern. The diameter of a circle is proportional to the size of the mark. *On the right*: The histogram of the marks.

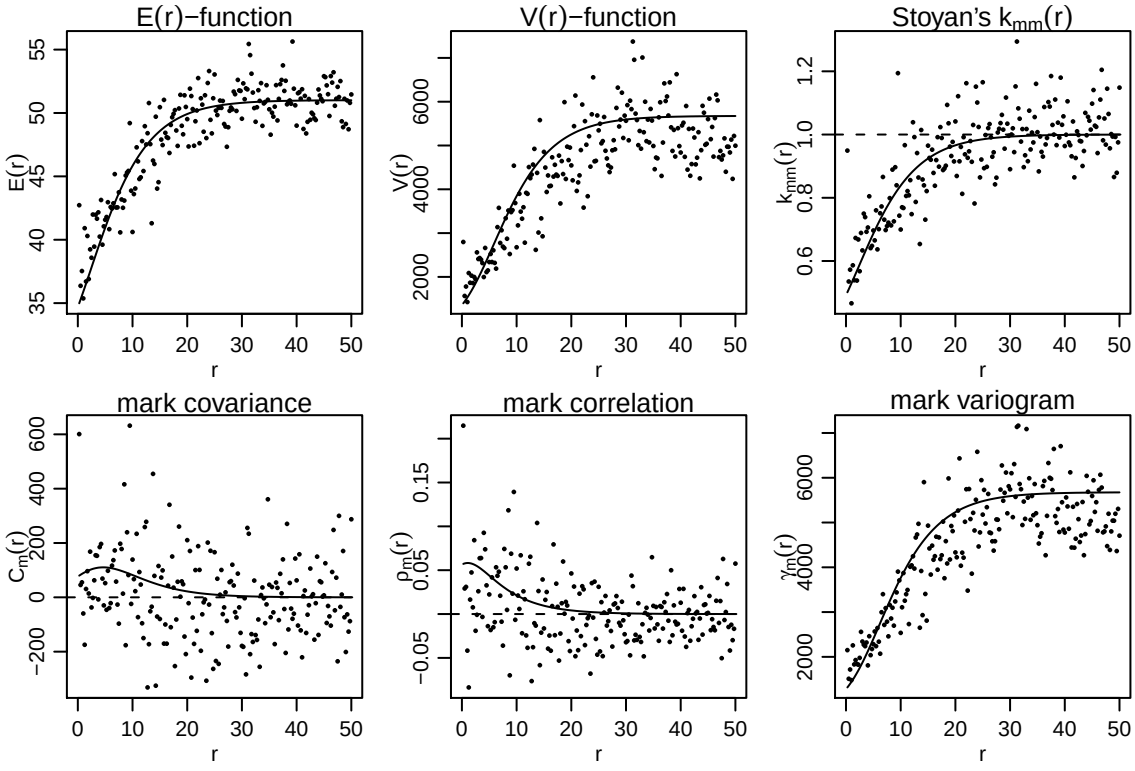

Figure 5.4: Mark summaries (*dots*) calculated from the realisation of the exponential intensity-marked Cox process of Figure 5.3 and their theoretical counterparts (*solid lines*) with parameters used in simulation. *Dashed* (horizontal) *lines* in the figures of $k_{mm}(r)$, $C_m(r)$ and $\rho_m(r)$ correspond to the case with no dependence.

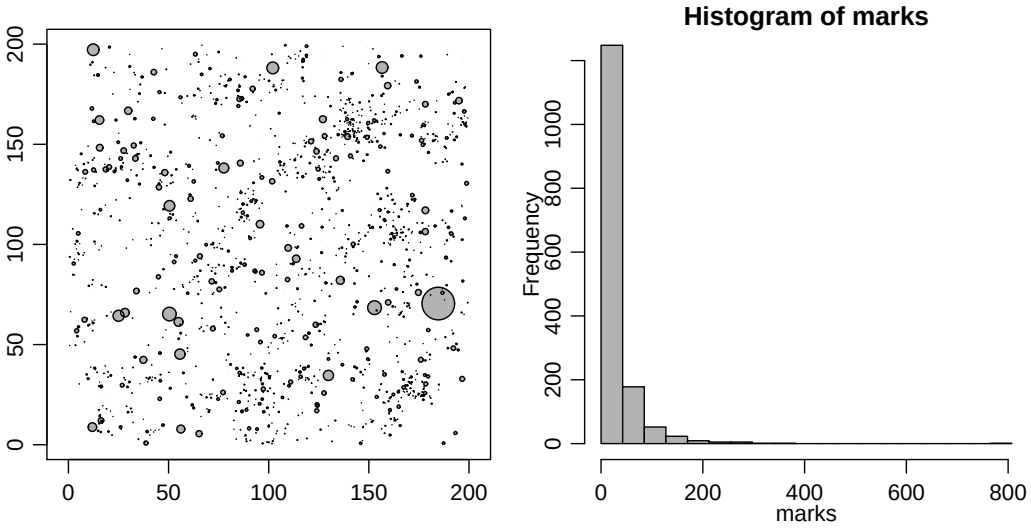

Figure 5.5: A realisation of the gamma intensity-marked Cox process of Example 5.2. *On the left*: The simulated marked point pattern. The diameter of a circle is proportional to the size of the mark. *On the right*: The histogram of the marks.

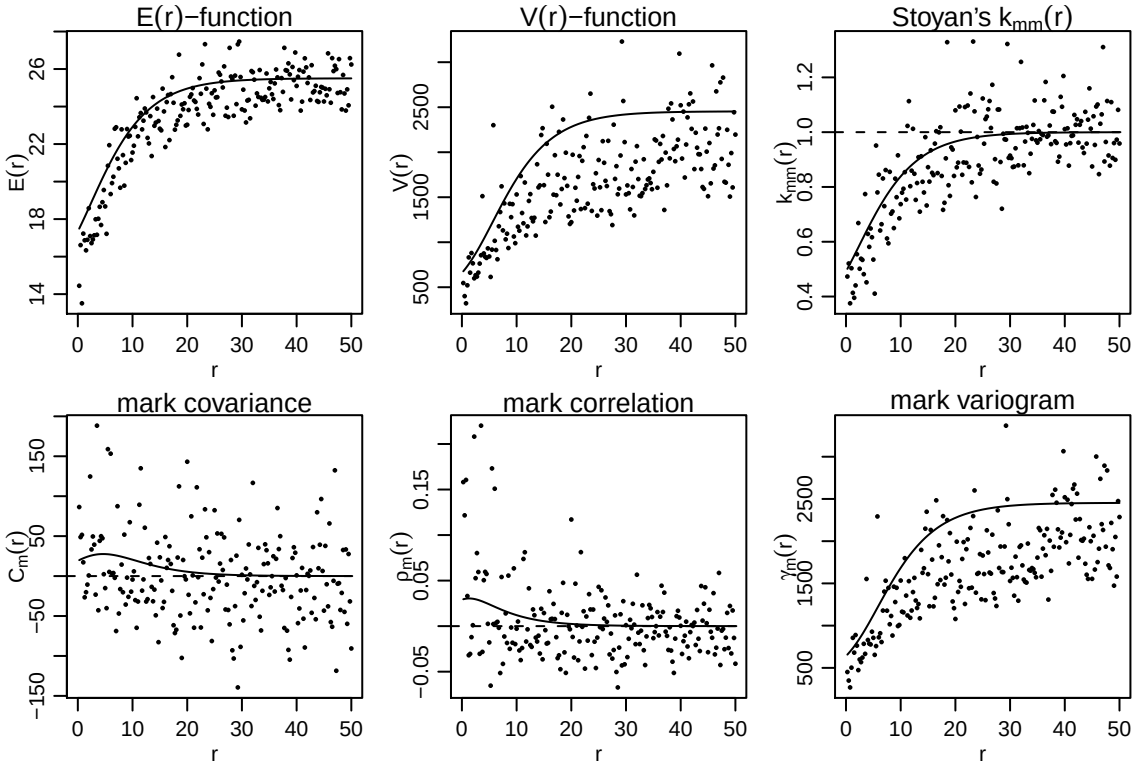

Figure 5.6: Mark summaries (*dots*) calculated from the realisation of the gamma intensity-marked Cox process of Figure 5.5 and their theoretical counterparts (*solid lines*) with parameters used in simulation. *Dashed* (horizontal) *lines* in the figures of $k_{mm}(r)$, $C_m(r)$ and $\rho_m(r)$ correspond to the case with no dependence.

Chapter 6

Statistical inference for conditionally marked Cox point processes

The likelihood of the conditionally marked Cox point process of Chapters 4 and 5 is of the form

$$\mathbb{E}\left\{\prod_{i=1}^{n}\left[\Lambda(x_i)f_{m|\Lambda}(m(x_i)|\,\Lambda(x_i))\right]e^{-\int_W \Lambda(s)\mathrm{d}s}\right\},$$

where the expectation is over the distribution of $\{\Lambda(s)\}$ and $f_{m|\Lambda}(m(x_i)|\,\Lambda(x_i))$ is the parametric density of the conditional mark distribution of the mark $m(x_i)$ at $x_i \in N$ given the intensity $\Lambda(x_i)$. This likelihood is analytically intractable. We start the development of model fitting techniques from simple practical procedures for preliminary analysis of data, before the suggestion of a more advanced Bayesian method. Our focus is in estimation of the mark distribution defined conditional on the intensity.

A simple idea for parameter estimation is to fit a non-linear regression model to the marks after conditioning by the empirical intensity function and the point locations. An obvious drawback of conditional fitting is that it may be sensitive to the smoothing level used in the (non-parametric) intensity estimation. A point process model is not specified in this model fitting. Our second idea is to base the inference on the second-order properties of the marked point process. The minimum contrast method (Diggle, 1979), called also a 'least squares' method, is used in a new context of log Gaussian Cox processes with intensity-dependent marks. This estimation is preceded by the estimation of the mean and the covariance function of the Gaussian random field generating the log Gaussian Cox process.

The minimum contrast method is a computationally easy approach, but it has been criticized, because it is based on some user-specified choices (see e.g. Cressie, 1993, p. 666 and Guan, 2006). Moreover, the two simple methods condition either by the empirical intensity function or directly by the estimated mean and covariance function of the log Gaussian Cox process. Since the marks are defined conditional on the intensity, the intensity should be estimated. Simultaneous estimation of intensity and marking would be more efficient than the sequential one, because under the assumption of intensity-dependence also the marks contain information on the intensity. We suggest to estimate the conditionally marked models using an empirical Bayesian approach where the Markov chain Monte Carlo (MCMC) methods are used in posterior simulation of the intensity and the parameters of the marking simultaneously. The

prior distribution of the intensity is assumed to be log Gaussian, the parameters of which are estimated using the point data.

All of the methods are presented, but we recommend the use of the Bayesian method in the final analysis of data. We also perform an extensive simulation study for the Bayesian method. The methods based on moments and minimum contrast estimation serve specifically as tools in empirical data analysis and in model identification. These methods are demonstrated for a few specific markings.

Model fitting has not been considered in the seminal paper Ho and Stoyan (2008) of intensity-marked Cox processes. However, Menezes (2005) considers estimation of parameters of the geostatistical model for preferential sampling allowing so-called *length-bias*, see Menezes (2005, p. 182), and Ho (2006) analysis a data set with the geostatistical model and the intensity-marked Cox process using a least squares method for $C_Z(r)$ and $k_{mm}(r)$ or $k_{mm}(r)$ and $\gamma_m(r)$, see Ho (2006, p. 83).

The rest of this chapter is organized as follows. Two estimators based on simple estimation of the intensity are presented in Section 6.1. In particular, an adaptive kernel estimation of the intensity, that uses also mark information, is introduced for the log-intensity marked Cox process. Section 6.2 considers the minimum contrast estimation, which is exploited for the log-intensity marked, Gaussian intensity-marked and gamma intensity-marked Cox processes. Section 6.3 introduces the (empirical) Bayesian method which is the main result of this chapter. Simulations studies for this method are performed in Section 6.4. Finally, the new models are applied to the marked point pattern from a tropical rainforest in Section 6.5. Section 6.6 is for discussion.

6.1 Empirical data analysis on marking

The simplest estimation method for the intensity-marked models estimates first the intensity $\Lambda(x_i)$, $x_i \in N$, and then considers the estimation of the marking parameters as a regression type problem. Of course, the estimation of the intensity is a non-trivial estimation problem of its own, but using a simple estimate for the intensity and studying its relationship to the marks may provide valuable information in the beginning of the analysis of real data.

The two methods presented in Section 6.1.2 are indeed meant to be used in preliminary analysis and they may be sensitive to the estimated intensity. We put the moment-based method a bit forward for the log-intensity marked Cox process and propose an adaptive kernel estimation method for the intensity which uses also mark information. Empirical evidence suggests that this method is competitive over the simple estimation of the intensity.

6.1.1 Kernel estimation of the intensity function

A kernel estimator of the intensity at location $s \in \mathbb{R}^d$ with kernel K is defined by

$$\hat{\Lambda}_K(s) = \frac{1}{h^d} \sum_{i=1}^{n} K\left(\frac{s - x_i}{h}\right), \tag{6.1}$$

where h is the *bandwidth* (or window width or smoothing parameter), see e.g. Silverman (1986) and Scott (1992). A kernel used here is the Epanechnikov kernel

$$K(u) = \begin{cases} \frac{1}{2b_d}(d+2)(1-\|u\|^2), & \text{if } \|u\|^2 \leq 1, \\ 0, & \text{otherwise,} \end{cases} \tag{6.2}$$

where b_d is the volume of the d-dimensional unit ball ($b_2 = \pi$) and $\|u\| = \|(s - x_i)/h\|$ is the distance of two points in $\mathbb{R}^d$ scaled by h.

Kernel estimators smooth the surface and the chosen bandwidth h can affect the results remarkably. Diggle (2003) and Stoyan and Stoyan (1994), for example, consider the choice of a suitable bandwidth. However, a 'suitably chosen' h may smooth the surface too much in areas with high intensity and too little in areas with low intensity. To solve the problem of smoothing adaptive kernel estimators have been developed. An idea to deal with long tailed intensities is to use a broader kernel in regions with low intensity than in regions with high intensity. Referring to Scott (1992), there are two types of adaptive kernel estimates: either a bandwidth changes for each estimation point $s \in W$ or for each $x_i \in N$. The latter case is recalled here.

An obvious practical problem is to decide whether an observation is in a region of low or high intensity. The adaptive kernel approach presented in Silverman (1986, p. 100) uses a two-stage procedure to solve this problem. The general idea is to first find a pilot estimate of the intensity, $\hat{\Lambda}_P(s)$, which satisfies $\hat{\Lambda}_P(x_i) > 0$ for all $x_i \in N$. Then the *local bandwidth factors* w_i can be defined by

$$w_i = \left\{ \frac{\hat{\Lambda}_P(x_i)}{g} \right\}^{-\xi},$$

where g is the geometric mean of the $\hat{\Lambda}_P(x_i)$ calculated through

$$\log(g) = \frac{1}{n} \sum_{i=1}^{n} \log(\hat{\Lambda}_P(x_i)),$$

and $0 \leq \xi \leq 1$ is a *sensitivity parameter*. The larger the power ξ is, the more sensitive the method will be to variations in the pilot density. The value $\xi = 0$ corresponds to the fixed bandwidth kernel approach since all w_i will then equal 1. Silverman (1986) recommends to use the value $\xi = 1/2$. An adaptive kernel estimator of the intensity can be defined by

$$\hat{\Lambda}_A(s) = \frac{1}{h^d} \sum_{i=1}^{n} \frac{1}{w_i^d} K\left(\frac{s - x_i}{h w_i} \right), \tag{6.3}$$

where h is the bandwidth, which is scaled for each $x_i \in N$ by w_i, and the kernel K in (6.2) can be used, for example.

6.1.2 Moment-based estimators and conditional likelihood for the marks

Assume $\Lambda(x_i)$, $x_i \in N$, is known. Then the following estimation methods, called Method 1 and 2, can be used. The moment-based estimator operates with the conditional mean of the

marks and the conditional likelihood method with the conditional mark distribution. Because the intensity is in practise unknown, an estimate obtained e.g. through kernel method must be used in the following equations instead of $\Lambda(x_i)$.

Method 1: Moment-based estimators

Assume a parametric formula is set for $\mathbb{E}_{x_i}[m(x_i)|\Lambda(x_i)]$. Then simple curve fitting can be done by minimizing the square sum

$$\sum_{[x_i;\, m(x_i)]\in N_m} \left(m(x_i) - \mathbb{E}_{x_i}[m(x_i)|\, \Lambda(x_i)]\right)^2 \tag{6.4}$$

or, alternatively, the weighted square sum

$$\sum_{[x_i;\, m(x_i)]\in N_m} \frac{1}{\Lambda(x_i)} \left(m(x_i) - \mathbb{E}_{x_i}[m(x_i)|\, \Lambda(x_i)]\right)^2 \tag{6.5}$$

with respect to parameters in $\mathbb{E}_{x_i}[m(x_i)|\Lambda(x_i)]$.

Method 2: Conditional likelihood for the marks

Assume that the marking is specified through a mark distribution which depends on the intensity $\Lambda(x_i)$ and the marks are assumed to be conditionally independent given $\{\Lambda(s)\}$. Let $f_{m|\Lambda}(m(x_i)|\, \Lambda(x_i))$ stand for the parametric density of the conditional mark distribution of the mark $m(x_i)$ at (fixed) $x_i \in N$ given the intensity $\Lambda(x_i)$. Then the conditional log likelihood for the marks given $\{x_i\}$ and $\Lambda(x_i)$, $x_i \in N$, is

$$\sum_{[x_i;\, m(x_i)]\in N_m} \log f_{m|\Lambda}(m(x_i)|\, \Lambda(x_i))\,. \tag{6.6}$$

Its maximization leads to parameter estimates.

6.1.3 Application to the log-intensity marked Cox process

In this section, Methods 1 and 2 (see Section 6.1.2) are illustrated for the marking

$$m(x_i)|\, \Lambda(x_i) \sim N(20 - 10\log(\Lambda(x_i)), 10), \tag{6.7}$$

see (4.1). First, estimates $\hat{\Lambda}(x_i)$, $x_i \in N$, are obtained by (6.1) with Epanechnikov kernel (6.2) and bandwidth $h = 10$. These estimates are then used to obtain estimates for the marking parameters α (true value 20) and β (true value -10) using Methods 1 and 2. The estimates are shown in Table 6.1.

It is obvious that the estimates are not very good, even though Method 1b provides surprisingly good point estimates for this particular simulation experiment. The variance σ_U^2 might be estimated from the residuals for Methods 1a and 1b. The problem arises because the simple kernel estimation can not produce such large variation (or range) in the intensity that the log Gaussian Cox process in fact has. In the following, a new adaptive kernel estimation method is described. This method utilizes the mark information in estimation by assuming $\mathbb{E}_{x_i}[m(x_i)|\Lambda(x_i)] = \alpha + \beta\log(\Lambda(x_i))$.

Table 6.1: Estimates for the mark parameters of the log-intensity marked Cox process.

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_U^2$
Method 1a: minimize (6.4)	14.72	-11.15	-
Method 1b: minimize (6.5)	19.46	-9.61	-
Method 2: conditional ML	14.74	-11.14	89.39
Value used in simulation	20.00	-10.00	10.00

Mark adapted kernel estimation

The adaptive kernel estimator of the intensity (6.3) is based only on the point data. Since mark information is available, it can be utilized as well. The use of mark information in estimation of the intensity is particularly interesting for intensity-dependently marked point processes, since in these models the marks and intensity are closely coupled. The following algorithm introduces a new adaptive kernel estimator. It assumes $\mathbb{E}_{x_i}[m(x_i)|\Lambda(x_i)] = \alpha + \beta \log(\Lambda(x_i))$, where α and β are the parameters to be estimated. The estimation is sensitive to this assumption, and therefore, it should not be used for model selection.

Algorithm 6.1. (mark adapted kernel estimation)

1. Find a (pilot) estimate of the intensity, $\hat{\Lambda}_0(s)$, which satisfies $\hat{\Lambda}_0(x_i) > 0$ for all $x_i \in N$.
2. Estimate the parameters α and β by minimizing (6.4) with $\Lambda = \hat{\Lambda}_0$. Let $\hat{\alpha}_0$ and $\hat{\beta}_0$ stand for the obtained estimates.
3. Calculate new estimates for the intensity at locations $x_i \in N$ by
$$\hat{\Lambda}_1(x_i) = \exp\left(\frac{m(x_i) - \hat{\alpha}_0}{\hat{\beta}_0}\right). \tag{6.8}$$
4. Calculate the local bandwidth factors w_i by
$$w_i = \left\{\frac{\hat{\Lambda}_1(x_i)}{g}\right\}^{-\xi}, \tag{6.9}$$
where g is the geometric mean of the $\hat{\Lambda}_1(x_i)$ and $0 \leq \xi \leq 1$ is a sensitivity parameter.
5. Calculate new intensity estimates $\hat{\Lambda}_{A,m}(x_i)$ for all $x_i \in N$ using the weights (6.9) in the adaptive kernel estimator (6.3) and the Epanechnikov kernel (6.2), for example.
6. Estimate the parameters α and β by minimizing (6.4) with $\hat{\Lambda}_{A,m}(x_i)$, $x_i \in N$. Let $\hat{\alpha}_1$ and $\hat{\beta}_1$ stand for the new estimates.
7. Iterate steps 3-6, if required.

If estimates for α and β are calculated by minimizing (6.5) (instead of (6.4)), then (6.8) in Algorithm 6.1 must be replaced by

$$\hat{\Lambda}_1(x_i) = \exp\left(\hat{\Lambda}_0(x_i) \cdot \frac{m(x_i) - \hat{\alpha}_0}{\hat{\beta}_0}\right).$$

For other purposes than here, $\hat{\Lambda}_{A,m}(s)$ may be calculated at any $s \in W \subseteq \mathbb{R}^d$.

Often the estimates $\hat{\Lambda}_0(x_i)$ are obtained by simple kernel estimation with fixed bandwidth h. Indeed, here the estimates obtained by Method 1a (see Table 6.1) are used as $\hat{\alpha}_0$ and $\hat{\beta}_0$. These estimates are further used in (6.8) to obtain new intensity estimates $\Lambda_1(x_i)$, which are then used as pilot estimates in adaptive kernel estimation. The idea is, that $\Lambda_1(x_i)$ are expected to provide more variation to the weights (6.9) than $\hat{\Lambda}_0(x_i)$.

The estimates of α and β are expected to stabilise after a few iterations of Algorithm 6.1. In fact, referring to Scott (1992) and Silverman (1986), using a nonadaptive pilot estimate for the intensity in adaptive kernel estimation should be adequate and the adaptive kernel method is insensitive to the fine detail of the pilot estimate. Accordingly, one iteration (steps $1-6$) should be enough. This seems to be relevant for estimation of the intensity and the parameters α and β.

The procedure described above is applied to the simulated data. First, in addition to intensity estimates $\hat{\Lambda}(x_i)$ obtained by (6.1), the adaptive kernel estimates $\hat{\Lambda}_A(x_i)$ (without mark information) are calculated through (6.3) with the Epanechnikov kernel and the pilot estimates $\hat{\Lambda}(x_i)$. Methods 1a and 1b are employed both with $\hat{\Lambda}(x_i)$ and $\hat{\Lambda}_A(x_i)$ yielding estimates $(\hat{\alpha}_0, \hat{\beta}_0)$ and $(\hat{\alpha}_A, \hat{\beta}_A)$, respectively, see Table 6.2. Further, Algorithm 6.1 is executed with $k = 1, \ldots, 10$ iterations. This gives estimates $(\hat{\alpha}_k, \hat{\beta}_k)$ for $k = 1, \ldots, 10$. The estimates for $k = 1, 5, 10$ are shown in Table 6.2.

Table 6.2: Estimates for the parameters α and β of the log-intensity marked Cox process.

	$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_A$	$\hat{\beta}_A$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_5$	$\hat{\beta}_5$	$\hat{\alpha}_{10}$	$\hat{\beta}_{10}$
Method 1a	14.25	-11.15	23.09	-7.89	20.06	-8.90	22.26	-8.03	22.30	-8.01
Method 1b	19.46	-9.61	33.09	-5.06	17.89	-10.17	17.93	-10.15	17.93	-10.15

Estimates $\hat{\alpha}_0$ and $\hat{\beta}_0$ obtained by Method 1b (minimize (6.5)) are close to the ones used in simulation, whereas estimates $\hat{\alpha}_A$ and $\hat{\beta}_A$ are not satisfactory. Algorithm 6.1 instead seems to lead to reasonable estimates. The estimates obtained by Method 1b stay the same after four iterations and the estimates obtained by Method 1a (minimize (6.4)) are also approximately the same (up to one decimal) after five iterations. Method 1b compared to Method 1a seems to produce slightly better results in this particular case. Of course, this simulation study is very limited and not clear conclusions can be drawn. However, Figure 6.1 clearly points out the improvement made in the estimation of the intensity, which was the main purpose of the adaptive method. It is clear that $\hat{\Lambda}_1(x_i)$ is closer to the 'true' intensity $\Lambda(x_i)$ than $\hat{\Lambda}(x_i)$ or $\hat{\Lambda}_A(x_i)$. The adaptive kernel estimation, which uses also mark information, performs well in the estimation of the high intensity values where the other kernel estimations are not so good. For $k = 5, 10$ iterations the relationship between $\hat{\Lambda}_k(x_i)$ and $m(x_i)$ is even closer to that of

$\Lambda(x_i)$ and $m(x_i)$, but this improvement does not seem to have a strong effect on the estimates of α and β anymore. The use of mark information decreases also correlation in the residuals $m(x_i) - \hat{\alpha}_t - \hat{\beta}_t \log(\hat{\Lambda}_t(x_i))$, where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the estimates obtained using $\hat{\Lambda}_t(x_i)$. The variance σ_U^2 could be estimated from these residuals.

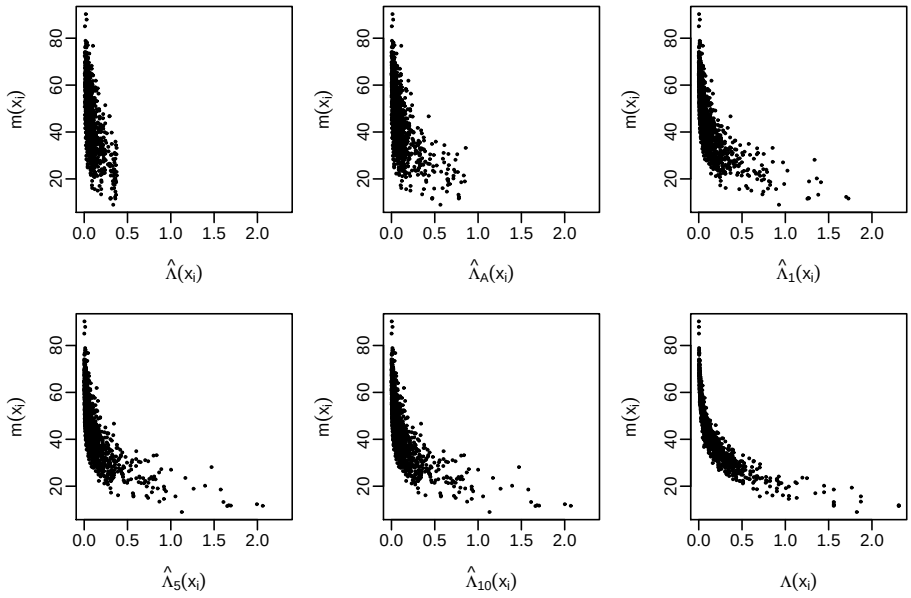

Figure 6.1: The marks $m(x_i)$ of the realisation of the log-intensity marked Cox process plotted against the estimated intensities $\hat{\Lambda}(x_i)$, $\hat{\Lambda}_A(x_i)$, $\hat{\Lambda}_1(x_i)$, $\hat{\Lambda}_5(x_i)$ and $\hat{\Lambda}_{10}(x_i)$ and the true intensity $\Lambda(x_i)$, $x_i \in N$.

6.2 Minimum contrast estimation

The minimum contrast estimation is a general procedure, see e.g. Diggle (1979), Diggle (1981), Møller et al. (1998), Diggle (2003, p. 86) and Illian et al. (2008, p. 451). It is also applicable for marked point processes. The idea is to minimize a discrepancy measure between a chosen (functional) theoretical characteristic $S(r;\theta)$ and its empirical version $\hat{S}(r)$ with respect to the unknown parameter vector θ. The minimization leads to an estimate for θ. It is demanded, that either a formula for $S(r;\theta)$ is known up to the model parameters or it can be obtained through simulation. Here the method is suggested to be applied in two steps to conditionally marked log Gaussian Cox processes.

More precisely, the suggested two-step method, called Method 3 in what follows, estimates first the parameters of the log Gaussian Cox process, and then the parameters of the marking equation conditional on the estimated log Gaussian Cox process. In the first step, the log Gaussian Cox process is estimated using existing methods: in this study the minimum contrast method (Møller et al., 1998; Møller and Waagepetersen, 2004; Illian et al., 2008) and an approximative likelihood method (Tanaka et al., 2008; see also Illian et al., 2008) are employed. The second step uses the minimum contrast method applied to the mark variogram.

One could apply the minimum contrast method to some other mark characteristic (see Table 2.1), but the choice of the mark variogram was made based on the following arguments: According to Schlather et al. (2004), there may be two different models with the same Stoyan's $k_{mm}(r)$-function even if their $C_m(r)$-functions differ. Therefore, $C_m(r)$ should be preferred to $k_{mm}(r)$. However, $\gamma_m(r) = V(r) - C_m(r)$ and also in geostatistics the covariance parameters are often estimated by fitting a variogram model instead of the covariance function. Of course, further studies would be necessary to justify the goodness of this choice. For us, the model fitting using the suggested minimum contrast method serves as a tool in identifying a suitable model for data. Then one can proceed with more advanced analysis using the Bayesian method introduced in Section 6.3.

The following section recalls the minimum contrast estimation for log Gaussian Cox processes and discusses also other estimation methods presented for log Gaussian Cox processes. Section 6.2.2 introduces the estimation of the marking parameters using the minimum contrast estimation and Section 6.2.3 modifies the method for the particular markings.

6.2.1 Estimation of log Gaussian Cox processes

For the estimation of the parameters of a log Gaussian Cox process, Møller et al. (1998) recommends the use of a minimum contrast estimation applied to the covariance function $C_Z(r)$ (see also Illian et al., 2008): If $\hat{C}_Z(\cdot)$ is a non-parametric estimator of the covariance function and $C_Z(r;\theta)$ a parametric model for the covariance function, then minimize

$$\int_{\epsilon}^{a_0} \left\{ \hat{C}_Z(r)^{\eta} - C_Z(r;\theta)^{\eta} \right\}^2 \mathrm{d}r \tag{6.10}$$

with respect to the parameter vector θ, including the variance σ_Z^2, using values $0 \leq \epsilon < a_0$ and $\eta > 0$ specified by the user. In computation, the integral (6.10) is approximated by a Riemann sum. Further, the intensity is estimated by

$$\hat{\lambda} = \frac{n}{\nu_d(W)}, \tag{6.11}$$

where n is the number of points in an observation window $W \subset \mathbb{R}^d$ and $\nu_d(W)$ is the volume of W. An estimate for the mean of the random field $\{Z(s)\}$ is obtained thereafter by

$$\hat{\mu}_Z = \log(\hat{\lambda}) - \frac{\hat{\sigma}_Z^2}{2}, \tag{6.12}$$

where $\hat{\sigma}_Z^2$ is the minimum contrast estimator of the variance.

It is also possible to use $L(r)$-function (or another cumulative characteristic) in parameter estimation. Guan and Sherman (2007) studies the minimum contrast estimation in terms of Ripley's K-function. In our simulation experiments, following the recommendation in Illian et al. (2008, p. 451), we use the (non-cumulative) $g(r)$-function (or $C_Z(r)$).

There are other competitive alternatives for the estimation of the log Gaussian Cox process. One can use composite likelihood method (Guan, 2006), approximative maximum likelihood method (Tanaka et al., 2008; see also Illian et al., 2008) or Bayesian method (Møller and Waagepetersen, 2004). In this work, the focus is in the estimation of the conditional mark distribution, and

thereby, the minimum contrast method, being a computationally easy approach, is applied to the log Gaussian Cox process. We also utilise the approximative maximum likelihood method in Section 6.2.3. In this estimation, the strategy explained in Tanaka et al. (2008) and Illian et al. (2008, p. 448) is used to estimate the parameters of $C_Z(r;\theta)$, and thereafter, estimates for λ and μ_Z are obtained from (6.11) and (6.12), respectively.

6.2.2 Minimum contrast estimation applied to the mark variogram

The parameters of the marking equations are estimated by minimizing the integral

$$\int_{\epsilon}^{a_0} \{\hat{\gamma}_m(r)^\eta - \gamma_m(r;\theta_M)^\eta\}^2 \,\mathrm{d}r \tag{6.13}$$

with respect to the mark parameter vector θ_M. Here the empirical variogram $\hat{\gamma}_m(r)$ is compared with its (parametric) theoretical counterpart $\gamma_m(r;\theta_M)$ known up to the marking parameters, and $0 \leq \epsilon < a_0$ and $\eta > 0$ are again values specified by the user. The parameters of the log Gaussian Cox process in $\gamma_m(r;\theta_M)$ are considered fixed (known) and, especially, $C_Z(r)$ is approximated either by a non-parametric estimate $\hat{C}_Z(r)$ or by a parametric estimate $C_Z(r;\hat{\theta})$. The latter is, for example, the exponential covariance function $C_Z(r;\theta) = \sigma_Z^2 \exp\{-r/\phi_Z\}$ with estimates of $\theta = (\sigma_Z^2, \phi_Z)$ plugged in, but this choice depends on the data to be analysed.

6.2.3 A simulation experiment

A simulation experiment is here performed for illustrating the parameter estimation methods of Sections 6.1.2 and 6.2.2. Particularly, the objective is to clarify how the Method 3, that is the minimum contrast method based on (6.13), can be implemented for specific markings. For comparison, also Methods 1 and 2 (see Section 6.1.2) are applied to simulated data. A more comprehensive simulation study for conditionally marked models is performed in Section 6.4.

One realisation of a log Gaussian Cox process $N = \{x_i\}$ is simulated in a window of size 200×200 in Example 4.1. Further, in Examples 4.1, 5.1 and 5.2, the point process is marked by the following marking models from 1 to 4:

$$\begin{aligned}
m^{(1)}(x_i)|\,\Lambda(x_i) &\sim \mathrm{Exp}(1/(30 + 0.8/\Lambda(x_i))) \\
m^{(2)}(x_i)|\,\Lambda(x_i) &\sim \mathrm{Gamma}(0.5, 1/(30 + 0.8/\Lambda(x_i))) \\
m^{(3)}(x_i)|\,\Lambda(x_i) &\sim N(20 + 0.2/\Lambda(x_i), 0.5^2/\Lambda(x_i) + 2^2) \\
m^{(4)}(x_i)|\,\Lambda(x_i) &\sim N(20 - 10\log(\Lambda(x_i)), 10).
\end{aligned}$$

The obtained marked point patterns are used as synthetic data sets.

First a log Gaussian Cox process is fitted to the simulated data. The exponential covariance function is fitted by means of the minimum contrast method (for short MinC) using the argument range $(0.25, 25.00)$ and $\eta = 1/2$ as in Møller et al. (1998). These estimates and, additionally, those obtained by the approximative maximum likelihood method (for short AML) are shown in Table 6.3. These estimates are very similar, but differ slightly from the ones used in simulation. The estimates affect the estimation of the mark parameters by Method 3.

Table 6.3: Estimation results for the log Gaussian Cox process simulation.

	$\hat{\sigma}_Z^2$	$\hat{\phi}_Z$	$\hat{\mu}_Z$	$\hat{\lambda}$
MinC	1.35	7.20	-3.94	0.038
AML	1.32	7.48	-3.93	0.038
Value used in simulation	1.50	6.00	-4.00	0.039

The estimates $\hat{\Lambda}(x_i)$ for $x_i \in N$ are obtained by (6.1) with Epanechnikov kernel and bandwidth 10. These are used in the estimation by Methods 1 and 2 and in the cusp-point method for Marking 3 (explained below). Note that Method 1 is especially considered for Marking 4 in Section 6.1.3.

Marking 1. The parameters a and b of the exponential intensity-marked Cox process are estimated by minimizing (6.13) with $\alpha = 1$, $\epsilon = 0.25$, $a_0 = 30$ and $\eta = 1$. Minimization is performed both with the parameters of the log Gaussian Cox process (for short LGCP) obtained by the minimum contrast method and by the approximative maximum likelihood, and both with $\hat{C}_Z(r)$ and $C_Z(r;\hat{\theta})$. In this experiment, $C_Z(r;\hat{\theta})$ is very close to $\hat{C}_Z(r)$ and, consequently, all the estimates are almost the same. Changing $\eta = 1$ to $\eta = 1/2$ does not affect the estimates much in this case. The estimates of Methods 2, 1a and 1b are obtained by maximizing (6.6) and by minimizing (6.4) and (6.5), respectively (all with $\alpha = 1$). All the estimates are shown in Table 6.4.

Table 6.4: Estimates for the mark parameters of the exponential marking.

	$\hat{a}$	$\hat{b}$	Model for points	Estimation method
Method 1a: minimize (6.4)	29.67	0.97	-	
Method 1b: minimize (6.5)	26.48	1.05	-	
Method 2: conditional ML	30.86	0.91	-	
Method 3a: MinC with $\hat{C}_Z(r)$	33.21	0.79	LGCP	MinC
Method 3b: MinC with $C_Z(r;\hat{\theta})$	32.33	0.80	LGCP	MinC
Method 3c: MinC with $\hat{C}_Z(r)$	33.18	0.80	LGCP	AML
Method 3d: MinC with $C_Z(r;\hat{\theta})$	32.14	0.83	LGCP	AML
Value used in simulation	30.00	0.80		

Theoretical mark characteristics with estimated parameters have been compared visually by plotting them with their empirical counterparts. The estimation based on the mark variogram (Method 1) seems to give the best fit with respect to mark characteristics of Table 2.1. Note that Methods 1a and 1b (and 2, not shown in the figure) do not utilize $C_Z(r,\hat{\theta})$ in estimation.

Marking 2. Assuming the shape parameter α to be known, the gamma intensity-marked Cox process is estimated in the same way as the exponential model with the difference that now $\alpha = 0.5$. These estimates are shown in Table 6.5. The estimation results have again been

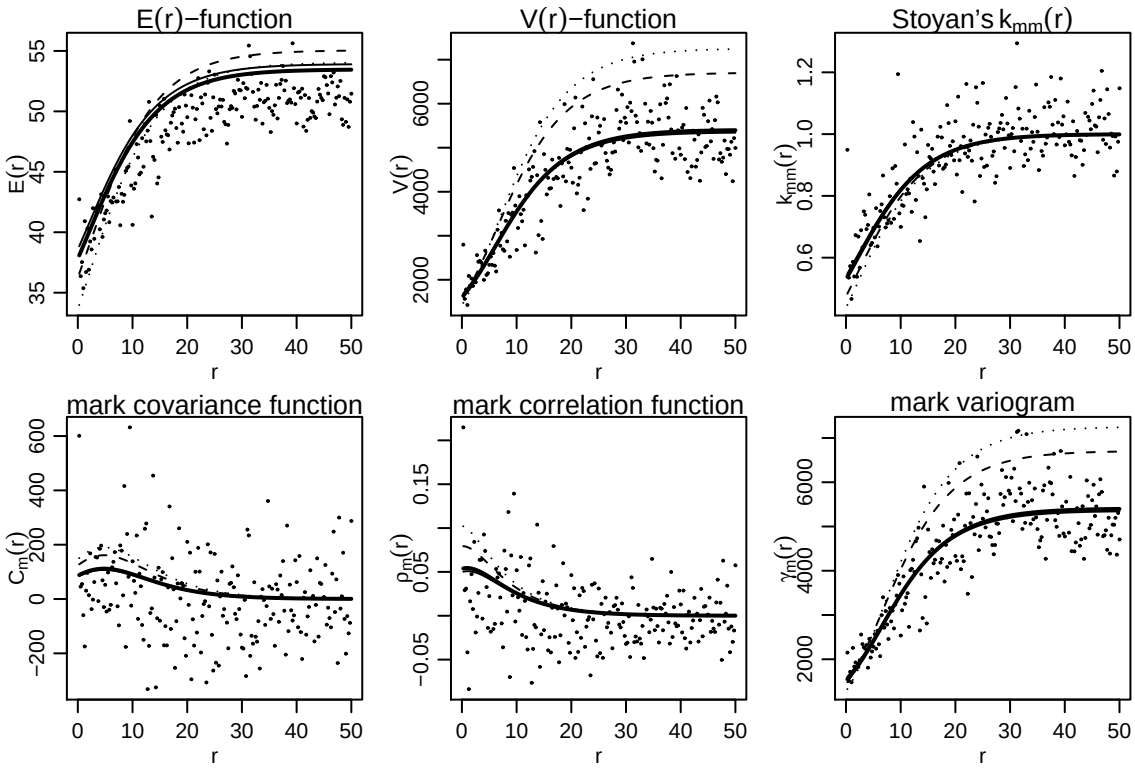

Figure 6.2: Mark summaries (*dots*) calculated from the realisation of the exponential intensity-marked Cox process of Figure 5.3 and their theoretical counterparts with estimated parameters $\hat{a}$, $\hat{b}$ and $C_Z(r;\hat{\theta})$. The *thick solid line* corresponds to Method 3b, the *thin solid line* to Method 3a, the *dashed line* to Method 1a and the *dotted line* to Method 1b.

compared visually in terms of the second-order mark characteristics of Table 2.1, and Method 3 (a, b, c, d) seems to give the best fit in this experiment, see Figure 6.3.

Table 6.5: Estimates for the mark parameters of the gamma marking with fixed $\alpha = 0.5$.

	$\hat{a}$	$\hat{b}$	Model for points	Estimation method
Method 1a: minimize (6.4)	31.71	0.84	-	
Method 1b: minimize (6.5)	41.20	0.60	-	
Method 2: conditional ML	24.63	1.21	-	
Method 3a: MinC with $\hat{C}_Z(r)$	30.24	0.70	LGCP	MinC
Method 3b: MinC with $C_Z(r;\hat{\theta})$	28.93	0.73	LGCP	MinC
Method 3c: MinC with $\hat{C}_Z(r)$	30.19	0.71	LGCP	AML
Method 3d: MinC with $C_Z(r;\hat{\theta})$	28.81	0.75	LGCP	AML
Value used in simulation	30.00	0.80		

Marking 3. For the marking (5.5), direct optimization of the integral (6.13) is problematic and, in addition, the model parameter a does not exist in the equation. A more stable solution,

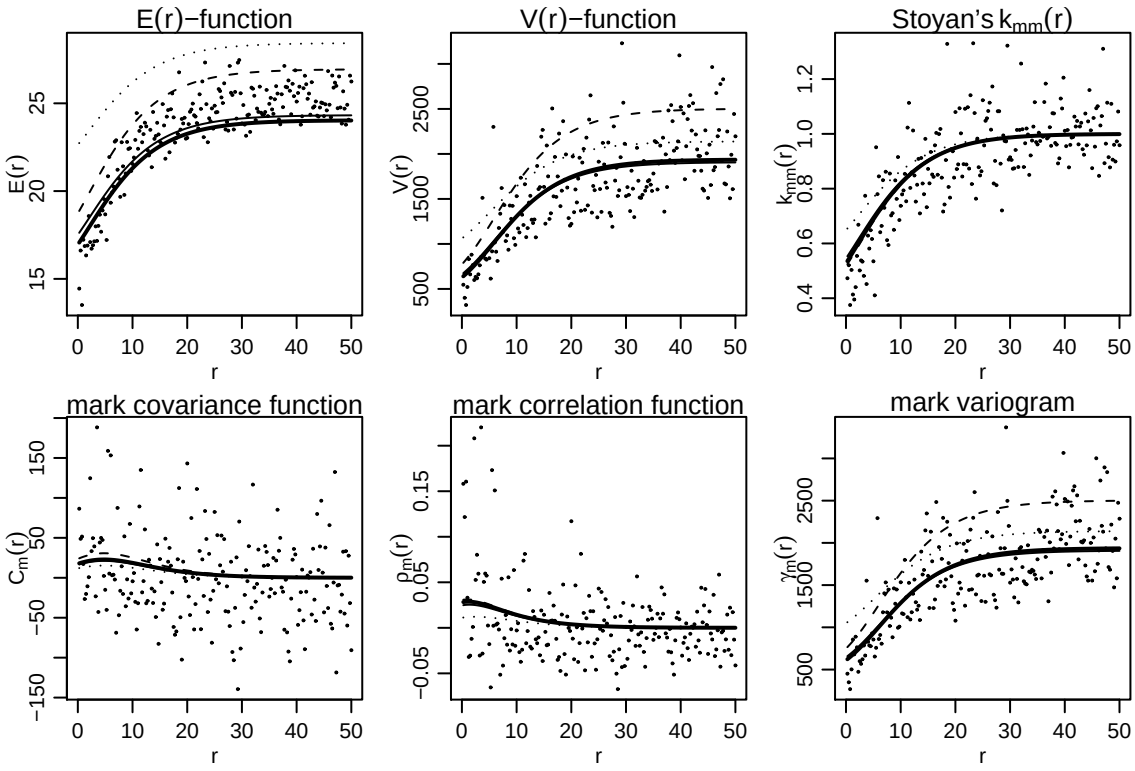

Figure 6.3: Mark summaries (*dots*) calculated from the realisation of the gamma intensity-marked Cox process of Figure 5.5 ($\alpha = 0.5$) and their theoretical counterparts with estimated parameters $\hat{a}$, $\hat{b}$ and $C_Z(r;\hat{\theta})$. The *thick solid line* corresponds to Method 3b, the *thin solid line* to Method 3a, the *dashed line* to Method 1a and the *dotted line* to Method 1b.

and also an estimate for a, is achieved by first estimating the parameters a and d^2 as the mean and variance of the marks in regions with high intensity. The point locations in high intensity can be determined by a cusp-point method: Calculate the mean and the variance of the marks in $H_h = \{[x_i; m(x_i)] \in N_m : \hat{\Lambda}(x_i) \geq h\}$ for several values of h. Then, choose a cusp-point $h = h_0$ for which the mean and variance are approximately constant for values $h \geq h_0$. The mean and variance of marks in H_{h_0} yield the estimates. Here $h_0 = 0.25$ gives the following estimates for a and d (value used in simulation are given in parentheses): $\hat{a} = 20.32$ (20.00) and $\hat{d} = 1.87$ (2.00).

Considering $\hat{d}^2$ as a fixed value for d^2 (and $\hat{a}$ for a), the parameters b and c are estimated by minimizing (6.13) with $\epsilon = 0.25$, $a_0 = 30$ and $\eta = 1/2$ (Method 3). Further, the conditional maximum likelihood method (Method 2) is applied. In principle, this method can be used for the marking (5.5) as such to obtain estimates for a, b, c and d if good estimates for $\Lambda(x_i)$, $x_i \in N$, exist. However, it seems, according to our limited experience, that better estimates are obtained if the estimates $\hat{a}$ and $\hat{d}$ obtained by the cusp-point method are used and (6.6) is maximized with fixed a and d to obtain estimates for b and c. This is a consequence of sensitivity to the obtained values of intensity. The estimates of b and c obtained by Methods 2 and 3 are shown in Table 6.6.

Again mark summaries have been compared visually with their theoretical counterparts of Table 2.1. With respect to these characteristics, Method 3 seems to give the best estimates for b and

c, see Figure 6.4. However, with this simulated data, the estimates closest to the ones used in simulation are obtained by Method 2.

Table 6.6: Estimates for the parameters b and c of the Gaussian marking.

	$\hat{b}$	$\hat{c}$	Model for points	Estimation method
Method 2: conditional ML	0.19	0.51	-	
Method 3a: MinC with $\hat{C}_Z(r)$	0.20	0.76	LGCP	MinC
Method 3b: MinC with $C_Z(r;\hat{\theta})$	0.21	0.69	LGCP	MinC
Method 3c: MinC with $\hat{C}_Z(r)$	0.21	0.78	LGCP	AML
Method 3d: MinC with $C_Z(r;\hat{\theta})$	0.22	0.70	LGCP	AML
Value used in simulation	0.20	0.50		

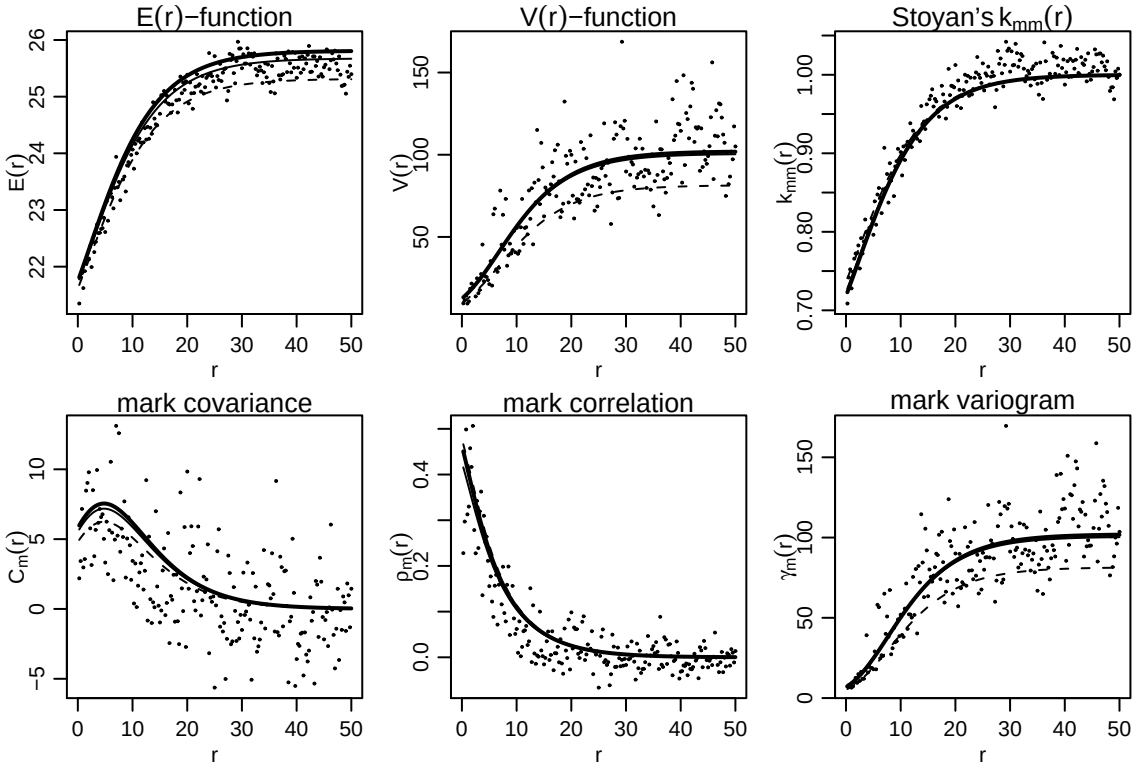

Figure 6.4: Mark summaries (*dots*) calculated from the realisation of the Gaussian intensity-marked Cox process of Figure 5.1 and their theoretical counterparts with estimated parameters $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{d}$ and $C_Z(r;\hat{\theta})$. The *thick solid line* corresponds to Method 3b, the *thin solid line* to Method 3a and the *dashed line* to Method 2.

Similar estimation procedure may be considered for the Gaussian intensity-marked Cox process (5.3), but the parameters a and d^2 must be estimated as the mean and variance of the marks in the regions with low intensity.

Marking 4. The parameters of Marking 4 to be estimated are α, β and σ_U^2. The mark variogram (4.6) for this model with uncorrelated errors is

$$\gamma_m(r;\theta_M) = \beta^2\sigma_Z^2 + \sigma_U^2 - \beta^2 C_Z(r) \quad \text{for } r > 0$$

and, therefore, the parameters β and σ_U^2 can be estimated by Method 3. In addition, the parameter α must be estimated. This estimation can be conducted through the mean mark or the $E(r)$-function, for example. The former uses first-order information, whereas the latter utilizes the second-order distributional property of marks. That is, the mean mark is estimated by

$$\hat{\mu}_m = \frac{1}{n}\sum_{i=1}^{n} m(x_i),$$

and the obtained estimate is then placed into the formula $\mu_m = \alpha + \beta\mu_Z + \beta\sigma_Z^2$. Then

$$\hat{\alpha} = \hat{\mu}_m - \hat{\beta}(\hat{\mu}_Z + \hat{\sigma}_Z^2) \tag{6.14}$$

is obtained. Alternatively, an estimate for α is obtained by minimizing

$$\int_{\epsilon}^{a_0} \left\{\hat{E}(r)^\eta - E(r;\alpha)^\eta\right\}^2 \mathrm{d}r \tag{6.15}$$

where $0 \le \epsilon < a_0$ and $\eta > 0$ are parameters specified by the user and $E(r;\alpha) = \alpha + \hat{\beta}(\hat{\mu}_Z - \hat{\sigma}_Z^2 + \tilde{C}_Z(r))$, where $\tilde{C}_Z(r) = \hat{C}_Z(r)$ or $\tilde{C}_Z(r) = C_Z(r;\hat{\theta})$.

The integral (6.13) with $\epsilon = 0.25$, $a_0 = 30$ and $\eta = 1$ is minimized. These estimates are shown in Table 6.7. The estimate $\hat{\beta} = -10.50$ is further used in estimation of α by (6.14) and (6.15). Since $\hat{\mu}_m = 45.89$, (6.14) gives $\hat{\alpha} = 18.65$, and minimization of (6.15) yields estimates $\hat{\alpha} = 18.98$ and $\hat{\alpha} = 18.46$ with $\tilde{C}_Z(r) = \hat{C}_Z(r)$ and $\tilde{C}_Z(r) = C_Z(r;\hat{\theta})$, respectively. (The true value used in simulation is $\alpha = 20$.) The fit with respect to mark characteristics is shown for parameter values $\hat{\beta} = -10.50$, $\hat{\sigma}_U^2 = 16.03$ (obtained by Method 3b) and $\hat{\alpha} = 18.46$ in Figure 6.5.

Table 6.7: Estimates for the parameters β and σ_U^2 of the log-intensity marked Cox process.

	$\hat{\beta}$	$\hat{\sigma}_U^2$	Model for points	Estimation method
Method 3a: MinC with $\hat{C}_Z(r)$	-10.50	15.87	LGCP	MinC
Method 3b: MinC with $C_Z(r;\hat{\theta})$	-10.50	16.03	LGCP	MinC
Method 3c: MinC with $\hat{C}_Z(r)$	-10.50	19.45	LGCP	AML
Method 3d: MinC with $C_Z(r;\hat{\theta})$	-10.60	17.94	LGCP	AML
Value used in simulation	-10.00	10.00		

6.3 Bayesian modelling

Assume that a realisation of an intensity-dependently marked log Gaussian Cox process, $\{[x_i; m(x_i)]; i = 1, \ldots, n\}$, denoted by $[X, M]$, is observed in a window $W \subset \mathbb{R}^2$. (A generalization to $\mathbb{R}^d$ can be obtained straightforwardly.) The objective is to fit the model to the observed data $[X, M]$, the interest being especially in the marking.

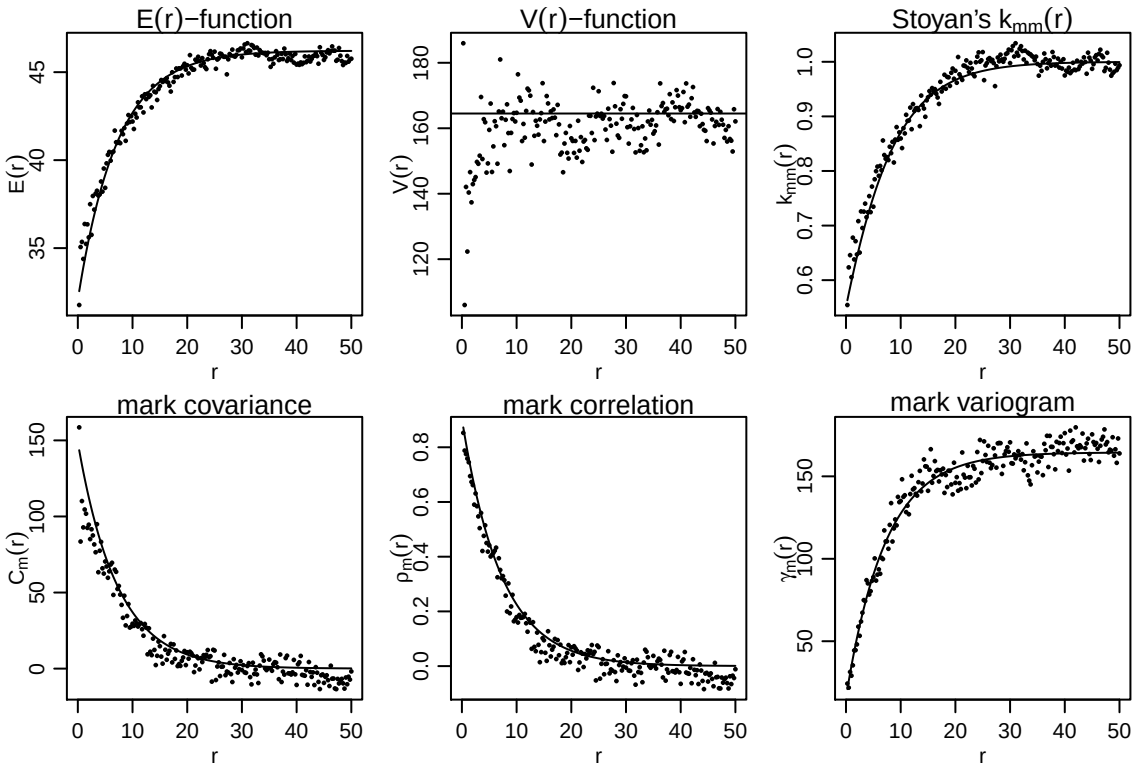

Figure 6.5: Mark summaries (*dots*) calculated from the realisation of the log-intensity marked Cox process of Figure 4.2 and their theoretical counterparts (*solid lines*) with estimated parameters $\hat{\alpha} = 18.46$, $\hat{\beta} = -10.50$, $\hat{\sigma}_U^2 = 16.03$ and $C_Z(r; \hat{\theta})$ ($\hat{\theta}$ obtained by MinC).

In the marking models of Sections 4 and 5, the marks depend on the underlying (Cox) point process though its (random) intensity $\{\Lambda(s)\}$. This intensity is assumed to be log Gaussian, that is $\Lambda(s) = \exp(Z(s))$, and it is characterized by the mean and the covariance function of the Gaussian random field $\{Z(s)\}$. The mean and the covariance property affect the values of the intensity, of course, but the intensity itself contains large variation. Taking this into account, we suggest an empirical Bayesian estimation method where the mean and the parameters of covariance function are estimated from the point data. These parameters are hyperparameters of the conditionally marked log Gaussian Cox model. Markov chain Monte Carlo (MCMC) methods are then used for posterior simulation of the intensity and the marking parameters. The mean and covariance parameters of $\{Z(s)\}$, that is the parameters of a log Gaussian Cox process, can be estimated using one of the known methods, see Section 6.2.1 and references there.

We experimented with a fully Bayesian method for the intensity-dependently marked log Gaussian Cox process, but this approach seems to be unstable. Møller and Waagepetersen (2002, p. 53), Møller and Waagepetersen (2003, p. 168) and Møller and Waagepetersen (2004) report disadvantages also in the estimation of the (non-marked) log Gaussian Cox process. The marked model may be over-parametrized due to the covariance parameter controlling the smoothness of the random field and due to the range parameter. This leads to unstable posterior MCMC computation. The problem is bypassed by estimating the covariance function from point data, which essentially leads to an empirical Bayesian procedure.

Conditional on $\{Z(s)\}$ or $\{\Lambda(s)\}$, the likelihood of the log Gaussian Cox process is the one of an inhomogeneous Poisson process,

$$p(\{x_i : i = 1, \ldots, n\} | \{\Lambda(s)\}) = \left[\prod_{i=1}^{n} \Lambda(x_i)\right] \exp\left\{-\int_W \Lambda(s)\mathrm{d}s\right\}, \tag{6.16}$$

see e.g. Daley and Vere-Jones (2003). The conditional likelihood for the conditionally marked log Gaussian Cox process is

$$p([X, M] |\, \theta_M, \{Z(s)\}) = \left[\prod_{i=1}^{n} \left[\Lambda(x_i) f_{m|\Lambda}(m(x_i) |\, \Lambda(x_i))\right]\right] \exp\left\{-\int_W \Lambda(s)\mathrm{d}s\right\}, \tag{6.17}$$

where $\Lambda(s) = \exp(Z(s))$, $f_{m|\Lambda}(m(x_i) |\, \Lambda(x_i))$ is the parametric density of the conditional mark distribution of the mark $m(x_i)$ at $x_i \in N$ given the intensity $\Lambda(x_i)$, and θ_M is the parameter vector of this density.

The prior for the marking parameters θ_M must be set according to the marking model and the application; some alternatives are presented in Sections 6.4 and 6.5. The prior distribution of $\{\Lambda(s)\}$ is determined by the model choice, which here implies that $\Lambda(s) = \exp(Z(s))$ for $s \in W$, where $\{Z(s)\}$ is a Gaussian random field with mean μ_Z and covariance function $C_Z(r)$.

In an empirical Bayesian approach, μ_Z and $C_Z(r)$ are estimated from the point data using one of the well-known methods mentioned above. The estimation of $C_Z(r)$ provides, that a suitable parametrized covariance function $C_Z(r)$ is chosen: in our simulation experiments and the rainforest application the Matérn covariance function

$$C_Z(r) = \sigma_Z^2 \cdot 2^{1-\nu}\Gamma(\nu)^{-1}(r/\phi_Z)^\nu K_\nu(r/\phi_Z) \quad \text{for } r > 0,\ \nu > 0, \tag{6.18}$$

where K_ν is the modified Bessel function, is employed. It defines a generic class of Gaussian random field models, where the parameter ν controls the roughness of the realisations of random fields. Other parameters are the variance σ_Z^2 and the scale parameter ϕ_Z. The latter controls the range of correlation. These parameters are estimated from the point data.

In the implementation, $\{Z(s)\}$ is approximated by a discrete Gaussian random field defined on a partition of W into disjoint rectangular sets of equal size A, called pixels, with centre points $s_1, \ldots, s_k$ forming a grid $\mathcal{G}$ on W. The values of $\{Z(s)\}$ on $\mathcal{G}$ are needed in the MCMC simulation for approximating the integral $\int_W \Lambda(s)\mathrm{d}s$ in the likelihood (6.17):

$$\int_W \Lambda(s)\mathrm{d}s = \int_W e^{Z(s)}\mathrm{d}s \approx A\sum_{i=1}^{k} e^{Z(s_j)}. \tag{6.19}$$

Because the marks depend on the intensity through $\Lambda(x_i)$, $i = 1, \ldots, n$, the Gaussian random field is also considered at the points x_i in the MCMC simulation. Note that the values $Z(x_i)$ at the observed points x_i are not used in (6.19), because $\{Z(x_i) : i = 1, \ldots, n\}$ is not an unbiased sample of $\{Z(s)\}$. Therefore, let $Z = (Z(x_1), \ldots, Z(x_n), Z(s_1), \ldots, Z(s_k))$. This vector Z replaces $\{Z(s)\}$ in posterior simulation. The prior distribution of Z, denoted hereafter by $p(Z)$, is the $(n + k)$-dimensional normal distribution with mean μ_Z and covariance matrix Σ_Z defined by $C_Z(r)$ and the distances between the locations $x_1, \ldots, x_n, s_1, \ldots, s_k$. The parameters of Z are fixed, hence the covariance matrix is fixed over simulations, too.

The latent vector Z is included in the set of unknowns and its values will be updated in the MCMC runs. Therefore, the posterior to be simulated is

$$p(\theta_M, Z|[X, M]) \propto p(\theta_M)p(Z)p([X, M]|\,\theta_M, Z)\,.$$

We update the parameters in the order θ_M, Z, where each value of θ_M and Z is updated one by one using Metropolis-Hastings steps. This requires specification of the proposal distributions for θ_M and $Z(\cdot)$s. Thereafter, the acceptance probabilities can be easily calculated. Alternatively, the Langevin-Hastings algorithm (see e.g. Møller et al., 1998, Møller and Waagepetersen, 2004) or block-updates (Rue and Held, 2005) could be used for updating the values of Z. These may lead to faster algorithms.

The Bayesian estimation results in the posterior distribution for the unknowns. In our implementation, the values of the marking parameters are stored, whilst $Z(x_i)$ and $Z(s_j)$ are recorded for examination only with two different values of i and j sampled from $\{1, 2, \ldots, n\}$ and $\{1, 2, \ldots, k\}$, respectively.

Because the intensity surface is usually assumed to be smooth and the marks depend on $\{\Lambda(s)\}$ only at $x_i, i = 1, \ldots, n$, we conclude that the grid $\mathcal{G}$ may not need to be very dense. However, if the objective is to predict the unobserved Gaussian random field or the intensity process, a dense grid is recommended. In this sense, our empirical Bayesian approach extends the method presented by Møller et al. (1998, p. 472) by using information also about marks in the calculation of the posterior distribution of the intensity field. An approximation of the intensity surface can be obtained as the posterior mean of $\exp(Z(s_j))$, $j = 1, \ldots, k$, from the posterior simulation.

6.4 A simulation study

Six realisations of a log Gaussian Cox process are simulated in a window W of size $[0, 200] \times [0, 200]$ using variable parameter combinations. These point processes are further marked using different intensity-dependent marking models with varying parameters. The obtained marked point patterns are used as synthetic data sets and their parameters are estimated using the empirical Bayesian approach of Section 6.3 to demonstrate the functionality of posterior computation.

We consider a Gaussian random field $\{Z(s)\}$ with mean $\mu_Z = -4.5$ and Matérn covariance function (6.18) with variance $\sigma_Z^2 = 1$ and values $\nu = 0.5, 1.0, 1.5$ of the roughness parameter. In addition, the values $\phi_Z = 5$ and $\phi_Z = 10$ are addressed to the scale parameter ϕ_Z in (6.18). The expected number of points in each point process is 733. The point processes are marked by the marking models (5.8) and (5.5). We consider the parameter values $a = 30$ and $b = 0.5, 1.0$, totally 12 marked point patterns, for the exponential marking (5.8). For the Gaussian marking (5.5), the following parameter values are considered: $a = 30$, $b = 0.2, 0.5$, $c^2 = 0.25, 1.0$ and $d^2 = 16$. This study design leads to 24 different Gaussian intensity-marked processes.

Assuming that N_m is observed, the parameters of the intensity-marked log Gaussian Cox processes are estimated using the method described in Section 6.3. First, the parameters σ_Z^2 and ϕ_Z are estimated from the point data using the minimum contrast method: the equation (6.10) is minimized with $\epsilon = 0.5$, $a_0 = 30$, $\eta = 1$ and fixed ν (which is assumed known) and an estimate for μ_Z is obtained thereafter by (6.12). The estimated mean and covariance function

are then used in the prior distribution of Z. Second, the intensity and the mark parameters are estimated using the Bayesian approach. For evaluating the sensitivity of the proposed method to the parameters of the log Gaussian Cox process, the (fully) Bayesian analysis for estimating a and b is performed also with the (known) true mean and covariance of Z. That is, the parameter values $(\mu_Z, \sigma_Z^2, \phi_Z)$ used in simulation enter the prior of Z.

The prior distributions for θ_M are set according to the marking model. We consider the following prior distributions for the exponential marking (5.8): $a \sim N(30, 100^2)$, $b \sim \text{Gamma}(A_b, B_b)$, a and b independent. The values $A_b = 0.0625, 0.25$ and $B_b = 0.125, 0.25$ are used corresponding to $b = 0.5, 1.0$, respectively. By this prior choice, the parameter b is restricted to be positive (but a is not). For the Gaussian marking (5.5), the following prior distributions are considered: $a \sim N(30, 10^2)$, $b \sim \text{Gamma}(A_b, B_b)$, $c^2 \sim \text{Unif}(0, 3)$ and $d^2 \sim \text{Gamma}(16, 1)$, where the values $A_b = 0.04, 0.25$ and $B_b = 0.2, 0.5$ are used corresponding to $b = 0.2, 0.5$, respectively. It is assumed that a, b, c^2 and d^2 are independent a priori.

The proposal distributions must be specified. Let a, b, c^2, d^2 and Z be the current values in the MCMC chains and let * refer to the proposal. The following proposal distributions are considered for the exponential model:

$$\begin{aligned} a^* &\sim N(a, \sigma_{aq}^2), \\ b^* &\sim \text{Unif}(\max(0, b - \delta_b), b + \delta_b), \\ Z^*(\cdot) &\sim N(Z(\cdot), \sigma^2), \end{aligned}$$

where $\sigma_{aq} = 5.0, \delta_b = 0.3$, and $\sigma = 1.5, 1.1, 0.7$ corresponding to $\nu = 0.5, 1.0, 1.5$, respectively. In addition, we use the proposal distributions

$$\begin{aligned} c^{2^*} &\sim \text{Unif}(\max(0, c^2 - \delta_c), c^2 + \delta_c), \\ d^{2^*} &\sim \text{Unif}(\max(0, d^2 - \delta_d), d^2 + \delta_d) \end{aligned}$$

for the Gaussian marking (5.5). The following parameter values are used in the proposal distributions for the Gaussian marking: $\sigma_{aq} = 1$, $\delta_b = 0.05$, $\delta_d = 6.0$, $\delta_c = 0.2, 0.5$ corresponding to $c^2 = 0.25, 1.00$, and $\sigma = 1.5, 1.1, 0.7$ corresponding to $\nu = 0.5, 1.0, 1.5$, respectively. For approximating the integral in the likelihood (6.17), the process $\{Z(s)\}$ is considered on $\mathcal{G}$. We employ 20×20 and 40×40 grids.

The MCMC algorithm is run for 50,000 iterations in each case, except for the Gaussian random field model with $\nu = 1.5$ and $\phi_Z = 10$ a total of 100,000 iterations is performed. According to our experimenting, the convergence of the MCMC algorithm becomes slower the smoother the random field is. Initial values of a and b for the exponential marking are simulated from the uniform distributions on the intervals $(5, 50)$ and $(0.1, 2)$, respectively, and for the Gaussian marking, the initial values of a, b, c^2 and d^2 are simulated from their corresponding prior distributions. The initial values of Z are simulated from the normal distribution with mean μ_Z and covariance matrix Σ_Z determined by the covariance model. These simulations of the Gaussian random field are performed using the function `GaussRF` in the R library `RandomFields`, see Schlather (2001b). The marginal posterior distributions of the marking parameters are described by means of their means and standard deviations. These descriptions calculated from the last 20,000 iterations are shown for the exponential marking in Tables 6.8 and 6.9 and for the Gaussian marking in Table 6.10. The results obtained using the 20×20 grid are omitted for the Gaussian marking.

Table 6.8: Results for the exponential marking with a 40×40 grid. Posterior means (standard deviations) for the marking parameters a and b. The third column shows the parameter values used in the prior of Z (first line: estimated; second line: values used in simulation). The value $a = 30$ is used in the simulation.

		$\mu_Z, \sigma_Z^2, \phi_Z$	$\hat{a}$	$\hat{b}$
$\nu = 0.5$	$b = 0.5$	-4.34, 0.90, 4.67	30.19 (2.95)	0.63 (0.09)
		-4.50, 1.00, 5.00	31.18 (3.02)	0.59 (0.09)
	$b = 1.0$	-4.34, 0.90, 4.67	36.35 (4.20)	0.92 (0.13)
		-4.50, 1.00, 5.00	38.31 (4.06)	0.83 (0.13)
	$b = 0.5$	-4.44, 0.81, 12.46	28.46 (3.03)	0.46 (0.08)
		-4.50, 1.00, 10.00	29.84 (2.91)	0.44 (0.07)
	$b = 1.0$	-4.44, 0.81, 12.46	31.92 (4.26)	1.06 (0.13)
		-4.50, 1.00, 10.00	34.91 (4.13)	1.02 (0.13)
$\nu = 1.0$	$b = 0.5$	-4.54, 0.90, 5.70	35.80 (3.28)	0.36 (0.07)
		-4.50, 1.00, 5.00	35.92 (3.17)	0.37 (0.07)
	$b = 1.0$	-4.54, 0.90, 5.70	39.40 (4.56)	0.74 (0.11)
		-4.50, 1.00, 5.00	40.52 (4.58)	0.72 (0.11)
	$b = 0.5$	-4.76, 0.74, 8.28	30.37 (4.76)	0.55 (0.09)
		-4.50, 1.00, 10.00	32.06 (4.86)	0.51 (0.09)
	$b = 1.0$	-4.76, 0.74, 8.28	20.91 (7.30)	1.14 (0.16)
		-4.50, 1.00, 10.00	24.18 (6.82)	1.18 (0.16)
$\nu = 1.5$	$b = 0.5$	-4.36, 0.83, 5.67	32.05 (2.83)	0.48 (0.07)
		-4.50, 1.00, 5.00	32.23 (2.87)	0.49 (0.08)
	$b = 1.0$	-4.36, 0.83, 5.67	27.22 (3.78)	0.97 (0.11)
		-4.50, 1.00, 5.00	29.14 (3.51)	0.97 (0.12)
	$b = 0.5$	-4.38, 0.82, 7.83	33.20 (2.65)	0.28 (0.05)
		-4.50, 1.00, 10.00	31.93 (3.04)	0.30 (0.06)
	$b = 1.0$	-4.38, 0.82, 7.83	28.28 (3.98)	0.91 (0.10)
		-4.50, 1.00, 10.00	29.40 (3.76)	0.75 (0.08)

Table 6.9: Results for the exponential marking with a 20×20 grid. Posterior means (standard deviations) for the marking parameters a and b. The third column shows the parameter values used in the prior of Z (first line: estimated; second line: values used in simulation). The value $a = 30$ is used in the simulation.

		$\mu_Z, \sigma_Z^2, \phi_Z$	$\hat{a}$	$\hat{b}$
$\nu = 0.5$	$b = 0.5$	-4.34, 0.90, 4.67	33.99 (2.76)	0.82 (0.13)
		-4.50, 1.00, 5.00	35.24 (2.70)	0.79 (0.12)
	$b = 1.0$	-4.34, 0.90, 4.67	41.85 (3.56)	1.18 (0.17)
		-4.50, 1.00, 5.00	43.70 (3.54)	1.12 (0.17)
	$b = 0.5$	-4.44, 0.81, 12.46	32.22 (2.73)	0.50 (0.09)
		-4.50, 1.00, 10.00	34.17 (2.57)	0.52 (0.09)
	$b = 1.0$	-4.44, 0.81, 12.46	38.73 (3.92)	1.23 (0.16)
		-4.50, 1.00, 10.00	43.26 (3.87)	1.28 (0.18)
$\nu = 1.0$	$b = 0.5$	-4.54, 0.90, 5.70	38.06 (2.92)	0.40 (0.08)
		-4.50, 1.00, 5.00	39.95 (2.84)	0.39 (0.08)
	$b = 1.0$	-4.54, 0.90, 5.70	45.17 (4.05)	0.76 (0.12)
		-4.50, 1.00, 5.00	46.76 (3.83)	0.80 (0.13)
	$b = 0.5$	-4.76, 0.74, 8.28	31.73 (4.66)	0.57 (0.10)
		-4.50, 1.00, 10.00	33.16 (4.51)	0.56 (0.09)
	$b = 1.0$	-4.76, 0.74, 8.28	28.87 (6.10)	1.14 (0.15)
		-4.50, 1.00, 10.00	28.80 (5.99)	1.18 (0.15)
$\nu = 1.5$	$b = 0.5$	-4.36, 0.83, 5.67	32.92 (2.72)	0.53 (0.08)
		-4.50, 1.00, 5.00	34.96 (2.56)	0.50 (0.08)
	$b = 1.0$	-4.36, 0.83, 5.67	31.28 (3.50)	0.99 (0.12)
		-4.50, 1.00, 5.00	36.07 (3.20)	0.95 (0.12)
	$b = 0.5$	-4.38, 0.82, 7.83	32.36 (2.54)	0.36 (0.06)
		-4.50, 1.00, 10.00	32.64 (2.60)	0.33 (0.06)
	$b = 1.0$	-4.38, 0.82, 7.83	31.75 (3.86)	0.95 (0.12)
		-4.50, 1.00, 10.00	30.46 (3.90)	0.91 (0.11)

Table 6.10: Results for the Gaussian marking with a 40×40 grid. Posterior means (standard deviations) for the marking parameters a, b, c^2 and d^2. The third column shows the parameter values used in the prior of Z (first line: estimated; second line: values used in simulation). The values $a = 30$ and $d^2 = 16$ are used in the simulation.

			$\mu_Z, \sigma_Z^2, \phi_Z$	$\hat{a}$	$\hat{b}$	$\hat{c^2}$	$\hat{d^2}$
$\nu = 0.5$	$b = 0.25$	$c^2 = 0.25$	-4.34, 0.90, 4.67	29.94 (0.48)	0.26 (0.02)	0.48 (0.22)	14.63 (2.57)
			-4.50, 1.00, 5.00	30.21 (0.44)	0.24 (0.02)	0.50 (0.19)	14.76 (2.36)
		$c^2 = 1.00$	-4.34, 0.90, 4.67	30.06 (0.48)	0.25 (0.02)	1.16 (0.21)	15.93 (2.74)
			-4.50, 1.00, 5.00	30.35 (0.47)	0.24 (0.02)	1.18 (0.22)	16.14 (2.73)
	$b = 0.50$	$c^2 = 0.25$	-4.34, 0.90, 4.67	29.76 (0.54)	0.63 (0.03)	0.33 (0.28)	16.63 (2.40)
			-4.50, 1.00, 5.00	30.37 (0.50)	0.60 (0.03)	0.39 (0.23)	17.08 (2.34)
		$c^2 = 1.00$	-4.34, 0.90, 4.67	30.73 (0.66)	0.59 (0.03)	1.08 (0.52)	21.06 (3.84)
			-4.50, 1.00, 5.00	31.14 (0.70)	0.57 (0.03)	1.16 (0.46)	21.22 (3.45)
	$b = 0.25$	$c^2 = 0.25$	-4.44, 0.81, 12.46	29.68 (0.53)	0.19 (0.02)	0.33 (0.12)	15.66 (2.53)
			-4.50, 1.00, 10.00	30.37 (0.41)	0.18 (0.01)	0.22 (0.12)	17.53 (2.34)
		$c^2 = 1.00$	-4.44, 0.81, 12.46	30.07 (0.59)	0.20 (0.02)	0.97 (0.17)	16.53 (3.20)
			-4.50, 1.00, 10.00	30.76 (0.51)	0.18 (0.02)	0.89 (0.18)	18.42 (3.16)
	$b = 0.50$	$c^2 = 0.25$	-4.44, 0.81, 12.46	30.21 (0.73)	0.48 (0.03)	0.75 (0.22)	11.35 (2.43)
			-4.50, 1.00, 10.00	31.80 (0.59)	0.45 (0.03)	0.54 (0.26)	13.58 (2.74)
		$c^2 = 1.00$	-4.44, 0.81, 12.46	30.79 (0.74)	0.46 (0.03)	1.73 (0.30)	14.56 (3.23)
			-4.50, 1.00, 10.00	32.07 (0.63)	0.44 (0.03)	1.48 (0.31)	17.95 (3.57)
$\nu = 1.0$	$b = 0.25$	$c^2 = 0.25$	-4.54, 0.90, 5.70	29.90 (0.48)	0.19 (0.01)	0.24 (0.10)	17.19 (2.42)
			-4.50, 1.00, 5.00	30.36 (0.46)	0.19 (0.01)	0.17 (0.10)	18.71 (2.23)
		$c^2 = 1.00$	-4.54, 0.90, 5.70	31.36 (0.52)	0.17 (0.01)	1.15 (0.17)	16.91 (3.19)
			-4.50, 1.00, 5.00	31.72 (0.51)	0.16 (0.02)	1.13 (0.19)	18.48 (3.36)
	$b = 0.50$	$c^2 = 0.25$	-4.54, 0.90, 5.70	30.47 (0.58)	0.46 (0.03)	0.33 (0.17)	14.77 (2.57)
			-4.50, 1.00, 5.00	31.05 (0.54)	0.48 (0.02)	0.19 (0.15)	16.20 (2.45)
		$c^2 = 1.00$	-4.54, 0.90, 5.70	30.26 (0.65)	0.46 (0.03)	0.93 (0.23)	15.94 (3.27)
			-4.50, 1.00, 5.00	31.11 (0.61)	0.45 (0.02)	0.77 (0.24)	18.33 (3.32)
	$b = 0.25$	$c^2 = 0.25$	-4.76, 0.74, 8.28	30.65 (0.63)	0.19 (0.02)	0.28 (0.08)	13.86 (2.65)
			-4.50, 1.00, 10.00	30.96 (0.54)	0.18 (0.01)	0.26 (0.08)	14.72 (2.65)
		$c^2 = 1.00$	-4.76, 0.74, 8.28	30.09 (1.05)	0.18 (0.02)	1.08 (0.18)	14.06 (3.14)
			-4.50, 1.00, 10.00	30.36 (0.76)	0.18 (0.02)	1.08 (0.15)	14.11 (3.25)
	$b = 0.50$	$c^2 = 0.25$	-4.76, 0.74, 8.28	31.69 (0.86)	0.44 (0.02)	0.24 (0.13)	16.53 (3.15)
			-4.50, 1.00, 10.00	32.36 (0.76)	0.43 (0.02)	0.55 (0.14)	13.88 (2.99)
		$c^2 = 1.00$	-4.76, 0.74, 8.28	29.13 (0.80)	0.49 (0.03)	0.84 (0.19)	15.08 (3.44)
			-4.50, 1.00, 10.00	29.82 (1.03)	0.47 (0.03)	0.93 (0.21)	14.81 (3.39)
$\nu = 1.5$	$b = 0.25$	$c^2 = 0.25$	-4.36, 0.83, 5.67	29.93 (0.38)	0.23 (0.01)	0.44 (0.09)	13.08 (1.99)
			-4.50, 1.00, 5.00	30.08 (0.49)	0.22 (0.02)	0.26 (0.09)	15.49 (2.17)
		$c^2 = 1.00$	-4.36, 0.83, 5.67	29.28 (0.49)	0.23 (0.01)	1.13 (0.16)	16.36 (3.09)
			-4.50, 1.00, 5.00	30.40 (0.43)	0.20 (0.01)	1.11 (0.17)	18.66 (3.22)
	$b = 0.50$	$c^2 = 0.25$	-4.36, 0.83, 5.67	30.44 (0.56)	0.50 (0.03)	1.04 (0.18)	13.62 (2.74)
			-4.50, 1.00, 5.00	31.45 (0.67)	0.47 (0.03)	0.86 (0.21)	14.84 (2.88)
		$c^2 = 1.00$	-4.36, 0.83, 5.67	28.80 (0.69)	0.54 (0.03)	1.35 (0.19)	12.92 (2.77)
			-4.50, 1.00, 5.00	30.84 (0.60)	0.46 (0.03)	1.43 (0.22)	13.36 (2.71)
	$b = 0.25$	$c^2 = 0.25$	-4.38, 0.82, 7.83	30.23 (0.42)	0.17 (0.01)	0.33 (0.07)	12.29 (1.99)
			-4.50, 1.00, 10.00	30.89 (0.36)	0.13 (0.01)	0.37 (0.06)	12.47 (1.90)
		$c^2 = 1.00$	-4.38, 0.82, 7.83	29.36 (0.46)	0.18 (0.01)	0.86 (0.12)	15.79 (2.90)
			-4.50, 1.00, 10.00	29.89 (0.47)	0.16 (0.01)	0.91 (0.13)	16.82 (3.22)
	$b = 0.50$	$c^2 = 0.25$	-4.38, 0.82, 7.83	31.54 (0.41)	0.35 (0.01)	0.51 (0.07)	10.49 (1.74)
			-4.50, 1.00, 10.00	29.88 (0.46)	0.37 (0.02)	0.39 (0.07)	11.25 (2.02)
		$c^2 = 1.00$	-4.38, 0.82, 7.83	30.11 (0.50)	0.42 (0.02)	1.03 (0.13)	14.60 (2.70)
			-4.50, 1.00, 10.00	29.67 (0.56)	0.40 (0.01)	0.96 (0.13)	17.15 (3.14)

Some limited conclusions can be done for the exponential marking from results shown in Tables 6.8 and 6.9. The use of the 40×40 grid seems to provide slightly better results than the 20×20 grid as expected. The posterior means obtained using the 40×40 grid are around the values used in the simulation. Even though the priors used for a and b are quite weak, the estimation worked well. One should note, that the normal prior used for a does not restrict the parameter $a + b/\Lambda(x_i)$ of the conditional exponential distribution to be positive. This did not cause any problems in our simulation experiments, but for real data one might prefer a distribution truncated to the positive values instead. Using the 20×20 grid we obtain almost

systematically larger posterior means for the parameters a and b in these simulation experiments than using the denser grid. The 20×20 grid may here have slightly too few grid points to obtain reliable results. We further applied the 80×80 grid to the first parameter combination (the first two lines in Table 6.8) and obtained the following posterior means (standard deviations): $\hat{a} = 28.90(3.03)$, $\hat{b} = 0.60(0.09)$ for ($\mu_Z = -4.34$, $\sigma_Z^2 = 0.90$, $\phi_Z = 4.67$) and $\hat{a} = 29.99(2.99)$, $\hat{b} = 0.55(0.08)$ for ($\mu_Z = -4.50$, $\sigma_Z^2 = 1.00$, $\phi_Z = 5.00$), which is again an improvement compared to the results obtained by the 40×40 grid. In general, the number of grid points is a compromise between the precision of the approximation and computational burden.

The parameters of the log Gaussian Cox process are another source affecting the posterior distributions of the marking parameters. On the basis of our limited experimenting, the posterior distributions obtained for the mark parameters do not seem to be very sensitive to small differences in mean and covariance parameters of the random field. Further simulation studies are necessary for studying this effect in more detail.

The prior of b was gamma distributed emphasizing small values of b. It is a question, can one replace this by a more non-informative alternative. Therefore, we further experimented with a uniform prior for b using the 40×40 grid. We obtained $\hat{a} = 30.12(2.92)$ and $\hat{b} = 0.64(0.09)$ for the first parameter combination (the first line in Table 6.8) using the prior $b \sim \text{Unif}(0, 3.0)$ and the proposal $b^* \sim \text{Unif}(\max(0, b - 0.3), \min(b + 0.3, 3.0))$. All the other prior and proposal distributions were kept the same as earlier. In this experiment, the posterior of b was not sensitive to the prior of b.

The results gathered into Table 6.10 show that the Bayesian estimation works also reasonably well for the Gaussian marking (5.5). Similarly as for the exponential marking, we observed that larger number of points in $\mathcal{G}$ produced better results, and there was a tendency to obtain larger posterior means for the marking parameters using the 20×20 grid than using the 40×40 grid. A bit problematic point in estimation seemed to be c^2. We first used a gamma prior for c^2, which for $c^2 = 0.25$ lead to a positively skew posterior distribution (longer right tail) having an unreasonably small mean. We replaced the gamma prior by a uniform prior, and obtained satisfactory results, which are shown in Table 6.10. For $c^2 = 1.00$ the estimation worked in both cases.

We also experimented with the Gaussian model changing d^2 to be 64 and it seemed to work equally well as with smaller $d^2 = 16$. Further, we experimented with less restrictive priors $a \sim N(30, 100^2)$ and $d^2 \sim \text{Gamma}(1.00, 0.0625)$. This seemed to work fine in most of our simulations, but the above mentioned problem for the value $c^2 = 0.25$ increased when the gamma prior was used.

We are concerned of the convergence of the smoothest random field structure with $\nu = 1.5$ and $\phi_Z = 10$. The convergence of the MCMC algorithm seems obvious for other parameter combinations, but for $\nu = 1.5$ even 100,000 iterations may not suffice. Therefore, we performed further runs of the algorithm (adjacent to the previous ones) with 200,000 iterations for the cases with $\nu = 1.5$ and $\phi_Z = 10$ using the 40×40 grid. However, the results calculated from the last 20,000 iterations of the longer runs were very close to the previous ones. Regardless, we would not recommend the use of a very smooth random field structure unless such choice is highly important for some particular applications.

For comparison, we further apply the minimum contrast estimation (see Section 6.2) and the conditional maximum likelihood method (see Section 6.1) to the simulated exponential intensity-marked Cox processes. The estimated parameters of the log Gaussian Cox process

enter the minimization of (6.13). Further, the minimization is performed with the values used in the simulation. In both cases, the integral (6.13) is minimized using $\epsilon = 0.5$, $a_0 = 30$, $\eta = 1$, and both $C_Z(r)$ (called Method 3a) and $C_Z(r;\hat{\theta})$ (called Method 3b), see Section 6.2.2. For Method 2 the estimates $\hat{\Lambda}(x_i)$ for $x_i \in N$ are obtained by (6.1) with Epanechnikov kernel and bandwidth $h = 10$. The obtained point estimates are shown in Table 6.11.

A clear advantage of the Bayesian approach is that it enables the calculation of uncertainties for the parameters in terms of the posterior standard deviation (or interval) of the marginal posterior distributions. It also otherwise performs better than Methods 2 and 3 according to our experimenting. The minimum contrast method may be sensitive to fine details in the covariance function $C_Z(r)$ and to the user-specified parameters (ϵ, a_0 and η), but the results of Table 6.11 show that it, and also Method 2, are able to give suggestive estimates.

Table 6.11: Point estimates for the exponential marking by Method 3a (MinC with $\hat{C}_Z(r)$) and 3b (MinC with $C_Z(r;\hat{\theta})$) using $\epsilon = 0.5$, $a_0 = 30$ and $\eta = 1$ in (6.13), and by Method 2 (using the Epanechnikov kernel with bandwidth $h = 10$ in kernel estimation of the intensity). The third column shows the parameter values used in the prior of Z (first line: estimated; second line: values used in simulation). The value $a = 30$ is used in the simulation.

		$\mu_Z, \sigma_Z^2, \phi_Z$	Method 3a: $(\hat{a}, \hat{b})$	Method 3b: $(\hat{a}, \hat{b})$	Method 2: $(\hat{a}, \hat{b})$
$\nu = 0.5$	$b = 0.5$	-4.34, 0.90, 4.67	(38.48, 0.47)	(38.48, 0.48)	(38.48, 0.49)
		-4.50, 1.00, 5.00	(34.17, 0.44)	(34.11, 0.46)	
	$b = 1.0$	-4.34, 0.90, 4.67	(33.68, 0.76)	(30.53, 0.80)	(31.03, 1.15)
		-4.50, 1.00, 5.00	(33.63, 0.65)	(38.00, 0.64)	
	$b = 0.5$	-4.44, 0.81, 12.46	(23.28, 0.48)	(22.18, 0.49)	(27.93, 0.58)
		-4.50, 1.00, 10.00	(23.21, 0.45)	(26.83, 0.41)	
	$b = 1.0$	-4.44, 0.81, 12.46	(24.81, 1.15)	(18.61, 1.21)	(29.37, 1.39)
		-4.50, 1.00, 10.00	(22.99, 1.07)	(30.98, 1.01)	
$\nu = 1.0$	$b = 0.5$	-4.54, 0.90, 5.70	(45.86, 0.27)	(47.23, 0.25)	(34.71 0.50)
		-4.50, 1.00, 5.00	(45.99, 0.28)	(47.53, 0.25)	
	$b = 1.0$	-4.54, 0.90, 5.70	(43.68, 0.59)	(48.38, 0.53)	(40.32, 0.90)
		-4.50, 1.00, 5.00	(43.75, 0.61)	(48.96, 0.54)	
	$b = 0.5$	-4.76, 0.74, 8.28	(24.55, 0.60)	(19.32, 0.64)	(24.98, 0.92)
		-4.50, 1.00, 10.00	(23.45, 0.77)	(24.98, 0.89)	
	$b = 1.0$	-4.76, 0.74, 8.28	(47.75, 0.67)	(44.04, 0.70)	(23.42, 1.68)
		-4.50, 1.00, 10.00	(47.71, 0.86)	(58.14, 0.89)	
$\nu = 1.5$	$b = 0.5$	-4.36, 0.83, 5.67	(24.93, 0.66)	(25.10, 0.65)	(29.23, 0.76)
		-4.50, 1.00, 5.00	(24.70, 0.57)	(29.41, 0.52)	
	$b = 1.0$	-4.36, 0.83, 5.67	(33.41, 0.92)	(34.28, 0.90)	(25.83, 1.39)
		-4.50, 1.00, 5.00	(33.05, 0.80)	(40.55, 0.72)	
	$b = 0.5$	-4.38, 0.82, 7.83	(37.34, 0.27)	(36.45, 0.29)	(31.11, 0.50)
		-4.50, 1.00, 10.00	(37.52, 0.24)	(38.22, 0.28)	
	$b = 1.0$	-4.38, 0.82, 7.83	(38.14, 0.96)	(35.92, 0.99)	(32.50, 1.22)
		-4.50, 1.00, 10.00	(37.89, 0.85)	(43.29, 0.96)	

6.5 Modelling the dbh of trees in a tropical rainforest data set

The marked point pattern of *Trichilia tuberculata*, shown in Figure 3.6, is here studied by the exponential intensity-marked Cox process. The model is fitted to the data using the (empirical) Bayesian method and its goodness-of-fit is evaluated.

First, the parameters of the log Gaussian Cox process are estimated using the minimum contrast method applied to the pair-correlation function. The Matérn covariance function (6.18) is fitted

using the values $\nu = 0.25, 0.5, 1.0, 1.5, 2.0$ to obtain estimates for σ_Z^2 and ϕ_Z. From these the parameter combination $\nu = 0.5$, $\sigma_Z^2 = 0.72$, $\phi_Z = 12.08$ minimizes the integral (6.10), where the argument range $(0.50, 30.00)$ and $\eta = 1$ are used. Note that the Matérn covariance function with $\nu = 0.5$ corresponds to the exponential covariance function $C_Z(r) = \sigma_Z^2 \exp(-r/\phi_Z)$, $r > 0$. Further, we obtain an estimate $\mu_Z = -4.30$.

The non-parametric estimate $\hat{C}_Z(r)$ and the fitted covariance function are plotted in Figure 6.6. The goodness-of-fit of the log Gaussian Cox process is evaluated by means of two other functions, the nearest neighbour distance distribution function $D(r)$ and the spherical contact distribution function $H(r)$, see for example Stoyan et al. (1995) or Illian et al. (2008). The envelopes for the $D(r)$- and $H(r)$-functions calculated from 99 simulations of the fitted model are shown in Figure 6.6. The functions `pcf`, `Fest` and `Gest` in R library `spatstat` have been used in estimation, see Baddeley and Turner (2005). These functions show that the fit is satisfactory, and therefore we continue with the analysis for the marks. The estimated parameters will enter the prior distribution of the intensity (or Z).

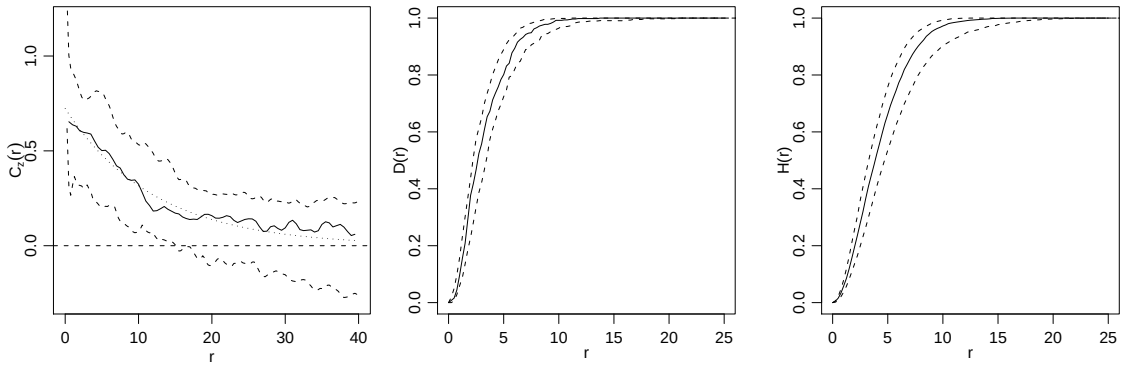

Figure 6.6: Estimated covariance function (*on the left*), nearest neighbour distance distribution function (*in the middle*) and spherical contact distribution function (*on the right*) of the *T. tuberculata* data set of Figure 3.6 (*solid lines*) with the envelopes (*dashed lines*) calculated from 99 simulations of the log Gaussian Cox process. The *dotted line* on the right is the fitted covariance function.

Next the marking model is fitted, which is of main interest. The range of the dbh is from 10 mm to 575 mm in the data set. The trees smaller than 10 mm have not been measured. We take the marks to be $m(x_i) = \text{dbh}(x_i) - 10$, where $\text{dbh}(x_i)$ is the dbh of the tree at x_i, and assume that these marks origin from (5.8). We set uniform priors for the marking parameters a and b: $a \sim \text{Unif}(0, 10.0)$ and $b \sim \text{Unif}(0, 3.0)$. These are weak priors which guarantee that $a + b/\Lambda(x_i) > 0$. The mean $\mu_Z = -4.30$ and the covariance function $C_Z(r) = 0.72 \exp(-r/12.08)$ are used in the prior distribution of the intensity. Further, we use the following proposal distributions: $a^* \sim \text{Unif}(\max(0, a - 2.0), \min(a + 2.0, 10.0))$, $b^* \sim \text{Unif}(\max(0, b - 0.3), \min(b + 0.3, 3.0))$ and $Z^*(\cdot) \sim N(Z(\cdot), 1.0)$. The 80×80 grid (corresponding to a 2.5×2.5 grid cell) is used for approximating the integral in the likelihood (6.17). The dense grid is especially necessary for performing accurate model assessment.

A total of 100,000 iterations of the MCMC algorithm is run for the data set. The initial values are simulated from $\text{Unif}(0, 10)$ and $\text{Unif}(0, 3)$ for a and b, respectively, and Z is initialized as explained in Section 6.4. The marginal posterior distributions of a and b are described through

their means, standard deviations and 90% posterior intervals. These descriptions, omitting the first 20,000 iterations as the burn-in, are given in Table 6.12. Further, the MCMC chains for a and b are shown in Figure 6.7. We also performed two additional runs of the algorithm using different initial values and these separate runs gave mutually consistent results.

Table 6.12: Results for the marked point pattern of *T. tuberculata* with marks $m(x_i) = \mathrm{dbh}(x_i) - 10$. Posterior means (standard deviations) and 90% posterior intervals for the model parameters a and b of the exponential marking. The first two columns tell the estimates for the parameters of the log Gaussian Cox process.

ν	$\mu_Z, \sigma_Z^2, \phi_Z$	$\hat{a}$	$\hat{b}$
0.5	-4.30, 0.72, 12.08	0.39 (0.39)	1.04 (0.06)
		0.02, 1.16	0.94, 1.15

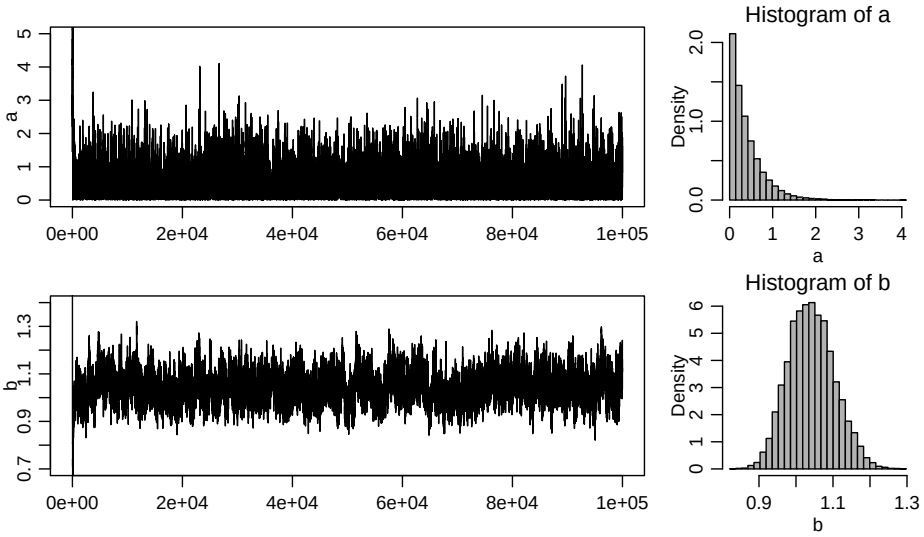

Figure 6.7: The plots of traces and marginal posterior distributions for the parameters a (*top*) and b (*bottom*) of the exponential marking fitted to the marked point pattern of *T. tuberculata* with marks $m(x_i) = \mathrm{dbh}(x_i) - 10$. The marginal posterior distributions are described by scaled histograms calculated omitting the 20,000 burn-in (dbh, diameter of tree at breast height).

Model assessment is performed utilizing predictive distributions. The algorithm is run forward 50,000 iterations, where every 1000th value of (a, b, Z) is stored. For each of the obtained 49 values of a, b and Z, a realisation of the intensity-marked Cox process is simulated: First, an inhomogeneous Poisson process is simulated in an area of size 200 m $\times$ 200 m using the intensity $\exp(Z(s))$, $s \in \mathcal{G}$. Let $x_1^s, \ldots, x_n^s$ be the points of this realisation. Next we simulate $Z(x_1^s), \ldots, Z(x_n^s)$ conditional on $\{Z(s) : s \in \mathcal{G}\}$, and the marks thereafter by (5.8) using $Z(x_1^s), \ldots, Z(x_n^s)$ and the corresponding values of a and b. The conditional simulation of the values of Z is performed using the function `CondSimu` in R library `RandomFields`, see Schlather (2001b). From the 49 simulated marked point patterns we calculate the characteristics $E(r)$, $V(r)$, $\kappa_{mm}(r)$ and $\gamma_m(r)$ (using R library `MarkedPointProcess`, see Schlather et al., 2004) and

construct the minimum and the maximum envelopes from these. These envelopes are shown in Figure 6.8 together with the mark summaries calculated from the rainforest data set. These characteristics show that the fitted intensity and marking can provide patterns that are in agreement with the marked point pattern of *T. tuberculata*.

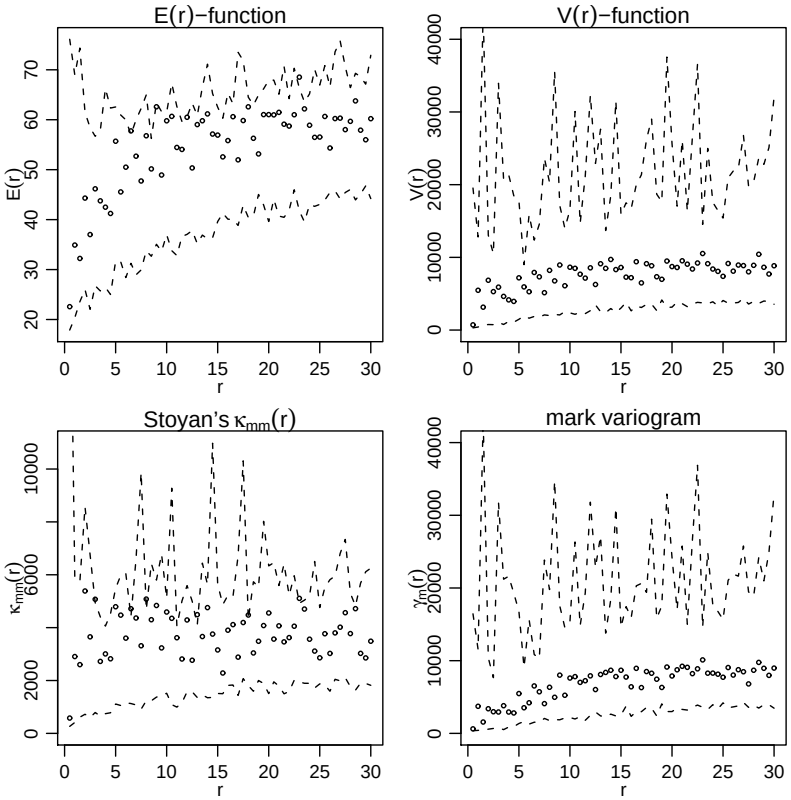

Figure 6.8: Mark summaries (*dots*) calculated from the marked point pattern of *T. tuberculata* with marks $m(x_i) = \text{dbh}(x_i) - 10$ and the envelopes (*dashed lines*) calculated from 49 marked point patterns simulated from the predictive distribution (dbh, diameter of tree at breast height).

The fitted model gives important information on interactions between the conspecific trees, which is one aspect of the forest structure. First, the point pattern of the locations of trees shows clustering, which is modelled by the log Gaussian Cox process. The scale parameter ϕ_Z of the Matérn covariance function (6.18) is estimated to be 12.08, which means that the spatial correlation vanishes (is less than 0.05) approximately for distances larger than 32 m. The fitted exponential marking shows that the tree size (dbh) depends on the local tree density. One possible explanation for the intensity-dependence is the seed dispersal. Further discussion on intensity-dependence can be found e.g. in Condit et al. (1992).

6.6 Discussion on intensity-dependent markings

The new markings of Section 5 for Cox processes extend the class of known marked point process models. In the new processes, the mark distribution varies along the intensity function,

and especially, intensity-dependence of the (conditional) mark variance is allowed. This feature is typical in forest science data, for example, and the new models are suitable for tree-wise forest modelling.

The heteroscedastic marking models considered here are obtained through intensity-dependent conditional mark variance. It should be noted that alternative constructions exist such as mark transformations. For example, if $\log(m(x_i)) = a + b\Lambda(x_i) + \epsilon(x_i)$ for $x_i \in N$, then both the conditional mean and variance of $m(x_i)$ given $\Lambda(x_i)$ depend on $\Lambda(x_i)$. The conditional specification of heteroscedasticity is more flexible, however. The choice of the mark distribution (5.2) is general. We have introduced normal, exponential and gamma intensity-marking keeping our application in mind.

We go beyond the seminal paper Ho and Stoyan (2008) and suggest the use of a Bayesian estimation method for the intensity-dependent markings. The simultaneous estimation of the intensity and marking parameters is supposed to be especially efficient, since under the assumption of intensity-dependence also the marks contain information on the intensity.

In this work, the log Gaussian Cox process is used as the model for point patterns. We prefer the use of the log Gaussian Cox process as a reasonable starting point at this early stage of development of marking strategies because many of its properties are analytically tractable. The point process model enters the estimation of the intensity and the marking parameters in the prior distribution of the intensity. Other Cox processes, such as cluster process or Poisson/gamma random field generated Cox process (Wolpert and Ickstadt, 1998), may be used as well.

From the point of view of applications, we have been successful in capturing the density-dependence which exists in the rainforest data. However, there is also a need for intensity-marked processes which allow different ranges of correlation for marks and points.

Besides the conditional markings considered in the previous chapters, other constructions are possible for creating intensity-dependent marks. For example, efficient use of point process transformations and generating structures (see Illian et al., 2008) can lead to interesting marking models. Some alternatives are presented in Chapters 7 and 8 of this thesis. A further generalisation is to consider bivariate (multivariate) markings along the current setup. This would allow to study the dependence of the marks of a tree species on the intensity of other species, for example.

When considering marking strategies, real-world data are essential sources of new ideas. For unmarked point pattern models no generic model family exists although Gibbs and Cox processes together form a flexible starting point. Accordingly, it is very unlikely to find a generic model for marked point processes. Instead, to develop a general marking strategy may be possible.

Chapter 7

Independent mark-dependent thinning of geostatistically marked point processes

The models of Chapters 4 and 5 start with a log Gaussian Cox point process and the marks are drawn from a conditional distribution which depends on the local intensity of the Cox process. These intensity-dependently marked log Gaussian Cox processes form a flexible model family for modelling intensity-dependence of marks, but also alternative constructions exist for creating marked point processes with intensity-dependent marks. One such construction is introduced in this chapter. The construction consists of two elements: geostatistical marking of an unmarked (stationary) point process and independent mark-dependent thinning. The two elements both have clear roles in the construction. First, the geostatistical marking creates dependence structure to the marks. Second, the mark-dependent thinning causes dependence between the intensity and marks.

This chapter is of theoretical nature. Properties of the new model are derived and examples of such processes are given. Especially, it is shown that the two simple mechanisms applied sequentially are able to create dependence between the marks and the point process.

The rest of this chapter is organized as follows. The new model is explained in Section 7.1 and results for its characteristics are given in Section 7.2. A simulation experiment is presented in Section 7.3 and Section 7.4 is reserved for discussion.

7.1 Mark-dependently thinned process

Let $N_b = \{y_j\}$ stand for a stationary and isotropic (unmarked) point process in $\mathbb{R}^d$. The point process N_b, called hereafter the *base process*, is subject to geostatistical marking: Let $\{U(s) : s \in \mathbb{R}\}$ be a real-valued stationary and isotropic second-order random field independent of N_b. The points of N_b are provided with marks $m(y_j) = U(y_j)$ resulting in a geostatistically marked point process $N_{b,U} = \{[y_j; m(y_j)]\}$. These marks inherit the correlation structure from $\{U(s)\}$.

Next a model for retention probabilities is introduced. The retention probability $q(m)$ is a function of the mark m. It is assumed to be a strictly monotonic function of m having values in a subset of $[0, 1]$. The marked point process $N_{b,U}$ undergoes independent mark-dependent thinning: each point y_j with mark $m(y_j) = U(y_j)$ is removed with probability $1 - q(m(y_j))$.

The removal of a point is independent of removals of other points. A new point process $N_m = \{[x_i; m(x_i)]\}$ is obtained with $\{x_i\} \subseteq \{y_j\}$ and $m(x_i) = U(x_i)$. Assuming stationarity and isotropy of N_b and $\{U(s)\}$, N_m is a stationary and isotropic process as well. Since thinning is mark-dependent and the marks are correlated, there will be dependence between the intensity and the marks.

7.2 First- and second-order characteristics for the thinned model

The important summary characteristics for the new marked point process are derived in this section. These characteristics show that the new model is able to create correlated marks that are not independent of the unmarked point process. First, the characteristics are given for the process N_m obtained by applying a retention function $q(m)$ to the geostatistically marked stationary and isotropic process $N_{b,U}$. Then, the characteristics for the specific case $q(m) = \exp(-am)$, $m > 0$, $a > 0$, are given in Section 7.2.2. Also a simple procedure to estimate the parameters of the model with $q(m) = \exp(-am)$ is presented.

7.2.1 General retention probability function

Assume that N_b is a stationary and isotropic point process with intensity $\lambda_b > 0$, second-order product density $\rho_b^{(2)}(r)$ and pair-correlation function $g_b(r)$. Further, assume that $\{U(s)\}$ is a real-valued stationary and isotropic second-order random field with mean μ_U, variance σ_U^2 and (valid) covariance function $C_U(r)$.

First-order characteristics

The intensity of N_m is

$$\lambda = \lambda_b \, \mathbb{E}[q(U(o))] \,, \tag{7.1}$$

and the mean mark

$$\mu_m = \frac{\mathbb{E}[U(o)\, q(U(o))]}{\mathbb{E}[q(U(o))]} = \frac{\lambda_b}{\lambda} \, \mathbb{E}[U(o)\, q(U(o))] \,. \tag{7.2}$$

The proofs for (7.1) and (7.2) are applications of the Campbell theorem (Stoyan et al., 1995, p. 103). Expressing the intensity of N_m in terms of the base process N_b conditioning by $\{U(s)\}$, using the Campbell theorem and stationarity, both of N_b and $\{U(s)\}$, the formula (7.1) is obtained for the intensity:

$$\begin{aligned}
\lambda &= \mathbb{E}\left(\sum_{[x_i;\, m_i] \in N_m} \mathbf{1}_{[0,1]^d}(x_i) \right) = \mathbb{E}\left[\mathbb{E}\left(\sum_{y_j \in N_b} \mathbf{1}_{[0,1]^d}(y_j) q(U(y_j)) \Big| \{U(s)\} \right) \right] \\
&= \mathbb{E}\left(\lambda_b \int \mathbf{1}_{[0,1]^d}(y) q(U(y)) \nu_d(\mathrm{d}y) \right) = \lambda_b \int \mathbf{1}_{[0,1]^d}(y) \mathbb{E}[q(U(y))] \nu_d(\mathrm{d}y) \\
&= \lambda_b \mathbb{E}[q(U(o))] \int \mathbf{1}_{[0,1]^d}(y) \nu_d(\mathrm{d}y) = \lambda_b \mathbb{E}[q(U(o))] \,,
\end{aligned}$$

where ν_d is the d-dimensional Lebesgue-measure. The mean mark (7.2) is obtained through the formulas (2.1), (2.2) and (2.3): Using the same arguments as above, the mean number of points in a Borel set $B \subset \mathbb{R}^d$ with mark in L satisfies

$$\begin{aligned}
\mathbb{E}[N_m(B \times L)] &= \mathbb{E}\left(\sum_{[x_i;\, m_i] \in N_m} \mathbf{1}_B(x_i)\mathbf{1}_L(m_i)\right) \\
&= \mathbb{E}\left[\mathbb{E}\left(\sum_{y_j \in N_b} \mathbf{1}_B(y_j)\mathbf{1}_L(U(y_j))q(U(y_j))\Big|\{U(s)\}\right)\right] \\
&= \mathbb{E}\left(\lambda_b \int_B \mathbf{1}_L(U(y))q(U(y))\nu_d(\mathrm{d}y)\right) \\
&= \lambda_b \int_B \mathbb{E}[\mathbf{1}_L(U(y))q(U(y))]\nu_d(\mathrm{d}y) \\
&= \lambda_b\, \nu_d(B)\, \mathbb{E}[\mathbf{1}_L(U(o))q(U(o))]\,.
\end{aligned}$$

Therefore, the mark distribution is

$$M(L) = \frac{\mathbb{E}[\mathbf{1}_L(U(o))q(U(o))]}{\mathbb{E}[q(U(o))]} = \frac{\lambda_b}{\lambda}\,\mathbb{E}[\mathbf{1}_L(U(o))q(U(o))]\,.$$

and the mark probability density function

$$f_m(m) = \frac{q(m)f_U(m)}{\mathbb{E}[q(U(o))]} = \frac{\lambda_b}{\lambda}\,q(m)f_U(m)\,,$$

where f_U is the density function of $U(o)$. Thus, the mean mark (7.2) follows.

The above derivations of the intensity and the mean mark are under the geostatistical marking. Similar formulas can be derived for these characteristics under the case where the marking of the base process is not specified. These apply the Campbell theorem (Stoyan et al., 1995, p. 106) for marked point processes, see Section 8.2 and the proofs for the intensity and the mean mark there.

Conditional on $\{U(s)\}$, the intensity of the thinned process is

$$\lambda(s) = \lambda_b q(U(s))$$

and the mark $m(x_i) = U(x_i)$ for $x_i \in N$. Vice versa, assuming q^{-1} exists,

$$m(x) = q^{-1}\left(\frac{\lambda(x)}{\lambda_b}\right).$$

Therefore, in realisations, the intensity and the marks are dependent.

The intensity-dependent markings of Chapters 4 and 5 are all for the log Gaussian Cox process and they are defined conditional on the intensity. A similar expression can be obtained for the model considered here if a Cox process is assumed. Suppose that N_b is a Cox process driven by the random intensity function $\{\Lambda_b(s)\}$. As above, $\{U(s)\}$ is a random field independent of N_b and, therefore, the transformed random field $\{q(U(s))\}$ is also independent of N_b. Then,

conditional on $\{q(U(s))\}$, independent thinning of the points in N_b with retention probabilities $\{q(U(s))\}$ results in a Cox process. The intensity function of the obtained Cox process is

$$\Lambda(s) = \Lambda_b(s)\, q(U(s)),$$

see Møller and Waagepetersen (2004, p. 58). If also $\{\Lambda_b(s)\}$ is fixed (that is $\{\Lambda(s)\}$ is fixed), then the relation between the marks and intensity can be obtained in the case q^{-1} exists:

$$m(x) = q^{-1}\left(\frac{\Lambda(x)}{\Lambda_b(x)}\right)$$

provided that $\Lambda_b(x) > 0$.

Second-order characteristics

The second-order behaviour of N_m is explained by the pair-correlation function

$$\begin{aligned} g(r) &= g_b(r)\,\frac{\mathbb{E}[q(U(o))q(U(\mathbf{r}))]}{[\mathbb{E}(q(U(o)))]^2} \\ &= g_b(r)\left\{1 + \frac{\mathrm{cov}(q(U(o)), q(U(\mathbf{r})))}{[\mathbb{E}(q(U(o)))]^2}\right\} \end{aligned} \tag{7.3}$$

and the f-mark correlation function

$$k_f(r) = \frac{\mathbb{E}[f(U(o), U(\mathbf{r}))\, q(U(o))q(U(\mathbf{r}))]}{\mathbb{E}[q(U(o))q(U(\mathbf{r}))]} \Big/ c_f, \tag{7.4}$$

where c_f is defined in (2.8). In order to obtain the formulas (7.3) and (7.4), the formulas for the second-order product density $\rho^{(2)}(r)$ and the second-order f-product density $\rho_f^{(2)}(r)$ are needed. These are derived below. Thereafter, the rest follows from the definitions (2.4), (2.6) and (2.9).

The functions $\rho^{(2)}(r)$ and $\rho_f^{(2)}(r)$ are the densities of the factorial moment measures $\alpha^{(2)}$ and $\alpha_f^{(2)}$, see Chapter 2. Using the formula (14.20) in Stoyan and Stoyan (1994, p. 246) to the base process, the following expressions are obtained for the factorial moment measures:

$$\begin{aligned} \alpha^{(2)}(B_1 \times B_2) &= \mathbb{E}\left(\sum_{[x_1;\, m_1],[x_2;\, m_2] \in N_m}^{\neq} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2)\right) \\ &= \mathbb{E}\left[\mathbb{E}\left(\sum_{y_1, y_2 \in N_b}^{\neq} \mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(U(y_1))q(U(y_2)) \,\Big|\, \{U(s)\}\right)\right] \\ &= \mathbb{E}\left[\int\int \mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(U(y_1))q(U(y_2))\varrho_b^{(2)}(y_1, y_2)\nu_d(\mathrm{d}y_1)\nu_d(\mathrm{d}y_2)\right] \\ &= \int_{B_1}\int_{B_2} \varrho_b^{(2)}(y_1, y_2)\mathbb{E}\left[q(U(y_1))q(U(y_2))\right]\nu_d(\mathrm{d}y_1)\nu_d(\mathrm{d}y_2) \end{aligned}$$

and

$$\begin{aligned}
\alpha_f^{(2)}(B_1 \times B_2) &= \mathbb{E}\left(\sum_{[x_1;\, m_1],[x_2;\, m_2] \in N_m}^{\neq} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2) f(m_1, m_2) \right) \\
&= \mathbb{E}\left[\mathbb{E}\left(\sum_{y_1, y_2 \in N_b}^{\neq} \mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2) f(U(y_1), U(y_2)) q(U(y_1)) q(U(y_2)) \Big| \{U(s)\} \right) \right] \\
&= \mathbb{E}\left(\int_{B_1} \int_{B_2} f(U(y_1), U(y_2)) q(U(y_1)) q(U(y_2)) \varrho_b^{(2)}(y_1, y_2) \nu_d(\mathrm{d}y_1) \nu_d(\mathrm{d}y_2) \right) \\
&= \int_{B_1} \int_{B_2} \varrho_b^{(2)}(y_1, y_2) \mathbb{E}[f(U(y_1), U(y_2)) q(U(y_1)) q(U(y_2))] \nu_d(\mathrm{d}y_1) \nu_d(\mathrm{d}y_2).
\end{aligned}$$

Therefore, assuming stationarity and isotropy,

$$\rho^{(2)}(r) = \rho_b^{(2)}(r) \mathbb{E}[q(U(o)) q(U(\mathbf{r}))] \tag{7.5}$$

and

$$\rho_f^{(2)}(r) = \rho_b^{(2)}(r) \mathbb{E}[f(U(o), U(\mathbf{r})) q(U(o)) q(U(\mathbf{r}))] \tag{7.6}$$

with $\|\mathbf{r}\| = r$. The formulas again assume geostatistical marking. Further, note that thinning is assumed to be independent: the probability that two points being at a distance r apart having marks m_1 and m_2 are both retained is simply $q(m_1) \cdot q(m_2)$. The formulas could be derived also in the case where neither geostatistical marking nor independent thinning is assumed, see the derivations of $\alpha^{(2)}$ and $\alpha_f^{(2)}$ in Section 8.2.

The mark characteristics of Table 2.1 for the thinned model can be obtained as direct consequences from (7.4). For example,

$$k_{mm}(r) = \frac{\mathbb{E}[U(o) U(\mathbf{r})\, q(U(o)) q(U(\mathbf{r}))]}{\mathbb{E}[q(U(o)) q(U(\mathbf{r}))]} \Big/ \mu_m^2 \,,$$

$$E(r) = \frac{\mathbb{E}[U(o)\, q(U(o)) q(U(\mathbf{r}))]}{\mathbb{E}[q(U(o)) q(U(\mathbf{r}))]} \,, \tag{7.7}$$

and

$$V(r) = \frac{\mathbb{E}[(U(o))^2\, q(U(o)) q(U(\mathbf{r}))]}{\mathbb{E}[q(U(o)) q(U(\mathbf{r}))]} - (E(r))^2 \,, \tag{7.8}$$

see Schlather (2001a) and Schlather et al. (2004). Interestingly, for the thinned model these characteristics do not depend on the characteristics of N_b, which is due to the geostatistical marking used in the construction of marks. Nevertheless, because of thinning, the functions (7.7) and (7.8) may not be constant anymore as the example in Section 7.2.2 shows.

Note that if N_b is a Cox process, the result (7.4) and its corollaries can also be obtained using the formula (5.4) (Ho and Stoyan, 2008).

7.2.2 An exponential retention probability function

Consider the retaining probability $q(m) = e^{-am}$, $m > 0$, $a > 0$. Then the following characteristics are obtained:

$$\begin{aligned}
\lambda &= \lambda_b \, e^{-a\mu_U + \frac{1}{2}a^2\sigma_U^2} \\
\mu_m &= \mu_U - a\sigma_U^2 \\
g(r) &= g_b(r) e^{a^2 C_U(r)} \\
E(r) &= \mu_U - a\sigma_U^2 - aC_U(r) \\
V(r) &= \sigma_U^2 \\
\kappa_{mm}(r) &= C_U(r) + (\mu_U - a\sigma_U^2 - aC_U(r))^2 \\
C_m(r) &= C_U(r) \\
\rho_m(r) &= C_U(r)/\sigma_U^2 \\
\gamma_m(r) &= \sigma_U^2 - C_U(r)
\end{aligned}$$

The proofs make use of $\mathbb{E}[e^{-aU(o)}] = e^{-a\mu_U + \frac{1}{2}a^2\sigma_U^2}$ and $\mathbb{E}[e^{-aU(o)-aU(\mathbf{r})}] = e^{-2a\mu_U + a^2\sigma_U^2 + a^2 C_U(r)}$, which are result of the use of the moment generating function of normally distributed variables, and the following expressions obtained by calculation:

$$\mathbb{E}[U(o)e^{-aU(o)-aU(\mathbf{r})}] = (\mu_U - a\sigma_U^2 - aC_U(r))e^{-2a\mu_U + a^2\sigma_U^2 + a^2}$$

$$\mathbb{E}[(U(o))^2 e^{-aU(o)-aU(\mathbf{r})}] = [\sigma_U^2 + (\mu_U - a\sigma_U^2 - aC_U(r))^2]e^{-2a\mu_U + a^2\sigma_U^2 + a^2}$$

$$\mathbb{E}[U(o)U(\mathbf{r})e^{-aU(o)-aU(\mathbf{r})}] = (C_U(r) + (\mu_U - a\sigma_U^2 - aC_U(r))^2)e^{-2a\mu_U + a^2\sigma_U^2 + a^2}.$$

These lead to the characteristics applying the formulas (7.1), (7.2), (7.3) and (7.4) and the relationships between the mark characteristics, see Table 2.1.

Maybe surprisingly, the above characteristics resemble the characteristics of the log-intensity marked Cox process (see Section 4.1), although the construction differs. For example, the $E(r)$-functions can have similar shape and $V(r)$-function is constant for both processes. However, while the mark characteristics of the log-intensity marked Cox process depend on the random field $\{Z(s)\}$ that generates the points, the characteristics of the thinned model origin from a geostatistical random field (and from thinning). The points and marks are very closely coupled in the log-intensity marked Cox process, whilst the thinned model may for example allow different ranges of correlation for the points and marks. Which of the models is more reasonable depends on the application.

Assume that a marked point pattern origins from a mark-dependently thinned process with the retention probability $q(m) = e^{-am}$, $m > 0$, $a > 0$. The following estimation procedure is the very first attempt to estimate the parameters of a mark-dependently thinned model. The estimation is strongly model-based.

Since $C_m(r) = C_U(r)$, the covariance function $C_U(r)$ and its parameters can be estimated directly from the marks. Usually fitting of a covariance model is done through the variogram. Therefore, estimates $\hat{C}_U(r)$ and $\hat{\sigma}_U^2$ (and for possible other covariance parameters) can be obtained by estimating the empirical mark variogram and fitting a model to it. The mean mark

μ_m is estimated as a mean of the marks, $\hat{\mu}_m = \sum_{i=1}^{n} m(x_i)/n$, and an estimate of a can be obtained by minimizing the integral

$$\int_{\epsilon}^{a_0} \left[(\hat{E}(r))^{\eta} - (\hat{\mu}_m - a\hat{C}_U(r))^{\eta} \right]^2 \mathrm{d}r,$$

where $\hat{E}(r)$ and $\hat{C}_U(r) = \hat{C}_m(r)$ are summaries estimated from the marks and ϵ, a_0 and η are parameters specified by the user. Then the mean μ_U can be estimated by

$$\hat{\mu}_U = \hat{\mu}_m + \hat{a}\hat{\sigma}_U^2.$$

Further, the intensity λ_b and the pair-correlation $g_b(r)$ of the point process N_b can be estimated by

$$\hat{\lambda}_b = \hat{\lambda} \exp(\hat{a}\hat{\mu}_U - \frac{1}{2}\hat{a}\hat{\sigma}_U^2)$$

and

$$\hat{g}_b(r) = \hat{g}(r) \exp(-\hat{a}^2 \hat{C}_U(r)), \tag{7.9}$$

where $\hat{\lambda}$ and $\hat{g}(r)$ are estimates for the intensity and pair-correlation of N_m, respectively. To estimate the pair-correlation by (7.9), similar type of estimators should be used for $\hat{g}(r)$ and $\hat{C}_U(r)$.

7.3 A simulation experiment

Let N_b be a Poisson process with intensity $\lambda_b = 2.7$ and $\{U(s)\}$ a stationary and isotropic Gaussian random field with mean $\mu_U = 6.0$ and Gaussian covariance function

$$C_U(r) = \sigma_U^2 \exp(-(r/\phi_U)^2)$$

with $\sigma_U^2 = 5.0$ and $\phi_U = 1.0$. The retaining probability function is $q(m) = \exp(-0.1m)$, $m > 0$. Figure 7.1 shows a realisation simulated in a square of size $[0, 10] \times [0, 10]$.

The mark characteristics for this construction are given in Section 7.2.2. These theoretical characteristics are plotted in Figure 7.2 together with the summaries calculated from the simulated data. Theoretical characteristics show that the marks are correlated and are not independent of the unmarked point pattern when $C_U(r)$ is not a constant function.

The results gathered in Table 7.1 are obtained employing the estimation procedure of Section 7.2.2. The Gaussian covariance (or variogram) model is fitted to the empirical mark variogram in order to get estimates for σ_U^2 and ϕ_U. The theoretical mark characteristics with the estimated parameters are also shown in Figure 7.2.

7.4 Discussion on the thinned process

Thinning is an operation that uses a specific rule to determine which points of a base process are deleted. It is a tool to construct new processes from a known base process. Well-known examples of point processes obtained through independent thinning are the inhomogeneous Poisson

Table 7.1: Estimation results for the mark-dependently thinned process with $q(m) = \exp(-0.1m)$, $m > 0$.

	σ_U^2	ϕ_U	a	μ_U	λ_b
Estimates	5.20	1.06	0.08	6.01	2.70
Value used in simulation	5.00	1.00	0.10	6.00	2.70

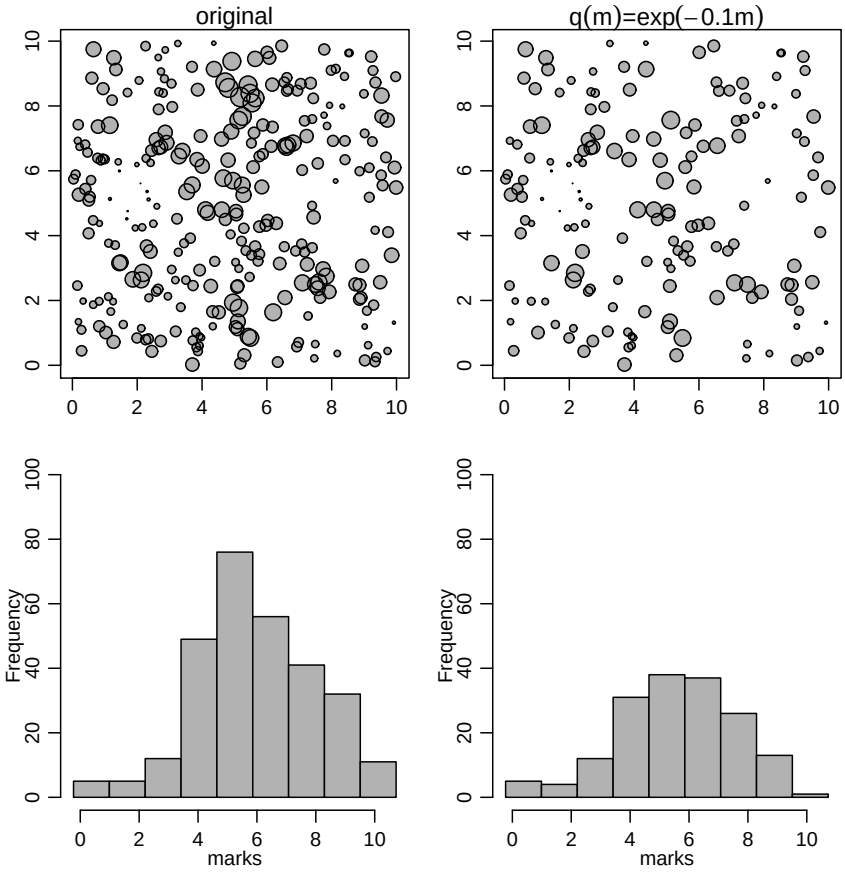

Figure 7.1: Realisations of the marked point processes $N_{b,U}$ (*on the left*) and N_m (*on the right*) and histograms of their marks.

process and the Cox process. The inhomogeneous Poisson process is obtained from a homogeneous Poisson process by applying an independent $p(x)$-thinning, where $p(x)$ is a deterministic function on $\mathbb{R}^d$ with $0 \leq p(x) \leq 1$, see e.g. Illian et al. (2008, p. 365). For the Cox process the thinning function is random. Similar strategy is here used for marked point processes: a point process is subject to independent random thinning where thinning probabilities origin from a geostatistical random field that generates the marks as well. Independent thinning is a reasonable starting point because of its simplicity. For example, some characteristics may be analytically tractable for models obtained through independent thinning.

Not all thinnings are independent. A well-known example of dependent thinning is Matérn type II process (Matérn, 1960, 1986; Illian et al., 2008).

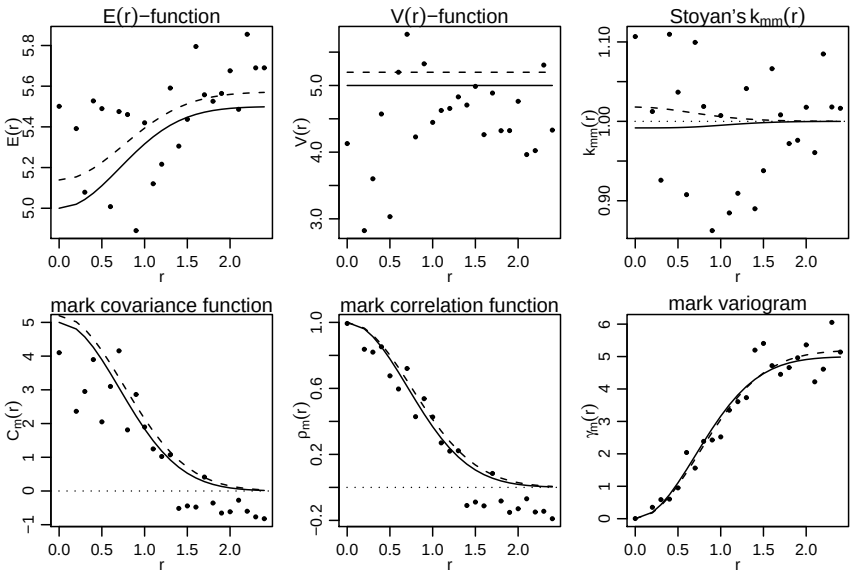

Figure 7.2: Mark summaries (*dots*) calculated from the realisation of N_m of Figure 7.1 and the corresponding theoretical characteristics with values used in simulation (*solid lines*) and with the estimated parameters (*dashed lines*).

Geostatistical marking and mark-dependent thinning applied together are able to create marked point processes with correlated marks which are not independent of the points. This new construction is most reasonable if it is motivated by the mechanisms underlying the observed data. Such data are an outcome of a dynamic process from the past to the present state. Often a reasonable starting point may be a Poisson process, the intensity of which is then modulated through the mark-dependent thinning.

The estimation of the thinned model may be based on its characteristics. Here such an estimation procedure is suggested for the case where thinning is according to the retention function $q(m) = \exp(-am), m > 0, a > 0$. This is very much based on the model assumption and is only the very first step in making inference for this kind of models.

Chapter 8

Mathematical characterization of Bitterlich forest: what is seen through a relascope

The main objective of this chapter is to provide mathematical characterization for the Bitterlich plot (or angle count plot or relascope plot) in terms of mark-dependent thinning. The Bitterlich plot is formed by the trees seen from a (random) location in a forest through a relascope, which is a simple optical instrument used to select the trees proportional to their diameter at a certain height, often 1.3 m above the ground level called breast height.

The geometrical principle of the relascope is presented in Figure 8.1. If the distance between the observer and a tree is less than $m/(2\sin\alpha)$, where m is the diameter of the tree at breast height (dbh), then the observer sees this tree with an angle larger than 2α, where α is a fixed angle. The relascope enables the observer to decide whether the angle subtended by a tree is larger than 2α. All trees fulfilling this criteria belong to the Bitterlich plot; they are retained in what is called relascope thinning.

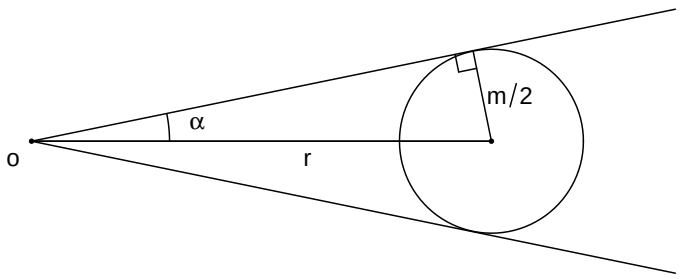

Figure 8.1: Geometrical principle of the relascope. The observer is at o and m is the diameter of the tree (at breast height).

In this work, the forest is described by a planar stationary marked point process $N_{b,m} =$

$\{[y_j; m(y_j)]\}$, where y_j stands for the location of a tree and $m(y_j)$ is its dbh (in metres). The Bitterlich plot is considered to be a realisation of a marked point process obtained by applying the relascope thinning to $N_{b,m}$ at a random location. A further objective is to create a new stationary marked point process which has intensity-dependent marks: applying the relascope thinning to $N_{b,m}$ at a number of random locations such a process is obtained.

Selection of trees by means of the relascope, called Bitterlich sampling or angle count sampling, is a technique developed for the estimation of the proportional (or mean) basal area of trees in a forest stand or for realising PPS (probability proportional to size) sampling (Bitterlich, 1948). In the terms of $N_{b,m}$, the proportional basal area is $\lambda_b \mathbb{E}_o A$, where λ_b is the intensity of $N_{b,m}$ and $\mathbb{E}_o A = \frac{\pi}{4}\mathbb{E}_o m^2$ is the expected basal area of a tree. (It is assumed that the cross-sections of trees are circular.) The popularity of Bitterlich sampling is based on the simple moment formula

$$\sin^2 \alpha \, \mathbb{E}T = \lambda_b \mathbb{E}_o A$$

for T defined as the number of trees in a Bitterlich plot, see e.g. Ripley (1981, p. 138) and Illian et al. (2008, p. 308). The relascope factor $\sin^2 \alpha$ is a quantity of the relascope. Therefore, since an unbiased estimator for the proportional basal area is obtained simply by $\sin^2 \alpha \cdot T$, Bitterlich sampling has become widely used in practical forest inventories, see e.g. Mandallaz (2008), Tomppo (2006) and Gregoire and Valentine (2008). If, in addition to counting the number of trees, the dbh of each tree in the Bitterlich plot has been measured, then it is possible to estimate also other proportional quantities of the forest (or $N_{b,m}$) based on the Bitterlich plot, see e.g. Mandallaz (2008). If the quantity is strongly correlated with the dbh, then the (unbiased) estimator based on Bitterlich sampling is expected to be efficient.

Despite of the popularity of Bitterlich sampling not much theoretical research has been done since Bitterlich (1948), see Penttinen (1988) and references there. The main objective has been the precision of the estimators for the proportional basal area and the comparison of efficiency between the uses of a Bitterlich plot and a sample plot of fixed size as an observation unit in sampling, see e.g. Holgate (1967), Matérn (1969), Matérn (1972) and Penttinen (1988). Here a new description of the Bitterlich plot is given from the point of view of marked point processes: it is studied what the Bitterlich plot itself, as a marked point pattern, looks like.

Bitterlich sampling is fundamentally considered as mark-related thinning of the process $N_{b,m}$, where the relascope defines the thinning rule. The emphasis is on the derivation of theoretical characteristics for the obtained process called point-centred Bitterlich point process. Assuming that $N_{b,m}$ is isotropic, the point-centred Bitterlich point process is isotropic having the sample point as its centre. Moreover, the marked process is an example of an isotropic centred marked point process. Its characteristics are considered in relation to the sample point, and it is shown that the mark distribution depends on the intensity function.

In addition, Bitterlich sampling is applied to $N_{b,m}$ at a set of random locations distributed according to a homogeneous Poisson process. This creates a new marked point process called here stationary Bitterlich point process. The construction of the process from a stationary marked point process $N_{b,m}$ is a two-step procedure where sample points are generated in the first step. Then, conditional on the generated sample points, deterministic thinning is applied to $N_{b,m}$. Important characteristics are derived for this new process, too, and examples are given. The characteristics for both new processes are mainly applications of Campbell theorems (see e.g. Stoyan et al., 1995 or Stoyan and Stoyan, 1994).

A base process is subject to mark-dependent thinning both in Bitterlich point processes and

in the process of Chapter 7. However, even if $N_{b,m}$ would be assumed to be geostatistically marked, the stationary Bitterlich point process is not a special case of the thinned model of Chapter 7 as such.

The rest of this chapter is arranged as follows. The finite and stationary Bitterlich point processes are studied in Sections 8.1 and 8.2, respectively. Section 8.3 is for examples and a real data is subject to Bitterlich sampling in Section 8.4. Section 8.5 is for discussion.

8.1 Point-centred Bitterlich point process

Assume that $N_{b,m} = \{[y_j; m_j]\}$ is a stationary marked point process in $\mathbb{R}^2$ with non-negative real-valued marks. The process $N_{b,m}$ is called the *base process*. No specific marking is assumed; this is a difference between $N_{b,m}$ and the base process $N_{b,U}$ of Chapter 7. However, the models introduced here are restricted to $\mathbb{R}^2$ and to non-negative marks. This is because the process $N_{b,m}$ describes a forest, trees with dbh-marks, and the thinning applies a geometric rule defined by the relascope (see Figure 8.1). From the point of view of the application, it is assumed that the cross-sections of trees are circular.

The intensity of the base process $N_{b,m}$ is denoted by λ_b. It is the mean number of points per unit area. Further, the mark distribution of the base process is denoted by M and the mark distribution function and its density by F_M and f_m, respectively. In what follows, the distances and dbh-marks are assumed to be in metres, and λ_b is the mean number of points per 1 m^2.

Define a *point-centred Bitterlich* (marked) *point process* N_m^ξ at $\xi \in \mathbb{R}^2$ as follows: a point $[y_j; m_j]$ of $N_{b,m}$ belongs to N_m^ξ if

$$|| y_j - \xi|| \leq \frac{m_j}{2 \sin \alpha}, \tag{8.1}$$

where α is a fixed relascope angle and $\xi \in \mathbb{R}^2$ is a sample point. The base process is subject to *relascope thinning*, where the retention/deletion of each point $[y_j; m_j] \in N_{b,m}$ depends on the distance from y_j to the sample point ξ and on the dbh-mark m_j at y_j according to (8.1). Each point $[y_j; m_j]$ has a limiting distance for inclusion in N_m^ξ: the point $[y_j; m_j]$ belongs to N_m^ξ if ξ falls into the circle $b(y_j, m_j/(2 \sin \alpha))$. If the mark distribution function F_M has a finite support, i.e. $F_M(m_0) = 1$ for some $m_0 < \infty$, then N_m^ξ is finite (with probability one) because $N_{b,m}$ is assumed to be locally finite.

A realisation of N_m^ξ is constructed by taking a realisation of $N_{b,m}$ and applying the thinning (retention) rule (8.1) to the points of $N_{b,m}$ at a location ξ. The realisation can be considered as the view seen from ξ through the relascope. It is called a *Bitterlich plot* (or angle count plot or relascope plot or angle gauge plot) in forestry. In the following, first-order characteristics of N_m^ξ are derived relative to the sample point ξ, which is considered fixed. Because the base process is assumed to be stationary, the o-centred process N_m^o is considered without loss of generality. Let N^o stand for the corresponding unmarked point process.

Assuming that the base process is isotropic, the process N^o is isotropic with respect to the sample point o. This *centred* point process (Daley and Vere-Jones, 2008, p. 466) is non-stationary, but it is θ-stationary on $\mathbb{R}^2$ as defined by Byth (1981). Also the marked point process N_m^o is isotropic: the mark distribution is invariant under rotations around the origin o. The process N_m^o is called *centred marked point process*.

The first-order behaviour of the unmarked process N^o is defined by an intensity function $\lambda(r)$, $r > 0$, which yields the mean number of points in the sphere $b(o, r) \subset \mathbb{R}^2$ (e.g. Illian et al., 2008, p. 114):

$$\mathbb{E}N^o(b(o,r)) = \int_0^r 2\pi R\lambda(R)\mathrm{d}R\,.$$

The value $\lambda(r)$ is proportional to the point density at a location being the distance r apart from o. It is the limit

$$\lambda(r) = \lim_{\Delta r \downarrow 0} \frac{\mathbb{E}N^o(b(o, r + \Delta r/2) \backslash b(o, r - \Delta r/2))}{2\pi r \Delta r}\,.$$

Applying the Campbell theorem (Stoyan et al., 1995, p. 108) for marked point processes to $N_{b,m}$ and changing to polar coordinates, we obtain

$$\begin{aligned}
\mathbb{E}N^o(b(o,r)) &= \mathbb{E}\left(\sum_{x\in N^o} \mathbf{1}_{b(o,r)}(x)\right) = \mathbb{E}\left(\sum_{[y;\,m]\in N_{b,m}} \mathbf{1}_{b(o,r)}(y)\mathbf{1}_{\{m\geq 2||\,y||\sin\alpha\}}(m)\right) \\
&= \lambda_b \int\int \mathbf{1}_{b(o,r)}(y)\mathbf{1}_{\{m\geq 2||\,y||\sin\alpha\}}(m)M(\mathrm{d}m)\nu_2(\mathrm{d}y) \\
&= \lambda_b \int_0^\pi \int_0^r \int_0^\infty \mathbf{1}_{\{m\geq 2R\sin\alpha\}}(m)M(\mathrm{d}m)R\;\mathrm{d}R\;\mathrm{d}\varphi \\
&= \int_0^r \lambda_b \int_0^\infty \mathbf{1}_{\{m\geq 2R\sin\alpha\}}(m)M(\mathrm{d}m)\;2\pi R\;\mathrm{d}R\,,
\end{aligned}$$

and thus,

$$\lambda(r) = \lambda_b(1 - F_M(2\sin\alpha\cdot r)) = \lambda_b S_M(2\sin\alpha\cdot r) \quad \text{for } r > 0\,, \tag{8.2}$$

where $S_M(t) = 1 - F_M(t)$ is the survival function of the mark distribution of the base process $N_{b,m}$. Recall that ν_d is the d-dimensional Lebesque measure (above $d = 2$). The intensity function $\lambda(r)$ depends on the distance r from o only, not on the direction. It is a decreasing function in r: the intensity of points is higher close to o, while also small trees are observed at short distances.

The mark density function of N_m^o depends on the distance r from o, and it is defined for $r > 0$ by

$$f_r(m) = \frac{\mathbf{1}_{\{m\geq 2r\sin\alpha\}}(m)f_m(m)}{\int_0^\infty \mathbf{1}_{\{l\geq 2r\sin\alpha\}}(l)f_m(l)\mathrm{d}l}\,,$$

where $f_m(m)$ is the mark density function of the base process. The formula origins from the fact that at the distance r from o only the points with mark $m \geq 2\sin\alpha\cdot r$ retain in thinning. The distribution $f_r(m)$ can further be rewritten as

$$f_r(m) = \frac{\mathbf{1}_{\{m\geq 2r\sin\alpha\}}(m)f_m(m)}{1 - F_M(2\sin\alpha\cdot r)} = \frac{\lambda_b}{\lambda(r)}\,\mathbf{1}_{\{m\geq 2r\sin\alpha\}}(m)f_m(m)\,, \tag{8.3}$$

where the equation (8.2) is utilized. From the formula (8.3) one can deduce the dependence between the intensity $\lambda(r)$ and the mark density $f_r(m)$. The corresponding cumulative mark distribution is

$$F_r(m) = \frac{\int_0^m \mathbf{1}_{\{t\geq 2r\sin\alpha\}}(t)f_m(t)\mathrm{d}t}{\int_0^\infty \mathbf{1}_{\{l\geq 2r\sin\alpha\}}(l)f_m(l)\mathrm{d}l} = \frac{F_M(m) - F_M(2\sin\alpha\cdot r)}{1 - F_M(2\sin\alpha\cdot r)} = 1 - \frac{\lambda_b}{\lambda(r)}(1 - F_M(m))\,,$$

for $m \geq 2r\sin\alpha$ and $r > 0$. The formula can also be expressed in terms of survival functions: it is

$$S_r(m) = \frac{S_M(m)}{S_M(2\sin\alpha \cdot r)} \quad \text{for } m \geq 2r\sin\alpha \text{ and } r > 0.$$

It tells the proportion of points larger than m for a fixed distance r.

The mean mark at a distance $r > 0$ from o is

$$\mu_m(r) \;=\; \int_0^\infty m f_r(m)\mathrm{d}m = \frac{\lambda_b}{\lambda(r)} \int_0^\infty m\mathbf{1}_{\{m\geq 2r\sin\alpha\}}(m) f_m(m)\mathrm{d}m\,,$$

and the mark variance at a distance $r > 0$ can be written as

$$\sigma_m^2(r) \;=\; \int_0^\infty m^2 f_r(m)\mathrm{d}m - (\mu_m(r))^2 = \frac{\lambda_b}{\lambda(r)} \int_0^\infty m^2\mathbf{1}_{\{m\geq 2r\sin\alpha\}}(m) f_m(m)\mathrm{d}m - (\mu_m(r))^2\,.$$

Intuitively, the mean mark $\mu_m(r)$ increases in r, but the mark variance $\sigma_m^2(r)$ decreases (if the marks are bounded), because at short distances both small and large trees are observed and at large distances only large trees.

8.2 Stationary Bitterlich point process

Assume that $N_{b,m} = \{[y_j; m_j]\}$ is the base process as above with intensity λ_b, second-order product density $\varrho_b(r)$ and with non-negative real-valued marks. Let $N_s = \{\xi_k\}$ be a homogeneous Poisson process with intensity λ_s. It is called (Poisson) *sampling process*, and it corresponds to uniform spatial sampling.

The base process $N_{b,m}$ is subject to relascope thinning: those points y_j with marks m_j, which are seen by means of the relascope of angle α from the sample points, are retained. Given N_s, the condition for the retention of a point $[y_j; m_j] \in N_{b,m}$ is

$$||\, y_j - \xi_k|| \leq \frac{m_j}{2\sin\alpha} \quad \text{for some } \xi_k \in N_s. \tag{8.4}$$

A new marked point process of the survived points, denoted by $N_m^B = \{[x_i; m_i]\}$, is obtained. It will be called *Bitterlich* (marked) *point process.*

A point y_j with mark m_j belongs to N_m^B, if there is at least one sample point in the circle $b(y_j, m_j/(2\sin\alpha))$. Therefore, assuming the sample points are distributed according to the homogeneous Poisson process of the intensity λ_s, the retention probability of a point with mark m equals

$$q(m, \alpha, \lambda_s) = 1 - \exp(-\lambda_s \pi m^2/(4\sin^2\alpha))\,, \tag{8.5}$$

the complement of the probability that there are no points of N_s in a circle with radius $m/(2\sin\alpha)$. For brevity of notation, let $Q(\alpha, \lambda_s)$ stand for the mean retention probability,

$$Q(\alpha, \lambda_s) = \int_0^\infty q(m, \alpha, \lambda_s) M(\mathrm{d}m)\,,$$

where M is the mark distribution of the base process. Note that the Bitterlich point process N_m^B is not a result of independent mark-dependent thinning with thinning probability (8.5).

8.2.1 First-order characteristics

The intensity of the stationary Bitterlich point process N_m^B is

$$\lambda = \lambda_b Q(\alpha, \lambda_s)\,, \tag{8.6}$$

which formula is similar to the intensity (7.1) of the thinned process, see Section 7.2. The proof of (8.6) is an application of the Campbell theorem (Stoyan et al., 1995, p. 108) to the marked point process $N_{b,m}$:

$$\begin{aligned}
\lambda &= \mathbb{E}\left(\sum_{[x;\,m]\in N_m^B} \mathbf{1}_{[0,1]^2}(x)\right) = \mathbb{E}\left(\sum_{[y;\,m]\in N_{b,m}} \mathbf{1}_{[0,1]^2}(y) q(m,\alpha,\lambda_s)\right) \\
&= \lambda_b \int\int \mathbf{1}_{[0,1]^2}(y) q(m,\alpha,\lambda_s) M(\mathrm{d}m)\nu_2(\mathrm{d}y) = \lambda_b Q(\alpha,\lambda_s)\int \mathbf{1}_{[0,1]^2}(y)\nu_2(\mathrm{d}y) \\
&= \lambda_b Q(\alpha,\lambda_s)\,.
\end{aligned}$$

The first line in the formula is obtained by first taking the conditional expectation over the sum given $N_{b,m}$, and then the expectation over $N_{b,m}$ is to be taken. The Campbell theorem is applied to the latter expectation.

The mark distribution of N_m^B is obtained through the equation (2.1): The expected number of points in a Borel set $B \subset \mathbb{R}^2$ with mark in $L \subset [0,\infty)$ satisfies

$$\begin{aligned}
\mathbb{E}[N_m(B\times L)] &= \mathbb{E}\left(\sum_{[x;\,m]\in N_m^B} \mathbf{1}_B(x)\mathbf{1}_L(m)\right) = \mathbb{E}\left(\sum_{[y;\,m]\in N_{b,m}} \mathbf{1}_B(y)\mathbf{1}_L(m) q(m,\alpha,\lambda_s)\right) \\
&= \lambda_b \int\int \mathbf{1}_B(y)\mathbf{1}_L(m) q(m,\alpha,\lambda_s) M(\mathrm{d}m)\nu_2(\mathrm{d}y) = \lambda_b \int_B \mathbb{E}[\mathbf{1}_L(m) q(m,\alpha,\lambda_s)]\nu_2(\mathrm{d}y) \\
&= \lambda_b \nu_2(B)\mathbb{E}[\mathbf{1}_L(m) q(m,\alpha,\lambda_s)]\,,
\end{aligned}$$

where again the Campbell theorem is used. It follows that the mark distribution of the Bitterlich process N_m^B is

$$M^B(L) = \frac{\lambda_b}{\lambda}\,\mathbb{E}[\mathbf{1}_L(m) q(m,\alpha,\lambda_s)] = \frac{\mathbb{E}[\mathbf{1}_L(m) q(m,\alpha,\lambda_s)]}{Q(\alpha,\lambda_s)}$$

and, especially, the mark distribution function is

$$F_M^B(m) = \frac{\lambda_b}{\lambda}\int_0^m q(l,\alpha,\lambda_s) f_m(l)\mathrm{d}l\,, \tag{8.7}$$

where $f_m(m)$ is the mark probability density function of the base process. Thus, the mark density function of the Bitterlich process is

$$f_m^B(m) = \frac{\lambda_b}{\lambda}\, q(m,\alpha,\lambda_s) f_m(m)$$

and the mean mark

$$\mu_m^B = \frac{\lambda_b}{\lambda}\int_0^\infty m q(m,\alpha,\lambda_s) f_m(m)\mathrm{d}m = \frac{\lambda_b}{\lambda}\,\mathbb{E}[m q(m,\alpha,\lambda_s)]\,.$$

These results are closely connected to the results obtained for the thinned model of Section 7: Consider a general retention probability $q(m)$ instead of $q(m, \alpha, \lambda_s)$ and $\mathbb{R}^d$ instead of $\mathbb{R}^2$ in the above formulas. Then the expressions for the thinned model will be obtained where the mark distribution $f_m(m)$ corresponds to the distribution of $U(o)$ under the geostatistical marking.

The intensity function and mark distribution can be considered locally with respect to the sample points ξ_k of N_s: Let $\{\xi_k\}$ be a fixed realisation of N_s. The local intensity function, conditional on $\{\xi_k\}$, is

$$\lambda(s|\{\xi_k\}) = \lambda_b(1 - F_M(2\sin\alpha \cdot \min_k\{\|s - \xi_k\|\})) \quad \text{for } s \in \mathbb{R}^2. \tag{8.8}$$

The formula (8.8) is due to that at $s \in \mathbb{R}^2$ the effect of the closest sample point matters: the point $[y_j; m_j] \in N_{b,m}$ is retained if $m_j \geq 2\sin\alpha \cdot \min_k\{\|y_j - \xi_k\|\}$. The area of influence of a sample point $\xi_k \in N_s$ is described by its Voronoi polygon

$$V_k = \{s : \|s - \xi_k\| \leq \|s - \xi_l\| \text{ for } l \neq k, \xi_l \in N_s\}.$$

These polygons define a Voronoi tessellation in $\mathbb{R}^2$, see e.g. Okabe et al. (2000). This tessellation tells the influence zone of each ξ_k. In the zone of ξ_k, the intensity is determined by (8.2) where r is the distance to ξ_k. The Voronoi tessellation is useful in practical calculation of (8.8).

The conditional mark distribution function of the mark m at $x \in N_m^B$ is

$$F_x(m|\{\xi_k\}) = 1 - \frac{\lambda_b}{\lambda(x|\{\xi_k\})}(1 - F_M(m)) \quad \text{for } m \geq 2\sin\alpha \cdot \min_k\{\|x - \xi_k\|\},$$

and the corresponding mark density function is

$$f_x(m|\{\xi_k\}) = \frac{\lambda_b}{\lambda(x|\{\xi_k\})}\mathbf{1}_{\left\{m \geq 2\sin\alpha \cdot \min_k\{\|x-\xi_k\|\}\right\}}(m) f_m(m).$$

These characteristics show that the marks of N_m^B depend locally on the intensity.

8.2.2 Second-order characteristics

Assume the base process is both stationary and isotropic. Then the second-order product densities $\varrho^{(2)}(r)$ and $\varrho_f^{(2)}(r)$ of the Bitterlich point process N_m are

$$\varrho^{(2)}(r) = \varrho_b^{(2)}(r) \int_0^\infty \int_0^\infty q(m_1, m_2, r) M_r(\mathrm{d}m_1, \mathrm{d}m_2) \tag{8.9}$$

and

$$\varrho_f^{(2)}(r) = \varrho_b^{(2)}(r) \int_0^\infty \int_0^\infty q(m_1, m_2, r) f(m_1, m_2) M_r(\mathrm{d}m_1, \mathrm{d}m_2), \tag{8.10}$$

respectively, where

$$q(m_1, m_2, r) = 1 - a - b + a\,b\,e^{\lambda_s \gamma(m_1, m_2, r)}, \tag{8.11}$$

where $a = e^{-\lambda_s \pi \frac{m_1^2}{4\sin^2\alpha}}$, $b = e^{-\lambda_s \pi \frac{m_2^2}{4\sin^2\alpha}}$ and

$$\gamma(m_1, m_2, r) = \nu_2\left(b\left(o, \frac{m_1}{2\sin\alpha}\right) \bigcap b\left(\mathbf{r}, \frac{m_2}{2\sin\alpha}\right)\right), \|\mathbf{r}\| = r, \tag{8.12}$$

and M_r is the two-point mark distribution of the base process $N_{b,m}$. The formula for (8.12) can be found e.g. in Stoyan and Stoyan (1994, p. 365): it is the area of the intersection of two discs. In other words, it is the set covariance of the two discs. The proofs of (8.9) and (8.10) follow.

The quantity (8.11) is the probability that two points of $N_{b,m}$ with marks m_1 and m_2 being the distance r apart are retained. Indeed, for a stationary and isotropic process this probability depends only on the marks and the distance r. The formula (8.11) is obtained as the complement of the probability that either $[y_1; m_1] \in N_{b,m}$ or $[y_2; m_2] \in N_{b,m}$ is removed, given that there are two points of $N_{b,m}$, one at y_1 and another at y_2 for which $\|y_1 - y_2\| = r$. Note that (8.11) depends on the sampling process through α and λ_s.

The probability (8.11) enters the formulas for the factorial moment measures $\alpha^{(2)}$ and $\alpha_f^{(2)}$, when the sums of these measures are again written in the terms of the base process. Further, the formula

$$\begin{aligned}
&\mathbb{E}\left(\sum_{[x_1;\,m_1],\,[x_2;\,m_2]\in\, N_m}^{\neq} f(m_1,m_2)\mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_n)\right) = \int_{B_1}\int_{B_2}\rho_f^{(2)}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\
&= \int_{B_1}\int_{B_2}\int\int f(m_1,m_2)M_{x_1,x_2}(\mathrm{d}m_1,\mathrm{d}m_2)\rho^{(2)}(x_1,x_2)\nu_d(\mathrm{d}x_1)\nu_d(\mathrm{d}x_2),
\end{aligned}$$

where N_m is a marked point process in $\mathbb{R}^d$, $\rho^{(2)}$ its second-order product density and M_{x_1,x_2} its two-point mark distribution, is applied to the base process. The formula is due to formulas in Stoyan and Stoyan (1994, p. 263), see also (2.7) and (2.6). The following expressions are obtained for the factorial moment measures:

$$\begin{aligned}
\alpha^{(2)}(B_1\times B_2) &= \mathbb{E}\left(\sum_{[x_1;\,m_1],[x_2;\,m_2]\in N_m^B}^{\neq}\mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2)\right)\\
&= \mathbb{E}\left(\sum_{[y_1;\,m_1],[y_2;\,m_2]\in N_{b,m}}^{\neq}\mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(m_1,m_2,r)\right)\\
&= \int\int\int\int\mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(m_1,m_2,r)M_r(\mathrm{d}m_1,\mathrm{d}m_2)\varrho_b^{(2)}(y_1,y_2)\nu_2(\mathrm{d}y_1)\nu_2(\mathrm{d}y_2)\\
&= \int_{B_1}\int_{B_2}\varrho_b^{(2)}(y_1,y_2)\mathbb{E}\left[q(m_1,m_2,r)\right]\nu_2(\mathrm{d}y_1)\nu_2(\mathrm{d}y_2)
\end{aligned}$$

and

$$\begin{aligned}
&\alpha_f(B_1\times B_2) = \mathbb{E}\left(\sum_{[x_1;\,m_1],[x_2;\,m_2]\in N_m^B}^{\neq}\mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2)f(m_1,m_2)\right)\\
&= \mathbb{E}\left(\sum_{[y_1;\,m_1],[y_2;\,m_2]\in N_{b,m}}^{\neq}\mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(m_1,m_2,r)f(m_1,m_2)\right)\\
&= \int\int\int\int\mathbf{1}_{B_1}(y_1)\mathbf{1}_{B_2}(y_2)q(m_1,m_2,r)f(m_1,m_2)M_r(\mathrm{d}m_1,\mathrm{d}m_2)\varrho_b^{(2)}(y_1,y_2)\nu_2(\mathrm{d}y_1)\nu_2(\mathrm{d}y_2)\\
&= \int_{B_1}\int_{B_2}\varrho_b^{(2)}(y_1,y_2)\mathbb{E}[q(m_1,m_2,r)f(m_1,m_2)]\nu_2(\mathrm{d}y_1)\nu_2(\mathrm{d}y_2)\,,
\end{aligned}$$

where $\|y_1 - y_2\| = r$. Therefore, the densities (8.9) and (8.10) are obtained.

Direct consequences of (8.9) and (8.10) are, by definition, the pair-correlation function

$$g(r) = \frac{g_b(r) \int_0^\infty \int_0^\infty q(m_1, m_2, r) M_r(\mathrm{d}m_1, \mathrm{d}m_2)}{(Q(\alpha, \lambda_s))^2}$$

where $g_b(r)$ is the pair-correlation function of the base process, and the f-mark correlation function

$$k_f(r) = \frac{\int_0^\infty \int_0^\infty q(m_1, m_2, r) f(m_1, m_2) M_r(\mathrm{d}m_1, \mathrm{d}m_2)}{\int_0^\infty \int_0^\infty q(m_1, m_2, r) M_r(\mathrm{d}m_1, \mathrm{d}m_2)},$$

from which follows, for example, that the $E(r)$-function is

$$E(r) = \frac{\int \int q(m_1, m_2, r) m_1 M_r(\mathrm{d}m_1, \mathrm{d}m_2)}{\int \int q(m_1, m_2, r) M_r(\mathrm{d}m_1, \mathrm{d}m_2)}.$$

8.3 A simulation experiment

The stationary base process is assumed to be an independently marked Poisson process with intensity $\lambda_b = 0.05$ and mark distribution Weibull$(a = 2, b = 0.25)$. The mark distribution has the density

$$f(m) = \left(\frac{a}{b}\right) \left(\frac{m}{b}\right)^{a-1} \exp\left\{-\left(\frac{m}{b}\right)^a\right\}, \quad m \geq 0,$$

and the mean mark of this distribution equals

$$\mathbb{E}(m) = b\Gamma\left(1 + \frac{1}{a}\right).$$

A realisation is simulated in a rectangular window of size $[0, 200] \times [0, 200]$. Consider this as a forest example: the tree density is 500 trees per hectare and the mean diameter (dbh) is approximately 0.22 m. The Weibull distribution is often used as a model for the dbh, see e.g. Kilkki and Päivinen (1986).

Assume that the trees are selected using Bitterlich sampling with $\alpha = 0.01$ rad ($0.57\ ^o$). For this relascope, each tree in the plot represents the basal area of size 1 m^2 per a hectare: an estimate for the proportional basal area per a hectare is simply the number of trees in the plot. Consider first that the observer goes to a random location in the plot. Here the location $(81.78, 64.80)$ was obtained by simulation. The finite Bitterlich process seen at this sample point by means of the relascope is shown in Figure 8.2 (in the middle). It has 27 points (trees). The maximum of the marks is 0.68 m in this simulation, and therefore, the maximum distance at which a tree can be seen by the relascope is 34.0 m.

The cumulative Weibull distribution is

$$F(m) = 1 - \exp\left\{-\left(\frac{m}{b}\right)^a\right\}, \quad m \geq 0,$$

and thus, here

$$\lambda(r) = \lambda_b \exp\left\{-\left(\frac{2r \sin\alpha}{b}\right)^a\right\}, \quad r > 0,$$

which is a decreasing function of r. The mark density function is

$$f_r(m) = \left(\frac{a}{b}\right)\left(\frac{m}{b}\right)^{a-1} \exp\left\{-\frac{m^a - (2\sin\alpha \cdot r)^a}{b^a}\right\} \mathbf{1}_{\{m \geq 2r\sin\alpha\}}(m)\,, \quad r > 0\,.$$

These functions are plotted in Figure 8.3. It shows how the intensity and the mark density function change in r.

Consider next the case where the marked point process is sampled using a Poisson sampling process N_s with intensity $\lambda_s = 0.001$. The sampling process is simulated in an extended window $[-34.0, 134.0] \times [-34.0, 134.0]$, and the obtained Bitterlich point process N_m is shown in Figure 8.2 (right). The number of points is $n = 578$. The mean of original marks is 0.22 (range $(0.006, 0.68)$), whilst the mean mark of the Bitterlich process is 0.31 (range $(0.04, 0.68)$). The relascope thinning (8.4) retains points with large dbh-marks, see Figure 8.4.

The theoretical local intensity function (8.8) is calculated in a grid in the observation window $[0, 200] \times [0, 200]$, and the intensity surface is shown in Figure 8.5 (left). The intensity is on its highest close to the sample points. The intensity (8.8) is calculated also at each point x_i of the realisation of N_m^B, and the marks of the realisation are plotted against these local intensity values in Figure 8.5 (right): in areas with high point density the mean of the marks is smaller and variance larger than in areas with low point density. Further, the mark summaries (see Table 2.1) calculated from the realisation of N_m^B are shown in Figure 8.6. Clearly, these summaries are not constants in r.

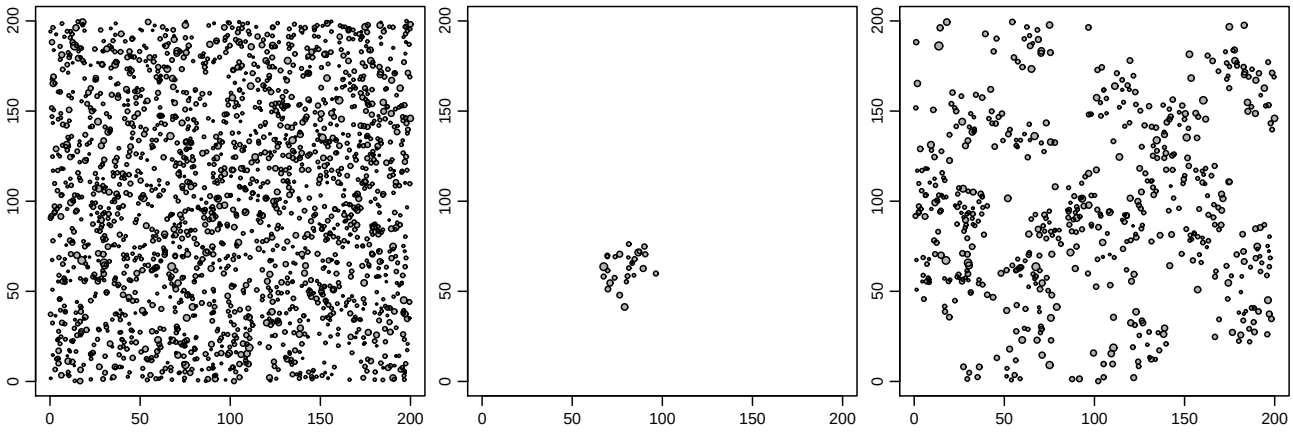

Figure 8.2: Realisations of the base process $N_{b,m}$ (*on the left*), point-centred Bitterlich point process N_m^o (*in the middle*) and stationary Bitterlich point process N_m^B (*on the right*). The diameter of a circle is proportional to the mark.

8.4 Bitterlich sampling applied to a primeval forest

Consider now a real data from a Finnish forest. It is a marked point pattern of a primeval forest in a window of size 50 m × 50 m, see Figure 8.7. The marks are the breast height diameters (dbh) and their range is from 0.01 m to 0.49 m. Consider again that the Bitterlich sampling is applied by a relascope with factor $q = 1$ (m^2/ha), that is $\alpha = 0.01$ rad. First, a sample point is

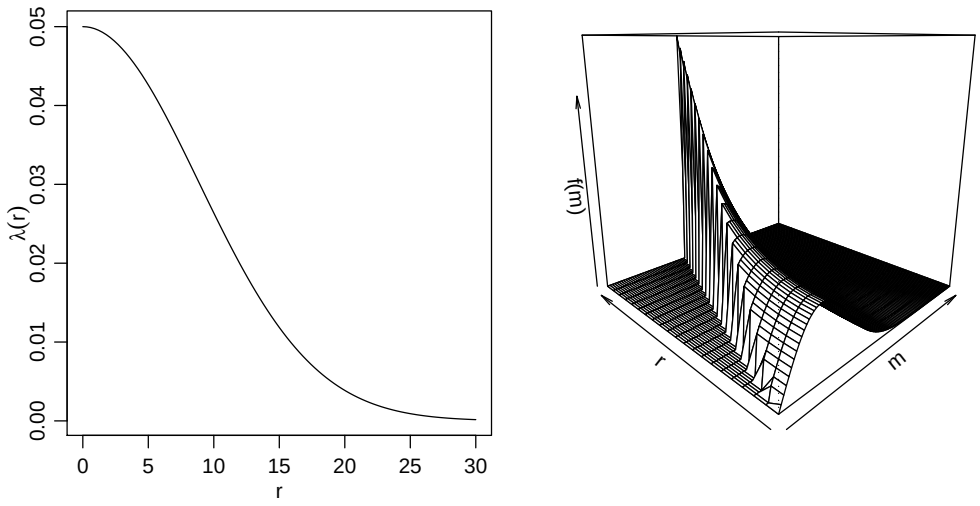

Figure 8.3: The theoretical intensity function $\lambda(r)$ (*on the left*) and mark density function $f_r(m)$ (*on the right*) of the simulation experiment.

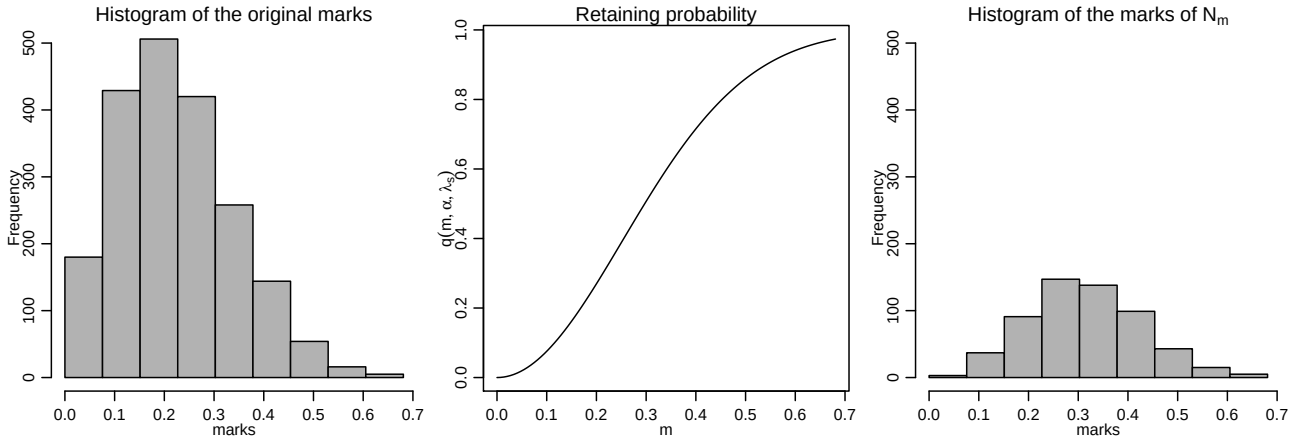

Figure 8.4: The histogram of the marks of the realisation of $N_{b,m}$ of Figure 8.2 (*on the left*), the retaining probability (8.5) as a function of m (*in the middle*) and the histogram of the marks of the realisation of N_m^B of Figure 8.2 (*on the right*).

generated uniformly in the area. The point $(21.41, 32.94)$ is obtained. This results in the finite Bitterlich point pattern shown in the Figure 8.7. Second, the marked point pattern is subject to relascope thinning according to the Poisson sampling process N_s with intensity $\lambda_s = 0.0025$. The obtained Bitterlich point process is shown in Figure 8.7 as well, and the histograms of the original and the Bitterlich point process are shown in Figure 8.8.

8.5 Discussion on the Bitterlich processes

We discuss two new marked point processes which are a result of mark-dependent thinning defined by a relascope. First of all, we give a mathematical characterization for the Bitterlich

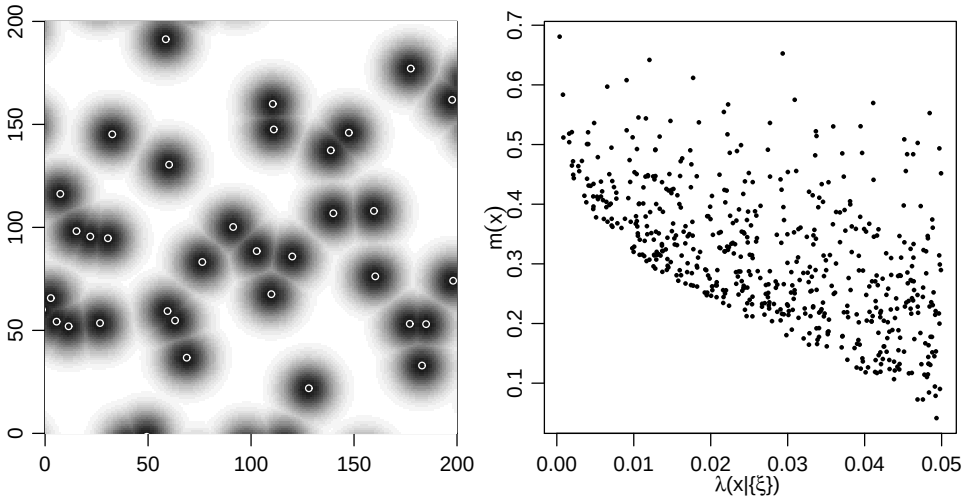

Figure 8.5: *On the left*: The local intensity function $\lambda(s|\{\xi_k\})$, $s \in [0, 200] \times [0, 200]$, of the realisation of N_m^B of Figure 8.2. Dark color corresponds to a high value of the intensity. The points are the sample points. *On the right*: The marks of the realisation of N_m^B are plotted against the values of the local intensity function (8.8) at the points x_i of the pattern.

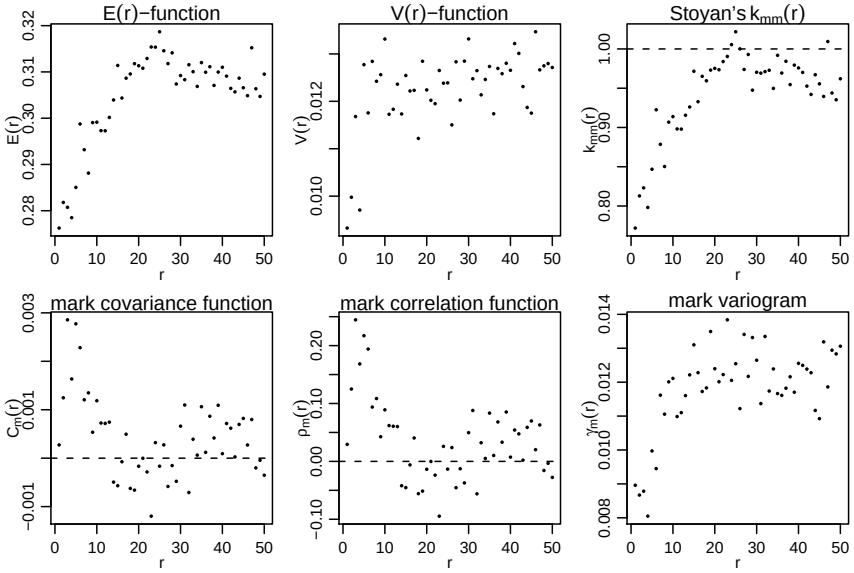

Figure 8.6: The mark summaries calculated from the realisation of N_m^B of Figure 8.2.

plot, which is widely met in practical forestry. The derived descriptions explain what kind of pattern of trees is seen through the relascope.

The point-centred Bitterlich point process is an example of a finite point process which is isotropic but not stationary. First-order characteristics of this process are derived here. Since any isotropic process is θ-stationary, one might further study the second-order properties of the point-centred Bitterlich point process using the approach suggested by Byth (1981) for θ-stationary processes. The marked point-centred Bitterlich point process is also a real example of centred marked point processes.

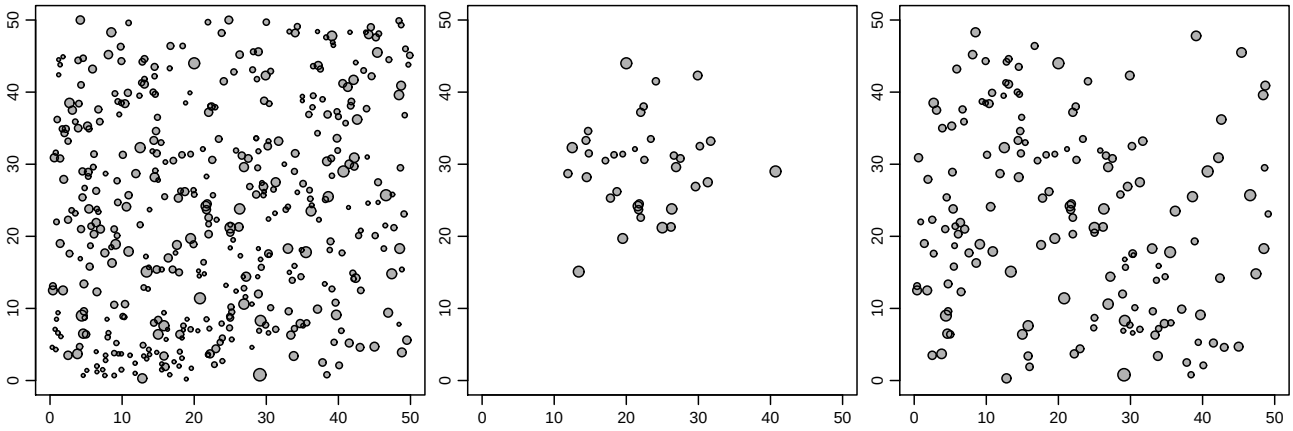

Figure 8.7: A marked point pattern of a primeval forest (*on the left*) and marked point patterns obtained through relascope thinning at a random sample point (*in the middle*) and at Poisson distributed sample points (*on the right*). The diameter of a circle is proportional to the diameter of a tree at breast height.

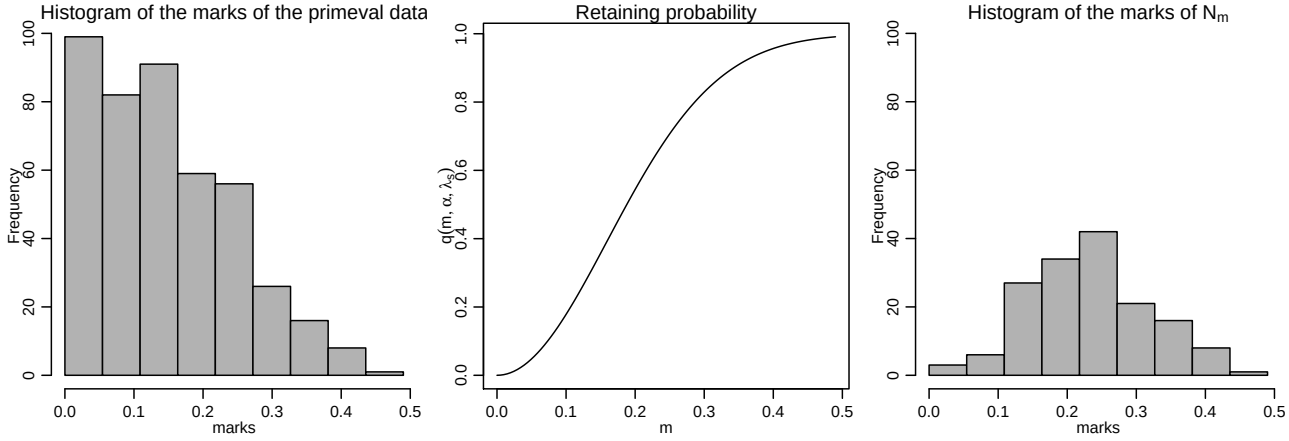

Figure 8.8: The histogram of the diameters (dbh) of the primeval forest of Figure 8.7 (*on the left*), the retaining probability (8.5) as a function of m (*in the middle*) and the histogram of the diameters (dbh) retained in relascope thinning (*on the right*).

The stationary Bitterlich point process is an example of a stationary marked point process where the marks depend locally on the intensity. The local intensity function of this process is defined relative to the sampling process N_s. Extensions for this stationary process can be obtained by replacing the Poisson sampling process N_s by another stationary process.

Chapter 9

Bayesian inference for Gaussian excursion set generated Cox processes with set-marking

Assume $N = \{x_i\}$ is a random set generated Cox process in $W \subseteq \mathbb{R}^d$ with random intensity function $\{\Lambda(s)\}$. This intensity is generated by a random closed set Θ in $\mathbb{R}^d$ as follows: $\Lambda(s) = \lambda_1$ if $s \in \Theta$, and λ_2 elsewhere, see Penttinen and Niemi (2007), Illian et al. (2008) and Section 2.2. This Cox process is a natural model for phenomena where the point intensity varies between two (random) phases that partition $\mathbb{R}^d$ and, conditional on the phases, the points are independently and uniformly distributed.

For a random set generated Cox process, the intensity is not observed, it is a latent structure in the hierarchical model. In some applications, however, the phase at the location of every point x_i of the process can be recorded, which gives information on the underlying random set. An example is the evaluation of forest regeneration outputs, where a random closed set is used as a model for patches caused by the soil treatment, see Section 3.3: The locations x_i of saplings are measured in the field inventory, and in addition, the field data consist of observations whether a sapling at x_i belongs to a treated patch or not. This type of observation provides each point $x_i \in N$ with a mark $m(x_i)$ such that $m(x_i) = 1$ if $x_i \in \Theta$, and $m(x_i) = 2$ if $x_i \in \Theta^c$, the complement of Θ. As a result, a marked point process $N_m = \{[x_i; m(x_i)]\}$, called random set marked Cox process, is obtained. This type of marking is called set-marking by Penttinen and Niemi (2007).

The main objective here is to consider statistical modelling and inference for set-marked Cox processes. The interest lies in the intensity parameters λ_1 and λ_2, describing the point intensities in the two phases, and in the characteristics of Θ, such as area fraction relative to W or covariance. Penttinen and Niemi (2007) suggests an estimation method based on second-order characteristics of point processes (pair correlation function or Ripley's K-function) but do not indicate the uncertainties of the parameter estimates. Their moment-based method does not assume any specific model for the random set.

The inferential approach of this study is fully Bayesian. Markov chain Monte Carlo (MCMC) methods are used in posterior simulation of the hierarchical model. A parametric random set generated Cox process model is adopted for the Bayesian inference. The random set is assumed to be a Gaussian excursion set, or a truncated Gaussian random function in the geostatistical

terminology, see e.g. Lantuéjoul (2002) and Adler (1981). The Gaussian excursion set is a statistical model for the random closed set, suitable to be used as a prior model. Its flexibility is supposed to be an advantage over some alternative constructions such as the germ-grain model (see e.g. Stoyan et al., 1995). The Bayesian modelling gives posterior distributions for the model parameters and allows estimation of the random set.

The new model is implemented and illustrated through simulated data sets. In addition, two pine sapling data sets shown in Figure 3.8 are studied using the new model.

In an earlier study by De Oliveira (2000), a similar Bayesian model was suggested. The difference is, that De Oliveira (2000) employs geostatistical marking where the marks are independent of the point process. Our model is more general allowing the point intensity to depend on the random set. This is an additional complexity in model fitting.

The rest of this chapter is arranged in the following way. The model is explained in Section 9.1. In Section 9.2, the Bayesian modelling of the Gaussian excursion set generated Cox process with information on the phases is explained in full detail, and simulation experiments are given in Section 9.3. Finally, Section 9.4 analyses two real-world data sets on forest research. Section 9.5 is for discussion.

9.1 Gaussian excursion set generated Cox process with set-marking

9.1.1 Gaussian excursion set generated Cox process

A flexible parametric model for the random closed set is obtained by thresholding a Gaussian random field $\{Z(s) : s \in \mathbb{R}^d\}$ by a fixed threshold level l as follows:

$$\Theta = \{s \in \mathbb{R}^d : Z(s) \geq l\}. \tag{9.1}$$

The random set (9.1) is called the *l-level excursion set* of $\{Z(s)\}$ (Lantuéjoul, 2002, p. 205). If the sample paths of $\{Z(s)\}$ are continuous almost surely, then Θ is closed. The mean square differentiability of a Gaussian random field is a sufficient condition for the sample paths to be continuous almost surely, see e.g. Revuz and Yor (1991, p. 26). Another sufficient condition can be found in Adler (1981, p. 62). Because μ_Z and l are linearly dependent, it is assumed without loss of generality that $l = 0$. This choice is also made by De Oliveira (2000).

All the distributional properties of the random set Θ are inherited from the generating Gaussian random field and the threshold level $l = 0$. In particular, if the Gaussian random field is stationary then the generated random closed set is stationary. The same holds for isotropy.

Assume that $\{Z(s) : s \in \mathbb{R}^d\}$ is a stationary Gaussian random field with mean μ_Z, variance σ_Z^2 and with a valid covariance function $C_Z(h)$, $h \in \mathbb{R}^d$, fulfilling $C_Z(o) = \sigma_Z^2$ for $o \in \mathbb{R}^d$, $\|o\| = 0$. Then, the area fraction p and the covariance function $C(h)$ of the random set Θ are

$$p = \mathbf{P}(o \in \Theta) = P(Z(o) \geq 0) = 1 - \Phi\left(-\frac{\mu_Z}{\sigma_Z}\right) = \Phi\left(\frac{\mu_Z}{\sigma_Z}\right) \tag{9.2}$$

and

$$C(h) = \mathbf{P}(o \in \Theta, h \in \Theta) = \mathbf{P}(Z(o) \geq 0, Z(h) \geq 0),$$

respectively, where $o, h \in \mathbb{R}^d$ and Φ is the distribution function of the standard normal distribution.

Assume that $\{Z(s)\}$ is stationary and isotropic. Examples of positive definite covariance functions $C_Z(r)$, $r = \|h\|$, $h \in \mathbb{R}^d$, can be found e.g. in Banerjee et al. (2004), Chilès and Delfiner (1999) and Cressie (1993). These, with variance $\sigma_Z^2 = 1$, include the stable class covariance function

$$C_Z(r) = \exp(-(\beta r)^a) \quad \text{for } r > 0,\ 0 < a \leq 2 \tag{9.3}$$

and the Matérn covariance function

$$C_Z(r) = 2^{1-a}\Gamma(a)^{-1}(\beta r)^a K_a(\beta r) \quad \text{for } r > 0,\ a > 0, \tag{9.4}$$

where $K_a(r)$ is the modified Bessel function. The parameter a in (9.3) and (9.4) controls the roughness of the realisations of the random field. The mean square differentiability of the random field is obtained for the stable covariance function for $a = 2$ only, and for Matérn covariance function for $a > 1$, see e.g. Stein (1999). The parameter $\beta > 0$ is the inverse scale parameter that controls the range of correlation. If $\sigma_Z^2 \neq 1$, then the covariance function (9.3) or (9.4) is multiplied by σ_Z^2.

In the *Gaussian excursion set generated Cox process*, the points are distributed according to the random intensity (2.12) generated by the Gaussian excursion set $\Theta = \{s \in \mathbb{R}^d : Z(s) \geq 0\}$. In addition to fixing the threshold level to $l = 0$, the standard deviation $\sigma_Z = \sqrt{C_Z(0)}$, being a scaling factor of the random field, is confounded with μ_Z and is therefore fixed to $\sigma_Z^2 = 1$.

9.1.2 Set-marking

Consider a point pattern, where the observations consist of the point locations originating from the random set generated Cox process and, in addition, of the knowledge whether a point belongs to the random set Θ or not. The marked point process

$$N_m = \{[x_i; m(x_i)]\}, \tag{9.5}$$

where marks are generated by $m(x_i) = 2 - \mathbf{1}(x_i \in \Theta)$, is called *random set marked Cox process*. An example of set-marking is shown in Figure 9.1.

If the points were independent of the random set, the set-marks would be a representative sample of the random field $\{2 - \mathbf{1}(s \in \Theta) : s \in \mathbb{R}^d\}$. If the interest is in the properties of the random set as in De Oliveira (2000), then the points indeed often serve as sampling points for the random set. This is a connection to mathematical morphology. The situation, where the points depend on the random set, is obviously different and more difficult to handle. The points of a random set generated Cox process depend on the random set, if $\lambda_1 \neq \lambda_2$. Then, in the marked process (9.5), the marks depend locally on the intensity.

9.2 Bayesian modelling

Assume the data $\{[x_i; m(x_i)], i = 1, \ldots, n\}$, denoted by $[X, M]$, are observed in a window $W \subset \mathbb{R}^2$ and let $|W|$ stand for the area of W. The restriction to the two-dimensional space is not essential: a generalization to $\mathbb{R}^d$ with $d > 2$ can be obtained straightforwardly. The objective is to infer the model parameters from $[X, M]$, and to give a prediction for the random set Θ. In what follows, the details and implementation of the Bayesian method are described.

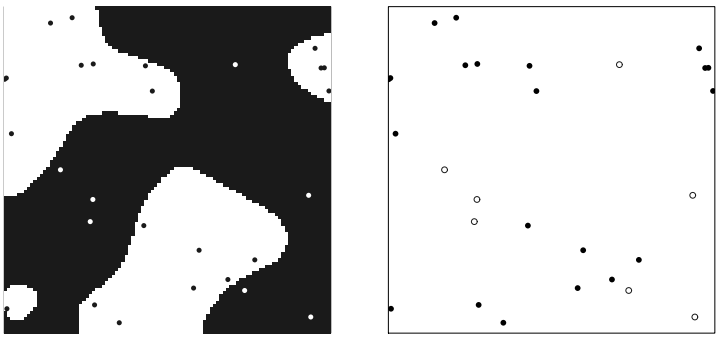

Figure 9.1: An example of set-marking. *On the left*: The random set (white area) and points of a realisation of a random set generated Cox process. Black points are points within the random set, white ones outside. *On the right*: The observed set-marked point pattern.

9.2.1 The likelihood for the set-marked process

Conditional on $\{Z(s)\}$ (or Θ), the likelihood of the Gaussian excursion set generated Cox process is the one of an inhomogeneous Poisson process, see (6.16), and it obtains the form

$$\begin{aligned} &p(\{x_i : i = 1, \ldots, n\} |\, \lambda_1, \lambda_2, \{Z(s)\}) = \\ &= \left[\prod_{i=1}^{n} \lambda_1^{\mathbf{1}(Z(x_i)\geq 0)} \lambda_2^{\mathbf{1}(Z(x_i)<0)}\right] \exp\left\{-\left[\lambda_1 \int_W \mathbf{1}(Z(u) \geq 0)\mathrm{d}u + \lambda_2 \int_W \mathbf{1}(Z(u) < 0)\mathrm{d}u\right]\right\}. \end{aligned}$$

The likelihood for the set-marked point process can be written as

$$\begin{aligned} p([X, M] |\, \lambda_1, \lambda_2, \{Z(s)\}) &= \left\{\prod_{i=1}^{n} [\mathbf{1}(Z(x_i) \geq 0)\mathbf{1}(m(x_i) = 1) + \mathbf{1}(Z(x_i) < 0)\mathbf{1}(m(x_i) = 2)]\right\} \\ &\times \left[\prod_{i=1}^{n} \lambda_1^{\mathbf{1}(m(x_i)=1)} \lambda_2^{\mathbf{1}(m(x_i)=2)}\right] \exp\left\{-\left[\lambda_1 \int_W \mathbf{1}(Z(u) \geq 0)\mathrm{d}u + \lambda_2 \int_W \mathbf{1}(Z(u) < 0)\mathrm{d}u\right]\right\} \\ &= \left\{\prod_{i=1}^{n} [\mathbf{1}(Z(x_i) \geq 0)\mathbf{1}(m(x_i) = 1) + \mathbf{1}(Z(x_i) < 0)\mathbf{1}(m(x_i) = 2)]\right\} \lambda_1^{n_1} \lambda_2^{n_2} \\ &\times \exp\left\{-\left[\lambda_1 \int_W \mathbf{1}(Z(u) \geq 0)\mathrm{d}u + \lambda_2 \int_W \mathbf{1}(Z(u) < 0)\mathrm{d}u\right]\right\}, \end{aligned} \tag{9.6}$$

where the first factor on the right is the condition: for all x_i, $i = 1, \ldots, n$, $Z(x_i) \geq 0$ if $m(x_i) = 1$ and $Z(x_i) < 0$ if $m(x_i) = 2$. If this condition does not hold, then the likelihood is zero. The shorthand notations $n_1 = \sum_{i=1}^{n} \mathbf{1}(m(x_i) = 1)$ and $n_2 = \sum_{i=1}^{n} \mathbf{1}(m(x_i) = 2)$ are employed.

9.2.2 Prior specification

Uniform priors are chosen for λ_1 and λ_2 such that $\lambda_1 \sim \mathrm{Unif}(0, \lambda_{\max,1})$ and $\lambda_2 \sim \mathrm{Unif}(0, \lambda_{\max,2})$, where $\lambda_{\max,1}$ and $\lambda_{\max,2}$ are pre-specified hyperparameters. These uniform priors, denoted

hereafter by $p(\lambda_1)$ and $p(\lambda_2)$, are quite uninformative unless they are strongly restricted by the choices of $\lambda_{\max,1}$ and $\lambda_{\max,2}$.

According to our model choice, $\{Z(s)\}$ is a Gaussian random field with mean μ_Z and covariance function $C_Z(r)$. A suitable parametrized covariance function $C_Z(r)$ is chosen (see the discussion in De Oliveira, 2000): in our forest science application (see Section 9.4) the Matérn covariance function (9.4) with $a = 1.5$ is employed. This correlation function with fixed $a = 1.5$ has one parameter, the inverse scale parameter β, which controls the range of correlation. Consequently, the prior of $\{Z(s)\}$ depends on the parameters β and μ_Z, and it is denoted by $p(\{Z(s)\}|\,\beta, \mu_Z)$. It is specifically dealt with in Section 9.2.3.

The hyperprior $p(\beta)$ is chosen to be Gamma(a_β, b_β) with the pre-specified shape parameter a_β and inverse scale parameter b_β. The normal distribution $N(\mu_\mu, \sigma_\mu^2)$ is used as the hyperprior for μ_Z. These choices imply that $p = \Phi(\mu_\mu)$ a priori.

9.2.3 Implementation

Similarly as in Bayesian modelling for intensity-marked Cox processes, see Section 6.3, the random field $\{Z(s)\}$ must be discretized for computation and it is considered at a finite number of locations. First, let $Z(x_1)$, $Z(x_2), \ldots, Z(x_n)$ be the values of the random field at $x_1, x_2, \ldots, x_n$, the points of the observed point pattern. Second, the Gaussian random field $\{Z(s)\}$ is augmented with the points $s_1, s_2, \ldots, s_k \in \mathcal{G}$. Here $\mathcal{G}$ is a grid covering the area W and being independent of the observed locations. The augmentation with these grid points is needed for approximating the integrals in the likelihood (9.6), and also in order to visualize the random set.

Consequently, let us denote

$$Z = (Z(x_1), Z(x_2), \ldots, Z(x_n), Z(s_1), Z(s_2), \ldots, Z(s_k))^T.$$

In MCMC simulation, the vector Z replaces $\{Z(s)\}$ in the likelihood (9.6). The integral $\int_W \mathbf{1}(Z(u) \geq 0)\mathrm{d}u$ in (9.6) is approximated by $|W|p_Z$, where

$$p_Z = \frac{1}{k} \sum_{s_j \in \mathcal{G}} \mathbf{1}(Z(s_j) \geq 0). \tag{9.7}$$

The approximation is obtained since $s_1, s_2, \ldots, s_k$ can be regarded as sampling locations of $\{Z(s)\}$ in W. Note again, similar to (6.19), that the values $Z(x_i)$ at the observed points x_i are not used in (9.7), because $\{Z(x_i) : i = 1, \ldots, n\}$ is not an unbiased sample of $\{Z(s)\}$. In addition, the prior $p(\{Z(s)\}|\,\beta, \mu_Z)$ is replaced by $p(Z|\,\beta, \mu_Z)$, which is the multinormal distribution with mean μ_Z and the covariance matrix Σ_Z determined by the covariance function $C_Z(r)$ and the distances between locations $x_1, \ldots, x_n, s_1, \ldots, s_k$.

The latent vector Z is included in the set of unknowns. Its values will be updated in the MCMC runs. The posterior is

$$p(\lambda_1, \lambda_2, \beta, \mu_Z, Z|[X, M]) \propto p(\lambda_1)p(\lambda_2)p(\beta)p(\mu_Z)p(Z|\,\beta, \mu_Z)p([X, M]|\,\lambda_1, \lambda_2, Z),$$

and the variables will be updated in the order

$$\lambda_1, \lambda_2, \beta, \mu_Z, Z(x_1), Z(x_2), \ldots, Z(x_n), Z(s_1), Z(s_2), \ldots, Z(s_k). \tag{9.8}$$

Here each variable is updated one at a time, but more efficient block-updates could be used for Z, see Rue and Held (2005). The updating is done using Metropolis-Hastings steps. Let the notation (9.8) stand for the current values in the MCMC chain and '*' refer to the proposal. The following proposal distributions are considered:

$$\begin{aligned}
\lambda_1^* &\sim \text{Unif}(\max(0, \lambda_1 - \delta_1), \min(\lambda_1 + \delta_1, \lambda_{\max,1})) \\
\lambda_2^* &\sim \text{Unif}(\max(0, \lambda_2 - \delta_2), \min(\lambda_2 + \delta_2, \lambda_{\max,2})) \\
\beta^* &\sim \text{Unif}(\max(0, \beta - \delta_\beta), \beta + \delta_\beta) \\
\mu_Z^* &\sim N(\mu_Z, \sigma_{\mu q}^2) \\
Z^*(x_i) &\sim N(Z(x_i), \sigma_x^2) \quad \text{for } i = 1, 2, \ldots, n \\
Z^*(s_j) &\sim N(Z(s_j), \sigma_s^2) \quad \text{for } j = 1, 2, \ldots, k,
\end{aligned}$$

where $\delta_1, \delta_2, \delta_\beta, \sigma_{\mu q}, \sigma_x, \sigma_s$ are proper (constant) values for the parameters of the proposal distributions set by the user. Note that every update of the value $Z(s_j)$, $j = 1, \ldots, k$, changes p_Z calculated from (9.7). This value p_Z again affects the updating of λ_1, λ_2, and $Z(s_j)$, $j = 1, \ldots, k$, through the acceptance probabilities. Further, note that $Z^*(x_i)$ can only be accepted if it fulfils the condition given by $m(x_i)$. The acceptance probabilities are given in Section 9.2.4.

The MCMC algorithm described above is applicable as such for small data. The updating of β becomes slow if the dimension of Σ_Z is high, which is due to large number observed points or a dense grid used in the augmentation of the Z-values. This computationally tedious updating is avoided by suggesting values for β only from a set $B = \{\beta_1, \beta_2, \ldots, \beta_h\}$, $\beta_1 < \beta_2 < \cdots < \beta_h$, $h \in \mathbb{N}, h \geq 3$. Often small changes in β are not meaningful and, considering only a finite number of values for β, the corresponding covariance matrices and their inverses can be calculated only once at the beginning of the algorithm. A disadvantage may be poorer mixing. We transform the gamma prior $p(\beta)$ on $(0, \infty)$ into a discrete prior distribution on B as $p_B(\beta_i) = p(\beta_i) / \sum_{j=1}^h p(\beta_j)$ for $i = 1, \ldots, h$. The scaling factor cancels out in the Metropolis-Hastings ratio. In the discretization of β, the proposal distribution is changed. Let β_t be the current value in the chain and Y a random variable for which $\mathbf{P}(Y = -1) = \mathbf{P}(Y = 1) = 1/2$. Then, the new value is suggested according to a random walk in B:

$$\beta^* = \begin{cases} \beta_{\max(1,\, t+Y)}, & \text{if } t = 1, \\ \beta_{t+Y}, & \text{if } 1 < t < h, \\ \beta_{\min(t+Y,\, h)}, & \text{if } t = h. \end{cases} \tag{9.9}$$

The Bayesian estimation results in posterior distributions for the two intensities and for the parameters of the random field. In addition, a prediction for the random set is obtained as the posterior mean of $\mathbf{1}(Z(s_j) \geq 0)$, $j = 1, \ldots, k$. Assume that T iterations have been performed (omitting the burn-in time) and $Z_1(s_j), Z_2(s_j), \ldots, Z_T(s_j)$ are the corresponding simulated values for $Z(s_j)$, $j = 1, \ldots, k$. The Bayesian estimate for the probability $\mathbf{P}(s_j \in \Theta)$ is then obtained by

$$\hat{\mathbf{P}}(s_j \in \Theta) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}(Z_t(s_j) \geq 0), \quad s_j \in \mathcal{G}. \tag{9.10}$$

These probabilities could be further used as a basis for classification.

The values for λ_1, λ_2, β and μ_Z are stored, whilst $Z(x_i)$ and $Z(s_j)$ are recorded for examination only with three different values of i and j due to the (large) dimension of Z. These three values are sampled from $\{1, \ldots, n\}$ and $\{1, \ldots, k\}$, respectively. For all $j \in \{1, \ldots, k\}$, the probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$ are calculated iteratively in the MCMC run.

9.2.4 Acceptance probabilities

In the following, let $\lambda_1, \lambda_2, \beta, \mu_Z, Z$ be the current values in the chain. In addition, let Σ_Z be the covariance matrix of Z determined by the covariance function with the range parameter β, $|\Sigma_Z|$ its determinant and p_Z the estimate (9.7) calculated from the current Z. The proposals for the new states are denoted by $\lambda_1^*, \lambda_2^*, \beta^*, \mu_Z^*$ and Z^*. The values of Z are updated one at a time and, thus, Z^* stands for the vector where one of the components have been changed with respect to Z. The updating of Z (at grid locations) affects p_Z. Thus, let the notation p_Z^* refer to (9.7) calculated from Z^*. Further, the updating of β affects Σ_Z: let Σ_Z^* be the covariance matrix of Z determined by the covariance function with β^*.

The acceptance probabilities for the proposals λ_1^*, λ_2^*, β^* by (9.9), μ_Z^* and Z^* are minimum of 1 and

$$r_{\lambda_1} = \left(\frac{\lambda_1^*}{\lambda_1}\right)^{n_1} e^{|W|p_Z(\lambda_1-\lambda_1^*)} \cdot \frac{\min(\lambda_1+\delta_1, \lambda_{\max,1}) - \max(0, \lambda_1-\delta_1)}{\min(\lambda_1^*+\delta_1, \lambda_{\max,1}) - \max(0, \lambda_1^*-\delta_1)},$$

$$r_{\lambda_2} = \left(\frac{\lambda_2^*}{\lambda_2}\right)^{n_2} e^{|W|(1-p_Z)(\lambda_2-\lambda_2^*)} \cdot \frac{\min(\lambda_2+\delta_2, \lambda_{\max,2}) - \max(0, \lambda_2-\delta_2)}{\min(\lambda_2^*+\delta_2, \lambda_{\max,2}) - \max(0, \lambda_2^*-\delta_2)},$$

$$r_\beta = \left(\frac{\beta^*}{\beta}\right)^{a_\beta - 1} e^{b_\beta(\beta-\beta^*)} \left(\frac{|\Sigma_Z|}{|\Sigma_Z^*|}\right)^{\frac{1}{2}} e^{\frac{1}{2}(Z-\mathbf{1_v}\mu_Z)^T(\Sigma_Z^{-1}-(\Sigma_Z^*)^{-1})(Z-\mathbf{1_v}\mu_Z)},$$

$$r_{\mu_Z} = e^{\frac{1}{2}\left[((\mu_Z-\mu_\mu)^2-(\mu_Z^*-\mu_\mu)^2)/\sigma_\mu^2+(\mathbf{1_v}(\mu_Z^*-\mu_Z))^T\Sigma_Z^{-1}(Z\cdot 2-\mathbf{1_v}(\mu_Z^*+\mu_Z))\right]}$$

and

$$r_Z = e^{\frac{1}{2}\left[(Z-Z^*)^T\Sigma_Z^{-1}(Z+Z^*-\mathbf{1_v}\cdot 2\mu_Z)\right]+|W|(\lambda_1-\lambda_2)(p_Z-p_Z^*)},$$

if $\prod_{i=1}^n [\mathbf{1}(Z^*(x_i) \geq 0)\mathbf{1}(m(x_i)=1) + \mathbf{1}(Z^*(x_i)<0)\mathbf{1}(m(x_i)=2)] = 1$, and 0 otherwise. Note that when $Z(x_i)$ is updated then $p_Z^* = p_Z$. Above $\mathbf{1_v} = (1, 1, \ldots, 1)^T$ is a (column) vector of length $n+k$.

9.3 A simulation study

Six realisations of the Gaussian excursion set marked Cox process are simulated in a window W of size $[0, 10] \times [0, 10]$ using different parameter combinations. The aim of the experimenting is to demonstrate the model and the functionality of the posterior computation. We consider a Gaussian random field $\{Z(s)\}$ with mean $\mu_Z = -0.25$ and Matérn covariance function (9.4) where $a = 1.5$. The values $\beta = 1$ and $\beta = 2$ are addressed to the inverse scale parameter of (9.4). This leads to two realisations of $\{Z(s)\}$, which are simulated on a fine 200×200 grid. After computing the excursion set $\Theta = \{s \in W : Z(s) \geq 0\}$ from a realisation of $\{Z(s)\}$, an inhomogeneous Poisson process having the intensity (2.12) is simulated. The following intensity parameter couples are considered: ($\lambda_1 = 6.5$, $\lambda_2 = 2.0$), ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$) and ($\lambda_1 = 3.8$, $\lambda_2 = 3.8$). In each case the total intensity is $\lambda = p\lambda_1 + (1-p)\lambda_2 \approx 3.8$ as the (expected) area fraction of the excursion sets equals $p = \Phi(-0.25) \approx 0.40$.

The study design leads to six different Gaussian excursion set marked Cox processes. The simulated random sets with $\beta = 1$ and $\beta = 2$ and two corresponding marked point patterns $N_m = \{[x_i; m(x_i)]\}$ with ($\lambda_1 = 6.5$, $\lambda_2 = 2.0$) are illustrated in Figures 9.2 and 9.3 (left). The corresponding patterns for ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$) are shown in Figures 9.4 and 9.5 (left), and for ($\lambda_1 = 3.8$, $\lambda_2 = 3.8$) in Figures 9.6 and 9.7 (left).

Assuming that N_m is observed, the Bayesian analysis is performed with the following parameter values of the prior and proposal distributions: $\delta_1 = 2.0, 2.0, 1.5$, $\delta_2 = 0.6, 0.8, 1.5$ corresponding respectively to ($\lambda_1 = 6.5$, $\lambda_2 = 2.0$), ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$) and ($\lambda_1 = 3.8$, $\lambda_2 = 3.8$), and $a_\beta = 400, 1600$, $b_\beta = 400, 800$, $B = \{0.20, 0.25, 0.30, \ldots, 1.80\}$, $\{1.20, 1.25, 1.30, \ldots, 2.80\}$ corresponding respectively to $\beta = 1$ and $\beta = 2$. In addition, the values $\lambda_{\max,1} = \lambda_{\max,2} = 20$, $\mu_\mu = -0.25, \sigma_\mu = 0.5, \sigma_{\mu q} = 0.5$, $\sigma_x = 0.3$ and $\sigma_s = 0.4$ are used throughout.

For approximating the integrals in the likelihood (9.6), the process $\{Z(s)\}$ is considered on $\mathcal{G}$. The number of grid points is a compromise between the precision of the approximation and computational burden. A good choice would be to relate it to the range of correlation of $\{Z(s)\}$. We employ a 20×20 grid (0.5×0.5 grid cell) for $\beta = 1$ and a 40×40 grid (0.25×0.25 grid cell) for $\beta = 2$. For comparison, we also employ the 40×40 grid for $\beta = 1$.

The MCMC algorithm is run for 20,000 iterations in each case. Initial values for λ_1, λ_2 and μ_Z are simulated from the corresponding prior distributions and for β uniformly in B. In addition, Z is initialized such that the likelihood (9.6) is positive for the initial values of $Z(x_1), \ldots, Z(x_n)$: n random variables are simulated from $N(0, 0.5)$, the absolute values of these are taken and the signs fixed in accordance with $m(x_i)$. The initial values for $Z(s_1), \ldots, Z(s_k)$ are then simulated conditional on $Z(x_1), \ldots, Z(x_n)$ using the initial parameter values of μ_Z and β. In this initialization, an additional nugget effect 0.05 is used to add flexibility to the initial field. Simulations of the Gaussian random field are performed using the function `GaussRF` and conditional simulation using the function `CondSimu`. Both of these functions exist in the R library `RandomFields`, see Schlather (2001b).

The chains seem to reach the balance, which is concluded by visual examination of the trace plots. Further three adjacent runs were made for each parameter combination using different initial values, and these separate runs gave mutually consistent results. The marginal posterior distributions are described by means of their means, standard deviations and 90% posterior intervals. These descriptions calculated from the last 10,000 iterations of the first runs are shown in Table 9.1. The posterior of p is computed using (9.2).

Some limited empirical observations can be drawn from the results shown in Table 9.1. In these simulations with $\beta = 1$, using the 20×20 grid λ_1 was estimated in each case larger and λ_2 smaller than using the 40×40 grid. Here the larger number of grid points leads to slightly better results, but the computation time is multiplied. The posterior mean fields of the probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$ corresponding to the parameter values $\beta = 1$ and $\beta = 2$, respectively, and ($\lambda_1 = 6.5$, $\lambda_2 = 2.0$) are plotted in Figures 9.2 (middle and right) and 9.3 (right). The corresponding posterior mean fields for ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$) are shown in Figures 9.4 and 9.5, and for ($\lambda_1 = 3.8$, $\lambda_2 = 3.8$) in Figures 9.6 and 9.7. It can be concluded that the uncertainty is highest on the borders of the random set, but the sets are reconstructed properly. Also the 20×20 grid used with $\beta = 1$ may be considered adequate for describing the random set, see Figures 9.2, 9.4 and 9.6 (middle).

We further experimented with the models with lower intensities ($\lambda_1 = 3.25$, $\lambda_2 = 1.0$) and ($\lambda_1 = 2.0$, $\lambda_2 = 2.0$) and these simulations worked as well, but a reduction in the information leads to greater variability, of course. We also experimented with less restrictive priors for β. Also a slightly less informative prior seems to work fine for models of Table 9.1, but much longer simulation chains are required due to lower mixing. For the model with $\beta = 1$ and ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$), we run the algorithm for 200,000 iterations using the 20×20 grid and values $a_\beta = 25$ and $b_\beta = 25$. The posterior means (standard deviations) from the last

Table 9.1: Results for the simulation experiment. Posterior means (standard deviations) and 90% posterior intervals for the model parameters. The value $\mu_Z = -0.25$ is used in the simulation.

parameters of simulation		λ_1	λ_2	β	μ_Z	p
$\beta = 1$, 20×20 grid	($\lambda_1 = 6.5$, $\lambda_2 = 2.0$)	7.07 (0.53) 6.24, 7.98	1.88 (0.18) 1.60, 2.19	1.01 (0.04) 0.95, 1.05	-0.24 (0.27) -0.69, 0.21	0.41 (0.10) 0.25, 0.58
	($\lambda_1 = 5.0$, $\lambda_2 = 3.0$)	5.28 (0.43) 4.62, 6.01	2.82 (0.22) 2.48, 3.20	1.00 (0.04) 0.95, 1.05	-0.27 (0.27) -0.72, 0.19	0.40 (0.10) 0.24, 0.57
	($\lambda_1 = 3.8$, $\lambda_2 = 3.8$)	4.45 (0.37) 3.84, 5.08	3.63 (0.26) 3.21, 4.07	0.98 (0.04) 0.90, 1.05	-0.17 (0.27) -0.62, 0.28	0.44 (0.10) 0.27, 0.61
$\beta = 1$, 40×40 grid	($\lambda_1 = 6.5$, $\lambda_2 = 2.0$)	6.70 (0.42) 6.02, 7.40	1.94 (0.18) 1.66, 2.26	1.01 (0.03) 0.95, 1.05	-0.23 (0.27) -0.66, 0.21	0.41 (0.10) 0.25, 0.58
	($\lambda_1 = 5.0$, $\lambda_2 = 3.0$)	5.00 (0.40) 4.38, 5.67	2.92 (0.23) 2.55, 3.32	1.00 (0.04) 0.95, 1.05	-0.19 (0.27) -0.63, 0.25	0.43 (0.10) 0.27, 0.60
	($\lambda_1 = 3.8$, $\lambda_2 = 3.8$)	4.27 (0.34) 3.73, 4.85	3.73 (0.26) 3.32, 4.16	1.09 (0.05) 1.00, 1.15	-0.13 (0.26) -0.57, 0.29	0.45 (0.10) 0.28, 0.61
$\beta = 2$, 40×40 grid	($\lambda_1 = 6.5$, $\lambda_2 = 2.0$)	7.02 (0.50) 6.23, 7.86	2.02 (0.19) 1.73, 2.35	2.00 (0.05) 1.95, 2.10	-0.23 (0.18) -0.54, 0.06	0.41 (0.07) 0.29, 0.52
	($\lambda_1 = 5.0$, $\lambda_2 = 3.0$)	5.26 (0.43) 4.56, 5.97	3.43 (0.25) 3.03, 3.85	1.99 (0.05) 1.90, 2.05	-0.25 (0.18) -0.55, 0.05	0.40 (0.07) 0.29, 0.52
	($\lambda_1 = 3.8$, $\lambda_2 = 3.8$)	3.18 (0.32) 2.69, 3.73	4.02 (0.31) 3.53, 4.53	2.00 (0.05) 1.90, 2.10	-0.18 (0.19) -0.48, 0.13	0.43 (0.07) 0.31, 0.55

100,000 iterations are: $\lambda_1 = 5.11\,(0.42)$, $\lambda_2 = 2.88\,(0.23)$, $\beta = 1.04\,(0.11)$, $\mu_Z = -0.21\,(0.27)$, $p = 0.42\,(0.10)$. The difference to the result shown in Table 9.1 is that, with the less informative prior, the standard deviation of the marginal posterior distribution of β is much larger.

In some simulations, we experienced that if the random initial value for β was unreasonably small, then β did not start to update properly. It might not be a good practice to start with a very small value of β (or strong spatial correlation).

In addition, we studied the effect of the discretization of β by allowing β to vary continuously for the model with $\beta = 1$ and ($\lambda_1 = 5.0$, $\lambda_2 = 3.0$) using the 20×20 grid. We used the same prior and proposal distribution parameters as in the case with discrete β except the proposal $\beta^* \sim \text{Unif}(\max(0, \beta - 0.2), \beta + 0.2)$ was used. The posterior means (standard deviations) from the last 10,000 iterations out of 20,000 are: $\lambda_1 = 5.28\,(0.43)$, $\lambda_2 = 2.82\,(0.23)$, $\beta = 1.00\,(0.04)$, $\mu_Z = -0.28\,(0.28)$, $p = 0.39\,(0.10)$. Thus, the means and standard deviations of the posterior distributions with a continuous distribution for β are very similar to those with a discrete distribution. On the basis of our very limited experience, we conclude that the discretization of β does not affect the results remarkably, and therefore, its use is recommended for computational reasons.

9.4 Modelling pine saplings growing in treated soil

The new Bayesian model is used to analyse two marked point patterns of pine saplings, both observed in a window of size 25 m $\times$ 25 m, see Figure 3.8. It is known a priori that the patches tend to be smaller than 1 m in diameter. Consequently, the grid $\mathcal{G}$ must be dense enough to

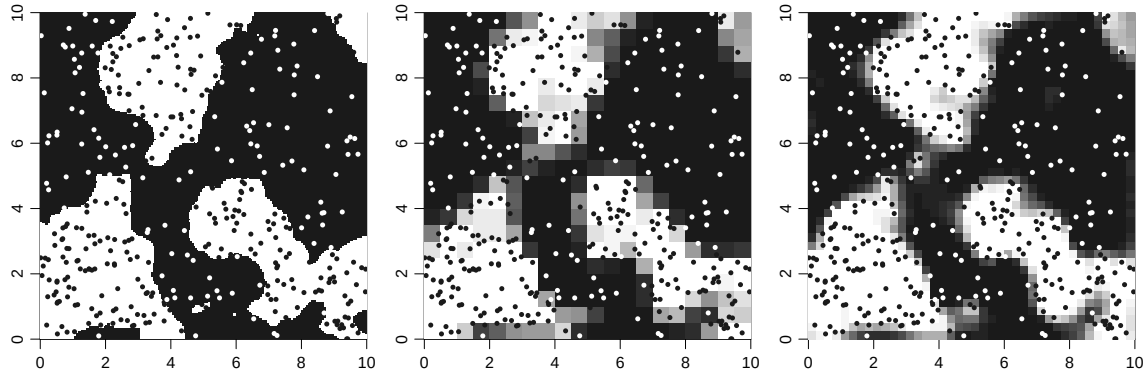

Figure 9.2: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 1$) with random set marked Cox process N_m where $\lambda_1 = 6.5$ and $\lambda_2 = 2.0$. Black points are points within the random set (white area), white ones outside. *In the middle* ($\mathcal{G}$ is 20×20 grid) and *on the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m. Light grey means high probability, dark grey small probability.

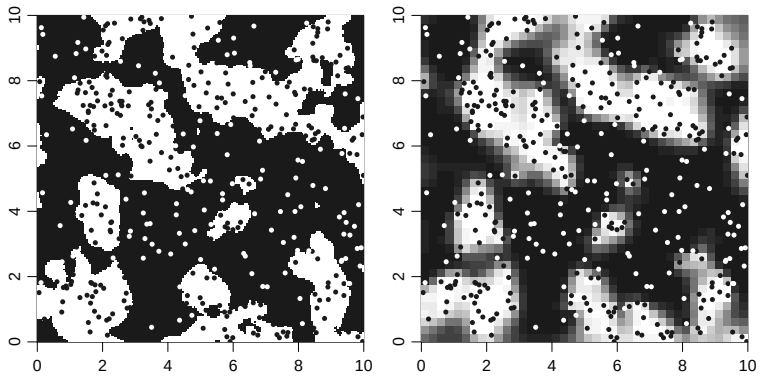

Figure 9.3: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 2$) with random set marked Cox process N_m where $\lambda_1 = 6.5$ and $\lambda_2 = 2.0$. *On the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m.

trace such small sets. The grid is taken to have 50×50 cells, or equivalently, a grid cell is of size 0.5 m $\times$ 0.5 m. Further, the Matérn covariance function (9.4) with $a = 1.5$ is used for the Gaussian random field.

A quite strong prior is used for β: $\beta \sim \text{Gamma}(1600, 800)$ with mean 2.00 and standard deviation 0.05, which is justified by a priori knowledge of the sizes of patches. In addition, the following prior distributions are considered: $\lambda_1 \sim \text{Unif}(0, 5)$, $\lambda_2 \sim \text{Unif}(0, 5)$ and $\mu_Z \sim N(-0.5, 0.25)$. The parameters of the proposal distributions are decided by preliminary testing to be the following: $\delta_1 = 0.5$, $\delta_2 = 0.1$, $\sigma_{\mu q} = 0.5$, $\sigma_x = 0.4$ and $\sigma_s = 0.7$. The values $B = \{1.00, 1.05, 1.10, 1.15, \ldots, 3.00\}$ are assigned for β. The initial values are simulated as in the simulation experiment of Section 9.3.

A total of 200,000 iterations is performed for the first sapling data set. The MCMC chains for the parameters λ_1, λ_2, β and μ_Z are shown in Figure 9.8. The marginal posterior distributions are described through the means, standard deviations and 90% posterior intervals obtained

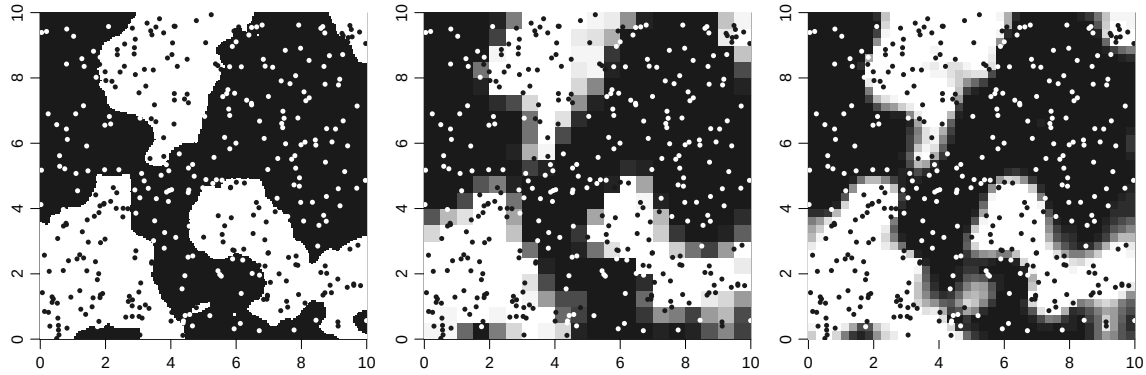

Figure 9.4: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 1$) with random set marked Cox process N_m where $\lambda_1 = 5.0$ and $\lambda_2 = 3.0$. Black points are points within the random set (white area), white ones outside. *In the middle* ($\mathcal{G}$ is 20×20 grid) and *on the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m. Light grey means high probability, dark grey small probability.

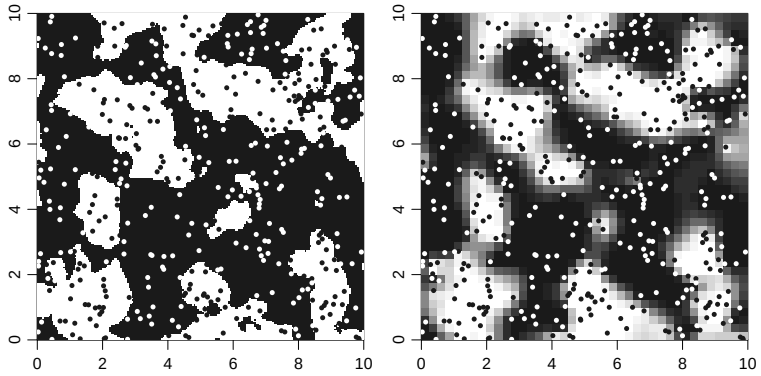

Figure 9.5: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 2$) with random set marked Cox process N_m where $\lambda_1 = 5.0$ and $\lambda_2 = 3.0$. *On the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m.

from the last 100,000 iterations, see Table 9.2. The probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$ are plotted in Figure 9.10 (left).

Similar chains are run for the second sapling data set. Again an MCMC sample of size 200,000 is collected, see Figure 9.9. The descriptions of the marginal posteriors shown in Table 9.2 are obtained discarding the first 100,000 iterations as burn-in. The result concerning the random set is shown in Figure 9.10 (right). For both data sets, three adjacent runs of the algorithm were performed and the runs gave consistent results.

In the second sapling data set, there are two vertical ditches where no saplings grow. This partly explains, why the estimates for λ_1 and λ_2 in the first data set are slightly larger than in the second one. The parameter μ_Z, and thus p, is estimated to be a bit larger in the second data set. The parameter β of the Matérn covariance function (9.4) is estimated to be around 2.05 and 2.08 in the data sets. This means that the spatial correlation vanishes (is less than 0.05) approximately at distances larger than 2.3 m. Despite the small differences in these data

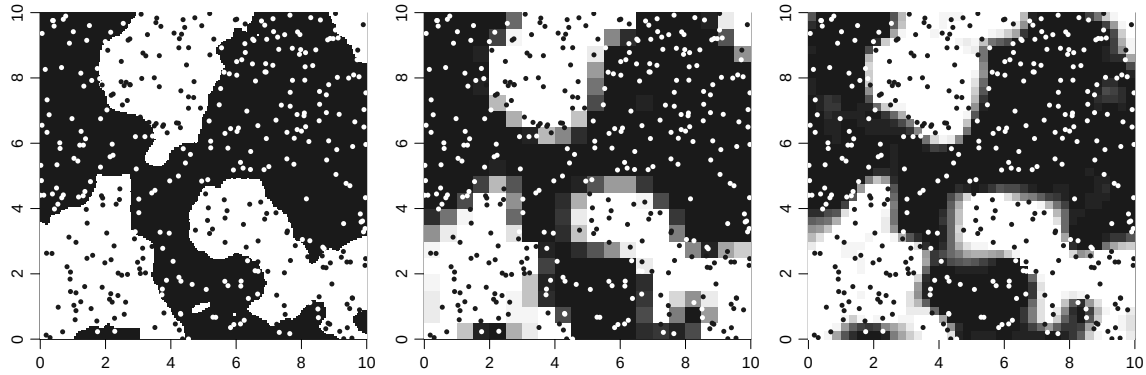

Figure 9.6: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 1$) with random set marked Cox process N_m where $\lambda_1 = 3.8$ and $\lambda_2 = 3.8$. Black points are points within the random set (white area), white ones outside. *In the middle* ($\mathcal{G}$ is 20×20 grid) and *on the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m. Light grey means high probability, dark grey small probability.

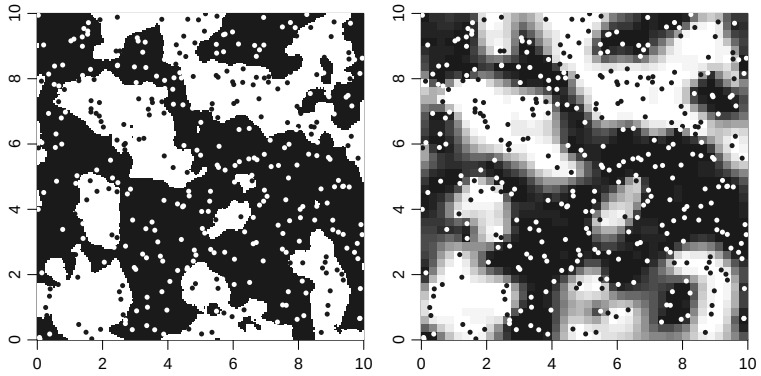

Figure 9.7: *On the left*: the simulated Gaussian excursion set ($\mu_Z = -0.25$, $\beta = 2$) with random set marked Cox process N_m where $\lambda_1 = 3.8$ and $\lambda_2 = 3.8$. *On the right* ($\mathcal{G}$ is 40×40 grid): posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, with points of N_m.

Table 9.2: Results for the pine sapling data sets. Posterior means (standard deviations) and 90% posterior intervals for the model parameters.

	λ_1	λ_2	β	μ_Z	p
sapling data set 1	1.73 (0.48)	0.18 (0.02)	2.08 (0.05)	-0.64 (0.16)	0.26 (0.05)
	1.11, 2.65	0.14, 0.22	2.00, 2.15	-0.90, -0.36	0.19, 0.36
sapling data set 2	1.32 (0.26)	0.13 (0.02)	2.05 (0.05)	-0.43 (0.15)	0.33 (0.05)
	0.98, 1.82	0.10, 0.17	1.95, 2.15	-0.68, -0.20	0.25, 0.42

sets, the results are mutually consistent. The obtained area fraction is also in accordance with the knowledge of the density of patches. The tree density in the tilled area is estimated to be

9 to 10 times larger than in the untilled area, see Table 9.2.

9.5 Discussion on Bayesian inference for set-marked data

Here a Bayesian method is proposed for statistical reasoning on point pattern data. The model employed is the random set generated Cox process but, in addition to the point locations, the observations consist of the knowledge of whether a point belongs to the random set or not. An interesting feature of this type of data, called here set-marked data, is that the knowledge of the phase at the point locations improves the possibilities of controlling global features of the random set, whilst the covariance of the random set has the role of (local) smoothing.

The Gaussian excursion set is used as a model for the random set. The Matérn covariance function (9.4) with a fixed smoothness parameter a is used as the covariance function of the generating random field. It is probably not possible to deduce the form of the covariance function from the point pattern data (with or without set-marking), which is discussed in Diggle et al. (1998) and in De Oliveira (2000). Instead, it is possible to estimate the inverse scale parameter (our β) from the data, but this may be sensitive to "odd" point configurations.

We have made a lot of simulations with models using the Gaussian covariance function, that is the stable covariance function (9.3) with $a = 2$, and our simulations always resulted in overestimation of the inverse scale parameter β, even if quite an informative prior was used for β. Overestimation of β further tends to lead to overestimation of λ_1 and underestimation of λ_2, μ_Z and p. Similar phenomena may be encountered if one uses a large value for the smoothness parameter a in the Matérn covariance function (9.4), which leads to smoother random fields. At least, according to our limited experience, very smooth structure may cause very slow convergence of the MCMC algorithm. The slow convergence for smooth random field structures was also observed in the MCMC algorithm of the intensity-marked log Gaussian Cox processes, see Section 6.4. The use of Gaussian covariance function has also been criticized by Diggle et al. (1998) because it may lead to singular covariance matrices for the finite distributions.

If there is knowledge of the size of the classification error made in measuring the set-marks, this misclassification can be included into the model. Let ϵ_1 be the probability to observe $m(x_i) = 2$ when $Z(x_i) \geq 0$ and ϵ_2 be the probability to observe $m(x_i) = 1$ when $Z(x_i) < 0$. Then the likelihood can be written as

$$\begin{aligned} p([X, M] |\, \lambda_1, \lambda_2, \{Z(s)\}) = {} & (1-\epsilon_1)^{\sum_{i=1}^n \mathbf{1}(Z(x_i)\geq 0)\mathbf{1}(m(x_i)=1)} \epsilon_1^{\sum_{i=1}^n \mathbf{1}(Z(x_i)\geq 0)\mathbf{1}(m(x_i)=2)} \\ & \times \epsilon_2^{\sum_{i=1}^n \mathbf{1}(Z(x_i)<0)\mathbf{1}(m(x_i)=1)} (1-\epsilon_2)^{\sum_{i=1}^n \mathbf{1}(Z(x_i)<0)\mathbf{1}(m(x_i)=2)} \\ & \times \lambda_1^{\sum_{i=1}^n \mathbf{1}(Z(x_i)\geq 0)} \lambda_2^{\sum_{i=1}^n \mathbf{1}(Z(x_i)<0)} \\ & \times \exp\left\{ -\left[\lambda_1 \int_W \mathbf{1}(Z(u) \geq 0) \mathrm{d}u + \lambda_2 \int_W \mathbf{1}(Z(u) < 0) \mathrm{d}u \right] \right\} . \end{aligned}$$

With $\epsilon_1 = \epsilon_2 = 0$ the likelihood (9.6) is obtained. The acceptance probabilities for λ_1, λ_2 and Z differ from those in the algorithm for the model without observation errors, but otherwise the algorithm described in Section 9.2 may be used. In our experimenting, this extension did not essentially affect the results with simulated data, where the classification was precise. As regards the sapling data sets, we believe that the measurements have been made with high precision.

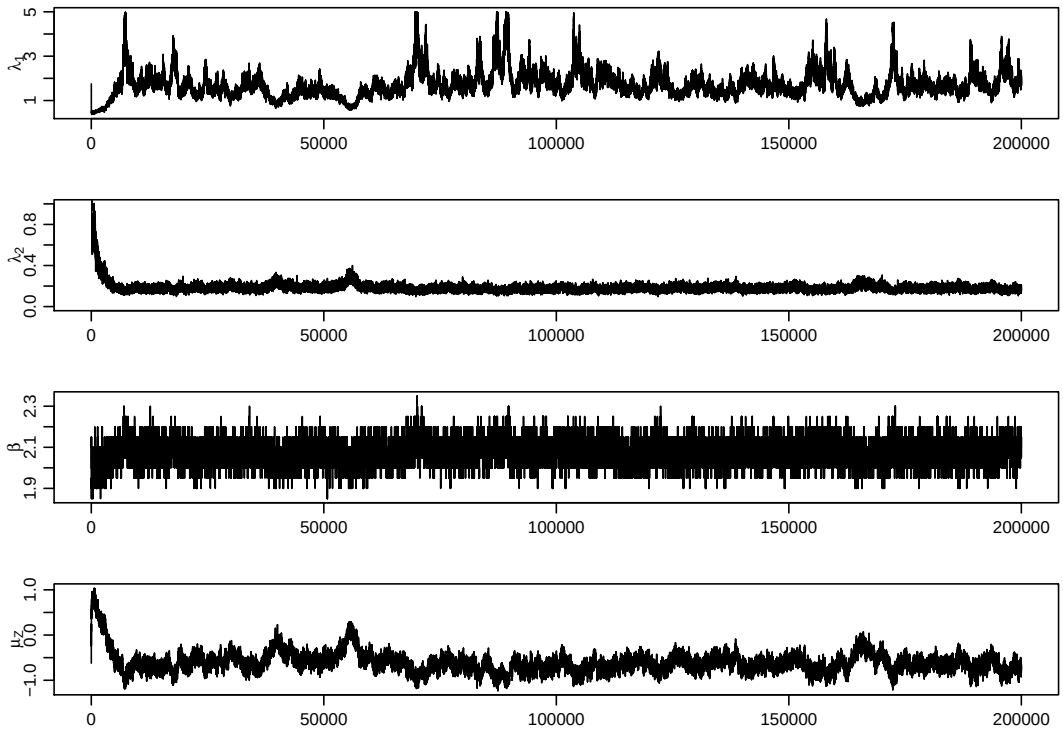

Figure 9.8: The MCMC chains for the first sapling data set.

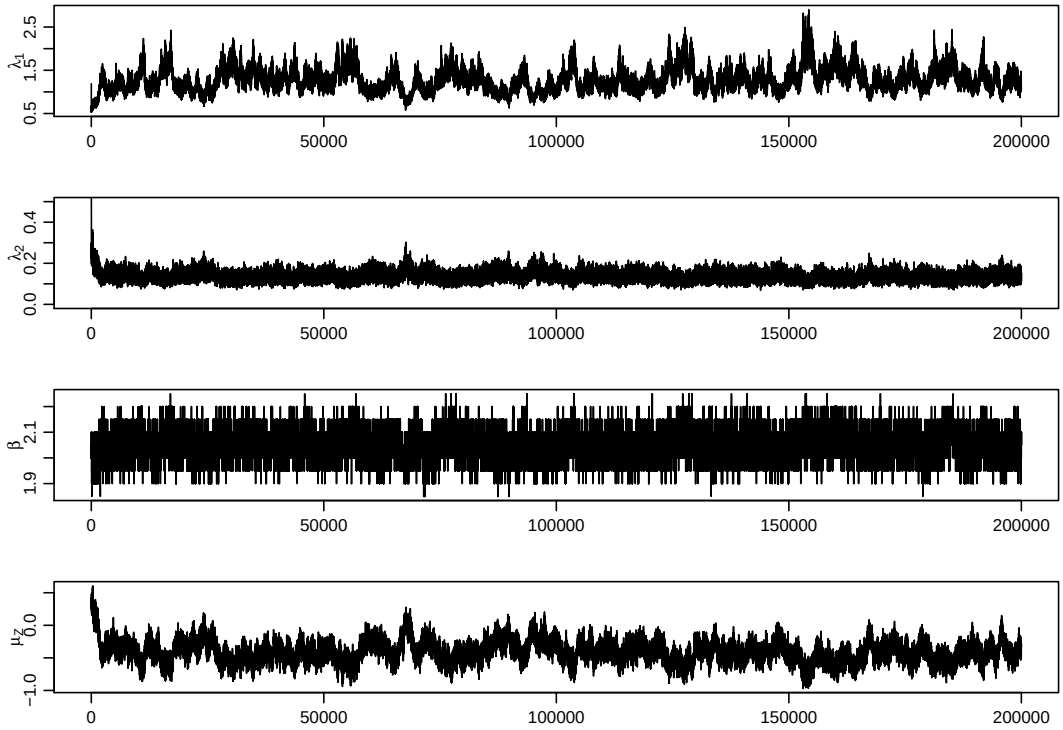

Figure 9.9: The MCMC chains for the second sapling data set.

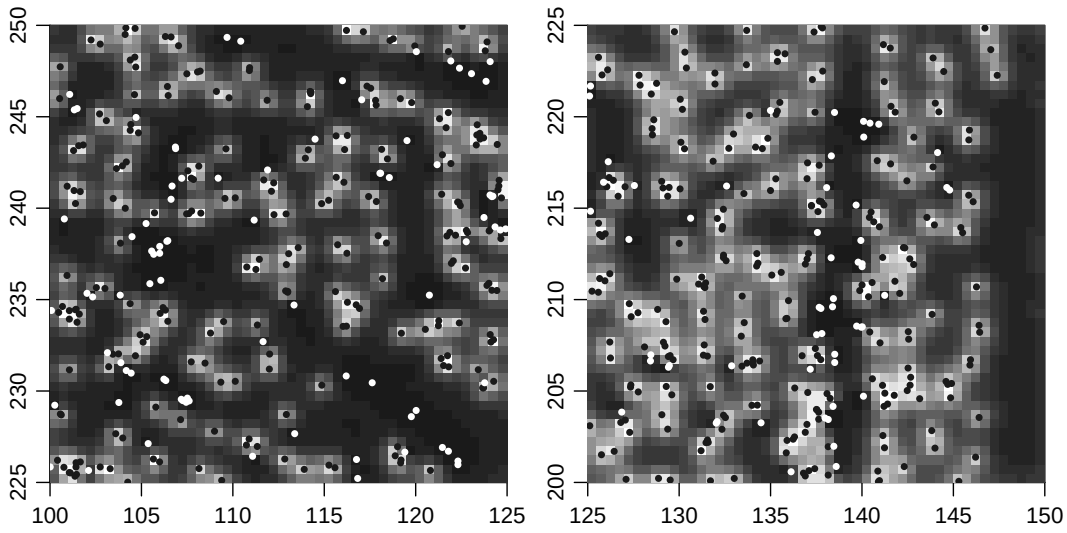

Figure 9.10: The posterior probabilities $\hat{\mathbf{P}}(s_j \in \Theta)$, $s_j \in \mathcal{G}$, for the first (*on the left*) and second (*on the right*) sapling data set. Light grey means high probability, dark grey small probability. The points are the observations.

Allowing observation errors, the germ-grain model may become a competitive alternative for the random set model, whereas it otherwise could be too rigid for real data sets.

Magnussen et al. (2006) considers the partition of a forest area to high and low intensity compartments using Voronoi tessellations. This is a further possibility for the leading measure of the random set generated Poisson process. Blackwell and Møller (2003) deals with Bayesian analysis for tessellations.

The Bayesian approach enables the calculation of uncertainties for the parameters in terms of the posterior interval or standard deviation of the marginal posterior distribution. In addition, it allows estimation of the random set, again with uncertainties. This is a great advantage of the Bayesian approach. The price is the computational load, especially when high resolution is used in the discretization of the random field or when the data are extensive, and possibly the sensitivity to the choice of prior distributions. We applied discretization of the inverse scale parameter β, which reduces the need for matrix inversion leading to a faster simulation algorithm. A drawback may be poorer mixing. For some applications, it might be reasonable to reparametrize the covariance model by $\phi = 1/\beta$ and address the values for ϕ symmetrically around the mean (or median or mode) of the prior distribution.

Our work was motivated by the saplings data, where independent and uniform scattering of points within the two phases is highly plausible because the competition between the young trees is very weak. In this setting, we believe, one will hardly find a modelling approach superseding the current one. However, it is a future task to study, how the situation is improved if the phase is observed both at the locations of points and at sampling locations, which are independent of the observed point pattern.

Bibliography

Adler, R. J. (1981). *The Geometry of Random Fields.* John Wiley & Sons Ltd., New York.

Baddeley, A. J. and Turner, R. (2005). Spatstat: An R package for analysing spatial point patterns. *Journal of Statistical Software* **12**(6), 1–42.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modelling and Analysis for Spatial Data.* Chapman & Hall/CRC, Boca Raton.

Bitterlich, W. (1948). Die Winzelzählprobe. *Allgemeine Forst- und Holzwirtschaftliche Zeitung* **59**, 4–5.

Blackwell, P. G. and Møller, J. (2003). Bayesian analysis of deformed tessellation models. *Advances in Applied Probability* **35**(1), 4–26.

Byth, K. (1981). θ-stationary point processes and their second-order analysis. *Journal of Applied Probability* **18**, 864–878.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* Wiley, New York.

Condit, R. (1998). *Tropical Forest Census Plots.* Springer-Verlag, Berlin.

Condit, R., Hubbell, S. P., and Foster, R. B. (1992). Recruitment near conspecific adults and the maintenance of tree and shrub diversity in a neotropical forest. *The American Naturalist* **140**(2), 261–286.

Condit, R., Hubbell, S. P., and Foster, R. B. (1996). Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology* **12**, 231–256.

Cressie, N. A. C. (1993). *Statistics for Spatial Data,* revised edition. John Wiley & Sons, Inc., New York.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods,* 2nd edition. Springer-Verlag, New York.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure,* 2nd edition. Springer-Verlag, New York.

De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis* **34**, 299–314.

Diggle, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35**, 87–101.

Diggle, P. J. (1981). Binary mosaics and the spatial pattern of heather. *Biometrics* **37**, 531–539.

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

Diggle, P. J., Menezes, R., and Su, T.-L. (2009). Geostatistical inference under preferential sampling. *Applied Statistics* (to appear).

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

Gregoire, T. G. and Valentine, H. T. (2008). *Sampling Strategies for Natural Resources and the Environment*. Chapman & Hall/CRC, Boca Raton.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* **101**(476), Theory and Methods, 1502–1512.

Guan, Y. and Sherman, M. (2007). On least squares fitting for stationary point processes. *Journal of the Royal Statistical Society Ser. B* **69**, 31–49.

Ho, L. P. (2006). Complete spatial randomness tests, intensity-dependent marking and neighborhood competition of spatial point processes with applications to ecology. Ph.D. Thesis, Hong Kong Baptist University.

Ho, L. P. and Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics and Probability Letters* **78**, 1194–1199.

Holgate, P. (1967). The angle-count method. *Biometrika* **53**, 615–623.

Hubbell, S. P. and Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In S. L. Sutton, T. C. Whitmore, and A. C. Chadwick (eds), *Tropical Rain Forest: Ecology and Management*. Blackwell Scientific Publications, Oxford, pp. 24–41.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons Ltd, Chichester.

Kilkki, P. and Päivinen, R. (1986). Weibull distribution in the estimation of the basal area dbh-distribution. *Silva Fennica* **20**, 149–156.

Lantuéjoul, C. (2002). *Geostatistical Simulation. Models and Algorithms*. Springer-Verlag, Berlin.

Magnussen, S., Allard, D., and Wulder, M. (2006). Poisson Voronoï tiling for finding clusters in spatial point patterns. *Scandinavian Journal of Forest Research* **21**, 239–248.

Mandallaz, D. (2008). *Sampling Techniques for Forest Inventories*. Chapman & Hall/CRS, Boca Raton.

Mase, S. (1996). The threshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics* **48**, 201–213.

Matérn, B. (1960). Spatial variation. *Meddelanden från Statens Skogsforskningsinstitut* **49**(5). (See also Matérn, 1986).

Matérn, B. (1986). *Spatial Variation,* Lecture Notes in Statistics **36**. Springer-Verlag, Berlin.

Matérn, B. (1969). Wie groß ist die "Relaskop-Fläche"? *Sonderdruck aus Allgemeine Forstzeitung* **2**.

Matérn, B. (1972). The precision of basal area estimates. *Forest Science* **18**, 123–125.

Menezes, R. (2005). Assessing spatial dependency under non-standard sampling. Ph.D. thesis (ISBN 84-9750-595-6). Universidad de Santiago de Compostela, Santiago de Compostela, Spain.

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

Møller, J. and Waagepetersen, R. P. (2002). Statistical inference for Cox processes. In A. B. Lawson and D. G. T. Denison (eds), *Spatial Cluster Modelling.* Chapman & Hall, Boca Raton, pp. 37–60.

Møller, J. and Waagepetersen, R. P. (2003). An introduction to simulation-based inference for spatial point processes. In J. Møller (eds), *Spatial Statistics and Computational Methods.* Springer-Verlag, New York, pp. 143–198.

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

Myllymäki, M. (2006). On intensity-dependent marking of log Gaussian Cox processes. M.Sc. thesis in Mathematics. University of Jyväskylä, Finland.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams,* 2nd edition. John Wiley & Sons Ltd, Chichester.

Penttinen, A. (1988). A random field approach to Bitterlich sampling. *Annales Academiæ Scientiarum Fennicæ, Series A. I. Mathematica* **13**, 256–268.

Penttinen, A. and Niemi, A. (2007). On Statistical inference for the random set generated Cox process with set-marking. *Biometrical Journal* **49**(2), 197–213.

Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin.

Ripley, B. D. (1981). *Spatial Statistics.* John Wiley & Sons, Inc., New York.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications.* Chapman & Hall/CRC, Boca Raton.

Schlather, M. (2001a). On the second-order characteristics of marked point processes. *Bernoulli* **7**(1), 99–117.

Schlather, M. (2001b). Simulation and analysis of random fields. *R-News* **1**(2), 18–20.

Schlather, M. S., Ribeiro, P. J., and Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society B* **66**, 79–93.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practise, and Visualization.* John Wiley & Sons, Inc., New York.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall Ltd, London.

Stein, M. (1999). *Interpolation of Spatial Data. Some Theory for Kriging.* Springer-Verlag, New York.

Stoyan, D. (1979). Interrupted point processes. *Biometrical Journal* **21**, 607–610.

Stoyan, D. (1984). On correlations of marked point processes. *Mathematische Nachrichten* **116**, 197–207.

Stoyan, D., Kendall, W. S., and Mecke, J. (1995). *Stochastic Geometry and its Applications,* 2nd edition. John Wiley & Sons Ltd, Chichester.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields.* John Wiley & Sons Ltd, Chichester.

Stoyan, D. and Wälder, O. (2000). On variograms in point process statistics, II: models of markings and ecological interpretation. *Biometrical Journal* **42**, 171–187.

Tanaka, U., Ogata, Y., and Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal* **50**, 43–57.

Tjetjep, A. and Seneta, E. (2006). Skewed normal variance-mean models for asset pricing and the method of moments. *International Statistical Review* **74**, 109–126.

Tomppo, E. (2006). The Finnish national forest inventory. In A. Kangas and M. Maltamo (eds), *Forest Inventory: Methodology and Applications.* Springer, Dordrecht, The Netherlands, pp. 179–194.

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika* **95**(2), 351–363.

Waagepetersen, R. and Guan, Y. (2007). Two-step estimation for inhomogeneous spatial point processes. R-2007-25, Department of Mathematical Sciences, Aalborg University.

Wackernagel, H. (1998). *Multivariate Geostatistics,* 2nd edition. Springer-Verlag, Berlin.

Wälder, O. and Stoyan, D. (1996). On variograms in point process statistics. *Biometrical Journal* **38**, 895–905.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85(2)**, 251–267.

Appendix A

Derivation of mark characteristics

Log-intensity marked Cox process

Proof of (4.2): Let Λ and Λ_p be the intensity measures of N_m and N, respectively, see Stoyan et al. (1995) for details of these measures. Since the measure $B \to \Lambda(B \times L)$ for fixed L in the mark space (here $\mathbb{R}$) is absolutely continuous with respect to Λ_p, referring to Stoyan et al. (1995), it can be shown that

$$\Lambda(\mathrm{d}(x,m)) = \lambda(x,m)\nu_d(\mathrm{d}x)\nu_1(\mathrm{d}m), \tag{A.1}$$

where $\lambda(x,m)$ is the intensity of the process N_m, and ν_d and ν_1 are Lebesgue measures in $\mathbb{R}^d$ and $\mathbb{R}$, respectively. Let $f_{zv}(z,v)$ stand for the 2-dimensional Gaussian distribution of $(Z(x), V(x))$, where $V(s) = \alpha + \beta Z(s) + U(s)$, $s \in W \subseteq \mathbb{R}^d$, as in Section 4.1. Further, let λ_p be the intensity of N. For the stationary log Gaussian Cox process N,

$$\lambda_p = \int f_z(z)e^z \,\mathrm{d}z = \int\int f_{zv}(z,v)e^z \,\mathrm{d}z\,\mathrm{d}v,$$

where $f_z(z) = \int f_{zv}(z,v)\,\mathrm{d}v$ is the marginal distribution of $Z(x)$. On the other hand,

$$\lambda_p = \int_{\mathbb{R}} \lambda(x,v)\,\mathrm{d}v.$$

Therefore, it follows that

$$\lambda(x,m) = \int f_{zv}(z,m)e^z \,\mathrm{d}z. \tag{A.2}$$

Assuming the density function f_m of the mark distribution M exists, $M(\mathrm{d}m) = f_m(m)\nu_1(\mathrm{d}m)$. Therefore, by combining (A.1) with (2.1), we obtain that $\lambda(x,m) = f_m(m)\lambda_p$. Since the intensity of a log Gaussian Cox process is $\lambda_p = e^{\mu_Z + \frac{1}{2}\sigma_Z^2}$ and (A.2) holds, we get the mark distribution

$$f_m(m) = \frac{\int f_{zv}(z,m)e^z \mathrm{d}z}{e^{\mu_Z + \frac{1}{2}\sigma_Z^2}}.$$

The result (4.2) follows directly from the calculation for the n-dimensional mark distribution by choosing $n = 1$, see the following proof below.

The proof of (4.3): For n marks $m = (m_1, \ldots, m_n)'$ at n points $x = (x_1, \ldots, x_n)'$ we get the same way as for a mark (see the previous proof), by replacing $d(x \times m)$ by $d(x_1 \times \ldots x_n \times m_1 \times \cdots \times m_n)$, that the mark distribution can be expressed in the form

$$f_m(m) = \frac{\int f_{zv}(z, m) e^{z'\mathbf{1}} \mathrm{d}z}{e^{\mu_Z' \mathbf{1} + \frac{1}{2}\mathbf{1}'\Sigma_Z \mathbf{1}}},$$

where $z = (Z(x_1), \ldots, Z(x_n))'$, $m = (V(x_1), \ldots, V(x_n))'$, $\mathbf{1} = (1, \ldots, 1)'$ (column matrix of length n), Σ_Z is the covariance matrix of z and we have used the notation μ_Z for $(\mu_Z(x_1), \ldots, \mu_Z(x_n))' = (\mu_Z, \ldots, \mu_Z)'$. The notations $\Sigma_Z = \Sigma_{Z(x)}$ and $\Sigma_V = \Sigma_{V(x)}$ are used for brevity, and in what follows, μ_V stands for $(\mu_V(x_1), \ldots, \mu_V(x_n))' = (\mu_V, \ldots, \mu_V)'$ in this proof. Let Σ_{VZ} be the cross-covariance matrix of v and z on condition z is known at all points $x_1, \ldots, x_n$. (The cross-covariance function is defined by $C_{VZ}(s_1, s_2) = \mathbb{E}[V(s_1) - \mu_V][Z(s_2) - \mu_Z] = \beta C_Z(s_1, s_2)$.) On condition that z is known at these locations, $\Sigma_{VZ} = \beta \mathrm{diag}(\Sigma_Z)$, where $\mathrm{diag}(\Sigma_Z)$ denotes that only the diagonal elements of Σ_Z are taken and other elements of the matrix are zeros. This is due to that, given $Z(x)$, V(x) is independent of other values of $\{Z(s)\}$. Since $f_{vz}(v, z) = f_v(v) f_{z|v}(z|\, v)$ and

$$z|\, v \sim N(\mu_Z + \Sigma_{VZ}' \Sigma_V^{-1}(v - \mu_V), \Sigma_Z - \Sigma_{VZ}' \Sigma_V^{-1} \Sigma_{VZ}),$$

it holds

$$\int f_{z|v}(z|\, v) e^{z'\mathbf{1}} \mathrm{d}z = M_{z|v}(\mathbf{1}) = e^{(\mu_Z + \Sigma_{VZ}' \Sigma_V^{-1}(v - \mu_V))'\mathbf{1} + \frac{1}{2}\mathbf{1}'(\Sigma_Z - \Sigma_{VZ}' \Sigma_V^{-1} \Sigma_{VZ})\mathbf{1}},$$

which is the moment-generating function of $z|\, v$ at point $\mathbf{1}$. Consequently, we get that

$$f_x(m) = \frac{f_v(m) \int f_{z|v}(z|\, m) e^{z'\mathbf{1}} \mathrm{d}z}{e^{\mu_Z' \mathbf{1} + \frac{1}{2}\mathbf{1}'\Sigma_Z \mathbf{1}}} = \frac{1}{2\pi\sqrt{\det \Sigma_V}} e^{-\frac{1}{2}\left[(m - \mu_V - \Sigma_{VZ}\mathbf{1})' \Sigma_V^{-1} (m - \mu_V - \Sigma_{VZ}\mathbf{1})\right]},$$

which is the density of the distribution $N(\mu_V + \Sigma_{VZ}\mathbf{1}, \Sigma_V)$.

The proof of (4.4) *and* (4.5): Following the structure of the proof that Ho and Stoyan (2008) have used to calculate $k_{m\cdot}(r)$, which is $E(r)$ divided by the mean of marks, for the intensity-marked Cox process, we can calculate $E(r)$ by

$$E(r) = \mathbb{E}_{or}[m(o)] = \frac{\mathbb{E}[m(o)\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]}, \tag{A.3}$$

being a special case of (5.4). The numerator of (A.3) is

$$\begin{aligned}\mathbb{E}[m(o)\Lambda(o)\Lambda(\mathbf{r})] &= \alpha \mathbb{E}[e^{Z(o) + Z(\mathbf{r})}] + \beta \mathbb{E}[Z(o) e^{Z(o) + Z(\mathbf{r})}] \\ &= \alpha e^{2\mu_Z + \sigma_Z^2 + C_Z(r)} + \beta \mathbb{E}_{Z(o)}\{\mathbb{E}[Z(o) e^{Z(o) + Z(\mathbf{r})} | Z(o)]\},\end{aligned}$$

where

$$\begin{aligned}&\mathbb{E}_{Z(o)}\{\mathbb{E}[Z(o) e^{Z(o) + Z(\mathbf{r})} | Z(o)]\} = \mathbb{E}_{Z(o)}\{Z(o) e^{Z(o)} \mathbb{E}[e^{Z(\mathbf{r})} | Z(o)]\} \\ &= e^{\mu_Z + \frac{1}{2}(\sigma_Z^2 - (C_Z(r))^2/\sigma_Z^2)} \mathbb{E}_{Z(o)}\{Z(o) e^{Z(o) + C_Z(r)/\sigma_Z^2 (Z(o) - \mu_Z)}\} \\ &= e^{\mu_Z + \frac{1}{2}(\sigma_Z^2 - (C_Z(r))^2/\sigma_Z^2)} e^{\frac{1}{2}(\sigma_Z^2 + (C_Z(r))^2/\sigma_Z^2) + C_Z(r) + \mu_Z} \int_{-\infty}^{\infty} Z(o) \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{1}{2\sigma_Z^2}(Z(o) - \mu_Z - \sigma_Z^2 - C_Z(r))^2} \mathrm{d}Z(o) \\ &= e^{2\mu_z + \sigma_Z^2 + C_Z(r)}(\mu_z + \sigma_Z^2 + C_Z(r)).\end{aligned}$$

Here $\mathbb{E}[e^{Z(\mathbf{r})}|Z(o)] = M_{Z(\mathbf{r})|Z(o)}(1)$ (the moment-generating function), the assumption of Z and U being independent and supplementing to a square have been used. Since $\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})] = \exp\{2\mu_z + \sigma_Z^2 + C_Z(r)\}$, we get $E(r) = \alpha + \beta(\mu_z + \sigma_Z^2 + C_Z(r))$. To calculate $V(r)$, we first calculate

$$\mathbb{E}_{or}[m(o)^2] = \frac{\mathbb{E}[m(o)^2\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} = \frac{\mathbb{E}[(\alpha + \beta Z(o) + U(o))^2\Lambda(o)\Lambda(\mathbf{r})]}{e^{2\mu_Z+\sigma_Z^2+C_Z(r)}}$$
$$= \frac{\alpha^2 e^{2\mu_Z+\sigma_Z^2+C_Z(r)} + \beta^2\mathbb{E}[Z(o)^2e^{Z(o)+Z(\mathbf{r})}] + \sigma_U^2 e^{2\mu_Z+\sigma_Z^2+C_Z(r)} + 2\alpha\beta e^{2\mu_Z+\sigma_Z^2+C_Z(r)}(\mu_Z + \sigma_Z^2 + C_Z(r))}{e^{2\mu_Z+\sigma_Z^2+C_Z(r)}}$$
$$= \alpha^2 + \frac{\beta^2\mathbb{E}[Z(o)^2e^{Z(o)+Z(\mathbf{r})}]}{e^{2\mu_Z+\sigma_Z^2+C_Z(r)}} + \sigma_U^2 + 2\alpha\beta(\mu_Z + \sigma_Z^2 + C_Z(r)).$$

A similar type of calculation as for $\mathbb{E}[Z(o)e^{Z(o)+Z(\mathbf{r})}]$ yields

$$\mathbb{E}[Z(o)^2e^{Z(o)+Z(\mathbf{r})}] = e^{2\mu_Z+\sigma_Z^2+C_Z(r)}(\sigma_Z^2 + (\mu_Z + \sigma_Z^2 + C_Z(r))^2).$$

Thus,

$$\mathbb{E}_{or}[m(o)^2] = \alpha^2 + \beta^2\left(\sigma_Z^2 + (\mu_Z + \sigma_Z^2 + C_Z(r))^2\right) + \sigma_U^2 + 2\alpha\beta(\mu_Z + \sigma_Z^2 + C_Z(r)).$$

Then, $V(r)$ is obtained by simple calculation:

$$V(r) = \mathbb{E}_{or}[m(o)^2] - [E(r)]^2 = \beta^2\sigma_Z^2 + \sigma_U^2.$$

Gaussian intensity-marked Cox process with marking (5.3)

Derivation of the mean mark and mark variance: Obviously

$$\mathbb{E}_o[m(o)] = \mathbb{E}_o[\mathbb{E}_o[m(o)|\,\Lambda(o)]] = a + b\mathbb{E}_o[\Lambda(o)]$$

and

$$\begin{aligned}\mathrm{var}_o[m(o)] &= \mathbb{E}_o\{\mathrm{var}_o[m(o)|\,\Lambda(o)]\} + \mathrm{var}_o\{\mathbb{E}_o[m(o)|\,\Lambda(o)]\}\\ &= c^2\mathbb{E}_o[\Lambda(o)] + d^2 + b^2\mathrm{var}_o[\Lambda(o)].\end{aligned}$$

The calculation of moments of the marks for the intensity-marked Cox process in Ho and Stoyan (2008) yields that

$$\mathbb{E}_o[\Lambda(o)] = \lambda e^{\sigma_Z^2}$$

and

$$\begin{aligned}\mathrm{var}_o[\Lambda(o)] &= \mathbb{E}_o[(\Lambda(o))^2] - [\mathbb{E}_o(\Lambda(o))]^2\\ &= \lambda^2e^{3\sigma_Z^2} - \lambda^2e^{2\sigma_Z^2}\\ &= \lambda^2e^{2\sigma_Z^2}\left(e^{\sigma_Z^2} - 1\right)\\ &= e^{2\sigma_Z^2}C_\Lambda(0).\end{aligned}$$

where λ is the intensity of the log Gaussian Cox process and $C_\Lambda(0) = \lambda^2(e^{\sigma_Z^2} - 1)$ is the variance of the intensity $\{\Lambda(s)\}$. The mean mark and the mark variance follow by simple calculation.

Derivation of the second-order characteristics: Characteristics $E(r)$ and $k_{mm}(r)$ can be calculated using (A.3) and

$$\mathbb{E}_{or}\{m(o)m(\mathbf{r})\} = \frac{\mathbb{E}[m(o)m(\mathbf{r})\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]}. \tag{A.4}$$

Since, because of conditional independence,

$$\begin{aligned}\mathbb{E}[m(o)m(\mathbf{r})\Lambda(o)\Lambda(\mathbf{r})] &= \mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\mathbb{E}[m(o)m(\mathbf{r})|\,\Lambda(o),\Lambda(\mathbf{r})]]\\ &= \mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})[a+b\Lambda(o)][a+b\Lambda(\mathbf{r})]]\end{aligned}$$

and

$$\mathbb{E}[m(o)\Lambda(o)\Lambda(\mathbf{r})] = \mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\mathbb{E}[m(o)|\,\Lambda(o),\Lambda(\mathbf{r})]] = \mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})[a+b\Lambda(o)]],$$

we get

$$\begin{aligned}\mathbb{E}[m(o)m(\mathbf{r})\Lambda(o)\Lambda(\mathbf{r})] &= a^2\mathbb{E}e^{Z(o)+Z(\mathbf{r})} + ab\mathbb{E}e^{Z(o)+2Z(\mathbf{r})} + ab\mathbb{E}e^{2Z(o)+Z(\mathbf{r})} + b^2\mathbb{E}e^{2Z(o)+2Z(\mathbf{r})}\\ &= a^2e^{2\mu_Z+\sigma_Z^2+C_Z(r)} + 2abe^{3\mu_Z+5/2\sigma_Z^2+2C_Z(r)} + b^2e^{4\mu_Z+4\sigma_Z^2+4C_Z(r)}\\ &= e^{2\mu_Z+\sigma_Z^2+C_Z(r)}\left(a + 2ab\lambda e^{\sigma_Z^2+C_Z(r)} + b^2\lambda^2e^{2\sigma_Z^2+3C_Z(r)}\right)\end{aligned}$$

and

$$\mathbb{E}[m(o)\Lambda(o)\Lambda(\mathbf{r})] = ae^{2\mu_Z+\sigma_Z^2+C_Z(r)} + be^{3\mu_Z+5/2\sigma_Z^2+2C_Z(r)}.$$

Hence

$$E(r) = a + b\lambda e^{\sigma_Z^2+C_Z(r)}$$

and

$$\mathbb{E}_{or}[m(o)m(\mathbf{r})] = a + 2ab\lambda e^{\sigma_Z^2+C_Z(r)} + b^2\lambda^2e^{2\sigma_Z^2+3C_Z(r)},$$

from which $k_{mm}(r)$ is obtained by dividing by μ_m^2. The value at zero is obtained by $k_{mm}(0) = \mathbb{E}_o[(m(o))^2]/\mu_m^2 = (\sigma_m^2+\mu_m^2)/\mu_m^2$. Since

$$\begin{aligned}&\mathbb{E}_{or}\{m(o)^2\} = \frac{\mathbb{E}[m(o)^2\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} = \frac{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\mathbb{E}[m(o)^2|\Lambda(o),\Lambda(\mathbf{r})]]}{e^{2\mu_Z+\sigma_Z^2+C_Z(r)}}\\ &= e^{-2\mu_Z-\sigma_Z^2-C_Z(r)}\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\{\mathrm{var}[m(o)|\,\Lambda(o),\Lambda(\mathbf{r})] + [\mathbb{E}[m(o)|\,\Lambda(o),\Lambda(\mathbf{r})]]^2\}]\\ &= e^{-2\mu_Z-\sigma_Z^2-C_Z(r)}\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\{c^2\Lambda(o)+d^2+(a+b\Lambda(o))^2\}]\\ &= e^{-2\mu_Z-\sigma_Z^2-C_Z(r)}\{(a^2+d^2)\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})] + (2ab+c^2)\mathbb{E}[\Lambda(o)^2\Lambda(\mathbf{r})] + b^2\mathbb{E}[\Lambda(o)^3\Lambda(\mathbf{r})]\}\\ &= a^2+d^2+(2ab+c^2)e^{\mu_Z+\frac{3}{2}\sigma_Z^2+C_Z(r)} + b^2e^{2\mu_Z+4\sigma_Z^2+2C_Z(r)},\end{aligned}$$

we get

$$\begin{aligned}V(r) &= \mathbb{E}_{or}\{m(o)^2\} - [E(r)]^2\\ &= d^2 + c^2\lambda e^{\sigma_Z^2+C_Z(r)} + b^2\lambda^2e^{2\sigma_Z^2+2C_Z(r)}\left(e^{\sigma_Z^2}-1\right).\end{aligned}$$

By similar calculations, we obtain the mark variogram

$$\gamma_m(r) = d^2 + c^2\lambda e^{2\sigma_Z^2+C_Z(r)} + b^2\lambda^2e^{2\sigma_Z^2+2C_Z(r)}\left(e^{\sigma_Z^2}-e^{C_Z(r)}\right) \quad \text{for } r>0.$$

Gaussian intensity-marked Cox process with marking (5.5)

Derivation of the mean mark and mark variance: Mean and variance of the mark (5.5) can be obtained by following the proof in Ho and Stoyan (2008) for the marking (5.1). First consider the marks $m(x_i) = 1/\Lambda(x_i)$ and denote the lognormal distribution of $\Lambda(o)$ by F. Then the corresponding mark distribution function is obtained in the following way using the Campbell and Fubini theorems and stationarity of the process:

$$\begin{aligned}
\mathbb{E}\{N_m(A \times (-\infty, m])\} &= \mathbb{E}\left\{\mathbb{E}\left[\sum_{x\in N} \mathbf{1}_A(x)\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(x)}\right) \mid \Lambda\right]\right\} \\
&\overset{\text{Campbell}}{=} \mathbb{E}\left[\int \mathbf{1}_A(x)\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(x)}\right)\Lambda(x)\mathrm{d}x\right] \\
&\overset{\text{Fubini}}{=} \int \mathbf{1}_A(x)\mathbb{E}\left[\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(x)}\right)\Lambda(x)\right]\mathrm{d}x \\
&\overset{\text{stationarity}}{=} \int \mathbf{1}_A(x)\mathbb{E}\left[\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(o)}\right)\Lambda(o)\right]\mathrm{d}x \\
&= \mathbb{E}\left[\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(o)}\right)\Lambda(o)\right]\int \mathbf{1}_A(x)\mathrm{d}x \\
&= \mathbb{E}\left[\mathbf{1}_{[1/m,\infty)}\left(\Lambda(o)\right)\Lambda(o)\right]\int \mathbf{1}_A(x)\mathrm{d}x \\
&= \nu(A)\int_{1/m}^{\infty} x\mathrm{d}F(x),
\end{aligned}$$

where the indicator function

$$\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(o)}\right) = \begin{cases} 1, & \text{if } \dfrac{1}{\Lambda(o)} \le m\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

has been rewritten as

$$\mathbf{1}_{(-\infty,m]}\left(\frac{1}{\Lambda(o)}\right) = \mathbf{1}_{[1/m,\infty)}\left(\Lambda(o)\right).$$

Because

$$\mathbb{E}[N_m(A \times L)] = \lambda\nu(A)M(L),$$

the distribution of the mark $m(x_i) = 1/\Lambda(x_i)$ is

$$F_M(m) = \frac{1}{\lambda}\int_{1/m}^{\infty} x\mathrm{d}F(x),$$

and the density is

$$f_m(m) = \frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}m}\int_{1/m}^{\infty} x\mathrm{d}F(x) = \frac{1}{\lambda}\frac{1}{m^3}f\left(\frac{1}{m}\right),$$

where f is the density of F. Thus the mean and the second moment for the marking $m(x_i) = 1/\Lambda(x_i)$ are

$$\mathbb{E}_o[m(o)] = \int_{-\infty}^{\infty} m f_m(m)\mathrm{d}m = \frac{1}{\lambda}\int_0^{\infty}\frac{1}{m^2}f\left(\frac{1}{m}\right)\mathrm{d}m = \frac{1}{\lambda}\int_0^{\infty} f(u)\mathrm{d}u = \frac{1}{\lambda}$$

and

$$\mathbb{E}_o[(m(o))^2] = \int_{-\infty}^{\infty} m^2 f_m(m)\mathrm{d}m = \frac{1}{\lambda}\int_0^{\infty} \frac{1}{m} f\left(\frac{1}{m}\right)\mathrm{d}m = \frac{1}{\lambda}\int_0^{\infty} \frac{1}{u} f(u)\mathrm{d}u$$
$$= \frac{1}{\lambda}\mathbb{E}_o\left[\frac{1}{\Lambda(o)}\right] = \frac{\exp\{-\mu_Z + \frac{1}{2}\sigma_Z^2\}}{\exp\{\mu_Z + \frac{1}{2}\sigma_Z^2\}} = \exp\{-2\mu_Z\},$$

where the change of variables $u = 1/m$ and $\Lambda(o) \geq 0$ have been used. Consequently,

$$\mathrm{var}_o[m(o)] = e^{-2\mu_Z} - \frac{1}{\lambda^2} = e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right).$$

Of course, it must hold $\mathrm{var}_o[m(o)] \geq 0$, which is now equivalent to $\sigma_Z^2 \geq 0$. Using the above results for $1/\Lambda(x_i)$, we get for the mark (5.5) that

$$\mu_m = \mathbb{E}_o[\mathbb{E}_o[m(o)|\,\Lambda(o)]] = \mathbb{E}_o\left[a + b\frac{1}{\Lambda(o)}\right] = a + \frac{b}{\lambda}$$

and

$$\sigma_m^2 = \mathrm{var}_o[\mathbb{E}_o[m(o)|\,\Lambda(o)]] + \mathbb{E}_o[\mathrm{var}_o[m(o)|\,\Lambda(o)]]$$
$$= \mathrm{var}_o\left[a + b\frac{1}{\Lambda(o)}\right] + \mathbb{E}_o\left[c^2\frac{1}{\Lambda(o)} + d^2\right]$$
$$= b^2 e^{-2\mu_Z}\left(1 - e^{-\sigma_Z^2}\right) + \frac{c^2}{\lambda} + d^2.$$

Derivation of the second-order characteristics: The characteristics are calculated same way as for (5.3) model. The $E(r)$ is obtained by

$$E(r) = \mathbb{E}_{or}[m(o)] = \frac{\mathbb{E}[m(o)\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} = \frac{a\lambda^2 e^{C_Z(r)} + b\lambda}{\lambda^2 e^{C_Z(r)}} = a + \frac{b}{\lambda}e^{-C_Z(r)}.$$

Since

$$\mathbb{E}_{or}[m(o)^2] = \frac{\mathbb{E}[m(o)^2\Lambda(o)\Lambda(\mathbf{r})]}{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]}$$
$$= \frac{1}{\lambda^2 e^{C_Z(r)}}\left[c^2\mathbb{E}[\Lambda(\mathbf{r})] + (a^2 + d^2)\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})] + 2ab\mathbb{E}[\Lambda(\mathbf{r})] + b^2\mathbb{E}\left[\frac{\Lambda(\mathbf{r})}{\Lambda(o)}\right]\right]$$
$$= \frac{c^2\lambda + (a^2 + d^2)\lambda^2 e^{C_Z(r)} + 2ab\lambda + b^2 e^{\sigma_Z^2 - C_Z(r)}}{\lambda^2 e^{C_Z(r)}}$$
$$= a^2 + d^2 + \frac{2ab + c^2}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{\sigma_Z^2 - 2C_Z(r)},$$

we get

$$V(r) = \mathbb{E}_{or}[m(o)^2] - [E(r)]^2 = d^2 + \frac{c^2}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}\left(e^{\sigma_Z^2} - 1\right)e^{-2C_Z(r)}.$$

The unscaled Stoyan's $\kappa_{mm}(r)$-function is

$$\kappa_{mm}(r) = \mathbb{E}_{or}[m(o)m(\mathbf{r})] = \frac{a^2\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})] + ab\mathbb{E}[\Lambda(o)] + ab\mathbb{E}[\Lambda(\mathbf{r})] + b^2}{\lambda^2 e^{C_Z(r)}}$$
$$= \frac{a^2\lambda^2 e^{C_Z(r)} + 2ab\lambda + b^2}{\lambda^2 e^{C_Z(r)}} = a^2 + \frac{2ab}{\lambda}e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{-C_Z(r)},$$

from which $k_{mm}(r)$-function is obtained by dividing by μ_m^2. The mark variogram can be calculated in the following way:

$$\begin{aligned}
\gamma_m(r) &= \frac{1}{2}\mathbb{E}_{or}\{[m(o) - m(\mathbf{r})]^2\} = \frac{\mathbb{E}\{[(m(o))^2 + (m(\mathbf{r}))^2 - 2m(o)m(\mathbf{r})]\Lambda(o)\Lambda(\mathbf{r})\}}{2\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} \\
&= \frac{\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})\mathbb{E}[(m(o))^2 + (m(\mathbf{r}))^2 - 2m(o)m(\mathbf{r})|\,\Lambda(o), \Lambda(\mathbf{r})]]}{2\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})]} \\
&= \frac{1}{2\lambda^2 e^{C_Z(r)}}\left[c^2\mathbb{E}[\Lambda(\mathbf{r})] + c^2\mathbb{E}[\Lambda(o)] + 2d^2\mathbb{E}[\Lambda(o)\Lambda(\mathbf{r})] + b^2\mathbb{E}\left[\frac{\Lambda(\mathbf{r})}{\Lambda(o)}\right] + b^2\mathbb{E}\left[\frac{\Lambda(o)}{\Lambda(\mathbf{r})}\right] - 2b^2\right] \\
&= \frac{1}{\lambda^2 e^{C_Z(r)}}\left[c^2\lambda + d^2\lambda^2 e^{C_Z(r)} + b^2 e^{\sigma_Z^2 - C_Z(r)} - b^2\right] = d^2 + \left(\frac{c^2}{\lambda} - \frac{b^2}{\lambda^2}\right)e^{-C_Z(r)} + \frac{b^2}{\lambda^2}e^{\sigma_Z^2 - 2C_Z(r)}.
\end{aligned}$$

The mark characteristics of the gamma intensity-marked Cox process are obtained by similar calculations.

83. PAPERS ON ANALYSIS: A VOLUME DEDICATED TO OLLI MARTIO ON THE OCCASION OF HIS 60TH BIRTHDAY. Edited by J. Heinonen, T. Kilpeläinen, and P. Koskela. (315 pp.) 2001
84. ONNINEN, JANI, Mappings of finite distortion: Continuity. (24 pp.) 2002
85. OLLILA, ESA, Sign and rank covariance matrices with applications to multivariate analysis. (42 pp.) 2002
86. KAUKO, VIRPI, Visible and nonexistent trees of Mandelbrot sets. (26 pp.) 2003
87. LLORENTE, JOSÉ G., Discrete martingales and applications to analysis. (40 pp.) 2002
88. MITSIS, THEMIS, Topics in harmonic analysis. (52 pp.) 2003
89. KÄRKKÄINEN, SALME, Orientation analysis of stochastic fibre systems with an application to paper research. (53 pp.) 2003
90. HEINONEN, JUHA, Geometric embeddings of metric spaces. (44 pp.) 2003
91. RAJALA, KAI, Mappings of finite distortion: Removable singularities. (23 pp.) 2003
92. FUTURE TRENDS IN GEOMETRIC FUNCTION THEORY. RNC WORKSHOP JYVÄSKYLÄ 2003. Edited by D. Herron. (262 pp.) 2003
93. KÄENMÄKI, ANTTI, Iterated function systems: Natural measure and local structure. (14 pp.) 2003
94. TASKINEN, SARA, On nonparametric tests of independence and robust canonical correlation analysis. (44 pp.) 2003
95. KOKKI, ESA, Spatial small area analyses of disease risk around sources of environmental pollution: Modelling tools for a system using high resolution register data. (72 pp.) 2004
96. HITCZENKO, PAWEŁ, Probabillistic analysis of sorting algorithms. (71 pp.) 2004
97. NIEMINEN, TOMI, Growth of the quasihyperbolic metric and size of the boundary. (16 pp.) 2005
98. HAHLOMAA, IMMO, Menger curvature and Lipschitz parametrizations in metric spaces. (8 pp.) 2005
99. MOLTCHANOVA, ELENA, Application of Bayesian spatial methods in health and population studies using registry data. (55 pp.) 2005
100. HEINONEN, JUHA, Lectures on Lipschitz analysis. (77 pp.) 2005
101. HUJO, MIKA, On the approximation of stochastic integrals. (19 pp.) 2005
102. LINDQVIST, PETER, Notes on the p-Laplace equation. (80 pp.) 2006
103. HUKKANEN, TONI, Renormalized solutions on quasi open sets with nonhomogeneous boundary values. (41 pp.) 2006
104. HÄHKIÖNIEMI, MARKUS, The role of representations in learning the derivative. (101 pp.) 2006
105. HEIKKINEN, TONI, Self-improving properties of generalized Orlicz–Poincaré inequalities. (15 pp.) 2006
106. TOLONEN, TAPANI, On different ways of constructing relevant invariant measures. (13 pp.) 2007
107. HORPPU, ISMO, Analysis and evaluation of cell imputation. (248 pp.) 2008
108. SIRKIÄ, SEIJA, Spatial sign and rank based scatter matrices with applications. (29 pp.) 2007
109. LEIKAS, MIKA, Projected measures on manifolds and incomplete projection families. (16 pp.) 2007
110. TAKKINEN, JUHANI, Mappings of finite distortion: Formation of cusps. (10 pp.) 2007
111. TOLVANEN, ASKO, Latent growth mixture modeling: A simulation study. (201 pp.) 2007
112. VARPANEN, HARRI, Gradient estimates and a failure of the mean value principle for p-harmonic functions. (66 pp.) 2008
113. MÄKÄLÄINEN, TERO, Nonlinear potential theory on metric spaces. (16 pp.) 2008
114. LUIRO, HANNES, Regularity properties of maximal operators. (11 pp.) 2008
115. VIHOLAINEN, ANTTI, Prospective mathematics teachers' informal and formal reasoning about the concepts of derivative and differentiability. (86 pp.) 2008
116. LEHRBÄCK, JUHA, Weighted Hardy inequalities and the boundary size. (21 pp.) 2008
117. NISSINEN, KARI, Small area estimation with linear mixed models from unit-level panel and rotating panel data. (230 pp.) 2009
118. BOJARSKI, B.V., Generalized solutions of a system of differential equations of the first order and elliptic type with discontinuous coefficients. (64 pp.) 2009
119. RAJALA, TAPIO, Porosity and dimension of sets and measures. (22 pp.) 2009

ISBN 978-951-39-3683-9
ISSN 1457-8905