

Henning Omre
Torstein M. Fjeldstad
Ole Bernhard Forberg

Bayesian Spatial Modelling with Conjugate Prior Models



Bayesian Spatial Modelling with Conjugate Prior Models

Henning Omre • Torstein M. Fjeldstad •
Ole Bernhard Forberg

Bayesian Spatial Modelling with Conjugate Prior Models



Springer

Henning Omre
Department of Mathematical Sciences
Norwegian University of Science
and Technology
Trondheim, Norway

Torstein M. Fjeldstad
Norwegian Computing Center
Oslo, Norway

Ole Bernhard Forberg
Department of Mathematical Sciences
Norwegian University of Science
and Technology
Trondheim, Norway

ISBN 978-3-031-65417-6 ISBN 978-3-031-65418-3 (eBook)
<https://doi.org/10.1007/978-3-031-65418-3>

Mathematics Subject Classification: 60-XX, 60Gxx, 62M40, 60G60, 62-XX, 60G55, 60J10, 62Fxx, 62F15, 62Hxx, 62H11, 62Mxx, 60G15

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

Spatial phenomena appear in numerous applications and disciplines and can be modelled using the language of probability and statistics within a spatial context. This branch of statistics is often referred to as spatial statistics. At the core of spatial statistics is the use of spatial, or regionalised, random variables, which are characterised by having a spatial reference within a specific spatial domain. Spatial statistics involve applying statistical models to understand the patterns, relationships and variations in the spatial phenomena distributed across space. Examples are found in agriculture (Oliver, 2010), earth and ocean sciences (Butler & Vance, 2022), hydrogeological mapping (Kitanidis, 1997), exploration of natural resources (Journel & Huijbregts, 1978), epidemiological modelling (Diggle & Giorgi, 2019), medical image diagnosis (Cootes & Taylor, 2001), among others.

Spatial statistics often refers to the probabilistic and statistical modelling of spatial variables, encompassing both spatial simulation and prediction, associated precision quantifications and spatial model inference. There are numerous books dedicated to the theory and applications of spatial statistics, including notable works by Journel and Huijbregts (1978), Winkler (2006), Illian et al. (2008), Gelfand et al. (2010), Gaetan and Guyon (2010), Cressie and Wikle (2011), Chiles and Delfiner (2012), Banerjee et al. (2014) and Lieshout (2019). This book draws inspiration from all these publications and the early work by Hjort and Omre (1994). However, as authors, we have made several unconventional choices in the presentations of these topics.

This book presents probabilistic and statistical models in a Bayesian spatial framework. This framework facilitates the separate modelling of the observation acquisition procedure in the likelihood model and the spatial variable characteristics in the prior model. The posterior model, which combines prior assumptions and data according to Bayes' rule, is the ultimate solution in Bayesian modelling and is uniquely defined by the likelihood and prior models. Model parameter inference is accomplished through a maximum marginal likelihood criterion or within a hierarchical Bayesian framework.

Assessment of the posterior model for high-dimensional spatial variables where there is strong spatial coupling and when precise observations are available at spe-

cific spatial locations is often challenging. This is because the normalising constant of the posterior model is very computationally demanding to calculate. Simulation-based assessments using brute force Markov chain Monte Carlo (McMC) techniques are generally unsuitable because their convergence rates are often unacceptably low. In this book, we focus on parametrised classes of spatial prior models that exhibit conjugate properties for certain classes of likelihood models. Consequently, the associated posterior spatial models belong to the same class of probability distributions as the prior models. This conjugate property provides very efficient analytical tractability of the posterior spatial model, and the marginal likelihood for the observations can also be calculated. In the context of more complex Bayesian spatial modelling problems, these classes of conjugate models may be viewed as approximations. Their analytical tractability enables efficient proposal models in more complex McMC algorithms.

Spatial variables can be categorised into three types: continuous, event and mosaic. Continuous spatial variables are real-valued and predominantly exhibit smooth variations across the spatial domain. Examples include the terrain elevation in an area and the temperature of ocean water in a specific volume. Event spatial variables are binary, signifying the background with one value, while the other value is assigned to particular locations within the spatial domain. The latter value indicates the spatial location of an event. These could, for example, define the locations of trees in a forest or the positions of cracks in a rock matrix. Mosaic spatial variables are categorical and assign each point in the spatial domain a single category. Such variables can represent, for example, land use in a geographical area or tissue classes in the brain. Although many spatial phenomena are spatio-temporal, as in Cressie and Wikle (2011), our book exclusively discusses spatial variables.

We have chosen to characterise random fields by employing a spatial discretisation on a regular grid that covers a finite spatial domain. This results in finite-dimensional spatial models, and all likelihood and prior models must be defined relative to this spatial discretisation. The mathematical developments are made on these spatially discretised models, providing exact results without using numerous stringent mathematical assumptions. The corresponding results for the underlying random fields can be derived through infill asymptotic analysis of the discretised fields, some of which we discuss briefly. Another advantage of spatial discretisation is the simplicity of combining the three types of spatial variables, a frequent requirement in various applications.

This book primarily comprises definitions and discussions of traditional conjugate spatial models. Additionally, concise presentations of several other frequently utilised spatial models are included. The book is structured in accordance with the principal dimensions of Bayesian models: likelihood, prior, and posterior, and finally, parameter inference. We discuss the spatial variable types for each principal dimension: continuous, event, and mosaic. This organisation is unconventional, as other extant books on these topics typically utilise the spatial variable type as the primary dimension in their presentation. The advantage of our organisation is that it exposes the line of thought in Bayesian spatial modelling, providing examples for

each spatial variable type. Moreover, a consistent notation is rigorously applied to the presentation of the Bayesian spatial model of each type.

Most spatially discretised spatial models are high-dimensional, which entails that calculating the posterior models can be very computationally demanding, even if they are analytically tractable. Therefore, a chapter that discusses computational challenges is included. In addition, the book covers several special topics related to the modelling of spatial variables. Spatial statistical models are used in various applications, and a selection of papers on applied studies is briefly discussed. Lastly, suitable project assignments and exercises are presented, and suggested solutions are available as supplementary material online.

The material in this book encompasses a typical curriculum for a graduate-level university course in spatial statistics. Students will benefit from having a background in university-level courses such as calculus, linear algebra, stochastic processes and multivariate statistics.

Declarations None of the three authors have direct competing interests related to the content of the book. We are active researchers in the field of spatial statistics and the content reflects our view of the field.

Trondheim, Norway
Oslo, Norway
Trondheim, Norway
December 2023

Henning Omre
Torstein M. Fjeldstad
Ole Bernhard Forberg

Reader's Guide

Readers unfamiliar with the Bayesian approach may benefit from an initial reading of specific sections before undertaking a comprehensive study. These sections introduce the concepts of Bayesian spatial modelling, with a continuous Gaussian random field case serving as a familiar example. The recommended preliminary reading is as follows:

- Chapter 1: Introduction
- Chapter 2: Bayesian Spatial Modelling (excluding the subsection on simulation-based assessment)
- Chapter 3: Conjugate Bayesian Models
- Chapter 4: Random Field Models (excluding the subsections on hierarchical models)

Moreover, readers should also read the introductory paragraphs and subsections on Gaussian random fields in the following chapters while omitting the subsections on hierarchical Gaussian random fields:

- Chapter 5: Likelihood Models
- Chapter 6: Prior Models
- Chapter 7: Posterior Models
- Chapter 8: Model Parameter Inference
- Chapter 9: Computational Challenges

A thorough study of the entire book benefits readers preparing for thesis work in spatial statistics. For readers seeking a comprehensive overview of the fundamentals of spatial modelling, we recommend studying most of the text, except the sections on hierarchical models. This suggested reading list aligns with the curriculum for a graduate-level course in a university statistics programme. The course should also incorporate three project assignments, one for each type of spatial model: continuous, event and mosaic. The suggested reading is:

- Chapter 1: Introduction
- Chapter 2: Bayesian Spatial Modelling

- Chapter 3: Conjugate Bayesian Models
- Chapter 4: Random Field Models (excluding the subsections on hierarchical models)
- Chapter 5: Likelihood Models
- Chapter 6: Prior Models (excluding the subsections on hierarchical models)
- Chapter 7: Posterior Models (excluding the subsections on hierarchical models)
- Chapter 8: Model Parameter Inference
- Chapter 9: Computational Challenges
- Chapter 10: Special Topics (excluding the subsection on hierarchical models)
- Chapter 11: Selected Applications (for interested readers)
- Chapter 12: Projects and Exercises

Acknowledgement

This book presents the senior author's view of spatial statistics after spending about 50 years in the field. Having obtained the status of professor emeritus, he has now found the time and motivation to express his views in the format of a book with the help of two of his former PhD students.

Several prominent statisticians have inspired the book's writing. In the early 1980s, Professor Paul Switzer emphasised the importance of predictive statistics at Stanford University in Stanford, California, while Professor Andre Journel insisted that every problem is unique. In this environment, the first results in Bayesian Kriging were developed.

At the Norwegian Computing Center (NR) in Oslo, Norway, a group of researchers explored the possible applications of a large variety of methodologies within spatial statistics during the late 1980s. The results are summarised in the publication 'Topics in Spatial Statistics' by Nils Lid Hjort and Henning Omre. The results reported in this publication set the stage for this book.

Most of the relevant research was conducted at the Department of Mathematical Sciences, Norwegian University of Science and Technology (NTNU), in Trondheim, Norway, where all three authors were employed. We appreciate many inspiring years in the statistics group at the department. Three professors active in spatial statistics: Håkon Tjelmeland, Jo Eidsvik and Geir-Arne Fuglstad, must be named. To quote Andre Journel: 'The quality of a professor should be judged by the quality of his/her students.' The senior author is very thankful to all his former students, PhD students in particular, for all their contributions.

Kristine Heiney and Eva Hiripi have been very supportive in preparing the book manuscript. Kristine has improved our English writing style, while Eva has been our contact at Springer Verlag. We also appreciate the financial support from the Department of Mathematical Sciences, NTNU, and NR.

Lastly, Henning Omre is very grateful to his wife Marit, his old mother, children and grandchildren for their patience while he prepared the book manuscript. Most of the writing took place in the mornings in Trondheim and Tønsberg—and on Crete, Greece. An upside of their patience was that they experienced a much happier old man for the remainder of the day.

Torstein M. Fjeldstad extends his deepest gratitude to Thea and his parents for their unwavering support while writing this book. Additionally, sincere appreciation is extended to colleagues at NR, with a special mention of Petter Abrahamsen. Without his generous allocation of time and resources to this project, participating in writing this book would not have been possible.

Ole Bernhard Forberg is very grateful to Luciana, his mother, and his sister for their continued patience and support.

Trondheim, Norway
Oslo, Norway
Trondheim, Norway
December 2023

Henning Omre
Torstein M. Fjeldstad
Ole Bernhard Forberg

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Bayesian Spatial Modelling | 3 |
| 2.1 | Motivating Examples | 6 |
| 2.2 | Posterior Model: Simulation-Based Assessment | 10 |
| 2.3 | Notation | 13 |
| 3 | Conjugate Bayesian Models | 17 |
| 4 | Random Field Models | 21 |
| 4.1 | Continuous Random Fields | 23 |
| 4.1.1 | Gaussian RF Models | 23 |
| 4.1.2 | Hierarchical Gaussian RF Models | 25 |
| 4.2 | Event Random Fields | 27 |
| 4.2.1 | Poisson RF Models | 27 |
| 4.2.2 | Hierarchical Poisson RF Models | 30 |
| 4.3 | Mosaic Random Fields | 31 |
| 4.3.1 | Markov RF Models | 32 |
| 4.3.2 | Hierarchical Markov RF Models | 35 |
| 5 | Likelihood Models | 37 |
| 5.1 | Continuous Spatial Variables | 38 |
| 5.1.1 | Example: Continuous Spatial Variable | 39 |
| 5.2 | Event Spatial Variables | 41 |
| 5.2.1 | Example: Event Spatial Variable | 42 |
| 5.3 | Mosaic Spatial Variables | 43 |
| 5.3.1 | Example: Mosaic Spatial Variable | 44 |
| 6 | Prior Models | 47 |
| 6.1 | Continuous Spatial Variables: Gaussian RF Models | 48 |
| 6.2 | More on Gaussian RF Prior Models | 49 |
| 6.2.1 | Model Parameters and Characteristics | 49 |
| 6.2.2 | Model Validation | 57 |

| | | |
|----------|--|------------|
| 6.2.3 | Explanatory Spatial Variables | 58 |
| 6.2.4 | Related Topics | 59 |
| 6.2.5 | Example: Gaussian RF | 62 |
| 6.3 | Hierarchical Gaussian RF Models | 63 |
| 6.4 | Event Spatial Variables: Poisson RF Models | 64 |
| 6.5 | More on Poisson RF Prior Models | 66 |
| 6.5.1 | Model Parameters and Characteristics | 66 |
| 6.5.2 | Model Validation | 69 |
| 6.5.3 | Explanatory Spatial Variables | 70 |
| 6.5.4 | Related Topics | 71 |
| 6.5.5 | Example: Poisson RF | 73 |
| 6.6 | Hierarchical Poisson RF Models | 74 |
| 6.7 | Mosaic Spatial Variables: Markov RF Models | 76 |
| 6.8 | More on Markov RF Prior Models | 77 |
| 6.8.1 | Model Parameters and Characteristics | 78 |
| 6.8.2 | Model Validation | 80 |
| 6.8.3 | Explanatory Spatial Variables | 81 |
| 6.8.4 | Related Topics | 82 |
| 6.8.5 | Example: Markov RF | 83 |
| 7 | Posterior Models | 85 |
| 7.1 | Continuous Spatial Variables | 86 |
| 7.1.1 | Gaussian RF Models | 87 |
| 7.1.2 | Example: Gaussian RF | 94 |
| 7.1.3 | Hierarchical Gaussian RF Models | 96 |
| 7.2 | Event Spatial Variables | 98 |
| 7.2.1 | Poisson RF Models | 98 |
| 7.2.2 | Example: Poisson RF | 103 |
| 7.2.3 | Hierarchical Poisson RF Models | 105 |
| 7.3 | Mosaic Spatial Variables | 106 |
| 7.3.1 | Markov RF Models | 107 |
| 7.3.2 | Example: Markov RF | 113 |
| 8 | Model Parameter Inference | 115 |
| 8.1 | Gaussian RF Models | 116 |
| 8.1.1 | Example: Gaussian RF | 120 |
| 8.2 | Poisson RF Models | 121 |
| 8.2.1 | Example: Poisson RF | 123 |
| 8.3 | Markov RF Models | 124 |
| 8.3.1 | Example: Markov RF | 125 |
| 9 | Computational Challenges | 127 |
| 9.1 | Gaussian RF Models | 127 |
| 9.1.1 | Sparse Matrix Representations | 129 |
| 9.1.2 | Localisation Approximations | 135 |

| | | |
|-------------------------|---|------------|
| 9.2 | Poisson RF Models | 137 |
| 9.3 | Markov RF Models | 142 |
| 10 | Special Topics | 143 |
| 10.1 | Classes of Simulation Algorithms | 143 |
| 10.1.1 | Transformation Algorithms | 144 |
| 10.1.2 | Sequential Algorithms | 144 |
| 10.1.3 | Rejection Algorithms | 145 |
| 10.1.4 | Iterative Algorithms | 146 |
| 10.1.5 | Approximate Algorithms | 147 |
| 10.2 | Geostatistics: Kriging Prediction Models | 148 |
| 10.3 | Gaussian Markov RF Models | 152 |
| 10.4 | Gaussian Basis Function RF Models | 157 |
| 10.5 | Kernel Predictors for Gaussian RF Models | 161 |
| 10.6 | Clustered/Repulsive Event RF | 167 |
| 10.7 | INLA Framework for Poisson RF | 171 |
| 10.8 | Markov Random Profile Versus Markov Random Chain | 174 |
| 10.9 | Model Parameter Inference in Hierarchical RF Models | 177 |
| 11 | Selected Applications | 183 |
| 11.1 | Continuous Spatial Variables | 183 |
| 11.2 | Event Spatial Variables | 188 |
| 11.3 | Mosaic Spatial Variables | 192 |
| 12 | Projects and Exercises | 195 |
| 12.1 | Projects in Spatial Modelling | 195 |
| 12.1.1 | Project Text: Continuous Spatial Variables | 196 |
| 12.1.2 | Project Text: Event Spatial Variables | 200 |
| 12.1.3 | Project Text: Mosaic Spatial Variables | 203 |
| 12.2 | Exercises in Spatial Modelling | 205 |
| 12.2.1 | Exercise Sets: Continuous Spatial Variables | 205 |
| 12.2.2 | Exercise Sets: Event Spatial Variables | 213 |
| 12.2.3 | Exercise Sets: Mosaic Spatial Variables | 218 |
| References | | 223 |

List of Abbreviations

| | |
|------|--|
| BLU | Best linear unbiased |
| CAR | Conditional autoregressive |
| cdf | Cumulative distribution function |
| INLA | Integrated nested Laplace approximation |
| LSE | Least squared error |
| MAP | Maximum a posteriori |
| McMC | Markov chain Monte Carlo |
| ML | Maximum likelihood |
| MMAP | Maximum marginal a posteriori |
| MML | Maximum marginal likelihood |
| MMPL | Maximum marginal pseudo-likelihood |
| pdf | Probability density function |
| RC | Random chain |
| RF | Random field |
| RP | Random profile |
| SPDE | Stochastic partial differential equation |

Chapter 1

Introduction



A coin rests at your thumb, ready to toss. You ask: After I toss, what are the probabilities for heads and tails? Here, probability characterises the outcome of a future randomised experiment. A coin rests at your thumb, you toss, and quickly cover it with your hand. You ask: What are the probabilities for heads and tails? Here, probability characterises the lack of information about the outcome of a past randomised experiment. Spatial probabilities usually correspond to the latter interpretation.

Probability assessment and statistical analysis are closely related activities. Probability assessments are based on abstract mathematical models. The random variable in focus is defined to take outcomes in a specific sample space on which a normalised probability measure is assigned. This abstract model can assess the probability of any event defined by the random variable. In practice, a parametric probability model is used such that the model can be adapted to the application at hand. Statistical modelling is based on a set of observations related to the variable in focus. Every datum tells a story. Hence, statistics is an applied discipline. The parameters of the probability model are assessed from the observations by statistical inference. The calibrated probability model and the set of observations can be used to predict future outcomes of the random variable in focus. In practice, probability assessments and statistical inference are jointly referred to as statistical analysis, and this analysis is used in different settings:

- **Descriptive Statistics:** Provides insight into the phenomenon under study by computing summary measures and visualising the observations. No probability model is required because the analysis does not extend beyond the set of observations. Descriptive statistics is often referred to as data analysis.
- **Predictive Statistics:** Provides predictions of new observations, forecasts of future events and interpolations of events of the phenomenon under study with associated uncertainty statements. A probability model is required because the analysis extends beyond the set of observations, but because the reliability of the predictions is the only objective, simple operational models will be preferred. Predictive statistics includes classification, forecasting and interpolation.

- **Confirmatory Statistics:** Provides basic understanding by testing hypotheses concerning the model parameters representative of the phenomenon under study. A probability model with a representative and suitable parametrisation is necessary, and a lack of representativeness may lead to misleading conclusions. Confirmatory statistics is often referred to as hypothesis testing.

Any statistical analysis depends on a reliable model for the relations between the observations and the phenomenon. Traditionally, the observations are acquired in a designed experiment and assumed to be independent and identically distributed. In modern technological society, with advanced computing facilities, sensor technology and graphical display units, a large variety of observation acquisition procedures are used, and observations are often indirectly collected, sometimes sequentially. These observations need to be associated with the phenomenon using likelihood models. Such models may incorporate physical relationships as convolutions and spatial blurring. Lastly, misclassification may occur during the gathering of the observations, and complex observation error structures may need to be taken into account. Given the spatial variable, these likelihood models are defined by probability models for the observations.

This book is presented in a predictive statistical spirit. We specify a probability model for the phenomenon under study that captures its primary characteristics known from experience and studies of comparable phenomena. We denote this model as the prior model. Given the available observations and based on the likelihood and prior models, Bayes' rule can be used to provide a posterior probability model for the phenomenon. This construction is termed Bayesian modelling and has proven exceptionally useful in predictive statistical settings. A Bayesian approach in this setting is non-controversial because the model is only used pragmatically to improve the predictions with associated uncertainty quantifications, and the prior model is defined for observable variables, which naturally can be elicited in an empirical Bayesian framework. As the variable of interest in principle is observable, though indirectly, expert knowledge based on previous experiments or physical models may dictate the choice of the prior model. In confirmatory statistical settings, Bayesian inference is more controversial when testing hypotheses concerning the characteristics of the phenomenon under study. Lastly, the prior model parameters can also be assessed in a Bayesian framework. This joint modelling is referred to as hierarchical Bayesian modelling.

Chapter 2

Bayesian Spatial Modelling



Through laws formulated by great scientists, we have gained a deeper understanding of phenomena in our universe: Newton gave us a law for gravitational behaviour, Maxwell described electromagnetic phenomena, Carnot explained thermodynamic processes and Einstein expressed mass–energy equivalence. In 1763, the mystical Reverend Bayes formulated a law quantifying and combining information from several sources—expert experience and observations—and gave us the probabilistic Bayes' rule. These fundamental principles, applicable across various scientific fields, lay the foundation for technological advances in modern societies.

Observations from the variable of interest are often indirectly collected by a complex acquisition procedure. The objective is to assess the variable of interest based on the available observations. In this book, we focus on spatial variables and consider the prediction and reconstruction of these spatial variables given a set of observations. Examples may be found in applications such as image analysis, remote sensing and geophysics.

Spatial statistics, in contrast to traditional statistics, is characterised by several special features that need to be accounted for in Bayesian spatial modelling:

- Spatial reference: The variables of interest are assigned coordinate references within a spatial domain, and spatial coupling may cause strong dependencies among the variables. The spatial coupling between spatial variables tends to decrease with increasing distance between their spatial references.
- High-dimensional variables: The spatial reference of a spatial variable is typically discretised onto a spatial grid. Each grid node on the discretised spatial grid defines a one-dimensional variable. These one-dimensional variables comprise the dimensions of the spatial variable. The dimensions of the variable exhibit spatial interdependence and may be strongly correlated. In cases involving three-dimensional spatial reference domains, variable dimensions in the order of 10^9 to 10^{12} are not uncommon.
- Single-realisation inference: Model parameter inference is frequently conducted using a single realisation of the spatial variable, which includes spatial coupling.

Acquiring a population of independent, identically distributed spatial variables is seldom possible. Therefore, the spatial model typically requires some form of ergodicity to ensure reliable estimators for model parameters.

- Expert knowledge: Spatial variables correspond to real-world phenomena and are observable, contrary to model parameters, which are mathematical constructs and therefore not directly observable. In Bayesian spatial modelling, the prior model can be informed by expert knowledge and previous observations, which makes empirical Bayesian approaches appealing.
- Observation acquisition: Spatial variables are commonly observed as averages over specific observation areas or volumes, often referred to as spatial support. Moreover, surface and sub-surface spatial variables are usually not directly observable, requiring indirect acquisition procedures. Furthermore, the spatial continuity of the spatial variable makes preferential data collection possible, which adds complexity to the parameter inference process. The likelihood model must capture these observational characteristics.
- Realisations versus predictions: Bayesian spatial modelling is defined in a predictive setting, with the variable of interest being directly or indirectly observable by some acquisition procedure. Hence, realisations of the variable represent the phenomenon, whereas predictions are typically based on a locationwise loss criterion. The former correctly represents the spatial heterogeneity of the phenomenon, whereas the latter appears spatially smoother and more homogeneous.
- Quality criterion: Bayesian spatial modelling is defined in a predictive setting. Therefore, the minimum cross-validated error is the preferred quality criterion. This is contrary to traditional model parameter inference, which typically relies on a maximum likelihood (ML) criterion. Even model parameter inference in Bayesian spatial modelling should be based on minimum cross-validated error.

The spatial variable of interest is $\{s(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^q\}$ with $s(\mathbf{x})$ being either a continuous field like the terrain height, an event field like the locations of trees in a forest or a mosaic field like the land use in an area. The spatial reference \mathbf{x} is defined on the finite reference domain $D \subset \mathbb{R}^q$, which in practice has dimension q equal to one, two or three. Whenever the reference domain D is of dimension two or higher, $q > 1$, we denote the spatial variable as a field. In contrast, the term profile is reserved for one-dimensional reference domains. It is important to distinguish these two cases because the latter has a naturally ordered reference domain, unlike the former. The variable is spatially discretised as $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$, where L is a regular grid of size n , covering D , and the spatial variable is represented by the finite-dimensional n -vector \mathbf{s} , which is a vector of length n . Assume that an m -vector of observations \mathbf{d} related to the variable of interest is collected. The Bayesian spatial modelling focuses on assessing \mathbf{s} given \mathbf{d} , denoted by $[\mathbf{s} | \mathbf{d}]$.

We phrase the assessment in a probabilistic setting by using a Bayesian spatial model and perform spatial variable reconstruction along the lines of Casella and Berger (2002) and Tarantola (2005), as

$$\begin{aligned} [\mathbf{s} | \mathbf{d}] &\sim p(\mathbf{s} | \mathbf{d}) = [p(\mathbf{d})]^{-1} \times p(\mathbf{d}, \mathbf{s}) \\ &= \left[\int p(\mathbf{d} | \mathbf{s}) p(\mathbf{s}) d\mathbf{s} \right]^{-1} \times p(\mathbf{d} | \mathbf{s}) p(\mathbf{s}) \\ &= \text{const} \times p(\mathbf{d} | \mathbf{s}) p(\mathbf{s}), \end{aligned} \quad (2.1)$$

where $\mathbf{s} \sim p(\mathbf{s})$ indicates that the random variable \mathbf{s} is distributed according to the probability density/mass function (pdf) $p(\mathbf{s})$. The pdf $p(\mathbf{s} | \mathbf{d})$ is the posterior pdf, which is the ultimate solution of Bayesian analysis. Assessment of this posterior is often referred to as Bayesian inversion. The likelihood function $p(\mathbf{d} | \mathbf{s})$, being a function of \mathbf{s} , represents the observation acquisition procedure. In contrast, the prior pdf $p(\mathbf{s})$ summarises prior information about the spatial variable of interest. The likelihood and prior models uniquely define the posterior model, although the integral defining the normalising constant is usually complicated to calculate.

A schematic graphical display of the Bayesian inversion methodology is presented in Fig. 2.1. In display (a), the sample space of the reconstruction problem

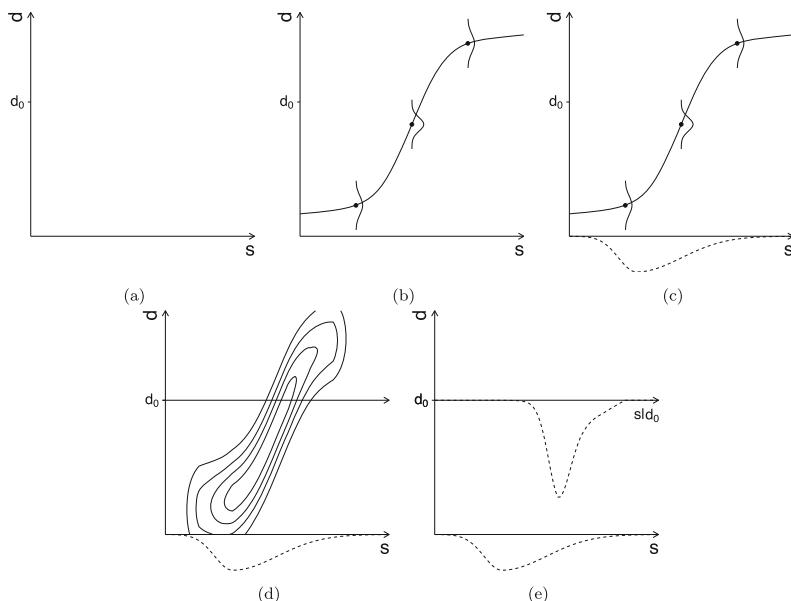


Fig. 2.1 Schematic presentation of Bayesian analysis: (a) dimension of study with observation, (b) likelihood model, (c) prior model, (d) joint variable and observation model and (e) posterior model

is defined as $\Omega_s \times \Omega_d$, and the observation \mathbf{d}_o is marked. Display (b) presents the likelihood function $p(\mathbf{d} | \mathbf{s}) = h(\mathbf{s}) + \mathbf{e}$ representing the observation acquisition procedure with additive observation error, while display (c) introduces the user-specified prior pdf $p(\mathbf{s})$, flipped on the axis for clarity. Display (d) contains the joint pdf $p(\mathbf{s}, \mathbf{d}) = p(\mathbf{d} | \mathbf{s})p(\mathbf{s})$, which fully defines all probabilistic relations of the variables involved $[\mathbf{s}, \mathbf{d}]$. The horizontal line at the observed value \mathbf{d}_o intersects the joint pdf at $p(\mathbf{s}, \mathbf{d}_o)$. Lastly, in display (e), the normalising constant $p(\mathbf{d}_o) = \int p(\mathbf{s}, \mathbf{d}_o) d\mathbf{s}$ is presented on the vertical axis, and the posterior pdf $p(\mathbf{s} | \mathbf{d}_o) = [p(\mathbf{d}_o)]^{-1} p(\mathbf{s}, \mathbf{d}_o)$ is presented flipped on the $[\mathbf{s} | \mathbf{d}_o]$ -axis. The latter pdf provides the ultimate solution in Bayesian reconstruction. By comparing the prior pdf $p(\mathbf{s})$ and the posterior pdf $p(\mathbf{s} | \mathbf{d}_o)$, one observes that by conditioning on the observed value \mathbf{d}_o , the expectation is shifted, and the variance is reduced, as anticipated.

2.1 Motivating Examples

The spatial field of interest is spatially discretised as $\{s(\mathbf{x}); \mathbf{x} \in L \subset D \subset \mathbb{R}^2\}$ and is represented by the n -vector \mathbf{s} . The spatial variable may be of continuous, event or mosaic type. The available observations, related to the variable of interest, are represented by the m -vector \mathbf{d} . We focus on assessment of \mathbf{s} from \mathbf{d} , and we formulate this as Bayesian spatial variable reconstruction. This section outlines three motivating examples, one for each variable type, while the relevant methodology is introduced later in the book.

Continuous Spatial Variable Let \mathbf{s} represent an unknown spatially discretised continuous field, as displayed in Fig. 2.2. The spatial variable to be reconstructed is presented in display (a). The likelihood model represents the observation acquisition procedure $p(\mathbf{d} | \mathbf{s})$, and the locations of five observed values are displayed as crosses. The observations are made without error and are represented by \mathbf{d} . To perform Bayesian reconstruction, we must assign a prior model $p(\mathbf{s})$ to the variable of interest. We specify a spatially stationary and isotropic Gaussian model. Displays (e), (f), (g) and (h) contain four realisations from $p(\mathbf{s})$. We observe different realisations of a continuous field that exhibit similar levels, variability and smoothness. The posterior model in Bayesian reconstruction is defined as $p(\mathbf{s} | \mathbf{d}) = \text{const} \times p(\mathbf{d} | \mathbf{s})p(\mathbf{s})$, and we may simulate realisations from this posterior model. Displays (i), (j), (k) and (l) contain four realisations from $p(\mathbf{s} | \mathbf{d})$. These realisations possess numerous characteristics inherited from the prior model and the conditioning observations. The observations are reproduced, and the prior model enforces the smoothness. Because of the conditioning on the observations, the set of posterior realisations exhibits less variability than the set of prior ones. Any of these posterior realisations could be the true spatial variable since they all display the supposedly correct spatial heterogeneity and reproduce the observations. We obtain estimates of the expectations and variances of $[\mathbf{s} | \mathbf{d}]$ by simulating a large

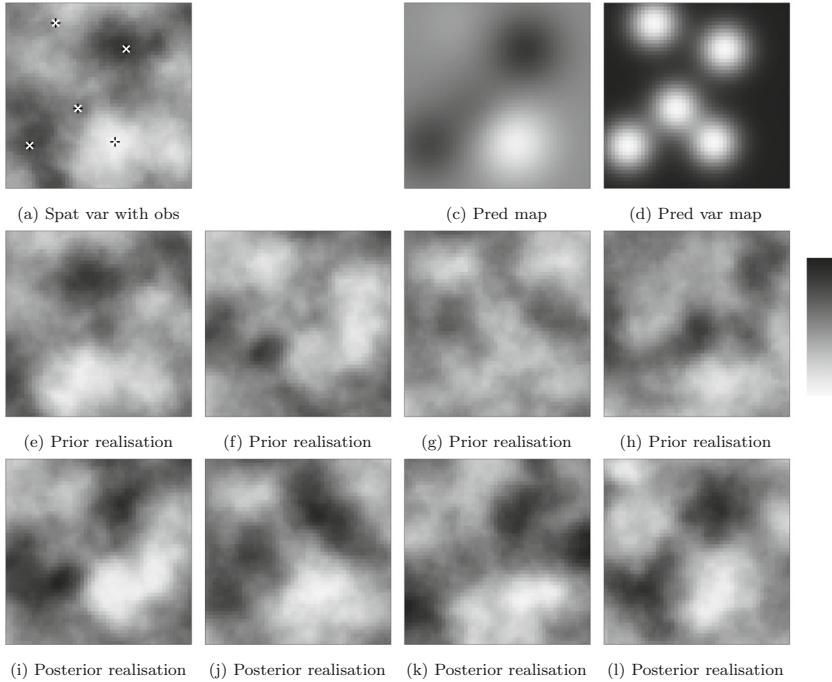


Fig. 2.2 Motivating example. Continuous spatial variable: (a) reference spatial variable with observations; (c) prediction map; (d) prediction variance map; (e), (f), (g), (h) four independent prior realisations; and (i), (j), (k), (l) four independent posterior realisations

set of posterior realisations and computing the locationwise averages and empirical variances. In displays (c) and (d) these estimates are displayed. The averages provide a locationwise prediction of $[s \mid d]$. Note that the observations are reproduced while regression towards the prior expectation function occurs in regions without observation influence, causing the prediction field to be smoothed. The prediction of $[s \mid d]$ appears as much smoother than each posterior realisation of $[s \mid d]$. The variances provide a locationwise prediction error measure, which is zero in observation locations and equal to the prior variance outside observation influence.

Event Spatial Variable Let s represent an unknown event field, Fig. 2.3. The spatial variable to be reconstructed is presented in display (a). The observation acquisition procedure is represented by the likelihood model $p(d \mid s)$, and the locations of the observed events are marked as circled points in display (a). The observations are collected with a probability for an event to be overlooked, and the observed events are denoted d . Bayesian reconstruction requires the assignment of a prior model on s , $p(s)$, and we specify a spatially stationary Poisson model. Displays (e), (f), (g) and (h) contain four realisations from this prior model, and different spatial event fields with similar event densities can be seen. The posterior model in Bayesian inversion is defined as $p(s \mid d) = \text{const} \times p(d \mid s)p(s)$,

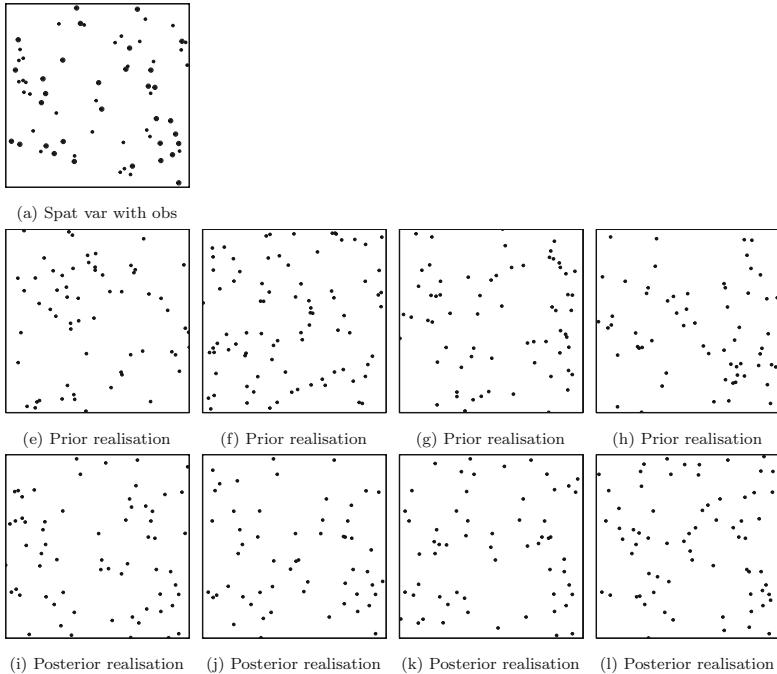


Fig. 2.3 Motivating example. Event spatial variable: (a) reference spatial variable with observations; (e), (f), (g), (h) four independent prior realisations; and (i), (j), (k), (l) four independent posterior realisations

and realisations may be simulated from this posterior model. Displays (i), (j), (k) and (l) contain four realisations from $p(\mathbf{s} \mid \mathbf{d})$. These realisations share some features because the observed events occur in all of them. Hence, the observations are reproduced, while the prior variability occurs elsewhere. Consequently, the set of posterior realisations exhibits less variability than the set of prior ones due to conditioning on the observations.

Mosaic Spatial Variable Let \mathbf{s} represent an unknown spatially discretised binary mosaic field with black and white labels, as displayed in Fig. 2.4. The spatial variable to be reconstructed is presented in display (a). The likelihood model $p(\mathbf{d} \mid \mathbf{s})$ represents the observation procedure, and the actual observations are displayed in display (b). The observations \mathbf{d} are generated by applying the following two steps on the spatial variable; first, a label-dependent misclassification process distorts the labels. Subsequently, considerable centred Gaussian observation noise is introduced. The prior model for \mathbf{s} , $p(\mathbf{s})$, is specified to be a spatially stationary coupled Markov model. Displays (e), (f), (g) and (h) contain four prior realisations from $p(\mathbf{s})$. Note that the spatial variables are very different, with the common feature of being predominantly one label with some small regions of the other label because of the spatial coupling in the prior model.

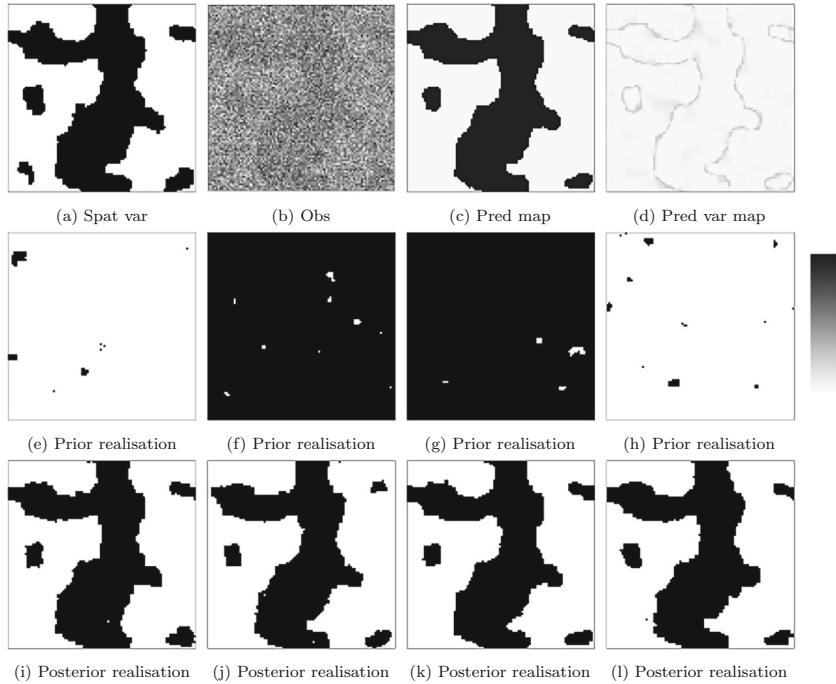


Fig. 2.4 Motivating example. Mosaic spatial variable: (a) reference spatial variable; (b) observations; (c) prediction map; (d) prediction variance map; (e), (f), (g), (h) four independent prior realisations; and (i), (j), (k), (l) four independent posterior realisations

The corresponding posterior model is $p(\mathbf{s} | \mathbf{d}) = \text{const} \times p(\mathbf{d} | \mathbf{s})p(\mathbf{s})$. Independent simulated realisations from this posterior pdf are included in displays (i), (j), (k) and (l). These posterior realisations are highly influenced by the observations, with the smoothness of the prior model enforced. The set of posterior realisations displays less variability than the set of prior ones. Various characteristics of the posterior model can be estimated by simulating a large set of posterior realisations. Display (c) contains the estimated maximum locationwise posterior predictions of the label. The corresponding prediction variances are presented in display (d). The former is obtained by selecting the most frequently occurring label by location, whereas the latter is estimated from a locationwise binomial model. The former provides a prediction of the label types of the mosaic field, whereas the latter represents classification uncertainty. As anticipated, the classification uncertainty is largest along the boundary between black and white labels, as seen in the latter display.

2.2 Posterior Model: Simulation-Based Assessment

The posterior model in Bayesian spatial modelling is usually assessed by simulation-based inference. By assigning a prior pdf $p(\mathbf{s})$ from which a realisation can be efficiently generated, one may define several brute-force simulation algorithms for generating realisations from the posterior pdf $p(\mathbf{s} | \mathbf{d})$. In Sect. 10.1, several classes of simulation algorithms are discussed, and three of these classes may be used to generate posterior realisations by brute-force.

The first alternative is the class of rejection inversion algorithms as defined in Algorithm 1. This algorithm provides an exact realisation from the posterior pdf, but it requires the identification of the maximum value of the likelihood function, which may be challenging to compute. The efficiency of the algorithm depends on the acceptance probabilities α , which, in turn, rely on the variability of the likelihood function values $p(\mathbf{d} | \mathbf{s})$. The acceptance probability α of each proposed sample is compared with a value sampled from the uniform pdf on the interval $[0, 1]$, denoted $\text{unif}[0, 1]$. The likelihood function involves a Dirac pdf if at least one continuous-valued observation is exact. In such cases, the acceptance probability will always be zero, and the algorithm no longer serves its purpose.

Algorithm 1: Rejection inversion algorithm

```

Initialise  $l_M = \max_{\mathbf{s}} p(\mathbf{d} | \mathbf{s})$ 
Define  $v = \text{const} \times l_M$ 
while  $b = 1, 2, \dots$  do
    Generate  $\mathbf{s}^p \sim p(\mathbf{s})$ 
    Calculate  $\alpha = \frac{p(\mathbf{s}^p | \mathbf{d})}{v \times p(\mathbf{s}^p)} = \frac{\text{const} \times p(\mathbf{d} | \mathbf{s}^p) p(\mathbf{s}^p)}{\text{const} \times l_M \times p(\mathbf{s}^p)} = \frac{p(\mathbf{d} | \mathbf{s}^p)}{l_M}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
        | break
    end
end
Result:  $\mathbf{s}^s = \mathbf{s}^b \sim p(\mathbf{s} | \mathbf{d})$ 

```

The class of iterative independent inversion algorithms, as defined in Algorithm 2, may be used. This algorithm provides an asymptotically exact realisation from the posterior pdf, but the convergence rate may be low. It follows an independent proposal Markov chain Monte Carlo (McMC) scheme, where the convergence rate relies solely on the acceptance probability α , as the independent proposals guarantee mixing. Thus, the variability of the likelihood function $p(\mathbf{d} | \mathbf{s})$ is crucial, and the presence of exact observations introducing the Dirac pdf in the likelihood model makes the algorithm unfeasible in practice.

Thirdly, the class of approximate inversion algorithms, as defined in Algorithm 3, can be employed. This algorithm provides an approximate realisation from the posterior pdf, and its approximation performance is expected to decrease with

Algorithm 2: Iterative independence inversion algorithm

```

Initialise  $\mathbf{s}^0$  such that  $p(\mathbf{s}^0 | \mathbf{d}) > 0$ 
Define  $g(\mathbf{s}'|\mathbf{s}) = p(\mathbf{s}')$ 
for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^p \sim p(\mathbf{s})$ 
    Calculate  $\alpha = \min \left\{ 1, \frac{p(\mathbf{s}^p|\mathbf{d})}{p(\mathbf{s}^{b-1}|\mathbf{d})} \times \frac{p(\mathbf{s}^{b-1})}{p(\mathbf{s}^p)} \right\} = \min \left\{ 1, \frac{p(\mathbf{d}|\mathbf{s}^p)}{p(\mathbf{d}|\mathbf{s}^{b-1})} \right\}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
    else
        | Set  $\mathbf{s}^b = \mathbf{s}^{b-1}$ 
    end
end

Result:  $\mathbf{s}^s = \mathbf{s}^{n_b} \sim p^{n_b}(\mathbf{s}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{s} | \mathbf{d})$ 

```

increasing dimensionality of \mathbf{s} . The accuracy of the approximation relies on the value of n_b and is susceptible to the curse of dimensionality. This algorithm is also sensitive to the variability of the likelihood function $p(\mathbf{d} | \mathbf{s})$.

Algorithm 3: Approximate inversion algorithm

```

Initialise  $n_b$  (large)
for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^{*b} \sim p(\mathbf{s})$ 
    Calculate  $\omega^b = \frac{p(\mathbf{s}^{*b}|\mathbf{d})}{p(\mathbf{s}^{*b})} = \text{const} \times p(\mathbf{d} | \mathbf{s}^{*b})$ 
end
Calculate  $\tilde{\omega}^b = \frac{\omega^b}{\sum_{c=1}^{n_b} \omega^c} = \frac{\text{const} \times p(\mathbf{d}|\mathbf{s}^{*b})}{\sum_{c=1}^{n_b} \text{const} \times p(\mathbf{d}|\mathbf{s}^{*c})} = \frac{p(\mathbf{d}|\mathbf{s}^{*b})}{\sum_{c=1}^{n_b} p(\mathbf{d}|\mathbf{s}^{*c})}$  for  $b = 1, 2, \dots, n_b$ 
Resample  $\mathbf{s}^{*s} \sim \{\mathbf{s}^{*1}, \mathbf{s}^{*2}, \dots, \mathbf{s}^{*n_b}\}$  with probability  $[\tilde{\omega}^1, \tilde{\omega}^2, \dots, \tilde{\omega}^{n_b}]$ 
Result:  $\mathbf{s}^s = \mathbf{s}^{*s}$  approximates  $\mathbf{s}^s \sim p(\mathbf{s} | \mathbf{d})$ 

```

The three brute-force simulation algorithms presented above are highly dependent on the shape of the likelihood function. The more informative the observations are, the poorer the algorithms are expected to perform. This is an unfortunate effect. If a general iterative McMC algorithm is used, one may define a more adaptive proposal mechanism than an independent generation from the prior pdf.

The class of iterative inversion algorithms, as defined in Algorithm 4, provides a sequence of asymptotically exact realisations from the posterior pdf. However, it is susceptible to a low convergence rate and poor mixing. The challenge is to specify an inversion-specific proposal mechanism. Experience tells us that using naive brute-force single-site proposal pdfs do not work well for high-dimensional spatial problems with strong spatial coupling. Developing more adaptive proposal mechanisms is challenging because the values of the pdf of the proposal and its reversal must be calculated.

Algorithm 4: Iterative inversion algorithm

```

Initialise  $\mathbf{s}^0$  such that  $p(\mathbf{s}^0 | \mathbf{d}) > 0$ 
Define  $g(\mathbf{s}' | \mathbf{s})$  adapted to the inversion
for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^p \sim g(\mathbf{s} | \mathbf{s}^{b-1})$ 
    Calculate  $\alpha = \min \left\{ 1, \frac{p(\mathbf{s}^p | \mathbf{d})}{p(\mathbf{s}^{b-1} | \mathbf{d})} \times \frac{g(\mathbf{s}^{b-1} | \mathbf{s}^p)}{g(\mathbf{s}^p | \mathbf{s}^{b-1})} \right\}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
    else
        | Set  $\mathbf{s}^b = \mathbf{s}^{b-1}$ 
    end
end

```

Result: $\mathbf{s}^s = \mathbf{s}^{n_b} \sim p^{n_b}(\mathbf{s}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{s} | \mathbf{d})$

Ideally, the observations \mathbf{d} should also be accounted for in the proposal mechanism. Consider the specified prior model $p(\mathbf{s})$ and define an approximate likelihood model $p_*(\mathbf{d} | \mathbf{s})$ such that the prior and likelihood model classes form a conjugate pair. With this choice of likelihood model, the corresponding posterior model $p_*(\mathbf{s} | \mathbf{d})$ becomes analytically tractable. This enables the design of efficient simulation algorithms and the calculation of the actual probabilities of the proposal and its reversal. The primary focus of this book is the definition and discussion of such conjugate classes of spatial models. An independent, conjugate proposal MCMC scheme with improved acceptance rates can then be defined.

The class of iterative independent conjugate inversion algorithms, as defined in Algorithm 5, provides a sequence of asymptotically exact realisations from the posterior model. A favourable convergence rate is obtained if the approximation of the likelihood model is reliable. The independent proposals ensure good mixing in practice. Note in particular that exact observations can be reproduced if they are reproduced in the proposal from $p_*(\mathbf{s} | \mathbf{d})$. Given that the approximate posterior model is analytically tractable, localised proposal kernels may also be defined.

Assume that the focus is on assessing a possibly non-linear function of \mathbf{s} , $s_w = w(\mathbf{s})$, with

$$[s_w | \mathbf{d}] \sim p(s_w | \mathbf{d}) = \int p(s_w | \mathbf{s}) p(\mathbf{s} | \mathbf{d}) d\mathbf{s}.$$

In a Bayesian setting, the solution may be approximated by a Monte Carlo algorithm, see Algorithm 6, and an asymptotically consistent estimate for $p(s_w | \mathbf{d})$ can be obtained. Later in this book, the focus primarily lies on linear functionals, for which analytical solutions are frequently attainable.

Algorithm 5: Iterative independence conjugate inversion algorithm

```

Initialise  $\mathbf{s}^0$  such that  $p(\mathbf{s}^0 | \mathbf{d}) > 0$ 
Define  $g(\mathbf{s}' | \mathbf{s}, \mathbf{d}) = p_*(\mathbf{s}' | \mathbf{d}) = \text{const} \times p_*(\mathbf{d} | \mathbf{s}') p(\mathbf{s}')$ 
for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^p \sim p_*(\mathbf{s} | \mathbf{d})$ 
    Calculate  $\alpha = \min \left\{ 1, \frac{p(\mathbf{s}^p | \mathbf{d})}{p(\mathbf{s}^{b-1} | \mathbf{d})} \times \frac{p_*(\mathbf{s}^{b-1} | \mathbf{d})}{p_*(\mathbf{s}^p | \mathbf{d})} \right\} = \min \left\{ 1, \frac{p(\mathbf{d} | \mathbf{s}^p)}{p_*(\mathbf{d} | \mathbf{s}^p)} \times \left[ \frac{p(\mathbf{d} | \mathbf{s}^{b-1})}{p_*(\mathbf{d} | \mathbf{s}^{b-1})} \right]^{-1} \right\}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
    else
        | Set  $\mathbf{s}^b = \mathbf{s}^{b-1}$ 
    end
end
Result:  $\mathbf{s}^s = \mathbf{s}^{n_b} \sim p^{n_b}(\mathbf{s}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{s} | \mathbf{d})$ 

```

Algorithm 6: Posterior function algorithm: Monte Carlo

```

Initialise  $n_s$  (large)
Define  $s_w = w(\mathbf{s})$ 
for  $s = 1, 2, \dots, n_s$  do
    Generate  $\mathbf{s}^s \sim p(\mathbf{s} | \mathbf{d})$ 
    Calculate  $s_w^s = w(\mathbf{s}^s)$ 
end
Empirically estimate  $p(s_w | \mathbf{d})$  using  $s_w^s$  for  $s = 1, 2, \dots, n_s$ 
Result: Empirical estimate of  $p(s_w | \mathbf{d})$ 

```

2.3 Notation

Lastly, we more clearly define the notation used in the previous sections and define the more general notation used in the remainder of the book. Generally, variables describing observable entities of the phenomenon under study are represented by lowercase Latin letters, whereas model parameters, describing non-observable model entities, are represented by Greek letters. Vectors are denoted by bold lowercase letters, whereas bold capital letters represent matrices. When removing a subset of elements from a vector, this is indicated by specifying the vector with a subscripted minus sign, followed by the removed subset. For two vectors, an inequality between them implies that the inequality holds for each pair of corresponding entries in the vectors. We use two vector–matrix operators: $\text{Vdiag}_n\{\cdot\}$ with an $(n \times n)$ matrix argument, which returns an n -vector containing the diagonal entries in the matrix, and $\text{Mdiag}_n\{\cdot\}$ with an n -vector argument, which returns an $(n \times n)$ diagonal matrix with the vector along the diagonal. Denote the trace of an $(n \times n)$ matrix \mathbf{A} by $\text{Tr}\{\mathbf{A}\}$.

The spatial variable of interest is denoted $\{s(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$, where the spatial reference \mathbf{x} in the reference domain D has dimension three. The variable $s(\mathbf{x})$ can be either real-valued, integer-valued or categorical. The three-dimensional reference domain D , with $|D| < \infty$, is, for notational convenience, assumed to be rectangular and axis-parallel. It is decomposed into a horizontal and a vertical component $D = [D_y, D_z] \subset [\mathbb{R}^2 \times \mathbb{R}]$ with corresponding spatial reference $\mathbf{x} = (\mathbf{y}, z)$. Furthermore, the reference domain D is spatially discretised into $L = [L_y, L_z]$, with L_y being a regular horizontal grid with quadratic grid nodes and L_z being a regular vertical grid. Here, the number of nodes is $n = n_y \times n_z$, and δ_n is the length of a grid unit side. The set of grid nodes in L is referred to by location or index. The corresponding volume of one cell of the three-dimensional spatial discretisation, the grid volume, is $\Delta_n = \Delta_{n_y} \times \Delta_{n_z}$. The grid volume centred at grid node i is the grid unit $\Delta_{ni} \subset D$. Hence, one has the set partition $\Delta_{ni} \cap \Delta_{nj} = \emptyset; i, j = 1, 2, \dots, n; i \neq j$ and $\cup_{i=1}^n \Delta_{ni} = D$ and the volume identity $n \Delta_n = |D|$.

The spatial variable of interest $\{s(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ can be spatially discretised into a vector of length n , which we refer to as the n -vector \mathbf{s} containing $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ in arbitrary order. Asymptotic evaluation of $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ can be performed in one of two ways, as defined in Stein (1999):

- Infill asymptotic analysis: The domain D is kept constant while the grid node volumes $\Delta_n \rightarrow 0$, such that $n \rightarrow \infty$. By using infill asymptotic analysis with suitable conditions, the limiting $\{s(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ of a given spatially discretised spatial variable can be evaluated.
- Extension asymptotic analysis: The grid volume Δ_n remains constant, while $D \rightarrow \infty$, consequently $n \rightarrow \infty$. Extension asymptotic analysis is mainly used to demonstrate model ergodicity, which ensures that each realisation from the model captures all model variability. Ergodicity ensures consistent inference of model parameters and predictions from estimators based on one single realisation.

Three types of spatial variables are discussed: continuous, event and mosaic spatial variables.

- Continuous spatial variables are denoted $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with $r(\mathbf{x}) \in \mathbb{R} : (-\infty, \infty)$. Continuous spatial variables represent almost continuous fields in space. Examples are abundant and include the temperature or pressure field in the atmosphere, the depth to geological horizons and the density of human tissues. The spatially discretised continuous variable on the grid $\{r(\mathbf{x}); \mathbf{x} \in L \subset D\}$ is represented by the n -vector $\mathbf{r} \in \mathbb{R}^n$. In this vector, each entry takes the value of the spatial variable in the corresponding grid node. A complete spatial representation based on the grid representation is usually obtained by a piecewise constant or piecewise planar function inside each grid unit. The former assigns the value in the grid node to the corresponding grid unit. The latter interpolates linearly from neighbouring grid node values.

- Event spatial variables are denoted $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with $k(\mathbf{x}) \in \mathbb{B} : \{0, 1\}$. Event spatial variables represent events that occur on a constant background. The spatial variable has a value of zero almost everywhere in the finite spatial domain D , except for in a finite number of singular locations in D with value one, representing the locations of a finite number of events. An event-location set $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}\}; \mathbf{x}_i \in D$ of size k_D may alternatively represent the event variable. Examples include the locations of trees in a forest and the locations of cracks in a rock sub-surface. The spatially discretised event count variable on the grid $\{k_\Delta(\mathbf{x}); \mathbf{x} \in L \subset D\}$ with $k_\Delta(\mathbf{x}) \in \mathbb{N}_+$ is represented by the event count n -vector $\mathbf{k} \in \mathbb{N}_+^n$. This vector contains the number of events occurring in each grid unit $\Delta_{ni}; i = 1, 2, \dots, n$. An approximate event-location set \mathbb{X}_D^n provides a complete spatial representation based on the spatially discretised representation. This set is obtained by generating k_i uniformly distributed locations within the volume of each grid unit Δ_i , where k_i corresponds to the event count in the grid unit. Hence, the spatial event intensity function is defined to be a piecewise constant approximation.
- Mosaic spatial variables are denoted $\{l(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with $l(\mathbf{x}) \in \mathbb{L} : \{1, 2, \dots, n_L\}$. Mosaic spatial variables represent categorical labels distributed across the spatial domain D . Examples are the land use on the Earth's surface, the rock types in the sub-surface and the types of tissue in the human brain. The spatially discretised mosaic variable on the grid $\{l(\mathbf{x}); \mathbf{x} \in L \subset D\}$ is represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. Each entry in the vector takes the value corresponding to the label that occurs in the spatial variable at the corresponding grid node. A complete spatial representation based on the grid representation is usually obtained by assigning the label of each grid node to its corresponding grid unit.

In the following, $p(\mathbf{s})$ denotes a probability density/mass function (pdf) of the random variable $\mathbf{s} \in \Omega_s$, where Ω_s is the sample space, whereas $\text{Prob}\{\mathbf{s} \in \mathcal{B}\}$ denotes the probability that the random variable is in the subset $\mathcal{B} \subset \Omega_s$. The conditional pdf for \mathbf{s} given \mathbf{t} is denoted $p(\mathbf{s} | \mathbf{t})$. To specify dependence on a certain model parameter, the format $p(\mathbf{s}; \theta)$ is sometimes used, with θ being the set of model parameters on which the pdf depends.

For the Gaussian random n -vector \mathbf{s} we use the specific notation

$$\begin{aligned} \mathbf{s} \sim p(\mathbf{s}) &= \phi_n(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= [2\pi]^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}[\mathbf{s} - \boldsymbol{\mu}]^T \boldsymbol{\Sigma}^{-1} [\mathbf{s} - \boldsymbol{\mu}]\right), \end{aligned} \quad (2.2)$$

where the model parameters are the n -vector $\boldsymbol{\mu}$ containing the expectations and the non-negative definite covariance matrix $\boldsymbol{\Sigma}$ of dimension $(n \times n)$. From a computational point of view, $|\boldsymbol{\Sigma}|^{-1}$ and $\boldsymbol{\Sigma}^{-1}$ are demanding, but these challenges are widely studied in the numerical analysis community, as discussed in Trefethen and Bau (1997). When the Gaussian random vector is centred at zero, meaning it has an expected value of zero, it is referred to as a centred Gaussian. The α -quantile of the univariate standardised Gaussian variable s with cumulative distribution

function (cdf) is denoted z_α ; that is, $\text{Prob}\{s \leq z_\alpha\} = \alpha$. We also use the notation \mathbf{i}_n for a unit n -vector, e.g. a vector of ones, and \mathbf{I}_n for an identity matrix of dimension $(n \times n)$. Next, $I(A)$ is an indicator function taking value 1 if A is true and 0 otherwise. For sets of real \mathbb{R}^q and natural \mathbb{N}^q numbers of dimension q , we append the subscripts ‘ \oplus ’ and ‘ $+$ ’ to denote sets consisting of positive values, where \oplus includes zero and $+$ excludes zero. Additional subsets of \mathbb{R}^q and \mathbb{N}^q are similarly defined.

Chapter 3

Conjugate Bayesian Models



Computers, sensors and methodology are the driving forces of modern technological advances. However, we remain far behind our technological ambitions, as we continuously adjust these ambitions and improve sensors to collect even more data. As the number of model dimensions increases, computer simulations become dramatically less efficient. Thus, maintaining analytical tractability, even if only for approximate models, remains a powerful tool.

Classical Bayesian inference, as defined in Casella and Berger (2002), concerns the estimation of model parameters, contrary to Bayesian spatial modelling, which is defined in a predictive setting for the spatial variable. The classical approach focuses on the posterior pdf of a low-dimensional vector of model parameters $\boldsymbol{\theta}$ given a set of observations \mathbf{d} , hence on $p(\boldsymbol{\theta} | \mathbf{d})$. To assess this posterior pdf, one must define a likelihood function $p(\mathbf{d} | \boldsymbol{\theta})$ and specify a prior pdf $p(\boldsymbol{\theta})$. Traditional Bayesian inference has introduced the concept of conjugate classes of prior pdfs to avoid calculating complex integrals. For a given likelihood function $p(\mathbf{d} | \boldsymbol{\theta})$, with a prior model $p(\boldsymbol{\theta})$ from the corresponding conjugate class of pdfs, the resulting posterior pdf $p(\boldsymbol{\theta} | \mathbf{d})$ will belong to the same conjugate class of pdfs. Consequently, the parameters of the posterior model can be analytically assessed from the hyperparameters of the conjugate prior and likelihood models together with the actual observations.

In Bayesian spatial modelling, the integral calculations are even more challenging than in traditional Bayesian inference because the spatial variable of interest is of a much higher dimension than the model parameters. In Bayesian spatial modelling, we introduce the concept of conjugate classes of prior pdfs for a given class of likelihood functions. The focus is on reconstructing a spatial variable spatially discretised into the n -vector \mathbf{s} , based on the observations represented in the m -vector \mathbf{d} . We use Bayesian spatial modelling, as defined in Expression (2.1), to perform this reconstruction. Consequently, the likelihood function $p(\mathbf{d} | \mathbf{s})$ must be specified and the prior pdf $p(\mathbf{s})$ assigned. Under these assumptions, the corresponding posterior pdf $p(\mathbf{s} | \mathbf{d})$ is fully defined. Inspired by the concept of conjugate prior pdfs in traditional Bayesian inference, we present Definition 1.

Definition 1 (Conjugate Class of Prior PDFs) Consider the Bayesian spatial model

$$p(\mathbf{s} \mid \mathbf{d}; \boldsymbol{\theta}_{s|d}) = \text{const} \times p(\mathbf{d} \mid \mathbf{s}; \boldsymbol{\psi}_d) p(\mathbf{s}; \boldsymbol{\theta}_s)$$

with likelihood function $p(\mathbf{d} \mid \mathbf{s}; \boldsymbol{\psi}_d)$ in a parametrised pdf class \mathcal{L}_{ψ} and prior pdf $p(\mathbf{s}; \boldsymbol{\theta}_s)$ in a parametrised pdf class \mathcal{P}_{θ} . If the associated posterior pdf $p(\mathbf{s} \mid \mathbf{d}; \boldsymbol{\theta}_{s|d})$ is also in the pdf class \mathcal{P}_{θ} , then the pdf class \mathcal{P}_{θ} is termed a conjugate class with respect to the likelihood function class \mathcal{L}_{ψ} . The posterior model parameter $\boldsymbol{\theta}_{s|d}$ is a function of $[\boldsymbol{\psi}_d, \boldsymbol{\theta}_s, \mathbf{d}]$.

Three examples of the conjugate characteristic for different variable types follow.

Example: Real-Valued Variable

Consider the unknown variable $s \in \mathbb{R}$ of interest, for example, the height/depth relative to sea level. Assume that one observation is collected as $[d \mid s] = \alpha + \beta s + e \in \mathbb{R}$, with $\alpha, \beta \in \mathbb{R}$ and the random observation error $e \in \mathbb{R}$ given by a centred Gaussian pdf with variance $\sigma_e^2 \in \mathbb{R}_+$. Hence, the observation appears as an affine transformation of the variable of interest with additive Gaussian error. We refer to the likelihood model as a Gauss-linear model because it is linear in the conditioning variable s and has an additive Gaussian error term. The likelihood function will then be in the Gauss-linear pdf class, $p(d \mid s; \boldsymbol{\psi}_d) = \phi_1(d; \alpha + \beta s, \sigma_e^2)$, with $\boldsymbol{\psi}_d = [\alpha, \beta, \sigma_e^2]$ known. Next, assign a prior model to s from the Gaussian pdf class, $p(s; \boldsymbol{\theta}_s) = \phi_1(s; \mu_s, \sigma_s^2)$, with $\mu_s \in \mathbb{R}$ and $\sigma_s^2 \in \mathbb{R}_+$ such that $\boldsymbol{\theta}_s = [\mu_s, \sigma_s^2]$. Furthermore, let $\boldsymbol{\theta}_s$ be known.

The posterior pdf $p(s \mid d; \boldsymbol{\theta}_{s|d}); s \in \mathbb{R}$ can be assessed by using Bayesian inversion as

$$\begin{aligned} p(s \mid d; \boldsymbol{\theta}_{s|d}) &= \left[\int p(d \mid s; \boldsymbol{\psi}_d) p(s; \boldsymbol{\theta}_s) ds \right]^{-1} \times p(d \mid s; \boldsymbol{\psi}_d) p(s; \boldsymbol{\theta}_s) \\ &= \text{const} \times \exp\left(-\frac{(d - (\alpha + \beta s))^2}{2\sigma_e^2}\right) \exp\left(-\frac{(s - \mu_s)^2}{2\sigma_s^2}\right) \\ &= \text{const} \times \exp\left(-\frac{(s - \mu_{s|d})^2}{2\sigma_{s|d}^2}\right) \end{aligned}$$

with

$$\begin{aligned} \mu_{s|d} &= \mu_s + \frac{\beta\sigma_s^2}{\beta^2\sigma_s^2 + \sigma_e^2} [d - (\alpha + \beta\mu_s)] \\ \sigma_{s|d}^2 &= \sigma_s^2 \left[1 - \frac{\beta^2\sigma_s^2}{\beta^2\sigma_s^2 + \sigma_e^2} \right]. \end{aligned}$$

Hence, the posterior pdf $p(s \mid d; \boldsymbol{\theta}_{s|d})$; $s \in \mathbb{R}$ belongs to the Gaussian pdf class. The parameters are $\boldsymbol{\theta}_{s|d} = [\mu_{s|d}, \sigma_{s|d}^2]$, which entails that $\boldsymbol{\theta}_{s|d}$ is a function of $[\boldsymbol{\psi}_d, \boldsymbol{\theta}_s, d]$. Since the prior pdf $p(s; \boldsymbol{\theta}_s)$ belongs to the Gaussian pdf class, the Gaussian prior pdf class is a conjugate class with respect to the Gauss-linear likelihood function class as per Definition 1.

Example: Infinite Countable-Valued Variable

Consider the unknown variable $s \in \mathbb{N}_\oplus$ of interest, for example, the number of trees in a forest. Assume that one observation is collected as $[d \mid s] = \sum_{i=1}^s I(u_i < \alpha) \in \mathbb{N}_{[0,s]}$ with probability $\alpha \in \mathbb{R}_{[0,1]}$ of being registered, and random terms $u_i \in \mathbb{R}_{[0,1]}; i = 1, 2, \dots, s$. Each u_i is assumed to be from the uniform $[0, 1]$ pdf and u_i and u_j are assumed to be independent when $i \neq j$. Consequently, the observation appears as the number of items after accounting for registration error, i.e. overlooked events. Then, the likelihood function belongs to the binomial pdf class, $p(d \mid s; \boldsymbol{\psi}_d) = \binom{s}{d} \alpha^d (1 - \alpha)^{s-d}$ with $\boldsymbol{\psi}_d = [\alpha]$ known. A prior model from the Poisson pdf class is assigned to s from the Poisson pdf class, $p(s; \boldsymbol{\theta}_s) = \frac{\lambda_s^s}{s!} \exp(-\lambda_s)$ with $\lambda_s \in \mathbb{R}_+$, hence $\boldsymbol{\theta}_s = [\lambda_s]$. Furthermore, let $\boldsymbol{\theta}_s$ be known.

The posterior pdf $p(s \mid d; \boldsymbol{\theta}_{s|d})$; $s \in \mathbb{N}_{[d,\infty]}$ can be assessed by using Bayesian inversion as

$$\begin{aligned} p(s \mid d; \boldsymbol{\theta}_{s|d}) &= \left[\sum_s p(d \mid s; \boldsymbol{\psi}_d) p(s; \boldsymbol{\theta}_s) \right]^{-1} \times p(d \mid s; \boldsymbol{\psi}_d) p(s; \boldsymbol{\theta}_s) \\ &= \text{const} \times \binom{s}{d} \alpha^d (1 - \alpha)^{s-d} \times \frac{\lambda_s^s}{s!} \exp(-\lambda_s) \\ &= \text{const} \times \frac{\lambda_{s|d}^{s-d}}{(s-d)!} \end{aligned}$$

with

$$\lambda_{s|d} = (1 - \alpha)\lambda_s.$$

The non-identified items are $\Delta s = s - d$. The pdf of $[(s = d + \Delta s) \mid d]$ is identical to the pdf $[\Delta s \mid d]$ being shifted by d . The posterior pdf for the non-identified items, $p(\Delta s \mid d; \boldsymbol{\theta}_{\Delta s|d})$; $\Delta s \in \mathbb{N}_\oplus$, is in the Poisson pdf class, as is the prior pdf $p(s; \boldsymbol{\theta}_s)$. The parameter is $\boldsymbol{\theta}_{\Delta s|d} = [\lambda_{s|d}]$, and $s = d + \Delta s$, hence $\boldsymbol{\theta}_{s|d} = [\lambda_{s|d}, d]$, which is a function of $[\boldsymbol{\psi}_d, \boldsymbol{\theta}_s, d]$. Consequently, according to Definition 1, the Poisson prior pdf class is a conjugate class with respect to the binomial likelihood function class.

Example: Finite Categorical-Valued Variable

Consider the unknown variable $s \in \mathbb{L} : \{l_1, l_2, \dots, l_{n_L}\}$ of interest with the number of labels n_L finite, for example, the possible categories of an item. Assume that one observation is collected as $[d \mid s] = \alpha_s + e \in \mathbb{R}$ with $\alpha_s \in \boldsymbol{\alpha} : \{\alpha_{l_i}; i = 1, 2, \dots, n_L\} \in \mathbb{R}^{n_L}$ and $e \in \mathbb{R}$ being a random observation error given by a centred

Gaussian pdf with variance $\sigma_e^2 \in \mathbb{R}_+$. Hence, each observation appears as a real value equal to the sum of a label level and a Gaussian error term. Because the expectation of $[d | s]$ depends on a set of label levels for s , we say that the likelihood model belongs to the Gaussian mixture pdf class. The expression for the likelihood function class is $p(d | s; \psi_d) = \phi_1(d; \alpha_s, \sigma_e^2); \alpha_s \in \boldsymbol{\alpha}$, with known $\psi_d = [\boldsymbol{\alpha}, \sigma_e^2]$. Next, assign a prior model to s from the Gibbs pdf class, $p(s; \theta_s) = \text{const} \times \exp(\mu_s)$ with $\mu_s \in \boldsymbol{\mu}_s : \{\mu_{l_i}; i = 1, 2, \dots, n_L\}$, thus $\theta_s = [\boldsymbol{\mu}_s]$. Furthermore, let θ_s be known.

The posterior pdf $p(s | d; \theta_{s|d}); s \in \mathbb{L}$ can be assessed by using Bayesian inversion as

$$\begin{aligned} p(s | d; \theta_{s|d}) &= \left[\sum_s p(d | s; \psi_d) p(s; \theta_s) \right]^{-1} \times p(d | s; \psi_d) p(s; \theta_s) \\ &= \text{const} \times \exp\left(-\frac{(d - \alpha_s)^2}{2\sigma_e^2}\right) \times \exp(\mu_s) \\ &= \text{const} \times \exp(\mu_{s|d}) \end{aligned}$$

with

$$\mu_{s|d} = \mu_s - \frac{(d - \alpha_s)^2}{2\sigma_e^2}.$$

Hence, both the prior pdf $p(s; \theta_s)$ and posterior pdf $p(s | d; \theta_{s|d}); s \in \mathbb{L}$ belong to the Gibbs pdf class. The parameter of the latter is $\theta_{s|d} = [\boldsymbol{\mu}_{s|d}]$, with $\boldsymbol{\mu}_{s|d} : \{\mu_{l_i|d}; i = 1, 2, \dots, n_L\}$, which entails that $\theta_{s|d}$ is a function of $[\psi_d, \theta_s, d]$. Consequently, according to Definition 1, the Gibbs prior pdf class is conjugate with respect to the Gaussian mixture likelihood function class.

For continuous spatial variables, the class of Gaussian prior models is known to be a conjugate class with respect to Gauss-linear likelihood functions. Hence, if the observations are collected through a linear forward model with additive Gaussian errors, and the prior pdf is specified as Gaussian, the posterior model is Gaussian. This characteristic is the basis for Kriging prediction and conditional simulation in geostatistics, as discussed in Chiles and Delfiner (2012). Moreover, for event spatial variables, the Poisson prior model class is conjugate with respect to binomial likelihood functions. For mosaic spatial variables, the class of Markov prior models is a conjugate class with respect to the Gaussian mixture likelihood function class. These conjugate characteristics are the primary reason for the frequent use of these spatial models.

Chapter 4

Random Field Models



Einstein once stated, “God does not play dice . . .”, and he may be right. But our lack of precise information about these purportedly deterministic creations of God still makes probability a suitable tool for us in constructing models of the universe. Probability provides us a vocabulary to express and quantify our lack of complete information.

A probabilistic interpretation of the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with continuous spatial reference $\mathbf{x} \in D$ is obtained through the concept of a random field (RF) in Definition 2. The Kolmogorov requirements are related to the consistency of pdfs for different subsets of variables of multivariate random variables. An in-depth definition of RFs is given in Yaglom (1987).

Definition 2 (Random Field (RF)) Consider the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D\}$ and define the pdfs $(s(\mathbf{x}_1), s(\mathbf{x}_2), \dots, s(\mathbf{x}_k)) \sim p(s(\mathbf{x}_1), s(\mathbf{x}_2), \dots, s(\mathbf{x}_k))$ for all configurations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) \in D^k$ and for all $k \in \mathbb{N}_+$. Let the Kolmogorov consistency requirements related to dimensionality hold for this set of pdfs. This infinite set of pdfs defines the random field.

RF models that are spatially stationary, or both spatially stationary and isotropic, as detailed in Definition 3, are commonly utilised as prior models in Bayesian spatial modelling. Stationarity implies that $p(s(\mathbf{x}_1), s(\mathbf{x}_2), \dots, s(\mathbf{x}_k)) = p(s(\mathbf{x}_1 + \mathbf{h}), s(\mathbf{x}_2 + \mathbf{h}), \dots, s(\mathbf{x}_k + \mathbf{h}))$ for all $\mathbf{h} \in \mathbb{R}^3$ and $\mathbf{x}, \mathbf{x} + \mathbf{h} \in D$. If the pdfs are also invariant to the rotation of the spatial references, then isotropy is ensured. Both stationarity and isotropy can be defined exclusively in the horizontal domain D_y .

Definition 3 (Random Field: Stationarity and Isotropy) If the pdfs defining the random field, $p(s(\mathbf{x}_1), s(\mathbf{x}_2), \dots, s(\mathbf{x}_k))$ for all configurations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$ and all k , are invariant to translations in the spatial reference $\mathbf{x} \in D$, then the random field is denoted stationary. If, in addition to stationarity, the pdfs defining the random field are invariant to rotations in the spatial reference $\mathbf{x} \in D$, then the random field is denoted stationary and isotropic. A random field that is not isotropic is denoted as anisotropic.

If the first and second moments of the pdfs specified by the expectations, the variances and the spatial correlation fields are all invariant to translations in the spatial reference $\mathbf{x} \in D$, then the RF is denoted second-order stationary. This entails that

$$\begin{aligned}\{E\{s(\mathbf{x})\} = \mu_s; \mathbf{x} \in D\} \\ \{\text{Var}\{s(\mathbf{x})\} = \sigma_s^2; \mathbf{x} \in D\} \\ \{\text{Corr}\{s(\mathbf{x}), s(\mathbf{x}')\} = \rho_s(\mathbf{x}, \mathbf{x}') = \rho_s(\tau); \tau = |\mathbf{x} - \mathbf{x}'| \in \mathbb{R}; \mathbf{x}, \mathbf{x}' \in D\}.\end{aligned}$$

If, in addition, the RF is isotropic, then the RF is denoted second-order stationary and isotropic. This entails that $\{\text{Corr}\{s(\mathbf{x}), s(\mathbf{x}')\} = \rho_s(\mathbf{x}, \mathbf{x}') = \rho_s(\tau); \tau = |\mathbf{x} - \mathbf{x}'| \in \mathbb{R}; \mathbf{x}, \mathbf{x}' \in D\}$. Both second-order stationarity and isotropy can be defined exclusively in the horizontal domain D_y .

The concept of ergodicity, as defined in Definition 4, is relevant for model parameter inference. Ergodicity of the RF is required for consistent estimators for the model parameters based on observations of one single realisation of the RF.

Definition 4 (Random Field: Ergodicity) If the random field is stationary, and the bivariate pdfs tend towards independence as $|\mathbf{x} - \mathbf{x}'| \rightarrow \infty$, which entails that $p(s(\mathbf{x}), s(\mathbf{x}')) \rightarrow p(s(\mathbf{x}))p(s(\mathbf{x}')) = [p(s(\mathbf{x}))]^2$, then the random field is denoted ergodic.

The three types of spatial variables defined in the previous chapter are assigned probabilistic RF models: continuous RF, event RF and mosaic RF models. The spatial variables are defined over the continuous domain D . Therefore, the corresponding sets of pdfs appear infinite-dimensional, which complicates the probabilistic interpretation. A comprehensive discussion of such RFs can be found in Yaglom (1987), Møller and Waagepetersen (2003), Adler and Taylor (2007) and Lieshout (2019). In this book, we choose to assign the probability models of the spatial variables to a spatial grid L of finite size n over D where $|D| < \infty$. This choice simplifies notation, calculations and interpretations. The spatially discretised RF $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ may be evaluated by infill asymptotic analysis to obtain the related RF $\{s(\mathbf{x}); \mathbf{x} \in D\}$ or by extension asymptotic analysis to validate ergodicity.

To demonstrate the importance of ergodicity, consider a spatially discretised, stationary, ergodic RF $\{s(\mathbf{x}); \mathbf{x} \in L\}$, represented by the n -vector s . The translation-invariant marginal pdf of the RF is $p(s_i)$, and assume that this RF has model parameter θ including, for example, the expectation level, the variance level or the intensity level. The corresponding hierarchical RF s_H is defined as

$$s_H \sim p(s_H) = \int p(s | \theta) p(\theta) d\theta,$$

where the model parameter θ is defined to be a random variable with prior model $p(\theta)$, which is dependent on a set of hyper-parameters. This hierarchical RF model

is not ergodic because the random model parameters prevent independence with increasing spatial variable inter-distance.

The observations \mathbf{d} from one realisation are, in theory, collected from the spatial variable $[\mathbf{s}_H | \boldsymbol{\theta}^t]$, where $\boldsymbol{\theta}^t$ contains the true model parameter values. Consequently, only the true parameter values can be inferred from \mathbf{d} . The non-observable hyper-parameters of the prior model $p(\boldsymbol{\theta})$, which defines the hierarchical model, cannot be elicited from the observations \mathbf{d} . These hyper-parameters must be user-specified or elicited from a super-population of comparable realisations.

Assume that the hyper-parameters are assigned specific values and that a single realisation is to be generated from the hierarchical model. This simulation is made sequentially. First generate $\boldsymbol{\theta}^s \sim p(\boldsymbol{\theta})$ and thereafter generate $\mathbf{s}_H^s \sim p(\mathbf{s} | \boldsymbol{\theta}^s)$. The resulting realisation \mathbf{s}_H^s is identical to a realisation of \mathbf{s} with model parameter value $\boldsymbol{\theta}^s$, and the within-realisation variability, termed spatial heterogeneity, will not capture the uncertainty in the model parameter $\boldsymbol{\theta}$. The corresponding spatial histogram $\text{Hist}\{\mathbf{s}_H^s\}$ is, in the extension limit $n \rightarrow \infty$, not identical to the translation-invariant marginal pdf $p(s_{Hi})$ of the hierarchical model. The former appears with less variability because of the lack of variability in $\boldsymbol{\theta}$.

If an ensemble of realisations $\{\mathbf{s}_H^s; s = 1, 2, \dots, n_s\}$ from the hierarchical model is generated, the across-realisation variability captures the uncertainty in the model parameter $\boldsymbol{\theta}$. Consequently, in an arbitrary grid node j , the location histogram $\text{Hist}\{s_{Hj}^s; s = 1, 2, \dots, n_s\}$ will, as $n_s \rightarrow \infty$, converge to be identical to the translation-invariant marginal pdf $p(s_{Hi})$.

Ergodicity in the RF model ensures that spatial heterogeneity and marginal variability are identical. This model consistency makes model parameter inference from one realisation more reliable and the interpretation of the results easier. Hence, the use of ergodic spatial RF models is recommended.

The remainder of this section presents probabilistic RF models for continuous, event and mosaic spatial variables.

4.1 Continuous Random Fields

Consider a continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}; r(\mathbf{x}) \in \mathbb{R}$. This section defines two classes of RF models for such spatial variables: the Gaussian RF and the hierarchical Gaussian RF models. Later, these two classes of RF models are demonstrated to be conjugate classes with respect to certain classes of likelihood models.

4.1.1 Gaussian RF Models

Consider a continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$, with the associated Gaussian RF model, defined in Definition 5.

Definition 5 (Gaussian Random Field) Consider the continuous RF $\{r(\mathbf{x}); \mathbf{x} \in D\}; r(\mathbf{x}) \in \mathbb{R}$ with an expectation field $\{\mu(\mathbf{x}); \mathbf{x} \in D\}; \mu(\mathbf{x}) \in \mathbb{R}$, and a variance field $\{\sigma^2(\mathbf{x}); \mathbf{x} \in D\}; \sigma^2(\mathbf{x}) \in \mathbb{R}_+$. Assume that the correlation field $\{\rho(\mathbf{x}, \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}$ satisfies $\rho(\mathbf{x}, \mathbf{x}) = 1.0$ and that $\rho(\mathbf{x}, \mathbf{x}') \in \mathbb{R}_{[-1, 1]}$ is a non-negative definite field.

If $(r(\mathbf{x}_1), r(\mathbf{x}_2), \dots, r(\mathbf{x}_k))$ is Gaussian, i.e. $p(\mathbf{r}) = \phi_k(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, with an expectation k -vector $\boldsymbol{\mu}$ and a $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}$ defined from the parameter fields above for all configurations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) \in D^k$ and $k \in \mathbb{N}_+$, then $\{r(\mathbf{x}); \mathbf{x} \in D\}$ is denoted a Gaussian RF with model parameters as specified above.

The non-negative definiteness of the correlation field ensures that the corresponding covariance matrices are also non-negative definite. Figure 4.1 schematically presents the definition of a stationary Gaussian RF. In display (a), one realisation of a Gaussian RF is presented, and four locations are marked $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$. In display (b) the three plots present the univariate Gaussian pdfs $p(r(\mathbf{x}_i))$ along the axes, and the bivariate Gaussian pdfs $p(r(\mathbf{x}_1), r(\mathbf{x}_i)); i = 2, 3, 4$ for all pairs of locations containing \mathbf{x}_1 . Note that the marginal pdfs are identical, and the correlation decreases with increasing spatial variable inter-distance for the joint pdfs. In display (c), the chosen spatial correlation function for the stationary Gaussian RF is presented with the spatial variable inter-distances $\tau_{1i} = |\mathbf{x}_1 - \mathbf{x}_i|$; $i = 2, 3, 4$ marked. The correlations specified by this function decrease with

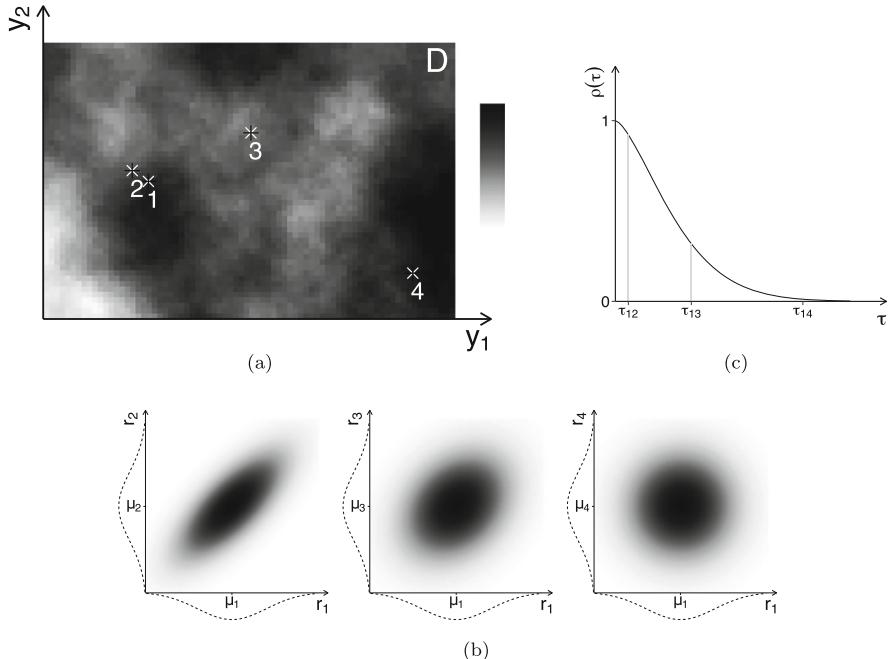


Fig. 4.1 Gaussian RF model. Schematic presentation: (a) realisation from a Gaussian RF model with four observations, (b) joint pdfs of pair of observations and (c) spatial correlation function with inter-observation correlations marked

increasing inter-distance. For additional reading on Gaussian RF models, see Chiles and Delfiner (2012) and Cressie and Wikle (2011).

Consider the continuous spatial variable spatially discretised on the grid $\{r(\mathbf{x}); \mathbf{x} \in L\}$ and represented by the n -vector $\mathbf{r} \in \mathbb{R}^n$. Then, the Gaussian model is defined from the Gaussian RF model on $\mathbf{r} \in \mathbb{R}^n$ as

$$\mathbf{r} \sim p(\mathbf{r}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \quad (4.1)$$

where the expectation n -vector $\boldsymbol{\mu}_r$ defines $\{\mu_r(\mathbf{x}); \mathbf{x} \in L\}$. The $(n \times n)$ covariance matrix can be decomposed as $\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_r^\sigma \boldsymbol{\Sigma}_r^\rho \boldsymbol{\Sigma}_r^\sigma$, where the diagonal standard deviation matrix $\boldsymbol{\Sigma}_r^\sigma$, of dimension $(n \times n)$, defines $\{\sigma_r^2(\mathbf{x}); \mathbf{x} \in L\}$. Lastly, the non-negative definite correlation matrix $\boldsymbol{\Sigma}_r^\rho$, of dimension $(n \times n)$, defines $\{\rho_r(\mathbf{x}, \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in L\}$. This Gaussian model class provides a natural spatial discretisation of the Gaussian RF model.

In Sects. 6.1 and 7.1, it is demonstrated that this Gaussian model class is a conjugate class with respect to certain likelihood models defined in Sect. 5.1.

The infill limit of the spatially discretised continuous spatial variable is $\{r(\mathbf{x}); \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{r(\mathbf{x}); \mathbf{x} \in D\}$. Consequently, the infill asymptotic limit of a Gaussian model is a Gaussian RF model with parameters obtained by the infill limit for the corresponding discretised expectation, variance and correlation fields.

4.1.2 Hierarchical Gaussian RF Models

The spatially discretised Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in L\}$, represented by the Gaussian n -vector $\mathbf{r} \in \mathbb{R}^n$, parametrised by the expectation n -vector $\boldsymbol{\mu}_r \in \mathbb{R}^n$ and the non-negative definite $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}_r \in \mathbb{R}^{n \times n}$, can be extended into a hierarchical Gaussian model. This hierarchical model considers the model parameters $[\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r]$ as random variables, and the marginal hierarchical Gaussian model $\{r_H(\mathbf{x}); \mathbf{x} \in L\}$ is represented by the n -vector $\mathbf{r}_H \in \mathbb{R}^n$. The objective is to define the marginal pdf for $p(\mathbf{r}_H)$, where the effect of random parameters is marginalised out. First, we develop an expression for the pdf of \mathbf{r}_H when $\boldsymbol{\Sigma}_r$ is fixed; that is, only the effect of $\boldsymbol{\mu}_r$ is marginalised, i.e. $p(\mathbf{r}_H | \boldsymbol{\Sigma}_r) = \int p(\mathbf{r}_H, \boldsymbol{\mu}_r | \boldsymbol{\Sigma}_r) d\boldsymbol{\mu}_r$. Second, we develop an expression for the marginal pdf $p(\mathbf{r}_H) = \int p(\mathbf{r}_H | \boldsymbol{\Sigma}_r) p(\boldsymbol{\Sigma}_r) d\boldsymbol{\Sigma}_r$ based on the former.

Assume that the prior expectation n -vector $\boldsymbol{\mu}_r$ is Gaussian with known hyper-parameters $\boldsymbol{\mu}_\mu \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}_\mu \in \mathbb{R}^{n \times n}$. Given $\boldsymbol{\mu}_r$, \mathbf{r}_H is assumed to be Gaussian with an unknown auxiliary covariance matrix $\boldsymbol{\Sigma}_{r|\mu} \in \mathbb{R}^{n \times n}$. That is,

$$\begin{aligned} [\mathbf{r}_H | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{r|\mu}] &\sim p(\mathbf{r}_H | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{r|\mu}) = \phi_n(\mathbf{r}_H; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{r|\mu}) \\ [\boldsymbol{\mu}_r | \boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu] &\sim p(\boldsymbol{\mu}_r | \boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu) = \phi_n(\boldsymbol{\mu}_r; \boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu). \end{aligned}$$

Because $[\mathbf{r}_H | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{r|\mu}]$ can be interpreted as $\mathbf{r}_H = \boldsymbol{\mu}_r + \boldsymbol{\varepsilon}_{r|\mu}$, where the latter is a centred Gaussian variable with a fixed covariance matrix $\boldsymbol{\Sigma}_{r|\mu}$ that is independent of $\boldsymbol{\mu}_r$, the pdf of $[\mathbf{r}_H | \boldsymbol{\Sigma}_r]$ is a Gaussian pdf,

$$\begin{aligned} [\mathbf{r}_H | \boldsymbol{\Sigma}_r] &\sim p(\mathbf{r}_H | \boldsymbol{\Sigma}_r) = \int_{\mathbb{R}^n} p(\mathbf{r}_H | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{r|\mu}) p(\boldsymbol{\mu}_r | \boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu) d\boldsymbol{\mu}_r \\ &= \phi_n(\mathbf{r}_H; \boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_r). \end{aligned}$$

Here, the expectation $\boldsymbol{\mu}_\mu$ and covariance $\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_\mu + \boldsymbol{\Sigma}_{r|\mu}$ are computed using the laws of double expectation and total covariance, respectively.

Next, we assume $\boldsymbol{\Sigma}_r$ to be a random variable with an inverse-Wishart prior pdf with known hyper-parameters $\boldsymbol{\Omega}_\Sigma \in \mathbb{R}^{n \times n}$ and $v_\Sigma \in \mathbb{R}_+$

$$\begin{aligned} \boldsymbol{\Sigma}_r &\sim p(\boldsymbol{\Sigma}_r) = 2^{-v_\Sigma n/2} \left[\Gamma_n \left(\frac{v_\Sigma}{2} \right) \right]^{-1} |\boldsymbol{\Omega}_\Sigma|^{v_\Sigma/2} \\ &\quad \times |\boldsymbol{\Sigma}_r|^{-(v_\Sigma+n+1)/2} \exp \left(-\frac{1}{2} \text{Tr} \left\{ \boldsymbol{\Omega}_\Sigma \boldsymbol{\Sigma}_r^{-1} \right\} \right) \end{aligned}$$

where $\Gamma_n(\cdot)$ is the n -variate gamma function. Here, $\boldsymbol{\Omega}_\Sigma$ is a non-negative definite spatial scale-coupling matrix, and v_Σ is the degrees of freedom parameter. Then the hierarchical Gaussian model $p(\mathbf{r}_H)$ for $\mathbf{r}_H \in \mathbb{R}^n$ is obtained through marginalisation,

$$\begin{aligned} \mathbf{r}_H &\sim p(\mathbf{r}_H) = \int_{\mathbb{R}^{n \times n}} p(\mathbf{r}_H | \boldsymbol{\Sigma}_r) p(\boldsymbol{\Sigma}_r) d\boldsymbol{\Sigma}_r \\ &= \left[\Gamma \left(\frac{v_\Sigma}{2} \right) \right]^{-1} [v_\Sigma \pi]^{-n/2} \Gamma \left(\frac{v_\Sigma + n}{2} \right) \\ &\quad \times |\boldsymbol{\Omega}_\Sigma|^{-1/2} \times \left[1 + v_\Sigma^{-1} (\mathbf{r}_H - \boldsymbol{\mu}_\mu)^T \boldsymbol{\Omega}_\Sigma^{-1} (\mathbf{r}_H - \boldsymbol{\mu}_\mu) \right]^{-(v_\Sigma+n)/2}. \end{aligned} \tag{4.2}$$

The hyper-parameter matrix $\boldsymbol{\Omega}_\Sigma$ must capture the joint effect of the spatial coupling represented in $\boldsymbol{\Sigma}_\mu$ and $\boldsymbol{\Sigma}_{r|\mu}$. This hierarchical model is an n -variate T-distribution with v_Σ degrees of freedom, parametrised by $[\boldsymbol{\mu}_\mu, \boldsymbol{\Omega}_\Sigma, v_\Sigma]$, which are defined by the hyper-parameters (Kotz and Nadarajah, 2004). Recall that the T-distribution is bell-shaped like the Gaussian pdf but with heavier tails. A particular case arises when there is one degree of freedom: the T-distribution becomes the Cauchy pdf, for which neither the variance nor the expectation exists. Recall that the T-distribution, and therefore the hierarchical Gaussian model, tends towards the regular Gaussian model when $v_\Sigma \rightarrow \infty$. The marginal pdf $p(r_{Hi})$ at an arbitrary grid node i is T-distributed with the scale $\omega_{\Sigma i}$ being the diagonal entry in $\boldsymbol{\Omega}_\Sigma$ and v_Σ being the degrees of freedom. Hence, the marginal pdf is parametrised by $[\mu_{\mu i}, \omega_{\Sigma i}, v_\Sigma]$.

In Sects. 6.3 and 7.1, it is demonstrated that this hierarchical Gaussian model class is also a conjugate class with respect to specific likelihood models defined in Sect. 5.1.

4.2 Event Random Fields

Consider an event spatial variable $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}; k(\mathbf{x}) \in \mathbb{B}$, which precisely specifies the locations of the events in D . This section defines two classes of RF models for such spatial variables: the Poisson RF and the hierarchical Poisson RF models. Later, these two classes of RF models are demonstrated to be conjugate classes for certain classes of likelihood models.

4.2.1 Poisson RF Models

Consider the spatial event variable $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$. The corresponding Poisson RF model is defined in Definition 6. Note that this definition of a Poisson RF deviates slightly from the traditional definition, as discussed in Møller and Waagepetersen (2003) because it includes a base event set with predefined locations within the reference domain D . This base event set establishes a predefined baseline, while the Poisson component introduces a layer of random variability. If the base event set is empty, Definition 6 aligns with the traditional one. We introduce this base event set extension in the definition of the Poisson RF model class to achieve conjugate properties.

Definition 6 (Poisson Random Field) Consider the event RF $\{k(\mathbf{x}); \mathbf{x} \in D\}; k(\mathbf{x}) \in \mathbb{B}$, with an intensity field $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$ and a base event set $\mathbb{E} : \{\mathbf{x}_1^e, \mathbf{x}_2^e, \dots, \mathbf{x}_{k^0}^e\}; \mathbf{x}_i^e \in D$. Let the event count variable $k_A \in \mathbb{N}_+$ be the number of events occurring in the domain $A \subset D$.

For $k_A \geq k_A^0$, let

$$k_A \sim p(k_A) = \frac{\lambda_A^{k_A - k_A^0}}{(k_A - k_A^0)!} \times \exp(-\lambda_A)$$

with $k_A^0 = \sum_{\mathbf{x} \in \mathbb{E}} I(\mathbf{x} \in A)$ and $\lambda_A = \int_{A \setminus \mathbb{E} \cap A} \lambda(\mathbf{x}) d\mathbf{x} < \infty$ for all sub-domains $A \subset D$. Moreover, let $[k_A, k_B]$ be independent for all sub-domains $A, B \subset D$ where $A \cap B = \emptyset$. Then, the set $\{k(\mathbf{x}); \mathbf{x} \in D\}$ is denoted a Poisson RF with an intensity field $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}$ and base event set \mathbb{E} .

In Fig. 4.2, the definition of a stationary Poisson RF, with constant intensity λ and empty base event set \mathbb{E} , is schematically presented. In display (a), one realisation of the Poisson RF is presented, and a sub-domain A is marked. The event count

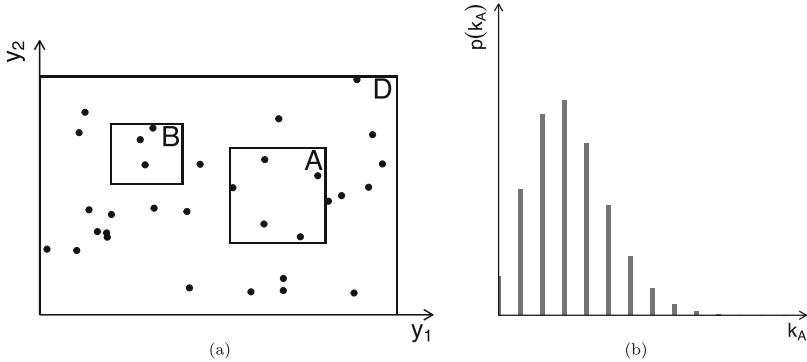


Fig. 4.2 Poisson RF model. Schematic presentation: (a) realisation from a Poisson RF model with two sub-domains and (b) Poisson pdf for an event count variable in given sub-domain

inside the sub-domain is k_A . A sub-domain B , disjoint from the sub-domain A , is also marked. Display (b) contains a presentation of the pdf $p(k_A)$ of the associated Poisson-distributed event count variable k_A . Note that the two event count variables (k_A, k_B) are independent because $A \cap B = \emptyset$. For alternative reading on Poisson RF models, see Møller and Waagepetersen (2003) and Lieshout (2019).

For a Poisson RF model on a finite domain D , i.e. $|D| < \infty$, the Poisson RF $\{k(\mathbf{x}); \mathbf{x} \in D\}$ may be rephrased by defining the event-location set $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}\}; \mathbf{x}_i \in D; k_D \in \mathbb{N}_{[k^0, \infty]}$. This set of event locations contains a random number of entries k_D . The associated pdf can be decomposed as $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}) = p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \mid k_D = k) \times p(k_D = k \mid k_D \geq k^0)$, which is simpler to interpret. This rephrasing leads to an alternative definition of a Poisson RF, as given in Definition 7.

Definition 7 (Poisson Random Field: Alternative) Consider the event-location set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}; \mathbf{x}_i \in D, |D| < \infty$, where $k \in \mathbb{N}_+$ is the random number of elements. Define the intensity field $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$, and base event set $\mathbb{E} : \{\mathbf{x}_1^e, \mathbf{x}_2^e, \dots, \mathbf{x}_{k^0}^e\}; \mathbf{x}_i^e \in D$. The associated pdf is $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) = p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k'} \mid k = k')p(k = k')$.

Let

$$k \sim p(k) = \frac{[\lambda_D]^k}{k!} \exp(-\lambda_D),$$

where $\lambda_D = \int_{D \setminus \mathbb{E}} \lambda(\mathbf{x}) d\mathbf{x} < \infty$ and

$$[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \mid k] \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \mid k) = \prod_{i=1}^k \frac{\lambda(\mathbf{x}_i)}{\lambda_D} = \frac{1}{[\lambda_D]^k} \prod_{i=1}^k \lambda(\mathbf{x}_i).$$

Thus, for $\mathbf{x}_i \in D$ and $k \in \mathbb{N}_{\oplus}$,

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) = \frac{1}{[\lambda_D]^k} \prod_{i=1}^k \lambda(\mathbf{x}_i) \times \frac{[\lambda_D]^k}{k!} \exp(-\lambda_D) = \frac{\prod_{i=1}^k \lambda(\mathbf{x}_i)}{k!} \times \exp(-\lambda_D).$$

Then $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \cup \mathbb{E}$ is the event-location set for a Poisson RF $\{k(\mathbf{x}); \mathbf{x} \in D\}$ with intensity field $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}$ and base event set \mathbb{E} .

Note that given the event count k_D , the event locations are independent, each with location pdf $p(\mathbf{x}) = \lambda(\mathbf{x})/\lambda_D; \mathbf{x} \in D$.

Consider the spatially discretised event spatial variable $\{k_{\Delta}(\mathbf{x}); \mathbf{x} \in L\}$ represented by the event count n -vector $\mathbf{k} \in \mathbb{N}_{\oplus}^n$ and Definition 6. The spatially discretised Poisson count RF model $\mathbf{k} \in \mathbb{N}_{\oplus}^n; \mathbf{k} \geq \mathbf{k}^o$ is defined as

$$\mathbf{k} \sim p(\mathbf{k}) = \prod_{i=1}^n \frac{[\lambda_{ki} \Delta_n]^{k_i - k_i^0}}{(k_i - k_i^0)!} \exp(-\lambda_{ki} \Delta_n). \quad (4.3)$$

The model parameter is the n -vector $\lambda_k = (\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kn})^T$, where $\lambda_{ki} = \Delta_n^{-1} \int_{\Delta_{ni}} \lambda(\mathbf{x}) d\mathbf{x}$ is the average event intensity in grid unit Δ_{ni} . The n -vector $\mathbf{k}^o = (k_1^o, k_2^o, \dots, k_n^o)^T$, where $k_i^o = \sum_{\mathbf{x} \in \mathbb{E}} I(\mathbf{x} \in \Delta_{ni})$ contains the number of base events in grid unit Δ_{ni} . Thus, the event count in the grid node i , representing the number of events in the grid unit centred at grid node i , Δ_{ni} , follows a Poisson pdf with parameter $\lambda_{ki} \Delta_n$, which is bounded below by k_i^o . Recall that $\Delta_{ni}; i = 1, 2, \dots, n$ constitutes a partition of D . Therefore, the event counts in the grid units are independent according to the definition of the Poisson RF. This Poisson count model class constitutes a natural spatial discretisation of the Poisson RF model.

In Sects. 6.4 and 7.2, it is demonstrated that this Poisson count model class is a conjugate class with respect to certain likelihood models, defined in Sect. 5.2.

Consider the infill asymptotic limit of the discretised spatial variable $\{k_{\Delta}(\mathbf{x}); \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{k(\mathbf{x}); \mathbf{x} \in D\}$ as $\Delta_n \rightarrow 0, n \rightarrow \infty$ and λ_k such that $\sum_{i=1}^n \lambda_{ki} \Delta_n$ is constant. Note that the n -vector \mathbf{k}^0 tends towards a binary vector as $n \rightarrow \infty$. It can then be demonstrated that $\{k(\mathbf{x}); \mathbf{x} \in D\}$ is a Poisson RF with the intensity field defined by $\{\lambda_k(\mathbf{x}); \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{\lambda_k(\mathbf{x}); \mathbf{x} \in D\}$ and $\{\|\mathbf{x}\|_{\mathbf{x}}^0 > 0; \mathbf{x} \in L\} \xrightarrow{\text{infill}} \mathbb{E}$. Thus, the constant $\sum_{i=1}^n \lambda_{ki} \Delta_n \xrightarrow{\text{infill}} \int_D \lambda_k(\mathbf{x}) d\mathbf{x} = \lambda_D$.

The approximate Poisson event-location set \mathbb{X}_D^n is defined by the spatially discretised Poisson count RF model represented by the n -vector $\mathbf{k} \in \mathbb{N}_{\oplus}^n$. Define the approximate event locations in each grid unit Δ_{ni} to be uniformly distributed. Then, for each grid unit $i = 1, 2, \dots, n$ the set of event locations in Δ_{ni} given the event count $k_i, \mathbf{x}_i : \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik_i}\}$ with $\mathbf{x}_{ij} \in D$ is

$$[\mathbf{x}_i | k_i] \sim p(\mathbf{x}_i | k_i) = \prod_{j=1}^{k_i} I(\mathbf{x}_{ij} \in \Delta_{ni}) \frac{1}{\Delta_n}. \quad (4.4)$$

The approximate Poisson event-location set given the Poisson count \mathbf{k} is

$$\begin{aligned} [\mathbb{X}_D^n | \mathbf{k}] &= \cup_{i=1}^n [\mathbf{x}_i | k_i] \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{k}) \\ &= \prod_{i=1}^n p(\mathbf{x}_i | k_i) = \prod_{i=1}^n \prod_{j=1}^{k_i} I(\mathbf{x}_{ij} \in \Delta_{ni}) \frac{1}{\Delta_n}. \end{aligned} \quad (4.5)$$

Hence, the approximate Poisson event-location model is

$$\begin{aligned} \mathbb{X}_D^n &\sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{\mathbf{k} \in \mathbb{N}_+^n} p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{k}) p(\mathbf{k}) \\ &= \prod_{i=1}^n \sum_{k_i \in \mathbb{N}_+} p(\mathbf{x}_i | k_i) p(k_i), \end{aligned} \quad (4.6)$$

where the first term is defined in Expression (4.5), while the latter term is defined in Expression (4.3). One realisation of the approximate Poisson event-location model can be generated sequentially. First, generate a realisation of the Poisson count model $\mathbf{k}^s \sim p(\mathbf{k})$. Next, the event locations are generated conditioned on \mathbf{k}^s , as $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^s \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{k}^s)$. Then $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^s \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ is a realisation of the approximate Poisson event-location model.

4.2.2 Hierarchical Poisson RF Models

The spatially discretised Poisson count RF $\{k_\Delta(\mathbf{x}); \mathbf{x} \in L\}$ is represented by the Poisson count n -vector $\mathbf{k} \in \mathbb{N}_+^n$. The model parameters are the event intensity n -vector $\lambda_k \in \mathbb{R}_+^n$ and the base event n -vector $\mathbf{k}^o \in \mathbb{N}_+^n$. This model can be extended into a hierarchical Poisson count model with the model parameter λ_k taken as a random variable. The marginal hierarchical Poisson count model $\{k_H(\mathbf{x}); \mathbf{x} \in L\}$ is then represented by the n -vector $\mathbf{k}_H \in \mathbb{N}_+^n$. The model parameter is re-parametrised to $\lambda_k = \lambda_k \beta_\lambda$ where $\lambda_k \in \mathbb{R}_+$ is an intensity-level random variable and $\beta_\lambda \in \mathbb{R}_+^n$ is a user-specified spatial relative-intensity n -vector. The prior model for λ_k is a gamma pdf with hyper-parameters $v_\lambda \in \mathbb{R}_+$ and $\xi_\lambda \in \mathbb{R}_+$. The hierarchical Poisson count model is then defined for $\mathbf{k}_H \in \mathbb{N}_+^n; \mathbf{k}_H \geq \mathbf{k}_n^o$ as

$$\begin{aligned} [\mathbf{k}_H | \lambda_k] &\sim p(\mathbf{k}_H | \lambda_k) = \prod_{i=1}^n \frac{(\lambda_k \beta_{\lambda i} \Delta_n)^{k_{Hi} - k_{ni}^0}}{(k_{Hi} - k_{ni}^0)!} \exp(-\lambda_k \beta_{\lambda i} \Delta_n) \\ \lambda_k &\sim p(\lambda_k) = [\Gamma(v_\lambda)]^{-1} \xi_\lambda^{v_\lambda} \lambda_k^{v_\lambda - 1} \times \exp(-\xi_\lambda \lambda_k). \end{aligned}$$

By marginalisation, the hierarchical Poisson count model becomes

$$\begin{aligned}\mathbf{k}_H \sim p(\mathbf{k}_H) &= \int_{\mathbb{R}_{+}} p(\mathbf{k}_H \mid \lambda_k) p(\lambda_k) d\lambda_k \\ &= [\Gamma(v_\lambda)]^{-1} \Gamma \left(\sum_{i=1}^n (k_{Hi} - k_{ni}^0) + v_\lambda \right) \kappa_\lambda^{v_\lambda} \prod_{i=1}^n \frac{[\omega_{\lambda,i}(1 - \kappa_\lambda)]^{k_{Hi} - k_{ni}^0}}{(k_{Hi} - k_{ni}^0)!},\end{aligned}\quad (4.7)$$

where $\kappa_\lambda = [1 + \xi_\lambda^{-1} \Delta_n(\boldsymbol{\beta}_\lambda^T \mathbf{i}_n)]^{-1} < 1$ and the weight n -vector is $\boldsymbol{\omega}_\lambda = \boldsymbol{\beta}_\lambda [\boldsymbol{\beta}_\lambda^T \mathbf{i}_n]^{-1}$. This hierarchical Poisson count model, for integer-valued v_λ , is an n -variate negative-multinomial pdf (Johnson et al., 1997). The model parameters are $(\kappa_\lambda, \boldsymbol{\omega}_\lambda, v_\lambda)$, which are defined by the hyper-parameters of the model. The negative-multinomial model has an urn interpretation, with $n + 1$ types of balls in the urn, each type with draw probability (p_0, p_1, \dots, p_n) . The numbers of balls of each type while drawing are registered (k_0, k_1, \dots, k_n) . If the drawing is completed when $k_0 = k_T$, the registered number of balls is negative multinomially distributed. In the hierarchical Poisson model, the entries in the n -vector $\boldsymbol{\omega}_\lambda$ are related to the probabilities, while the parameter v_λ is related to the stopping criterion k_T .

The marginal pdf, $p(k_{Hi})$, for arbitrary grid node i , is a negative-binomial pdf parametrised by $[\kappa_\lambda, \omega_{\lambda,i}, v_\lambda]$. Recall that whenever $v_\lambda \rightarrow \infty$ and $\xi_\lambda \rightarrow \infty$ with $v_\lambda/\xi_\lambda = \lambda_k$, the hierarchical Poisson model tends towards the regular Poisson model.

In Sects. 6.6 and 7.2, it is demonstrated that this hierarchical Poisson count model class is also a conjugate class with respect to certain likelihood models, defined in Sect. 5.2.

A full hierarchical Poisson RF model is obtained if in addition the n -vector $[\boldsymbol{\beta}_\lambda | \lambda_k]$ is specified as random with a prior Dirichlet model. The associated hierarchical Poisson RF model class exhibits conjugate properties with respect to a particular likelihood model class having a constant misclassification probability across the spatial domain.

4.3 Mosaic Random Fields

Consider a mosaic spatial variable $\{l(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}; l(\mathbf{x}) \in \mathbb{L}$. This section contains a definition of one class of RF models for such spatial variables, the Markov RF model. Later, this class of RF models is demonstrated to be a conjugate class with respect to one certain class of likelihood models. There is no simple natural extension of this Markov RF model into a corresponding hierarchical version that can be easily assessed.

4.3.1 Markov RF Models

The concept of a clique system on the spatial grid is a critical factor in Markov RF models, and it is defined in Definition 8. We consider clique systems consisting of the maximal clique units, where each clique is not a subset of any other clique. The simplest example of a clique design on a two-dimensional grid L is the two nearest nodes in each direction. The corresponding clique system contains all pairs of nearest nodes in the grid. The next level of clique design in a two-dimensional grid consists of a (2×2) -node clique unit. The corresponding clique system contains all unique sets of four nearest nodes in the grid. Figure 4.3 contains one example of a clique design of four-nearest-node in display (a).

Definition 8 (Clique System) Consider a spatial grid system $L \subset D$. A clique system $\mathbf{c}_L : \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_c}\}$ consists of clique units $\mathbf{c}_i \subset L$ and represents any fully connected subset of L .

Consider a mosaic spatial variable $\{l(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the spatially discretised spatial variable $\{l(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. The associated Markov RF model is defined in Definition 9. For alternative reading on Markov RF models, see Besag (1974) and Winkler (2006).

Definition 9 (Markov Random Field) Consider the mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L\}$ with $l(\mathbf{x}) \in \mathbb{L}$, defined on the grid L covering D , represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. Define a clique system $\mathbf{c}_L : \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_c}\}$ with cliques $\mathbf{c}_i \subset L$. If

$$\mathbf{l} \sim p(\mathbf{l}) = \text{const} \times \prod_{i=1}^n v_{0i}(l_i) \times \prod_{\mathbf{c} \in \mathbf{c}_L} v_{1\mathbf{c}}(l_j; j \in \mathbf{c}),$$

where $v_{0i}(\cdot) \in \mathbb{R}_+$, $v_{1\mathbf{c}}(\cdot) \in \mathbb{R}_+$ and const is the normalising constant, then $\{l(\mathbf{x}); \mathbf{x} \in L\}$ is a Markov RF on the grid L with Gibbs formulation defined on clique system \mathbf{c}_L .

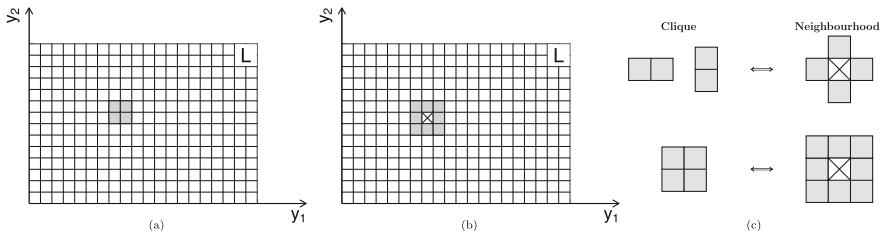


Fig. 4.3 Markov RF model. Schematic presentation: (a) grid discretised domain with one four-nearest-node clique unit, (b) grid discretised domain with the corresponding eight-nearest-node neighbourhood unit for one given grid node and (c) examples of corresponding clique and neighbourhood designs

The parameters $v_{0i}(\cdot); i = 1, 2, \dots, n$ are related to the label proportions in each grid unit, whereas the parameters $v_{1c}(\cdot); c \in \mathbf{c}_L$ specify the spatial coupling of the labels in neighbouring nodes.

The Hammersley–Clifford theorem, given in Theorem 1, ensures the equivalence between a Gibbs formulation using the clique system and a Markov formulation using the corresponding neighbourhood system for a Markov RF model. A Markov RF is generally specified through the n -variate Gibbs formulation, as it is challenging to ensure a consistent and valid set of n one-dimensional pdfs for the Markov formulation. We refer the reader to Besag (1974) for further details.

Theorem 1 (Hammersley–Clifford) Consider a spatial mosaic variable $\{l(\mathbf{x}); \mathbf{x} \in L\}$ with $l(\mathbf{x}) \in \mathbb{L}$ represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. The Markov RF model is defined in the Gibbs formulation by clique system \mathbf{c}_L . There exists a corresponding Markov formulation of the Markov RF model based on the neighbourhood system $\mathbf{n}_L : \{\mathbf{n}_x; \mathbf{x} \in L\}$, where $\mathbf{n}_x \subset L \setminus \mathbf{x}$, such that for all $i = 1, 2, \dots, n$

$$[l_i | \mathbf{l}_{-i}] \sim p(l_i | \mathbf{l}_{-i}) = \text{const} \times v_{0i}(l_i) \times \omega_i(l_i | l_j; j \in \mathbf{n}_i),$$

where $v_{0i}(\cdot) \in \mathbb{R}_+$, $\omega_i(\cdot) \in \mathbb{R}_+$, and const is the normalising constant. This set of expressions is termed the Markov formulation of the Markov RF on grid L with neighbourhood system \mathbf{n}_L .

Consider the mosaic spatial variable discretised to the grid $\{l(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. Then, the Gibbs formulation of the Markov RF model is defined as

$$\begin{aligned} \mathbf{l} \sim p(\mathbf{l}) &= \text{const} \times \prod_{i=1}^n v_{0i}(l_i) \times \prod_{c \in \mathbf{c}_L} v_{1c}(l_j; j \in c) \\ \text{const} &= \left[\sum_{\mathbf{l}' \in \mathbb{L}^n} \prod_{i=1}^n v_{0i}(l'_i) \times \prod_{c \in \mathbf{c}_L} v_{1c}(l'_j; j \in c) \right]^{-1}. \end{aligned} \quad (4.8)$$

The model parameters are the proportion functions $v_{0i}(l_i) \in \mathbb{R}_+; l_i \in \mathbb{L}; i = 1, 2, \dots, n$, the clique system $\mathbf{c}_L : \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_c}\}$ and the functional forms of $v_{1c}(\cdot) \in \mathbb{R}_+; c \in \mathbf{c}_L$. The clique designs are typically defined as spatially convex sets of nodes within the grid. The normalising constant is defined by summation over $\mathbf{l} \in \mathbb{L}^n$, which in practice is unfeasible to compute.

The Hammersley–Clifford theorem states that for the Markov RF model, there exists an equivalent parametrisation of the Gibbs formulation, known as the Markov

formulation for $i = 1, 2, \dots, n$

$$[l_i | \mathbf{l}_{-i}] \sim p(l_i | \mathbf{l}_{-i}) = \text{const} \times v_{0i}(l_i) \times \omega_i(l_i | l_j; j \in \mathbf{n}_i) \quad (4.9)$$

$$\text{const} = \left[\sum_{l'_i \in \mathbb{L}} v_{0i}(l'_i) \times \omega_i(l'_i | l_j; j \in \mathbf{n}_i) \right]^{-1}.$$

The model parameters are the proportion functions $v_{0i}(l_i) \in \mathbb{R}_+; l_i \in \mathbb{L}; i = 1, 2, \dots, n$, the neighbourhood system $\mathbf{n}_L : \{\mathbf{n}_x; x \in L\}$ and the Markov interaction functions $\omega_i(l_i | \cdot) \in \mathbb{R}_+; l_i \in \mathbb{L}; i = 1, 2, \dots, n$. Each neighbourhood design is defined using the clique system in the Gibbs formulation and is typically also a spatially convex set of nodes, excluding the reference location. The neighbourhood design appears to be somewhat larger than the clique designs. The normalising constants are defined by sums only over $l_i \in \mathbb{L}$ and can usually be efficiently computed.

Figure 4.3 contains examples of corresponding clique and neighbourhood designs. In display (a), a four-nearest-node clique unit is presented, whereas the corresponding eight-nearest-node neighbourhood unit for one given node is presented in display (b). In addition, the simplest two-nearest-node clique and corresponding four-nearest-node neighbourhood designs are presented in display (c).

The Gibbs formulation consists of one n -dimensional pdf, whereas the Markov formulation consists of n univariate conditional pdfs. Moreover, the normalising constant in the former is usually very computationally demanding, whereas the latter's constants can easily be computed.

In Sects. 6.7 and 7.3, it is demonstrated that this Markov model class is a conjugate class with respect to specific likelihood models, defined in Sect. 5.3.

The traditional three-dimensional Markov RF model can be recast as a profile Markov RF, which is a multivariate, two-dimensional Markov RF. The spatial reference is assigned in the horizontal plane L_y , whereas the multivariate variable is represented along the vertical reference L_z . The profile Markov RF is denoted $\{\mathbf{l}(\mathbf{y}); \mathbf{y} \in L_y\}$, where $\mathbf{l}(\mathbf{y}) \in \mathbb{L}^{n_z}$. Note that for each node $\mathbf{y} \in L_y$, the n_z -vector \mathbf{l}_y will be a Markov random profile (RP), which is one-dimensional. As in Sect. 10.8, it can be shown that any Markov RP can be expressed on sequential form as a Markov random chain (RC).

The profile Markov RF has the following Markov formulation for each $\mathbf{y} \in L_y$:

$$[\mathbf{l}_y | \mathbf{l}_{-y}] \sim p(\mathbf{l}_y | \mathbf{l}_{-y}) = p(\mathbf{l}_y | \mathbf{l}_u; \mathbf{u} \in \mathbf{n}_y), \quad (4.10)$$

where the profile neighbourhood \mathbf{n}_y is defined by the neighbourhood system \mathbf{n}_L of the original Markov RF as $\mathbf{n}_y = \cup_{i \in L_z} \mathbf{n}_{(y,i)}$. The conditional Markov RP $[\mathbf{l}_y | \mathbf{l}_u; \mathbf{u} \in \mathbf{n}_y]$ can also be recast as a Markov RC in sequential, factorial form, which is suitable for the efficient sequential simulation algorithm, as defined in Sect. 10.1.

The spatially discretised Markov RF model cannot be easily evaluated by infill asymptotic analysis. Therefore, the model appears as grid-specific. Consistent models across different spatial discretisations are challenging to define.

4.3.2 Hierarchical Markov RF Models

The spatially discretised Markov RF $\{l(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$, with model parameter $\boldsymbol{\theta}_{pM}$, may also be extended into a hierarchical model. The set of model parameters may describe the clique systems, such as the size and shape of each clique unit, as well as the associated parameters that define the strength of the interactions. The Gibbs formulation of the Markov RF model can be expressed as $p(\mathbf{l}; \boldsymbol{\theta}_{pM}) = \text{const}_{\boldsymbol{\theta}_{pM}} \times h(\mathbf{l}; \boldsymbol{\theta}_{pM})$, where $h(\mathbf{l}; \boldsymbol{\theta}_{pM})$ defines the pdf shape and $\text{const}_{\boldsymbol{\theta}_{pM}}$ is the normalising constant, which is computationally demanding to calculate. To assess the Markov RF model with the given set of model parameter value, realisations from $p(\mathbf{l}; \boldsymbol{\theta}_{pM})$ are typically generated by the iterative McMC algorithm, as defined in Sect. 10.1. In this class of algorithms, the constant $\text{const}_{\boldsymbol{\theta}_{pM}}$ is not calculated explicitly. The troublesome constant cancels in the computation of the acceptance probabilities.

In the hierarchical setting, the set of model parameters $\boldsymbol{\theta}_{pM}$ is considered to be a random variable and is assigned a pdf $p(\boldsymbol{\theta}_{pM})$. The focus is on the marginal hierarchical Markov RF model $\{l_H(\mathbf{x}); \mathbf{x} \in L\}$, represented by the n -vector $\mathbf{l}_H \in \mathbb{L}^n$. The corresponding Gibbs formulation can be expressed as

$$\begin{aligned}\mathbf{l}_H \sim p(\mathbf{l}_H) &= \int_{\Omega_{\boldsymbol{\theta}_{pM}}} p(\mathbf{l}, \boldsymbol{\theta}_{pM}) d\boldsymbol{\theta}_{pM} = \int_{\Omega_{\boldsymbol{\theta}_{pM}}} p(\mathbf{l} | \boldsymbol{\theta}_{pM}) p(\boldsymbol{\theta}_{pM}) d\boldsymbol{\theta}_{pM} \\ &= \int_{\Omega_{\boldsymbol{\theta}_{pM}}} \text{const}_{\boldsymbol{\theta}_{pM}} \times h(\mathbf{l} | \boldsymbol{\theta}_{pM}) p(\boldsymbol{\theta}_{pM}) d\boldsymbol{\theta}_{pM}.\end{aligned}$$

The natural procedure for assessing this marginal pdf $p(\mathbf{l}_H)$ is to generate joint realisations $\{[\mathbf{l}, \boldsymbol{\theta}_{pM}]^s; s = 1, 2, \dots, n_s\}$ from $p(\mathbf{l}, \boldsymbol{\theta}_{pM})$ and consider the \mathbf{l} -components only. If an iterative McMC algorithm is used to simulate from $p(\mathbf{l}, \boldsymbol{\theta}_{pM})$, proposals of $\boldsymbol{\theta}_{pM}$ must be generated. To calculate the acceptance probabilities, the quotient of $\text{const}_{\boldsymbol{\theta}_{pM}}$ for two different values of $\boldsymbol{\theta}_{pM}$ must be calculated. Computing $\text{const}_{\boldsymbol{\theta}_{pM}}$ must be performed in each iteration, which is typically too computationally demanding for spatial models.

Chapter 5

Likelihood Models



Observations divorced from their acquisition procedure are merely numbers. They carry no relevant information. The likelihood model specifies the acquisition procedure and thus links observations to the variable of interest. The likelihood model and its associated observed values provide the most valuable information in spatial modelling, as they represent the only connection to the spatial variable under study.

Consider the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D\}$ of interest, spatially discretised to $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ and represented by the n -vector \mathbf{s} . Assume that an m -vector \mathbf{d} of related observations is available. The likelihood model $p(\mathbf{d} | \mathbf{s})$ is defined by the observation collection procedure and links the observations to the variable of interest. The variable of interest \mathbf{s} may be a continuous, event or mosaic spatial variable. Note that in the likelihood model, the observation \mathbf{d} is known, whereas \mathbf{s} is the unknown variable, and $p(\mathbf{d} | \mathbf{s})$ is not necessarily a pdf with respect to \mathbf{s} since it need not integrate to one. Therefore, no normalisation is required, and only the relative values of $p(\mathbf{d} | \mathbf{s})$ when varying \mathbf{s} impact the posterior model. The model formulation will, however, also be dependent on the representation of the actual spatial variable and should consider the spatial discretisation level. Although there are a variety of observation acquisition procedures, we primarily focus on some particular classes of likelihood models for which corresponding conjugate classes of prior models exist for the continuous, event and mosaic spatial variables of interest. The likelihood models are defined relative to the discretised spatial variable to obtain exact probabilistic results without enforcing complex assumptions. In addition, we present extensions of these likelihood models that may be useful in practice.

The conditionally independent, single-site response likelihood model is frequently used in practice

$$p(\mathbf{d} | \mathbf{s}) = \prod_{i=1}^m p(d_i | \mathbf{s}) = \prod_{i=1}^m p(d_i | s_i^d)$$

where the m -vector $\mathbf{s}^d = (s_1^d, s_2^d, \dots, s_m^d)$ contains the values in the grid units where observations are available. The former identity follows from the conditional independence assumption, whereas the latter follows from the single-site response assumption. Making these assumptions on the likelihood function can, for some models, simplify the assessment of the posterior model.

5.1 Continuous Spatial Variables

Consider the continuous spatial variable being spatially discretised to the grid $\{r(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\mathbf{r} \in \mathbb{R}^n$. We define the Gauss-linear likelihood model relative to this spatially discretised spatial variable and observe $\mathbf{d} \in \mathbb{R}^m$ according to

$$[\mathbf{d} | \mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e}_{d|r} \sim p(\mathbf{d} | \mathbf{r}) = \phi_m(\mathbf{d}; \mathbf{H}\mathbf{r}, \Sigma_{d|r}), \quad (5.1)$$

where the $(m \times n)$ matrix \mathbf{H} is the observation design matrix. The m -vector $\mathbf{e}_{d|r}$ captures the observation error, which is assumed to be a centred Gaussian variable with $(m \times m)$ covariance matrix $\Sigma_{d|r}$ and independent of \mathbf{r} . The observation design matrix \mathbf{H} may capture various observation procedures, such as exact observations in certain locations, spatial integrals over sub-domains, derivatives in various directions or mixtures of these.

One particular version of the Gauss-linear likelihood model is also useful to define. The Gauss-point likelihood model represents point observations with independent centred Gaussian error terms independent of \mathbf{r} and is given by

$$[\mathbf{d}_o | \mathbf{r}] = \mathbf{H}_o\mathbf{r} + \mathbf{e}_{d|r} \sim p(\mathbf{d}_o | \mathbf{r}) = \phi_m(\mathbf{d}_o; \mathbf{H}_o\mathbf{r}, \sigma_{d|r}^2 \mathbf{I}_m), \quad (5.2)$$

where the observation design matrix \mathbf{H}_o is binary, with each row and column adding to unity. This entails that m point observations with an error are made in a subset of the grid nodes, $\{d_o(\mathbf{x}) = r(\mathbf{x}) + e_{d|r}; \mathbf{x} \in L^o \subset L\}$, where $e_{d|r}$ are independent, centred Gaussian errors with variance $\sigma_{d|r}^2$. This likelihood model is recognised as being conditionally independent with single-site response.

Moreover, two limiting cases of the Gauss-linear and Gauss-point likelihood models are defined. The Dirac-linear and the Dirac-point likelihood models are given by

$$[\mathbf{d}_e | \mathbf{r}] = \mathbf{H}\mathbf{r} \sim p(\mathbf{d}_e | \mathbf{r}) = \delta_m(\mathbf{d}_e; \mathbf{H}\mathbf{r})$$

$$[\mathbf{d}_{oe} | \mathbf{r}] = \mathbf{H}_o\mathbf{r} \sim p(\mathbf{d}_{oe} | \mathbf{r}) = \delta_m(\mathbf{d}_{oe}; \mathbf{H}_o\mathbf{r}),$$

where $\delta_m(\cdot; \cdot)$ is a Dirac delta function. Note that these models appear as the limiting case of the Gauss-linear and Gauss-point likelihood models, respectively, when $\text{Tr}\{\Sigma_{d|r}\} \rightarrow 0$ and $\sigma_{d|r}^2 \rightarrow 0$.

The class of Gaussian RF and hierarchical Gaussian RF prior models are later demonstrated to be a conjugate class with respect to some of the likelihood models defined above.

Spatial Discretisation Uncertainty

The likelihood model is specified relative to the discretised spatial variable $\{r(\mathbf{x}); \mathbf{x} \in L\}$, which introduces an error beyond the observation error. Consider one observation $d^o = r(\mathbf{x}^o) + e_{d|r}$ in an arbitrary location $\mathbf{x}^o \in D$. The grid node located closest to \mathbf{x}^o is denoted $\mathbf{x}_{i^o} \in L$. Recall that δ_n is the length of the grid unit side. Thus, the distance between the observation location and its closest grid node is $|\mathbf{x}^o - \mathbf{x}_{i^o}| \leq \Delta x_M^n$, with $\Delta x_M^n = 3^{1/2}\delta_n/2$ for three-dimensional reference domains $D \subset \mathbb{R}^3$ and $\Delta x_M^n = 2^{1/2}\delta_n/2$ for two-dimensional reference domains $D \subset \mathbb{R}^2$.

The likelihood model is based on $d = r(\mathbf{x}_{i^o}) + e_{d|r}$ due to the spatial discretisation. The likelihood difference $\Delta d^o = d^o - d$ caused by this spatial discretisation is

$$[\Delta d^o | \{r(\mathbf{x}); \mathbf{x} \in D\}] = [d^o - d | \{r(\mathbf{x}); \mathbf{x} \in D\}] = r(\mathbf{x}^o) - r(\mathbf{x}_{i^o}),$$

thus,

$$E[\Delta d^o | \{r(\mathbf{x}); \mathbf{x} \in D\}] = r(\mathbf{x}^o) - r(\mathbf{x}_{i^o}) = \Delta r^o$$

$$\text{Var}[\Delta d^o | \{r(\mathbf{x}); \mathbf{x} \in D\}] = 0.$$

The discretisation of the likelihood model introduces a bias in the model, which depends on the grid density as well as the variability and smoothness of the spatial variable. By infill asymptotic analysis when $n \rightarrow \infty$ and $\delta_n \rightarrow 0$ the bias Δr^o tends towards zero. Similar expressions can be developed for the full Gauss-linear likelihood model in Expression (5.1).

5.1.1 Example: Continuous Spatial Variable

Consider a continuous spatial variable spatially discretised on a two-dimensional grid with n quadratic grid cells, each having a side length of δ_n . The values in the grid nodes are sequentially represented in the n -vector \mathbf{r} . The observation design matrix \mathbf{H} , of dimension $(m \times n)$, can represent almost any linear observation acquisition design. As an example, consider the $(3 \times n)$ matrix

$$\mathbf{H} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/2\delta_n & 0 & -1/2\delta_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1/4 & 1/4 & 0 & \cdots & 0 & 1/4 & 1/4 & 0 & \cdots & \cdots & 0 \end{pmatrix},$$

which, combined with the observation error 3-vector $\mathbf{e}_{d|r}$, provides the observation 3-vector $\mathbf{d} = (d_1, d_2, d_3)^T$. The observation d_1 captures the spatial variable value in one particular grid node, including an additive error. Observations from several locations can be represented by adding rows in the matrix. The observation d_2 captures the directional derivative of the spatial variable in a specific grid node by differentiation of the values in neighbouring grid nodes. Higher-order stencils may also be used to approximate the derivative, along with sequential differentiation to represent higher-order derivatives. By adding rows in the matrix, multiple derivative observations can be represented. The observation d_3 captures the average value of the spatial variable in a sub-domain. The value is represented by the average values in the grid nodes within the sub-domain, which is 2×2 in this example. Note that the neighbouring nodes in the grid need not be neighbours in the representation of the spatial variable. Weighted average observations can be represented accordingly. Multiple average observations can be included by adding rows in the matrix. Repeated observations can be included by copying the row several times, provided observation errors are present. The observations may overlap spatially without causing complications; for example, observations from specific locations can be located inside the sub-domain of average observations.

The following example of the reconstruction of a continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$ is used throughout the book. Figure 5.1 contains illustrations for the example. In display (a), the target spatial variable is given, and a spatially discretised version is presented in display (b). The discretisation grid has dimension (50×50) , and the spatial variable is represented by a n -vector \mathbf{r} , with $n = 2500$. The task is to reconstruct the discretised spatial variable, also known as the reference spatial variable, which is unknown.

Observations of the spatial variable of interest, as presented in display (c), are available, and the acquisition design must be specified in the corresponding likelihood model $p(\mathbf{d} | \mathbf{r})$. Five nodewise observations and two (5×5) -node averages are collected, all with observation errors. The observed values are represented by the m -vector \mathbf{d} , with $m = 7$. These values are presented in shades of grey in display (c). The observation errors are assumed to be additive, independent, centred Gaussian, and unrelated to the spatial variable. The error variance for the five

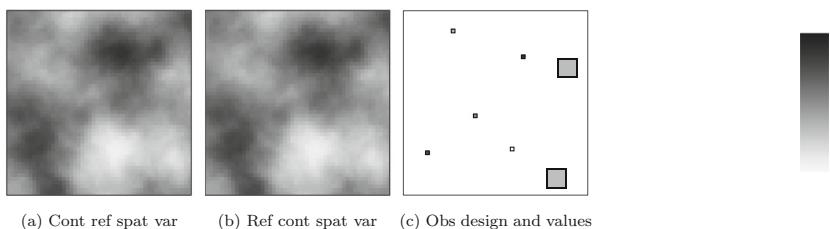


Fig. 5.1 Example Gaussian RF model. Gauss-linear likelihood model: (a) continuous reference spatial variable, (b) reference continuous spatial variable gridded and (c) observation design and values

nodewise observations is $\sigma_{d|r,1}^2 = 1.0$, whereas for the two node-average ones it is $\sigma_{d|r,2}^2 = 0.04$. These assumptions provide the Gauss-linear likelihood model $p(\mathbf{d} | \mathbf{r})$.

The challenge is to reconstruct the reference spatial variable in display (b) based on the observation set in display (c) with the associated Gauss-linear likelihood model. This assessment cannot be made without additional assumptions on the spatial variable, which the prior model provides.

5.2 Event Spatial Variables

Consider the event spatial variable $\{k_\Delta(\mathbf{x}); \mathbf{x} \in L\}$ spatially discretised to the grid, represented by the event count n -vector $\mathbf{k} \in \mathbb{N}_\oplus^n$. The observations are represented by a set of event locations $\mathbb{D} : \{\mathbf{x}_1^d, \mathbf{x}_2^d, \dots, \mathbf{x}_{n_d}^d\}; \mathbf{x}_i^d \in D$. The misclassification likelihood model is defined relative to the spatially discretised spatial variable. The spatially discretised observations $\{d_\Delta(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\mathbf{d} \in \mathbb{N}_\oplus^n; \mathbf{d} \leq \mathbf{k}$ are defined such that $d_i = \sum_{\mathbf{x} \in \mathbb{D}} I(\mathbf{x} \in \Delta_{ni})$. Each d_i is the number of observed events in each grid unit Δ_{ni} , which is less than or equal to the actual event count because some events may have been overlooked.

For each grid node i , the likelihood model follows a binomial with $d_i \leq k_i$, and is defined as

$$[d_i | k_i] \sim p(d_i | k_i) = \binom{k_i}{d_i} \alpha_i^{d_i} (1 - \alpha_i)^{k_i - d_i},$$

with the probability field for correctly observing an event being $\{\alpha(\mathbf{x}); \mathbf{x} \in D\}; \alpha(\mathbf{x}) \in \mathbb{R}_{[0,1]}$. The probability field is spatially discretised into $\{\alpha(\mathbf{x}); \mathbf{x} \in L\}$ and represented by the n -vector $\boldsymbol{\alpha} \in \mathbb{R}_{[0,1]}^n$. The observations are assumed to be made independently from one event to another, and thus, the likelihood model with $\mathbf{d} \leq \mathbf{k}$ is

$$[\mathbf{d} | \mathbf{k}] \sim p(\mathbf{d} | \mathbf{k}) = \prod_{i=1}^n p(d_i | k_i) = \prod_{i=1}^n \binom{k_i}{d_i} \alpha_i^{d_i} (1 - \alpha_i)^{k_i - d_i}. \quad (5.3)$$

This likelihood model is recognised as being conditionally independent with single-site response. It is possible to represent the likelihood model for cases where only parts of the grid L are observed by assigning $d_i = 0$ and $\alpha_i = 0$ for the unobserved nodes.

The class of Poisson RF and hierarchical Poisson RF prior models is later demonstrated to be a conjugate class with respect to the likelihood models defined above.

Instead of observing events in each grid unit, the accumulated number of events within a spatial sub-domain for event spatial variables may be observed. For a set of disjoint sub-domains $A_i \subset D; i = 1, 2, \dots, m$ with sub-grid representation $L_{A_i} \subset L$

of size n_{A_i} ; $i = 1, 2, \dots, m$, the sum of observations in each of these sub-domains is registered as $d_{A_i} = \sum_{i \in L_A} d_i$; $i = 1, 2, \dots, m$. This is denoted as accumulated observations. The union of these areas is denoted A and the corresponding set of grid nodes is denoted L_A . Note that the sub-domains may consist of only one grid node and that for all grid nodes $i \in L_A^c$, it is necessary that both $\alpha_i = 0$ and $d_i = 0$ if L_A contains all known observations.

Denote the m -vector of accumulated observations \mathbf{d}_A and consequently the likelihood $p(\mathbf{d}_A | \mathbf{k}) = p(\mathbf{d}_A | \mathbf{k}_A)$, with \mathbf{k}_A being the event counts in the grid nodes within sub-grid L_A . Although no conjugate class of prior models exists for this accumulated likelihood model, we demonstrate later that relatively simple expressions for posterior models can be developed.

Spatial Discretisation Uncertainty

The observation event-location set \mathbb{D} is exact and contains the actual event locations detected according to a user-specified misclassification field. The likelihood model is, however, specified relative to the discretised event count spatial variable $\{k_\Delta(\mathbf{x}); \mathbf{x} \in L\}$, representing the number of observations in each grid unit. This likelihood model is accurate, but the spatial discretisation reduces its resolution. One specific observation in location $\mathbf{x}^o \in D$ within grid unit Δ_{ni} can only be reproduced with location precision equal to the grid unit volume Δ_n . By using infill asymptotic analysis with D constant and letting $n \rightarrow \infty$, thus having $\Delta_n \rightarrow 0$, an exact reproduction of the observations can be made.

5.2.1 Example: Event Spatial Variable

The following example of reconstruction of an event spatial variable $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$ is used throughout the book. Figure 5.2 contains illustrations for the example. In display (a), the target event-location spatial variable is given, and the corresponding spatially discretised target event count spatial variable, represented by the n -vector \mathbf{k} with $n = 20 \times 20$, is presented in display (b). These reference event spatial variables are unknown, and we want to assess them.

The observation acquisition procedure causes some events to be overlooked. The registration probability $\{\alpha(\mathbf{x}); \mathbf{x} \in D\}$ is spatially varying, as specified in display (c). Almost every event is observed in the centre of the domain, whereas most events are overlooked in the northeastern corner. The observed event locations are presented in display (g), and the corresponding spatially discretised event count observations, represented by the n -vector \mathbf{d} , are presented in display (h). The misclassification likelihood model $p(\mathbf{d} | \mathbf{k})$ is defined by this observation acquisition procedure.

The challenge is reconstructing the reference event spatial variables in displays (a) and (b) based on the observation sets in displays (g) and (h) with the associated misclassification likelihood model. This assessment is made in a Bayesian framework by introducing a prior model for the spatial variable.

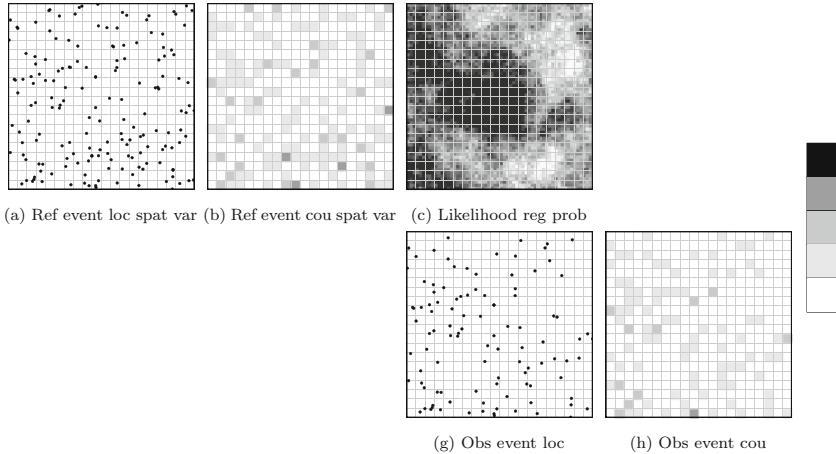


Fig. 5.2 Example Poisson RF model. Misclassification likelihood model: **(a)** reference event-location spatial variable, **(b)** reference event count spatial variable, **(c)** likelihood registration probability, **(g)** observed event locations and **(h)** observed event counts

5.3 Mosaic Spatial Variables

Consider the mosaic spatial variable being spatially discretised on the grid $\{l(\mathbf{x}); \mathbf{x} \in \mathcal{L}\}$, represented by the n -vector $\mathbf{l} \in \mathbb{L}^n$. The response likelihood model provides a response for each distinct label in every grid node. The n -vector \mathbf{d} represents the corresponding observations on the grid, with $\mathbf{d} \in \mathbb{R}^n$ or $\mathbf{d} \in \mathbb{L}^n$. The observations are assumed to be conditionally independent from one node to the other and to have a single-site response. Thus, the likelihood model is in factorial form

$$[\mathbf{d} | \mathbf{l}] \sim p(\mathbf{d} | \mathbf{l}) = \prod_{i=1}^n p(d_i | l_i) = \prod_{i=1}^n p(d_i | l_i). \quad (5.4)$$

The response may be signals $\mathbf{d} \in \mathbb{R}^n$ of the type

$$[d_i | l_i] = h_i(l_i) + e_{d_i | l_i} \sim p(d_i | l_i) = p_e(d_i - h_i(l_i) | l_i),$$

where $h_i(l_i) \in \mathbb{R}$ is the signal and $e_{d_i | l_i} \in \mathbb{R}$ is a random error term with label-dependent centred pdf $p_e(e | l_i)$, for each grid node. For $i \neq j$, $e_{d_i | l_i}$ and $e_{d_j | l_j}$ are assumed to be independent. Alternatively, the response may be labels $\mathbf{d} \in \mathbb{L}^n$ of the type

$$[d_i | l_i] \sim p(d_i | l_i) = p_q(d_i | l_i),$$

where the acquisition procedure may introduce misclassifications, represented by the $(n_L \times n_L)$ matrix \mathbf{Q} with elements $[\mathbf{Q}]_{l,l'} = p_q(l' | l)$ being the probability for observing label l' if label l is the correct one. Signal and misclassification likelihood models may be combined with $\mathbf{d} \in \mathbb{R}^n$,

$$[d_i | l_i] \sim p(d_i | l_i) = \sum_{l' \in \mathbb{L}} p(d_i | l') p(l' | l_i) = \sum_{l' \in \mathbb{L}} p_e(d_i - h_i(l') | l') p_q(l' | l_i).$$

If observations are available in only parts of the grid L , the unobserved grid nodes must be assigned a non-zero constant likelihood value independent of the label l_i .

The class of Markov RF prior models is later demonstrated to be a conjugate class with respect to these likelihood models.

It is also possible to define convolved observations, also called blurred observations, which implies that the observations have a multi-site response. The corresponding likelihood model is then

$$[\mathbf{d}_b | \mathbf{l}] \sim p(\mathbf{d}_b | \mathbf{l}) = \prod_{i=1}^n p(d_{bi} | \mathbf{l}) = \prod_{i=1}^n p(d_{bi} | l_j; j \in \mathbf{n}_i^d), \quad (5.5)$$

with \mathbf{n}_i^d representing the convolution neighbourhood of the observation in grid node i . Although no conjugate class of prior models exists, according to this book's definition, we later demonstrate that relatively simple expressions for the posterior models can be developed for this convolved likelihood model.

Spatial Discretisation Uncertainty

The likelihood model of the mosaic spatial variable is specified relative to the discretised spatial variable $\{l(\mathbf{x}); \mathbf{x} \in L\}$. This likelihood accurately represents the observations in the grid nodes according to the observation acquisition procedure. However, the precision of representing any label transition location is limited to the length of the grid unit side δ_n . Increasing the grid density enhances the precision of reproducing these transition locations.

5.3.1 Example: Mosaic Spatial Variable

The following example of reconstruction of a mosaic spatial variable $\{l(\mathbf{x}); \mathbf{x} \in L \subset D \subset \mathbb{R}^2\}$ is used throughout the book. The mosaic variable consists of two labels, $l(\mathbf{x}) \in \mathbb{L} : \{W, B\}$, representing white and black. Figure 5.3 contains illustrations for the example. In display (a), the target mosaic variable on a (50×50) spatial grid is presented, and it is represented by the 2500-vector \mathbf{l} .

Observations of the spatial variable are collected at each grid node, and the two labels produce different signals, $h_i(l) \in \mathbb{R}; l \in \{W, B\}$, with $h_i(W) = -1.0$ and $h_i(B) = 1.0$ for $i = 1, 2, \dots, n$. The observation acquisition introduces errors of two types: misclassifications and signal noise. The former error type

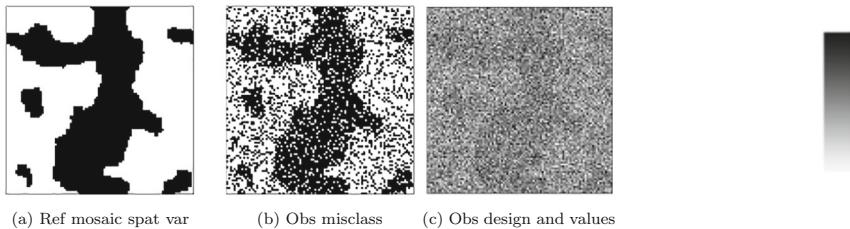


Fig. 5.3 Example Markov RF model. Response likelihood model: **(a)** reference mosaic spatial variable, **(b)** misclassification observations and **(c)** observation design and values

causes the labels to be interchanged with probability 0.25 for B to change to W and 0.15 for the reverse for each grid node. The signal noise is centred Gaussian with a variance of 3.0. Both error types occur independently across the spatial domain. Display (b) contains the misclassifications that occur during the observation acquisition, and display (c) presents the observations collected, including signal noise and misclassifications. Given this set of observations, represented by the 2500-vector \mathbf{d} , the reference mosaic is challenging to identify. The observation acquisition procedure outlined above defines the associated likelihood model $p(\mathbf{d} | \mathbf{I})$.

The challenge is to restore the reference mosaic spatial variable in display (a) based on the observed values in display (c), given the associated response likelihood model. This assessment is made in a Bayesian setting by introducing a prior model for the spatial variable.

Chapter 6

Prior Models



Predictive statistics, unlike statistical inference, focuses on real-world phenomena. These phenomena are in principle observable, and one may envision a super-population of phenomena with similar characteristics. Experts may quantify these characteristics to define a prior probabilistic model for the phenomenon under study. This prior model, representing expert knowledge, becomes indispensable when dealing with scarce and imprecise observations. Merged with site-specific observations and the associated likelihood models, the prior model provides statistical predictions with quantified levels of uncertainty.

Consider the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D\}$, spatially discretised to $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ and represented by the n -vector \mathbf{s} . Let the associated prior model be defined from a parametric class of pdfs $p(\mathbf{s}; \boldsymbol{\theta}_p)$. This prior model must be user-specified based on prior experience and expert knowledge. The characteristics of the variable will usually make either a continuous, event or mosaic model suitable. A combination of the spatial variable types may be required for more complex spatial variables to define an adequate model. The prior model may also depend on spatial scale. One example is the porosity of rock: at a small scale it may be modelled as a mosaic model with labels for the rock and voids, but at a larger scale, it is usually modelled as a continuous proportion of voids. Furthermore, the observation collection procedure should also be taken into account because some likelihood models align well with specific prior models. In fact, some pairings of prior and likelihood models yield analytically assessable posterior models that are on a known format, and we identify such pairings in sections devoted to the concept of conjugate prior models.

In Bayesian spatial modelling, the prior model represents prior knowledge of the spatial variable under study. It is advisable to be conservative when specifying the model rather than introducing very detailed information. An empirical Bayesian framework is often natural to use when eliciting the prior model, and observations from natural variations of comparable spatial variables elsewhere may be used in the specification. If comparable spatial variables are used, one would usually

specify prior models that are spatially stationary horizontally, possibly with some vertical trends, and let the observations introduce local effects through the likelihood model. In some studies, introducing explanatory spatial variables as covariates for the spatial variable under study may be beneficial. These covariates can be included in the prior model in a spatial regression framework, usually without influencing the conjugate characteristic.

More realistic uncertainty quantifications can be provided by also taking model parameter uncertainty into account. This incorporation leads to a hierarchical prior model where the model parameters in θ_p are considered to be random variables that must also be assigned a prior pdf $p(\theta_p)$.

In Chap. 5, three classes of likelihood models are presented, corresponding to each of the spatial variable types: continuous, event and mosaic. These likelihood classes have corresponding classes of conjugate prior models, which are defined in this section, and together they define the posterior model, as discussed in Chap. 2. The following sections present various characteristics of these conjugate prior models and related topics.

6.1 Continuous Spatial Variables: Gaussian RF Models

The class of Gaussian RF models is defined in Sect. 4.1.1, with the corresponding class of spatially discretised Gaussian RF models in Expression (4.1). The class of Gauss-linear likelihood models is defined in Sect. 5.1. Later, Sect. 7.1 demonstrates that this pair of likelihood and prior model classes has conjugate characteristics.

The Gaussian RF prior model is typically defined on the spatially discretised representation $\mathbf{r} \in \mathbb{R}^n$ as

$$\mathbf{r} \sim p(\mathbf{r}) = \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \boldsymbol{\Sigma}_r^\rho), \quad (6.1)$$

which is a spatially discretised stationary Gaussian RF model with expectation and variance levels $\mu(\mathbf{x}) = \mu_r$ and $\sigma^2(\mathbf{x}) = \sigma_r^2$, respectively, and with $(n \times n)$ correlation matrix $\boldsymbol{\Sigma}_r^\rho$. The correlation matrix is defined by the translation invariant non-negative definite spatial correlation function $\rho(\mathbf{x}, \mathbf{x}') = \rho_r(\mathbf{x} - \mathbf{x}')$. For an overview of the use of Gaussian models in spatial statistics, see Gelfand and Schliep (2016).

Note that all marginal pdfs $p(r_i)$ and $p(r_i, r_j)$ are invariant with respect to translations on the grid L. Furthermore, both the stationarity assumption and non-negative definiteness requirement imply that $p(r_i, r_j) \rightarrow p(r_i)p(r_j)$ as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$, which entails ergodicity. Thus, marginal variability and spatial heterogeneity are identical in the extension limit.

Spatial Discretisation Uncertainty

The prior model is specified to be a spatially discretised, stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in L\}$, and this discretisation introduces an error. To quantify this error,

assume that the full spatial representation is obtained as a piecewise constant function based on the grid representation, and consider the value of the spatial variable $r(\mathbf{x}^o)$ at an arbitrary location $\mathbf{x}^o \in D$. Let $\mathbf{x}_{i^o} \in L$ be the grid node located nearest to \mathbf{x}^o and recall that δ_n is the length of the grid unit sides. Then, the distance between the observation location and its nearest grid node is $|\mathbf{x}^o - \mathbf{x}_{i^o}| \leq \Delta x_M^n$, with $\Delta x_M^n = 3^{1/2}\delta_n/2$ for three-dimensional reference domains $D \subset \mathbb{R}^3$ and $\Delta x_M^n = 2^{1/2}\delta_n/2$ for two-dimensional reference domains $D \subset \mathbb{R}^2$. The spatial variable in the location \mathbf{x}^o is, due to the spatial discretisation, assigned the value $r(\mathbf{x}_{i^o})$. Consequently, the error is

$$\Delta r^o = r(\mathbf{x}^o) - r(\mathbf{x}_{i^o}),$$

which is Gaussian for Gaussian RFs, with

$$\begin{aligned} E\{\Delta r^o\} &= 0 \\ \text{Var}\{\Delta r^o\} &= 2\sigma_r^2[1 - \rho_r(\mathbf{x}^o - \mathbf{x}_{i^o})] = \sigma_\Delta^2. \end{aligned}$$

The discretisation of the prior model introduces additional variability in the model, which depends on the grid density and the variance and spatial correlation function of the Gaussian RF. However, the error variance is bounded for isotropic, stationary Gaussian RF with monotonically decreasing correlation functions, $\sigma_\Delta^2 \leq 2\sigma_r^2[1 - \rho_r(\Delta x_M^n)]$. It can be shown that the error variance σ_Δ^2 tends towards zero for spatial correlation functions that are continuous at zero by use of infill asymptotic analysis with $n \rightarrow \infty$ and $\delta_n \rightarrow 0$.

6.2 More on Gaussian RF Prior Models

The general definition of a Gaussian RF is presented in Definition 5. Consider the stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^q\}$, normally with $q = 3$, on which the spatially discretised Gaussian prior model is based. The model parameters are the expectation and variance levels, $\{\mu(\mathbf{x}) = \mu_r; \mathbf{x} \in D\}$ and $\{\sigma^2(\mathbf{x}) = \sigma_r^2; \mathbf{x} \in D\}$, and the spatial correlation function $\{\rho(\mathbf{x}, \mathbf{x}') = \rho_r(\boldsymbol{\tau}); \mathbf{x}, \mathbf{x}' \in D\}$ with $\boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$.

6.2.1 Model Parameters and Characteristics

The vector of model parameters is $\boldsymbol{\theta}_{pG} = [\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ which contains the expectation $\mu_r \in \mathbb{R}$, variance $\sigma_r^2 \in \mathbb{R}_+$ and spatial correlation function $\rho_r(\boldsymbol{\tau}) \in \mathbb{R}_{[-1, 1]}$ being a non-negative definite function. This latter requirement ensures that all covariance matrices, for all spatial configurations and all dimensions, are positive definite matrices in the definition of the stationary Gaussian RF.

Non-negative Definiteness

The class of non-negative definite functions is defined in Definition 10. This function class is closed under positive scaling. Hence, we can assume that $\rho_r(0\mathbf{i}_q) = 1$. The necessary constraint $\rho_r(\boldsymbol{\tau}) \in \mathbb{R}_{[-1,1]}$ is then implicitly ensured by the non-negative definiteness constraint, as is the symmetry requirement $\rho_r(\boldsymbol{\tau}) = \rho_r(-\boldsymbol{\tau})$. Non-negative definiteness is related to the dimension of the reference domain. If a function is non-negative definite for dimension q , it is so for all $q' \leq q$. The non-negative definiteness constraint on the spatial correlation function is a severe constraint, and identifying valid functions is not necessarily easy. These functions share some characteristics, however, which are discussed later.

Definition 10 (Non-negative (Positive) Definite Functions) Consider a function $c(\boldsymbol{\tau}) : \mathbb{R}^q \rightarrow \mathbb{R}$. If

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j c(\mathbf{x}_i - \mathbf{x}_j) \geq 0$$

for all configurations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{q \times n}$, for all weights $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^T \in \mathbb{R}^n$, and for all $n \in \mathbb{N}_{[2,\infty]}$, then $c(\boldsymbol{\tau})$ is a non-negative definite function for $\boldsymbol{\tau} \in \mathbb{R}^q$.

Moreover, the class of non-negative definite correlation functions is closed under addition with normalised weights and closed under multiplication, which considerably extends the class of valid correlation functions. These properties of closedness can be more formally expressed if we consider a set of non-negative definite correlation functions $\{\rho_{ri}(\boldsymbol{\tau}); i = 1, 2, \dots, k\}; \boldsymbol{\tau} \in \mathbb{R}^q$. The following non-negative definite correlation functions can be defined:

$$\begin{aligned} \rho(\boldsymbol{\tau}) &= \sum_{i=1}^k \alpha_i \rho_{ri}(\boldsymbol{\tau}) \text{ with } \alpha_i \in \mathbb{R}_{\oplus} \text{ and } \sum_{i=1}^k \alpha_i = 1 \\ \rho(\boldsymbol{\tau}) &= \prod_{i=1}^k \rho_{ri}(\boldsymbol{\tau}). \end{aligned}$$

Closedness under addition with normalised weights is frequently used to combine two correlation functions: one capturing small-scale variability, such as white noise, and one capturing larger-scale spatial dependence. This construction makes it possible to add a so-called nugget effect to the model. Closedness under multiplication can be employed to enforce finite-range characteristics on a selected correlation function by defining the second factor to be a finite-range correlation function. It may also be applied to construct a spatial correlation function with spatial separability, i.e. each factor represents spatial dependencies specific to a sub-dimension of \mathbb{R}^q . One factor can, for example, represent spatial dependencies in the horizontal dimensions with the other representing dependencies in the vertical one.

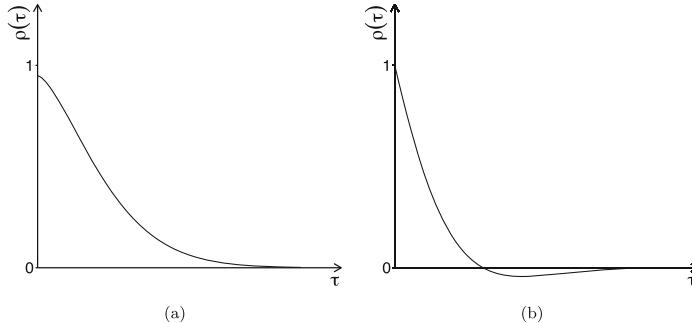


Fig. 6.1 Gaussian RF model: schematic presentation of spatial correlation functions: (a) regular spatial correlation function and (b) hole-effect spatial correlation function

The interpretations of the parameters μ_r and σ_r^2 in the Gaussian prior model are clear. The former defines the level around which the RF fluctuates, and the latter defines the magnitude of fluctuation. The spatial correlation function $\rho_r(\tau)$ is more complicated to interpret, but the functional shape for small $|\tau|$ has the most impact. A slow decay from $\rho_r(0\mathbf{i}_q) = 1$ provides fairly smooth realisations of the RF, whereas a rapid decline provides rugged realisations. In Fig. 6.1, two typical non-negative definite isotropic correlation functions are displayed. The isotropy $\tau = |\tau|$ is enforced only to simplify presentation and discussion. For all correlation functions $\rho_r(0) = 1$ and $\rho_r(\tau) \in \mathbb{R}_{[-1, 1]}; \tau \in \mathbb{R}_+$ because the function represents correlation between two random variables. The function is continuous everywhere, except at $\tau = 0$ where a step may occur. This discontinuity, known as the nugget effect, represents a white noise component in the RF. The small-scale variability of the RF is represented in $\rho_r(\tau)$ for small τ . If the correlation function is continuous at $\tau = 0$, then the RF is continuous almost everywhere. If the function is $2k$ times differentiable at $\tau = 0$, then the RF is k times differentiable almost everywhere, as given in Adler and Taylor (2007). Away from $\tau = 0$, the function must be smooth, and as $\tau \rightarrow \infty$, the function $\rho_r(\tau) \rightarrow 0$. This asymptotic behaviour ensures that two spatial variables at locations \mathbf{x} and \mathbf{x}' tend towards being uncorrelated as $|\mathbf{x} - \mathbf{x}'| \rightarrow \infty$. For Gaussian RFs, being uncorrelated entails independence. Thus, using extension asymptotic analysis, one can demonstrate that the Gaussian RFs are ergodic. Certain correlation functions are said to have a finite range, meaning that independence is obtained for $\tau < \infty$. The convergence of $\rho_r(\tau)$ to zero as $\tau \rightarrow \infty$ may occur monotonically as in display (a), or it may occur with oscillations or pseudo-periodicity, referred to as the hole effect, as in display (b). This oscillation implies that the spatial variables exhibit negative correlation at certain inter-distances.

We now return to general non-negative definite functions, maintaining the assumption of isotropy $\tau = |\tau|$ for presentational simplicity, and specify a procedure for identifying them along the lines of Gneiting and Sasvari (1999). Theorem 2 is particularly useful when specifying correlation functions.

Theorem 2 (Schoenberg Theorem (1938)) Consider a function $c(\tau) : \mathbb{R}_+ \rightarrow \mathbb{R}; \tau = |\tau|; \tau \in \mathbb{R}^q$. If $c(\tau)$ can be expressed as the scale mixture for $\tau \in \mathbb{R}_+$

$$c(\tau) = \int_{\mathbb{R}_+} \Omega_q(\xi \tau) p(\xi) d\xi,$$

where $\Omega_q(u) = \Gamma(q/2) \left[\frac{2}{u} \right]^{(q-2)/2} \times J_{(q-2)/2}(u); u \in \mathbb{R}_+$, $p(\xi)$ is an arbitrary pdf for $\xi \in \mathbb{R}_+$, and $J_{(q-2)/2}(u)$ is the Bessel function, then $c(\tau)$ is a non-negative (positive) definite function.

The dimension of the reference domain $D \subset \mathbb{R}^q$ is important for the existence of non-negative definite functions and the corresponding spatial correlation functions. First, we consider some classes of non-negative definite functions that exist for all dimensions, i.e. for $q = \infty$ and, therefore, for the current model with $q = 3$. It can be demonstrated that $\Omega_\infty(u) = \exp(-u^2) \in \mathbb{R}_{[0,1]}; u \in \mathbb{R}_+$, and thus $c(\tau) \in \mathbb{R}_+; \tau \in \mathbb{R}_+$. Consequently, for $q = \infty$, no spatial correlation functions with pseudo-periodicity exist satisfying Theorem 2. We obtain parametric classes of eligible spatial correlation functions by assigning the arbitrary pdfs $p(\xi)$ parametric pdf models. The most commonly used classes are the white noise, powered exponential and Matérn classes. These model classes are presented in Table 6.1 and Fig. 6.2.

None of the previously defined correlation functions appear with pseudo-periodicity or finite-range dependence. To obtain correlation functions with these effects, we need to search among non-negative definite functions for dimension $q < \infty$. Because we are concerned with $D \subset \mathbb{R}^3$, we use $q = 3$. We obtain the damped cosine and the spherical classes, as seen in Table 6.1 and Fig. 6.2, by assigning different parametric pdfs for $p(\xi)$. The former appears with pseudo-periodicity, whereas the latter exhibits finite-range dependence.

In Table 6.1 and Fig. 6.2, several classes of non-negative definite spatial correlation functions are presented. In the table, τ_ξ represents a normalised inter-distance, and v is a shape parameter. The normalised inter-distance for $\tau \in \mathbb{R}^3$ with anisotropy factor $\xi = (\xi_1, \xi_2, \xi_3)$ is specified to be $\tau_\xi = [[\tau_1/\xi_1]^2 + [\tau_2/\xi_2]^2 + [\tau_3/\xi_3]^2]^{1/2}$. In the table, the column “Validity” refers to the maximum dimension of the reference domain for which the spatial correlation function is non-negative definite. The column “Continuity” refers to the order of differentiability of the corresponding Gaussian RF. In the figure, four classes of non-negative definite spatial correlation functions are displayed together with one realisation for each of the three different parameter sets. The realisations are generated by the transformation algorithm, as defined in Sect. 10.1, with identical white noise realisations. Therefore, all differences between the realisations are caused by the different spatial correlation functions. Observe the differences in small-scale variability and long-range spatial dependence.

In practice, a significant amount of subjectivity is involved in assigning the spatial correlation function model. The expected smoothness characteristics of the

Table 6.1 Gaussian RF spatial correlation functions for $\{r(\mathbf{x}); \mathbf{x} \in \mathbb{R}^q\}$ with $\tau_\xi^2 = \boldsymbol{\tau}^T \boldsymbol{\Psi}_\xi^{-1} \boldsymbol{\tau}$, where the diagonal ($q \times q$) matrix $\boldsymbol{\Psi}_\xi = \text{Mdiag}_\xi \left\{ \xi^2 \right\}$ with ξ being the anisotropy q -vector. Model parameters, valid dimensionality and continuity of the Gaussian RF are also listed

| Class | Expression | Parameters | Validity | Continuity |
|---------------------|---|--|-----------------|---------------------------------|
| White noise | $\rho(\tau_\xi) = I(\tau_\xi = 0)$ | $\xi \in \mathbb{R}_\oplus^q$ | $q \leq \infty$ | — |
| Powered exponential | $\rho(\tau_\xi) = \exp(-\tau_\xi^\nu)$ | $\xi \in \mathbb{R}_\oplus^q$ | $q \leq \infty$ | $\nu = 2$ ∞ diff. |
| Matérn | $\rho(\tau_\xi) = \frac{2^{1-\nu}}{F^{(v)}} \tau_\xi^\nu \mathcal{B}_v(\tau_\xi)$ | $\nu \in (0, 2]$ $\xi \in \mathbb{R}_\oplus^q$ $\nu \in \mathbb{R}_+$ | $q \leq \infty$ | $\nu \geq 2$ $\nu - 1$ diff. |
| Damped cosine | $\rho(\tau_\xi) = \cos(\nu \tau_\xi) \exp(-\tau_\xi)$ | $\xi \in \mathbb{R}_\oplus^q$ $\nu \in \mathbb{R}_+$ $\nu \xi < 3 \tan \pi / 2q$ | $q \leq 3$ | — |
| Spherical | $\rho(\tau_\xi) = I(\tau_\xi \leq 1) \times (1 + \frac{\tau_\xi}{2})(1 - \tau_\xi)^2$ | $\xi \in \mathbb{R}_\oplus^q$ | $q \leq 3$ | — |

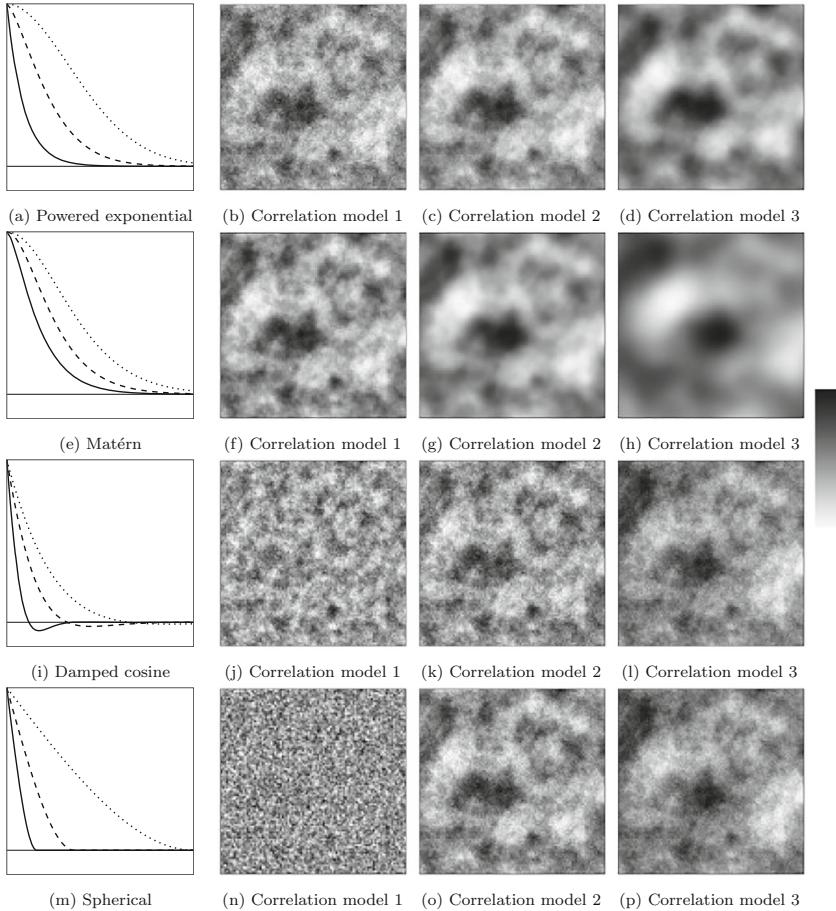


Fig. 6.2 Gaussian RF models with varying spatial correlation models: correlation model 1 (solid line), correlation model 2 (dashed line) and correlation model 3 (dotted line). **(a)** powered exponential model; **(b), (c), (d)** three realisations with model parameters $(\nu, \xi) \in \{(1.0, 1.5\mathbf{i}_2), (1.4, 2.5\mathbf{i}_2), (1.9, 3.5\mathbf{i}_2)\}$; **(e)** Matérn model; **(f), (g), (h)** three realisations with model parameters $\xi = 1.7\mathbf{i}_2$ and $\nu \in \{1, 2, 4\}$; **(i)** damped cosine model; **(j), (k), (l)** three realisations with model parameters $(\nu, \xi) \in \{(0.9, 1.3\mathbf{i}_2), (0.7, 1.9\mathbf{i}_2), (0.55, 2.4\mathbf{i}_2)\}$; **(m)** spherical model; and **(n), (o), (p)** three realisations with model parameters $\xi \in \{2\mathbf{i}_2, 3\mathbf{i}_2, 5\mathbf{i}_2\}$

Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ play a central role. A basis spatial correlation function $\rho_B(\tau_\xi)$ will normally be assigned based on experience. The properties of closedness can be used to adjust this model for small-scale variability or finite spatial influence. We can adjust for small-scale variability by adding a white noise model to the basis model, and we can adjust for finite spatial influence by multiplying the basis model with a finite-range spherical model. If a nugget effect is expected to be

present in the Gaussian RF, one may add a white noise model $\rho_W(\tau_\xi)$ to the basis model to obtain the spatial correlation function

$$\begin{aligned}\rho_r(\tau_\xi) &= \alpha\rho_W(\tau_\xi) + (1 - \alpha)\rho_B(\tau_\xi) \\ &= \alpha I(\tau_\xi = 0) + (1 - \alpha)\rho_B(\tau_\xi)\end{aligned}$$

with $\alpha \in \mathbb{R}_{[0,1]}$. The nugget variance will eventually be $\alpha\sigma_r^2$, and the remaining variance $(1 - \alpha)\sigma_r^2$ appears in the spatially smooth component.

If a finite spatial correlation effect is desired, one may multiply the basis model by a spherical model $\rho_S(\tau_\xi)$ to obtain the spatial correlation function

$$\begin{aligned}\rho_r(\tau_\xi) &= \rho_S(\tau_\xi) \times \rho_B(\tau_\xi) \\ &= I(\tau_\xi \leq 1) \times (1 + \tau_\xi/2)(1 - \tau_\xi)^2 \times \rho_B(\tau_\xi).\end{aligned}$$

The finite-range characteristic is inherited from the spherical factor. This effect will typically make the correlation matrix Σ_r^ρ sparse, which may reduce the computational demands later in spatial prediction.

The selected spatial correlation function class $\rho_r(\boldsymbol{\tau}; \boldsymbol{\eta}_r)$ is typically dependent on a set of model parameter values $\boldsymbol{\eta}_r$. This entails that the Gaussian RF prior model will have parametrisation $\theta_{pG} = [\mu_r, \sigma_r^2, \boldsymbol{\eta}_r]$. Inference of these model parameter values is usually made from the available observations of the spatial variable under study, as discussed in Sect. 8.1.

Linear Operators

The class of Gaussian RF models is closed with respect to linear operators, such as integration and differentiation. We now present the Gaussian RFs resulting from applying these linear operations to a Gaussian RF. Consider the stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D\}$, and for presentational convenience let it be isotropic with $\tau = |\boldsymbol{\tau}| = |\mathbf{x} - \mathbf{x}'|$. Define the related δ -average filtered Gaussian RF, when it exists, by

$$\left\{ r_\delta(\mathbf{x}) = \frac{1}{|\delta|} \int_{\delta_x} r(\mathbf{u}) d\mathbf{u}; \mathbf{x} \in D_\delta \right\}, \quad (6.2)$$

where $\delta \subset \mathbb{R}^3$ is a relatively small rotation-invariant volume, δ_x represents this volume centred at $\mathbf{x} \in D_\delta$ and D_δ is the reference domain D eroded by the volume δ . This filtered Gaussian RF exists whenever $\rho_r(\tau)$ is continuous at $\tau = 0$. The model parameters of this filtered Gaussian RF are

$$\begin{aligned}E\{r_\delta(\mathbf{x})\} &= \mu_r \\ \text{Cov}\{r_\delta(\mathbf{x}), r_\delta(\mathbf{x}')\} &= \frac{\sigma_r^2}{|\delta|^2} \int_{\delta_x} \int_{\delta_{x'}} \rho_r(|\mathbf{u} - \mathbf{v}|) d\mathbf{u} d\mathbf{v}.\end{aligned}$$

Hence,

$$\begin{aligned}\mu_{r_\delta} &= \mu_r \\ \sigma_{r_\delta}^2 &= \frac{\sigma_r^2}{|\delta|^2} \int_{\delta_x} \int_{\delta_x} \rho_r(|\mathbf{u} - \mathbf{v}|) d\mathbf{u} d\mathbf{v} \\ \rho_{r_\delta}(\tau) &= \left[\int_{\delta_x} \int_{\delta_x} \rho_r(|\mathbf{u} - \mathbf{v}|) d\mathbf{u} d\mathbf{v} \right]^{-1} \times \int_{\delta_x} \int_{\delta_{x'}} \rho_r(|\mathbf{u} - \mathbf{v}|) d\mathbf{u} d\mathbf{v}.\end{aligned}$$

Note that the cross-covariance of the filtered Gaussian RF with the basis Gaussian RF is

$$\text{Cov}\{r_\delta(\mathbf{x}), r(\mathbf{x})\} = \frac{\sigma_r^2}{|\delta|} \int_{\delta_x} \rho_r(|\mathbf{x} - \mathbf{u}|) d\mathbf{u}.$$

Similarly, define the differential Gaussian RF in direction $\mathbf{x}_i \in \mathbb{R}^3$, if it exists, by

$$\left\{ \dot{r}(\mathbf{x}) = \frac{d\mathbf{r}(\mathbf{x})}{d\mathbf{x}_i}; \mathbf{x} \in D \right\}. \quad (6.3)$$

Due to the assumed isotropy, the differential RF is independent of the direction of \mathbf{x}_i . The model parameters of this Gaussian RF are

$$\begin{aligned}E\{\dot{r}(\mathbf{x})\} &= 0 \\ \text{Cov}\{\dot{r}(\mathbf{x}), \dot{r}(\mathbf{x}')\} &= \sigma_r^2 \frac{d^2 \rho_r(|\mathbf{x} - \mathbf{x}'|)}{d\mathbf{x}_i d\mathbf{x}'_i} = -\sigma_r^2 \frac{d^2 \rho_r(\tau)}{d\tau^2}.\end{aligned}$$

The differential RF exists if the second derivative of $\rho_r(\tau)$ exists at $\tau = 0$, and further

$$\begin{aligned}\mu_{\dot{r}} &= 0 \\ \sigma_{\dot{r}}^2 &= -\sigma_r^2 \left. \frac{d^2 \rho_r(\tau)}{d\tau^2} \right|_{\tau=0} \\ \rho_{\dot{r}}(\tau) &= \left[\left. \frac{d^2 \rho_r(\tau)}{d\tau^2} \right|_{\tau=0} \right]^{-1} \times \frac{d^2 \rho_r(\tau)}{d\tau^2}.\end{aligned}$$

Furthermore, the cross-covariance of the filtered Gaussian RF with the basis Gaussian RF is

$$\text{Cov}\{\dot{r}(\mathbf{x}), r(\mathbf{x})\} = \sigma_r^2 \left. \frac{d\rho_r(\tau)}{d\tau} \right|_{\tau=0} = 0$$

because $\rho_r(\tau)$ must have a maximum at $\tau = 0$. Consequently, a Gaussian RF is independent of its derivative at an arbitrary location.

The presented linear family of RFs defines $\{[r(\mathbf{x}), r_\delta(\mathbf{x}), \dot{r}(\mathbf{x})]; \mathbf{x} \in D\}$ as a stationary trivariate Gaussian RF, and it may be used in the joint evaluation of the continuous spatial variable, the corresponding δ -averaged filter and the differential variables.

6.2.2 Model Validation

Consider a continuous spatial variable and assume that a set of exact observations from it is available, given as $\mathbf{r}^o : \{r(\mathbf{x}_i^o); \mathbf{x}_i^o \in D; i = 1, 2, \dots, n_o\}$. The task is to investigate whether a stationary Gaussian RF is a suitable model for the spatial variable.

We first evaluate spatial stationarity by defining the statistics

$$\left\{ \begin{array}{l} \hat{\mu}_B(\mathbf{x}) = \frac{1}{n_{B_x}} \times \sum_{i=1}^{n_o} I(\mathbf{x}_i^o \in B_x) \times r(\mathbf{x}_i^o); \mathbf{x} \in L^e \\ \hat{\sigma}_B^2(\mathbf{x}) = \frac{1}{n_{B_x}} \times \sum_{i=1}^{n_o} I(\mathbf{x}_i^o \in B_x) \times [r(\mathbf{x}_i^o) - \hat{\mu}_B(\mathbf{x})]^2; \mathbf{x} \in L^e \end{array} \right\},$$

where B_x refers to a ball of suitable size centred at $\mathbf{x} \in D$, n_{B_x} is the number of observations that is located within it and L^e is a grid of suitable density. The ball size and grid density must be adjusted to the density of observations. We may now evaluate contour maps of $\{\hat{\mu}_B(\mathbf{x}); \mathbf{x} \in L^e\}$ and $\{\hat{\sigma}_B^2(\mathbf{x}); \mathbf{x} \in L^e\}$ to determine whether the expectation or the variance exhibits spatial trends.

Define also the empirical variogram as

$$\left\{ \hat{\gamma}(\boldsymbol{\tau}) = \frac{1}{2n_{\Delta_\tau}} \times \sum_{(\mathbf{x}_i^o, \mathbf{x}_j^o) \in \Delta_\tau} [r(\mathbf{x}_i^o) - r(\mathbf{x}_j^o)]^2; \boldsymbol{\tau} \in L_\tau \right\},$$

where

$$\Delta_\tau : \{(\mathbf{x}_i^o, \mathbf{x}_j^o) | [\mathbf{x}_i^o - \mathbf{x}_j^o] \in [\boldsymbol{\tau} - \boldsymbol{\delta}, \boldsymbol{\tau} + \boldsymbol{\delta}]\}$$

with $\boldsymbol{\delta}$ representing some suitable tolerance in the distance and angle and L_τ being a suitable discretisation of $\boldsymbol{\tau}$. The degree of tolerance and discretisation must be adapted to the density of observations. A diagram of $\{\hat{\gamma}(\boldsymbol{\tau}); \boldsymbol{\tau} \in L_\tau\}$ can be evaluated to determine whether the surface becomes flat for large values of $|\boldsymbol{\tau}|$, which is a requirement for stationary continuous RFs.

Next, we validate Gaussianity. Define the estimate of the marginal pdf of the continuous RF, as

$$\hat{p}(r(\mathbf{x})) = \text{Hist}\{r(\mathbf{x}_i^o); i = 1, 2, \dots, n_o\}.$$

A plot of $\hat{p}(r(\mathbf{x}))$ is informative with respect to the number of modes and the degree of symmetry, as well as the heavy-tailedness and the peakedness of the marginal pdf. Stationary Gaussian RFs will have Gaussian marginal pdfs.

The sampling uncertainty and boundary effects must be considered in the final evaluation. A stationary Gaussian RF model is appropriate if spatial stationarity and marginal Gaussianity appear fulfilled. If spatial stationarity is apparent, but marginal Gaussianity is not, then either a transformation must be applied or a more complex prior model must be specified. If spatial stationarity is lacking, one should seek explanatory spatial variables such that the residual spatial variable exhibits spatial stationary and appears marginally Gaussian.

Alternatively, one may perform simulation-based Monte Carlo testing of a Gaussian RF model with a given parameter set $\boldsymbol{\theta}_{pG} = [\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$, for a specified significance level α , as described in Diggle and Ribeiro (2007). The procedure includes the simulation of a large set of realisations of the Gaussian RF with parameter $\boldsymbol{\theta}_{pG}$, providing the set $\mathbf{r}^s; s = 1, 2, \dots, n_s$. For each realisation \mathbf{r}^s , select the sub-realisation $\mathbf{r}^{so} : \{r^s(\mathbf{x}_i^o); \mathbf{x}_i^o \in D; i = 1, 2, \dots, n_o\}$ and calculate $(\hat{\mu}_B^s(\mathbf{x}), \hat{\sigma}_B^{2s}(\mathbf{x}), \hat{\gamma}^s(\boldsymbol{\tau}), \hat{p}^s(r(\mathbf{x}))$. The set of realisations defines a set of values for each of these statistics. Accept the Gaussian RF model with parameter $\boldsymbol{\theta}_{pG}$ if the observed statistics $[\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\rho}_r(\cdot)]$ fall within the empirical $(1 - \alpha)$ -confidence interval of the simulated ones.

6.2.3 Explanatory Spatial Variables

For the unknown spatial variable of interest $\{r(\mathbf{x}); \mathbf{x} \in D\}$, assume that a set of related spatial variables $\{g_j(\mathbf{x}); \mathbf{x} \in D\}; j = 1, 2, \dots, n_g$ is fully observed in the reference domain D . One may include these related spatial variables in the prior model as explanatory variables in a spatial regression setting. The parameters of the Gaussian RF are as follows:

$$\begin{aligned} \{\mu(\mathbf{x}) &= \mu_r^0 + \sum_{j=1}^{n_g} \beta_r^j g_j(\mathbf{x}); \mathbf{x} \in D\} \\ \{\sigma^2(\mathbf{x}) &= \sigma_r^2; \mathbf{x} \in D\} \\ \{\rho(\mathbf{x}, \mathbf{x}') &= \rho_r(\boldsymbol{\tau}; \boldsymbol{\eta}_r); \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'; \mathbf{x}, \mathbf{x}' \in D\} \end{aligned}$$

with the regression coefficients μ_r^0 and $\boldsymbol{\beta}_r = (\beta_r^1, \beta_r^2, \dots, \beta_r^{n_g})^T$. Note that whenever the explainability of the expectation is increased, one needs to reduce

the variance parameter accordingly. It can be demonstrated that this extended non-stationary Gaussian model is also a conjugate prior model for the Gauss-linear likelihood model. The corresponding Gaussian posterior model is developed in Sect. 7.1.

6.2.4 Related Topics

Due to its mathematical tractability, the Gaussian RF model is frequently specified as a prior model for continuous spatial variables. Several related topics are also in frequent use:

- Geostatistical modelling: The non-stationary regression-extended Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the model parameters $[\mu_r^0, \boldsymbol{\beta}_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ is closely related to the Kriging models in geostatistics, as discussed in Journel and Huijbregts (1978), Chiles and Delfiner (2012) and Sect. 10.2. In geostatistics, spatial evaluation is conventionally done using a Kriging predictor and conditional simulation.

The classical simple Kriging model is based on a continuous RF model with only the parameters for the two first moments $[\mu_r^0, \boldsymbol{\beta}_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ known. No Gaussian assumptions are imposed. The less assumptive universal Kriging model is also based on a continuous RF model as above, but only the parameters for the second moment $[\sigma_r^2, \rho_r(\boldsymbol{\tau})]$ are assumed to be known, whereas the parameters for the first moment $[\mu_r^0, \boldsymbol{\beta}_r]$ are considered unknown. Similarly, no Gaussian assumptions are imposed.

The Kriging predictor is defined to be the best linear unbiased (BLU) predictor, which, under the assumptions above, can be evaluated only in the least squared error (LSE) sense. In the universal Kriging case, certain unbiasedness constraints must be satisfied. The associated Kriging prediction variance can also be assessed. Note that neither the $(1-\alpha)$ prediction intervals nor realisations from conditional simulation can be provided since only second-order assumptions are made. Consequently, there is a long tradition in geostatistical practice to add the Gaussian assumption to obtain the prediction intervals and conditional realisations. These additional assumptions make the Kriging model identical to the Gaussian RF prior model used in Bayesian spatial modelling. Section 10.2 contains a more comprehensive discussion of geostatistical modelling.

- Entropy: The stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ aligns with the maximum-entropy RF with corresponding parameters, as discussed in Cressie and Wikle (2011). Consequently, the realisations under the two models are the smoothest continuous spatial variables with the given model parameter set.
- Markov property: The stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ cannot exhibit the Markov property in the traditional sense within the reference domain $D \in \mathbb{R}^3$. However, the discretised

Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in L \subset D\}$ represented on an orthogonal grid with a spatial correlation function defined by unidirectional exponential models in factorial form $\rho_r(\boldsymbol{\tau}) = \prod_{i=1}^3 \exp(-\alpha_i \tau_i)$ with $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3) \in \mathbb{R}_+^3$ will exhibit the Markovian characteristic. This characteristic appears as $p(r(x_{ijk}) | r(\mathbf{x}'); \mathbf{x}' \in L_{-(ijk)}) = p(r(x_{ijk}) | r(\mathbf{x}'); \mathbf{x}' \in \{x_{i\pm 1, j\pm 1, k\pm 1}\})$ and is referred to as the screening or Gaussian Markov effect. For more details, see Rue and Held (2005). If this non-isotropic grid-directional spatial correlation function is appropriate, the computational demands can be reduced because $[\Sigma_r^\rho]^{-1}$ will appear as a sparse matrix; see Rue and Held (2005). Section 10.3 contains a more comprehensive discussion of discretised Gaussian RF with Markovian characteristics.

- Frequency representation: The stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ is spatially parametrised in the correlation domain. Alternatively, the Gaussian RF can be spatially parametrised in the frequency domain using the spectrum parameter $\xi_r(\mathbf{v}); \mathbf{v} \in \mathbb{R}^3$. The relationship between these two parametrisations is, for $\mathbf{v} \in \mathbb{R}^3$, as follows:

$$\xi_r(\mathbf{v}) = [2\pi]^3 \int_{\boldsymbol{\tau} \in \mathbb{R}^3} \rho_r(\boldsymbol{\tau}) \exp(-\mathbf{v}^T \boldsymbol{\tau}) d\boldsymbol{\tau}.$$

For non-negative definite correlation functions $\rho_r(\boldsymbol{\tau})$, the corresponding spectrum appears as $\xi_r(\mathbf{v}) \geq 0; \mathbf{v} \in \mathbb{R}^3$, which is actually a characterising property for non-negative definiteness. Moreover, the spectrum appears with $\int \xi_r(\mathbf{v}) d\mathbf{v} = 1$. Thus, $\xi_r(\mathbf{v}); \mathbf{v} \in \mathbb{R}^3$ may be interpreted as a pdf, as given in Yaglom (1987). This re-parametrisation of the Gaussian RF can be used to provide extremely computationally efficient simulation algorithms based on fast Fourier transforms, as discussed in Borgman et al. (1984) and Buland et al. (2003).

- Fractal properties: The stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ cannot represent RFs with fractal characteristics like self-similarity and self-affinity. The latter, and larger, class requires that the spatial covariance function exhibit the property $\text{Cov}\{r(c\mathbf{x}), r(c\mathbf{x}')\} = \sigma_c^2 \text{Cov}\{r(\mathbf{x}), r(\mathbf{x}')\}; c \in \mathbb{R}_+$ for arbitrary $\sigma_c^2 \in \mathbb{R}_+$. This entails that the covariance function retains its shape after rescaling the reference variable \mathbf{x} . If $\sigma_c^2 = c$, the RF is denoted self-similar. This requirement cannot be fulfilled for Gaussian RFs because $\sigma_r^2 < \infty$ and $\rho_r(\boldsymbol{\tau}) \rightarrow 0$ as $|\boldsymbol{\tau}| \rightarrow \infty$, the latter being a consequence of the non-negative definiteness requirement.

One may, however, define an intrinsic Gaussian RF $\{r_I(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ with the difference RF $\{\Delta_h(\mathbf{x}) = [r_I(\mathbf{x}-\mathbf{h}/2) - r_I(\mathbf{x}+\mathbf{h}/2)]; \mathbf{x} \in D_h \subset \mathbb{R}^3\}$ a Gaussian RF for all $\mathbf{h} \in \mathbb{R}^3$ and D_h being D eroded by \mathbf{h} , as defined in Chiles and Delfiner (2012). Note that all stationary Gaussian RFs are intrinsic Gaussian RFs with expectation $E\{\Delta_h(\mathbf{x})\} = 0$ and variance $\text{Var}\{\Delta_h(\mathbf{x})\} = 2\sigma_r^2(1 - \rho_r(\mathbf{h}))$. However, the intrinsic Gaussian RF class is larger because the marginal pdfs need not be Gaussian. The spatial coupling parameter for the intrinsic Gaussian RF is termed the variogram function $\gamma_r(\mathbf{h}) = 1/2 \times \text{Var}\{\Delta_h(\mathbf{x})\} = 1/2 \times E\{[\Delta_h(\mathbf{x})]^2\}; \mathbf{h} \in \mathbb{R}^3$. Traditionally, the term semivariogram is used, but we use the shorter form. The

existence requirement is that $[-\gamma_r(\mathbf{h})]$ is conditionally non-negative definite, which corresponds to Definition 10 with the phrase “all $\boldsymbol{\alpha} \in \mathbb{R}^n$ ” replaced by “all $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\boldsymbol{\alpha}^T \mathbf{i}_n = 0$ ”. Note that stronger constraints on $\boldsymbol{\alpha}$ leave a larger class of functions, and $\gamma_r(\mathbf{h}) = |\mathbf{h}|^\nu; \nu \in \mathbb{R}_{[0,2]}$ is a valid variogram function. This variogram function fulfils $\gamma_r(c\mathbf{h}) = c^\nu \gamma_r(\mathbf{h})$, and hence $\{r_I(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ has the self-affine fractal property. Moreover, expressions for the fractal parameters can be developed as the Hurst exponent $H = \nu/2$ and fractal dimension $d = 3 - \nu/2$. Similarly, if $\gamma_r(\mathbf{h}) = |\mathbf{h}|$, then $\sigma_c^2 = c$ and the intrinsic Gaussian RF is self-similar, with Markovian properties like the factorial exponential correlation function.

Note that the intrinsic Gaussian RF $\{r_I(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ is not entirely probabilistically defined since the marginal pdf is left unspecified. Hence, the model as such cannot be used as a prior model in a Bayesian spatial modelling framework. Realisations from the model cannot be generated by simulation. If the marginal is assumed to be Gaussian, the RF can be demonstrated to be a Gaussian RF with the usual non-negative definite constraints on $[-\gamma_r(\mathbf{h})] = -\sigma_r^2(1 - \rho_r(\mathbf{h}))$. The associated full RF model cannot usually be developed if only an arbitrary marginal pdf is assumed. One may address this lack of full specification by obtaining one exact observation in the RF at an arbitrary location $\mathbf{x}_0 \in D$, i.e. $r(\mathbf{x}_0)$. Then, the conditional RF $\{[r(\mathbf{x}) \mid r(\mathbf{x}_0)]; \mathbf{x} \in D \subset \mathbb{R}^3\}$ is a Gaussian RF and completely probabilistically defined by the intrinsic Gaussian RF model.

- Multivariate Gaussian RF: Sometimes, we are interested in a set of spatial variables represented by the k -variate RF $\{\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), r_2(\mathbf{x}), \dots, r_k(\mathbf{x})); \mathbf{x} \in D\}$. The expectation levels then parametrise the corresponding stationary Gaussian RF prior model $\boldsymbol{\mu}_r = (\mu_{r1}, \mu_{r2}, \dots, \mu_{rk})^T \in \mathbb{R}^k$, the variance levels $\boldsymbol{\sigma}_r^2 = (\sigma_{r1}^2, \sigma_{r2}^2, \dots, \sigma_{rk}^2)^T \in \mathbb{R}_+^k$ and a non-negative definite set of spatial correlation and cross-correlation functions $\{\rho_{rij}(\boldsymbol{\tau}); i, j = 1, 2, \dots, k\}; \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$. The construction of such a set of non-negative definite functions is not trivial, as discussed in Gneiting et al. (2010).

A straightforward decomposition of the multivariate Gaussian RF model is possible, however:

$$\{\mathbf{r}(\mathbf{x}) = \boldsymbol{\mu}_r + \mathbf{W}^u \mathbf{u}(\mathbf{x}) + \mathbf{W}^n \mathbf{e}_r; \mathbf{x} \in D\},$$

where the k -variate centred, normalised Gaussian RF $\{\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}), \dots, u_k(\mathbf{x})); \mathbf{x} \in D\}$ has independent components with individual non-negative definite spatial correlation functions $\rho_{ui}(\boldsymbol{\tau}); \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'; i = 1, 2, \dots, k$. The Gaussian k -vector \mathbf{e}_r has independent, centred and normalised entries. The weight matrices, of dimension $(k \times k)$, are \mathbf{W}^u and \mathbf{W}^n . The former specifies each entry in $\mathbf{r}(\mathbf{x})$ to be a linear combination of the entries in $\mathbf{u}(\mathbf{x})$. In contrast, the latter is a diagonal matrix defining the nugget component in the spatial correlation function. The resulting k -variate RF $\{\mathbf{r}(\mathbf{x}); \mathbf{x} \in D\}$ is a Gaussian RF since it is defined as a linear combination of Gaussian RFs. Expressions for the corresponding model

parameters can easily be developed. The k -variate set of spatial covariance and cross-covariance functions is by construction non-negative definite.

- Multimodal RF: The marginal pdfs of Gaussian RF models are Gaussian pdfs, and thus unimodal and symmetric. However, many spatial variables have multimodal spatial histograms. An underlying unobservable mosaic spatial variable typically causes the modes. The Gaussian RF model can be generalised to represent such multimodal spatial phenomena. In Fjeldstad et al. (2021), a Gaussian mixture RF model is defined, and it is demonstrated that it exhibits conjugate properties. Alternatively, a selection-Gaussian RF model as discussed in Omre and Rimstad (2021) and Forberg et al. (2021) may be used to capture multimodal effects. This model also exhibits conjugate properties.

6.2.5 Example: Gaussian RF

The example discussed is introduced in Sect. 5.1 and Fig. 5.1. The unknown continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$ to be reconstructed is presented in display (a). The evaluation is made on the reference spatial variable in display (b), which is spatially discretised and represented by the n -vector \mathbf{r} . The available observations are presented in display (c), and a Gauss-linear likelihood model $p(\mathbf{d} | \mathbf{r})$ is specified.

A prior model $p(\mathbf{r})$ must be assigned to assess the reference spatial variable by Bayesian inversion. Let the prior model be a stationary Gaussian RF model with expectation and variance levels $\mu_r = 5.0$ and $\sigma_r^2 = 1.0$, respectively. The spatial dependence is defined by the spatial correlation function $\rho_r(\boldsymbol{\tau}; \eta_r) = \exp(-|\boldsymbol{\tau}|^{\eta_r})$; $\boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$ with $\eta_r = 1.8$. The prior model on the n -vector \mathbf{r} is then Gaussian with expectation n -vector $\boldsymbol{\mu}_r = 5.0\mathbf{i}_n$ and $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_r^\rho$. This matrix is defined by $\rho_r(\boldsymbol{\tau}; \eta_r)$. These model parameters define the Gaussian prior model $p(\mathbf{r})$.

Figure 6.3 contains the reference spatial variable to be assessed in display (a). The three displays (b), (c) and (d) present three independent realisations from the Gaussian RF prior model generated by the transformation algorithm in Sect. 10.1.

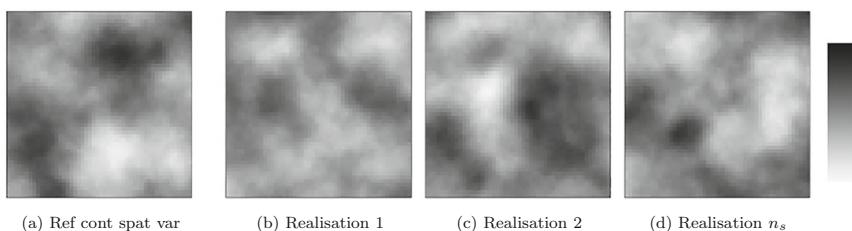


Fig. 6.3 Example Gaussian RF model. Gaussian prior model: (a) reference continuous spatial variable and (b), (c), (d) three independent realisations

Each of these realisations is considered to be a possible reference spatial variable. The realisations vary around the same level with the same variance and spatial smoothness but have different appearances and extrema in different locations.

Now, the Gauss-linear likelihood model $p(\mathbf{d} \mid \mathbf{r})$ is specified and the Gaussian prior model $p(\mathbf{r})$ is assigned. In a Bayesian inversion setting, as discussed in Chap. 2, these two models uniquely define the posterior model $p(\mathbf{r} \mid \mathbf{d})$.

6.3 Hierarchical Gaussian RF Models

The hierarchical Gaussian prior model is specified as a spatially stationary version of the hierarchical Gaussian model as defined in Sect. 4.1.2. The expectation is defined by $\{\mu(\mathbf{x}) = \mu_r; \mathbf{x} \in L\}$, which gives $\boldsymbol{\mu}_r = \mu_r \mathbf{i}_n$, and the spatial covariance function is defined by $\{\sigma(\mathbf{x})\sigma(\mathbf{x}')\rho(\mathbf{x}, \mathbf{x}') = \sigma_r^2 \rho_r(\mathbf{x} - \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in L\}$, which gives $\boldsymbol{\Sigma}_{r|\mu} = \sigma_r^2 \boldsymbol{\Sigma}_{r|\mu}^\rho$, as discussed in Røislien and Omre (2006). The spatial correlation function $\rho_r(\tau)$, which defines $\boldsymbol{\Sigma}_{r|\mu}^\rho$, is assumed to be known. The prior models for the other model parameters, $p(\mu_r \mid \sigma_r^2)$ and $p(\sigma_r^2)$, are assumed to be Gaussian and inverse-gamma pdfs with hyper-parameters $\mu_\mu \in \mathbb{R}$ and $\gamma_\mu \in \mathbb{R}_+$ for the former and $v_{\sigma^2} \in \mathbb{R}_+$ and $\xi_{\sigma^2} \in \mathbb{R}_+$ for the latter. The correlation function may also be parametrised as $\rho_r(\tau; \boldsymbol{\eta}_r)$, and the parameters in $\boldsymbol{\eta}_r$ can be considered random. A prior pdf $p(\boldsymbol{\eta}_r)$ must then be assigned to the hyper-parameters. By making this extension, however, the analytical tractability of the model is lost and computationally demanding McMC approaches must be used later in the process.

The discretised hierarchical Gaussian prior model $\{r_H(\mathbf{x}); \mathbf{x} \in L\}$, represented by the n -vector $\mathbf{r}_H \in \mathbb{R}^n$, is then defined by

$$\begin{aligned} [\mathbf{r}_H \mid \mu_r, \sigma_r^2] &\sim p(\mathbf{r}_H \mid \mu_r, \sigma_r^2) = \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \boldsymbol{\Sigma}_{r|\mu}^\rho) \\ [\mu_r \mid \sigma_r^2] &\sim p(\mu_r \mid \sigma_r^2) = \phi_1(\mu; \mu_\mu, \gamma_\mu \sigma_r^2) \\ \sigma_r^2 &\sim p(\sigma_r^2) = \left[\Gamma\left(\frac{v_{\sigma^2}}{2}\right) \right]^{-1} \left[\frac{v_{\sigma^2} \xi_{\sigma^2}^2}{2} \right]^{v_{\sigma^2}/2} \\ &\quad \times [\sigma_r^2]^{-(v_{\sigma^2}+2)/2} \times \exp\left(-\frac{v_{\sigma^2} \xi_{\sigma^2}^2}{2\sigma_r^2}\right). \end{aligned}$$

The prior model on the n -vector \mathbf{r}_H is obtained by marginalisation as

$$\begin{aligned} \mathbf{r}_H \sim p(\mathbf{r}_H) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}} p(\mathbf{r}_H \mid \mu_r, \sigma_r^2) p(\mu_r \mid \sigma_r^2) p(\sigma_r^2) d\mu_r d\sigma_r^2 \\ &= \left[\Gamma\left(\frac{v_{\sigma^2}}{2}\right) \right]^{-1} [v_{\sigma^2} \pi]^{-n/2} \Gamma\left(\frac{v_{\sigma^2} + n}{2}\right) \\ &\quad \times |\boldsymbol{\Omega}_{r_H}|^{-1/2} \times \left[1 + v_{\sigma^2}^{-1} (\mathbf{r}_H - \mu_\mu \mathbf{i}_n)^T \boldsymbol{\Omega}_{r_H}^{-1} (\mathbf{r}_H - \mu_\mu \mathbf{i}_n) \right]^{-(v_{\sigma^2}+n)/2} \end{aligned} \tag{6.4}$$

with spatial coupling matrix $\Omega_{r_H} = \xi_{\sigma^2}^2 [\gamma_\mu \mathbf{i}_n \mathbf{i}_n^T + \Sigma_{r|\mu}^\rho]$. The substitution $\theta = [\sigma_r^2]^{-1}$ and the integral identity

$$\int_{\mathbb{R}_+} \theta^\alpha \exp(-\theta\beta) d\theta = \Gamma(\alpha + 1)\beta^{-(\alpha+1)} \quad (6.5)$$

are used in the development of the expression above. This expression is an n -variate T-distribution with v_{σ^2} degrees of freedom, parametrised by $[\mu_\mu \mathbf{i}_n, \Omega_{r_H}, v_{\sigma^2}]$. The marginal pdf $p(r_{Hi})$ in an arbitrary grid node i is T-distributed with v_{σ^2} degrees of freedom with model parameters $[\mu_\mu, \xi_{\sigma^2}^2 (\gamma_\mu + 1), v_{\sigma^2}]$. These marginal pdfs are independent of location i and are thus invariant to translations in location. The bivariate pdf of (r_{Hi}, r_{Hj}) will not factorise into the product of the corresponding marginal pdfs as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$, because the global random model parameters $[\mu_r, \sigma_r^2]$ cause global spatial coupling. Consequently, the T-distributed RF is not ergodic.

A sequential algorithm provides realisations from the hierarchical Gaussian RF: first generate $\sigma_r^{2s} \sim p(\sigma_r^2)$, thereafter generate $\mu_r^s \sim p(\mu_r | \sigma_r^{2s})$ and lastly generate $\mathbf{r}_H^s \sim p(\mathbf{r}_H | \mu_r^s, \sigma_r^{2s})$. This procedure provides one realisation from the model, $\mathbf{r}_H^s \sim p(\mathbf{r}_H)$. If an ensemble of realisations $\{\mathbf{r}_H^s; s = 1, 2, \dots, n_s\}$ is generated, the marginal pdf represented by the location histogram will be according to a T-distribution with v_{σ^2} degrees of freedom parametrised by $[\mu_\mu, \xi_{\sigma^2}^2 (\gamma_\mu + 1), v_{\sigma^2}]$. The spatial heterogeneity represented by the spatial histogram in an arbitrary realisation will, however, be according to a Gaussian distribution parametrised by $[\mu_r^s, \sigma_r^{2s}]$, which are the actual realisations of the model parameters. This inconsistency between marginal variability and spatial heterogeneity may cause problems in applying the hierarchical Gaussian RF model.

The hierarchical Gaussian model can be extended to include explanatory spatial variables. For a more comprehensive discussion, see Røislien and Omre (2006).

6.4 Event Spatial Variables: Poisson RF Models

The class of Poisson RF models is defined in Sect. 4.2.1, with the corresponding class of spatially discretised Poisson count RF models in Expression (4.3). The class of misclassification likelihood models is defined in Sect. 5.2. Later, in Sect. 7.2, it is demonstrated that this pair of likelihood and prior model classes exhibits conjugate characteristics.

The Poisson RF prior model is usually defined on the spatially discretised event count representation $\mathbf{k} \in \mathbb{N}_\oplus^n$ as

$$\begin{aligned}\mathbf{k} \sim p(\mathbf{k}) &= \prod_{i=1}^n p(k_i) = \prod_{i=1}^n \frac{[\lambda_k^n]^{k_i}}{k_i!} \exp(-\lambda_k^n) \\ &= \frac{[\lambda_k^n]^{k_D}}{\prod_i k_i!} \times \exp(-n\lambda_k^n)\end{aligned}\quad (6.6)$$

with grid unit event intensity $\lambda_k^n = \lambda_k \Delta_n$ and total number of events $k_D = \sum_{i=1}^n k_i$. This model is a spatially discretised stationary Poisson count RF model with intensity level $\{\lambda(\mathbf{x}) = \lambda_k; \mathbf{x} \in D\}$ and base event set $\mathbb{E} = \emptyset$. We refer to Møller and Waagepetersen (2003) for a traditional introduction to Poisson RF models.

The corresponding approximate stationary Poisson event-location model, as defined in Sect. 4.2.1, is

$$\begin{aligned}\mathbb{X}_D^n = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) &\sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ &= \sum_{\mathbf{k} \in \mathbb{N}_\oplus^n} p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{k}) p(\mathbf{k}) \\ &= \prod_{i=1}^n \sum_{k_i \in \mathbb{N}_\oplus} \left[\frac{1}{\Delta_n} \right]^{k_i} \prod_{j=1}^{k_i} I(\mathbf{x}_{ij} \in \Delta_{ni}) p(k_i).\end{aligned}\quad (6.7)$$

This event-location set represents an exact stationary Poisson RF model because the intensity function is constant over D . Thus, λ_k^n is identical for all grid units. Every marginal pdf $p(k_i)$ is invariant with respect to translations on the grid L . Moreover, $p(k_i, k_j) = p(k_i)p(k_j)$ for every pair of grid nodes, causing the Poisson prior RF model to be ergodic. Consequently, the marginal variability and the spatial heterogeneity become identical in the extension limit.

Spatial Discretisation Uncertainty

The prior model is a stationary Poisson RF and is represented by a spatially discretised event count model $\{k_\Delta(\mathbf{x}); \mathbf{x} \in L\}$. Each event location is registered to the resolution of a grid unit volume Δ_n . For an event located at $\mathbf{x}^o \in D$ assigned to grid unit Δ_{ni^o} , the location of the event can only be retrieved from the event count model with the precision of a grid unit volume Δ_n . By infill asymptotic analysis with D constant and $n \rightarrow \infty$, hence $\Delta_n \rightarrow 0$, exact reproduction of the event location is obtained.

6.5 More on Poisson RF Prior Models

The general definition of a Poisson RF is presented in Definition 6. Consider the stationary Poisson RF $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ on which the spatially discretised Poisson prior model is based. The model parameters are the intensity level $\{\lambda(\mathbf{x}) = \lambda_k; \mathbf{x} \in D\}$ and an empty base event set $\mathbb{E} = \emptyset$. Let the corresponding event-location set of random length be $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}\}; \mathbf{x}_i \in D$.

6.5.1 Model Parameters and Characteristics

The vector of model parameters is $\boldsymbol{\theta}_{PP} = [\lambda_k, \emptyset]$ which contains the event intensity $\lambda_k \in \mathbb{R}_+$ and base event set $\mathbb{E} = \emptyset$. The corresponding intensity λ_{kA} for an arbitrary sub-domain $A \subset D$ is defined as $\lambda_{kA} = \int_A \lambda_k d\mathbf{x} = \lambda_k |A|$. From the definition of the Poisson RF, it follows that the event count in domain A , denoted k_A , is distributed according to a Poisson pdf for $k \in \mathbb{N}_+$, as

$$k_A \sim p(k_A = k) = \frac{[\lambda_k |A|]^k}{k!} \exp(-\lambda_k |A|).$$

Thus, the probability for no events in the domain A is

$$p(k_A = 0) = \exp(-\lambda_k |A|).$$

Consider a partition of the reference domain D into subsets $A_1, A_2, \dots, A_m \subset D; A_i \cap A_j = \emptyset; i, j = 1, 2, \dots, m; i \neq j$ and $\cup_{i=1}^m A_i = D$. Each subset is associated with one event count $k_{A_1}, k_{A_2}, \dots, k_{A_m} \in \mathbb{N}_+$. From the definition of the Poisson RF, the joint pdf for the event counts is in factorial form for $k_i \in \mathbb{N}_+$, as

$$\begin{aligned} p(k_{A_1} = k_1, k_{A_2} = k_2, \dots, k_{A_m} = k_m) &= \prod_{i=1}^m p(k_{A_i} = k_i) \\ &= \prod_{i=1}^m \frac{[\lambda_k |A_i|]^{k_i}}{k_i!} \times \exp(-\lambda_k |D|). \end{aligned}$$

Moreover, given that the total number of events in D equals k_D , the pdf for the spatial distribution of events among the partitioning sets is multinomial for $k_i \in \mathbb{N}_+; \sum_{i=1}^m k_i = k_D$, as

$$\begin{aligned} p(k_{A_1} = k_1, k_{A_2} = k_2, \dots, k_{A_m} = k_m \mid \sum_{i=1}^m k_i = k_D) \\ = k_D! \prod_{i=1}^m [k_i!]^{-1} \times \prod_{i=1}^m \left[\frac{|A_i|}{|D|} \right]^{k_i}. \end{aligned}$$

Inter-event Distance

Consider an arbitrary location $\mathbf{x}_0 \in D \subset \mathbb{R}^3$, and define the distance

$$t_{\mathbf{x}_0} = \min_{\mathbf{x} \in \mathbb{X}_D} \{|\mathbf{x}_0 - \mathbf{x}|\}.$$

This distance $t_{\mathbf{x}_0} \in \mathbb{R}_+$ between \mathbf{x}_0 and the nearest event location is a random variable, and its pdfs for $D \subset \mathbb{R}^3$ and $D \subset \mathbb{R}^2$ are displayed in Fig. 6.4. The associated probability is

$$\text{Prob}\{t_{\mathbf{x}_0} > t\} = p(k_{B_{\mathbf{x}_0}(t)} = 0) = \exp(-\lambda_k |B_{\mathbf{x}_0}(t)|),$$

where $B_{\mathbf{x}_0}(t)$ is a ball centred at \mathbf{x}_0 with radius t . Thus, $|B_{\mathbf{x}_0}(t)| = (4/3)\pi t^3$ if border effects due to the finiteness of domain D are ignored. The cdf for the distance $t_{\mathbf{x}_0}$ is then

$$t_{\mathbf{x}_0} \sim P(t_{\mathbf{x}_0}) = 1 - \exp\left(-\lambda_k \times \frac{4}{3}\pi t_{\mathbf{x}_0}^3\right),$$

and the corresponding pdf, as presented in display (a), is given by

$$t_{\mathbf{x}_0} \sim p(t_{\mathbf{x}_0}) = 4\lambda_k \pi t_{\mathbf{x}_0}^2 \times \exp\left(-\frac{4}{3}\lambda_k \pi t_{\mathbf{x}_0}^3\right),$$

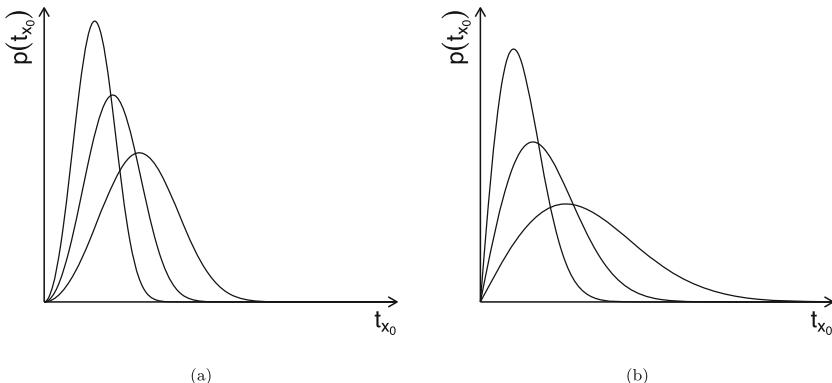


Fig. 6.4 Poisson RF model: pdf of distance between arbitrary location and closest event for varying intensity parameters λ_k : (a) for $D \subset \mathbb{R}^3$ and (b) for $D \subset \mathbb{R}^2$

with

$$\begin{aligned} E\{t_{x_0}\} &= \Gamma\left(\frac{4}{3}\right) \times \left[\frac{4}{3} \times \pi \lambda_k\right]^{-1/3} \\ \text{Var}\{t_{x_0}\} &= \left[\Gamma\left(\frac{5}{3}\right) - \Gamma^2\left(\frac{4}{3}\right) \right] \times \left[\frac{4}{3} \times \pi \lambda_k\right]^{-2/3}. \end{aligned}$$

Both the expected distance and the variance decrease with increasing intensity. The corresponding pdf for a Poisson RF on a two-dimensional reference domain $D \subset \mathbb{R}^2$ presented in display (b) is

$$t_{x_0} \sim p(t_{x_0}) = 2\lambda_k \pi t_{x_0} \times \exp(-\lambda_k \pi t_{x_0}^2),$$

with

$$\begin{aligned} E\{t_{x_0}\} &= \frac{1}{2} \lambda_k^{-1/2} \\ \text{Var}\{t_{x_0}\} &= [4 - \pi] \times [4\pi\lambda_k]^{-1}. \end{aligned}$$

Expressions for the probabilities of distance from an arbitrary location to the second-nearest event, third-nearest and so on can also be derived.

In the derivation above, the reference is assumed to be at an arbitrary location $x_0 \in D$. Alternatively, we may assume that the arbitrary reference location is $x_0 \in \mathbb{X}_D$, i.e. the location of an arbitrary event in the spatial variable. Expressions for the probability of the distance to the nearest of the other events can be developed as above. These calculations are somewhat more complicated, and they define the Palm distribution for the event RF, as defined in Møller and Waagepetersen (2003). It turns out that for stationary Poisson RFs these expressions are identical to the ones developed above. This identity is a characterising property of a stationary Poisson RF. It holds if and only if the event RF is a stationary Poisson RF.

Moments

Consider again the event count variable k_A for an arbitrary domain $A \subset D$. For a stationary Poisson RF model with pdf $p(k_A)$, it follows that

$$E\{k_A\} = \lambda_k |A|$$

$$\text{Var}\{k_A\} = \lambda_k |A|.$$

Then, consider the count variable k_B for $B \subset D$. It can be demonstrated that

$$\text{Cov}\{k_A, k_B\} = \lambda_k |A \cap B|.$$

Note that $\text{Cov}\{k_A, k_A\} = \lambda_k |A| = \text{Var}\{k_A\}$ and $\text{Cov}\{k_A, k_B\} = 0$ for $A \cap B = \emptyset$, as expected.

Descriptive Statistics

Define the interaction function, often referred to as the J -function, for $t \in \mathbb{R}_+$ as

$$J(t) = \frac{\mathbb{E}\{k_{B_{x_o}(t)} - 1\}}{|B_{x_o}(t) \cap D|},$$

where $B_{x_o}(t)$ is defined as above with $x_o \in \mathbb{X}_D$. The reduction of the numerator by one corrects for the fact that one event is given to occur at location x_o . This interaction function measures the clustering or repulsion effects present in the event RF. For the stationary Poisson RF, the interaction function for $t \in \mathbb{R}_+$ is

$$J(t) = \frac{\lambda_k |B_{x_o}(t) \cap D|}{|B_{x_o}(t) \cap D|} = \lambda_k.$$

For a stationary Poisson RF, the interaction function is constant and independent of t . In cases where the event RF exhibits clustering, the value of the interaction function is above average when t is small and below average for large t . Conversely, for a repulsive event RF, the pattern is reversed.

Alternatively, one may use the dimension-dependent L-interaction functions. For $q = 3$ and $q = 2$, they appear respectively, for $t \in \mathbb{R}_+$, as

$$\begin{aligned} L_3(t) &= \left[\frac{\mathbb{E}\{k_{B_{x_o}(t)} - 1\}}{\lambda_k 4/3\pi} \right]^{1/3} \approx J(t) \times t/\lambda_k \\ L_2(t) &= \left[\frac{\mathbb{E}\{k_{B_{x_o}(t)} - 1\}}{\lambda_k \pi} \right]^{1/2} \approx J(t) \times t/\lambda_k. \end{aligned}$$

These interaction functions for a stationary Poisson RF are $L_3(t) = t$ and $L_2(t) = t$.

6.5.2 Model Validation

Consider an event spatial variable and assume that representative observations of the event locations are available as $\mathbb{X}_D^o : \{x_1^o, x_2^o, \dots, x_{n_o}^o\}; x_i^o \in D$. The task is to justify whether a stationary Poisson RF is appropriate as a model for the spatial variable.

First, evaluate the spatial stationarity. Define the following statistic:

$$\left\{ \hat{\lambda}_B(\mathbf{x}) = \frac{1}{|B_{\mathbf{x}} \cap D|} \sum_{i=1}^{n_o} I(x_i^o \in B_{\mathbf{x}}); \mathbf{x} \in L^e \right\},$$

where $B_{\mathbf{x}}$ is a ball of suitable size centred at $\mathbf{x} \in D$ and L^e is a grid with suitable density. The ball size and the grid density must be adapted to the density of

observations. Generate a contour map based on $\{\hat{\lambda}_B(\mathbf{x}); \mathbf{x} \in L^e\}$ and evaluate whether there appears to be a spatial trend in the intensity.

Second, evaluate the Poisson assumption. Define the empirical interaction function

$$\left\{ \hat{J}(t) = \frac{1}{|B_{\mathbf{x}}(t) \cap D|} \frac{1}{n_o} \sum_{i=1}^{n_o} \left[\sum_{j=1}^{n_o} I(\mathbf{x}_j^o \in B_{\mathbf{x}_i^o}(t)) - 1 \right]; t \in T \right\},$$

where T defines a suitable discretisation of $t \in \mathbb{R}_+$. Generate a plot of the function $\{\hat{J}(t); t \in T\}$. Evaluate whether the function plot deviates significantly from being constant, which it should be for Poisson RF interactions. If the function plot appears above average for small t , it indicates a clustered event RF. If it appears below average for small t , it indicates a repulsive event RF.

In the final evaluation, sampling uncertainty and boundary effects must be considered. The stationary Poisson RF model is appropriate if both spatial stationarity and Poisson interaction are observed. If spatial stationarity is observed, but Poisson interaction cannot be justified, other stationary event RF models capturing clustering or repulsion should be considered, as discussed in Sect. 10.6. If spatial stationarity is lacking, one should seek explanatory spatial variables such that the residual spatial variable appears spatially stationary with Poisson interaction.

Alternatively, simulation-based Monte Carlo testing of a Poisson RF model with given parameters $\theta_{PP} = [\lambda_k, \emptyset]$, with significance level α , can be performed, as discussed in Diggle and Ribeiro (2007). Simulate a large set of realisations of the Poisson RF with parameter θ_{PP} , $[X_D | k_D = n_o]^s; s = 1, 2, \dots, n_s$. For each realisation, calculate estimates for the parameters $\hat{\lambda}_B^s(\mathbf{x})$ and $\hat{J}^s(t)$. The set of realisations provides a set of simulated values for each of these statistics. Accept the Poisson RF model with parameter θ_{PP} if the statistics based on the observations $[\hat{\lambda}_B(\mathbf{x}), \hat{J}(t)]$ fall within the empirical $(1 - \alpha)$ -confidence interval based on the realisations. The test is conditional on the number of events in the Poisson RF being $k_D = n_o$.

6.5.3 Explanatory Spatial Variables

The unknown spatial variable of interest is $\{k(\mathbf{x}); \mathbf{x} \in D\}$. Assume that a set of related spatial variables $\{g_j(\mathbf{x}); \mathbf{x} \in D\}; j = 1, 2, \dots, n_g$ are observed in every location in the reference domain D . These related spatial variables, or covariates, can be included into the prior model as explanatory variables in a spatial regression

setting. The intensity parameter of the Poisson RF can be expressed as

$$\left\{ \lambda(\mathbf{x}) = \lambda_k^0 \times \prod_{j=1}^{n_g} \exp(\beta_k^j g_j(\mathbf{x})); \mathbf{x} \in D \right\}.$$

Denote the regression coefficients $\boldsymbol{\beta}_k = (\beta_k^1, \beta_k^2, \dots, \beta_k^{n_g})^T$. It can be demonstrated that this extended non-stationary Poisson model is also a conjugate prior model for the misclassification likelihood model. The corresponding Poisson posterior model is developed in Sect. 7.2.

6.5.4 Related Topics

The Poisson RF model is considered a reference model for event spatial variables. Other important topics related to the Poisson RF model are:

- Binomial event models: The binomial event RF model is defined as a spatially discretised event count RF represented by the n -vector $\mathbf{k}_B \in \mathbb{N}_{\oplus}^n$. The model parameters are the population size n -vector $\boldsymbol{\kappa}_P \in \mathbb{N}_{\oplus}^n$ and the marking probability $\pi_B \in \mathbb{R}_{[0,1]}$. Diggle and Ribeiro (2007) contains a thorough discussion of the model. The binomial event RF model is defined for $\mathbf{k}_B \leq \boldsymbol{\kappa}_P$ as

$$\begin{aligned} \mathbf{k}_B \sim p(\mathbf{k}_B) &= \prod_{i=1}^n p(k_{Bi}) \\ &= \prod_{i=1}^n \frac{\kappa_{Pi}!}{(\kappa_{Pi} - k_{Bi})! k_{Bi}!} \pi_B^{k_{Bi}} (1 - \pi_B)^{\kappa_{Pi} - k_{Bi}}. \end{aligned}$$

Note that the pdf is in factorial form. Thus, the event counts are independent from one grid unit to another. The event count in a grid unit is the number of marked events if each member in the event population κ_{Pi} is marked with probability π_B . Recall that the binomial pdf approaches the Poisson pdf as the population size increases and the marking probability decreases, whereas their product is constant and identical to the intensity. Consequently, the binomial event RF approaches the Poisson event count RF in this limit. Unfortunately, no known likelihood model for the observations has the binomial event RF as a conjugate model.

A hierarchical binomial event RF model can be defined along the lines of the log-Gaussian Cox model discussed in Sect. 6.6. The marking probabilities are considered a spatially discretised RF, and a logistic transformation is assigned a Gaussian RF model. Thus, the probability model is a spatially discretised

logit-Gaussian RF model. The assessment of this hierarchical RF model faces computational challenges similar to the ones for the log-Gaussian Cox model.

- Cluster/repulsive event models: The stationary Poisson RF $\{k(\mathbf{x}); \mathbf{x} \in D\}$ with model parameters $[\lambda_k, \emptyset]$ exhibit neither event clustering nor event repulsion. In the Poisson RF model, the events are independently and uniformly located in D , given the number of events k_D . The corresponding interaction function is $J(t) = \lambda_k$. Thus, it takes a constant value and is independent of t . In practice, most event spatial variables appear with either clustering or repulsion. A comprehensive discussion of cluster and repulsive event models is found in Illian et al. (2008), and Sect. 10.6 contains further discussion on both model types and simulation algorithms.

If the events are clustered, $J(t)$ appears above average for small t . The typical models for event clustering define underlying hidden RF models representing the intensity function $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}$ in a non-stationary Poisson RF. Two such hidden RF models are frequently applied:

- Cox event models assign a stationary log-Gaussian RF to the intensity field $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$, with parameters $(\mu_\lambda, \sigma_\lambda^2, \rho_\lambda(\tau))$. The marginal event RF $\{k_C(\mathbf{x}); \mathbf{x} \in D\}$ appears with event clustering.
- Neyman–Scott (mother–child) event models employ a stationary mother Poisson RF model with intensity λ_M and an associated kernel function centred at the locations of mothers to generate the intensity function $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$. The marginal event RF $\{k_{NS}(\mathbf{x}); \mathbf{x} \in D\}$ appears with event clustering.

Both the Cox and Neyman–Scott models are primarily evaluated by simulation, and realisations can easily be generated by the sequential algorithm, as defined in Sect. 10.6.

If the events are repulsive, $J(t)$ appears below average for small t . The commonly used model for event repulsion is the Strauss model.

- Strauss event models are defined for a given number of events $k_D = k$. The pdf for the event locations is typically defined as $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \mid k_D = k) = \text{const} \times \prod_{i,j=1; i \neq j}^k \exp(-\varphi_s(\mathbf{x}_i - \mathbf{x}_j))$, where $\varphi_s(\tau) \in \mathbb{R}_+$ is a pairwise repulsion potential that penalises events for being located near each other. Consequently, $\varphi_s(\tau)$ is a declining function with increasing $|\tau|$. The resulting event RF $\{k_S(\mathbf{x}); \mathbf{x} \in D\}$ appears with event repulsion.

The Strauss model can only be evaluated by simulation, and realisations for $k_D = k$ can be generated by the iterative McMC algorithm, as defined in Sect. 10.6. If, in addition, the number of events k_D is random with pdf $p(k_D = k)$, the iterative reversible-jump McMC algorithm, as defined in Green (1995), can be applied.

- Markov properties: The Markov property of event RF models is defined by the intensity of an event to occur in a specific location $\mathbf{x}_o \in D$ given the event locations in the remains of the domain. An event RF model with a Markov range

$\delta \in \mathbb{R}_+$ possesses the following characteristic, according to Ripley and Kelly (1977):

$$[\lambda(\mathbf{x}_o) \mid \mathbb{X}_D \setminus \mathbf{x}_o] = [\lambda(\mathbf{x}_o) \mid \{\mathbf{x} \mid \mathbf{x} \in \mathbb{X}_D \setminus \mathbf{x}_o; |\mathbf{x} - \mathbf{x}_o| \leq \delta\}].$$

This holds for all $\mathbf{x}_o \in D$ where the set $\mathbb{X}_D \setminus \mathbf{x}_o$ contains the realisation of the event locations in the domain D excluding a possible event in location \mathbf{x}_o . Consequently, the intensity depends only on the event locations within a distance δ of \mathbf{x}_o .

A stationary Poisson RF model appears as an extreme Markov event model with range $\delta = 0.0$. The intensity in a specific location is independent of the number and locations of the other events in the realisation. The Strauss event RF model with repulsion potential $\varphi_s(\tau) = 0.0$ for $|\tau| > \delta$ will have Markov characteristics with range δ .

- **Marked event RF:** The stationary Poisson RF $\{k(\mathbf{x}); \mathbf{x} \in D\}$ with model parameters $[\lambda_k, \emptyset]$ defines the pdf of the event-location set $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}\}$. In practice, these events may have other characteristics of interest, such as type, shape or orientation. These characteristics, or marks, for event i are denoted $\mathbf{m}_i = (m_{i1}, m_{i2}, \dots, m_{in_m}) \in \Omega_1 \times \Omega_2 \times \dots \times \Omega_{n_m}$. A pdf can be assigned to these marks. The marked event-location set is defined as $\mathbb{X}_D^M : \{(\mathbf{x}_1, \mathbf{m}_1), (\mathbf{x}_2, \mathbf{m}_2), \dots, (\mathbf{x}_{k_D}, \mathbf{m}_{k_D})\}$, as specified in Illian et al. (2008). The mark pdf may depend on the location in D . The sequential algorithm can generate realisations from marked Poisson RF models.

6.5.5 Example: Poisson RF

The example discussed is introduced in Sect. 5.2.1 and Fig. 5.2. The unknown event spatial variable $\{k(\mathbf{x}); \mathbf{x} \in D\}$ to be reconstructed is presented in display (a), with corresponding event count spatial variable, represented by the n -vector \mathbf{k} , in display (b). The available observations \mathbf{d} are presented in displays (g) and (h), and the misclassification likelihood model $p(\mathbf{d} \mid \mathbf{k})$ is defined.

A stationary Poisson RF prior model is assigned, with intensity parameter $\lambda_k = 30$, entailing that the expected number of events in the domain is 187.5. That is, the prior model $p(\mathbf{k})$ on the spatially discretised event count spatial variable is a Poisson pdf.

Figure 6.5 contains the reference spatial variable presented in event-location and event count format in displays (a) and (e), respectively. Displays (b), (c), (d) and (f), (g), (h) contain three independent realisations in each of the two formats. Each realisation is considered a possible outcome of the reference spatial variable. These realisations have similar event densities, and the events are relatively evenly distributed in the domain.

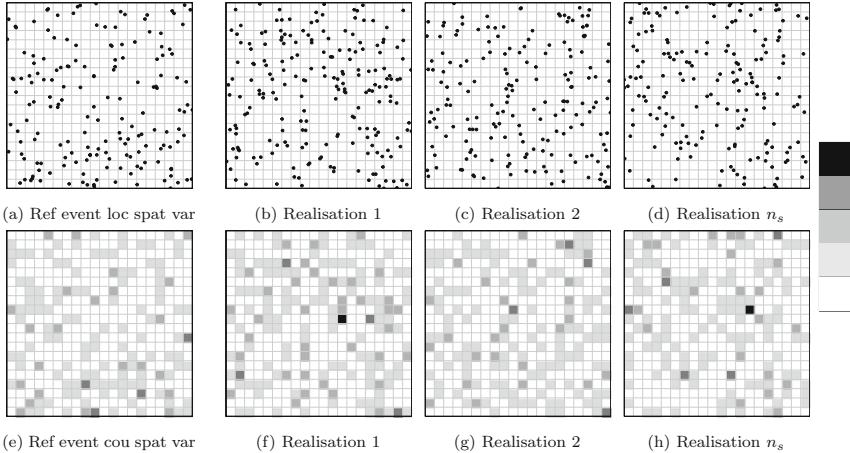


Fig. 6.5 Example Poisson RF model. Poisson prior model: **(a)** reference event-location spatial variable; **(b), (c), (d)** three independent realisations; **(e)** reference event count spatial variable; and **(f), (g), (h)** three independent realisations

Now, the misclassification likelihood model $p(\mathbf{d} \mid \mathbf{k})$ is specified, and the Poisson prior model $p(\mathbf{k})$ is assigned. Consequently, the corresponding posterior model $p(\mathbf{k} \mid \mathbf{d})$ is defined; this model constitutes the ultimate solution to the Bayesian inversion problem.

6.6 Hierarchical Poisson RF Models

The hierarchical Poisson count prior model is specified to be a spatially stationary version of the hierarchical Poisson count model, as defined in Sect. 4.2.2. The relative intensity n -vector is defined to be $\beta_\lambda = i_n$. The associated random variable for intensity-level $\lambda_k \in \mathbb{R}_+$ is assigned a gamma prior pdf $p(\lambda_k)$ with hyperparameters $v_\lambda \in \mathbb{R}_+$ and $\xi_\lambda \in \mathbb{R}_+$. Furthermore, the base event set is defined to be empty, $\mathbb{E} = \emptyset$. The prior spatially discretised hierarchical Poisson count model $\{k_H(\mathbf{x}); \mathbf{x} \in \mathbb{L}\}$, represented by the n -vector $\mathbf{k}_H \in \mathbb{N}_+^n$, is then defined by

$$\begin{aligned} [\mathbf{k}_H \mid \lambda_k] &\sim p(\mathbf{k}_H \mid \lambda_k) = \prod_{i=1}^n \frac{[\lambda_k \Delta_n]^{k_{Hi}}}{k_{Hi}!} \exp(-\lambda_k \Delta_n) \\ \lambda_k &\sim p(\lambda_k) = [\Gamma(v_\lambda)]^{-1} \xi_\lambda^{v_\lambda} \lambda_k^{v_\lambda - 1} \exp(-\xi_\lambda \lambda_k). \end{aligned}$$

The prior model for the n -vector \mathbf{k}_H is obtained by marginalisation as

$$\begin{aligned}\mathbf{k}_H \sim p(\mathbf{k}_H) &= \int_{\mathbb{R}_{+}^n} p(\mathbf{k}_H | \lambda_k) p(\lambda_k) d\lambda_k \\ &= [\Gamma(v_\lambda)]^{-1} \Gamma \left(\sum_i k_{Hi} + v_\lambda \right) \kappa_\lambda^{v_\lambda} \prod_{i=1}^n \frac{[(1 - \kappa_\lambda)/n]^{k_{Hi}}}{k_{Hi}!},\end{aligned}\quad (6.8)$$

where $\kappa_\lambda = [1 + \xi_\lambda^{-1} |\mathcal{D}|]^{-1} < 1$. In the development of the expression, the integral identity in Expression (6.5) is used.

This expression is, for integer-valued v_λ , a spatially stationary negative-multinomial pdf parametrised by $(\kappa_\lambda, n^{-1} \mathbf{i}_n, v_\lambda)$. Due to the symmetry, the marginal pdfs $p(k_{Hi})$ are identical and negative-binomially distributed. The bivariate pdf of (k_{Hi}, k_{Hj}) does not factorise into the marginal pdfs as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$, however, because the global random model parameter λ_k causes global spatial coupling. Consequently, this negative-binomial RF is not ergodic.

Simulation of realisations from the hierarchical Poisson count RF can be made by a sequential procedure. First, generate $\lambda_k^s \sim p(\lambda_k)$ and thereafter generate $\mathbf{k}_H^s \sim p(\mathbf{k}_H | \lambda_k^s)$. This procedure provides one realisation $\mathbf{k}_H^s \sim p(\mathbf{k}_H)$. If an ensemble of realisations $\{\mathbf{k}_H^s; s = 1, 2, \dots, n_s\}$ is generated, the marginal variability represented by the location histogram will be according to a negative-binomial pdf. However, the spatial heterogeneity represented by the spatial histogram in an arbitrary realisation will be according to a Poisson pdf with parameter λ_k^s , which is the actual realisation of the model parameter. This inconsistency between marginal variability and spatial heterogeneity may cause problems in applications of the hierarchical Poisson count RF model.

An alternative, ergodic hierarchical Poisson count RF can be defined by letting the spatial intensity variable be an ergodic, continuous non-negative RF, such as a log-Gaussian RF, as discussed in Møller et al. (1998). Let $\{\lambda(\mathbf{x}); \mathbf{x} \in L\}$ be a discretised log-Gaussian RF, represented by the n -vector λ . The discretised hierarchical Poisson count RF, also called a log-Gaussian Cox RF, is then defined by

$$\begin{aligned}[\mathbf{k}_H | \lambda] \sim p(\mathbf{k}_H | \lambda) &= \prod_{i=1}^n \frac{[\lambda_i \Delta_n]^{k_{Hi}}}{k_{Hi}!} \exp(-\lambda_i \Delta_n) \\ \lambda \sim p(\lambda) &= \log \phi_n(\lambda; \beta_\lambda \mathbf{i}_n, \gamma_\lambda \Sigma_\lambda^\rho).\end{aligned}$$

The hyper-parameters are $\beta_\lambda \in \mathbb{R}$, $\gamma_\lambda \in \mathbb{R}_{+}$, and the spatial coupling matrix Σ_λ^ρ is defined by the correlation function $\rho_\lambda(\tau)$. The parametrisation is then $[\beta_\lambda, \gamma_\lambda, \rho_\lambda(\tau)]$. These expressions define the discretised hierarchical Poisson count RF $\{k_H(\mathbf{x}); \mathbf{x} \in L\}$

$$\mathbf{k}_H \sim p(\mathbf{k}_H) = \int_{\mathbb{R}_{+}^n} p(\mathbf{k}_H | \lambda) p(\lambda) d\lambda. \quad (6.9)$$

This expression does not have a nice analytical solution but can be simulated sequentially similarly to the hierarchical Poisson model defined above. The resulting realisations will appear as clustered event spatial variables.

Since the log-Gaussian RF is ergodic, the hierarchical Poisson count model is also ergodic. This model, known as the log-Gaussian Cox model, is frequently applied in practice. In Sects. 9.2 and 10.7, approximations to this model are further discussed.

6.7 Mosaic Spatial Variables: Markov RF Models

The class of Markov RF models is defined in Sect. 4.3.1 and in Expression (4.8). Section 5.3 defines the class of response likelihood models. Later, Sect. 7.3 demonstrates that this pair of likelihood and prior model classes has conjugate characteristics.

The Markov RF prior model is defined on the spatially discretised representation $\mathbf{l} \in \mathbb{L}^n$ with the Gibbs formulation, as

$$\mathbf{l} \sim p(\mathbf{l}) = \text{const} \times \prod_{i=1}^n v_{0l}(l_i) \times \prod_{\mathbf{c}' \in \mathbf{c}_L^l} v_{1l}(l_j; j \in \mathbf{c}') \quad (6.10)$$

with clique system \mathbf{c}_L^l where all cliques have identical geometry, except for boundary effects. Moreover, the proportion and interaction functions, $v_{0l}(\cdot) \in \mathbb{R}_+$ and $v_{1l}(\cdot) \in \mathbb{R}_+$, are spatially stationary. Consequently, the Markov RF model is approximately spatially stationary. The normalising constant is usually not feasible to calculate because it requires summation over $\mathbf{l} \in \mathbb{L}^n$. Besag (1974) is the classical reference for Markov RF models.

The corresponding Markov formulation, for $i = 1, 2, \dots, n$, is

$$[l_i | \mathbf{l}_{-i}] \sim p(l_i | \mathbf{l}_{-i}) = p(l_i | l_j; j \in \mathbf{n}_i^l) = \text{const} \times v_{0l}(l_i) \times \omega_l(l_i | l_j; j \in \mathbf{n}_i^l)$$

with neighbourhood system \mathbf{n}_L^l and interaction function $\omega_l(\cdot | \cdot) \in \mathbb{R}_+$, both spatially stationary, except for boundary effects. The normalising constant is feasible to calculate because it requires summation over $l_i \in \mathbb{L}$ only.

The marginal pdfs are invariant with respect to translations on the grid L , except for boundary effects. Also, the bivariate marginal pdfs $p(l_i, l_j)$ are translation-invariant. Furthermore, the Markov property entails that the bivariate marginal pdf $p(l_i, l_j) \rightarrow p(l_i)p(l_j)$ as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$. Consequently, the bivariate marginal pdf factorises in the limit, which entails that the Markov prior model exhibits ergodicity.

The corresponding expression for the particular case of the Markov RF model with the two-nearest-node clique system with different horizontal and vertical couplings is

$$\mathbf{l} \sim p(\mathbf{l}) = \text{const} \times \prod_{i=1}^n v_{0l}(l_i) \times \prod_{\substack{h < i, j \\ i, j \in L}} v_{1lh}(l_i, l_j) \times \prod_{\substack{v < i, j \\ i, j \in L}} v_{1lv}(l_i, l_j),$$

where $h < \cdot, \cdot >$ and $v < \cdot, \cdot >$ are the nearest nodes in the horizontal plane and the vertical direction, respectively. The functions $v_{1lh}(\cdot, \cdot) \in \mathbb{R}_+$ and $v_{1lv}(\cdot, \cdot) \in \mathbb{R}_+$ are the Gibbs interaction functions for the horizontal plane and the vertical dimension, respectively. The interaction functions take $[(n_L \times (n_L + 1))/2]$ values for the corresponding combinations of outcomes for (l_i, l_j) .

The corresponding Markov formulation, for $i = 1, 2, \dots, n$, is

$$\begin{aligned} [l_i | \mathbf{l}_{-i}] \sim p(l_i | \mathbf{l}_{-i}) &= \text{const} \times v_{0l}(l_i) \times \omega_{lh}(l_i | l_j; h < i, j >, j \in L) \\ &\quad \times \omega_{lv}(l_i | l_j; v < i, j >, j \in L), \end{aligned}$$

where the Markov interaction function $\omega_l(\cdot | \cdot) \in \mathbb{R}_+$ factorises into a horizontal and a vertical factor.

Spatial Discretisation Uncertainty

The prior model is a stationary Markov RF defined on a spatially discretised grid $\{l(\mathbf{x}); \mathbf{x} \in L\}$. A common approach to achieve a spatial representation is to assign each grid unit the label that is registered at the corresponding grid node. This approach accurately reproduces the label proportions without biases; however, it can cause other distortions of the mosaic spatial variable. The label bodies, defined as spatially closed volumes of the same label, will appear increasingly larger with decreasing grid density. The body numbers will, however, decline such that the label proportions remain constant.

6.8 More on Markov RF Prior Models

The general definition of a Markov RF is presented in Definition 9. Consider the stationary Markov RF $\{l(\mathbf{x}); \mathbf{x} \in L\}$ which corresponds to the Markov RF prior model, as discussed in Sect. 6.7. The model parameters are the clique designs in the clique system $\mathbf{c}^l \in \mathbf{c}_L^l$ and the proportion and interaction functions $v_{0l}(\cdot)$ and $v_{1l}(\cdot)$.

6.8.1 Model Parameters and Characteristics

The model parameters are $\boldsymbol{\theta}_{pM} = [\mathbf{c}^l, v_{0l}(\cdot), v_{1l}(\cdot)]$ with clique design \mathbf{c}^l usually being a convex subset of grid nodes. This subset can, for example, be one node with the four nearest nodes in the horizontal plane and the two nearest nodes in the vertical direction, resulting in a clique design of the seven nearest nodes. The proportion and interaction functions $v_{0l}(\cdot), v_{1l}(\cdot) \in \mathbb{R}_+$ are defined for discrete arguments associated with the categorical label set \mathbb{L} . These parameters represent the Gibbs formulation for the RF, and according to the Hammersley–Clifford theorem, there exists a corresponding Markov formulation with neighbourhood design \mathbf{n}^l and coupling function $\omega_l(\cdot \mid \cdot) \in \mathbb{R}_+$ uniquely defined by the Gibbs formulation parameters. To develop the links between the two parameter sets for an arbitrary node i , define the subset of the clique system \mathbf{c}_L^l that includes node i as follows

$$\mathbf{c}_{L|i}^l : \{\mathbf{c} \mid \mathbf{c} \in \mathbf{c}_L^l; i \in \mathbf{c}\}.$$

The Markov formulation of the Markov RF model is

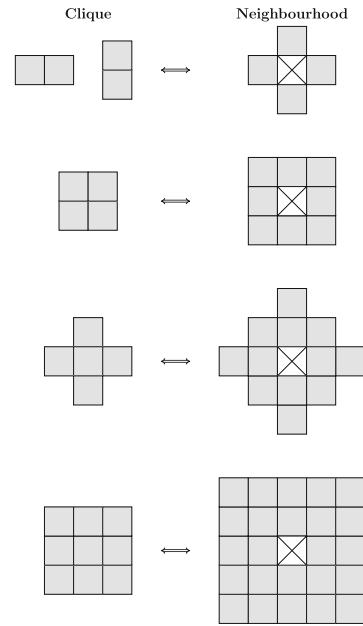
$$\begin{aligned} [l_i \mid \mathbf{l}_{-i}] &\sim p(l_i \mid \mathbf{l}_{-i}) = \frac{p(\mathbf{l})}{\sum_{l'_i \in \mathbb{L}} p(l'_i, \mathbf{l}_{-i})} \\ &= \frac{\text{const} \times \prod_{j=1, j \neq i}^n v_{0l}(l_j) \times \prod_{\mathbf{c} \in \mathbf{c}_L^l \setminus \mathbf{c}_{L|i}^l} v_{1l}(l_j; j \in \mathbf{c})}{\text{const} \times \prod_{j=1, j \neq i}^n v_{0l}(l_j) \times \prod_{\mathbf{c} \in \mathbf{c}_L^l \setminus \mathbf{c}_{L|i}^l} v_{1l}(l_j; j \in \mathbf{c})} \\ &\quad \times \frac{v_{0l}(l_i) \times \prod_{\mathbf{c} \in \mathbf{c}_{L|i}^l} v_{1l}(l_j; j \in \mathbf{c})}{\sum_{l'_i \in \mathbb{L}} \left[v_{0l}(l'_i) \times \prod_{\mathbf{c} \in \mathbf{c}_{L|i}^l} v_{1l}(l'_i, l_j; j \in \mathbf{c} \setminus i) \right]} \\ &= \left[\sum_{l'_i \in \mathbb{L}} [v_{0l}(l'_i) \times \prod_{\mathbf{c} \in \mathbf{c}_{L|i}^l} v_{1l}(l'_i, l_j; j \in \mathbf{c} \setminus i)] \right]^{-1} \times v_{0l}(l_i) \times \prod_{\mathbf{c} \in \mathbf{c}_{L|i}^l} v_{1l}(l_j; j \in \mathbf{c}) \\ &= \left[\sum_{l'_i \in \mathbb{L}} [v_{0l}(l'_i) \times \omega_l(l'_i \mid l_j; j \in \mathbf{n}_i^l)] \right]^{-1} \times v_{0l}(l_i) \times \omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l). \end{aligned} \tag{6.11}$$

Furthermore, the neighbourhood of node i is defined as

$$\mathbf{n}_i^l = \bigcup_{\mathbf{c} \in \mathbf{c}_{L|i}^l} \{j \mid j \in \mathbf{c}\} \setminus i. \tag{6.12}$$

Fig. 6.6 Markov RF model.

Examples of corresponding clique and neighbourhood designs



Consequently, $i \notin \mathbf{n}_i^l$, and because the geometry of \mathbf{c}^l is stationary on L for a stationary Markov RF, the geometry of \mathbf{n}^l is also stationary on L , except for boundary effects. Moreover, the relation of the coupling functions in node i is

$$\omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l) = \text{const} \times \prod_{\mathbf{c} \in \mathbf{c}_{L|i}^l} v_{1l}(l_j; j \in \mathbf{c}).$$

Figure 6.6 presents the relation between clique design and neighbourhood design for some simple two-dimensional designs. Because the arguments of the two interaction functions are discrete, so are their functional values. Consequently, they can be parametrised by a finite set of values. For the simplest isotropic, two-nearest-node clique design, the number of parameters required in the interaction function $v_{1l}(\cdot)$ is $[n_L \times (n_L + 1)]/2 - 1$, with the final (-1) accounting for the normalisation by the constant. Therefore, the number of required parameters is two for a {black, white} model.

Ising Model

The Ising model, as introduced in Ising (1925), has $l_i \in \{\text{black, white}\}$ with \mathbf{c}^l containing the two nearest nodes, as discussed in Besag (1974). The proportion and interaction functions are for the grid node references $i, j \in L$

$$\begin{aligned} v_{0l}(l_i) &= \text{const} \\ v_{1l}(l_i, l_j; \langle i, j \rangle) &= \beta^{I(l_i=l_j)}, \end{aligned}$$

where $\langle i, j \rangle$ represents pairs of nearest nodes in the grid and the interaction parameter is $\beta \in \mathbb{R}_\oplus$. The model implicitly assumes symmetry between the black and white labels; hence, there is only one model parameter. This Ising model is frequently studied, particularly in quantum physics with black and white representing negative and positive electron charges.

Realisations from the Ising model appear very different for different values of β . For large β , there is an overwhelming probability of having a realisation $\{l(\mathbf{x}); \mathbf{x} \in L\}$ either almost all black or all white. These realisations appear from a pdf with two distinct modes situated at the extremes of the sample space. On the other hand, for β close to one, the realisation $\{l(\mathbf{x}); \mathbf{x} \in L\}$ appears with random mixing of black and white almost without spatial coupling. These realisations appear from a pdf with a single mode situated centrally in the sample space. A distinctive characteristic of the Ising model is the abrupt transition between these two regimes, which occurs at a particular value of β known as the critical temperature β_c . This flipping effect makes the model less attractive from a spatial statistics point of view because, in practice, only two regimes occur; either almost entirely one label, or a random mixture of both labels. Somewhat surprisingly, however, the first regime of the Ising model has proven to be very well suited as a prior model for Bayesian spatial modelling in image analysis. It provides a regularising spatial effect on the posterior model, as demonstrated in Besag (1974).

6.8.2 Model Validation

Consider a mosaic spatial variable and assume that a representative set of exact observations $\mathbf{l}^o = \{l(\mathbf{x}_i^o); \mathbf{x}_i^o \in D; i = 1, 2, \dots, n_o\}$ is available. The task is to evaluate whether a stationary Markov RF is appropriate as a model for the spatial variable.

First, evaluate spatial stationarity. Define the statistic for each label $l \in \mathbb{L}$ as

$$\left\{ \hat{\nu}_{0XB}(l) = \frac{1}{n_{Bx}} \sum_{i=1}^{n_o} I(\mathbf{x}_i^o \in B_x) \times I(l(\mathbf{x}_i^o) = l); \mathbf{x} \in L^e \right\},$$

where B_x is a ball of suitable size centred at $\mathbf{x} \in D$, n_{Bx} is the number of observations located within it and L^e is a grid of suitable density. The ball size and grid density must be adjusted to the density of observations. Generate contour maps based on $\{\hat{\nu}_{0XB}(l); \mathbf{x} \in L^e\}$ for each $l \in \mathbb{L}$ and evaluate whether there appears to be spatial trends in the label proportions.

Second, validate the Markovian assumption. Define the empirical interaction function for each label $l \in \mathbb{L}$ as

$$\begin{aligned} \hat{J}_l(t) = & \frac{1}{|\mathcal{B}_{\mathbf{x}}(t) \cap \mathcal{D}|} \frac{1}{n_{ol}} \sum_{i=1}^{n_o} I(l(\mathbf{x}_j^o) = l) \\ & \times \left[\sum_{j=1}^{n_o} I(\mathbf{x}_j^o \in \mathcal{B}_{\mathbf{x}}(t)) I(l(\mathbf{x}_i^o) = l) - 1 \right]; t \in T \end{aligned},$$

where n_{ol} is the number of observations of label l and T defines a suitable discretisation of $t \in \mathbb{R}_+$. Generate a plot of the function $\{\hat{J}_l(t); t \in T\}$ for each $l \in \mathbb{L}$ and evaluate whether the values of any of the functions appear to be above average for small values of t , which indicates clustering and Markovian behaviour of the labels.

In the final evaluation, sampling uncertainty and boundary effects must be considered. The stationary Markov RF model is appropriate if spatial stationarity and Markov interaction are observed. If spatial stationarity is observed, but Markovian interaction is lacking, a particular Markov RF without spatial interaction may be used. If spatial stationarity cannot be justified, one should seek explanatory spatial variables such that the residual spatial variable appears spatially stationary with Markovian interaction.

6.8.3 Explanatory Spatial Variables

The unknown spatial variable of interest is $\{l(\mathbf{x}); \mathbf{x} \in L\}$, represented by the n -vector \mathbf{l} . Assume that a set of related spatial variables $\{g_j(\mathbf{x}); \mathbf{x} \in L\}; j = 1, 2, \dots, n_g$ are observed everywhere on the grid L for one of the labels $l^* \in \mathbb{L}$, and let them be represented by the n -vectors $\mathbf{g}_j; j = 1, 2, \dots, n_g$. These related spatial variables, or covariates, can be included as explanatory variables in the prior model in a spatial regression setting. The parameters in the Markov formulation of the Markov RF can be expressed as

$$\begin{aligned} \left\{ v_{0lx}(l_x) = v_{0l}^0(l_x) \times \left[\prod_{j=1}^{n_g} \exp(\beta_l^j g_{jx}) \right]^{I(l_x=l^*)} ; \mathbf{x} \in L \right\} \\ \omega_l(l_i | l_j; j \in \mathbf{n}_i); i = 1, 2, \dots, n. \end{aligned}$$

Denote the regression coefficients $\boldsymbol{\beta}_l = (\beta_l^1, \beta_l^2, \dots, \beta_l^{n_g})^T$. It can be demonstrated that this extended non-stationary Markov model is also a conjugate prior model

for the response likelihood model. The corresponding Markov posterior model is developed in Sect. 7.3.

6.8.4 Related Topics

The Markov RF model appears as almost the only viable model for mosaic spatial variables. Other important topics related to the Markov RF model are:

- Physics notation: The Ising model is extensively studied in statistical physics and is discussed in more detail in Ising (1925). This model has $l_i \in \mathbb{L} : \{+1, -1\}$ representing electrical charges, with equal marginal probability for each charge, and two-nearest-node horizontal and vertical cliques. The model depends on one single parameter $\beta \in \mathbb{R}$, and the associated neighbourhood design is the six-nearest-nodes in three-dimensional models. The notation for the Gibbs formulation in the physics community is as follows:

$$h(\mathbf{l}) = Z^{-1} \times \prod_{c \in c_L} \eta_c(\mathbf{l}) = Z^{-1} \times \prod_{\substack{h < i, j > \\ v < i, j >}} \exp(\beta \times l_i l_j),$$

where $h < i, j >$ and $v < i, j >$ are the nearest horizontal and vertical nodes. The function $h(\mathbf{l})$ is referred to as the energy function, Z is the partition function, and $\eta_c(\mathbf{l})$ is the potential function. Observe that Z^{-1} corresponds to the normalising constant, which is very computationally demanding to calculate because it requires a summation over $\mathbf{l} \in \mathbb{L}^n$.

The origin of the iterative McMC algorithm, as defined in Sect. 10.1, is the Metropolis algorithm, first introduced in the physics literature by Metropolis et al. (1953). This algorithm is a brute-force single-site proposal McMC algorithm, initially used to generate realisations from the Ising model. Later, the more efficient Swendsen–Wang algorithm, as defined in Swendsen and Wang (1987), was developed in the physics community.

- Compact classes: Binary Markov RF models with simple pairwise clique systems, and hence six-nearest-node neighbourhood systems, tend to generate one out of two realisation types: one of the two labels everywhere with some noise, or a random mixture of the two labels. To obtain realisations with compact regions of each label, higher-order spatial interactions must be defined, which can dramatically reduce the convergence rates in the McMC algorithm. Moreover, the model parametrisation becomes more complicated.

Realisations with compact label regions can be obtained by specifying the Markov RF in Gibbs form with higher-order cliques, as discussed in Tjelmeland and Besag (1998). In a binary model, foreground and background labels are specified, and the focus is on the parametrisation of boundaries and corners inside the clique geometry. Hexagonal grids are commonly used to provide angular

flexibility. The simulation of realisations and model parameter inference requires the use of McMC algorithms, which are generally computationally demanding. Informative prior models like these can be valuable in applications where the observations have a low signal-to-noise ratio.

6.8.5 Example: Markov RF

The example discussed is introduced in Sect. 5.3.1 and Fig. 5.3, and the unknown mosaic spatial variable to be reconstructed $\{l(\mathbf{x}); \mathbf{x} \in L \subset D\}$, represented in the n -vector \mathbf{l} , is given in display (a). The observations, represented in the n -vector \mathbf{d} , are presented in the corresponding display (c). The associated likelihood model for the observation acquisition procedure $p(\mathbf{d} | \mathbf{l})$ is also defined.

The prior model $p(\mathbf{l})$ captures the user experience and knowledge with the reference spatial variable. A Markov RF prior model is used since the mosaic is expected to contain white and black regions. An Ising RF prior model is assigned with a two-nearest-node clique design and, hence, a four-nearest-node neighbourhood design. The white and black proportions are specified to be equal, and the spatial interaction parameter is set to $\beta = 1.8$.

Figure 6.7 contains the reference spatial variable to be assessed in display (a). The three displays (b), (c) and (d) present three independent realisations from the Markov RF prior model. The realisations are generated by the iterative algorithm, as defined in Sect. 10.1, with a complete restart to ensure independence. The prior pdf appears to have two dominant modes. Thus, it has two preferred regimes: almost all white and almost all black. Despite this simple feature, the Ising RF model has proven to be suitable as a prior model for restoring mosaic spatial variables.

Now, the response likelihood model $p(\mathbf{d} | \mathbf{l})$ is specified, and the Markov RF prior model $p(\mathbf{l})$ is assigned. For a Bayesian spatial model, as defined in Chap. 2, these two models uniquely define the posterior model $p(\mathbf{l} | \mathbf{d})$. This posterior model provides the ultimate probabilistic reconstruction of the reference mosaic spatial variable.

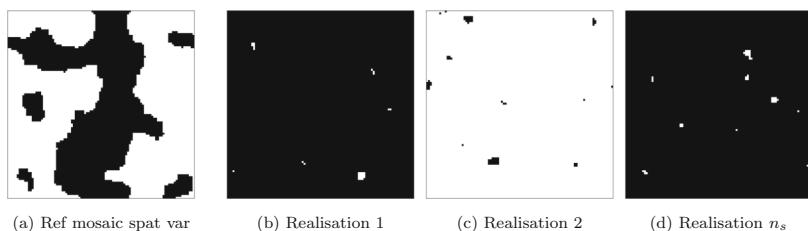


Fig. 6.7 Example Markov RF model. Markov prior model: (a) reference mosaic spatial variable and (b), (c), (d) three independent realisations

Chapter 7

Posterior Models



Predictions accompanied by their quantified uncertainties provide the basis for robust decision making. Better decisions lead to more favourable outcomes, a characteristic of a winner. The prediction uncertainties reflect the volume of information that supports the decisions. Without quantified uncertainties, the collection of observations becomes a meaningless task.

Consider the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D\}$ of interest, spatially discretised to $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ and represented by the n -vector \mathbf{s} . Assume that an m -vector \mathbf{d} of related observations is available. We have defined the likelihood model $p(\mathbf{d} | \mathbf{s})$ that links the observations and the variable of interest in Chap. 5. The prior model $p(\mathbf{s})$, which represents general experience and knowledge about the spatial variable of interest, is assigned in Chap. 6. The ultimate solution to Bayesian spatial modelling is the posterior model

$$p(\mathbf{s} | \mathbf{d}) = \left[\int p(\mathbf{d} | \mathbf{s}) p(\mathbf{s}) d\mathbf{s} \right]^{-1} \times p(\mathbf{d} | \mathbf{s}) p(\mathbf{s}). \quad (7.1)$$

The likelihood and prior models uniquely define the posterior model, which also contains a normalising constant that is complicated to calculate in the general case because it involves integration over the n -dimensional variable \mathbf{s} . We consider three classes of likelihood functions, one for each spatial variable type. Furthermore, we have chosen to use conjugate classes of prior models for each likelihood class, ensuring that the posterior models belong to the same conjugate class. Consequently, we may calculate the normalising constant analytically from the parameters in the likelihood and prior models. We demonstrate this conjugate property for the three spatial variable types: continuous, event and mosaic.

The posterior model can be represented by a set of realisations $\{[\mathbf{s}|\mathbf{d}]^s; s = 1, 2, \dots, n_s\}$ generated from the posterior pdf $p(\mathbf{s} | \mathbf{d})$. The simulation algorithms employed for this purpose are adapted versions of those discussed in Sect. 10.1. Each realisation reproduces the observations \mathbf{d} to the precision of the likelihood model and captures the spatial heterogeneity of the prior model. Hence, it represents

a possible spatial variable under study. However, realisations are not optimal for locationwise prediction of the spatial variable. Alternatively, one may provide optimal spatial predictors, typically in a LSE sense, $\hat{\mathbf{s}} = E\{\mathbf{s}|\mathbf{d}\}$, with associated prediction error measures, $\Sigma_p = \text{Var}\{\mathbf{s}|\mathbf{d}\}$. The locationwise prediction variances appear on the diagonal of this matrix, and these variances are represented in the n -vector σ_p^2 . Both the prediction and its variance depend on the spatial location and they are optimal for locationwise prediction of the spatial variable. The prediction reproduces the observations \mathbf{d} with the precision of the likelihood model but causes regression towards the prior model. Therefore, it expresses less spatial heterogeneity than the spatial variable under study. Note that the set of realisations and the predictions with associated errors are related as $1/n_s \sum_s [\mathbf{s}|\mathbf{d}]^s \xrightarrow{n_s \rightarrow \infty} E\{\mathbf{s}|\mathbf{d}\} = \hat{\mathbf{s}}$ and $1/n_s \sum_s [[\mathbf{s}|\mathbf{d}]^s - \hat{\mathbf{s}}][[\mathbf{s}|\mathbf{d}]^s - \hat{\mathbf{s}}]^T \xrightarrow{n_s \rightarrow \infty} \text{Var}\{\mathbf{s}|\mathbf{d}\} = \Sigma_p$.

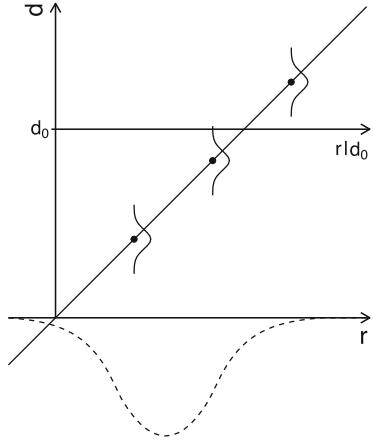
In practice, functional expressions involving the spatial variable are often of interest and, in the spatially discretised representation, can be expressed as $s_w = w(\mathbf{s})$. Such functionals may represent spatial averages, the proportion above a threshold value, penetration time for a fluid or other similar quantities. The associated posterior pdf is denoted $p(s_w | \mathbf{d})$, and in the general case, this posterior model must be assessed by Monte Carlo simulation. Consider the set of realisations from $p(\mathbf{s} | \mathbf{d})$ as previously defined. The set $\{s_w^s = w([\mathbf{s}|\mathbf{d}]^s); s = 1, 2, \dots, n_s\}$ will then be a set of realisations from $p(s_w | \mathbf{d})$. Predictions with associated error measures can be assessed from this set of realisations. Linear functionals represented by the n -vector \mathbf{w} can be defined as $s_w = \mathbf{w}^T \mathbf{s}$. Spatial averages over sub-domains and spatial slopes in a given location and direction are examples of linear functionals of the spatial variable. The associated posterior pdf $p(s_w | \mathbf{d})$ can typically be assessed analytically.

This section covers simulations from the posterior pdf, predictors with associated errors and prediction of linear functionals for the various spatial variable types. Furthermore, posterior models with explanatory variables and posterior models for the hierarchical prior models are developed.

7.1 Continuous Spatial Variables

The class of Gauss-linear likelihood models is defined in Sect. 5.1, and the corresponding classes of conjugate continuous prior models are defined in Sect. 6.1. These spatially discretised models define the posterior model, and the conjugate property makes it possible to develop expressions for the exact posterior model by considering only the parameters of the likelihood and prior models and the actual observations. Therefore, complex high-dimensional numerical integration is avoided. We demonstrate this conjugate characteristic in this section. In Fig. 7.1, a graphical presentation of the Gaussian inversion challenge is displayed in a format comparable to Fig. 2.1. The challenge is to determine the pdf along the $[\mathbf{r}|\mathbf{d}_0]$ -axis.

Fig. 7.1 Bayesian inversion.
Schematic presentation for a
continuous Gaussian RF



7.1.1 Gaussian RF Models

The joint pdf for the spatial variable of interest and the observations is, for $[\mathbf{r}, \mathbf{d}] \in \mathbb{R}^{n+m}$,

$$\begin{aligned} \begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix} &\sim p(\mathbf{r}, \mathbf{d}) = p(\mathbf{d} | \mathbf{r})p(\mathbf{r}) = \phi_m(\mathbf{d}; \mathbf{H}\mathbf{r}, \Sigma_{d|r}) \times \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \Sigma_r^\rho) \\ &= \phi_{n+m} \left(\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \mu_r \mathbf{i}_n \\ \mu_r \mathbf{H}\mathbf{i}_n \end{bmatrix}, \begin{bmatrix} \sigma_r^2 \Sigma_r^\rho & \sigma_r^2 \Sigma_r^\rho \mathbf{H}^T \\ \sigma_r^2 \mathbf{H} \Sigma_r^\rho & \sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r} \end{bmatrix} \right). \end{aligned} \quad (7.2)$$

This joint pdf is Gaussian with the model parameters defined by the model parameters of the likelihood and prior models. Consequently, the posterior model is also a Gaussian pdf, as given in Mardia et al. (1979)

$$[\mathbf{r}|\mathbf{d}] \sim p(\mathbf{r} | \mathbf{d}) = \phi_n(\mathbf{r}; \mu_{r|d}, \Sigma_{r|d}) \quad (7.3)$$

with model parameters

$$\begin{aligned} \mu_{r|d} &= \mu_r \mathbf{i}_n + \Sigma_r \mathbf{H}^T \Sigma_d^{-1} [\mathbf{d} - \mu_r \mathbf{H}\mathbf{i}_n] \\ &= \mu_r \mathbf{i}_n + \sigma_r^2 \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} [\mathbf{d} - \mu_r \mathbf{H}\mathbf{i}_n] \\ \Sigma_{r|d} &= \Sigma_{r|d}^\sigma \Sigma_{r|d}^\rho \Sigma_{r|d}^\sigma \\ &= \Sigma_r - \Sigma_r \mathbf{H}^T \Sigma_d^{-1} \mathbf{H} \Sigma_r \\ &= \sigma_r^2 \Sigma_r^\rho - \sigma_r^2 \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} \sigma_r^2 \mathbf{H} \Sigma_r^\rho, \end{aligned}$$

where $\Sigma_{r|d}^\sigma$ is a diagonal standard deviation matrix of dimension $(n \times n)$ and $\Sigma_{r|d}^\rho$ is the associated correlation matrix, of dimension $(n \times n)$. Both matrices are non-negative definite by construction. The conditional expectation n -vector $\mu_{r|d}$ is linear in the conditioning values \mathbf{d} , whereas the conditional covariance matrix $\Sigma_{r|d}$ is dependent on the observation design and is independent of the actual observed values \mathbf{d} .

The parameters in Expression (7.3) can be rephrased by the matrix identities from Lindgren et al. (2011)

$$\begin{aligned}\Sigma_r \mathbf{H}^T \Sigma_d^{-1} &= \Sigma_{r|d} \mathbf{H}^T \Sigma_{d|r}^{-1} \\ \Sigma_{r|d} &= [\Sigma_r^{-1} + \mathbf{H}^T \Sigma_{d|r}^{-1} \mathbf{H}]^{-1}.\end{aligned}$$

This posterior model belongs to the class of Gaussian RF model, as defined in Expression (4.1), and so does the prior model. Consequently, the Gaussian RF model constitutes a conjugate class with respect to Gauss-linear likelihood models. It can be demonstrated that this conjugate property holds for non-discretised Gaussian RF models by using infill asymptotic analysis. The model parameters for the posterior Gaussian RF are defined by the expectation vector $\mu_{r|d}$ in $\{\mu_{r|d}(\mathbf{x}); \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{\mu_{r|d}(\mathbf{x}); \mathbf{x} \in D\}$, the diagonal standard deviation matrix $\Sigma_{r|d}^\sigma$ in $\{\sigma_{r|d}^2(\mathbf{x}); \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{\sigma_{r|d}^2(\mathbf{x}); \mathbf{x} \in D\}$ and the correlation matrix $\Sigma_{r|d}^\rho$ in $\{\rho_{r|d}(\mathbf{x}, \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in L\} \xrightarrow{\text{infill}} \{\rho_{r|d}(\mathbf{x}, \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}$. The posterior model appears as a non-stationary Gaussian RF with the non-stationarity introduced by conditioning on the observations.

Spatial Discretisation Uncertainty

The posterior model is defined as a spatially discretised, non-stationary Gaussian RF $\{[r(\mathbf{x})|\mathbf{d}]; \mathbf{x} \in L\}$, and the discretisation in both the likelihood and the prior models introduces errors. The crucial terms in the expression for the posterior model are the expectation n -vector $\mu_d = \mu_r \mathbf{H} \mathbf{i}_n$ and the covariance matrix $\Sigma_d = \sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}$ for the observations.

The discretisation errors at a specific location \mathbf{x}^o , denoted by Δd^o and Δr^o in the likelihood and prior models, respectively, are discussed in Sects. 5.1 and 6.1. The expectation and variance of the differences in the observations, which are of particular interest, are calculated using the laws of double expectation and total variance,

$$\begin{aligned}E\{\Delta d^o\} &= E\{E\{\Delta d^o | \Delta r^o\}\} = 0, \\ \text{Var}\{\Delta d^o\} &= E\{\text{Var}\{\Delta d^o | \Delta r^o\}\} + \text{Var}\{E\{\Delta d^o | \Delta r^o\}\} \\ &= 2\sigma_r^2 [1 - \rho_r(\mathbf{x}^o - \mathbf{x}_{i^o})] = \sigma_\Delta^2.\end{aligned}$$

Recall that, from the model defined in these sections, $\Delta d^o = d^o - d$ and that $E\{d^o\} = \mu_r$ while $\text{Var}\{d^o\} = \sigma_r^2 + \sigma_{d|r}^2$. Thus,

$$E\{d\} = \mu_r,$$

$$\text{Var}\{d\} = \sigma_r^2 + \sigma_{d|r}^2 + \sigma_\Delta^2.$$

Consequently, the expectation μ_d remains unaffected by the discretisation, whereas the covariance Σ_d appears with a variance level that is too low. For an isotropic and stationary Gaussian prior model without a hole effect in the spatial correlation function, the variance error is bounded by $\sigma_\Delta^2 \leq 2\sigma_r^2[1 - \rho_r(\Delta x_M^n)]$. However, this extra discretisation error is not considered in the current developments of the posterior expressions.

Simulation

Simulation of a realisation $\{[r(\mathbf{x})|\mathbf{d}]; \mathbf{x} \in L\}$, represented by the n -vector $[\mathbf{r}|\mathbf{d}]^s$, from the non-stationary posterior Gaussian RF, can be generated by a transformation simulation algorithm, as defined in Algorithm 7. The algorithm requires a Cholesky decomposition of the posterior covariance matrix $\Sigma_{r|d}$, of dimension $(n \times n)$, which may be computationally demanding. Cholesky decompositions are extensively studied in the field of numerical analysis, and many efficient algorithms exist, as discussed in Sect. 9.1.

Algorithm 7: Posterior Gaussian RF: simultaneous simulation

Define $\mathbf{r}^T = \Sigma_{r|d}^{-1/2}[\mathbf{r} - \mu_{r|d}] \sim p(\mathbf{r}^T) = \phi_n(\mathbf{r}; \mathbf{0}, \mathbf{I}_n)$

Generate $\mathbf{r}^{Ts} \sim p(\mathbf{r}^T) = \phi_n(\mathbf{r}; \mathbf{0}, \mathbf{I}_n)$

Calculate $[\mathbf{r} | \mathbf{d}]^s = \mu_{r|d} + \Sigma_{r|d}^{1/2} \mathbf{r}^{Ts} \sim p(\mathbf{r} | \mathbf{d}) = \phi_n(\mathbf{r}; \mu_{r|d}, \Sigma_{r|d})$

Result: $[\mathbf{r} | \mathbf{d}]^s \sim p(\mathbf{r} | \mathbf{d}) = \phi_n(\mathbf{r}; \mu_{r|d}, \Sigma_{r|d})$

Alternatively, the simulation of $[\mathbf{r}|\mathbf{d}]^s$ can be generated by a sequential algorithm, as defined in Algorithm 8. For a Gaussian vector $[\mathbf{r}|\mathbf{d}]$, all marginal conditional pdfs are Gaussian and analytically tractable. However, they may be computationally demanding to obtain for high-dimensional conditioning. A localised approximation, with conditioning on only the spatially nearest simulated values, is often used to improve the algorithm's efficiency, as discussed in Sect. 9.1.

The computational challenges in the algorithms consist of the Cholesky decomposition of unstructured matrices up to dimension $(n \times n)$, corresponding to the grid dimension. This challenging decomposition can be avoided by rephrasing the simulation task. One realisation of the posterior non-stationary Gaussian RF can alternatively be decomposed as

$$[\mathbf{r}|\mathbf{d}]^s = \mu_{r|d} + [\mathbf{r}_*^s - \mu_{r|d_*^s}] \quad (7.4)$$

Algorithm 8: Posterior Gaussian RF: sequential decomposition simulation

```

Generate  $r_1^s \sim p(r_1 | \mathbf{d}) = \phi_1(r; \mu_{r|d}, \sigma_{r|d}^2)$ 
for  $i = 2, 3, \dots, n$  do
     $\left| \text{Generate } r_i^s \sim p(r_i | \mathbf{d}, r_{1:(i-1)}^s) = \phi_1\left(r; \mu_{r_i|d, r_{1:(i-1)}^s}, \sigma_{r_i|d, r_{1:(i-1)}^s}^2\right)$ 
end
Result:  $[\mathbf{r} | \mathbf{d}]^s = [r_1^s, r_2^s, \dots, r_n^s] \sim p(\mathbf{r} | \mathbf{d}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d})$ 

```

with

$$\begin{aligned} E\{[\mathbf{r} | \mathbf{d}]^s\} &= \boldsymbol{\mu}_{r|d} + 0\mathbf{i}_n = \boldsymbol{\mu}_{r|d} \\ \text{Var}\{[\mathbf{r} | \mathbf{d}]^s\} &= 0\mathbf{I}_n + \boldsymbol{\Sigma}_{r|d} = \boldsymbol{\Sigma}_{r|d}. \end{aligned}$$

The n -vector \mathbf{r}_*^s is a simulated realisation of the prior stationary Gaussian RF. The associated expectation $\boldsymbol{\mu}_{r|d_*^s}$ is conditioned on the set of simulated observations $\mathbf{d}_*^s = \mathbf{H}\mathbf{r}_*^s + \mathbf{e}_{r|d}$, which is based on this realisation. One realisation of the Gaussian posterior model based on this decomposition requires only the generation of one realisation from the Gaussian prior model. Because the latter is spatially stationary, efficient simulation algorithms are available. In addition, the conditional expectation must be calculated, which requires the Cholesky decomposition of an unstructured matrix, of dimension $(m \times m)$, corresponding to the dimension of the observations. The dimension of the observations is typically much smaller than the dimension of the grid.

Finally, one realisation from the Gaussian posterior model can alternatively be obtained by an optimisation procedure, as defined in Oliver (1996),

$$[\mathbf{r} | \mathbf{d}]^s = \arg \min_{\mathbf{r}} \{(\mathbf{r}^* - \mathbf{r})^T \boldsymbol{\Sigma}_r^{-1} (\mathbf{r}^* - \mathbf{r}) + (\mathbf{d}^* - \mathbf{H}\mathbf{r})^T \boldsymbol{\Sigma}_{d|r}^{-1} (\mathbf{d}^* - \mathbf{H}\mathbf{r})\} \quad (7.5)$$

where

$$\begin{aligned} \mathbf{r}^* &\sim \phi_n(\mathbf{r}^*; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\ \mathbf{d}^* &\sim \phi_m(\mathbf{d}^*; \mathbf{d}, \boldsymbol{\Sigma}_{d|r}). \end{aligned}$$

This approach to simulation appears as a maximum randomised posterior expression. In the expression, $\boldsymbol{\mu}_r$ and \mathbf{d} are randomised as \mathbf{r}^* and \mathbf{d}^* , respectively. In the case of the Gaussian model, the optimisation can be solved analytically, and it can be demonstrated that $E\{[\mathbf{r} | \mathbf{d}]^s\} = \boldsymbol{\mu}_{r|d}$ and $\text{Var}\{[\mathbf{r} | \mathbf{d}]^s\} = \boldsymbol{\Sigma}_{r|d}$, as expected for a posterior realisation. This simulation approach is suitable as an approximation whenever the response function in the likelihood model is non-linear.

Prediction

Prediction of the spatial variable $\{\hat{r}(\mathbf{x}); \mathbf{x} \in L\}$ represented by the n -vector $\hat{\mathbf{r}}$ is usually defined based on a LSE criterion. Thus, the prediction is

$$\hat{\mathbf{r}} = E\{\mathbf{r}|\mathbf{d}\} = \boldsymbol{\mu}_r|_d = \mu_r \mathbf{i}_n + \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1} [\mathbf{d} - \mu_r \mathbf{H} \mathbf{i}_n]. \quad (7.6)$$

The associated prediction covariance is

$$\boldsymbol{\Sigma}_p = \text{Var}\{\mathbf{r}|\mathbf{d}\} = \boldsymbol{\Sigma}_{r|d} = \sigma_r^2 \boldsymbol{\Sigma}_r^\rho - \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1} \sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho$$

with

$$\sigma_p^2 = \text{Vdiag}_n\{\boldsymbol{\Sigma}_p\}. \quad (7.7)$$

Thus, the prediction variance n -vector σ_p^2 is determined by the diagonal entries of the prediction covariance matrix $\boldsymbol{\Sigma}_p$, and the associated set of marginal prediction $(1 - \alpha)$ -intervals is

$$MPI_\alpha = \hat{\mathbf{r}} \pm z_{\alpha/2} \boldsymbol{\sigma}_p. \quad (7.8)$$

The predictor has alternative formulations, however. Consider one specific location $\mathbf{x}_0 \in D$ at which the associated variable $r_0 = r(\mathbf{x}_0)$ is to be predicted. The LSE predictor with prediction variance is

$$\begin{aligned} \hat{r}_0 &= E\{r_0|\mathbf{d}\} = \mu_r + \sigma_r^2 \boldsymbol{\rho}_{0d}^T \boldsymbol{\Sigma}_d^{-1} (\mathbf{d} - \boldsymbol{\mu}_d) \\ \sigma_0^2 &= \text{Var}\{r_0 - \hat{r}_0|\mathbf{d}\} = \text{Var}\{r_0\} + \text{Var}\{\hat{r}_0\} - 2\text{Cov}\{r_0, \hat{r}_0\} \\ &= \sigma_r^2 [1 - \sigma_r^2 \boldsymbol{\rho}_{0d}^T \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\rho}_{0d}], \end{aligned}$$

where $\boldsymbol{\mu}_d$ contains expectations of \mathbf{d} and $\boldsymbol{\rho}_{0d} = \mathbf{H} \boldsymbol{\rho}_{0r}$ contains the prior correlations between r_0 and \mathbf{d} . In the latter expression, the vector $\boldsymbol{\rho}_{0r}$ represents the prior correlations between r_0 and \mathbf{r} , which is defined by the spatial correlation function $\rho_r(\cdot)$. The matrix $\boldsymbol{\Sigma}_d = [\sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]$ contains the prior covariances in the observation design.

The traditional predictor formulation, as given in Chiles and Delfiner (2012), corresponding to the traditional Kriging formulation, is

$$\hat{r}_0 = \mu_r + \boldsymbol{\omega}_{0d}^T (\mathbf{d} - \boldsymbol{\mu}_d),$$

where the m -vector $\boldsymbol{\omega}_{0d} = \sigma_r^2 \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\rho}_{0d}$ contains the weights and is dependent on the location \mathbf{x}_0 and the observation design. The weights, however, are independent of the

observed values \mathbf{d} . The predictor is linear in \mathbf{d} , and the weights must be recalculated for each new location \mathbf{x}_0 to be evaluated. The corresponding prediction variance is

$$\begin{aligned}\sigma_0^2 &= \text{Var}\{r_0\} + \text{Var}\{\hat{r}_0\} - 2\text{Cov}\{r_0, \hat{r}_0\} \\ &= \text{Var}\{r_0\} + \boldsymbol{\omega}_{0d}^T \text{Var}\{\mathbf{d}\} \boldsymbol{\omega}_{0d} - 2\text{Cov}\{r_0, \mathbf{d}\} \boldsymbol{\omega}_{0d} \\ &= \sigma_r^2 [1 - \boldsymbol{\omega}_{0d}^T \boldsymbol{\rho}_{0d}].\end{aligned}$$

Thus, the variance σ_0^2 is independent of the observed values \mathbf{d} .

The alternative predictor formulation, which corresponds to the dual Kriging formulation, as given in Chiles and Delfiner (2012), is expressed as

$$\hat{r}_0 = \mu_r + \boldsymbol{\rho}_{0d}^T \boldsymbol{\omega}_d,$$

where the m -vector $\boldsymbol{\omega}_d = \sigma_r^2 \boldsymbol{\Sigma}_d^{-1} (\mathbf{d} - \boldsymbol{\mu}_d)$ contains the weights and is dependent on the observation design and the observed values \mathbf{d} . The weights, however, are independent of the location \mathbf{x}_0 . The predictor is linear in the elements of the vector $\boldsymbol{\rho}_{0d} = \mathbf{H} \boldsymbol{\rho}_{0r}$, which can be analytically calculated based on the spatial correlation function $\rho_r(\cdot)$. Consequently, the weights can be calculated once for each observation set \mathbf{d} and can afterwards be used for prediction in an arbitrary location $\mathbf{x}_0 \in \mathcal{D}$. The corresponding prediction variance, which is identical to the one above, can be expressed as

$$\begin{aligned}\sigma_0^2 &= \text{Var}\{r_0\} + \text{Var}\{\hat{r}_0\} - 2\text{Cov}\{r_0, \hat{r}_0\} \\ &= \text{Var}\{r_0\} + \boldsymbol{\rho}_{0d}^T \text{Var}\{\boldsymbol{\omega}_d\} \boldsymbol{\rho}_{0d} - 2\text{Cov}\{r_0, \boldsymbol{\omega}_d\} \boldsymbol{\rho}_{0d} \\ &= \sigma_r^2 [1 - \sigma_r^2 \boldsymbol{\rho}_{0d}^T \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\rho}_{0d}].\end{aligned}$$

Thus, the variance σ_0^2 is independent of the observed values \mathbf{d} .

Consequently, the spatially non-discretised predictor and prediction variance over \mathcal{D} can be defined as

$$\begin{aligned}\{\hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\rho}_{\mathbf{x}d}^T \boldsymbol{\omega}_d = \mu_r + \sum_{i=1}^m \omega_d^i \mathbf{h}_i^T \boldsymbol{\rho}_{\mathbf{x}r}; \mathbf{x} \in \mathcal{D}\} \quad (7.9) \\ \{\sigma_p^2(\mathbf{x}) = \sigma_r^2 [1 - \boldsymbol{\rho}_{\mathbf{x}r}^T \mathbf{H}^T \sigma_r^2 \boldsymbol{\Sigma}_d^{-1} \mathbf{H} \boldsymbol{\rho}_{\mathbf{x}r}]; \mathbf{x} \in \mathcal{D}\},\end{aligned}$$

where the vectors \mathbf{h}_i^T are rows in the observation design matrix \mathbf{H} and the vector $\boldsymbol{\rho}_{\mathbf{x}r}$ contains the correlations between the prediction value and the grid values. These expressions for the predictor and the associated prediction variance are almost everywhere continuous functions in $\mathbf{x} \in \mathcal{D}$. Thus, they provide a spatially non-discretised representation of the prediction and the associated prediction variance. For the particular case with a Gauss-point likelihood model, as defined in Sect. 5.1,

with m observations \mathbf{d}_o made in a subset of the grid nodes $L^o \subset L$ with independent Gaussian errors, the non-discretised predictor can be expressed as

$$\{\hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\rho}_{\mathbf{x}d}^T \boldsymbol{\omega}_d = \mu_r + \sum_{\mathbf{x}' \in L^o} \omega_d^{\mathbf{x}'} \rho_r(\mathbf{x} - \mathbf{x}'); \mathbf{x} \in D\}.$$

The m -vector $\boldsymbol{\omega}_d = \sigma_r^2 \boldsymbol{\Sigma}_d^{-1} (\mathbf{d} - \boldsymbol{\mu}_d)$ contains the weights. This predictor appears as a weighted, linear combination of m spatial correlation functions, each centred at one of the m observation grid nodes L^o . The corresponding prediction variance can be expressed similarly in a continuous quadratic form.

This spatially non-discretised predictor is referred to as the kernel predictor in Omre and Spremec (2023). See Sect. 10.5 for a more comprehensive discussion.

Special Case

Consider the Dirac-linear likelihood model $p(\mathbf{d}_e | \mathbf{r})$ with exact observations, as defined in Sect. 5.1. The posterior model is then expressed as

$$[\mathbf{r} | \mathbf{d}_e] \sim p(\mathbf{r} | \mathbf{d}_e) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d_e}, \boldsymbol{\Sigma}_{r|d_e}) \quad (7.10)$$

with

$$\begin{aligned} \boldsymbol{\mu}_{r|d_e} &= \mu_r \mathbf{i}_n + \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T [\mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T]^{-1} [\mathbf{d}_e - \mathbf{H} \mu_r \mathbf{i}_n] \\ \boldsymbol{\Sigma}_{r|d_e} &= \boldsymbol{\Sigma}_{r|d}^\sigma \boldsymbol{\Sigma}_{r|d}^\rho \boldsymbol{\Sigma}_{r|d}^\sigma = \sigma_r^2 [\boldsymbol{\Sigma}_r^\rho - \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T [\mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T]^{-1} \mathbf{H} \boldsymbol{\Sigma}_r^\rho] \end{aligned}$$

with the previously specified model parameters. For this likelihood model without observation error, the predictor $\hat{\mathbf{r}} = \boldsymbol{\mu}_{r|d_e}$ is independent of the prior variance σ_r^2 , whereas the prediction covariance matrix $\boldsymbol{\Sigma}_p = \boldsymbol{\Sigma}_{r|d_e}$ is proportional to the prior variance σ_r^2 .

Functionals

The posterior model for the linear functional $r_w = \mathbf{w}^T \mathbf{r}$ is developed from the joint pdf

$$\begin{aligned} \begin{bmatrix} r_w \\ \mathbf{r} \\ \mathbf{d} \end{bmatrix} &\sim p(r_w, \mathbf{r}, \mathbf{d}) \\ &= \phi_{1+n+m} \left(\begin{bmatrix} r_w \\ \mathbf{r} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \mu_r \mathbf{w}^T \mathbf{i}_n \\ \mu_r \mathbf{i}_n \\ \mu_r \mathbf{H} \mathbf{i}_n \end{bmatrix}, \begin{bmatrix} \sigma_r^2 \mathbf{w}^T \boldsymbol{\Sigma}_r^\rho \mathbf{w} & \sigma_r^2 \mathbf{w}^T \boldsymbol{\Sigma}_r^\rho & \sigma_r^2 \mathbf{w}^T \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T \\ \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{w} & \sigma_r^2 \boldsymbol{\Sigma}_r^\rho & \sigma_r^2 \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T \\ \sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{w} & \sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho & \sigma_r^2 \mathbf{H} \boldsymbol{\Sigma}_r^\rho \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r} \end{bmatrix} \right). \end{aligned}$$

The posterior model of interest is Gaussian as well:

$$[r_w | \mathbf{d}] \sim p(r_w | \mathbf{d}) = \phi_1(r_w; \mu_{r_w|d}, \sigma_{r_w|d}^2) \quad (7.11)$$

with

$$\begin{aligned}\mu_{r_w|d} &= \mu_r \mathbf{w}^T \mathbf{i}_n + \sigma_r^2 \mathbf{w}^T \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} [\mathbf{d} - \mu_r \mathbf{H} \mathbf{i}_n] \\ \sigma_{r_w|d}^2 &= \sigma_r^2 \mathbf{w}^T \Sigma_r^\rho \mathbf{w} - \sigma_r^2 \mathbf{w}^T \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} \sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{w}.\end{aligned}$$

Thus, a set of realisations $\{[r_w|\mathbf{d}]^s; s = 1, 2, \dots, n_s\}$ from the Gaussian posterior model $p(r_w | \mathbf{d})$ can be generated by Algorithms 7 and 8. The predictor is $\hat{r}_w = E\{r_w|\mathbf{d}\} = \mu_{r_w|d}$, the prediction variance is $\sigma_p^2 = \text{Var}\{r_w|\mathbf{d}\} = \sigma_{r_w|d}^2$, and the associated marginal prediction $(1 - \alpha)$ -interval is $MPI_\alpha = \hat{r}_w \pm z_{\alpha/2} \sigma_{r_w|d}$.

Explanatory Variables

Explanatory spatial variables, or covariates, may be available, and these can be included into the Gaussian prior model, as specified in Sect. 6.5. The grid representation of the explanatory variables is $\{g_j(\mathbf{x}); \mathbf{x} \in L\}; j = 1, 2, \dots, n_g$, and they are represented in the n -vectors $\mathbf{g}_j; j = 1, 2, \dots, n_g$, which in turn are stacked into the $(n_g \times n)$ matrix $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{n_g}]^T$. The Gaussian regression prior model is then defined with regression coefficients: intercept μ_r^0 and slope n_g -vector $\boldsymbol{\beta}_r$. The corresponding posterior model is expressed as

$$[\mathbf{r}|\mathbf{d}] \sim p(\mathbf{r} | \mathbf{d}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \Sigma_{r|d}) \quad (7.12)$$

with

$$\begin{aligned}\boldsymbol{\mu}_{r|d} &= [\mu_r^0 \mathbf{i}_n + \mathbf{G}^T \boldsymbol{\beta}_r] + \sigma_r^2 \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} [\mathbf{d} - \mathbf{H}[\mu_r^0 \mathbf{i}_n + \mathbf{G}^T \boldsymbol{\beta}_r]] \\ \Sigma_{r|d} &= \Sigma_{r|d}^\sigma \Sigma_{r|d}^\rho \Sigma_{r|d}^\sigma = \sigma_r^2 \Sigma_r^\rho - \sigma_r^2 \Sigma_r^\rho \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_r^\rho \mathbf{H}^T + \Sigma_{d|r}]^{-1} \sigma_r^2 \mathbf{H} \Sigma_r^\rho.\end{aligned}$$

This posterior model is a spatially discretised Gaussian RF. Thus, this regression-extended Gaussian prior model is also a conjugate class with respect to a Gauss-linear likelihood model. Simulation and predictions can be obtained similarly to the case with a stationary prior model.

7.1.2 Example: Gaussian RF

The example discussed is introduced in Sect. 5.1.1 and Fig. 5.1. The unknown continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$ to be reconstructed is given in display (a). The evaluation is made on the spatially discretised reference spatial variable presented in display (b), represented by the n -vector \mathbf{r} . The available observations are presented in display (c) and a Gauss-linear likelihood model $p(\mathbf{d} | \mathbf{r})$ is specified. In Sect. 6.1, a Gaussian RF prior model is assigned, and thus, the prior model $p(\mathbf{r})$ is Gaussian.

The Gaussian prior model $p(\mathbf{r})$ is demonstrated to be conjugate with respect to the Gauss-linear likelihood $p(\mathbf{d} | \mathbf{r})$. Consequently, the posterior model $p(\mathbf{r} | \mathbf{d})$ is

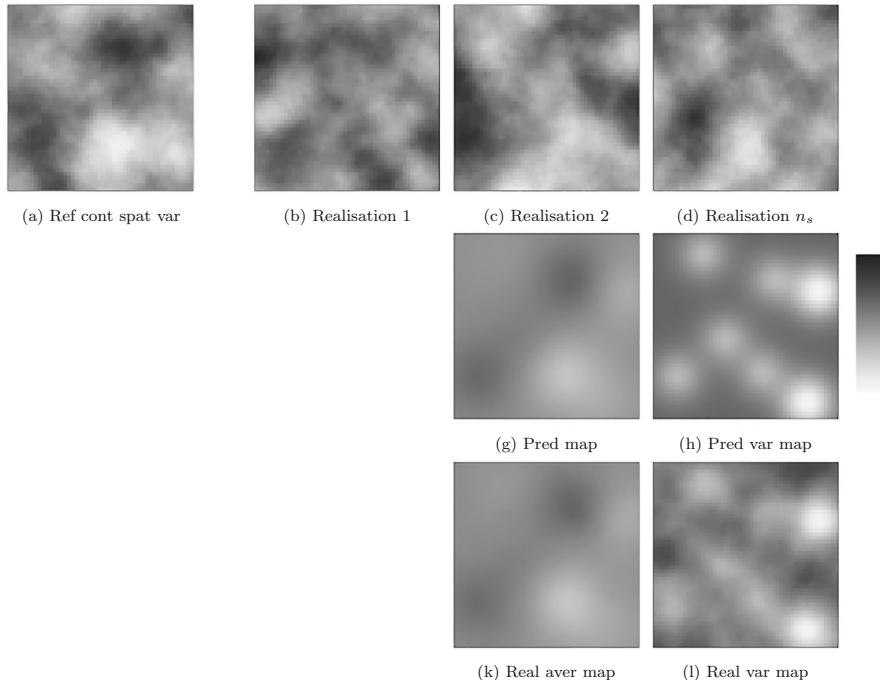


Fig. 7.2 Example Gaussian RF model. Gaussian posterior model: (a) reference continuous spatial variable; (b), (c), (d) three independent realisations; (g) prediction map; (h) prediction variance map; (k) realisation average map; and (l) realisation variance map

also Gaussian, with expectation $\mu_{r|d}$ and covariance $\Sigma_{r|d}$. These model parameters are analytically assessed from the model parameters of the likelihood and prior models, as well as the observed values.

Figure 7.2 contains the reference spatial variable in display (a) and three independent realisations from the posterior Gaussian RF model $p(\mathbf{r} | \mathbf{d})$ in displays (b), (c) and (d). These realisations are conditioned on the observations \mathbf{d} and reproduce these observations to the precision of the observation errors. Furthermore, given the model and the observations \mathbf{d} , they appear as possible outcomes of the continuous spatial variable. The best prediction of the reference spatial variable, $\{\hat{r}(\mathbf{x}); \mathbf{x} \in D\}$, in the LSE sense, is the conditional expectation. Thus, the spatially discretised prediction is represented by the n -vector $\hat{\mathbf{r}} = \mu_{r|d}$, which is presented in display (g). The influence of the observed values \mathbf{d} is evident, but the prediction is spatially smoother than the reference spatial variable. Likewise, the spatially discretised prediction variances in the n -vector σ_p^2 , presented in display (h), are defined by the diagonal entries of the posterior covariance matrix $\Sigma_{r|d}$. The prediction variance is smallest at the observation locations and increases towards the prior variance level with increasing distance from the observation locations.

The realisations are related to the prediction and prediction variance because they are all based on the posterior model $p(\mathbf{r} \mid \mathbf{d})$. Displays (k) and (l) contain the average and empirical variance of 200 independent realisations. The average converges towards the prediction as the number of realisations increases, whereas the empirical variance converges towards the prediction variance. As expected, the convergence rate is faster for the average than for the empirical variance.

7.1.3 Hierarchical Gaussian RF Models

The stationary Gaussian RF is cast in a hierarchical setting in Sect. 6.3. The discretised stationary hierarchical Gaussian RF $\{r_H(\mathbf{x}); \mathbf{x} \in L\}$ is represented by the n -vector \mathbf{r}_H . This RF is not ergodic. Thus, care should be taken when performing model parameter inference and when interpreting the prediction results from this model.

Consider a Dirac-linear likelihood model, as defined in Sect. 5.1, which entails that the observations are without errors. Let the prior model be a discretised stationary hierarchical Gaussian RF model as in Sect. 6.3. The posterior model is then an n -variate T-distributed pdf, as demonstrated in Røislien and Omre (2006). The posterior model for $\mathbf{r}_H \in \mathbb{R}^n$ is expressed as

$$\begin{aligned} [\mathbf{r}_H | \mathbf{d}_e] &\sim p(\mathbf{r}_H \mid \mathbf{d}_e) \\ &= \int_{\mathbb{R}^{\oplus}} \int_{\mathbb{R}} p(\mathbf{r}_H \mid \mathbf{d}_e, \mu_r, \sigma_r^2) p(\mu_r \mid \mathbf{d}_e, \sigma_r^2) p(\sigma_r^2 \mid \mathbf{d}_e) d\mu_r d\sigma_r^2 \\ &= \left[\Gamma \left(\frac{v_{\sigma^2|d}}{2} \right) \right]^{-1} [v_{\sigma^2|d} \pi]^{-n/2} \Gamma \left(\frac{v_{\sigma^2|d} + n}{2} \right) \\ &\quad \times |\boldsymbol{\Omega}_{r_H|d}|^{-1/2} \times \left[1 + v_{\sigma^2|d}^{-1} (\mathbf{r}_H - \boldsymbol{\mu}_{r_H|d})^T \boldsymbol{\Omega}_{r_H|d}^{-1} (\mathbf{r}_H - \boldsymbol{\mu}_{r_H|d}) \right]^{-(v_{\sigma^2|d} + n)/2} \end{aligned} \tag{7.13}$$

with

$$\begin{aligned} \boldsymbol{\mu}_{r_H|d} &= \mu_\mu \mathbf{i}_n + \boldsymbol{\Sigma}_{r_H}^o \mathbf{H}^T [\mathbf{H} \boldsymbol{\Sigma}_{r_H}^o \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ \boldsymbol{\Omega}_{r_H|d} &= \xi_{\sigma^2|d}^2 \left[\boldsymbol{\Sigma}_{r_H}^o - \boldsymbol{\Sigma}_{r_H}^o \mathbf{H}^T [\mathbf{H} \boldsymbol{\Sigma}_{r_H}^o \mathbf{H}^T]^{-1} \mathbf{H} \boldsymbol{\Sigma}_{r_H}^o \right], \end{aligned}$$

where $\boldsymbol{\Sigma}_{r_H}^o = \gamma_\mu \mathbf{i}_n \mathbf{i}_n^T + \boldsymbol{\Sigma}_{r|\mu}^\rho$. In the expression, the first two factors are conditional Gaussian pdfs, and the latter is a conditional inverse-gamma pdf. This inverse-gamma posterior pdf for the variance follows from the inverse-gamma prior pdf and the Dirac-linear likelihood model, as they define a conjugate pair, as demonstrated in

Sect. 8.1. The model parameters in the posterior inverse-gamma pdf for the variance are

$$\begin{aligned} v_{\sigma^2|d} &= v_{\sigma^2} + m \\ \xi_{\sigma^2|d}^2 &= \frac{v_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)}{v_{\sigma^2} + m} \end{aligned}$$

with

$$\sigma(\mathbf{d}_e) = (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n)^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n).$$

The corresponding posterior model $p(\mathbf{r}_H \mid \mathbf{d}_e)$ is then an n -variate T-distributed pdf parametrised by $[\boldsymbol{\mu}_{r_H|d}, \boldsymbol{\Omega}_{r_H|d}, v_{\sigma^2|d}]$. This posterior model belongs to the class of hierarchical Gaussian models, as defined in Sect. 4.1.2, as does the prior model. Consequently, the hierarchical Gaussian RF models constitute a conjugate class with respect to the class of Dirac-linear likelihood models.

The corresponding predictor and prediction variance are

$$\begin{aligned} \hat{\mathbf{r}}_H &= E\{\mathbf{r}_H | \mathbf{d}_e\} = \boldsymbol{\mu}_{r_H|d} \\ \sigma_p^2 &= V\text{diag}_n\{\boldsymbol{\Sigma}_{r_H|d}\} = \frac{v_{\sigma^2|d}}{v_{\sigma^2|d} - 2} \times V\text{diag}_n\{\boldsymbol{\Omega}_{r_H|d}\}. \end{aligned}$$

The marginal prediction $(1 - \alpha)$ -intervals can be assessed by using the univariate T-distribution $\alpha/2$ -points, denoted $t_{\alpha/2}$. For a Dirac-point likelihood model, as defined in Sect. 5.1, with exact observations in a subset of the grid node, the predictor reproduces the observations exactly in these grid nodes. The predictions and prediction variances in this subset of nodes are $\mathbf{H}\hat{\mathbf{r}}_H = \mathbf{d}_{oe}$ and $\mathbf{H}\sigma_p^2 = 0\mathbf{i}_m$, as expected.

The linear functional predictor and prediction variance for $r_w = \mathbf{w}^T \mathbf{r}_H$ are expressed as

$$\begin{aligned} \hat{r}_{w_H} &= E\{\mathbf{w}^T \mathbf{r}_H | \mathbf{d}_e\} = \mathbf{w}^T \boldsymbol{\mu}_{r_H|d} \\ \sigma_p^2 &= \frac{v_{\sigma^2|d}}{v_{\sigma^2|d} - 2} \times \mathbf{w}^T \boldsymbol{\Omega}_{r_H|d} \mathbf{w}. \end{aligned}$$

Furthermore, explanatory variables can be included into the hierarchical Gaussian RF model without losing the conjugate characteristics, as demonstrated in Røislien and Omre (2006).

This reproduction of the T-distribution does not follow the more general class of Gauss-linear likelihood models. This is because the sum of two independent T-distributed random variables will not appear as T-distributed, except in the limiting case of the Gaussian pdf. For the particular case where the prior variance of the Gaussian RF is proportional to the variances of the observation errors. It can be demonstrated that the conjugate properties hold.

7.2 Event Spatial Variables

The class of misclassification likelihood models is defined in Sect. 5.2, and the corresponding classes of conjugate event prior models are defined in Sect. 6.4. The conjugate property of the prior model with respect to this likelihood model makes it easier to assess the posterior model. The posterior model belongs to the same class of RFs as the prior model, and the associated model parameters can be calculated analytically. This conjugate characteristic is demonstrated in this section. Figure 7.3 displays a graphical presentation of the Poisson inversion challenge in a format comparable to Fig. 2.1. The challenge is to determine the pdf along the $[\mathbf{k} | \mathbf{d}_0]$ -axis.

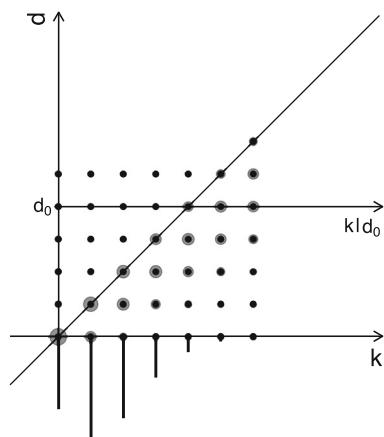
7.2.1 Poisson RF Models

The joint pdf for the spatial variable and the observations for $\mathbf{k}, \mathbf{d} \in \mathbb{N}_{\oplus}^n$ and $\mathbf{k} \geq \mathbf{d}$ is expressed as

$$\begin{aligned} \begin{bmatrix} \mathbf{k} \\ \mathbf{d} \end{bmatrix} &\sim p(\mathbf{k}, \mathbf{d}) = p(\mathbf{d} | \mathbf{k})p(\mathbf{k}) \\ &= \prod_{i=1}^n p(d_i | k_i) \prod_{i=1}^n p(k_i) = \prod_{i=1}^n p(d_i, k_i) \\ &= \prod_{i=1}^n \left[\binom{k_i}{d_i} \alpha_i^{d_i} [1 - \alpha_i]^{k_i - d_i} \times \frac{[\lambda_k^n]^{k_i}}{k_i!} \exp(-\lambda_k^n) \right] \\ &= \prod_{i=1}^n \left[\frac{[\alpha_i \lambda_k^n]^{d_i}}{d_i!} \times \frac{[(1 - \alpha_i) \lambda_k^n]^{k_i - d_i}}{(k_i - d_i)!} \times \exp(-\lambda_k^n) \right] \end{aligned} \quad (7.14)$$

with $\lambda_k^n = \lambda_k \Delta_n$ being the prior expected number of events in one grid unit.

Fig. 7.3 Bayesian inversion. Schematic presentation for an event count Poisson RF



This joint pdf is in factorial form, which entails independence between nodes. Hence, the normalising constant can be computed independently for each factor by summing over $k_i \geq d_i$ to obtain $p(d_i)$. The posterior model can be expressed for $\mathbf{k} \in \mathbb{N}_\oplus^n; \mathbf{k} \geq \mathbf{d}$ as

$$\begin{aligned} [\mathbf{k} | \mathbf{d}] \sim p(\mathbf{k} | \mathbf{d}) &= \prod_{i=1}^n [p(d_i)]^{-1} \times p(d_i, k_i) = \prod_{i=1}^n p(k_i | d_i) \\ &= \prod_{i=1}^n \frac{[(1 - \alpha_i)\lambda_k^n]^{k_i - d_i}}{(k_i - d_i)!} \exp(-(1 - \alpha_i)\lambda_k^n). \end{aligned} \quad (7.15)$$

This posterior model belongs to the class of Poisson count models, as defined in Expression (4.3), as does the prior model. Consequently, the Poisson count models constitute a conjugate class for the class of misclassification likelihood models.

The corresponding approximate posterior Poisson event-location set model is defined to be

$$\begin{aligned} [\mathbb{X}_{\mathbb{D}}^n | \mathbf{d}] \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{d}) &= \sum_{\substack{\mathbf{k} \in \mathbb{N}_\oplus^n \\ \mathbf{k} \geq \mathbf{d}}} p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{k}) p(\mathbf{k} | \mathbf{d}) \\ &= \prod_{i=1}^n \sum_{\substack{k_i \in \mathbb{N}_\oplus \\ k_i \geq d_i}} p(\mathbf{x}_i | k_i) p(k_i | d_i), \end{aligned} \quad (7.16)$$

where $p(\mathbf{x}_i | k_i)$ is defined in Expression (4.4).

By using infill asymptotic analysis, it can be demonstrated that this conjugate property also holds for Poisson RFs in general. The conditional probabilities in the nodes define the intensity field of the posterior Poisson RF $\{\lambda_{k|d}(\mathbf{x}) = (1 - \alpha(\mathbf{x}))\lambda_k; \mathbf{x} \in L\} \xrightarrow{\text{infill}} \{\lambda_{k|d}(\mathbf{x}) = (1 - \alpha(\mathbf{x}))\lambda_k; \mathbf{x} \in \mathbb{D}\}$ and the base event set $\{[\mathbf{x}|d_{\mathbf{x}} > 0]; \mathbf{x} \in L\} \xrightarrow{\text{infill}} \mathbb{E} = \mathbb{D}$, where \mathbb{D} is the observed event-location set. The posterior model becomes a non-stationary Poisson RF with the non-stationarity introduced by the variations in the observation probabilities and the conditioning on the observations.

The likelihood model for the accumulated event observation \mathbf{d}_A is defined in Sect. 5.2. The events are observed in m disjoint areas represented by sub-grid sets

$L_{A_i} \subset L$ of size n_{A_i} for $i = 1, 2, \dots, m$. The corresponding event count posterior model for a stationary Poisson prior model for $\mathbf{k} \in \mathbb{N}_{\oplus}^n$ and $\mathbf{k} \geq \mathbf{d}$ is

$$\begin{aligned} p(\mathbf{k} | \mathbf{d}_A) &= \prod_{i \in L_A^c} p(k_i) \times \prod_{j=1}^m p(k_i; i \in L_{A_j} | d_{A_j}) \\ &= \prod_{i \in L_A^c} p(k_i) \times \prod_{j=1}^m \sum_{\substack{d_i \\ i \in L_{A_j}}} p(k_i; i \in L_{A_j} | d_i; i \in L_{A_j}) p(d_i; i \in L_{A_j} | d_{A_j}) \\ &= \prod_{i \in L_A^c} p(k_i) \times \prod_{j=1}^m \sum_{\substack{d_i \\ i \in L_{A_j}}} \prod_{i \in L_{A_j}} p(k_i | d_i) p(d_i; i \in L_{A_j} | d_{A_j}). \end{aligned} \quad (7.17)$$

Note that the accumulated observation in each sub-domain d_{A_j} does not influence the event count in the grid units outside L_{A_j} . The event count in grid units inside L_{A_j} appears coupled with the weight factor $p(d_i; i \in L_{A_j} | d_{A_j})$, which is a multinomial pdf with a set of parameters $\{p_i = \text{const} \times \alpha_i \lambda_k^n; i \in L_{A_j}\}$. Therefore, the posterior model is a weighted average of the marginal posterior pdfs defined in Expression (7.15).

Alternatively, the prior stationary Poisson RF model can be represented by the m disjoint areas with event counts $k_{A_i} = \sum_{j \in L_{A_i}} k_j$ and with $\lambda_{A_i}^n = \sum_{j \in L_{A_i}} \lambda_k^n = n_{A_i} \lambda_k^n$ for $i = 1, 2, \dots, m$. The observation probabilities in the areas are $\alpha_{A_i} = n_{A_i}^{-1} \sum_{j \in L_{A_i}} \alpha_j$ for $i = 1, 2, \dots, m$. The corresponding event count posterior model for the m -vector $\mathbf{k}_A = (k_{A_1}, k_{A_2}, \dots, k_{A_m})^T; \mathbf{k}_A \in \mathbb{N}_{\oplus}^m, \mathbf{k}_A \geq \mathbf{d}_A$ is

$$p(\mathbf{k}_A | \mathbf{d}_A) = \prod_{i=1}^m \frac{[(1 - \alpha_{A_i}) \lambda_{A_i}^n]^{k_{A_i} - d_{A_i}}}{(k_{A_i} - d_{A_i})!} \exp(-(1 - \alpha_{A_i}) \lambda_{A_i}^n).$$

Note that the elements in \mathbf{k}_A are independent because the areas $A_i; i = 1, 2, \dots, m$ are disjoint. Moreover, the posterior model in the complement area A^c is identical to the prior stationary Poisson count RF model.

Spatial Discretisation Uncertainty

The posterior model is a non-stationary Poisson RF represented by a spatially discretised event count model $\{[k_{\Delta}(\mathbf{x}) | \mathbf{d}]; \mathbf{x} \in L\}$. Consider an actual event-location $\mathbf{x}^o \in D$ occurring in grid unit Δ_{ni^o} . Because the Poisson RF is non-stationary with spatially varying intensity, there are two sources of uncertainty: the intensity variations within the grid unit Δ_{ni^o} are unknown, and the exact location \mathbf{x}^o within Δ_{ni^o} is unknown. The uncertainties related to both effects decrease as the grid unit volume Δ_n decreases, as for infill asymptotic analysis with D constant and $n \rightarrow \infty$. Recall that in the corresponding event-location representation of the non-stationary posterior Poisson RF $[X_D^n | \mathbf{d}]$, the observed events will be correctly located. However,

the new event locations will be generated according to a piecewise constant intensity function based on the grid representation, which introduces uncertainty.

Simulation

A realisation of the event counts $\{[k_{\Delta}(\mathbf{x})|\mathbf{d}]; \mathbf{x} \in \mathcal{L}\}$, based on the non-stationary posterior Poisson RF model, can be generated by a sequential simulation algorithm, as defined in Algorithm 9. The realisation is represented by the spatially discretised count n -vector $[\mathbf{k}|\mathbf{d}]^s$ and the corresponding approximate event-location set $[\mathbb{X}_{\mathcal{D}}^n|\mathbf{d}]^s$. The variables in each grid node are independent; thus, the algorithm takes a trivial form.

Algorithm 9: Posterior Poisson RF: independent simulation

```

Initialise  $n_k = 0$ 
for  $i = 1, 2, \dots, n$  do
    Generate  $k_i^s \sim p(k_i | d_i)$ 
    for  $j = 1, 2, \dots, k_i^s$  do
        | Generate  $\mathbf{x}_{n_k+j}^s \sim \text{unif}[\Delta_{ni}]$ 
    end
    Set  $n_k = n_k + k_i^s$ 
end
Result:  $[\mathbf{k} | \mathbf{d}]^s = (k_1^s, k_2^s, \dots, k_n^s) \sim p(\mathbf{k} | \mathbf{d})$  and
 $[\mathbb{X}_{\mathcal{D}}^n | \mathbf{d}]^s = (\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{n_k}^s) \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{d})$ 

```

Alternatively, a rejection simulation algorithm, as defined in Algorithm 10, which is based on the definition of the limiting Poisson RF, can be used. The algorithm provides the realisation n -vector $[\mathbf{k}|\mathbf{d}]^s$ and set $[\mathbb{X}_{\mathcal{D}}^n|\mathbf{d}]^s$. Recall that the set of observed event locations is $\mathcal{D} : \{\mathbf{x}_1^d, \mathbf{x}_2^d, \dots, \mathbf{x}_{n_d}^d\}; \mathbf{x}_i^d \in \mathcal{D}$. This algorithm is expected to be more efficient than the previous one if the misclassification field $\{\alpha(\mathbf{x}); \mathbf{x} \in \mathcal{D}\}$ does not vary too much in value.

Prediction

Prediction of the spatial count variable $\{\hat{k}_{\Delta}(\mathbf{x}); \mathbf{x} \in \mathcal{L}\}$ represented by the n -vector $\hat{\mathbf{k}}$ is commonly defined by using a LSE criterion. Thus,

$$\hat{\mathbf{k}} = E\{\mathbf{k}|\mathbf{d}\} = \boldsymbol{\mu}_{\mathbf{k}|d} = \mathbf{d} + [i_n - \boldsymbol{\alpha}] \lambda_k^n. \quad (7.18)$$

The associated prediction variance matrix, of dimension $(n \times n)$, is diagonal because of the independence between grid nodes as

$$\boldsymbol{\Sigma}_p = \text{Var}\{\mathbf{k}|\mathbf{d}\} = \text{Mdiag}_n \{(1 - \alpha_i) \lambda_k^n\}_{i=1,2,\dots,n}. \quad (7.19)$$

The locationwise prediction variances appear on the diagonal. Furthermore, the associated prediction $(1 - \alpha)$ -intervals can be independently developed for each grid node from the respective Poisson pdfs.

Algorithm 10: Posterior Poisson RF: rejection sampling

```

Set  $k_i^s = d_i$  for  $i = 1, 2, \dots, n$ 
Set  $\mathbf{x}_i^s = \mathbf{x}_i^d$  for  $i = 1, 2, \dots, n_d$ 
 $\lambda_\delta = \lambda_k^n \sum_{i=1}^n (1 - \alpha_i)$ 
 $\lambda_M = \max_{i=1,2,\dots,n} \{(1 - \alpha_i) \lambda_k^n\}$ 
Generate  $k_\delta \sim p(k) = \frac{\lambda_\delta^k}{k!} \exp\{-\lambda_\delta\}$ 
for  $i = 1, 2, \dots, k_\delta$  do
    while  $b = 1, 2, \dots$  do
        Generate  $j \sim \text{unif}[1, 2, \dots, n]$ 
        Calculate  $\alpha = \frac{\lambda_j}{\lambda_M} = \frac{(1-\alpha_j)\lambda_k^n}{\lambda_M}$ 
        Generate  $u \sim \text{unif}[0, 1]$ 
        if  $u \leq \alpha$  then
            Set  $k_j^s = k_j^s + 1$ 
            Generate  $\mathbf{x}_{n_d+i}^s \sim \text{unif}[\Delta_{nj}]$ 
            break
        end
    end
end

Result:  $[\mathbf{k} \mid \mathbf{d}]^s = (k_1^s, k_2^s, \dots, k_n^s) \sim p(\mathbf{k} \mid \mathbf{d})$ ,  $n_k = n_d + k_\delta$  and
 $[\mathbb{X}_D^n \mid \mathbf{d}]^s = [\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{n_k}^s] \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \mid \mathbb{D})$ 

```

Functionals

The posterior model for the linear functional $k_w = \mathbf{w}^T \mathbf{k}$ is evaluated for binary n -vectors \mathbf{w} ; $w_i \in \{0, 1\}$ and $k_w \in \mathbb{N}_\oplus$. This linear functional corresponds to the spatial integral over $A_w = \bigcup_i I(w_i = 1) \Delta_{ni}$. The expression of interest is

$$\begin{aligned} [k_w \mid \mathbf{d}] &\sim p(k_w \mid \mathbf{d}) = \sum_{\mathbf{k} \geq \mathbf{d}} p(k_w \mid \mathbf{k}) p(\mathbf{k} \mid \mathbf{d}) \\ &= \sum_{\mathbf{k} \geq \mathbf{d}} p(k_w \mid \mathbf{k}) \prod_{i=1}^n p(k_i \mid d_i). \end{aligned}$$

Recall that $p(k_i \mid d_i); i = 1, 2, \dots, n$ are Poisson pdfs. Thus, $[k_w \mid \mathbf{d}]$ is a sum of independent Poisson random variables. It can be demonstrated that this sum is also a Poisson random variable. Therefore, the pdf of interest for $k_w \in \mathbb{N}_\oplus; k_w \geq d_w$, is

$$[k_w \mid \mathbf{d}] \sim p(k_w \mid \mathbf{d}) = \frac{[\lambda_{k_w|d}^n]^{k_w-d_w}}{(k_w - d_w)!} \exp(-\lambda_{k_w|d}^n) \quad (7.20)$$

with

$$\begin{aligned} \lambda_{k_w|d}^n &= \mathbf{w}^T [\mathbf{i}_n - \boldsymbol{\alpha}] \lambda_k^n \\ d_w &= \mathbf{w}^T \mathbf{d}. \end{aligned}$$

Therefore, a set of realisations $\{[k_w | \mathbf{d}]^s; s = 1, 2, \dots, n_s\}$ from $p(k_w | \mathbf{d})$ can be generated by a standard Poisson pdf simulation algorithm. The predictor for the linear combination is $\hat{k}_w = E\{k_w | \mathbf{d}\} = d_w + \lambda_{k_w | d}^n$, and the associated prediction variance is $\sigma_p^2 = \text{Var}\{k_w | \mathbf{d}\} = \lambda_{k_w | d}^n$. These results follow from the properties of the Poisson pdf.

Explanatory Variables

Explanatory spatial variables may be available, and these can be included into the Poisson prior model, as defined in Sect. 6.5. The grid representation of the explanatory variables are $\{g_j(\mathbf{x}); \mathbf{x} \in L\}; j = 1, 2, \dots, n_g$ represented in the n -vectors $\mathbf{g}_j; j = 1, 2, \dots, n_g$. The regression Poisson prior model is defined with regression coefficients: intensity level λ_k^0 and an n_g -vector $\boldsymbol{\beta}_k$ with slopes. In the discretised representation, we use $\lambda_k^{0n} = \lambda_k^0 \Delta_n$, and the corresponding posterior model for $\mathbf{k} \in \mathbb{N}_{\oplus}; \mathbf{k} \geq \mathbf{d}$ is then

$$\begin{aligned} [\mathbf{k} | \mathbf{d}] \sim p(\mathbf{k} | \mathbf{d}) &= \prod_{i=1}^n [p(d_i)]^{-1} \times p(d_i, k_i) = \prod_{i=1}^n p(k_i | d_i) \\ &= \prod_{i=1}^n \frac{[(1 - \alpha_i)\lambda_k^{0n} \prod_{j=1}^{n_g} \exp(\beta_k^j g_{ji})]^{k_i - d_i}}{(k_i - d_i)!} \\ &\quad \times \exp(-(1 - \alpha_i)\lambda_k^{0n} \prod_{j=1}^{n_g} \exp(\beta_k^j g_{ji})). \end{aligned} \quad (7.21)$$

This posterior model represents a spatially discretised Poisson count RF; hence, this extended regression Poisson prior model is also conjugate with respect to a misclassification likelihood model. Simulation and prediction can be performed, as in the case of a stationary prior model.

7.2.2 Example: Poisson RF

The example discussed is introduced in Sect. 5.2.1 and Fig. 5.2. The unknown event spatial variable $\{k(\mathbf{x}); \mathbf{x} \in D\}$ to be reconstructed is given in display (a). The available observations are presented in display (g). The model is specified on the spatially discretised event count spatial variable, represented by the n -vector \mathbf{k} . The misclassification likelihood model is $p(\mathbf{d} | \mathbf{k})$ with the observation n -vector \mathbf{d} presented in display (h). In Sect. 6.5.5 the prior stationary Poisson RF model is assigned. Thus, the prior model $p(\mathbf{k})$ is according to a Poisson pdf.

The Poisson RF prior model $p(\mathbf{k})$ is demonstrated to be conjugate with respect to misclassification likelihood models $p(\mathbf{d} | \mathbf{k})$. Therefore, the posterior model $p(\mathbf{k} | \mathbf{d})$ is also Poisson distributed, with the model parameter event intensity n .

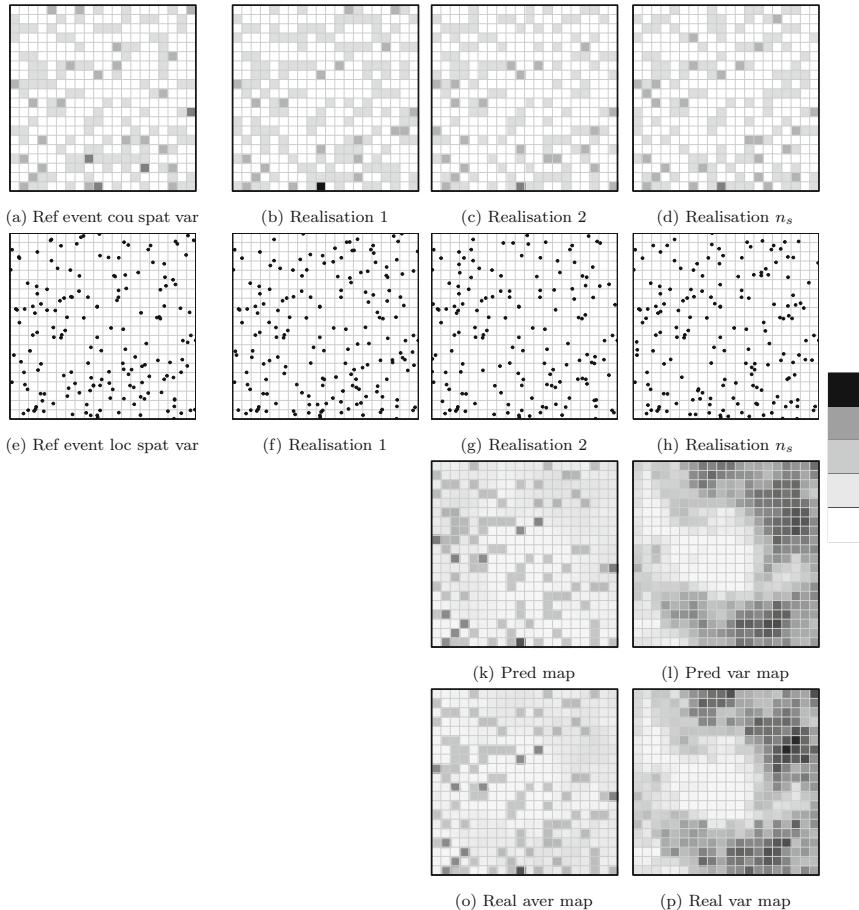


Fig. 7.4 Example Poisson RF model. Poisson posterior model: **(a)** reference event count spatial variable; **(b), (c), (d)** three independent realisations; **(e)** reference event-location spatial variable; **(f), (g), (h)** three independent realisations; **(k)** prediction map; **(l)** prediction variance map; **(o)** realisation average map; and **(p)** realisation variance map

vector $\lambda_{k|d}$. The posterior model parameter is analytically assessed from the model parameters of the likelihood and prior models and the observations.

Figure 7.4 contains the reference spatial variable in event count format in display (a) and in event-location format in display (e). Three independent realisations from the posterior model $p(\mathbf{k} | \mathbf{d})$ are generated in event count format and are presented in displays (b), (c) and (d). The corresponding realisations in event-location format are presented in displays (f), (g) and (h). These posterior realisations reproduce the observed events and appear as possible outcomes of the reference spatial variable given the observations. The event counts prediction $E\{\mathbf{k}|\mathbf{d}\}$ and the corresponding prediction variance $\text{Var}\{\mathbf{k}|\mathbf{d}\}$ are analytically assessed and presented

in displays (k) and (l). The prediction map exhibits less variability than the reference spatial variable and is favourable for locationwise prediction. The misclassification probability map highly influences the corresponding prediction variance map in the likelihood model. Based on 200 posterior realisations, the prediction and prediction variance maps are estimated using the realisation average and empirical variance and they are presented in displays (o) and (p). The estimates typically improve with an increasing number of posterior realisations.

7.2.3 Hierarchical Poisson RF Models

The stationary Poisson RF is cast in a hierarchical setting in Sect. 6.6. The discretised stationary hierarchical Poisson count RF $\{k_H(\mathbf{x}); \mathbf{x} \in L\}$ is represented by the n -vector \mathbf{k}_H . Recall that this RF model is not ergodic. Thus, care should be taken when performing model parameter inference and when interpreting the prediction results from this model. Consider a misclassification likelihood model, as defined in Sect. 5.2, and let the prior model be a discretised stationary hierarchical Poisson RF model as in Sect. 6.6. The posterior model is then demonstrated to be a negative-multinomial pdf. The posterior model, for $\mathbf{k}_H \in \mathbb{N}_{\oplus}$; $\mathbf{k}_H \geq \mathbf{d}$, can be expressed as

$$\begin{aligned} [\mathbf{k}_H | \mathbf{d}] &\sim p(\mathbf{k}_H | \mathbf{d}) = \int_{\mathbb{R}_{\oplus}} p(\mathbf{k}_H | \mathbf{d}, \lambda_k) p(\lambda_k | \mathbf{d}) d\lambda_k \\ &= [\Gamma(v_{\lambda|d})]^{-1} \Gamma\left(\sum_{i=1}^n (k_{Hi} - d_i) + v_{\lambda|d}\right) \kappa_{\lambda|d}^{v_{\lambda|d}} \prod_{i=1}^n \frac{[\omega_{(\lambda|d)i}(1 - \kappa_{\lambda|d})]^{k_{Hi} - d_i}}{(k_{Hi} - d_i)!} \end{aligned} \quad (7.22)$$

with $\kappa_{\lambda|d} = [1 + \xi_{\lambda|d}^{-1} \Delta_n [\mathbf{i}_n - \boldsymbol{\alpha}]^T \mathbf{i}_n]^{-1} < 1$ and $\omega_{\lambda|d} = (\mathbf{i}_n - \boldsymbol{\alpha})[[\mathbf{i}_n - \boldsymbol{\alpha}]^T \mathbf{i}_n]^{-1}$. In the expression, the former factor is a Poisson pdf, and the latter factor is a conditional gamma pdf. This posterior gamma pdf for the intensity appears from the gamma prior pdf and the misclassification likelihood model because they define a conjugate pair, as demonstrated in Sect. 8.2. The model parameters in the posterior gamma pdf for the intensity are

$$\begin{aligned} v_{\lambda|d} &= v_{\lambda} + n_d \\ \xi_{\lambda|d} &= \xi_{\lambda} + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n, \end{aligned}$$

where n_d is the number of event observations in the set \mathbb{D} and $\boldsymbol{\alpha}$ is the n -vector of probabilities of correctly observing an event.

The corresponding posterior model $p(\mathbf{k}_H | \mathbf{d})$ is then a negative-multinomial pdf parametrised by $[v_{\lambda|d}, \kappa_{\lambda|d}, \omega_{\lambda|d}]$. This posterior model belongs to the class of hierarchical Poisson count models, as defined in Sect. 4.2.2, as does the prior model.

Thus, the hierarchical Poisson count models constitute a conjugate class with respect to the class of misclassification likelihood models.

The corresponding predictor and prediction variance are

$$\begin{aligned}\hat{\mathbf{k}}_H &= E\{\mathbf{k}_H | \mathbf{d}\} = \boldsymbol{\mu}_{k_H|d} = \mathbf{d} + \kappa_{\lambda|d}^{-1} v_{\lambda|d} [1 - \kappa_{\lambda|d}] \boldsymbol{\omega}_{\lambda|d} \\ \sigma_p^2 &= V\text{diag}_n \{ \boldsymbol{\Sigma}_{k_H|d} \} \\ &= V\text{diag}_n \left\{ \kappa_{\lambda|d}^{-1} v_{\lambda|d} [1 - \kappa_{\lambda|d}] M\text{diag}_n \{ \boldsymbol{\omega}_{\lambda|d} \} + \kappa_{\lambda|d}^{-2} v_{\lambda|d} [1 - \kappa_{\lambda|d}]^2 \boldsymbol{\omega}_{\lambda|d} \boldsymbol{\omega}_{\lambda|d}^T \right\} \\ &= \kappa_{\lambda|d}^{-1} v_{\lambda|d} [1 - \kappa_{\lambda|d}] \boldsymbol{\omega}_{\lambda|d} + \kappa_{\lambda|d}^{-2} v_{\lambda|d} [1 - \kappa_{\lambda|d}]^2 V\text{diag}_n \left\{ \boldsymbol{\omega}_{\lambda|d} \boldsymbol{\omega}_{\lambda|d}^T \right\}.\end{aligned}$$

For an observation likelihood model without misclassification $\boldsymbol{\alpha} = \mathbf{i}_n$, then $\kappa_{\lambda|d} = 1$ and the predictor is $\hat{\mathbf{k}}_H = \mathbf{d}$ with prediction variance $\sigma_p^2 = 0\mathbf{i}_n$, as expected. If misclassification occurs, $\kappa_{\lambda|d} < 1$, but if $\alpha_i = 1$ for a particular grid unit i , then $\omega_{\lambda|d_i} = 0$ and the predictor is $\hat{k}_{H_i} = d_i$ with $\sigma_{pi}^2 = 0$.

The linear functional predictor and prediction variance for $k_{w_H} = \mathbf{w}^T \mathbf{k}_H$ are

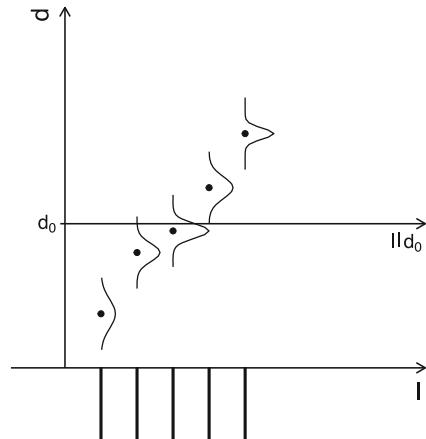
$$\begin{aligned}\hat{k}_{w_H} &= E\{\mathbf{w}^T \mathbf{k}_H | \mathbf{d}_e\} = \mathbf{w}^T \boldsymbol{\mu}_{k_H|d} \\ \sigma_p^2 &= \mathbf{w}^T \boldsymbol{\Sigma}_{k_H|d} \mathbf{w}.\end{aligned}$$

Furthermore, explanatory variables can be included into the hierarchical Poisson RF model without losing the conjugate characteristics.

7.3 Mosaic Spatial Variables

The class of response likelihood models is defined in Sect. 5.3, and the corresponding class of conjugate mosaic prior models is defined in Sect. 6.7. This conjugate characteristic also entails that the posterior model is in the same class of RFs as the prior model, with the model parameters analytically assessable. This conjugate characteristic is demonstrated in this section. In Fig. 7.5, a graphical presentation of the Markov inversion challenge is displayed, in a format comparable to Fig. 2.1. The challenge is to determine the pdf along the $[\mathbf{l}|\mathbf{d}_0]$ -axis.

Fig. 7.5 Bayesian inversion.
Schematic presentation for a
mosaic Markov RF



7.3.1 Markov RF Models

The joint pdf for the spatial variable of interest and the observations is based on the Gibbs formulation of the Markov RF prior model

$$\begin{aligned} \left[\begin{matrix} \mathbf{l} \\ \mathbf{d} \end{matrix} \right] &\sim p(\mathbf{l}, \mathbf{d}) = p(\mathbf{d} \mid \mathbf{l}) p(\mathbf{l}) \\ &= \text{const} \times \prod_{i=1}^n p(d_i \mid l_i) \times \prod_{i=1}^n v_{0l}(l_i) \prod_{\mathbf{c}' \in \mathcal{C}'_{\mathbf{l}}} v_{1l}(l_j; j \in \mathbf{c}'_l). \end{aligned} \quad (7.23)$$

Recall that the normalising constant is very computationally demanding to calculate, as is the marginal pdf $p(\mathbf{d})$, as they both require summation over $\mathbf{l} \in \mathbb{L}^n$.

The associated Markov formulation of the joint model, for $i = 1, 2, \dots, n$, is

$$\begin{aligned} \left[\begin{matrix} l_i \\ \mathbf{d} \end{matrix} \mid \mathbf{l}_{-i} \right] &\sim p(l_i, \mathbf{d} \mid \mathbf{l}_{-i}) = p(\mathbf{d} \mid \mathbf{l}) p(l_i \mid \mathbf{l}_{-i}) \\ &= \text{const} \times \prod_{j=1}^n p(d_j \mid l_j) \times v_{0l}(l_i) \times \omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l). \end{aligned}$$

The normalising constants can be calculated by summation over only $l_i \in \mathbb{L}$. The corresponding posterior model in the Markov formulation is obtained, for $i = 1, 2, \dots, n$, as

$$\begin{aligned} \left[l_i \mid \mathbf{d} \right]_{\mathbf{l}_{-i}} &\sim p(l_i \mid \mathbf{d}, \mathbf{l}_{-i}) = \frac{p(l_i, \mathbf{d} \mid \mathbf{l}_{-i})}{p(\mathbf{d} \mid \mathbf{l}_{-i})} \\ &= \text{const} \times \frac{p(l_i, d_i, \mathbf{d}_{-i} \mid \mathbf{l}_{-i})}{p(\mathbf{d}_{-i} \mid \mathbf{l}_{-i})} \\ &= \text{const} \times p(l_i, d_i \mid \mathbf{d}_{-i}, \mathbf{l}_{-i}) \\ &= \text{const} \times p(d_i \mid l_i) \times v_{0l}(l_i) \times \omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l) \end{aligned} \quad (7.24)$$

with

$$\text{const} = \left[\sum_{l'_i \in \mathbb{L}} p(d_i \mid l'_i) \times v_{0l}(l'_i) \times \omega_l(l'_i \mid l_j; j \in \mathbf{n}_i^l) \right]^{-1}.$$

Recall that $p(\mathbf{d}_{-i} \mid \mathbf{l}_{-i}) = \prod_{j=1, j \neq i}^n p(d_j \mid l_j)$. The normalising constant can easily be calculated, as it only requires summation over $l_i \in \mathbb{L}$. The posterior Markov formulation appears as a non-stationary Markov RF with proportion functions $v_{0i}(l_i \mid d_i) = p(d_i \mid l_i)v_{0l}(l_i)$ and interaction functions $\omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l)$. The former is location-dependent because of the conditioning on the observations \mathbf{d} . The latter is, however, identical to that in the prior model and is thus dependent only on the variables in the neighbourhood \mathbf{n}_i^l . Recall that the prior model is a spatially discretised Markov RF. Consequently, the spatially discretised Markov RF constitutes a class of conjugate prior models with respect to the response likelihood model class.

The corresponding Markov formulation for the particular case with a two-nearest-node clique system having different horizontal and vertical couplings, for $i = 1, 2, \dots, n$, is

$$\begin{aligned} \left[l_i \mid \mathbf{d} \right]_{\mathbf{l}_{-i}} &\sim p(l_i \mid \mathbf{d}, \mathbf{l}_{-i}) = \text{const} \times p(d_i \mid l_i)v_{0l}(l_i) \\ &\quad \times \omega_{lh}(l_i \mid l_j; h < i, j >, j \in \mathbf{L})\omega_{lv}(l_i \mid l_j; v < i, j >, j \in \mathbf{L}) \end{aligned}$$

with the notation defined in Sect. 6.7. The normalising constant can easily be computed by summing over $l_i \in \mathbb{L}$.

For the convolved mosaic spatial variable observations \mathbf{d}_b , defined in Sect. 5.3, with identical neighbourhood design, the corresponding posterior model for a stationary Markov prior model is

$$\begin{aligned} \left[l_i \mid \mathbf{d}_b \right] \sim p(l_i \mid \mathbf{d}_b, \mathbf{l}_{-i}) &= \frac{p(l_i, \mathbf{d}_b \mid \mathbf{l}_{-i})}{p(\mathbf{d}_b \mid \mathbf{l}_{-i})} \\ &= \text{const} \times \prod_{k \in \mathbf{n}_i^d} p(d_{bk} \mid l_j; j \in \mathbf{n}_k^d) \times v_{0l}(l_i) \\ &\quad \times \omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l) \\ &= \text{const} \times v_{0l}(l_i) \times \omega_{ld}(l_i \mid l_j, d_{bk}; j \in \mathbf{n}_i^{ld}, k \in \mathbf{n}_i^d). \end{aligned}$$

The expression contains both the convolution neighbourhood in the likelihood, \mathbf{n}_i^d , and the coupling neighbourhood in the prior \mathbf{n}_i^l . The Markov posterior model in grid node i must account for the influence of the labels in both these neighbourhood systems to comply with the prior model and to condition on the observations. Therefore, the extended posterior neighbourhood is $\mathbf{n}_i^{ld} = \{j \mid \cup_{k \in \mathbf{n}_i^d} j \in \mathbf{n}_k^d\} \cup \{j \mid j \in \mathbf{n}_i^l\} \setminus i$. The normalising constant can easily be computed by summing over $l_i \in \mathbb{L}$. This posterior model belongs to the class of Markov RF models but has an extended neighbourhood relative to the Markov RF prior model.

Spatial Discretisation Uncertainty

The posterior model is a non-stationary Markov RF represented on a spatial grid $\{[l(\mathbf{x})|\mathbf{d}]; \mathbf{x} \in \mathcal{L}\}$. The lack of precision in the spatially discretised likelihood model and the distortions of the discretised prior model influence the discretisation uncertainty in the posterior model. If the full spatial representation is obtained by assigning each grid unit the label of the associated grid node, the label proportions are reproduced without bias. The label bodies, defined as spatially closed volumes of the same label, are distorted as for the prior model. The bodies are fewer but larger than in the correct posterior mosaic model. This distortion effect is reduced by increasing the grid density.

Simulation

A realisation of the non-stationary posterior RF $\{[l(\mathbf{x})|\mathbf{d}]; \mathbf{x} \in \mathcal{L}\}$, represented by the n -vector $[\mathbf{l}|\mathbf{d}]^s$, can be generated by an iterative simulation algorithm, as defined in Sect. 10.1. The algorithm most frequently used is an McMC/Gibbs algorithm with a single-site proposal pdf based on the Markov formulation for the Markov RF, as specified in Algorithm 11. The acceptance probability α is one for all proposals, which is demonstrated by using the factorisation $p(\mathbf{l}) = p(l_i \mid \mathbf{l}_{-i})p(\mathbf{l}_{-i})$ in the computation of the acceptance probability in a general McMC algorithm.

The McMC algorithms are only asymptotically correct, and the rate of distributional convergence is important. For an arbitrary initial state, the number of iterations b required for \mathbf{s}^b to be distributed approximately according to the pdf $p(\mathbf{s})$ is referred to as the burn-in time. Formal calculations of this burn-in time

Algorithm 11: Posterior Markov RF: single-site McMC/Gibbs simulation

```

Initialise  $\mathbf{l}^0$  such that  $p(\mathbf{l}^0 | \mathbf{d}) > 0$ 

define  $g(\mathbf{l}' | \mathbf{l})$ :
  | Generate  $i \sim \text{unif}[1, 2, \dots, n]$ 
  | Generate  $l'_i \sim p(l_i | d_i, l_j; j \in \mathbf{n}_i')$ 
  | return  $\mathbf{l}' = (l_1, \dots, l_{i-1}, l'_i, l_{i+1}, \dots, l_n)$ 
end

for  $b = 1, 2, \dots, n_b$  do
  | Generate  $\mathbf{l}^b \sim g(\mathbf{l} | \mathbf{l}^{b-1})$ 
end

Result:  $\mathbf{l}^{n_b} \sim p^{n_b}(\mathbf{l}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{l} | \mathbf{d})$ 

```

for general models are not feasible, but many heuristic approaches are used in practice. One common approach is based on a convergence indicator, such as label proportions if the variable is a mosaic spatial variable, and then displaying the trace of this indicator during the iterations. Consider several extreme initial states, such as all the same label or random spatial mixing, and plot all the corresponding indicator traces in one display. The burn-in time is defined as the iteration number at which these traces approximately merge. In spatial studies, the term “map sweeps”, corresponding to the iteration number divided by the grid size n , is often used. The convergence indicator is a function of the realisations of the RF. Thus, the indicator appears as a realisation of a random variable. Consequently, equality of the indicators will not occur, only similar variability. After convergence, subsequent \mathbf{s}^b will remain dependent. The iteration difference $|b - b'|$ for which \mathbf{s}^b and $\mathbf{s}^{b'}$ appear as approximately independent is referred to as the mixing time. The mixing time can be determined heuristically by evaluating long iteration sequences of \mathbf{s}^b after convergence and testing for independence. The rates of convergence and mixing for three-dimensional Markov RFs discretised to a grid are typically very low, often unacceptably so.

Rephrasing the posterior three-dimensional Markov RF as a profile Markov RF, as demonstrated in Sect. 4.3, only the McMC iterations are required in the horizontal dimensions. A very efficient sequential simulation algorithm can be used in the vertical dimension, as specified in Fjellveit (2021). To reduce notation, this algorithm is only presented for the particular case with a neighbourhood design consisting of the six nearest nodes, corresponding to the two-nearest-node clique design. The algorithm appears as a McMC/block Gibbs algorithm, with each vertical profile generated sequentially from the conditioned Markov RC, which is assessed recursively, as specified in Sect. 10.8 and in Algorithm 12.

This profile-based algorithm’s convergence and mixing rates are comparable to those for two-dimensional models, as demonstrated in Fjellveit (2021), and hence are much faster than for the single-site proposal algorithm. This profile-based algorithm can be extended to a Markov RF model with arbitrary clique systems, at the expense of more complex notation.

Algorithm 12: Posterior Markov RF: profile McMC/Gibbs simulation

Special case: six-nearest-node neighbourhood $\mathbf{n}^l \rightarrow \mathbf{n}'_y = 4, \mathbf{n}'_z = 2$

Initialise \mathbf{l}^0 such that $p(\mathbf{l}^0 | \mathbf{d}) > 0$

define $g(\mathbf{l}' | \mathbf{l})$:

Generate $y^p \sim \text{unif}[1, 2, \dots, n_y]$

Calculate

$$\begin{aligned} p(l_{y^p n_z} | l_{y^p(n_z-1)}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p}) &= \text{const} \times p(d_{y^p n_z} | l_{y^p n_z}) v_{0l}(l_{y^p n_z}) \\ &\times \omega_{lh}(l_{y^p n_z} | l_j; j \in \mathbf{n}'_{y^p n_z}) \times \omega_{lv}(l_{y^p n_z} | l_{y^p(n_z-1)}) \end{aligned}$$

where

$$\begin{aligned} \text{const} &= \left[\sum_{l'_{y^p n_z} \in \mathbb{L}} p(d_{y^p n_z} | l'_{y^p n_z}) v_{0l}(l'_{y^p n_z}) \right. \\ &\quad \left. \times \omega_{lh}(l'_{y^p n_z} | l_j; j \in \mathbf{n}'_{y^p n_z}) \times \omega_{lv}(l'_{y^p n_z} | l_{y^p(n_z-1)}) \right]^{-1} \end{aligned}$$

for $i = (n_z - 1), \dots, 2$ **do**

Calculate

$$\begin{aligned} p(l_{y^p i} | l_{y^p(i-1)}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p}) &= \text{const} \times p(d_{y^p i} | l_{y^p i}) v_{0l}(l_{y^p i}) \\ &\times \frac{\omega_{lh}(l_{y^p i} | l_j; j \in \mathbf{n}'_{y^p i}) \omega_{lv}(l_{y^p i} | l_{y^p(i-1)}, l_{y^p(i+1)})}{p(l_{y^p(i+1)} | l_{y^p i}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})} \end{aligned}$$

where

$$\begin{aligned} \text{const} &= \left[\sum_{l'_{y^p i} \in \mathbb{L}} p(d_{y^p i} | l'_{y^p i}) v_{0l}(l'_{y^p i}) \right. \\ &\quad \left. \times \frac{\omega_{lh}(l'_{y^p i} | l_j; j \in \mathbf{n}'_{y^p i}) \omega_{lv}(l'_{y^p i} | l_{y^p(i-1)}, l_{y^p(i+1)})}{p(l_{y^p(i+1)} | l'_{y^p i}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})} \right]^{-1} \end{aligned}$$

end

Calculate

$$\begin{aligned} p(l_{y^p 1} | \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p}) &= \text{const} \times p(d_{y^p 1} | l_{y^p 1}) v_{0l}(l_{y^p 1}) \\ &\times \frac{\omega_{lh}(l_{y^p 1} | l_j; j \in \mathbf{n}'_{y^p 1}) \omega_{lv}(l_{y^p 1} | l_{y^p 2})}{p(l_{y^p 2} | l_{y^p 1}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})} \end{aligned}$$

where

$$\text{const} = \left[\sum_{l'_{y^p 1} \in \mathbb{L}} p(d_{y^p 1} | l'_{y^p 1}) v_{0l}(l'_{y^p 1}) \times \frac{\omega_{lh}(l'_{y^p 1} | l_j; j \in \mathbf{n}'_{y^p 1}) \omega_{lv}(l'_{y^p 1} | l_{y^p 2})}{p(l_{y^p 2} | l'_{y^p 1}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})} \right]^{-1}$$

Algorithm 12: (Continued)

```

Generate  $l'_{y^p1} \sim p(l_{y^p1} | \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})$ ;
for  $i = 2, 3, \dots, n_z$  do
| Generate  $l'_{y^pi} \sim p(l_{y^pi} | l'_{y^{p(i-1)}}, \mathbf{d}_{y^p}, \mathbf{l}_u; u \in \mathbf{n}'_{y^p})$ ;
end
return  $\mathbf{l}' = (\mathbf{l}_1, \dots, \mathbf{l}_{y^{p-1}}, \mathbf{l}'_{y^p}, \mathbf{l}_{y^p+1}, \dots, \mathbf{l}_{n_y})$ 
end
for  $b = 1, 2, \dots, n_b$  do
| Generate  $\mathbf{l}^b \sim g(\mathbf{l} | \mathbf{l}^{b-1})$ ;
end
Result:  $\mathbf{l}^{n_b} \sim p^{n_b}(\mathbf{l}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{l} | \mathbf{d})$ 

```

Prediction

Prediction of the spatial variable $\{\hat{l}(\mathbf{x}); \mathbf{x} \in \mathbb{L}\}$ represented by the n -vector $\hat{\mathbf{l}}$ is usually defined based on a maximum a posteriori (MAP) criterion, specifically a maximum marginal a posteriori (MMAP) criterion to make computation feasible. The MMAP predictor is

$$\hat{\mathbf{l}} = [\hat{l}_i = \operatorname{argmax}_{l_i \in \mathbb{L}} \{p(l_i | \mathbf{d})\}]_{i=1,2,\dots,n}.$$

The probabilities $p(l_i = l | \mathbf{d})$ for $i = 1, 2, \dots, n$ and $l \in \mathbb{L}$ provide the associated locationwise uncertainty. For each label $l \in \mathbb{L}$, we define the probability spatial variable $\{p(l(\mathbf{x}) = l | \mathbf{d}); \mathbf{x} \in \mathbb{L}\}$, which is represented by the n -vector \mathbf{p}_l and defined as follows:

$$\mathbf{p}_l = [p_{l_i} = p(l_i = l | \mathbf{d})]_{i=1,2,\dots,n}.$$

Both the prediction and probability spatial variables must be assessed by simulation-based inference, which entails generating a set of realisation $\{[\mathbf{l} | \mathbf{d}^s; s = 1, 2, \dots, n_s\}$ and assessing them by counting estimators. The estimated prediction is the most frequently occurring label in each grid node, and the estimated label probabilities are the relative frequency of each label in each grid node.

Alternatively, one may specify the prediction uncertainty by the variance map, represented by the n -vector σ_p^2 , for each $l \in \mathbb{L}$, as

$$\sigma_p^2 = [\sigma_{p_i}^2 = p_{l_i}(1 - p_{l_i})]_{i=1,2,\dots,n}.$$

The estimate is based on a binomial model and the estimated label probabilities in each grid node.

Explanatory Variables

Explanatory variables may be available, which can be included into the Markov prior model, as defined in Sect. 6.8. The explanatory variables are represented by a set of n -vectors $\mathbf{g}_j; j = 1, 2, \dots, n_g$ for label $l^* \in \mathbb{L}$. The regression coefficients thereby define the regression Markov RF prior in the Markov formulation: a level

$v_{0l}^0(\cdot)$ and slopes β_l , along with the spatial coupling functions $\omega_l(\cdot|\cdot)$. Consequently, the corresponding posterior model, for $i = 1, 2, \dots, n$, is

$$\begin{aligned} \left[l_i \mid \mathbf{d}, \mathbf{l}_{-i} \right] &\sim p(l_i \mid \mathbf{d}, \mathbf{l}_{-i}) = p(l_i \mid d_i, l_j; j \in \mathbf{n}_i^l) \\ &= \text{const} \times p(d_i \mid l_i) \times v_{0l}^0(l_i) \times \left[\prod_{j=1}^{n_g} \exp(\beta_l^j g_{ji}) \right]^{\mathbb{I}(l_i=l^*)} \\ &\quad \times \omega_l(l_i \mid l_j; j \in \mathbf{n}_i^l) \end{aligned} \quad (7.25)$$

with

$$\text{const} = \left[\sum_{l'_i \in \mathbb{L}} p(d_i \mid l'_i) \times v_{0l}^0(l'_i) \times \left[\prod_{j=1}^{n_g} \exp(\beta_l^j g_{ji}) \right]^{\mathbb{I}(l'_i=l^*)} \times \omega_l(l'_i \mid l_j; j \in \mathbf{n}_i^l) \right]^{-1}.$$

This posterior model is a Markov RF model. Thus, this extended regression Markov prior model is also a conjugate prior model with respect to the class of response likelihood models. Simulation and prediction can be obtained, as in the case of a stationary prior model.

7.3.2 Example: Markov RF

The example discussed is introduced in Sect. 5.3.1 and Fig. 5.3. The unknown mosaic spatial variable $\{l(\mathbf{x}); \mathbf{x} \in \mathcal{L} \subset \mathcal{D}\}$ to be reconstructed is given in display (a), represented by the n -vector \mathbf{l} . The available observations, represented by the n -vector \mathbf{d} , are presented in display (c), and the associated response likelihood model $p(\mathbf{d} \mid \mathbf{l})$ is specified. In Sect. 6.8.5, the Markov RF prior model $p(\mathbf{l})$ is assigned.

The Markov RF prior model $p(\mathbf{l})$ is demonstrated to be conjugate with respect to the response likelihood model $p(\mathbf{d} \mid \mathbf{l})$. Therefore, the posterior model $p(\mathbf{l} \mid \mathbf{d})$ is also a Markov RF model but with spatially non-stationary proportions of white and black. In this example, the spatial coupling is identical to the coupling in the prior model.

Figure 7.6 contains the reference spatial variable in display (a), as well as three approximately independent realisations from the posterior Markov RF model $p(\mathbf{l} \mid \mathbf{d})$ in displays (b), (c) and (d). The realisations are generated by the iterative McMC/Gibbs algorithm, as specified in Algorithm 11. Figure 7.7 presents the associated distribution convergence plot. The proportion of white serves as a convergence indicator, and traces with different initial states are plotted. The initial states are all white, 10% white, all black and a random spatial mix with equal proportions of each. Convergence, corresponding to the burn-in time, is considered to occur after 100,000 map sweeps, whereas independence mixing time is assumed

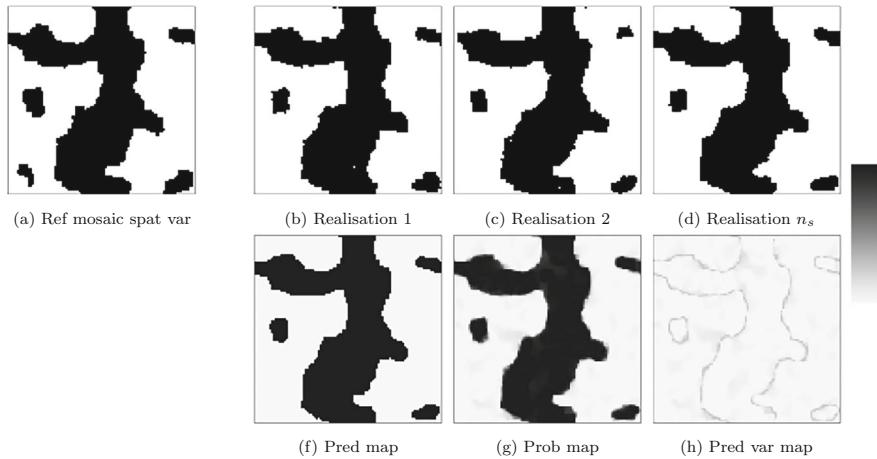
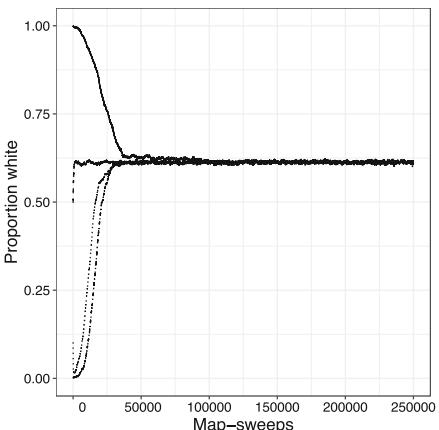


Fig. 7.6 Example Markov RF model. Markov posterior model: (a) reference mosaic spatial variable; (b), (c), (d) three independent realisations; (f) prediction map; (g) probability map; and (h) prediction variance map

Fig. 7.7 Example Markov RF model. Convergence indicator plot of the proportion of the white label based on various initial proportions



to be approximately 10,000 map sweeps. The three realisations are selected from the MCMC sequence to be approximately independent. These realisations appear as possible outcomes of the reference mosaic spatial variable based on the probabilistic model and given the observations. The MMAP prediction map, inferred from 200 approximately independent realisations, is presented in display (f). This prediction appears smoother than the reference, and this effect may be observed along the white/black boundary. The black-probability map in display (g) represents the prediction uncertainty and the prediction variance map in display (h) is inferred from 200 approximately independent realisations. The classification uncertainty is the largest along the white/black boundary.

Chapter 8

Model Parameter Inference



Model parameters are a part of our theoretically constructed model, and they can only be asymptotically determined as we collect infinitely many observations. These parameters must, however, be quantified to have a useful model. The best we can do is to estimate them based on the available information and quantify the associated uncertainty. This parameter uncertainty should preferably be accounted for in the final prediction uncertainty.

Consider the spatial variable $\{s(\mathbf{x}); \mathbf{x} \in D\}$ of interest, spatially discretised to $\{s(\mathbf{x}); \mathbf{x} \in L \subset D\}$ and represented by the n -vector \mathbf{s} . Assume that an m -vector \mathbf{d} of related observations is available. The likelihood model $p(\mathbf{d} | \mathbf{s}; \boldsymbol{\theta}_l)$, which depends on a set of model parameters $\boldsymbol{\theta}_l$, is defined in Chap. 5. Furthermore, the prior model $p(\mathbf{s}; \boldsymbol{\theta}_p)$ with model parameter $\boldsymbol{\theta}_p$ is assigned in Chap. 6. The associated posterior model is expressed as

$$\begin{aligned} p(\mathbf{s} | \mathbf{d}; \boldsymbol{\theta}_l, \boldsymbol{\theta}_p) &= \left[\int p(\mathbf{d} | \mathbf{s}; \boldsymbol{\theta}_l) p(\mathbf{s}; \boldsymbol{\theta}_p) d\mathbf{s} \right]^{-1} \times p(\mathbf{d} | \mathbf{s}; \boldsymbol{\theta}_l) p(\mathbf{s}; \boldsymbol{\theta}_p) \quad (8.1) \\ &= [p(\mathbf{d}; \boldsymbol{\theta}_l, \boldsymbol{\theta}_p)]^{-1} \times p(\mathbf{d} | \mathbf{s}; \boldsymbol{\theta}_l) p(\mathbf{s}; \boldsymbol{\theta}_p). \end{aligned}$$

Assume that the set of likelihood model parameters $\boldsymbol{\theta}_l$ can be assessed from the observation acquisition procedure and hence is known. The prior model, however, is assigned based on previous experience, and the set of model parameters $\boldsymbol{\theta}_p$ is more challenging to assess. Empirical Bayesian elicitation of $\boldsymbol{\theta}_p$ is generally recommended, as discussed in Efron and Morris (1973). This method uses observations of similar spatial variables from other locations and previous experience to determine the prior model parameter $\boldsymbol{\theta}_p$.

Alternatively, the prior model parameter $\boldsymbol{\theta}_p$ may be assessed from the set of available observations by using a maximum marginal likelihood (MML) criterion,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_p &= \operatorname{argmax}_{\boldsymbol{\theta}_p} \{ p(\mathbf{d}; \boldsymbol{\theta}_l, \boldsymbol{\theta}_p) \} \\ &= \operatorname{argmax}_{\boldsymbol{\theta}_p} \left\{ \log \int p(\mathbf{d} | \mathbf{s}; \boldsymbol{\theta}_l) p(\mathbf{s}; \boldsymbol{\theta}_p) d\mathbf{s} \right\}.\end{aligned}\quad (8.2)$$

Thus, $\hat{\boldsymbol{\theta}}_p$ is the set of parameter values that make it most likely to observe exactly \mathbf{d} . Observe that the marginal likelihood corresponds to the inverse of the normalising constant in the posterior model. Calculating $p(\mathbf{d}; \boldsymbol{\theta}_l, \boldsymbol{\theta}_p)$ for a given value of $\boldsymbol{\theta}_p$ requires integration over the n -dimensional \mathbf{s} . This calculation is computationally demanding and unfeasible for most spatial RF models.

For Bayesian spatial modelling in a conjugate setting, calculating $p(\mathbf{d}; \boldsymbol{\theta}_l, \boldsymbol{\theta}_p)$ is typically less demanding. However, the maximisation with respect to $\boldsymbol{\theta}_p$ typically requires numerical optimisation techniques. The MML estimators for $\boldsymbol{\theta}_p$ are discussed for Gaussian, Poisson and Markov RF models with appropriate likelihood models.

The hierarchical Gaussian RF and hierarchical Poisson RF models presented in Chap. 6 are also conjugate with respect to the model parameters for certain likelihood model classes. Consequently, the posterior models $p(\boldsymbol{\theta}_p | \mathbf{d})$ for the model parameters are analytically tractable and can be used for model parameter inference.

8.1 Gaussian RF Models

The model parameters for the stationary Gaussian RF prior are the expectation level $\mu_r \in \mathbb{R}$, variance level $\sigma_r^2 \in \mathbb{R}_+$ and correlation function parameter set $\boldsymbol{\eta}_r \in \mathbb{R}^{q_p}$, assuming that a parametrised correlation function $\rho_r(\boldsymbol{\tau}; \boldsymbol{\eta}_r)$ is assigned. Hence, $\boldsymbol{\theta}_{pG} = [\mu_r, \sigma_r^2, \boldsymbol{\eta}_r]$ can be determined by MML estimation as

$$\hat{\boldsymbol{\theta}}_{pG} = \operatorname{argmax}_{\boldsymbol{\theta}_p} \{ \log p(\mathbf{d}; \boldsymbol{\theta}_p) \}. \quad (8.3)$$

From Expression (7.2), we obtain

$$[\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\boldsymbol{\eta}}_r] = \operatorname{argmax}_{[\mu_r, \sigma_r^2, \boldsymbol{\eta}_r]} \left\{ \log \phi_m(\mathbf{d}; \mathbf{H}\mu_r \mathbf{i}_n, \sigma_r^2 \mathbf{H}\boldsymbol{\Sigma}_r^{\rho(\boldsymbol{\eta}_r)} \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}) \right\}.$$

Therefore,

$$[\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\eta}_r] = \operatorname{argmax}_{[\mu_r, \sigma_r^2, \eta_r]} \left\{ -\frac{m}{2} \log 2\pi - \frac{1}{2} \log |\sigma_r^2 \mathbf{H} \Sigma_r^{\rho(\eta_r)} \mathbf{H}^T + \Sigma_{d|r}| \right. \\ \left. - \frac{1}{2} [\mathbf{d} - \mathbf{H} \mu_r \mathbf{i}_n]^T [\sigma_r^2 \mathbf{H} \Sigma_r^{\rho(\eta_r)} \mathbf{H}^T + \Sigma_{d|r}]^{-1} [\mathbf{d} - \mathbf{H} \mu_r \mathbf{i}_n] \right\}. \quad (8.4)$$

This objective function can be differentiated with respect to μ_r and σ_r^2 , whereas the dependence on η_r is more complicated. The optimisation with respect to η_r may be performed numerically, as described in Algorithm 13. However, this numerical optimisation involves a computationally demanding Cholesky decomposition of an $(m \times m)$ matrix to compute the matrix determinant and the matrix inverse. Because η_r typically is low-dimensional, the optimisation is often feasible.

Algorithm 13: MML Gaussian RF prior parameters

```

Initialise  $\kappa \in \mathbb{R}_+$  and  $\eta_r^0 \in \mathbb{R}^{q_p}$ 
while  $b = 1, 2, \dots$  do
    Calculate
     $\mu_r^b = [\mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_r^{\rho(\eta_r^{b-1})} \mathbf{H}^T + \kappa \Sigma_{d|r}]^{-1} \mathbf{H} \mathbf{i}_n]^{-1} \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_r^{\rho(\eta_r^{b-1})} \mathbf{H}^T + \kappa \Sigma_{d|r}]^{-1}$ 
    Calculate  $\sigma_r^{2b} = m^{-1} [\mathbf{d} - \mathbf{H} \mu_r^b \mathbf{i}_n]^T [\mathbf{H} \Sigma_r^{\rho(\eta_r^{b-1})} \mathbf{H}^T + \kappa \Sigma_{d|r}]^{-1} [\mathbf{d} - \mathbf{H} \mu_r^b \mathbf{i}_n]$ 
    Set  $\kappa = [\sigma_r^{2b}]^{-1}$ 
    Optimise  $\eta_r^b = \operatorname{arg min}_{\eta_r} \left\{ \log |\sigma_r^{2b} \mathbf{H} \Sigma_r^{\rho(\eta_r)} \mathbf{H}^T + \Sigma_{d|r}| \right. \\ \left. + [\mathbf{d} - \mathbf{H} \mu_r^b \mathbf{i}_n]^T [\sigma_r^{2b} \mathbf{H} \Sigma_r^{\rho(\eta_r)} \mathbf{H}^T + \Sigma_{d|r}]^{-1} [\mathbf{d} - \mathbf{H} \mu_r^b \mathbf{i}_n] \right\}$ 
    if  $[\Delta \mu_r^b, \Delta \sigma_r^{2b}, \Delta \eta_r^b] < \epsilon$  then
        break
    end
end
Result:  $\hat{\mu}_r = \mu_r^b, \hat{\sigma}_r^2 = \sigma_r^{2b}$  and  $\hat{\eta}_r = \eta_r^b$ 

```

Confidence intervals for model parameters are typically assessed by bootstrapping in traditional statistics. This assessment is more challenging in spatial statistics because the observations in the m -vector \mathbf{d} are interdependent and, therefore, not exchangeable. The calculation of the Hessian matrix and the use of this matrix for assessing the confidence intervals by the delta approach is still possible, as outlined in Casella and Berger (2002).

A Dirac-linear likelihood model without observation error, denoted $p(\mathbf{d}_e \mid \mathbf{r})$, has an MML estimator that is based on the posterior model $p(\mathbf{r} \mid \mathbf{d}_e)$ given in Expression (7.10). Because μ_r and σ_r^2 are separable in the posterior model, $\hat{\mu}_r$ and $\hat{\sigma}_r^2$ can be assessed sequentially by closed-form estimators, given the spatial correlation function parameters. Consequently, the parameter inference can

be considerably simplified. This can be observed by letting $\text{Tr}\{\boldsymbol{\Sigma}_{d|r}\} \rightarrow 0$ in Algorithm 13.

Explanatory spatial variables may be available, which can be included into the Gaussian prior model, as discussed in Sect. 6.5. The explanatory variables are stacked into the $((n_g + 1) \times n)$ matrix $\mathbf{G}^+ = [\mathbf{i}_n, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{n_g}]^T$. The Gaussian regression prior model is then defined by the stacked regression coefficients in the $(n_g + 1)$ -vector $\boldsymbol{\beta}_r^+ = (\mu_r^0, \beta_r^1, \dots, \beta_r^{n_g})^T$. The inference of the model parameter vector $\boldsymbol{\theta}_{pG} = (\boldsymbol{\beta}_r^+, \sigma_r^2, \boldsymbol{\eta}_r)$ is obtained as

$$[\hat{\boldsymbol{\beta}}_r^+, \hat{\sigma}_r^2, \hat{\boldsymbol{\eta}}_r] = \underset{[\boldsymbol{\beta}_r^+, \sigma_r^2, \boldsymbol{\eta}_r]}{\text{argmax}} \left\{ \log \phi_m(\mathbf{d}; \mathbf{H}\mathbf{G}^{+T} \boldsymbol{\beta}_r^+, \sigma_r^2 \mathbf{H}\boldsymbol{\Sigma}_r^{\rho(\eta_r)} \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}) \right\}, \quad (8.5)$$

which can be assessed by Algorithm 14. For a Dirac-linear likelihood model, the estimators of $\boldsymbol{\beta}_r^+$ and σ_r^2 can be calculated sequentially because the former is independent of the latter, given the value on $\boldsymbol{\eta}_r$.

Algorithm 14: MML Gaussian RF prior parameters with explanatory variables

Initiate $\kappa \in \mathbb{R}_+$ and $\boldsymbol{\eta}_r^0 \in \mathbb{R}^{q_p}$

while $b = 1, 2, \dots$ **do**

 Set

$$\mathbf{H}_{\boldsymbol{\eta}_r^{b-1}} = \left[\mathbf{H}\boldsymbol{\Sigma}_r^{\rho(\eta_r^{b-1})} \mathbf{H}^T + \kappa \boldsymbol{\Sigma}_{d|r} \right]^{-1}$$

 Calculate

$$\boldsymbol{\beta}_r^{+b} = \left[\mathbf{G}^+ \mathbf{H}^T \mathbf{H}_{\boldsymbol{\eta}_r^{b-1}} \mathbf{H}\mathbf{G}^{+T} \right]^{-1} \mathbf{G}^+ \mathbf{H}^T \mathbf{H}_{\boldsymbol{\eta}_r^{b-1}} \mathbf{d}$$

 Calculate

$$\sigma_r^{2b} = m^{-1} [\mathbf{d} - \mathbf{H}\mathbf{G}^{+T} \boldsymbol{\beta}_r^{+b}]^T \mathbf{H}_{\boldsymbol{\eta}_r^{b-1}} [\mathbf{d} - \mathbf{H}\mathbf{G}^{+T} \boldsymbol{\beta}_r^{+b}]$$

 Set $\kappa = [\sigma_r^{2b}]^{-1}$

 Optimise

$$\begin{aligned} \boldsymbol{\eta}_r^b &= \arg \min_{\boldsymbol{\eta}_r} \left\{ \log |\sigma_r^{2b} \mathbf{H}\boldsymbol{\Sigma}_r^{\rho(\eta_r)} \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}| \right. \\ &\quad \left. + [\mathbf{d} - \mathbf{H}\mathbf{G}^{+T} \boldsymbol{\beta}_r^{+b}]^T [\sigma_r^{2b} \mathbf{H}\boldsymbol{\Sigma}_r^{\rho(\eta_r)} \mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1} [\mathbf{d} - \mathbf{H}\mathbf{G}^{+T} \boldsymbol{\beta}_r^{+b}] \right\} \end{aligned}$$

if $[\Delta \boldsymbol{\beta}_r^{+b}, \Delta \sigma_r^{2b}, \Delta \boldsymbol{\eta}_r^b] < \epsilon$ **then**
 break

end

end

Result: $\hat{\boldsymbol{\beta}}_r^+ = \boldsymbol{\beta}_r^{+b}, \hat{\sigma}_r^2 = \sigma_r^{2b}$ and $\hat{\boldsymbol{\eta}}_r = \boldsymbol{\eta}_r^b$

In the hierarchical Gaussian prior model, as defined in Sect. 6.3, the model parameters are assigned prior models $p(\mu_r \mid \sigma_r^2)$ and $p(\sigma_r^2)$, which are Gaussian and inverse-gamma pdfs, respectively. The latter is the univariate special case of the multivariate inverse-Wishart pdf. Recall that the spatial correlation function $\rho_r(\tau)$ is considered known. For Dirac-linear likelihood models without observation errors, as defined in Sect. 5.1, these prior models are conjugate priors, and hence the corresponding posterior models are also Gaussian and inverse-gamma, as demonstrated in Sect. 10.9. The posterior pdf for the expectation μ_r , given the observations \mathbf{d}_e and variance parameter σ_r^2 , is

$$p(\mu_r \mid \mathbf{d}_e, \sigma_r^2) = \phi_1(\mu_r; \mu_{\mu|d, \sigma^2}, \sigma_{\mu|d, \sigma^2}^2)$$

with

$$\begin{aligned}\mu_{\mu|d, \sigma^2} &= \mu_\mu + \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\gamma_\mu \mathbf{H} \mathbf{i}_n \mathbf{i}_n^T \mathbf{H}^T + \mathbf{H} \Sigma_{r|\mu}^\rho \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ \sigma_{\mu|d, \sigma^2}^2 &= \sigma_r^2 \gamma_\mu [1 - \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\gamma_\mu \mathbf{H} \mathbf{i}_n \mathbf{i}_n^T \mathbf{H}^T + \mathbf{H} \Sigma_{r|\mu}^\rho \mathbf{H}^T]^{-1} \mathbf{H} \mathbf{i}_n].\end{aligned}$$

The posterior pdf for the variance σ_r^2 is

$$\begin{aligned}p(\sigma_r^2 \mid \mathbf{d}_e) &= \left[\Gamma \left(\frac{\nu_{\sigma^2|d}}{2} \right) \right]^{-1} \left[\frac{\nu_{\sigma^2|d} \xi_{\sigma^2|d}^2}{2} \right]^{\nu_{\sigma^2|d}/2} [\sigma_r^2]^{-(\nu_{\sigma^2|d}+2)/2} \\ &\quad \times \exp \left(-\frac{\nu_{\sigma^2|d} \xi_{\sigma^2|d}^2}{2\sigma_r^2} \right)\end{aligned}$$

with

$$\begin{aligned}\nu_{\sigma^2|d} &= \nu_{\sigma^2} + m \\ \xi_{\sigma^2|d}^2 &= \frac{\nu_{\sigma^2} \xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)}{\nu_{\sigma^2} + m} \\ \sigma(\mathbf{d}_e) &= (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n)^T [\gamma_\mu \mathbf{H} \mathbf{i}_n \mathbf{i}_n^T \mathbf{H}^T + \mathbf{H} \Sigma_{r|\mu}^\rho \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n).\end{aligned}$$

Consequently, a reasonable estimator for the model parameter μ_r is

$$\begin{aligned}\hat{\mu}_r &= E\{\mu_r \mid \mathbf{d}_e\} = \mu_\mu + \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\gamma_\mu \mathbf{H} \mathbf{i}_n \mathbf{i}_n^T \mathbf{H}^T + \mathbf{H} \Sigma_{r|\mu}^\rho \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ &= [1 - \boldsymbol{\xi}^T \mathbf{H} \mathbf{i}_n] \mu_\mu + \boldsymbol{\xi}^T \mathbf{d}_e\end{aligned}\tag{8.6}$$

with the m -vector $\boldsymbol{\xi} = [\gamma_\mu \mathbf{i}_n^T \mathbf{H}^T (\gamma_\mu \mathbf{H} \mathbf{i}_n \mathbf{i}_n^T \mathbf{H}^T + \mathbf{H} \Sigma_{r|\mu}^\rho \mathbf{H}^T)^{-1}]^T$ containing the weights. Observe that the estimator for the expectation is independent of σ_r^2 , as

is usual for Gauss-linear/Gaussian models. If the grid is extended by extension asymptotic analysis and the number of observations m increases, the term $\xi^T \mathbf{H} \mathbf{i}_n$ tends towards 1.0. Consequently, all weight is assigned to the observations \mathbf{d}_e , as in the traditional ML estimator for μ_r .

A reasonable estimator for the model parameter σ_r^2 is

$$\hat{\sigma}_r^2 = E\{\sigma_r^2 | \mathbf{d}_e\} = \frac{v_{\sigma^2|d}\xi_{\sigma^2|d}^2}{v_{\sigma^2|d} - 2} = \frac{v_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)}{v_{\sigma^2} + m - 2}. \quad (8.7)$$

Let the grid be extended by extension asymptotic analysis, and let the number of observations m increase accordingly. Because $\sigma(\mathbf{d}_e)$ is approximately proportional to m , their ratio dominates the variance estimator. Thus, in the extension limit, the estimator coincides with the traditional ML estimator for σ_r^2 . Bayesian $(1 - \alpha)$ -credibility regions for the model parameters can be computed from the joint posterior model $p(\mu_r, \sigma_r^2 | \mathbf{d}_e) = p(\mu_r | \mathbf{d}_e, \sigma_r^2)p(\sigma_r^2 | \mathbf{d}_e)$. These credibility regions are the regions within the parameter space of the bivariate pdf that integrate to $(1 - \alpha)$.

The expectation function can be extended to include regression on explanatory variables. Posterior pdfs for the associated regression coefficients can be developed along the same lines.

8.1.1 Example: Gaussian RF

The example discussed is introduced in Sect. 5.1.1 and Fig. 5.1. The reference spatial variable represented by the n -vector \mathbf{r} is given in display (b). The available observations are presented in display (c), and the Gauss-linear likelihood model $p(\mathbf{d} | \mathbf{r}; \boldsymbol{\theta}_l)$ is specified with model parameter $\boldsymbol{\theta}_l$, which are assumed to be known. In Sect. 6.2.5, a Gaussian RF prior model is assigned. Hence, the prior model $p(\mathbf{r}; \boldsymbol{\theta}_p)$ is Gaussian with model parameter $\boldsymbol{\theta}_p = [\mu_r, \sigma_r^2, \eta_r]$.

Consider the prior model parameter $\boldsymbol{\theta}_p = [\mu_r, \sigma_r^2, \eta_r]$ to be unknown and estimate it based on an MML criterion. The resulting estimate is $\hat{\boldsymbol{\theta}}_p = [\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\eta}_r] = [4.90, 1.25, 1.38]$. The unknown spatial variable is defined as a realisation of a Gaussian RF model with model parameter $\boldsymbol{\theta}_p = [5.0, 1.0, 1.8]$. Therefore, based on seven observations only, the estimate $\hat{\boldsymbol{\theta}}_p$ is fairly close to the true values.

Figure 8.1 contains the reference spatial variable in display (a), the observation design in display (b) and the prediction with associated prediction variance based on the model parameter value $\hat{\boldsymbol{\theta}}_p$, instead of the true value $\boldsymbol{\theta}_p$, in displays (c) and (d). The latter two displays can be compared to similar displays in Fig. 7.2, and the differences are minor.

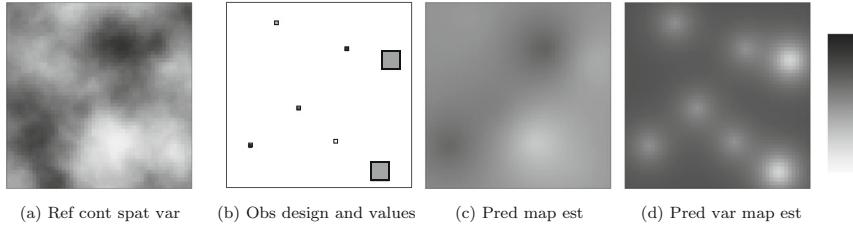


Fig. 8.1 Example Gaussian RF model. Maximum marginal likelihood model parameter inference: (a) reference continuous spatial variable, (b) observation design and values, (c) prediction map based on estimated parameters and (d) prediction variance map based on estimated parameters

8.2 Poisson RF Models

The model parameter for the stationary Poisson RF prior is the intensity $\lambda_k = \lambda_k^n / \Delta_n$. The intensity λ_k^n corresponds to the expected event count in a grid unit in the discretised field represented by the n -vector \mathbf{k} . Hence, the model parameter for \mathbf{k} is $\boldsymbol{\theta}_{pP} = [\lambda_k^n, \emptyset]$. This parameter may be assessed by MML estimation as

$$\hat{\boldsymbol{\theta}}_{pP} = \operatorname{argmax}_{\boldsymbol{\theta}_p} \{ \log p(\mathbf{d}; \boldsymbol{\theta}_p) \}. \quad (8.8)$$

From Expression (7.14), using the factorial form of the likelihood, and because each term is a sum over $k_i \in \mathbb{N}_\oplus$ with $k_i \geq d_i$, the marginal likelihood expression is

$$\begin{aligned} p(\mathbf{d}; \lambda_k^n) &= \sum_{\mathbf{k} \geq \mathbf{d}} p(\mathbf{k}, \mathbf{d}; \lambda_k^n) = \prod_{i=1}^n \sum_{k_i \geq d_i} p(k_i, d_i; \lambda_k^n) \\ &= \prod_{i=1}^n \frac{[\alpha_i \lambda_k^n]^{d_i}}{d_i!} \exp(-\alpha_i \lambda_k^n). \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{\lambda}_k^n &= \arg \max_{\lambda_k^n} \left\{ \sum_{i=1}^n [-\log d_i! + d_i (\log \alpha_i + \log \lambda_k^n) - \alpha_i \lambda_k^n] \right\} \\ &= \frac{n_d}{\boldsymbol{\alpha}^T \mathbf{i}_n}, \end{aligned} \quad (8.9)$$

where n_d is the number of event observations in the set \mathbb{D} and $\boldsymbol{\alpha}$ is the n -vector of correct classification probabilities. The corresponding estimated intensity in the stationary Poisson RF prior is $\hat{\lambda}_k = \hat{\lambda}_k^n / \Delta_n$.

Based on infill asymptotic analysis, $\hat{\lambda}_k = \hat{\lambda}_k^n / \Delta_n \xrightarrow{\text{infill}} n_d / \alpha_D$ with $\alpha_D = \int_D \alpha(\mathbf{x}) d\mathbf{x}$. Observe that if all observations are exact, $n_d = k_D$ and $\int_D \alpha(\mathbf{x}) d\mathbf{x} = |D|$, and hence the traditional estimator for intensity λ_k is obtained.

A confidence interval for the model parameter can be obtained by calculating the second-order derivative of the objective function using the delta approach.

Explanatory spatial variables may be available, and these can be included into the Poisson prior model, as discussed in Sect. 6.5. The grid representation of the explanatory variables is $\{g_j(\mathbf{x}); \mathbf{x} \in L\}; j = 1, 2, \dots, n_g$, and they can be represented in all grid nodes by the n_g -vectors $\mathbf{g}_i; i = 1, 2, \dots, n$. The regression marginal likelihood model can then be defined with intensity level λ_k^0 and regression n_g -vector β_k . In the discretised representation, we use $\lambda_k^{0n} = \lambda_k^0 \Delta_n$, and the corresponding model is then

$$p(\mathbf{d}; \lambda_k^{0n}, \beta_k) = \prod_{i=1}^n \frac{[\alpha_i \lambda_k^{0n} \exp(\mathbf{g}_i^T \beta_k)]^{d_i}}{d_i!} \times \exp(-\alpha_i \lambda_k^{0n} \exp(\mathbf{g}_i^T \beta_k)).$$

By using the MML criterion, the vector of model parameters $\theta_{PP} = [\lambda_k^{0n}, \beta_k]$ may be assessed iteratively, as specified in Algorithm 15.

Algorithm 15: MML Poisson RF prior parameters with explanatory variables

```

Initialize  $\lambda_k^{0n0} \in \mathbb{R}_+$  and  $\beta_k^0 \in \mathbb{R}^{n_g}$ 
while  $b = 1, 2, \dots$  do
     $\lambda_k^{0nb} = \left[ \sum_{i=1}^n \alpha_i \exp(\mathbf{g}_i^T \beta_k^{b-1}) \right]^{-1} \times n_d$ 
     $\beta_k^b = \operatorname{argmin}_{\beta_k} \left\{ \sum_{i=1}^n [d_i \mathbf{g}_i^T \beta_k - \alpha_i \lambda_k^{0nb} \exp(\mathbf{g}_i^T \beta_k)] \right\}$ 
    if  $[\Delta \lambda_k^{0nb}, \Delta \beta_k^b] < \epsilon$  then
        break
    end
end
Result:  $\hat{\lambda}_k^{0n} = \lambda_k^{0nb}$  and  $\hat{\beta}_k = \beta_k^b$ 

```

In the hierarchical Poisson prior model, as defined in Sect. 6.6, the model parameter is assigned a gamma prior pdf $p(\lambda_k)$ with hyper-parameters $v_\lambda \in \mathbb{R}_+$ and $\xi_\lambda \in \mathbb{R}_+$. For the locationwise misclassification likelihood model, this prior model is conjugate. Thus, the corresponding posterior model is also a gamma pdf, as demonstrated in Sect. 10.9. The posterior pdf for the model parameter is:

$$p(\lambda_k | \mathbf{d}) = [\Gamma(v_{\lambda|d})]^{-1} [\xi_{\lambda|d}]^{v_{\lambda|d}} \lambda_k^{v_{\lambda|d}} \exp(-\lambda_k \xi_{\lambda|d})$$

with

$$v_{\lambda|d} = v_\lambda + n_d$$

$$\xi_{\lambda|d} = \xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n.$$

Here, n_d is the number of event observations in the set \mathbb{D} , and $\boldsymbol{\alpha}$ is the n -vector of correct classification probabilities. A reasonable estimator for the model parameter λ_k is

$$\hat{\lambda}_k = E\{\lambda_k | \mathbf{d}\} = \frac{v_{\lambda|d}}{\xi_{\lambda|d}} = \frac{v_\lambda + n_d}{\xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n}. \quad (8.10)$$

Assume that the grid extends through extension asymptotic analysis, with a corresponding increase in the number of observations n_d . Then $\boldsymbol{\alpha}^T \mathbf{i}_n$ increases approximately proportional to n_d . Thus, in the extension limit, the estimator converges towards the traditional ML estimator for λ_k .

A Bayesian $(1 - \alpha)$ -credibility region for the model parameter can be determined from the posterior model $p(\lambda_k | \mathbf{d})$ using the $(1 - \alpha)$ -quantiles of the posterior.

8.2.1 Example: Poisson RF

The example discussed is introduced in Sect. 5.2.1 and Fig. 5.2. The reference event location spatial variable is presented in display (a), with the corresponding event count spatial variable represented by the n -vector \mathbf{k} in display (b). The available observations \mathbf{d} are presented in display (g), and the likelihood model $p(\mathbf{d} | \mathbf{k}; \boldsymbol{\theta}_l)$ is specified with known model parameter $\boldsymbol{\theta}_l$. A Poisson RF prior model is assigned in Sect. 6.5.5. Hence, the prior model $p(\mathbf{k}; \boldsymbol{\theta}_p)$ is Poisson distributed with model parameter $\boldsymbol{\theta}_p = [\lambda_k]$.

The prior model parameter $\boldsymbol{\theta}_p = [\lambda_k]$ is unknown and to be estimated based on an MML criterion. The resulting estimate is $\hat{\boldsymbol{\theta}}_p = \hat{\lambda}_k = 26.72$. The reference event spatial variable to be reconstructed is a realisation of a stationary Poisson RF model with intensity parameter $\lambda_k = 30$. Thus, this estimate is fairly accurate based on observations constituting only a subset of the events.

Figure 8.2 contains the reference event location spatial variable in display (a) and the observed events in display (b). The event count prediction and prediction variance maps based on the estimated model parameter value $\hat{\boldsymbol{\theta}}_p$ are presented in displays (c) and (d), respectively. These two displays can be compared to similar displays in Fig. 7.4, and the differences are minor.

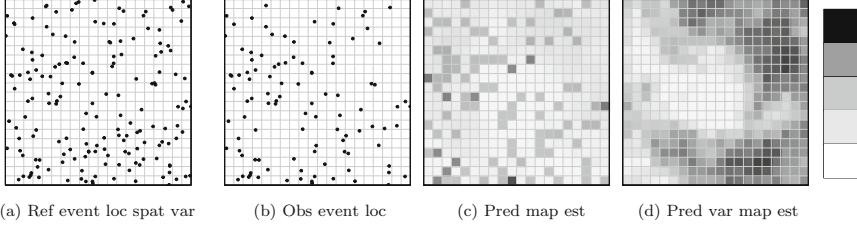


Fig. 8.2 Example Poisson RF model. Maximum marginal likelihood model parameter inference: (a) reference event location spatial variable, (b) observed event locations, (c) prediction map based on estimated parameters and (d) prediction variance map based on estimated parameters

8.3 Markov RF Models

The model parameters in the stationary Markov RF prior are the functions $v_{0l}(\cdot)$ and $v_{1l}(\cdot)$ in the Gibbs formulation, and those in the Markov formulation are $v_{0l}(\cdot)$ and $\omega_l(\cdot|\cdot)$. The function $v_{0l}(\cdot; \theta_{pM0})$ depends on the parameter θ_{pM0} of dimension n_L , being the number of labels. The functions $v_{1l}(\cdot; \theta_{pM1})$ and $\omega_l(\cdot; \theta_{pM1})$, which are dependent on the same parameters, have parametrisations according to the design of the corresponding clique and neighbourhood systems \mathbf{c}_L and \mathbf{n}_L .

The MML estimators are defined by maximising $p(\mathbf{d}; \theta_{pM0}, \theta_{pM1})$ with respect to $[\theta_{pM0}, \theta_{pM1}]$. This marginal likelihood is identical to the inverse of the normalising constant in the posterior model. This normalising constant is typically exceptionally computationally demanding to calculate for given parameter values and even more challenging to optimise with respect to the parameters. Alternatively, one can define the marginal pseudo-likelihood, which can be applied in the inference procedure,

$$\begin{aligned}
\hat{p}(\mathbf{d}; \theta_{pM0}, \theta_{pM1}) &= \prod_{i=1}^n \hat{p}(d_i, d_j; j \in \mathbf{n}_i; \theta_{pM0}, \theta_{pM1}) \\
&= \prod_{i=1}^n \sum_{[l'_i, l'_j; j \in \mathbf{n}_i] \in \mathbb{L}} p(d_i, d_j; j \in \mathbf{n}_i \mid l'_i, l'_j; j \in \mathbf{n}_i) \\
&\quad \times p(l'_i \mid l'_j; j \in \mathbf{n}_i; \theta_{pM0}, \theta_{pM1}) \\
&\quad \times p(l'_j; j \in \mathbf{n}_i; \theta_{pM0}, \theta_{pM1}) \\
&\approx \prod_{i=1}^n \sum_{[l'_i, l'_j; j \in \mathbf{n}_i] \in \mathbb{L}} \prod_{j=i, j \in \mathbf{n}_i} p(d_j \mid l'_j) \times v_{0l}(l'_j; \theta_{pM0}) \\
&\quad \times \omega_l(l'_i | l'_j; j \in \mathbf{n}_i; \theta_{pM1}) \times \text{const.}
\end{aligned}$$

The last term, which is dependent on $(\boldsymbol{\theta}_{pM0}, \boldsymbol{\theta}_{pM1})$, is approximated as a constant. Consequently, the maximum marginal pseudo-likelihood (MMPL) estimator is defined as

$$[\hat{\boldsymbol{\theta}}_{pM0}, \hat{\boldsymbol{\theta}}_{pM1}] = \operatorname{argmax}_{\boldsymbol{\theta}_{pM0}, \boldsymbol{\theta}_{pM1}} \{\log \hat{p}(\mathbf{d}; \boldsymbol{\theta}_{pM0}, \boldsymbol{\theta}_{pM1})\}. \quad (8.11)$$

The complexity of optimising this objective function depends on the actual parametrisation of the Markov prior model.

Consider the special case with the Markov prior model having a two-nearest-node clique design and let $v_{0l}(l_i)$ be identical for all labels. Moreover, let the coupling functions be $v_{1lh}(l_i, l_j; \beta_h) = \beta_h^{I(l_i=l_j)}$ and $v_{1lv}(l_i, l_j; \beta_v) = \beta_v^{I(l_i=l_j)}$ with $\beta_h, \beta_v \in \mathbb{R}_+$. Then the MMPL estimator for the model parameters $[\beta_h, \beta_v]$ can be expressed as

$$\begin{aligned} [\hat{\beta}_h, \hat{\beta}_v] &= \arg \max_{\beta_h, \beta_v} \left\{ \sum_{i=1}^n \log \left[\sum_{[l'_i, l'_j; j \in \mathbf{n}_i] \in \mathbb{L}} \prod_{j=i, j \in \mathbf{n}_i} p(d_j \mid l'_j) \right. \right. \\ &\quad \times \left. \prod_{\substack{h < i, j > \\ j \in \mathbb{L}}} \beta_h^{I(l'_i=l'_j)} \prod_{\substack{v < i, j > \\ j \in \mathbb{L}}} \beta_v^{I(l'_i=l'_j)} \right] \left. \right\}. \end{aligned}$$

This optimisation includes no computationally demanding terms but requires summation over n_L^7 terms for each grid node. These calculations may be unfeasible if the model has many labels and the grid is large. If the observations are exact, a training image is available, and the likelihood model appears as $p(d_i \mid l_i) = I(d_i = l_i)$. Therefore, the MMPL estimator coincides with the traditional MMPL estimator for Markov RF models

$$[\hat{\beta}_h, \hat{\beta}_v] = \arg \max_{\beta_h, \beta_v} \left\{ \sum_{i=1}^n \log \left[\sum_{[l'_i; j \in \mathbf{n}_i] \in \mathbb{L}} \prod_{h < i, j >, j \in \mathbb{L}} \beta_h^{I(d_i=l'_j)} \prod_{v < i, j >, j \in \mathbb{L}} \beta_v^{I(d_i=l'_j)} \right] \right\}.$$

The delta approach can be used to obtain confidence intervals for the model parameters based on the associated Hessian matrix.

8.3.1 Example: Markov RF

The example discussed is introduced in Sect. 5.3.1 and Fig. 5.3. The reference mosaic spatial variable, presented in display (a), is represented by the n -vector \mathbf{l} . The available observations represented by the n -vector \mathbf{d} are given in display (c).

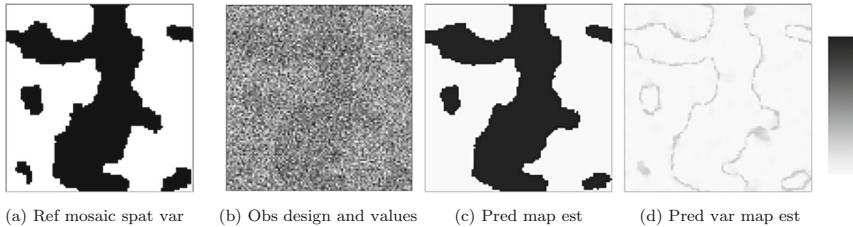


Fig. 8.3 Example Markov RF model. Maximum marginal likelihood model parameter inference: (a) reference mosaic spatial variable, (b) observation design and values, (c) prediction map based on estimated parameters and (d) prediction variance map based on estimated parameters

The associated response likelihood model $p(\mathbf{d} | \mathbf{l}; \boldsymbol{\theta}_l)$ with known model parameter vector $\boldsymbol{\theta}_l$ is specified. In Sect. 6.8.5, a Markov RF prior model $p(\mathbf{l}; \boldsymbol{\theta}_p)$ is assigned with the spatial coupling model parameter $\boldsymbol{\theta}_p = [\beta]$.

Consider the prior model parameter $\boldsymbol{\theta}_p = [\beta]$ to be unknown and estimate it from the observations \mathbf{d} based on a MMPL criterion. The resulting estimate is $\hat{\boldsymbol{\theta}}_p = \hat{\beta} = 1.64$. The reference mosaic spatial variable is manually constructed. Thus, no reference value for β is available. In the prior model assigned above, the value 1.8 is used.

Figure 8.3 contains the reference mosaic spatial variable in display (a), and the available observations are presented in display (b). The prediction and associated prediction variance maps, estimated from 200 approximately independent realisations based on a model using $\hat{\beta}$, are presented in displays (c) and (d). These two latter displays can be compared to similar displays in Fig. 7.6. Observe that the displays are fairly similar, which indicates that the approach is relatively robust with respect to varying values of β in this example.

Chapter 9

Computational Challenges



Advances in technology resulting in more powerful computers may enable us to use more advanced models by simply throwing more computer power at these models. However, methodological and algorithmic advances can provide quantum leaps forward by solving the underlying bottlenecks that fundamentally delay us.

Bayesian spatial modelling discretised on a spatial grid, as defined in this book, is represented by $[\mathbf{s} \mid \mathbf{d}]$. The variable of interest is represented by the n -vector \mathbf{s} , whereas the observations are represented by the m -vector \mathbf{d} . In applications, both the grid size n and the observation dimension m may be large, and hence the computational demands can be significant. Both assessing the posterior model $p(\mathbf{s} \mid \mathbf{d})$ and estimating the model parameter $\boldsymbol{\theta}_p$ in the prior model $p(\mathbf{s}; \boldsymbol{\theta}_p)$ may require substantial computational resources.

Some computational challenges for the Gaussian, Poisson and Markov models are discussed in this section. In Sects. 10.3, 10.4 and 10.5, alternative methodologies for spatial prediction suitable for studies with large observation sets are presented.

9.1 Gaussian RF Models

The computational challenges for the Gaussian model occur primarily in two expressions. Firstly, the assessment of the inverse of the observation covariance matrix $\boldsymbol{\Sigma}_d^{-1} = [\mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1}$, of dimension $(m \times m)$, is computationally challenging. This matrix appears both in the expression for the posterior model, as given in Expression (7.3), and in the MML expression in model parameter inference, as given in Expression (8.4). Secondly, the calculation of the determinant of the same matrix $|\boldsymbol{\Sigma}_d| = |\mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}|$ appearing in the MML expression in the model parameter inference, as given in Expression (8.4), requires significant computational resources. Consequently, the dimension m of the observation vector \mathbf{d} is a critical factor. Inversion of general $(m \times m)$ matrices, performed by Cholesky

decomposition, has computational requirements proportional to m^3 . The number n of nodes in the spatial discretisation is less important for the computational demands since only storage of the variable covariance matrix Σ_r , of dimension $(n \times n)$, is challenging. There is no need for either inversion or determinant evaluation of this matrix.

This section is primarily based on the following model assumptions. The stationary Gaussian prior RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ is parametrised by expectation and variance levels $\mu_r \in \mathbb{R}$ and $\sigma_r^2 \in \mathbb{R}_+$, respectively, and the non-negative definite spatial correlation function $\rho_r(\boldsymbol{\tau}) \in \mathbb{R}_{[-1,1]}; \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$. When the Gaussian prior model is spatially discretised on a grid $\{r(\mathbf{x}); \mathbf{x} \in L \subset D\}$, represented in the n -vector \mathbf{r} , the correlation parameter is the grid value correlation matrix Σ_r^ρ defined by $\rho_r(\boldsymbol{\tau})$.

The m -vector \mathbf{d}_0 contains the observations collected in locations $M : \{\mathbf{x}_i^d \in D; i = 1, 2, \dots, m\}$. The observations are typically acquired according to a Gauss-point likelihood model, i.e. locationwise with additive independent centred Gaussian errors with variance $\sigma_{d|r}^2$. The observation covariance matrix of primary interest is then $\Sigma_d = \sigma_r^2 \Sigma_d^\rho + \sigma_{d|r}^2 \mathbf{I}_m$. The matrix Σ_d^ρ contains the correlations between the observations with locations in the set M and is defined by the spatial correlation function $\rho_r(\boldsymbol{\tau})$. In a spatially discretised model, the Gauss-point likelihood expression is $[\mathbf{d}_0 \mid \mathbf{r}] = \mathbf{H}_o \mathbf{r} + \mathbf{e}_{d|r}$ with binary observation design matrix \mathbf{H}_o , as defined in Sect. 5.1. The corresponding observation covariance matrix is $\Sigma_d = \sigma_r^2 \mathbf{H}_o \Sigma_r^\rho \mathbf{H}_o^T + \sigma_{d|r}^2 \mathbf{I}_m$.

If repeated observations with identical observation designs are collected, often referred to as replicates, the number of observations m can far exceed the number of grid nodes n . In this case, computational efficiency can be improved by pre-processing the observations. For each observation design, the repeated observations that belong to this observation design are combined into one as a weighted linear combination, typically based on a LSE criterion. The associated term in the likelihood model is variance-adjusted accordingly. Consequently, the number of observations is identical to the number of unique observation designs.

All models in this section can also be extended to incorporate spatial explanatory variables in the expectation model, as outlined in Sect. 6.2.3. However, such extensions come at the expense of more complex notation. Additionally, all models discussed here can be utilised for simulation from the posterior pdf using the decomposition specified in Expression (7.4). Recall that this simulation approach only involves one realisation from the Gaussian prior model and one prediction from a spatial predictor.

The evaluation of large spatial observation sets has drawn considerable attention since the introduction of computers in statistics in the 1970s. Two decades later, the increase in the usage of automatic sensors has further increased the importance of designing computationally efficient models and algorithms for large spatial datasets. Survey papers such as Banerjee et al. (2008) and Heaton et al. (2018) offer introductions to recent developments. In presenting the methodology, we use notation consistent with the rest of the book and also provide references to advancements in the traditional geostatistical community.

9.1.1 Sparse Matrix Representations

For studies with a large set of observations, i.e. a large m , inversion of the observation covariance matrix Σ_d , of dimension $(m \times m)$, is computationally unfeasible. One mitigation strategy is to define the matrix to be sparse. The sparse matrix inversion is made by Cholesky decomposition with computer requirements proportional to $m^{3/2}$ for $D \subset \mathbb{R}^2$ and typically m^2 for $D \subset \mathbb{R}^3$, as given in Rue and Held (2005). The Cholesky decomposition of a matrix is used to assess both the inverse and determinant of the matrix. Hence, both prediction and parameter inference are made efficiently. This section contains a discussion of various strategies for sparsification of the observation covariance matrix Σ_d .

Finite-Range Prior Model

A classical reference for geostatistical modelling is Chiles and Delfiner (2012), and a brief introduction to geostatistics is presented in Sect. 10.2. The prior model includes a non-negative definite spatial correlation function $\rho_r(\tau); \tau = \mathbf{x} - \mathbf{x}'$. In traditional geostatistics a sparse observation covariance matrix Σ_d , of dimension $(m \times m)$, is usually obtained by assuming a finite-range Gaussian prior RF model. The spatial correlation function is specified to have a finite range $\tau_0 \in \mathbb{R}_+$ with $\tau_0 < \infty$, which entails that $\rho_r(\tau) = 0.0; |\tau| > \tau_0$. An important class of finite-range spatial correlation functions, as defined in Sect. 6.2 and Gneiting (2002), is constructed from a base correlation function $\rho_B(\tau)$ and a finite-range correlation function $\rho_R(\tau)$. Both functions are non-negative definite correlation functions, and the latter function takes the value of zero for $|\tau| > \tau_0$. Because non-negative definiteness is closed under multiplication, the class is defined as

$$\rho_r(\tau) = \rho_B(\tau) \times \rho_R(\tau) = \begin{cases} 1.0 & |\tau| = 0.0 \\ \rho_B(\tau) \times \rho_R(\tau) & 0.0 < |\tau| < \tau_0 \\ 0.0 & |\tau| \geq \tau_0. \end{cases}$$

Consequently, the corresponding prior covariance matrix $\Sigma_r = \sigma_r^2 \Sigma_r^\rho$, of dimension $(n \times n)$, is sparse, and so is the observation covariance matrix $\Sigma_d = \sigma_r^2 \mathbf{H}_o \Sigma_r^\rho \mathbf{H}_o^T + \sigma_{d|r}^2 \mathbf{I}_m$. The matrix Σ_d can then be efficiently Cholesky decomposed, and the inverse and determinant of the matrix can be calculated.

Finite-range prior models are also useful when combined with a spatially separable correlation function in the Gaussian RF prior model, as defined in Sect. 6.2. Consider a likelihood model of the form $[\mathbf{d} \mid \mathbf{r}] = \mathbf{r} + \mathbf{e}_{d|r}$, which entails a Gauss-point likelihood model with observations in every grid node. The observations are contained in the n -vector \mathbf{d} . The covariance matrix of the observations is given by $\Sigma_d = \sigma_r^2 \Sigma_r^\rho + \sigma_{d|r}^2 \mathbf{I}_m$. Because the prior correlation function is separable, the grid value correlation matrix Σ_r^ρ is expressed as a separable matrix represented by a Kronecker product:

$$\Sigma_r^\rho = \Sigma_r^{\rho y} \otimes \Sigma_r^{\rho z}.$$

Here, $\Sigma_r^{\rho y}$ typically is the grid value correlation matrix, of dimension $(n_y \times n_y)$, for the horizontal component in D_y , and $\Sigma_r^{\rho z}$ is the grid value correlation matrix, of dimension $(n_z \times n_z)$, for the vertical component in D_z .

The Kronecker product of two invertible matrices $\Sigma_y \in \mathbb{R}^{n_y \times n_y}$ and $\Sigma_z \in \mathbb{R}^{n_z \times n_z}$ has the following properties: $[\Sigma_y \otimes \Sigma_z]^{-1} = \Sigma_y^{-1} \otimes \Sigma_z^{-1}$ and $|\Sigma_y \otimes \Sigma_z| = |\Sigma_y|^{n_z} |\Sigma_z|^{n_y}$. Moreover, eigenvalue decomposition of $\Sigma_y \otimes \Sigma_z$ is performed by decomposing the two matrices separately and combining the respective eigenvalues and eigenvectors.

The separate eigenvalue decompositions of the two correlation matrices $\Sigma_r^{\rho y}$ and $\Sigma_r^{\rho z}$ provide the $(n_y \times n_y)$ matrix Λ_y and $(n_z \times n_z)$ matrix Λ_z . The corresponding matrices of eigenvectors are \mathbf{U}_y and \mathbf{U}_z . Eigenvalue decomposition is the most computationally demanding operation because it involves Cholesky decomposition. In the separable model, only two matrices of dimensions $(n_y \times n_y)$ and $(n_z \times n_z)$ need to be eigenvalue decomposed rather than one large matrix of dimension $(n \times n)$ with $n = n_y \times n_z$.

The following expression for the observation covariance matrix is developed:

$$\Sigma_d = \left(\sigma_r^2 \Sigma_r^{\rho y} \otimes \Sigma_r^{\rho z} + \sigma_{d|r}^2 \mathbf{I}_n \right) = (\mathbf{U}_y \otimes \mathbf{U}_z) \left(\sigma_r^2 \Lambda_y \otimes \Lambda_z + \sigma_{d|r}^2 \mathbf{I}_n \right) (\mathbf{U}_y^T \otimes \mathbf{U}_z^T).$$

The calculation of the inverse and determinant of this matrix, which are required in the spatial predictor and the model parameter inference, is made very efficient because the matrix is in eigenvalue form. The matrix separation technique presented above can be extended to cover other more general likelihood models, as discussed in Rakitsch et al. (2013).

Matrix Tapering

An extensive reference for matrix tapering techniques is Stein (2013). The observation covariance matrix Σ_d , of dimension $(m \times m)$, usually has few zero entries but many near zero because distant observations tend to be almost independent. In the numerical analysis community, the tapering concept is frequently used to provide sparse matrix approximations:

$$\tilde{\Sigma}_d = \Sigma_d \circ \mathbf{T},$$

where \mathbf{T} is a user-specified non-negative definite tapering matrix, of dimension $(m \times m)$, with compact support and \circ is the Hadamar matrix product entailing elementwise multiplication of the matrices. The resulting approximate matrix $\tilde{\Sigma}_d$, of dimension $(m \times m)$, is non-negative definite by construction and sparse.

The finite-range prior model assumption and the matrix tapering approximation are identical procedures when the tapering matrix is constructed from the finite-range correlation function $\rho_R(\tau)$. The matrix Σ_d is sparse by construction for a finite-range Gaussian prior model, and this assumption is generally favoured over a numerical approximative matrix tapering approach. Various tapering approaches are presented and discussed in Furrer et al. (2006) and Stein (2013), and their use on large observation sets is demonstrated.

Alternatively, one can use a tapering-like blocking approach based on the experience that spatially distant observations tend to be almost independent. One can partition the spatial domain D into n_D sub-domains $\{D_i; i = 1, 2, \dots, n_D\}$ and assume independence between observations located in non-neighbouring sub-domains. Consequently, the corresponding approximate observation covariance matrix $\tilde{\Sigma}_d$ appears sparse. In Varin et al. (2011) and Eidsvik et al. (2014), this partitioning approach is used for prediction and model parameter pseudo-likelihood inference.

Gaussian Markov RF Models

An in-depth reference for the Gaussian Markov RF model is Rue and Held (2005), and a brief introduction is presented in Sect. 10.3. The Gaussian Markov RF model is a spatially discretised Gaussian RF model defined with an alternative parametrisation. The Gaussian Markov prior model, as defined in Rue and Held (2005), is parametrised by $[\mu_r, \Psi_r]$ with expectation level $\mu_r \in \mathbb{R}$ and grid value precision matrix Ψ_r of dimension $(n \times n)$. This matrix is defined as the inverse of the grid value covariance matrix, $\Psi_r = \Sigma_r^{-1} = [\sigma_r^2 \Sigma_r^\rho]^{-1}$. Therefore, Ψ_r is also a non-negative definite matrix.

Consider the spatially discretised observation m -vector $[\mathbf{d}_{oe} \mid \mathbf{r}] = \mathbf{H}_o \mathbf{r} = \mathbf{r}^d$ assuming a Dirac-point likelihood model which entails that the observations are exact without error, as defined in Sect. 5.1. The spatial variable in the unobserved grid nodes is represented by the $(n - m)$ -vector \mathbf{r}_o . The precision matrix Ψ_r is decomposed accordingly into $(n - m, m)$ -elements as

$$\Psi_r = \begin{bmatrix} \Psi_o & \Psi_{od} \\ \Psi_{do} & \Psi_d \end{bmatrix}.$$

The focus is on the posterior model for the $(n - m)$ -vector \mathbf{r}_o given the observation set \mathbf{d}_{oe} . In the Gaussian Markov parametrisation, this entails

$$[\mathbf{r}_o \mid \mathbf{d}_{oe}] \sim p(\mathbf{r}_o \mid \mathbf{d}_{oe}) = \phi_{n-m}(\mathbf{r}_o; \boldsymbol{\mu}_{o|d}, \boldsymbol{\Psi}_{o|d}^{-1}).$$

The model parameters, as given in Rue and Held (2005), are

$$\begin{aligned} \boldsymbol{\mu}_{o|d} &= \mu_r \mathbf{i}_{n-m} - \boldsymbol{\Psi}_o^{-1} \boldsymbol{\Psi}_{od} (\mathbf{d}_{oe} - \mu_r \mathbf{i}_m) \\ \boldsymbol{\Psi}_{o|d} &= \boldsymbol{\Psi}_o. \end{aligned}$$

The Gaussian Markov precision parametrisation is challenging to interpret. Therefore, typically $\boldsymbol{\Sigma}_{o|d} = \boldsymbol{\Psi}_{o|d}^{-1}$ is calculated to obtain the prediction variance. The corresponding spatially discretised predictor is $\hat{\mathbf{r}} = (\boldsymbol{\mu}_{o|d}^T, \mathbf{d}_{oe}^T)^T$. The prediction variances appear as the diagonal entries of $\boldsymbol{\Sigma}_{o|d}$ with zeros in the grid nodes containing observations.

Computational efficiency is obtained by assigning a sparse precision matrix Ψ_r of dimension $(n \times n)$. Consequently, the diagonal sub-matrices are also sparse, and

the sparse matrix Ψ_o used in the predictor can be efficiently Cholesky decomposed. It is, however, challenging to specify matrices that are both sparse and non-negative definite. The observation precision matrix Ψ_d that appears in the parameter inference expressions is also sparse. Thus, its determinant can be assessed efficiently by Cholesky decomposition. Computational efficiency can be improved by defining a prior stationary Gaussian RF model where the precision matrix is Kronecker separable $\Psi_r = \Psi_r^y \otimes \Psi_r^z$. This decomposition provides computational savings, as discussed earlier.

Spatial prediction based on observations from a Dirac-point likelihood model requires a sparse $((n-m) \times (n-m))$ matrix to be inverted in the Gaussian Markov RF parametrisation. Comparably, the same spatial prediction requires an $(m \times m)$ matrix to be inverted in the traditional Gaussian RF parametrisation. If the latter is restricted to finite-range Gaussian RF models, this matrix also appears as sparse. Recall that typically, the grid density is significantly higher than the observation density, which entails $n \gg m$. However, the resulting computational demands depend on the actual computer implementation because the Cholesky decomposition of sparse matrices is technically complicated.

The general posterior model in the Gaussian Markov RF parametrisation, corresponding to Expression (7.3) in the Gaussian parametrisation, can be developed as demonstrated in Lindgren et al. (2011). The prior stationary Gaussian pdf for the vector \mathbf{r} is as previously defined. The general Gauss-linear likelihood model, as defined in Sect. 5.1, is specified as $[\mathbf{d} \mid \mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e}_{d|r}$, with the arbitrary observation design matrix \mathbf{H} and the observation error vector $\mathbf{e}_{d|r}$ being a centred Gaussian with precision matrix $\Psi_{d|r}$. The resulting Gaussian posterior model, in the Gaussian Markov parametrisation, is expressed as

$$[\mathbf{r} \mid \mathbf{d}] \sim p(\mathbf{r} \mid \mathbf{d}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Psi}_{r|d}^{-1})$$

with

$$\begin{aligned} \boldsymbol{\mu}_{r|d} &= \mu_r \mathbf{i}_n + \boldsymbol{\Psi}_{r|d}^{-1} \mathbf{H}^T \boldsymbol{\Psi}_{d|r} (\mathbf{d} - \mu_r \mathbf{H} \mathbf{i}_n) \\ \boldsymbol{\Psi}_{r|d} &= \boldsymbol{\Psi}_r + \mathbf{H}^T \boldsymbol{\Psi}_{d|r} \mathbf{H}. \end{aligned}$$

When the observations are collected according to a Gauss-linear likelihood model, it becomes necessary to invert a matrix of full grid dimension $(n \times n)$. Despite this matrix possibly being sparse, the computational demands for Cholesky decomposition across the entire grid system can be significant. Recall that the observation correlation matrix of dimension $(m \times m)$ must be inverted in the traditional Gaussian RF model parametrisation.

In the book Rue and Held (2005) and the survey paper Lindgren et al. (2022), several case studies involving the Gaussian Markov RF model can be found.

Basis Function Representation

An extensive reference for the basis function model is Cressie and Johannesson (2008), and a brief introduction is presented in Sect. 10.4. The basis function representation provides an approximation of the prior stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D\}$,

$$\left\{ r_f(\mathbf{x}) = \mu_r + \sum_{i=1}^{n_f} a_i f_i(\mathbf{x}) = \mu_r + \mathbf{f}(\mathbf{x})^T \mathbf{a}; \mathbf{x} \in D \right\}.$$

The functional n_f -vector $\{\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_{n_f}(\mathbf{x})]^T; \mathbf{x} \in D\}$ contains the user-specified basis functions. The n_f -vector $\mathbf{a} = (a_1, a_2, \dots, a_{n_f})^T$ containing the random weights is defined to be centred Gaussian with $(n_f \times n_f)$ covariance matrix Σ_a . This model involving the weight vector must be assigned as a part of the prior model. This basis function approach provides a functional representation of the prior stationary Gaussian RF, apparently without spatial grid discretisation. However, the n_f basis functions need to be specified, and they should not remain constant across the reference domain D . Therefore, each basis function $f_i(\mathbf{x})$ requires a separate reference location \mathbf{x}_i^b to be defined, and the collection of reference locations defines an underlying grid design.

The basis function representation of the Gaussian RF has expectation and spatial covariance functions

$$\{\mu_r(\mathbf{x}) = E\{r_f(\mathbf{x})\} = \mu_r; \mathbf{x} \in D\} \quad (9.1)$$

$$\{\sigma_r(\mathbf{x}, \mathbf{x}') = \text{Cov}\{r_f(\mathbf{x}), r_f(\mathbf{x}')\} = \mathbf{f}(\mathbf{x})^T \Sigma_a \mathbf{f}(\mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}.$$

For the Gaussian RF prior model to exhibit stationarity, it is necessary to have $\sigma_r(\mathbf{x}, \mathbf{x}') = \sigma_r(\mathbf{x} - \mathbf{x}')$, entailing $\sigma_r(\mathbf{x}, \mathbf{x}) = \sigma_r^2$. Obtaining stationarity for non-trivial models is difficult, even approximately, unless n_f is very large. A large variety of basis functions are used in spatial modelling. A Fourier basis function is used in Borgman et al. (1984). Non-orthogonal, finite-support basis functions, like splines and wavelets, are used in Cressie et al. (2022). Furthermore, Lindgren et al. (2011) discusses piecewise linear basis functions based on a triangulation of the area of interest.

Consider the observation m -vector \mathbf{d}_o collected by a Gauss-point likelihood model. The likelihood model cast in the basis function representation is

$$[\mathbf{d}_o | \mathbf{a}] = \mu_r \mathbf{i}_m + \mathbf{F}_d \mathbf{a} + \mathbf{e}_{d|r} \sim p(\mathbf{d}_o | \mathbf{a}) = \phi_m(\mathbf{d}_o; \mu_r \mathbf{i}_m + \mathbf{F}_d \mathbf{a}, \sigma_{d|r}^2 \mathbf{I}_m)$$

with basis function matrix $\mathbf{F}_d = [\mathbf{f}(\mathbf{x}_1^d), \mathbf{f}(\mathbf{x}_2^d), \dots, \mathbf{f}(\mathbf{x}_m^d)]^T$ of dimension $(m \times n_f)$. To represent the observations reliably, the basis function representation typically needs many more terms than observations, which entails $n_f \gg m$.

The predictor with associated prediction variances in the Gaussian basis function representation is

$$\begin{aligned}\{\hat{r}_f(\mathbf{x}) = \mu_r + \mathbf{f}^T(\mathbf{x})\boldsymbol{\mu}_{a|d}; \mathbf{x} \in D\} \\ \{\sigma_p^2(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\Sigma}_{a|d}\mathbf{f}(\mathbf{x}); \mathbf{x} \in D\},\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{\mu}_{a|d} &= 0\mathbf{i}_{n_f} + \boldsymbol{\Sigma}_a \mathbf{F}_d^T [\mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m]^{-1} [\mathbf{d}_o - \mu_r \mathbf{i}_m] \\ \boldsymbol{\Sigma}_{a|d} &= \boldsymbol{\Sigma}_a - \boldsymbol{\Sigma}_a \mathbf{F}_d^T [\mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m]^{-1} \mathbf{F}_d \boldsymbol{\Sigma}_a,\end{aligned}$$

because the prior model is $\mathbf{a} \sim p(\mathbf{a}) = \phi_{n_f}(\mathbf{a}; 0\mathbf{i}_{n_f}, \boldsymbol{\Sigma}_a)$. The predictor with associated prediction variances takes on a functional representation without explicit spatial grid discretisation.

To assess the posterior model, the observation covariance matrix $\boldsymbol{\Sigma}_d = \mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m$, of dimension $(m \times m)$, must be inverted, which is computationally demanding for large observation sets. To ensure a sparse $\boldsymbol{\Sigma}_d$, the basis functions in the set $\{\mathbf{f}(\mathbf{x}); \mathbf{x} \in D\}$ must have narrow spatial support and the coefficient covariance matrix $\boldsymbol{\Sigma}_a$ must be sparse. These constraints must, however, be balanced against a realistic spatial structure in the Gaussian RF prior model.

A low-order basis function representation is used in Cressie and Johannesson (2008), demonstrating that the corresponding posterior model is computed by only inverting an $(n_f \times n_f)$ matrix. The associated predictor is termed fixed rank Kriging. However, the low-order representation cannot capture the small-scale variability of the prior model and, therefore, cannot reproduce exact observations. In Nychka et al. (2015), a multi-resolution basis function approach is introduced, which captures the small-scale variability, and hence more reliable approximations of the posterior model are provided. The predictor is termed lattice Kriging. Lastly, Katzfuss (2017) redefines the multi-resolution basis function representation such that user-specified spatial correlation structures are reproduced. The model formulation enables the use of parallel computing algorithms. The survey paper Cressie et al. (2022) discusses recent advances in basis function representations of Gaussian RF with applications.

An alternative basis function approach is introduced in Lindgren et al. (2011). The prior stationary Gaussian model is defined on piecewise linear triangulations, and the associated weights are obtained based on an stochastic partial differential equation (SPDE). This solution is demonstrated to be equivalent to a centred Gaussian RF with a Matérn spatial correlation structure. The precision matrix for the weights is defined to be sparse, and thus efficient numerical algorithms can be used to assess the posterior model. This SPDE approach is also available for spatial prediction based on large observation sets, and recent advances with applications are discussed in Lindgren et al. (2022).

9.1.2 Localisation Approximations

For studies with a very large set of observations, i.e. very large m , even sparsifying the observation covariance matrix Σ_d , of dimension $(m \times m)$, may not be sufficient to make inversion computationally feasible. One mitigation strategy is to define localised predictors that approximate the global ones. For an arbitrary location $\mathbf{x}_o \in D$, only the m_Δ observations located in a user-specified spatial neighbourhood of \mathbf{x}_o are used in the predictor. The localised observation covariance matrix, of dimension $(m_\Delta \times m_\Delta)$, must be inverted, with $m_\Delta \ll m$. This inversion must, however, be made for each grid node in the spatial discretisation. Thus, the computational demand is proportional to the number of grid nodes n but often with a rather large proportionality constant. Recent methodological developments require matrix inversions in the observation locations only. Thus, the computational demand is proportional to the number of observations m .

Localised Kriging Predictors

In traditional geostatistics, the computational challenges in prediction and prediction variance calculation are addressed by defining localised approximations, as discussed in Journel and Huijbregts (1978) and Chiles and Delfiner (2012). The spatial variable of interest $r(\mathbf{x}_o)$ in an arbitrary location $\mathbf{x}_o \in D$ is assessed by

$$\begin{aligned}\hat{r}(\mathbf{x}_o) &= E\{r(\mathbf{x}_o) \mid \mathbf{d}_{\mathbf{x}_o}^\Delta\} \\ \hat{\sigma}_p^2(\mathbf{x}_o) &= \text{Var}\{r(\mathbf{x}_o) \mid \mathbf{d}_{\mathbf{x}_o}^\Delta\}.\end{aligned}$$

Here, the \mathbf{x}_o -neighbourhood observation m_Δ -vector $\mathbf{d}_{\mathbf{x}_o}^\Delta$ contains the m_Δ observation values with the locations nearest to \mathbf{x}_o . The number of neighbours defines the spatial neighbourhood, either as observations within a given areal neighbourhood or as a directionally balanced number of neighbours. This localisation approach is also robust against deviations from the global stationarity assumptions in the model. The localised predictor must be activated in all grid nodes in the spatial discretisation. The localised Kriging predictor reproduces observations located in grid nodes exactly. However, the spatial continuity of the predictor will be distorted by extreme-valued observations being included/excluded from the neighbourhoods, and quantifying this approximation error is complicated. The computational demand is then proportional to the size n of the grid, with the proportionality constant controlled by the size m_Δ of the neighbourhood.

Traditionally, a cross-validation criterion is used in the model parameter inference, and this criterion is localised as

$$\hat{\theta}_{pG} = \operatorname{argmax}_{\theta_p} \left\{ \prod_{i=1}^m p(d_i \mid \mathbf{d}_{\mathbf{x}_i}^\Delta \setminus d_i; \theta_p) \right\}.$$

The pdf $p(\cdot \mid \cdot; \theta_p)$ is Gaussian with cross-validation expectation and variance based on the sub-set $\mathbf{d}_{\mathbf{x}_i^d}^\Delta \setminus d_i$. This approximate inference approach is highly computationally efficient and can be applied to large observation sets.

In Datta et al. (2016), a Vecchia concept is used to define a hierarchical, localised, sequential decomposition of the posterior model, termed the nearest-neighbour Gaussian process model. In Asfaw and Omre (2016), a non-stationary Gaussian model is defined with localised model parameter inference in a hierarchical conjugate setting. Consequently, all inference and spatial predictions are analytically tractable. The corresponding predictor is termed localised/shrinkage Kriging.

Gaussian Kernel Predictors

A recent reference for the kernel predictors is Omre and Spremec (2023), and a brief introduction is presented in Sect. 10.5. Consider an arbitrary location $\mathbf{x}_o \in \mathcal{D}$. Under the Gaussian RF prior model assumptions with a Gauss-point likelihood model, the conditional expectation and variance are

$$\begin{aligned} E\{r(\mathbf{x}_o) \mid \mathbf{d}_o\} &= \mu_{o|d} = \mu_r + \boldsymbol{\sigma}_{od} \boldsymbol{\Sigma}_d^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m) \\ \text{Var}\{r(\mathbf{x}_o) \mid \mathbf{d}_o\} &= \sigma_{o|d}^2 = \sigma_r^2 - \boldsymbol{\sigma}_{od} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\sigma}_{do}. \end{aligned}$$

The prediction value and observation cross-covariance m -vector $\boldsymbol{\sigma}_{od}$ is defined as $[\boldsymbol{\sigma}_{od}]_i = \sigma_r^2 \rho_r(\mathbf{x}_o - \mathbf{x}_i^d); i = 1, 2, \dots, m$. Define the m -vector $\mathbf{w} = \boldsymbol{\Sigma}_d^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m) = (w_1, w_2, \dots, w_m)^T$ containing the random weights which is independent of the prediction location $\mathbf{x}_o \in \mathcal{D}$. Furthermore, $E\{\mathbf{w}\} = 0\mathbf{i}_m$ and $\text{Var}\{\mathbf{w}\} = \boldsymbol{\Sigma}_d^{-1}$.

The spatial kernel predictor for the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in \mathcal{D}\}$ based on the observations in \mathbf{d}_o is defined as

$$\left\{ \hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\sigma}_{xd} \mathbf{w}^d = \mu_r + \sigma_r^2 \sum_{i=1}^m w_i^d v_i(\mathbf{x}); \mathbf{x} \in \mathcal{D} \right\}$$

with the associated prediction variances

$$\left\{ \sigma_p^2(\mathbf{x}) = \sigma_r^2 - \boldsymbol{\sigma}_{xd} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\sigma}_{dx} = \sigma_r^2 \left[1 - \sigma_r^2 \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} v_i(\mathbf{x}) v_j(\mathbf{x}) \right]; \mathbf{x} \in \mathcal{D} \right\}.$$

The m -vector $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}), \dots, v_m(\mathbf{x}))^T$ containing the observation kernel functions is defined by $v_i(\mathbf{x}) = \rho_r(\mathbf{x} - \mathbf{x}_i^d); i = 1, 2, \dots, m$, and the coefficients are $\beta_{ij} = [\boldsymbol{\Sigma}_d^{-1}]_{ij}; i, j = 1, 2, \dots, m$. The dual Kriging predictor discussed in Matheron (1971) and Royer and Vieira (1984) is of a similar form.

The kernel predictor is in a functional representation and appears as a linear combination of the m spatial correlation functions centred at the observation locations. The prediction variance is also in a functional representation, but it is quadratic in the spatial correlation functions. Note that no spatial discretisation is

required. The kernel predictor represents the asymptotic limit of infilling a grid using Kriging. The computational demands for the kernel and Kriging predictors are largely the same, dominated by the inversion of the matrix Σ_d of dimension $(m \times m)$. This matrix is, however, sparse if a finite-range Gaussian RF prior model is used. Moreover, for finite-range prior models, the kernel predictor is local because the predictor in an arbitrary location only depends on the observations located within the finite range of this location, given the predictor weights \mathbf{w} .

For very large observation sets, i.e. very large m , inversion of even sparse $(m \times m)$ matrices is computationally unfeasible. In such cases, approximate predictors based on localisation offer a solution, as discussed in Vigsnes et al. (2017). The localised approximation used for the kernel predictor is designed to replace the troublesome matrix Σ_d^{-1} , of dimension $(m \times m)$, by an approximate matrix Σ_d^{-1*} . Based on this approximate matrix, approximate kernel predictions and prediction variances are obtained from the respective expressions. Moreover, approximate model parameter inference can be made. The approximate matrix Σ_d^{-1*} is obtained as follows. For each observation $i = 1, 2, \dots, m$, construct a localised observation covariance matrix $\Sigma_{d_i^A}$, of dimension $(m_A \times m_A)$, based on observation i and its m_A nearest observations. Invert each of these sub-matrices. The approximate matrix Σ_d^{-1*} , of dimension $(m \times m)$, is constructed by combining these m inverted sub-matrices. This approximate matrix is sparse and symmetric for a finite-range Gaussian RF prior model, but it need not be non-negative definite.

The localised kernel predictor and prediction variance are obtained by inserting the approximate Σ_d^{-1*} into the kernel predictor and predictor variance to obtain $\{\hat{r}^*(\mathbf{x}); \mathbf{x} \in D\}$ and $\{\sigma_p^{2*}(\mathbf{x}); \mathbf{x} \in D\}$, respectively. According to the Gaussian RF prior model, the approximate predictor has correct spatial continuity, but exact observations are not reproduced. The approximation error can be quantified by inspecting the deviation between the exact observations and the corresponding predictions. The approximate prediction variances may fall outside the eligible $[0.0, \sigma_r^2]$ range. Therefore, some post-processing may be required. The computational demands for the localised kernel predictor are proportional to the number of observations m , because one fixed-dimension neighbourhood matrix must be inverted for each observation. Observe in particular that spatial discretisation is avoided; hence, the computation requirements are independent of grid size and largely independent of the dimension of the reference domain D . The prior expectation and variance levels can be estimated based on this approximate model.

The spatial kernel predictor is discussed in Omre and Spremic (2023), which also contains examples.

9.2 Poisson RF Models

The computational challenges for the Poisson RF model with a misclassification likelihood model, as defined in Expressions (7.15) and (7.17), are minor. The poste-

rior model is in factorial form; hence, each grid unit can be assessed independently. Furthermore, model parameter inference is computationally efficient.

On the other hand, the hierarchical Poisson model, as defined in Sect. 6.6, with an accumulated observation likelihood model, as specified in Sect. 5.2, and the hierarchical prior model for the intensity being a log-Gaussian RF is computationally demanding to assess. This model, referred to as the log-Gaussian Cox model, is frequently used, as discussed in Diggle and Ribeiro (2007). The posterior model in the event count representation can be decomposed as

$$[\mathbf{k}_H \mid \mathbf{d}_A] \sim p(\mathbf{k}_H \mid \mathbf{d}_A) = \int_{\mathbb{R}_{+}^n} \sum_{\mathbf{d} \in \mathbb{N}_{+}^n} p(\mathbf{k}_H \mid \mathbf{d}, \boldsymbol{\lambda}) p(\mathbf{d} \mid \boldsymbol{\lambda}, \mathbf{d}_A) p(\boldsymbol{\lambda} \mid \mathbf{d}_A) d\boldsymbol{\lambda}. \quad (9.2)$$

Therefore, based on Expression (7.15), for $\mathbf{k}_H \in \mathbb{N}_{+}^n$, $\mathbf{k}_H \geq \mathbf{d}$,

$$p(\mathbf{k}_H \mid \mathbf{d}, \boldsymbol{\lambda}) = \prod_{i=1}^n \frac{[(1 - \alpha_i)\lambda_i \Delta_n]^{k_{H_i} - d_i}}{(k_{H_i} - d_i)!} \exp(-(1 - \alpha_i)\lambda_i \Delta_n). \quad (9.3)$$

The pdf $p(\mathbf{d} \mid \boldsymbol{\lambda}, \mathbf{d}_A)$ is defined by $p(d_j; j \in L_{A_i} \mid \lambda_j; j \in L_{A_i}, d_{A_i}); i = 1, 2, \dots, m$. Each of these terms is a multinomial distribution with a set of parameters $\{p_j = \text{const} \times \alpha_j \lambda_j \Delta_n; j \in L_{A_i}\}$ and $d_{A_i} = \sum_{j \in L_{A_i}} d_j$. Recall that the sub-grid representation $L_{A_i}; i = 1, 2, \dots, m$ is disjoint and that for $i \in L_{A^c}$ the likelihood is defined as $\alpha_i = 0$ and $d_i = 0$. The assessment of the posterior pdf $p(\boldsymbol{\lambda} \mid \mathbf{d}_A)$ for the intensity n -vector $[\boldsymbol{\lambda} \mid \mathbf{d}_A]$ is computationally demanding, however, and tailored McMC algorithms are typically used, as discussed in Taylor and Diggle (2014). The convergence rates for these algorithms are often unacceptably low for large models with accumulated observations.

For large-scale models, approximations in a Gauss-linear/Gauss framework can be defined. Consider the Bayesian decomposition

$$p(\boldsymbol{\lambda} \mid \mathbf{d}_A) = [p(\mathbf{d}_A)]^{-1} \times p(\mathbf{d}_A \mid \boldsymbol{\lambda}) p(\boldsymbol{\lambda}).$$

Recall that if $p(\mathbf{d}_A \mid \boldsymbol{\lambda})$ is Gauss-linear and $p(\boldsymbol{\lambda})$ is Gaussian, the expression is analytically tractable by the conjugate property because both $p(\boldsymbol{\lambda} \mid \mathbf{d}_A)$ and $p(\mathbf{d}_A)$ are also Gaussian.

The likelihood model for the accumulated observations in the m -vector \mathbf{d}_A can be decomposed as

$$p(\mathbf{d}_A \mid \boldsymbol{\lambda}) = \prod_{i=1}^m p(d_{A_i} \mid \boldsymbol{\lambda}),$$

because the conditional model is a Poisson RF, and the areas $A_j; i = 1, 2, \dots, m$ are by definition disjoint. The definition of the Poisson RF entails further that $[d_{A_i} \mid \boldsymbol{\lambda}]$ is

Poisson distributed with parameter $\lambda_{A_i} \Delta_n = \sum_{j \in L_{A_i}} \alpha_j \lambda_j \Delta_n$. A reasonable Gauss-linear approximation is therefore

$$p(d_{A_i} | \boldsymbol{\lambda}) \approx \phi_1(d; \mu_{[d_A|\lambda]_i}, \sigma_{[d_A|\lambda]_i}^2).$$

The associated parameters are $\mu_{[d_A|\lambda]_i} = \mathbf{h}_i^{\alpha T} \boldsymbol{\lambda}$ and $\sigma_{[d_A|\lambda]_i}^2 = \mathbf{h}_i^{\alpha T} \boldsymbol{\mu}_{\lambda} i_n$. The n -vector \mathbf{h}_i^α has value $\alpha_j \Delta_n$ in entries $j \in L_{A_i}$ and value 0 elsewhere. Moreover, $\boldsymbol{\mu}_{\lambda}$ is the expectation of λ_i in the corresponding stationary prior model $p(\boldsymbol{\lambda})$. In the Gauss-linear model, the variance is independent of the value of the conditioning variable. Thus, the expectation of the intensity is inserted in the variance expression earlier.

Because of the independence of the elements in $[\mathbf{d}_A | \boldsymbol{\lambda}]$, a reasonable Gauss-linear approximation takes the form

$$p(\mathbf{d}_A | \boldsymbol{\lambda}) \approx \phi_m(\mathbf{d}; \mathbf{H}^{\alpha T} \boldsymbol{\lambda}, \boldsymbol{\Sigma}_{d_A|\lambda}^\alpha).$$

The regression matrix is $\mathbf{H}^\alpha = [\mathbf{h}_1^\alpha, \mathbf{h}_2^\alpha, \dots, \mathbf{h}_m^\alpha]$, of dimension $(n \times m)$, and the covariance matrix $\boldsymbol{\Sigma}_{d_A|\lambda}^\alpha$, of dimension $(m \times m)$, is a diagonal matrix with entries $\mathbf{h}_i^{\alpha T} \boldsymbol{\mu}_{\lambda} i_n; i = 1, 2, \dots, m$.

The prior model on the intensity n -vector $\boldsymbol{\lambda}$ is assigned a log-Gaussian RF model and the Gaussian RF model approximates it

$$\begin{aligned} p(\boldsymbol{\lambda}) &= \log \phi_n(\boldsymbol{\lambda}; \beta_\lambda i_n, \gamma_\lambda^2 \boldsymbol{\Sigma}_{\lambda}^\rho) \\ &\approx \phi_n(\boldsymbol{\lambda}; \boldsymbol{\mu}_{\lambda} i_n, \sigma_{\lambda}^2 \boldsymbol{\Sigma}_{\lambda}^\rho). \end{aligned}$$

In this approximation, the natural hyper-parameter values are the expectation and variance of the log-Gaussian pdf, $\boldsymbol{\mu}_{\lambda} = \exp(\beta_\lambda + \gamma_\lambda^2/2)$ and $\sigma_{\lambda}^2 = (\exp(\gamma_\lambda^2) - 1) \times \exp(2\beta_\lambda + \gamma_\lambda^2)$. The log-Gaussian model has support on \mathbb{R}_{+}^n , whereas the Gaussian model has support on \mathbb{R}^n . Therefore, $\boldsymbol{\mu}_{\lambda}$ should be greater than $2\sigma_{\lambda}$ to ensure a reasonable approximation. The hyper-parameters may alternatively be elicited from the observations \mathbf{d}_A by a maximum approximate marginal likelihood approach or by naive estimators, as outlined below.

In the approximate model, $p(\mathbf{d}_A | \boldsymbol{\lambda})$ is Gauss-linear and $p(\boldsymbol{\lambda})$ is Gaussian. Consequently, they constitute a pair of conjugate models, and by following the developments in Sect. 7.1, one obtains

$$p(\boldsymbol{\lambda} | \mathbf{d}_A) \approx \phi_n(\boldsymbol{\lambda}; \boldsymbol{\mu}_{\lambda|d_A}, \boldsymbol{\Sigma}_{\lambda|d_A}) \tag{9.4}$$

with

$$\begin{aligned} \boldsymbol{\mu}_{\lambda|d_A} &= \boldsymbol{\mu}_{\lambda} i_n + \sigma_{\lambda}^2 \boldsymbol{\Sigma}_{\lambda}^\rho \mathbf{H}^\alpha [\sigma_{\lambda}^2 \mathbf{H}^{\alpha T} \boldsymbol{\Sigma}_{\lambda}^\rho \mathbf{H}^\alpha + \boldsymbol{\Sigma}_{d_A|\lambda}]^{-1} [\mathbf{d}_A - \boldsymbol{\mu}_{\lambda} \mathbf{H}^{\alpha T} i_n] \\ \boldsymbol{\Sigma}_{\lambda|d_A} &= \sigma_{\lambda}^2 \boldsymbol{\Sigma}_{\lambda}^\rho - \sigma_{\lambda}^2 \boldsymbol{\Sigma}_{\lambda}^\rho \mathbf{H}^\alpha [\sigma_{\lambda}^2 \mathbf{H}^{\alpha T} \boldsymbol{\Sigma}_{\lambda}^\rho \mathbf{H}^\alpha + \boldsymbol{\Sigma}_{d_A|\lambda}]^{-1} \sigma_{\lambda}^2 \mathbf{H}^{\alpha T} \boldsymbol{\Sigma}_{\lambda}^\rho. \end{aligned}$$

Because the pair of likelihood and prior models is conjugate, the normalising constant can be calculated analytically. Thus, a maximum approximate marginal likelihood criterion may be used to elicit the hyper-parameter $\boldsymbol{\theta}_\lambda = [\mu_\lambda, \sigma_\lambda^2, \rho_\lambda(\cdot)]$, as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_\lambda &= \operatorname{argmax}_{\boldsymbol{\theta}_\lambda} \{p(\mathbf{d}_A; \boldsymbol{\theta}_\lambda)\} \\ &= \operatorname{argmin}_{\boldsymbol{\theta}_\lambda} \left\{ m \log 2\pi + \log |\sigma_\lambda^2 \mathbf{H}^{\alpha T} \boldsymbol{\Sigma}_\lambda^\rho \mathbf{H}^\alpha + \boldsymbol{\Sigma}_{d_A|\lambda}| \right. \\ &\quad \left. + [\mathbf{d}_A - \mu_\lambda \mathbf{H}^{\alpha T} \mathbf{i}_n] [\sigma_\lambda^2 \mathbf{H}^{\alpha T} \boldsymbol{\Sigma}_\lambda^\rho \mathbf{H}^\alpha + \boldsymbol{\Sigma}_{d_A|\lambda}]^{-1} [\mathbf{d}_A - \mu_\lambda \mathbf{H}^{\alpha T} \mathbf{i}_n]^T \right\}.\end{aligned}$$

Observe that this hyper-parameter estimator is not exactly the same as the corresponding estimator in the Gaussian case because the expectation μ_λ is involved in the likelihood model covariance matrix $\boldsymbol{\Sigma}_{d_A|\lambda}$.

The naive estimators may be expressed as

$$\begin{aligned}\hat{\mu}_\lambda &= \frac{1}{m} \sum_{i=1}^m \frac{d_{A_i}}{n_{A_i} \Delta_n} \\ \hat{\sigma}_\lambda^2 &= \frac{1}{m} \sum_{i=1}^m \left(\frac{d_{A_i}}{n_{A_i} \Delta_n} - \hat{\mu}_\lambda \right)^2.\end{aligned}$$

The associated spatial correlation function $\rho_\lambda(\tau)$ must be user-specified based on experience with the spatial smoothness of the intensity.

All factors in the event count posterior model $p(\mathbf{k}_H \mid \mathbf{d}_A)$, as specified in Expression (9.2), are now approximated. Thus, the event count variable $[\mathbf{k}_H \mid \mathbf{d}_A]$ can be simulated by a sequential algorithm. The intensity λ must be non-negative, but the approximative model does not ensure positivity. Hence, a transformation must be utilised to ensure positivity, for instance, an exponential transform, which is inherently suitable for a log-Gaussian Cox model. The corresponding event location variable $[\mathbb{X}_{HD}^n \mid \mathbf{d}_A]$ can be simulated by an algorithm similar to the one defined for Expression (7.16).

An estimate of the event intensity n -vector λ given the observations \mathbf{d}_A is

$$\hat{\lambda} = \hat{E}\{\lambda \mid \mathbf{d}_A\} = \boldsymbol{\mu}_{\lambda|d_A},$$

where \hat{E} indicates the expectation under the approximate model. The associated estimation variances appear as the diagonal elements of $\boldsymbol{\Sigma}_{\lambda|d_A}$. Observe that both these estimates are analytically available in the approximate model for $p(\lambda \mid \mathbf{d}_A)$. Because the conditioning observations \mathbf{d}_A are non-negative, the estimator $\hat{\lambda}$ will most likely also provide non-negative estimates, as expected. The approximate posterior models for $p(\mathbf{k}_H \mid \mathbf{d}_A)$ and $p(\lambda \mid \mathbf{d}_A)$ are both analytically tractable. Therefore, they may define proposal pdfs in either the rejection or iterative simulation algorithms, as defined in Sects. 2.2 and 10.1. These adapted algorithm

proposal pdfs facilitate the efficient generation of realisations from the correct posterior pdf.

First, consider the special case with observation n -vector \mathbf{d} recorded in each grid node with registration probabilities in n -vector $\boldsymbol{\alpha}$. Here, the registration is made with event censoring in each grid unit. Thus, no accumulated observations are made. Recall that missing observations in grid node i can be represented by setting $\alpha_i = 0$ and $d_i = 0$. The posterior model for the event count n -vector is then

$$[\mathbf{k}_H \mid \mathbf{d}] \sim p(\mathbf{k}_H \mid \mathbf{d}) = \int_{\mathbb{R}_{\oplus}^n} p(\mathbf{k}_H \mid \mathbf{d}, \boldsymbol{\lambda}) p(\boldsymbol{\lambda} \mid \mathbf{d}) d\boldsymbol{\lambda}. \quad (9.5)$$

The pdf $p(\mathbf{k}_H \mid \mathbf{d}, \boldsymbol{\lambda})$ is identical to Expression (9.3). The approximate pdf $p(\boldsymbol{\lambda} \mid \mathbf{d})$ is identical to Expression (9.4), where the regression matrix \mathbf{H}^α , of dimension $(n \times n)$, is diagonal with the n -vector $\boldsymbol{\alpha}$ as diagonal entries. Further, the observation covariance matrix $\boldsymbol{\Sigma}_{d|\lambda}$, of dimension $(n \times n)$, is also a diagonal matrix with the n -vector $\mu_\lambda \boldsymbol{\alpha}$ as diagonal elements. The model hyper-parameters can be inferred as previously outlined.

Second, consider the case with observation n -vector \mathbf{d} containing exact recordings of every event. Thus, the registration probabilities are specified by the n -vector $\boldsymbol{\alpha} = \mathbf{i}_n$. The posterior model for the event count n -vector can be expressed as

$$[\mathbf{k}_H \mid \mathbf{d}] \sim p(\mathbf{k}_H \mid \mathbf{d}) = I(\mathbf{k}_H = \mathbf{d}).$$

Observe that the event count n -vector is distributed according to an n -variate discrete Dirac pdf, with parameters identical to the observation n -vector \mathbf{d} , as expected. The approximate posterior model for the event intensity n -vector $p(\boldsymbol{\lambda} \mid \mathbf{d})$ is identical to Expression (9.4), where the regression matrix is $\mathbf{H}^\alpha = \mathbf{I}_n$, of dimension $(n \times n)$, and the observation covariance matrix is $\boldsymbol{\Sigma}_{d|\lambda} = \mu_\lambda \mathbf{I}_n$ of dimension $(n \times n)$. The observations are exact in all grid units in this case. Thus, the approximate posterior variance is identical in all grid units. Model hyper-parameter inference can be performed as previously outlined.

The approximate posterior model for $p(\boldsymbol{\lambda} \mid \mathbf{d})$ can be extended to include explanatory spatial variables, as discussed in Sect. 6.2.3. A double hierarchical model may also be defined considering the hyper-parameters $[\mu_\lambda, \sigma_\lambda^2]$ as random variables with suitable prior models, as defined in Sect. 6.3. Because the approximate model is a Gaussian RF model, both these extensions can be made while maintaining the analytical tractability, as demonstrated in Sect. 7.1. In large-scale studies, especially when the number of observations m is substantial, an additional level of approximation is required to evaluate the approximate posterior model, as detailed in Sect. 9.1.

The approximate approach described above is similar to the frequently used integrated nested Laplace approximation (INLA) approach defined in Rue et al. (2009) and briefly presented in Sect. 10.7. The INLA model is defined as a double hierarchical model with the focus on the assessment of the posterior model for hyper-parameter event intensity, $p(\boldsymbol{\lambda} \mid \mathbf{d})$. In the INLA framework, the model for

$p(\boldsymbol{\lambda} \mid \mathbf{d}, \boldsymbol{\theta}_{hh})$, with hyper-hyper-parameter $\boldsymbol{\theta}_{hh}$, is approximated by a discretised Gaussian RF model. Likewise, a Gaussian-inspired approximation for the posterior model, $p(\boldsymbol{\theta}_{hh} \mid \mathbf{d})$, is defined. Finally, the marginalisation with respect to $\boldsymbol{\theta}_{hh}$ of the joint pdf, $p(\boldsymbol{\lambda}, \boldsymbol{\theta}_{hh} \mid \mathbf{d}) = p(\boldsymbol{\lambda} \mid \mathbf{d}, \boldsymbol{\theta}_{hh})p(\boldsymbol{\theta}_{hh} \mid \mathbf{d})$, provides the posterior pdf of interest $p(\boldsymbol{\lambda} \mid \mathbf{d})$. This marginalisation is performed numerically by an efficient algorithm. Alternatives to INLA for latent Gaussian models are based on automatic differentiation, as discussed in Kristensen et al. (2016), and on Hamiltonian Monte Carlo sampling, as discussed in Carpenter et al. (2017).

9.3 Markov RF Models

Numerous computational challenges exist in assessing the Markov posterior model, as defined in Expression (7.24). An iterative McMC algorithm is typically used to generate realisations, but it is challenging to attain favourable convergence and mixing rates. Moreover, all predictions with associated prediction uncertainties must be simulation-based. Furthermore, model parameter inference typically uses a pseudo-likelihood criterion to avoid summation over the complete sample space. Observe also that many advanced versions of the McMC algorithm are unavailable in the Markov RF context because derivatives are undefined for categorical spatial variables.

The Markov mesh model is closely related to the Markov RF model, as discussed in Cressie and Davidson (1998), although the two are not equivalent. Recognise that, when available, the sequential simulation algorithm, as defined in Sect. 10.1, is the most efficient. Sequential algorithms generate one realisation in a single pass through the grid nodes. Recall that an equivalent Markov RC can be defined for a Markov RP on a one-dimensional grid L , as specified in Sect. 10.8. Realisations from the latter Markov RC can be generated sequentially. For a Markov RF on a higher-dimensional grid L , no unique equivalent Markov RC can be defined because of a lack of unique ordering. The grid nodes can, however, be ordered according to certain rules to define a chain. This chain of nodes is denoted a Markov random mesh. Consequently, the Markov RF model can be transformed into a Markov mesh model. In principle, the Markov random mesh model can be simulated sequentially, and an ML criterion can perform model parameter inference. One challenge in this context is that spatially convex neighbourhood designs in the Markov RF model will not appear as such in the Markov random mesh model. As a consequence, the computational efficiency may be significantly reduced. To mitigate this reduction in efficiency, intelligent approximate parametrisations, as outlined in Luo and Tjelmeland (2019), can be employed.

Chapter 10

Special Topics



You have a problem to solve. Identifying a suitable notation is half the battle to finding a solution. Casting alternative methodologies in a comparable notation is half the battle to selecting a suitable methodology.

A short introduction to various methods and algorithms frequently used in spatial modelling is included. The topics are presented in the book's notation to aid comprehension and facilitate comparison with the methodology in the book's main body. Additional references are provided for further reading.

10.1 Classes of Simulation Algorithms

Consider a discretised RF $\{s(\mathbf{x}); \mathbf{x} \in L \in D \subset \mathbb{R}^3\}$ on a grid L of size n , represented by the n -vector $\mathbf{s} = (s_1, s_2, \dots, s_n) \in \Omega_s^n$ with pdf $p(\mathbf{s})$. The dimension n can be very large for spatial variables, often in the 10^9 – 10^{12} range. The challenge is to generate one realisation $s^* \sim p(\mathbf{s})$, and the procedure for obtaining this realisation is referred to as simulation. A more comprehensive discussion of simulation techniques can be found in Gamerman (2006).

The class of univariate inverse simulation algorithms is defined in Algorithm 16. This algorithm provides one realisation of the univariate $s \in \Omega_s$ with associated pdf $p(s)$ and cdf $P(s)$.

Algorithm 16: Univariate inverse algorithm

Generate $u \sim \text{unif}[0, 1]$

$$s^* = P^{-1}(u)$$

Result: $s^* \sim p(s)$

Simulation of multivariate random variables $\mathbf{s} \sim p(\mathbf{s})$ is more complicated because the inter-dependence characteristics must be reproduced. This section introduces five classes of multivariate simulation algorithms: transformation, sequential, rejection, iterative and approximate algorithms.

10.1.1 Transformation Algorithms

The class of transformation algorithms is defined in Algorithm 17. Given that a one-to-one transformation of the n -variate random variable of interest into n independent random variables with known pdfs exists and can be identified, the transformation algorithm contains two steps: first, n intermediate univariate realisations are independently generated from their respective pdfs, and second, this set of realisations is back-transformed into the n -variate realisation of interest. The algorithm provides one realisation of \mathbf{s} from the correct pdf $p(\mathbf{s})$.

In practice, the algorithm has limited use because it is challenging to identify an appropriate transformation for general pdfs. Additionally, the transformation, if it exists, may be computationally demanding to evaluate. The transformation algorithm is particularly useful for generating $\mathbf{r} \sim p(\mathbf{r}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ because the Gaussian class is closed under linear transformation. A realisation is typically constructed by the linear transformation $\mathbf{z} = \boldsymbol{\Sigma}_r^{-1/2}[\mathbf{r} - \boldsymbol{\mu}_r]$, where the n -vector \mathbf{z} has independent entries $z_i \sim \phi_1(z_i; 0, 1)$ for $i = 1, 2, \dots, n$.

Algorithm 17: Transformation algorithm

```

Identify a one-to-one transformation  $T(\cdot)$  such that  $\mathbf{s}' = T(\mathbf{s}) \sim p(\mathbf{s}') = \prod_{i=1}^n p(s'_i)$ 
for  $i = 1, 2, \dots, n$  do
    | Generate  $s_i^{ts} \sim p(s_i')$ 
end
Calculate  $\mathbf{s}' = T^{-1}(s_1^{ts}, s_2^{ts}, \dots, s_n^{ts})$ 
Result:  $\mathbf{s}' \sim p(\mathbf{s})$ 

```

10.1.2 Sequential Algorithms

The class of sequential algorithms is defined in Algorithm 18. The pdf of an arbitrary n -vector \mathbf{s} can be decomposed as $p(\mathbf{s}) = p(s_1) \prod_{i=2}^n p(s_i | s_{i-1}, s_{i-2}, \dots, s_1)$. The algorithm sequentially generates realisations of the elements in the n -variate random variable, each conditioned upon the realisations already generated in preceding entries of the vector. Sequential algorithms provide exact realisations of $\mathbf{s} \sim p(\mathbf{s})$. Sequential algorithms may be computationally efficient, given that the conditional pdfs can be identified and that realisations from them can be efficiently generated. However, identifying conditional pdfs from general pdfs is typically a challenging

task. Sequential algorithms tend to be numerically unstable for high-dimensional pdfs.

Sequential algorithms are frequently used in the simulation of Gaussian, Poisson and Markov RF models. They are particularly useful if the model is defined in a sequential manner, as for hierarchical models. Furthermore, RF models with a one-dimensional reference domain $D \subset \mathbb{R}$ can typically be recast as sequential pdfs. The algorithm is also suitable for general RF models if they can be described in sequential forms and the associated conditional pdfs can be calculated.

Algorithm 18: Sequential algorithm

```

Generate  $s_1^s \sim p(s_1)$ 
for  $i = 2, 3, \dots, n$  do
  | Generate  $s_i^s \sim p(s_i \mid s_{1:i-1}^s)$ 
end
Result:  $\mathbf{s}^s = (s_1^s, s_2^s, \dots, s_n^s) \sim p(\mathbf{s})$ 
```

10.1.3 Rejection Algorithms

The class of rejection algorithms is defined in Algorithm 19. This class of algorithms is very flexible and can, in principle, be utilised to simulate a realisation from an arbitrary pdf. Assume that a proposal pdf $p^*(\mathbf{s})$, similar to the target pdf $p(\mathbf{s})$, is available and easy to generate realisations from. A realisation is proposed from $p^*(\mathbf{s})$ and accepted or rejected as a realisation from $p(\mathbf{s})$ based on the value of the acceptance probability, $\alpha = p(\mathbf{s}) / (\nu p^*(\mathbf{s}))$, where ν is a constant that ensures $\alpha \in \mathbb{R}_{[0,1]}$ for all $\mathbf{s} \in \Omega_s$. Rejection algorithms generate an exact realisation of \mathbf{s} from the pdf $p(\mathbf{s})$ and provide a general approach that, in principle, can be used for most pdfs. The simplicity of a rejection algorithm makes it easy to implement, even for complex pdfs.

Algorithm 19: Rejection algorithm

```

Specify the proposal pdf  $p^*(\mathbf{s})$  and constant  $\nu$  such that  $\frac{p(\mathbf{s})}{\nu p^*(\mathbf{s})} \leq 1.0$  for all  $\mathbf{s} \in \Omega_s^n$ 
while  $b = 1, 2, \dots$  do
  | Generate  $\mathbf{s}^p \sim p^*(\mathbf{s})$ 
  | Calculate  $\alpha = \frac{p(\mathbf{s}^p)}{\nu p^*(\mathbf{s}^p)}$ 
  | Generate  $u \sim \text{unif}[0, 1]$ 
  | if  $u \leq \alpha$  then
    |   | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
    |   | break
  | end
end
Result:  $\mathbf{s}^s = \mathbf{s}^b \sim p(\mathbf{s})$ 
```

In practice, identifying the value of v may be challenging. Rejection algorithms are typically very inefficient for high-dimensional variables because the sample space is dominated by low acceptance probabilities α . Heavy-tailed target pdfs, in particular, tend to provide large values of v , leading to low acceptance probabilities.

Rejection algorithms are frequently used for Bayesian inversion. Often the proposal pdf $p^*(\mathbf{s})$ is defined as the prior pdf, and thus v is the product of the normalising constant and the maximum value of the likelihood function. The latter can, in practice, be challenging to identify.

10.1.4 Iterative Algorithms

The class of iterative algorithm is defined in Algorithm 20. This class contains the McMC algorithms and their subclass of Metropolis–Hastings algorithms. Iterative algorithms are mostly used for simulation from pdfs of the form $p(\mathbf{s}) = \text{const} \times h(\mathbf{s})$, where the normalising constant const is computationally unfeasible. The iterative algorithm requires a proposal function $g(\cdot|\cdot)$, possibly dependent on the current state of the variable, to be specified. Moreover, the initial state of the variable \mathbf{s}^0 must be generated such that $p(\mathbf{s}^0) > 0$. At each step b of the iterative algorithm, a proposed realisation \mathbf{s}^p is generated from the proposal function $g(\mathbf{s}|\mathbf{s}^{b-1})$. The proposed realisation is accepted or rejected based on the acceptance probability α . The determination of this acceptance probability is contingent on the ratio between the target pdf and the proposal function, which is computed for both the existing and the proposed states. If the proposed realisation is rejected, the current realisation persists.

A realisation from a McMC algorithm is only asymptotically correct, and the distributional convergence rate is important for the applicability of the algorithm. For an arbitrary initial state, the number of iterations b required for \mathbf{s}^b to be distributed approximately according to the pdf $p(\mathbf{s})$ is denoted as the burn-in time. Formal calculations of the burn-in time for general models are not possible, but many heuristic approaches are used in practice. Heuristic approaches typically depend on a convergence indicator, such as the label proportions if the variable is a mosaic spatial variable, and monitor this indicator throughout the iterations. These convergence indicators should be evaluated when starting the algorithm from various extreme initial states, such as one label only or random spatial mixing of labels, and their corresponding convergence indicators should exhibit similar characteristics after burn-in. The burn-in time is defined as the iteration number at which these traces approximately merge. Since the convergence indicators are random variables, they share characteristics but are not identical after the burn-in period. After convergence, subsequent realisations in the iterative procedure tend to be autocorrelated. The minimum number of iterations between approximately independent realisations \mathbf{s}^b and $\mathbf{s}^{b'}$ is referred to as the mixing time. The mixing time can be determined heuristically by evaluating long iteration sequences of \mathbf{s}^b after convergence and testing for independence.

Iterative algorithms define a broad class of simulation algorithms that are widely applicable because the possibly unknown normalising constant of the target pdf cancels in the acceptance probability. Iterative algorithms are typically easy to implement, but their convergence rate and mixing time are model-specific and highly dependent on the specification of the proposal function $g(\cdot|\cdot)$ and the initial state. Iterative algorithms are frequently used for Bayesian inversion, where the normalising constant is complicated to calculate. Still, for Bayesian spatial models in high dimensions, algorithms based on brute-force/single-site proposal pdfs typically have far too slow convergence and poor mixing to be applicable. In practice, it is necessary to provide a model-specific proposal function for iterative algorithms to be useful for complex spatial problems.

Algorithm 20: Iterative algorithm

```

Initialise  $\mathbf{s}^0$  such that  $p(\mathbf{s}^0) > 0$ 
Define a conditional proposal pdf  $g(\mathbf{s}'|\mathbf{s})$ 
for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^p \sim g(\mathbf{s} | \mathbf{s}^{b-1})$ 
    Calculate  $\alpha = \max \left\{ 1, \frac{p(\mathbf{s}^p)}{p(\mathbf{s}^{b-1})} \frac{g(\mathbf{s}^{b-1} | \mathbf{s}^p)}{g(\mathbf{s}^p | \mathbf{s}^{b-1})} \right\}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        | Set  $\mathbf{s}^b = \mathbf{s}^p$ 
    else
        | Set  $\mathbf{s}^b = \mathbf{s}^{b-1}$ 
    end
end

Result:  $\mathbf{s}^* = \mathbf{s}^{n_b} \sim p^{n_b}(\mathbf{s}) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{s})$ 

```

10.1.5 Approximate Algorithms

Various approximations can be used to assess the target pdf $p(\mathbf{s})$ without the need to rely on the asymptotic convergence of iterative algorithms. One alternative is the class of sampling importance resampling (SIR) algorithms defined in Algorithm 21.

The SIR algorithms are suitable for pdfs of the form $p(\mathbf{s}) = \text{const} \times h(\mathbf{s})$, where the normalising constant is unfeasible to calculate. The algorithm is initiated by specifying a pdf $p^*(\mathbf{s})$ similar to the target pdf and easy to simulate from. The sample space is discretised by n_b realisations from this pdf, and weights are assigned to each realisation. The weights are based on the ratio between $p(\mathbf{s})$ and $p^*(\mathbf{s})$ for each realisation. The weight vector is normalised and interpreted as probabilities. Lastly, re-simulation from the discretised pdf with associated probabilities is performed to obtain one realisation from the target pdf. The SIR algorithm is simple to implement and efficiently yields an approximate realisation from the target pdf. In practice, assessing whether the approximate realisation is of good or poor quality is challenging. The approximation, which is dependent on the

density of the n_b sample values, is subject to the curse of dimensionality. Hence, the approximation can be poor for high-dimensional models. If the normalised weights for a large proportion of the n_b sample values are much less than $1/n_b$, it is an indication of a poor representation of the target pdf $p(\mathbf{s})$.

The SIR algorithm is frequently used for Bayesian inversion, where $p^*(\mathbf{s})$ is specified as the prior model. However, the curse of dimensionality becomes severe for Bayesian spatial modelling in high dimensions, and the approximations are expected to be unacceptable.

Algorithm 21: Approximate algorithm

Specify a pdf $p^*(\mathbf{s})$ that is similar to the pdf $p(\mathbf{s})$ and easy to simulate from.

```

for  $b = 1, 2, \dots, n_b$  do
    Generate  $\mathbf{s}^{*b} \sim p^*(\mathbf{s})$ 
    Calculate  $w^b = \frac{p(\mathbf{s}^{*b})}{p^*(\mathbf{s}^{*b})}$ 
end
Calculate  $\bar{w}^b = \frac{w^b}{\sum_{c=1}^{n_b} w^c}$  for  $b = 1, 2, \dots, n_b$ 
Resample  $\mathbf{s}^{*s} \sim [\mathbf{s}^{*1}, \mathbf{s}^{*2}, \dots, \mathbf{s}^{*n_b}]$  with probability  $[\bar{w}^1, \bar{w}^2, \dots, \bar{w}^{n_b}]$ 
Result:  $\mathbf{s}^s = \mathbf{s}^{*s}$  approximates  $\mathbf{s}^s \sim p(\mathbf{s})$ 
```

10.2 Geostatistics: Kriging Prediction Models

This section contains a presentation of the fundamentals of traditional Kriging prediction techniques cast in the notation of the book. More comprehensive presentations can be found in Journel and Huijbregts (1978) and Chiles and Delfiner (2012).

Consider a continuous RF $\{r(\mathbf{x}); \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^3\}$ and assume that a set of related spatial variables $\{g_j(\mathbf{x}); \mathbf{x} \in \mathcal{D}; j = 1, 2, \dots, n_g\}$ is observed in the reference domain \mathcal{D} . Here, each $g_j(\mathbf{x}); j = 1, 2, \dots, n_g$ is an explanatory spatial variable. Define the model assumptions

$$\{E\{r(\mathbf{x})\} = \mu(\mathbf{x}) = \mu_r^0 + \sum_{j=1}^{n_g} \beta_r^j g_j(\mathbf{x}); \mathbf{x} \in \mathcal{D}\}$$

$$\{\text{Var}\{r(\mathbf{x})\} = \sigma^2(\mathbf{x}) = \sigma_r^2; \mathbf{x} \in \mathcal{D}\}$$

$$\{\text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} = \rho(\mathbf{x}, \mathbf{x}') = \rho_r(\boldsymbol{\tau}); \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'; \mathbf{x}, \mathbf{x}' \in \mathcal{D}\}.$$

The function $\gamma_r(\boldsymbol{\tau})$ is referred to as the variogram, although it is traditionally called the semivariogram, and is defined by

$$\{\text{Var}\{r(\mathbf{x}) - r(\mathbf{x}')\} = 2\gamma_r(\boldsymbol{\tau}) = 2\sigma_r^2[1 - \rho_r(\boldsymbol{\tau})]; \boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'; \mathbf{x}, \mathbf{x}' \in \mathcal{D}\}$$

with the model parameters $[\boldsymbol{\beta}_r^+, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$, where the $(n_g + 1)$ -vector is defined as $\boldsymbol{\beta}_r^+ = (\mu_r^0, \beta_r^1, \beta_r^2, \dots, \beta_r^{n_g})^T$. The spatial correlation function $\rho_r(\boldsymbol{\tau})$ is a non-negative definite function. The model is often parametrised by the equivalent conditionally non-positive definite variogram function $\gamma_r(\boldsymbol{\tau}) = \sigma_r^2[1 - \rho_r(\boldsymbol{\tau})]$ instead. The eligible class of variogram functions is larger than the corresponding class of spatial correlation functions. This subject is further discussed in Sect. 6.2.4. Note that only the first two moments are specified and that no Gaussian assumption is made.

Assume that a set of exact observations $\{r(\mathbf{x}_i^d); \mathbf{x}_i^d \in D; i = 1, 2, \dots, m\}$ is available with an associated Dirac-point likelihood function. Note that the Kriging technique can be extended to also cover observations that are linear operators of the spatial variable, corresponding to the Gauss-linear likelihood function.

The Kriging predictor can be specified for an arbitrary unobserved location $\mathbf{x}_0 \in D$, thereby providing a prediction of $r_0 = r(\mathbf{x}_0)$. The predictor is based on the set of observations represented by the m -vector $\mathbf{r}^d = (r(\mathbf{x}_1^d), r(\mathbf{x}_2^d), \dots, r(\mathbf{x}_m^d))^T$. The Kriging predictor is defined to be the BLU predictor under a LSE criterion and can be shown to be of the form

$$\hat{r}_0 = \boldsymbol{\alpha}^T \mathbf{r}^d, \quad (10.1)$$

where the m -vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)^T; \alpha_i \in \mathbb{R}$ contains the Kriging weights to be determined to define the predictor. The unbiasedness requirement and the LSE criterion are used to identify these weights. The spatial variable and the associated prediction variances are sequentially calculated and stored on the grid representation $\{\hat{r}(\mathbf{x}); \mathbf{x} \in L \in D\}$ and $\{\sigma_p^2(\mathbf{x}); \mathbf{x} \in L \in D\}$, respectively. Two classes of traditional Kriging predictors are discussed in some detail: simple Kriging and universal Kriging.

Simple Kriging

The simple Kriging predictor is based on a model assuming all model parameters $[\boldsymbol{\beta}_r^+, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$ to be known, and takes the form of Expression (10.1). The unbiasedness requirement for the predictor is

$$E\{\hat{r}_0 - r_0\} = \boldsymbol{\alpha}^T E\{\mathbf{r}^d\} - E\{r_0\} = 0.$$

The requirement is fulfilled by defining the predictor on the residual RF

$$\hat{r}_{0\Delta} = \boldsymbol{\alpha}^T \mathbf{r}_\Delta^d$$

with $\{r_\Delta(\mathbf{x}) = r(\mathbf{x}) - \mathbf{g}_x^T \boldsymbol{\beta}_r^+; \mathbf{x} \in D\}$, and the $(n_g + 1)$ -vector $\mathbf{g}_x = (1, g_1(\mathbf{x}), \dots, g_{n_g}(\mathbf{x}))^T$. Thus, both $\hat{r}_{0\Delta}$ and $r_{0\Delta}$ are centred at zero, and therefore, the unbiasedness criterion is fulfilled. The LSE criterion requires

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \{\text{Var}\{\hat{r}_{0\Delta} - r_{0\Delta}\}\}$$

$$\begin{aligned}
&= \arg \min_{\boldsymbol{\alpha}} \{\text{Var}\{\boldsymbol{\alpha}^T \mathbf{r}_d - r_{0d}\}\} \\
&= \arg \min_{\boldsymbol{\alpha}} \{\sigma_r^2 - 2\boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\rho}_{0d} + \boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\Sigma}_d^\rho \boldsymbol{\alpha}\} \\
&= [\boldsymbol{\Sigma}_d^\rho]^{-1} \boldsymbol{\rho}_{0d},
\end{aligned}$$

which constitutes an optimisation expression with a quadratic objective function that can be solved analytically. In the expression, the correlation m -vector $\boldsymbol{\rho}_{0d}$ is defined as $\rho_r(\mathbf{x}_0 - \mathbf{x}_i^d)$; $i = 1, 2, \dots, m$, i.e. the correlation between the variable to be predicted r_0 and the observations \mathbf{r}^d . The correlation observation matrix $\boldsymbol{\Sigma}_d^\rho$, of dimension $(m \times m)$, is defined by $\rho_r(\mathbf{x}_i^d - \mathbf{x}_j^d)$; $i, j = 1, 2, \dots, m$. The simple Kriging predictor with associated prediction variance is then

$$\begin{aligned}
\hat{r}_0 &= \mathbf{g}_{\mathbf{x}_0}^T \boldsymbol{\beta}_r^+ + \boldsymbol{\rho}_{0d}^T [\boldsymbol{\Sigma}_d^\rho]^{-1} [\mathbf{r}^d - \mathbf{G}_d \boldsymbol{\beta}_r^+] \\
\sigma_p^2 &= \sigma_r^2 [1 - \boldsymbol{\rho}_{0d}^T [\boldsymbol{\Sigma}_d^\rho]^{-1} \boldsymbol{\rho}_{0d}]
\end{aligned}$$

with the matrix $\mathbf{G}_d = [\mathbf{g}_{\mathbf{x}_1^d}, \mathbf{g}_{\mathbf{x}_2^d}, \dots, \mathbf{g}_{\mathbf{x}_m^d}]^T$ of dimension $(m \times (n_g + 1))$.

Observe that the expression for the Kriging predictor \hat{r}_0 does not involve the variance σ_r^2 , while the expression for the prediction variance σ_p^2 does not involve the actual observations \mathbf{r}^d . These are familiar characteristics of linear regression.

Universal Kriging

The universal Kriging predictor is based on a model with unknown parameter $\boldsymbol{\beta}_r^+$ and known parameters $[\sigma_r^2, \rho_r(\boldsymbol{\tau})]$, and takes the form of Expression (10.1). The unbiasedness of the predictor requires

$$\mathbb{E}\{\hat{r}_0 - r_0\} = \boldsymbol{\alpha}^T \mathbb{E}\{\mathbf{r}^d\} - \mathbb{E}\{r_0\} = 0$$

which leads to

$$\mathbf{G}_d^T \boldsymbol{\alpha} = \mathbf{g}_{\mathbf{x}_0}.$$

These unbiasedness requirements combined with the LSE criterion provide the following constrained minimisation

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \arg \min_{\boldsymbol{\alpha}} \{\text{Var}\{\hat{r}_0 - r_0\}\} \\
&= \arg \min_{\boldsymbol{\alpha}} \{\sigma_r^2 - 2\boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\rho}_{0d} + \boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\Sigma}_d^\rho \boldsymbol{\alpha}\} \\
\text{such that } &\mathbf{G}_d^T \boldsymbol{\alpha} = \mathbf{g}_{\mathbf{x}_0}.
\end{aligned}$$

This expression constitutes an optimisation problem for the m -vector $\boldsymbol{\alpha}$ with quadratic objective function and linear equality constraints. Hence, the solution can be analytically obtained using Lagrange multipliers. The solution is

$$\hat{\boldsymbol{\alpha}} = [\boldsymbol{\Sigma}_d^\rho]^{-1} \left(\boldsymbol{\rho}_{0d} - \mathbf{G}_d^T [\boldsymbol{\Sigma}_d^\rho]^{-1} \mathbf{G}_d \right)^{-1} (\mathbf{G}_d^T [\boldsymbol{\Sigma}_d^\rho]^{-1} \boldsymbol{\rho}_{0d} - \mathbf{g}_{\mathbf{x}_0}).$$

Consequently, the universal Kriging predictor with associated prediction variance is

$$\begin{aligned}\hat{r}_0 &= \hat{\boldsymbol{\alpha}}^T \mathbf{r}^d \\ \sigma_p^2 &= \sigma_r^2 \left[1 - 2\hat{\boldsymbol{\alpha}}^T \boldsymbol{\rho}_{0d} + \hat{\boldsymbol{\alpha}}^T \boldsymbol{\Sigma}_d^\rho \hat{\boldsymbol{\alpha}} \right].\end{aligned}$$

Observe that the expression for the Kriging predictor \hat{r}_0 does not involve the variance σ_r^2 and the expression for the prediction variance σ_p^2 does not involve the observed values in \mathbf{r}^d . These are familiar characteristics of linear regression. In geostatistical terminology, the case with unknown constant expectation μ_r^0 and without explanatory spatial variables is called the ordinary Kriging predictor.

Model parameter inference of the parameter $(n_g + 1)$ -vector $\boldsymbol{\beta}_r^+$ can be made by a LSE criterion. Consider the following linear estimator for model parameter $\boldsymbol{\beta}_r^{+i}$

$$\hat{\boldsymbol{\beta}}_r^{+i} = \boldsymbol{\alpha}^T \mathbf{r}^d,$$

where the m -vector $\boldsymbol{\alpha} \in \mathbb{R}^m$ containing the estimator weights must be determined. The unbiasedness requirement of the predictor entails

$$E\{\hat{\boldsymbol{\beta}}_r^{+i} - \boldsymbol{\beta}_r^{+i}\} = \boldsymbol{\alpha}^T E\{\mathbf{r}^d\} - \boldsymbol{\beta}_r^{+i} = 0$$

which leads to

$$\mathbf{G}_d^T \boldsymbol{\alpha} = \delta_{n_g+1}^i,$$

where the $(n_g + 1)$ -vector $\delta_{n_g+1}^i$ is an indicator vector with all entries except the i th equal to zero, $\delta_{n_g+1}^i = (0, \dots, 0, 1, 0, \dots, 0)^T$. These unbiasedness requirements combined with the LSE criterion provide the following constrained minimisation problem:

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \arg \min_{\boldsymbol{\alpha}} \{\text{Var}\{\hat{\boldsymbol{\beta}}_r^{+i} - \boldsymbol{\beta}_r^{+i}\}\} \\ &= \arg \min_{\boldsymbol{\alpha}} \{\boldsymbol{\alpha}^T \sigma_r^2 \boldsymbol{\Sigma}_d^\rho \boldsymbol{\alpha}\} \\ \text{such that } \mathbf{G}_d^T \boldsymbol{\alpha} &= \delta_{n_g+1}^i.\end{aligned}$$

This minimisation can be solved by using Lagrange multipliers to obtain the estimator weights

$$\hat{\boldsymbol{\alpha}} = [\boldsymbol{\Sigma}_d^\rho]^{-1} \mathbf{G}_d [\mathbf{G}_d^T [\boldsymbol{\Sigma}_d^\rho]^{-1} \mathbf{G}_d]^{-1} \boldsymbol{\delta}_{n_g+1}^i.$$

Consequently, the estimator for the model parameters with associated estimation variance is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_r^{+i} &= \hat{\boldsymbol{\alpha}}^T \mathbf{r}^d \\ \sigma_e^2 &= \sigma_r^2 \hat{\boldsymbol{\alpha}}^T \boldsymbol{\Sigma}_r^\rho \hat{\boldsymbol{\alpha}}.\end{aligned}$$

By repeating this estimation procedure for $i = 1, 2, \dots, n_g + 1$, the $(n_g + 1)$ -vector including the model parameters can be assessed as $\hat{\boldsymbol{\beta}}_r^+$. The corresponding estimation variances can be calculated accordingly.

Closing Remarks

The traditional Kriging predictor model is only specified up to second-order moments, and no Gaussian assumption is made. Consequently, exact prediction intervals cannot be provided, nor can simulations from the conditional RF be generated. By adding a Gaussian assumption and rephrasing the universal Kriging prediction model in a hierarchical framework with a conjugate Gaussian prior pdf on $\boldsymbol{\beta}_r^+$, it is possible to provide exact prediction intervals and to generate realisations of the conditional RF. A Bayesian bridge between the simple and universal Kriging models can be established, as demonstrated in Omre and Halvorsen (1989).

10.3 Gaussian Markov RF Models

This section presents the fundamentals of the Gaussian Markov RF models used for spatial prediction and model parameter inference, cast in the notation of the book. A more comprehensive presentation with numerical algorithms and applications can be found in Rue and Held (2005), in which the methodology is defined in a more general graph setting.

Consider a continuous RF spatially discretised to a grid L of size n , $\{r(\mathbf{x}); \mathbf{x} \in L \subset D \subset \mathbb{R}^3\}$, represented by the n -vector \mathbf{r} . Let \mathbf{r} be Gaussian with expectation n -vector $\boldsymbol{\mu}_r$ and covariance matrix $\boldsymbol{\Sigma}_r$, of dimension $(n \times n)$. The latter must be non-negative definite. In traditional Gaussian RF modelling, the stationary prior model is defined by the expectation and variance levels, $\mu_r \in \mathbb{R}$ and $\sigma_r^2 \in \mathbb{R}_+$, respectively, and a positive definite spatial correlation function $\rho_r(\tau) \in \mathbb{R}_{[-1,1]}$. The model parameters of the Gaussian prior pdf are $\boldsymbol{\mu}_r = \mu_r \mathbf{i}_n$ and $\boldsymbol{\Sigma}_r = \sigma_r^2 \boldsymbol{\Sigma}_r^\rho$.

In the Gaussian Markov RF model, the parametrisation is different. The expectation $\mu_r \mathbf{i}_n$ is identical but may be extended to contain explanatory variables, as specified in Sect. 6.2.3. The spatial dependence structure is represented by the

precision matrix $\Psi_r = [\Sigma_r]^{-1}$ of dimension $(n \times n)$. Note that the precision matrix is defined as the inverse of the grid value covariance matrix and is also non-negative definite. The precision matrix is related to the conditional independence characteristics of the vector \mathbf{r} as follows:

$$p(r_i, r_j | \mathbf{r}_{-ij}) = p(r_i | \mathbf{r}_{-ij})p(r_j | \mathbf{r}_{-ij})$$

equivalent to

$$[\Psi_r]_{ij} = [\Psi_r]_{ji} = 0.$$

The precision matrix can express the conditional moments of the Gaussian model as

$$\begin{aligned} E\{r_i | \mathbf{r}_{-i}\} &= \mu_r - [\Psi_r]_{ii}^{-1} \sum_{j=1; j \neq i}^n [\Psi_r]_{ij}(r_j - \mu_r) \\ \text{Var}\{r_i | \mathbf{r}_{-i}\} &= [\Psi_r]_{ii}^{-1} \\ \text{Corr}\{r_i, r_j | \mathbf{r}_{-ij}\} &= \frac{[\Psi_r]_{ij}}{[(\Psi_r)_{ii}(\Psi_r)_{jj}]^{1/2}}. \end{aligned}$$

Observe that these expressions do not involve the inverse of the $((n - 1) \times (n - 1))$ precision matrix formed by excluding the i th row and column from the precision matrix Ψ_r . In contrast, the corresponding inverse of the $((n - 1) \times (n - 1))$ covariance matrix does appear in the corresponding expressions based on the traditional Gaussian model parametrisation.

The purpose of using the Gaussian Markov RF formulation is to define a spatially local predictor with favourable computational characteristics. These favourable features occur if the associated precision matrix Ψ_r , of dimension $(n \times n)$, is locally sparse. One challenge is that Ψ_r must be non-negative definite, which severely limits modelling flexibility.

The Gaussian Markov RF model, with a sparse precision matrix Ψ_r , is well suited as a prior model for spatial prediction in studies with Dirac-point observation likelihood models, as defined in Sect. 5.1. This likelihood model entails that the observations are acquired locationwise in a subset of grid nodes without observation errors. They appear in the m -vector $[\mathbf{d}_{oe} | \mathbf{r}] = \mathbf{H}_o \mathbf{r} = \mathbf{r}^d$, where the observation design matrix \mathbf{H}_o , of dimension $(m \times n)$, is binary with each row and column adding to one. Reorder the n -vector \mathbf{r} as a concatenation of the values in the unobserved grid nodes and the values in the observed grid nodes, $\mathbf{r} = (\mathbf{r}_o^T, \mathbf{r}^{dT})^T$. Let the precision matrix be decomposed accordingly:

$$\Psi_r = \begin{bmatrix} \Psi_o & \Psi_{od} \\ \Psi_{do} & \Psi_d \end{bmatrix}.$$

The sub-matrices on the diagonal in this decomposition are non-negative definite, and they are usually also sparse. Thus, the conditional expectations and variances of the values in the unobserved grid nodes given the observed ones are according to Rue and Held (2005)

$$\begin{aligned} E\{\mathbf{r}_o | \mathbf{r}^d\} &= \boldsymbol{\mu}_{o|d} = \mu_r \mathbf{i}_{n-m} - \boldsymbol{\Psi}_o^{-1} \boldsymbol{\Psi}_{od} (\mathbf{r}^d - \mu_r \mathbf{i}_m) \\ \text{Var}\{\mathbf{r}_o | \mathbf{r}^d\} &= \boldsymbol{\Sigma}_{o|d} = \boldsymbol{\Psi}_o^{-1}. \end{aligned}$$

The complicated term in these expressions is the inverse of the matrix $\boldsymbol{\Psi}_o$, of dimension $((n-m) \times (n-m))$, but because the matrix is specified to be sparse, computationally efficient Cholesky decomposition algorithms are available. The matrix can be Cholesky decomposed in a number of operations proportional to $(n-m)^{3/2}$ for $D \subset \mathbb{R}^2$ and to $(n-m)^2$ for $D \subset \mathbb{R}^3$, as given in Rue and Held (2005).

The discretised prediction $\{\hat{r}(\mathbf{x}); \mathbf{x} \in L \subset D\}$ is represented by the n -vector $(\boldsymbol{\mu}_{o|d}^T, \mathbf{r}^{dT})^T$. The corresponding prediction variances in the unobserved grid nodes are given by the diagonal entries of the matrix $\boldsymbol{\Sigma}_{o|d}$ and are zero in the observed grid nodes. Furthermore, conditional realisations can be generated on the grid representation by using the decomposition simulation algorithm as defined in Expression (7.4).

For the Dirac-point observation likelihood model specified above, consider parameter inference in a Gaussian Markov RF model parametrised by $[\mu_r, \sigma_r^2, \boldsymbol{\eta}_r]$. The precision matrix $\boldsymbol{\Psi}_r = [\sigma_r^2]^{-1} \boldsymbol{\Psi}_r^\rho$ is parametrised by the variance level $\sigma_r^2 \in \mathbb{R}_+$, and $\boldsymbol{\Psi}_r^\rho = [\boldsymbol{\Sigma}_r^\rho]^{-1}$ is parametrised by $\boldsymbol{\eta}_r \in \mathbb{R}^q$. The MML criterion may assess these model parameters

$$\begin{aligned} [\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\boldsymbol{\eta}}_r] &= \underset{\mu_r, \sigma_r^2, \boldsymbol{\eta}_r}{\text{argmax}} \left\{ p(\mathbf{r}^d; \mu_r, \sigma_r^2, \boldsymbol{\eta}_r) \right\} \\ &= \underset{\mu_r, \sigma_r^2, \boldsymbol{\eta}_r}{\text{argmax}} \left\{ [2\pi]^{-m/2} [\sigma_r^2]^{-m/2} |\boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)}|^{1/2} \right. \\ &\quad \times \exp \left(-[2\sigma_r^2]^{-1} (\mathbf{r}^d - \mu_r \mathbf{i}_m)^T \boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)} (\mathbf{r}^d - \mu_r \mathbf{i}_m) \right) \left. \right\}. \end{aligned}$$

The three corresponding maximum conditional marginal likelihood estimators are

$$\begin{aligned} [\hat{\mu}_r | \sigma_r^2, \boldsymbol{\eta}_r] &= [\mathbf{i}_m^T \boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)} \mathbf{i}_m]^{-1} \times \mathbf{i}_m^T \boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)} \mathbf{r}^d \\ [\hat{\sigma}_r^2 | \mu_r, \boldsymbol{\eta}_r] &= m^{-1} (\mathbf{r}^d - \mu_r \mathbf{i}_m)^T \boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)} (\mathbf{r}^d - \mu_r \mathbf{i}_m) \\ [\hat{\boldsymbol{\eta}}_r | \mu_r, \sigma_r^2] &= \underset{\boldsymbol{\eta}_r}{\text{arg min}} \{ \ln |\boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)}| + (\mathbf{r}^d - \mu_r \mathbf{i}_m)^T [\sigma_r^2]^{-1} \boldsymbol{\Psi}_d^{\rho(\boldsymbol{\eta}_r)} (\mathbf{r}^d - \mu_r \mathbf{i}_m) \}. \end{aligned}$$

The first two optimisations are analytically tractable, and thus $[\hat{\mu}_r, \hat{\sigma}_r^2 | \boldsymbol{\eta}_r]$ can be assessed analytically. The latter optimisation must be performed numerically, but because the q -vector $\boldsymbol{\eta}_r$ containing parameters is usually low-dimensional this

optimisation is typically feasible. Therefore, the MML estimator $[\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\eta}_r]$ can be assessed iteratively.

The likelihood calculation involves the computation of the determinant of the normalised observation precision matrix $\Psi_d^{\rho|\eta_r}$ of dimension $(m \times m)$. If this matrix is sparse, as it usually is, a Cholesky decomposition algorithm can calculate the determinant efficiently. Observe that the expressions for $[\mu_r \mid \sigma_r^2, \eta_r]$ and $[\sigma_r^2 \mid \mu_r, \eta_r]$ only involve the normalised observation precision matrix $\Psi_d^{\rho|\eta_r}$ itself; no other matrix inverses are needed.

Consider the general Gaussian RF model. The prior stationary Gaussian pdf for the n -vector \mathbf{r} is as previously defined. The general Gauss-linear likelihood model is specified as $[\mathbf{d}|\mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e}_{d|r}$, with an arbitrary observation design matrix \mathbf{H} , of dimension $(m \times n)$, and the m -vector $\mathbf{e}_{d|r}$ including observation error is centred Gaussian with precision matrix $\Psi_{d|r}$ of dimension $(m \times m)$. The general posterior model in the Gaussian Markov RF parametrisation, corresponding to Expression (7.3) in the Gaussian parametrisation, can be developed as demonstrated in Lindgren et al. (2011)

$$[\mathbf{r}|\mathbf{d}] \sim p(\mathbf{r} \mid \mathbf{d}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \Psi_{r|d}^{-1})$$

with

$$\begin{aligned} \boldsymbol{\mu}_{r|d} &= \mu_r \mathbf{i}_n + \Psi_{r|d}^{-1} \mathbf{H}^T \Psi_{d|r} (\mathbf{d} - \mu_r \mathbf{H} \mathbf{i}_n) \\ \Psi_{r|d} &= \Psi_r + \mathbf{H}^T \Psi_{d|r} \mathbf{H}. \end{aligned}$$

This expression requires the Cholesky decomposition of an $(n \times n)$ matrix for the entire grid. This Cholesky decomposition can be computationally demanding, even when this matrix is sparse. The same applies to the computation of the likelihood function in model parameter inference. Therefore, assigning a Gaussian Markov RF prior model for spatial prediction in cases with general Gauss-linear likelihood models will not necessarily ensure computational efficiency.

The Gaussian Markov RF model formulation is grid-specific. In grid-specific studies, where the objective is the prediction of a discretised spatial variable based on exact observations in a subset of grid nodes on a specified grid, the Gaussian Markov RF model formulation is suitable. The task involves the Cholesky decomposition of a sparse matrix with dimensions $((n-m) \times (n-m))$. Efficient numerical algorithms are available to facilitate this process. However, the spatially continuous Gaussian RF prior model, as defined in Sect. 6.1, provides much more flexibility since consistent grid designs can be defined and full Gauss-linear likelihood models can be specified. In this model formulation, the observation covariance matrix Σ_d , of dimension $(m \times m)$, is typically required to be Cholesky decomposed to make a spatial prediction. Furthermore, by assuming a finite-range spatial correlation function, as defined in Sect. 9.1, the grid value covariance matrix Σ_r is a sparse matrix. If, in addition, the observation design matrix \mathbf{H} has narrow

spatial support, the observation covariance matrix Σ_d is also sparse. Then many of the computational advantages of Gaussian Markov RF models will also hold for the Gaussian RF model. In practical applications, it is often important to ensure that the grid density is significantly higher than the observation density to efficiently capture the information from the observations. This typically implies that $n \gg m$.

A close connection between the Gaussian Markov RF model and the Gaussian RF model with the Matérn spatial correlation function is demonstrated in Lindgren et al. (2011). This relation follows from the Gaussian RF with a Matérn spatial correlation function being the solution of a specific SPDE. Consequently, a spatially discretised RF can be approximated by a Gaussian Markov RF. The spatial RF representation is typically based on a triangular grid and basis functions with narrow spatial support, which, in combination, is particularly useful for obtaining sparse precision matrices. Efficient numerical algorithms designed for solving the SPDE can be used for spatial prediction based on the Gaussian Markov RF model.

The conditional autoregressive (CAR) model, introduced in Besag (1974), constitutes a special case of the Gaussian Markov RF model. For notational simplicity, the expectation is assumed to be spatially stationary, $\mu_r = \mu_r \mathbf{i}_n$, while the precision matrix Ψ_r , of dimension $(n \times n)$, takes the form

$$[\Psi_r]_{ij} = \begin{cases} \kappa_i & i = j \\ \kappa_i \beta_{ij} & i \neq j \end{cases}$$

where $\kappa_i \in \mathbb{R}_+; i = 1, 2, \dots, n$ are the scale levels and $\beta_{ij} \in \mathbb{R}; i, j = 1, 2, \dots, n; i \neq j$ are the coupling parameters. Furthermore, the precision matrix must be symmetric, i.e. $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}; i = 1, 2, \dots, n; i \neq j$. These specifications do not ensure that the precision matrix Ψ_r is non-negative definite but ensure that the conditional moments for each grid node $i \in L$ are according to the CAR model,

$$\text{E}\{r_i | \mathbf{r}_{-i}\} = \mu_r - \sum_{\substack{j=1 \\ j \neq i}}^n \beta_{ij} (r_j - \mu_r)$$

$$\text{Var}\{r_i | \mathbf{r}_{-i}\} = [\kappa_i]^{-1}.$$

Not all model parameter values for the scale levels and coupling parameters correspond to a valid pdf for the complete n -vector \mathbf{r} , which is required. If the corresponding precision matrix Ψ_r is non-negative definite, then this n -variate pdf exists, as given in Rue and Held (2005).

The usual sufficient, but not necessary, conditions for the entries of the precision matrix to ensure non-negative definiteness are termed the diagonal dominance constraints. These additional constraints are, for each $i = 1, 2, \dots, n$,

$$1.0 - \sum_{\substack{j=1 \\ j \neq i}}^n |\beta_{ij}| > 0.0.$$

So far, sparsity has not been enforced on the CAR model, and in order to do so, a neighbourhood system $\mathbf{n}_L : \{\mathbf{n}_x; \mathbf{x} \in L\}$ must be defined on the spatial grid $L \subset D$, as discussed in Sect. 4.3. The neighbourhood for an arbitrary grid node $\mathbf{x} \in L$ is defined as a set of grid nodes that are near the location \mathbf{x} , excluding the grid node \mathbf{x} itself. Thus, $\mathbf{n}_x \subset L \setminus \mathbf{x}$. To ensure the symmetry constraint $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}; i = 1, 2, \dots, n; i \neq j$, these neighbourhoods need to have the same spatial design in all $\mathbf{x} \in L$ except for regular boundary adjustments. The corresponding localised CAR model is

$$\begin{aligned} E\{r_i | \mathbf{r}_{-i}\} &= \mu_r - \sum_{j \in \mathbf{n}_i} \beta_{ij}(r_j - \mu_r) \\ \text{Var}\{r_i | \mathbf{r}_{-i}\} &= [\kappa_i]^{-1}. \end{aligned}$$

This expression implies that $\beta_{ij} = \beta_{ji} = 0; j \notin \mathbf{n}_i$. Consequently, the corresponding Gaussian Markov RF model has a sparse precision matrix Ψ_r .

The Gaussian Markov RF model is also extended to an intrinsic Gaussian Markov RF version, as defined in Rue and Held (2005), similar to the intrinsic Gaussian RF model discussed in Sect. 6.2.4.

10.4 Gaussian Basis Function RF Models

This section contains a presentation of the fundamentals of Gaussian basis function prediction techniques cast in the notation of the book. A more comprehensive presentation with examples can be found in Cressie and Johannesson (2008) and Cressie et al. (2022).

The Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ is represented by the basis function expression

$$\{r_f(\mathbf{x}) = \mu_r + \sum_{i=1}^{n_f} a_i f_i(\mathbf{x}) = \mu_r + \mathbf{f}(\mathbf{x})^T \mathbf{a}; \mathbf{x} \in D\}$$

with the set of functional n_f -vectors $\{\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_{n_f}(\mathbf{x})]^T; \mathbf{x} \in D\}$ containing the user-specified basis functions. The n_f -vector $\mathbf{a} = [a_1, a_2, \dots, a_{n_f}]^T$ contains random weights that are assumed to be centred Gaussian with $(n_f \times n_f)$ covariance matrix Σ_a . Additional short-range terms are sometimes included to represent small-scale variability, but for notational simplicity, we omit such terms.

The basis function RF model is Gaussian because it is a linear combination of Gaussian random variables. It represents a functional, spatially non-discretised

Gaussian RF, and the model parameters are

$$\{\mu_r(\mathbf{x}) = E\{r_f(\mathbf{x})\} = \mu_r; \mathbf{x} \in D\}$$

$$\{\sigma_r^2(\mathbf{x}, \mathbf{x}') = \text{Cov}\{r_f(\mathbf{x}), r_f(\mathbf{x}')\} = \mathbf{f}(\mathbf{x})^T \boldsymbol{\Sigma}_a \mathbf{f}(\mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}.$$

Consider the spatially discretised Gaussian RF in the n -vector \mathbf{r} , which is based on a grid representation of size n as per the definition provided in Sect. 4.1. Utilise an approach that uses piecewise constant spatial interpolation within the grid. This representation of the RF can be expressed in a basis function format. The number of basis functions corresponds to the grid size, $n_f = n$, and the basis functions are $f_i(\mathbf{x}) = I(\mathbf{x} \in \Delta_i); i = 1, 2, \dots, n_f$, each of which equals one inside the grid unit centred at the corresponding grid node i and zero elsewhere. The weight covariance matrix must be assigned as $\boldsymbol{\Sigma}_a = \sigma_r^2 \boldsymbol{\Sigma}_r^\rho$. The grid infill asymptotic limit as $n_f \rightarrow \infty$ is then the stationary Gaussian RF prior $\{r(\mathbf{x}); \mathbf{x} \in D\}$.

The set of basis functions may have global support and be orthogonal, as, for example, the set of Fourier functions, as defined in Borgman et al. (1984). The asymptotic limit as $n_f \rightarrow \infty$ is then the stationary Gaussian RF prior if the weight covariance matrix $\boldsymbol{\Sigma}_a$ is assigned from the spectral transform of the spatial covariance function $\sigma_r^2 \rho_r(\tau)$.

In Cressie and Johannesson (2008), non-orthogonal basis functions with finite spatial support are used, such as splines or wavelets. However, their motivation is to keep the number of basis functions n_f low to reduce the computational demands. For the representation to be useful, each basis function within the set $\{f_i(\mathbf{x}); \mathbf{x} \in D\}; i = 1, 2, \dots, n_f$ must have a specific spatial reference location $\mathbf{x}_i^b \in D$. Thus, the corresponding reference location set $\{\mathbf{x}_i^b; i = 1, 2, \dots, n_f\}$, which must be user-specified, constitutes an underlying grid design. The shape of the basis functions and their underlying grid design has a crucial impact on the spatial covariance structure of $\{r_f(\mathbf{x}); \mathbf{x} \in D\}$, as specified above.

Consider the expressions for expectation and covariance of the Gaussian basis function RF. The expectation model is stationary but can easily be extended to account for explanatory spatial variables, as discussed in Sect. 6.2.3. The spatial covariance model is more challenging since the Gaussian RF prior model preferably should be stationary with $\sigma_r(\mathbf{x}, \mathbf{x}') = \sigma_r(\mathbf{x} - \mathbf{x}')$, which entails $\sigma_r(\mathbf{x}, \mathbf{x}) = \sigma_r^2$. Any spatial non-stationarity in the prior model should be supported by evidence about the phenomenon under study, and it is unfortunate if this is an artefact of a convenient model parametrisation. The spatial reference set for the basis functions should preferably be spatially dense to provide an approximate stationary prior model, and thus n_f should be large. The prior model is determined by specifying the expectation level μ_r and the covariance matrix $\boldsymbol{\Sigma}_a$, of dimension $(n_f \times n_f)$, for the random weights.

Consider a set of observations with a Gauss-point likelihood model, as defined in Sect. 5.1, $[d_{oi}| \{r(\mathbf{x}); \mathbf{x} \in D\}] = r(\mathbf{x}_i^d) + e_{di|r}; i = 1, 2, \dots, m$. This observation

set in the representation of basis functions is, for $i = 1, 2, \dots, m$,

$$[d_{oi} | \mathbf{a}] = \mu_r + \mathbf{f}(\mathbf{x}_i^d)^T \mathbf{a} + e_{d_i|r}.$$

The observation errors are assumed to be independent of each other, each being a centred Gaussian variable with variance $\sigma_{d|r}^2$. The corresponding Gauss-point likelihood model in basis function representation is

$$[\mathbf{d}_o | \mathbf{a}] = \mu_r \mathbf{i}_m + \mathbf{F}_d \mathbf{a} + \mathbf{e}_{d|r} \sim p(\mathbf{d}_o | \mathbf{a}) = \phi_m(\mathbf{d}_o; \mu_r \mathbf{i}_m + \mathbf{F}_d \mathbf{a}, \sigma_{d|r}^2 \mathbf{I}_m),$$

where the matrix $\mathbf{F}_d = [\mathbf{f}(\mathbf{x}_1^d), \mathbf{f}(\mathbf{x}_2^d), \dots, \mathbf{f}(\mathbf{x}_{n_f}^d)]^T$ of dimension $(m \times n_f)$. The error m -vector is centred Gaussian with covariance matrix $\sigma_{d|r}^2 \mathbf{I}_m$ of dimension $(m \times m)$. The accuracy of this Gauss-point likelihood model will improve with increasing spatial density of basis functions, i.e. with increasing value of n_f . It is recommended to assign $n_f \gg m$ to represent the observations reliably. Expressions similar to the ones above can also be developed for a full Gauss-linear likelihood model at the expense of more complex expressions.

The joint pdf for the random basis function weights and the observations is

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{d}_o \end{bmatrix} \sim p(\mathbf{a}, \mathbf{d}_o) = \phi_{n_f+m} \left(\begin{bmatrix} \mathbf{a} \\ \mathbf{d}_o \end{bmatrix}; \begin{bmatrix} 0\mathbf{i}_{n_f} \\ \mu_r \mathbf{i}_m \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_a \mathbf{F}_d^T \\ \mathbf{F}_d \boldsymbol{\Sigma}_a & \mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m \end{bmatrix} \right).$$

Consequently, the posterior pdf for the random weight vector is

$$[\mathbf{a} | \mathbf{d}_o] \sim p(\mathbf{a} | \mathbf{d}_o) = \phi_{n_f}(\mathbf{a}; \boldsymbol{\mu}_{a|d}, \boldsymbol{\Sigma}_{a|d})$$

with

$$\begin{aligned} \boldsymbol{\mu}_{a|d} &= 0\mathbf{i}_{n_f} + \boldsymbol{\Sigma}_a \mathbf{F}_d^T [\mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m]^{-1} [\mathbf{d}_o - \mu_r \mathbf{i}_m] \\ \boldsymbol{\Sigma}_{a|d} &= \boldsymbol{\Sigma}_a - \boldsymbol{\Sigma}_a \mathbf{F}_d^T [\mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m]^{-1} \mathbf{F}_d \boldsymbol{\Sigma}_a. \end{aligned}$$

The observation covariance matrix $\boldsymbol{\Sigma}_d = \mathbf{F}_d \boldsymbol{\Sigma}_a \mathbf{F}_d^T + \sigma_{d|r}^2 \mathbf{I}_m$, of dimension $(m \times m)$, is typically Cholesky decomposed to calculate the posterior model for the weights. The posterior model for the RF is therefore a Gaussian RF

$$\{[r_f(\mathbf{x}) | \mathbf{d}_o] = \mu_r + \mathbf{f}(\mathbf{x})^T [\mathbf{a} | \mathbf{d}_o]; \mathbf{x} \in D\}$$

with model parameter functions

$$\begin{aligned} \{\mu_{r|d}(\mathbf{x}) &= \mu_r + \mathbf{f}(\mathbf{x})^T \boldsymbol{\mu}_{a|d}; \mathbf{x} \in D\} \\ \{\sigma_{r|d}(\mathbf{x}) \sigma_{r|d}(\mathbf{x}') \rho_{r|d}(\mathbf{x}, \mathbf{x}') &= \mathbf{f}(\mathbf{x})^T \boldsymbol{\Sigma}_{a|d} \mathbf{f}(\mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}. \end{aligned}$$

The posterior Gaussian RF can be simulated in the basis function representation. One realisation $\{[r_f(\mathbf{x})|\mathbf{d}_o] = \mu_r + \mathbf{f}(\mathbf{x})^T [\mathbf{a}|\mathbf{d}_o]; \mathbf{x} \in D\}$ is obtained based on one realisation of the vector of conditional weights $[\mathbf{a}|\mathbf{d}_o]$ from the conditional pdf $p(\mathbf{a}|\mathbf{d}_o)$. The realisation of the Gaussian RF is in functional form and, thus, apparently without any spatial discretisation.

The spatial predictor with associated prediction variance is

$$\{\hat{r}_f(\mathbf{x}) = \mu_r + \mathbf{f}(\mathbf{x})^T \boldsymbol{\mu}_{a|d}; \mathbf{x} \in D\}$$

$$\{\sigma_{pf}^2(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\Sigma}_{a|d} \mathbf{f}(\mathbf{x}); \mathbf{x} \in D\},$$

where $\boldsymbol{\mu}_{a|d}$ and $\boldsymbol{\Sigma}_{a|d}$ are defined earlier. The set of basis functions is assigned a corresponding set of reference locations that define an underlying spatial grid design. Moreover, each basis function is typically specified with finite spatial support. Thus, the predictor in an arbitrary location $\mathbf{x}_o \in D$ is local in the sense that only basis functions and grid nodes defined within their region of spatial support influence the prediction, given the posterior weight vector $\boldsymbol{\mu}_{a|d}$.

Parameter inference under the Gaussian basis function RF model can be performed by using an MML criterion. Let the model parameters be $[\mu_r, \sigma_a^2, \boldsymbol{\eta}_a]$, where $\mu_r \in \mathbb{R}$ is the expectation level, $\sigma_a^2 \in \mathbb{R}_+$ is the coefficient scale in $\boldsymbol{\Sigma}_a = \sigma_a^2 \boldsymbol{\Sigma}_a^0$ and the standardised coefficient covariance matrix $\boldsymbol{\Sigma}_a^0$, of dimension $(n_f \times n_f)$, is parametrised by $\boldsymbol{\eta}_a \in \mathbb{R}^q$. Assume that the latter parametrisation ensures non-negative definiteness, which is not an easy task.

The MML estimator is then defined from the joint pdf

$$[\hat{\mu}_r, \hat{\sigma}_a^2, \hat{\boldsymbol{\eta}}_a] = \operatorname{argmax}_{\mu_r, \sigma_a^2, \boldsymbol{\eta}_a} \left\{ \phi_m(\mathbf{d}_o; \mu_r \mathbf{i}_m, \sigma_a^2 \boldsymbol{\Sigma}_d^{0(\boldsymbol{\eta}_a)}) \right\}$$

with

$$\boldsymbol{\Sigma}_d^{0(\boldsymbol{\eta}_a)} = \mathbf{F}_d \boldsymbol{\Sigma}_a^{0(\boldsymbol{\eta}_a)} \mathbf{F}_d^T + v_{d/a} \mathbf{I}_m.$$

The relative scale coefficient $v_{d/a} = \sigma_{d|a}^2 / \sigma_a^2$ is assumed to be known, which entails that $\boldsymbol{\Sigma}_d^{0|\boldsymbol{\eta}_a} = [\boldsymbol{\Sigma}_d^{0(\boldsymbol{\eta}_a)} | \boldsymbol{\eta}_a]$ is also known. Consequently, the conditional MML estimators can be calculated as

$$[\hat{\mu}_r | \sigma_a^2, \boldsymbol{\eta}_a] = [\mathbf{i}_m^T [\boldsymbol{\Sigma}_d^{0|\boldsymbol{\eta}_a}]^{-1} \mathbf{i}_m]^{-1} \times \mathbf{i}_m^T [\boldsymbol{\Sigma}_d^{0|\boldsymbol{\eta}_a}]^{-1} \mathbf{d}_o$$

$$[\hat{\sigma}_a^2 | \mu_r, \boldsymbol{\eta}_a] = m^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m)^T [\boldsymbol{\Sigma}_d^{0|\boldsymbol{\eta}_a}]^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m)$$

$$[\hat{\boldsymbol{\eta}}_a | \mu_r, \sigma_a^2] = \arg \min_{\boldsymbol{\eta}_a} \{ \ln |\boldsymbol{\Sigma}_d^{0(\boldsymbol{\eta}_a)}| + (\mathbf{d}_o - \mu_r \mathbf{i}_m)^T [\sigma_a^2 \boldsymbol{\Sigma}_d^{0(\boldsymbol{\eta}_a)}]^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m) \}.$$

The first two expressions are in closed analytical form, whereas the latter expression must be calculated numerically. The conditional MML can be obtained iteratively based on the expressions above.

In practice, Σ_a^0 is specified based on experience, which entails that the expectation level and the coefficient scale can be estimated analytically:

$$\begin{aligned} [\hat{\mu}_r | \sigma_a^2, \Sigma_a^0] &= [\mathbf{i}_m^T [\Sigma_d^0]^{-1} \mathbf{i}_m]^{-1} \times \mathbf{i}_m^T [\Sigma_d^0]^{-1} \mathbf{d}_o \\ [\hat{\sigma}_a^2 | \mu_r, \Sigma_a^0] &= m^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m)^T [\Sigma_d^0]^{-1} (\mathbf{d}_o - \mu_r \mathbf{i}_m). \end{aligned}$$

An approximately stationary prior model with reliable representation of the observations requires $n_f \gg m$. Thus, the computational demand of the Gaussian basis function predictor is comparable to that of the traditional Gaussian predictor. The observation covariance matrix Σ_d , of dimension $(m \times m)$, is typically Cholesky decomposed, which requires calculations proportional to m^3 in the general case. If the basis functions are assigned with narrow spatial support, and the covariance matrix for the weights is specified to be sparse, then the observation covariance matrix Σ_d appears as sparse. In such cases, efficient Cholesky decomposition algorithms can be used to assess the matrix inverse. If the flexibility of the prior model and a reliable representation of the observations are of less importance than computational efficiency, one may specify $n_f < m$, as done in Cressie and Johannesson (2008). If so, the Woodbury matrix inversion identity, as given in Guttmann (1946), is

$$[\mathbf{F}_d \Sigma_a \mathbf{F}_d^T + \sigma_{d|a}^2 \mathbf{I}_m]^{-1} = [\sigma_{d|a}^2]^{-1} [\mathbf{I}_m - \mathbf{F}_d [\sigma_{d|a}^2 \Sigma_a^{-1} + \mathbf{F}_d^T \mathbf{F}_d]^{-1} \mathbf{F}_d^T]$$

can be used to reduce the computational demands of the Cholesky decomposition of an $(n_f \times n_f)$ matrix.

The standard narrow spatial support basis function model is extended to a multi-level version in Nychka et al. (2015), which is more computationally efficient.

10.5 Kernel Predictors for Gaussian RF Models

This section contains a presentation of the fundamentals of kernel predictor techniques for Gaussian RF models. A more comprehensive discussion with examples is available in Omre and Spremec (2023).

The prior stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$, or more compactly $r(\cdot)$, is parametrised by $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$. The model parameters are the expectation and variance levels $\mu_r(\mathbf{x}) = \mu_r \in \mathbb{R}$ and $\sigma_r^2(\mathbf{x}) = \sigma_r^2 \in \mathbb{R}_+$ and the non-negative definite spatial correlation function $\rho_r(\boldsymbol{\tau}) \in \mathbb{R}_{[-1,1]}$; $\boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$. Extensions to expectation models with explanatory spatial variables, as discussed in Sect. 6.2.3, can be made. Note in particular that spatial discretisation is avoided, and thus, no spatial grid representation is enforced.

Consider a set of m observations with a corresponding set of reference locations $M : \{\mathbf{x}_i^d; i = 1, 2, \dots, m\}$ and the m -vector $\mathbf{d} = (d_1, d_2, \dots, d_m)^T$ containing the observed values. The likelihood models for the observations are assumed to be, for

$i = 1, 2, \dots, m$, of the form

$$[d_i | r(\cdot)] = l_i(r(\cdot)) + e_{d|r,i}.$$

The m -vector $\mathbf{l}(r(\cdot)) = (l_1(r(\cdot)), l_2(r(\cdot)), \dots, l_m(r(\cdot)))^T$ containing the response functions has entries that are linear operators on the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$ in the location set M . The observation error m -vector $\mathbf{e}_{d|r}$ is independent centred Gaussian with variances $\sigma_{d|r}^2$. The response functions are linear operators on the continuous spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$. Thus, no spatial discretisation is required. There are three important types of such response functions: locationwise, integration and differentiation functions. In the observation location $\mathbf{x}^d \in D$, provided the response functions exist, they are typically given by

$$\begin{aligned} l_L(r(\cdot)) &= r(\mathbf{x}^d) \\ l_I(r(\cdot)) &= \int_{D_{\mathbf{x}^d}} h_{\mathbf{x}^d}(\mathbf{u}) r(\mathbf{u}) d\mathbf{u} \\ l_D(r(\cdot)) &= \left. \frac{dr(\mathbf{u})}{d\mathbf{u}_z} \right|_{\mathbf{u}=\mathbf{x}^d}, \end{aligned}$$

with the domain $D_{\mathbf{x}^d} \subset D$ and the weighting function $h_{\mathbf{x}^d}(\mathbf{u}) < \infty$ for $\mathbf{u} \in D_{\mathbf{x}^d}$. Furthermore, the differentiation is made along the \mathbf{z} -direction of \mathbb{R}^3 in location $\mathbf{u} \in D$, denoted by \mathbf{u}_z . The former expression provides the value of the spatial variable in one specific location $\mathbf{x}^d \in D$. The middle provides a weighted integral of the spatial variable over a sub-domain $D_{\mathbf{x}^d} \subset D$ with reference location $\mathbf{x}^d \in D$. The latter provides the \mathbf{z} -direction derivative of the spatial variable at the location $\mathbf{x}^d \in D$.

The complete Gauss-linear likelihood model is

$$[\mathbf{d}|r(\cdot)] \sim p(\mathbf{d} | r(\cdot)) = \phi_m(\mathbf{d}; \boldsymbol{\mu}_{d|r}, \boldsymbol{\Sigma}_{d|r})$$

with the m -vector $\boldsymbol{\mu}_{d|r} = \mathbf{l}(r(\cdot))$ containing the observation expectations and with error covariance matrix $\boldsymbol{\Sigma}_{d|r} = \sigma_{d|r}^2 \mathbf{I}_m$ of dimension $(m \times m)$.

Consider spatial prediction in one arbitrary location $\mathbf{x}_o \in D$ with value $r(\mathbf{x}_o) = r_o$. The $(1+m)$ -vector $(r(\mathbf{x}_o), \mathbf{d})$ is jointly Gaussian:

$$\begin{bmatrix} r(\mathbf{x}_o) \\ \mathbf{d} \end{bmatrix} \sim p(r_o, \mathbf{d}) = \phi_{1+m} \left(\begin{bmatrix} r_o \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_r \\ \boldsymbol{\mu}_d \end{bmatrix}, \begin{bmatrix} \boldsymbol{\sigma}_r^2 & \boldsymbol{\sigma}_{od} \\ \boldsymbol{\sigma}_{do} & \boldsymbol{\Sigma}_d \end{bmatrix} \right),$$

where the elements in $\boldsymbol{\mu}_d$, $\boldsymbol{\sigma}_{od} = \boldsymbol{\sigma}_{do}^T$ and $\boldsymbol{\Sigma}_d$, for $i, j = 1, 2, \dots, m$, are

$$\begin{aligned} [\boldsymbol{\mu}_d]_i &= l_i^{\mathbf{u}}(\boldsymbol{\mu}_r(\mathbf{u})) \\ [\boldsymbol{\sigma}_{od}]_i &= \sigma_r^2 l_i^{\mathbf{u}}(\rho_r(\mathbf{x}_o - \mathbf{u})) \\ [\boldsymbol{\Sigma}_d]_{ij} &= \sigma_r^2 l_i^{\mathbf{u}}(l_j^{\mathbf{v}}(\rho_r(\mathbf{v} - \mathbf{u}))) + \sigma_{d|r}^2 I(i = j). \end{aligned}$$

The upper index \mathbf{u} in $l_i^{\mathbf{u}}(\cdot)$ refers to the argument variable on which the linear operator acts.

The posterior model is also Gaussian:

$$[r(\mathbf{x}_o)|\mathbf{d}] \sim p(r_o | \mathbf{d}) = \phi_1(r_o; \mu_{o|d}, \sigma_{o|d}^2)$$

with

$$\begin{aligned}\mu_{o|d} &= E\{r(\mathbf{x}_o)|\mathbf{d}\} = \mu_r + \boldsymbol{\sigma}_{od} \boldsymbol{\Sigma}_d^{-1} (\mathbf{d} - \boldsymbol{\mu}_d) \\ \sigma_{o|d}^2 &= \text{Var}\{r(\mathbf{x}_o)|\mathbf{d}\} = \sigma_r^2 - \boldsymbol{\sigma}_{od} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\sigma}_{do}.\end{aligned}$$

The associated LSE predictor of $r(\mathbf{x}_o)$, based on the observation values \mathbf{d} , and its corresponding prediction variance are

$$\begin{aligned}\hat{r}(\mathbf{x}_o) &= E\{r(\mathbf{x}_o)|\mathbf{d}\} = \mu_r + \boldsymbol{\sigma}_{od} \mathbf{w}^d \\ \sigma_p^2(\mathbf{x}_o) &= \text{Var}\{r(\mathbf{x}_o)|\mathbf{d}\} = \sigma_r^2 - \boldsymbol{\sigma}_{od} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\sigma}_{do}.\end{aligned}$$

The m -vector $\mathbf{w}^d = \boldsymbol{\Sigma}_d^{-1}(\mathbf{d} - \boldsymbol{\mu}_d) = (w_1^d, w_2^d, \dots, w_m^d)^T$ contains the random weights, which depends on the random observation vector \mathbf{d} but not on the prediction location $\mathbf{x}_o \in D$. The expectation vector and covariance matrix of the random weight vector are $E\{\mathbf{w}^d\} = 0_{m \times 1}$ and $\text{Var}\{\mathbf{w}^d\} = \boldsymbol{\Sigma}_d^{-1}$. The prediction value and observation cross-covariance m -vector $\boldsymbol{\sigma}_{od}$ is deterministic because it is a function of the model parameters only.

The spatial kernel predictor for the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$, based on the observation \mathbf{d} , is defined as

$$\left\{ \hat{r}(\mathbf{x}) = \mu_r + \boldsymbol{\sigma}_{xd} \mathbf{w}^d = \mu_r + \sigma_r^2 \sum_{i=1}^m w_i^d v_i(\mathbf{x}); \mathbf{x} \in D \right\}.$$

The associated prediction variances are

$$\left\{ \sigma_p^2(\mathbf{x}) = \sigma_r^2 - \boldsymbol{\sigma}_{xd} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{\sigma}_{dx} = \sigma_r^2 [1 - \sigma_r^2 \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} v_i(\mathbf{x}) v_j(\mathbf{x})]; \mathbf{x} \in D \right\}.$$

The observation kernel function m -vector $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}), \dots, v_m(\mathbf{x}))^T$ is defined as $v_i(\mathbf{x}) = l_i^{\mathbf{u}}(\rho_r(\mathbf{x} - \mathbf{u})); i = 1, 2, \dots, m$. The coefficients in the expression are $\beta_{ij} = [\boldsymbol{\Sigma}_d^{-1}]_{ij}; i, j = 1, 2, \dots, m$.

The response functions are typically either locationwise, integrated or differential. These functions provide the typical observation kernel functions centred in the observation location $\mathbf{x}^d \in D$, provided they exist

$$\begin{aligned}\{v_L(\mathbf{x}) = l_L^{\mathbf{u}}(\rho_r(\mathbf{x} - \mathbf{u})) = \rho_r(\mathbf{x} - \mathbf{x}^d) ; \mathbf{x} \in D\} \\ \{v_I(\mathbf{x}) = l_I^{\mathbf{u}}(\rho_r(\mathbf{x} - \mathbf{u})) = \int_{D_{\mathbf{x}^d}} h_{\mathbf{x}^d}(\mathbf{u}) \rho_r(\mathbf{x} - \mathbf{u}) d\mathbf{u} ; \mathbf{x} \in D\} \\ \{v_D(\mathbf{x}) = l_D^{\mathbf{u}}(\rho_r(\mathbf{x} - \mathbf{u})) = \left. \frac{d\rho_r(\mathbf{x} - \mathbf{u})}{d\mathbf{u}_z} \right|_{\mathbf{u}=\mathbf{x}^d} ; \mathbf{x} \in D\}.\end{aligned}$$

For certain spatial correlation functions, the observation kernel functions set in $v(\mathbf{x})$ are analytically tractable, and the kernel functions are expressed in closed form. In such cases, the spatial kernel predictor appears as a weighted average of the analytical observation kernels with observation reference locations from the set M . The associated prediction variances are quadratic in the kernel expressions. Consequently, both the spatial kernel predictor and the associated prediction variance are in a functional representation, avoiding any spatial discretisation.

The weight vector $\mathbf{w}^d = \Sigma_d^{-1}(\mathbf{d} - \mu_d)$ is independent of the predictor location $\mathbf{x} \in D$ and thus only needs to be calculated once. The primary computational demand comes from the Cholesky decomposition of the observation covariance matrix Σ_d of dimension $(m \times m)$.

The spatial kernel predictor with prediction variances is optimal in a LSE sense for a Gaussian RF model. The traditional Kriging predictor shares the same characteristics for the predictions in the grid nodes of a grid representation. In fact, the kernel predictor provides the grid infill asymptotic limit for the Kriging predictor as the grid density approaches infinity. Observe that both predictors have computational demands that primarily depend on the Cholesky decomposition of the observation covariance matrix Σ_d of dimension $(m \times m)$.

In some studies, the focus is on linear operations of the spatial variable $\{r(\mathbf{x}); \mathbf{x} \in D\}$, such as the Δ -filtered spatial variable or the \mathbf{z} -directional derivative of the spatial variable. The kernel predictor is in functional representation. Therefore, the actual linear operator can be activated directly on the kernel predictor. The optimality of the resulting predictor under Gaussian model assumptions is ensured.

The kernel predictor in functional representation can also be used in simulation from the posterior Gaussian RF model, providing realisations in functional representation. Consider the decomposition

$$\{[r(\mathbf{x})|\mathbf{d}]^s = E\{r(\mathbf{x})|\mathbf{d}\} + [r_*^s(\mathbf{x}) - E\{r(\mathbf{x})|\mathbf{d}_*^s\}]; \mathbf{x} \in D\},$$

where $\{r_*^s(\mathbf{x}); \mathbf{x} \in D\}$ represents a realisation from the Gaussian RF prior model and \mathbf{d}_*^s is the observation value m -vector collected from this realisation using the observation design of the study. The expression appears as a decomposition of a realisation from the Gaussian posterior model. Let the realisation $\{r_*^s(\mathbf{x}); \mathbf{x} \in D\}$ be

in a Fourier representation, which can be provided efficiently for the prior stationary Gaussian RF model. If the kernel predictor provides the expectations, all three right-hand side expressions are in functional form. Consequently, the realisation of the posterior Gaussian RF model is in a functional representation.

Inference of the model parameters $[\mu_r, \sigma_r^2, \eta_r]$ can be performed by using a MML criterion:

$$[\hat{\mu}_r, \hat{\sigma}_r^2, \hat{\eta}_r] = \operatorname{argmax}_{\mu_r, \sigma_r^2, \eta_r} \{ \phi_m(\mathbf{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d) \}.$$

The optimisation can be partly analytically made, as demonstrated in Sect. 8.1.

The kernel predictor involves the Cholesky decomposition of the observation covariance matrix $\boldsymbol{\Sigma}_d$, of dimension $(m \times m)$, with computational demands proportional to m^3 . For large observation sets, that is m large, this matrix inversion may not be computationally feasible. Sparse $(m \times m)$ matrices can be Cholesky decomposed more efficiently, with the number of steps proportional to $m^{3/2}$ for $D \subset \mathbb{R}^2$ and proportional to m^2 for $D \subset \mathbb{R}^3$. The kernel predictor is optimal for any non-negative definite spatial correlation function under Gaussian assumptions. By limiting the model class to finite-range correlation functions with $\rho_r(\tau) = 0.0$ for $|\tau| > \tau_0$ with $\tau_0 < \infty$, the matrix $\boldsymbol{\Sigma}_d$ is sparse. Thus, the Cholesky decomposition can be performed more efficiently while optimality for this subclass of Gaussian models is maintained. Observe that, in this case, the observation kernels are zero at some distance from the actual observation location because the finite-range correlation functions are zero beyond τ_0 . Consequently, the kernel predictor appears as local in the sense that only observations within a distance of τ_0 from the prediction location influence the predictor, given the weight vector \mathbf{w}^d .

For very large m , not even sparsification of the matrix $\boldsymbol{\Sigma}_d$, of dimension $(m \times m)$, can make Cholesky decomposition computationally feasible. An approximate version of the kernel predictor is designed for such big data applications. The spatially localised kernel predictor is constructed to have computational demands proportional to m for $D \subset \mathbb{R}^3$.

In the localised kernel predictor, the troublesome inverse observation covariance matrix $\boldsymbol{\Sigma}_d^{-1}$ is replaced by an approximation $\boldsymbol{\Sigma}_d^{-1*}$. This approximation is based on the Cholesky decomposition of localised observation covariance sub-matrices at each of the m observation locations. These sub-matrices are combined to provide the inverse of the full observation covariance matrix. The algorithm for obtaining this approximate matrix is specified in Algorithm 22.

The observation covariance matrix $\boldsymbol{\Sigma}_d$ has rows and columns corresponding to the entries in the observation location set M . The operator Sub-matrix $\{\boldsymbol{\Sigma}_d; M_x^\Delta\}$ extracts the sub-matrix of $\boldsymbol{\Sigma}_d$ containing the rows and columns corresponding to the observation location sub-set M_x^Δ . The support matrix Ψ is sparse because the location set $M_{x_i^q}$ is a sub-set of M . Consequently, not all entries in row i are updated. Finally, $\boldsymbol{\Sigma}_d^{-1*}$ is both sparse and symmetric but typically not non-negative definite. The localised kernel prediction $\{\hat{r}^*(\mathbf{x}); \mathbf{x} \in D\}$ and prediction variance $\{\sigma_p^{2*}(\mathbf{x}); \mathbf{x} \in D\}$ are obtained from the regular expressions with replacing the matrix $\boldsymbol{\Sigma}_d^{-1}$ by its

Algorithm 22: Observation covariance matrix approximation

Define a localisation range $\Delta \in \mathbb{R}_+$

Define the support ($m \times m$)-matrix $\Psi = 0\mathbf{I}_m$

for $i = 1, 2, \dots, m$ **do**

Define the neighbourhood set: $\mathbb{M}_{\mathbf{x}_i^d}^\Delta = \{\mathbf{y} | \mathbf{y} \in \mathbb{M}, |\mathbf{x}_i^d - \mathbf{y}| < \Delta\}$ of dimension $m_{\mathbf{x}_i^d}^\Delta$

Define the matrix $\Sigma_{d_\mathbf{x}^\Delta} = \text{Sub-matrix}\{\Sigma_d; \mathbb{M}_{\mathbf{x}_i^d}^\Delta\}$ of dimension $(m_{\mathbf{x}_i^d}^\Delta \times m_{\mathbf{x}_i^d}^\Delta)$

Compute $\Sigma_{d_\mathbf{x}^\Delta}^{-1}$ of dimension $(m_{\mathbf{x}_i^d}^\Delta \times m_{\mathbf{x}_i^d}^\Delta)$

Insert $[\Psi]_{ij} = \left[\Sigma_{d_\mathbf{x}^\Delta}^{-1} \right]_{\mathbf{x}_i^d, \mathbf{x}_j^d}$ for corresponding j and $\mathbf{y} \in \mathbb{M}_{\mathbf{x}_i^d}^\Delta$ matrix entries

end

Result: Define $\Sigma_d^{-1*} = 1/2 \times [\Psi + \Psi^T]$

approximation Σ_d^{-1*} . Model parameter inference can be performed accordingly to obtain $\hat{\mu}_r^*$ and $\hat{\sigma}_r^{2*}$.

The computational requirements are dominated by m Cholesky decompositions of $(m_{\mathbf{x}}^\Delta \times m_{\mathbf{x}}^\Delta)$ matrices, where Δ defining the size of $m_{\mathbf{x}}^\Delta$ is a user-specified approximation parameter. Because the dimensions of the sub-matrices are approximately fixed at $m_{\mathbf{x}}^\Delta$, the computational demands are proportional to the number of observations m . If Δ is very large, such that $\mathbb{M}_{\mathbf{x}}^\Delta = \mathbb{M}$ for all $\mathbf{x} \in \mathcal{D}$, it can be demonstrated that $\Sigma_d^{-1*} = \Sigma_d^{-1}$. The approximation becomes rougher as the size of Δ decreases. Observe further that the algorithm is simple to implement on a computer, and that parallel processing is possible.

The resulting localised kernel predictor is not optimal in the LSE sense under the specified Gaussian RF prior model. However, the predictor appears with the correct spatial continuity. No artificial discontinuities due to the approximation occur. If exact observations are available, the predictor does not correctly reproduce these. The deviation between the exact observations and the respective predictions is represented by the corresponding deviation variance σ_*^2 . This variance provides a quantification of the approximation error. Because of the approximation, the prediction variances are not ensured to be strictly non-negative. A more representative prediction variance, which also captures the approximation error, can be defined as $\{\hat{\sigma}_p^2(\mathbf{x}) = \sigma_p^{2*}(\mathbf{x}) + \sigma_*^2; \mathbf{x} \in \mathcal{D}\}$. Alternative localised predictors based on a patchwork of kernel predictors are discussed in Aunon and Gomez-Hernandez (2000) and Vigsnes et al. (2017). These predictors will, however, introduce artefacts across patch borders.

10.6 Clustered/Repulsive Event RF

This section contains a presentation of models for event RFs with spatial interactions, termed clustered and repulsive event RFs. Consider an event RF $\{k(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^3\}$ and let it be represented by the event location set $\mathbb{X}_D : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_D}\}; \mathbf{x}_i \in D$, where the set size k_D is random. For stationary Poisson RFs, the distribution of event locations in a ball with radius t centred at \mathbf{x}_0 is independent of whether an event is located at \mathbf{x}_0 . Consequently, the interaction J -function is independent of t . For cluster and repulsive event RFs, the interaction J -function is respectively higher and lower than average for small values of t .

Clustered Event RF

Two alternative models for clustered event spatial variables are presented. Both models are based on a non-stationary Poisson RF model with a randomly defined intensity function $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$ and base event set $E = \emptyset$. Thus, they are based on a hierarchical Poisson RF model.

Cox Event RF The Cox event RF model, as defined in Møller et al. (1998), is a hierarchical Poisson RF model with intensity function defined as $\{\lambda(\mathbf{x}) = \exp(r(\mathbf{x}); \mathbf{x} \in D\}$. The RF $\{r(\mathbf{x}); \mathbf{x} \in D\}$ is a stationary Gaussian RF with model parameters $[\mu_r, \sigma_r^2, \rho_r(\boldsymbol{\tau})]$. In practice, a spatial discretisation is made, and the discretised Cox event count RF $\{k_\Delta^C(\mathbf{x}); \mathbf{x} \in L \in D\}$ represented by the event count n -vector $\mathbf{k}^C \in \mathbb{N}_+^n$ is

$$\mathbf{k}^C \sim p(\mathbf{k}^C) = \int_{\mathbb{R}_+^n} p(\mathbf{k}^C | \boldsymbol{\pi}_n) p(\boldsymbol{\pi}_n) d\boldsymbol{\pi}_n,$$

where the n -vector $\boldsymbol{\pi}_n = \exp(\mathbf{r}_{\Delta_n})$ and the $r_{\Delta_n} = \frac{1}{\Delta_n} \int_{\Delta_n} r(\mathbf{u}) d\mathbf{u}$. The latter is a Δ_n -average filtered Gaussian RF, as defined in Sect. 6.2, with model parameters $[\mu_{\Delta_n}, \sigma_{\Delta_n}^2, \rho_{\Delta_n}(\boldsymbol{\tau})]$. The corresponding approximate Cox event location set \mathbb{X}_D^{Cn} is defined as in Sect. 4.2.

The Cox event RF model is only partly analytically tractable, as discussed in Møller et al. (1998). However, it may be evaluated by simulation based on the sequential simulation algorithm, as defined in Sect. 10.1. Algorithm 23 defines a sequential algorithm for the Cox event RF. The output from the algorithm is the n -vector \mathbf{k}^{Cs} and the n_k -set \mathbb{X}_D^{Cns} , which are realisations of the event count vector and the approximate event location set for the Cox event RF, respectively.

Neyman–Scott Event RF The Neyman–Scott (mother–child) event RF model, as discussed in Illian et al. (2008), is a hierarchical Poisson RF model with the mother intensity function specified by a stationary Poisson event RF model with intensity λ_M . Centred at each mother location $\{\mathbf{x}_j^M; j = 1, 2, \dots, k_D^M\}$, a set of child events are independently located, with set size according to the count pdf $p(k_j^c)$ and the intensity pdf $p(\mathbf{x} | \mathbf{x}_j^M)$. The intensity pdf is typically a Gaussian pdf

Algorithm 23: Sequential simulation: Cox event RF

```

Initiate  $n_k = 0$ 
Generate  $\mathbf{r}_{\Delta_n}^s \sim \phi_n(\mathbf{r}; \mu_{\Delta_n} \mathbf{i}_n, \sigma_{\Delta_n}^2 \boldsymbol{\Sigma}_{\Delta_n}^\rho)$ 
Compute  $\pi_n^s = \exp\{\mathbf{r}_{\Delta_n}^s\}$ 
for  $i = 1, 2, \dots, n$  do
    Generate  $k_i^s \sim p(k_i) = \frac{[\pi_{ni}^s]^{k_i}}{k_i!} \exp\{-\pi_{ni}^s\}$ 
    if  $k_i^s > 0$  then
        for  $j = 1, \dots, k_i^s$  do
            Generate  $\mathbf{x}_{n_k+j}^s \sim \text{unif}[\Delta_{ni}]$ 
        end
    end
    Set  $n_k = n_k + k_i^s$ 
end
Result:  $\mathbf{k}^{Cs} = (k_1^s, k_2^s, \dots, k_n^s)$  and  $\mathbb{X}_{\mathcal{D}}^{Cns} = (\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{n_k}^s)$ 

```

Algorithm 24: Sequential simulation: Neyman–Scott event RF

```

Initiate  $k = 0$ 
Generate  $k_{\mathcal{D}}^{Ms} \sim p(k) = \frac{[\lambda_M |\mathcal{D}|]^k}{k!} \exp\{-\lambda_M |\mathcal{D}|\}$ 
for  $j = 1, 2, \dots, k_{\mathcal{D}}^{Ms}$  do
    Generate  $\mathbf{x}_j^{Ms} \sim \text{unif}[\mathcal{D}]$ 
    Generate  $k_j^{Cs} \sim p(k^c)$ 
    for  $i = 1, 2, \dots, k_j^{Cs}$  do
        Generate  $\mathbf{x}_{ji}^{ps} \sim \phi_3(\mathbf{x}; \mathbf{x}_j^{Ms}, \sigma_c^2 \mathbf{I}_3)$ 
        Generate  $\mathbf{x}_{ji}^s = \text{LocTorus-3D}(\mathbf{x}_{ji}^{ps}; \mathcal{D})$ 
    end
    Set  $k = k + k_j^{Cs}$ 
end
Result:  $k_{\mathcal{D}}^s = k$  and  $\mathbb{X}_{\mathcal{D}}^{Nss} : \{\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{k_{\mathcal{D}}^s}^s\} = \{\mathbf{x}_{ji}^s; j = 1, 2, \dots, k_{\mathcal{D}}^{Ms}; i = 1, 2, \dots, k_j^{Cs}\}$ 

```

$\phi_3(\mathbf{x}; \mathbf{x}^M, \sigma_c^2 \mathbf{I}_3)$. The vector of model parameters is then $\boldsymbol{\theta}_{pNS} = [\lambda_M, \sigma_c^2, p(k)]$. Furthermore, there are boundary effects to consider because of the finite domain \mathcal{D} .

The Neyman–Scott event RF is represented by the event location set $\mathbb{X}_{\mathcal{D}}^{Ns} : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_{\mathcal{D}}}^s\} : \{\mathbf{x}_{ji}; j = 1, 2, \dots, k_{\mathcal{D}}^M; i = 1, 2, \dots, k_j^c\}; \mathbf{x}_{ji} \in \mathcal{D}$ with $k_{\mathcal{D}} = \sum_{j=1}^{k_{\mathcal{D}}^M} k_j^c$. This Neyman–Scott event RF model is not analytically tractable, but realisations can be generated by a sequential algorithm, as defined in Algorithm 24. Note, however, that boundary corrections must be made by using a torus, or wrapping, representation of the event locations on \mathcal{D} . In the algorithm, the LocTorus-3D($\cdot; \cdot$) function enforces torus boundary conditions, as described in Algorithm 25. In this algorithm, the modulo operation applied to the two real variables yields the signed remainder resulting from the division of the first variable by the second. Thus, the event location set $\mathbb{X}_{\mathcal{D}}^{Ns}$ appears as a realisation of the Neyman–Scott event RF with model parameter $\boldsymbol{\theta}_{pNS} = [\lambda_M, \sigma_c^2, p(k)]$.

Algorithm 25: LocTorus-3D (\mathbf{x} ; D)

Input coordinate: $\mathbf{x} \in \mathbb{R}^3$
 Lower-left domain D corner: $\mathbf{x}_L \in \mathbb{R}^3$
 Upper-right domain D corner: $\mathbf{x}_H \in \mathbb{R}^3$ where $\mathbf{x}_H > \mathbf{x}_L$

```

for  $i = 1, 2, 3$  do
  Set  $x_T^i = x_L^i + [(x^i - x_L^i) \bmod (x_H^i - x_L^i)]$ 
  if  $x_T^i < 0$  then
    | Set  $x_T^i = x_H^i + x_T^i$ 
  end
end
Result: Coordinate  $\mathbf{x}_T = (x_T^1, x_T^2, x_T^3) \in D \subset \mathbb{R}^3$ 
```

Repulsive Event RF

The Strauss model for repulsive event spatial variables, as discussed in Illian et al. (2008), is presented. The Strauss model is typically specified in conditional form, given the event count $k_D = k$, as

$$\begin{aligned} [\mathbb{X}_D^S | k_D = k] &\sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | k_D = k) \\ &= \text{const} \times \prod_{\substack{i,j=1 \\ i \neq j}}^k \exp(-\phi(\mathbf{x}_i - \mathbf{x}_j)). \end{aligned}$$

The pairwise interaction function $\phi(\tau_{ij}) \in \mathbb{R}_+$; $\tau_{ij} = \mathbf{x}_i - \mathbf{x}_j$ is a declining function with increasing argument $|\tau_{ij}|$. The interaction function is often isotropic, with $\tau_{ij} = |\tau_{ij}| \in \mathbb{R}_+$, and typically takes the form

$$\phi(\tau) = \begin{cases} \phi_0 & 0 \leq \tau \leq \tau_0 \\ \phi_0 \exp(-\phi_1(\tau - \tau_0)) & \tau \geq \tau_0 \end{cases}$$

with $\tau_0 \in \mathbb{R}_+$ and $\phi_0, \phi_1 \in \mathbb{R}_+$. Thus, the model parameter is $\boldsymbol{\theta}_{PS|k} = [\tau_0, \phi_0, \phi_1]$. Note that whenever two events are within a distance of τ_0 , the pdf is multiplied by a penalty of $\exp(-\phi_0)$. Moreover, the pdf becomes zero if $\phi_0 = \infty$. Thus, with $\phi_0 = \infty$, the Strauss model defines a hard-core repulsion model. The Strauss model generally imposes penalties whenever events are located near each other. Boundary effects may occur because the domain D is finite.

The Strauss event RF model is not analytically tractable, but realisations can be generated by an iterative algorithm, as defined in Algorithm 26. In the iterative algorithm, the proposal pdf is symmetric: $g(\mathbf{x}|\mathbf{x}') = g(\mathbf{x}'|\mathbf{x})$. Thus, the proposal pdf does not impact the acceptance probability α because it cancels in the acceptance ratio. The event location set $\mathbb{X}_D^{n_b}$ will, in the limit $n_b \rightarrow \infty$, approach $\mathbb{X}_D^{S_s}$, which

is a realisation of the Strauss event RF. The algorithm provides a realisation with boundary corrections using a torus distance between events as provided by the algorithm DistTorus-3D($\cdot, \cdot; \cdot$). This algorithm is described in Algorithm 27.

Algorithm 26: Iterative McMC simulation: Strauss event RF

```

Initiate  $\mathbb{X}_D^0 : \{\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_k^0\}$  such that  $p(\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_k^0 | k_D = k) > 0$ 
for  $b = 1, 2, \dots, n_b$  do
    Generate  $i \sim \text{unif}[1, 2, \dots, k]$ 
    Generate  $\mathbf{x}^p \sim \text{unif}[D]$ 
    for  $j = 1, 2, \dots, k$  do
         $d_j^p = \text{DistTorus-3D}(\mathbf{x}^p, \mathbf{x}_j^b; D)$ 
         $d_j^i = \text{DistTorus-3D}(\mathbf{x}_i^b, \mathbf{x}_j^b; D)$ 
    end
    Calculate  $\alpha = \min \left\{ 1, \exp \left\{ - \sum_{j=1; j \neq i}^k (\phi(d_j^p) - \phi(d_j^i)) \right\} \right\}$ 
    Generate  $u \sim \text{unif}[0, 1]$ 
    if  $u \leq \alpha$  then
        Set  $\mathbb{X}_D^b : \{\mathbf{x}_1^b, \dots, \mathbf{x}_{i-1}^b, \mathbf{x}^p, \mathbf{x}_{i+1}^b, \dots, \mathbf{x}_k^b\}$ 
    else
        Set  $\mathbb{X}_D^b = \mathbb{X}_D^{b-1}$ 
    end
end

Result:  $\mathbb{X}_D^{n_b} \sim p^{n_b}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | k_D = k) \xrightarrow{n_b \rightarrow \infty} p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | k_D = k)$ 

```

Algorithm 27: DistTorus-3D ($\mathbf{x}, \mathbf{x}_C; D$)

```

Proposal coordinate:  $\mathbf{x} \in D \subset \mathbb{R}^3$ 
Current event coordinate:  $\mathbf{x}_C \in D \subset \mathbb{R}^3$ 
Lower-left domain D corner:  $\mathbf{x}_L \in \mathbb{R}^3$ 
Upper-right domain corner:  $\mathbf{x}_H \in \mathbb{R}^3$  where  $\mathbf{x}_H > \mathbf{x}_L$ 
Diagonal vector  $\mathbf{x}_\Delta = \mathbf{x}_H - \mathbf{x}_L \in \mathbb{R}^3$ 
Compute relative proposal coordinate:  $\mathbf{x}^o = \mathbf{x} - \mathbf{x}_L$ 
Compute relative current coordinate:  $\mathbf{x}_C^o = \mathbf{x}_C - \mathbf{x}_L$ 
for  $i = -1, 0, 1$  do
    for  $j = -1, 0, 1$  do
        for  $k = -1, 0, 1$  do
            Set Torus-shift vector  $\mathbf{x}_{TS} = (i \ x_\Delta^1, j \ x_\Delta^2, k \ x_\Delta^3)^T$ 
            Compute  $\Delta_{ijk} = |\mathbf{x}^o - \mathbf{x}_C^o + \mathbf{x}_{TS}|$ 
        end
    end
end
Set  $\Delta = \min_{ijk} \{\Delta_{ijk}\}$ 
Result: Torus-distance  $\Delta \in \mathbb{R}_+$  between  $\mathbf{x}$  and  $\mathbf{x}_C$ 

```

In cases where the event count $k_D \in \mathbb{R}_+$ is considered random with pdf $p(k_D)$, the Strauss event RF model takes the form

$$\mathbb{X}_D^S \sim p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \mid k_D = k) p(k_D = k).$$

A reversible-jump iterative algorithm can be defined to generate realisations from this model, as discussed in Green (1995).

10.7 INLA Framework for Poisson RF

This section presents the fundamentals of the integrated nested Laplace approximation (INLA) technique cast in the book's notation. A more comprehensive presentation with examples can be found in Rue and Martino (2007) and van Niekerk et al. (2023).

The INLA model framework can be cast in a hierarchical Bayesian model framework. The variable of primary interest is the continuous RF $\{r(\mathbf{x}); \mathbf{x} \in D\}$, usually spatially discretised on the grid $L \subset D$ of size n . The n -vector $\mathbf{r} \in \mathbb{R}^n$ represents the RF. The prior model assigned to \mathbf{r} is $p(\mathbf{r} \mid \boldsymbol{\theta}_p)$, with random parameter q_p -vector $\boldsymbol{\theta}_p$. Note that typically $q_p \ll n$. The observations are represented by the m -vector \mathbf{d} , acquired according to the likelihood model $p(\mathbf{d} \mid \mathbf{r}, \boldsymbol{\theta}_l)$ with random parameter q_l -vector $\boldsymbol{\theta}_l$. The random model parameter q -vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_p, \boldsymbol{\theta}_l)$, with $q = q_p + q_l$, is assigned prior model $p(\boldsymbol{\theta})$. In the INLA framework, the prior expectation may include a complex function of the model parameters based on a set of explanatory variables. Similarly, the likelihood model may appear as non-linear in the variable of interest.

In the INLA framework, the focus is typically on the two posterior pdfs $p(\mathbf{r} \mid \mathbf{d})$ and $p(\boldsymbol{\theta} \mid \mathbf{d})$. The INLA decompositions of these two pdfs are

$$\begin{aligned} [\mathbf{r}|\mathbf{d}] &\sim p(\mathbf{r} \mid \mathbf{d}) = \int p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{d}) d\boldsymbol{\theta} \\ [\boldsymbol{\theta}|\mathbf{d}] &\sim p(\boldsymbol{\theta} \mid \mathbf{d}) = \text{const} \times \frac{p(\mathbf{d} \mid \mathbf{r}, \boldsymbol{\theta}_l) p(\mathbf{r} \mid \boldsymbol{\theta}_p) p(\boldsymbol{\theta})}{p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta})} \Big|_{\mathbf{r}=\mathbf{r}^*(\boldsymbol{\theta})}. \end{aligned}$$

The former expression is easy to justify, although both factors in the integrand are challenging to assess. The latter expression appears from two alternative factorisations of the joint model:

$$\begin{aligned} p(\mathbf{r}, \mathbf{d}, \boldsymbol{\theta}) &= p(\mathbf{d} \mid \mathbf{r}, \boldsymbol{\theta}_l) p(\mathbf{r} \mid \boldsymbol{\theta}_p) p(\boldsymbol{\theta}) \\ &= p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{d}) p(\mathbf{d}). \end{aligned}$$

Rearranging the latter identity, recognising that $p(\mathbf{d})$ is a constant, yields the latter decomposition. Note that the identity is valid for all $\mathbf{r} \in \mathbb{R}^n$, including $\mathbf{r}^*(\boldsymbol{\theta})$ which

is the mode of the prior model $p(\mathbf{r} \mid \boldsymbol{\theta}_p)$. The three factors in the numerator coincide with the likelihood and two prior models. The term in the denominator is, however, challenging to assess.

The challenging term is $p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta})$, which represents the posterior model for the n -vector \mathbf{r} given the parameter q -vector $\boldsymbol{\theta}$. This expression must be calculated for all eligible values of $\boldsymbol{\theta}$. Traditionally, the INLA framework relies on the assumption that $p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta})$ is Gaussian:

$$p^*(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d,\theta}^*, \boldsymbol{\Sigma}_{r|d,\theta}^*).$$

The model parameters, the expectation n -vector $\boldsymbol{\mu}_{r|d,\theta}^*$ and $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}_{r|d,\theta}^*$, are traditionally assigned the mode vector and associated Hessian matrix of the posterior model $p(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta}) = \text{const} \times p(\mathbf{d} \mid \mathbf{r}, \boldsymbol{\theta}_l) p(\mathbf{r} \mid \boldsymbol{\theta}_p)$, as demonstrated in Eidsvik et al. (2009). These calculations are demanding because they operate in the full grid dimension n . This approximation is recognised as the Laplace approximation in Tierney et al. (1989). Alternatively, a variational Bayes approach can be used, as discussed in van Niekerk and Rue (2023). The n -vector $\mathbf{r}^*(\boldsymbol{\theta})$, which is the mode of the prior model, is typically simple to assess because the prior model is spatially stationary.

The posterior model $p(\boldsymbol{\theta} \mid \mathbf{d})$ is herein defined up to a normalising constant, but the exact value of that constant can usually not be calculated. In the traditional INLA approach, the sample space of the model parameter-vector $\boldsymbol{\theta}$ is discretised by a grid L_θ . This discretisation is feasible if the dimension q is very small. The normalised values of the expression for the posterior model, with an inserted approximation for $p^*(\mathbf{r} \mid \mathbf{d}, \boldsymbol{\theta})$, for $\boldsymbol{\theta}$ in the grid nodes of L_θ , are computed. These values are represented by $\{p_{\boldsymbol{\theta}|\mathbf{d}}^*; \boldsymbol{\theta} \in L_\theta\}$, from which $p^*(\boldsymbol{\theta} \mid \mathbf{d})$ can be assessed.

The corresponding INLA approach for assessing the posterior model for the RF of primary interest is based on numerical integration as

$$p^*(\mathbf{r} \mid \mathbf{d}) = \sum_{\boldsymbol{\theta} \in L_\theta} \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d,\theta}^*, \boldsymbol{\Sigma}_{r|d,\theta}^*) p_{\boldsymbol{\theta}|\mathbf{d}}^*.$$

This approximate posterior model appears as a spatially discretised mixture of Gaussian RFs. Recent versions of the INLA package also include a marginal transformation to account for marginal skewness. The INLA framework focuses on the marginal posterior models of $p^*(\boldsymbol{\theta} \mid \mathbf{d})$ and $p^*(\mathbf{r} \mid \mathbf{d})$. Thus, the focus is on $p^*(\theta_i \mid \mathbf{d}); i = 1, 2, \dots, q$ and $p^*(r_i \mid \mathbf{d}); i = 1, 2, \dots, n$. These marginal posterior distributions can be assessed from the joint pdfs.

The INLA framework relies heavily on Gaussian assumptions. However, an analytically tractable and fully Gaussian approximation can be formulated, as follows:

$$p^*(\mathbf{r} \mid \boldsymbol{\theta}_p = (\boldsymbol{\alpha}, \sigma^2)) = \phi_n(\mathbf{r}; \mathbf{G}^* \boldsymbol{\alpha}, \sigma^2 \boldsymbol{\Sigma}_r^{\rho*})$$

$$p^*(\mathbf{d} \mid \mathbf{r}, \boldsymbol{\theta}_l = (\sigma^2)) = \phi_m(\mathbf{d}; \mathbf{H}^* \mathbf{r}, v^* \sigma^2 \mathbf{I}_m)$$

with the $(q - 1)$ -vector $\boldsymbol{\alpha} \in \mathbb{R}^{q-1}$ containing the random model parameters and with the scalar $\sigma^2 \in \mathbb{R}_+$. The approximate linear prior expectation is specified by the $(n \times (q - 1))$ matrix \mathbf{G}^* . The approximate prior spatial correlation function $\rho_r^*(\tau)$ that defines the correlation matrix $\Sigma_r^{\rho*}$ must be specified. Likewise, the approximate linear likelihood specified by the $(m \times n)$ matrix \mathbf{H}^* and the corresponding approximate variance scaling $v^* \in \mathbb{R}_+$ must be calculated. Observe that INLA requires similar approximations to justify that the posterior $p(\mathbf{r} | \mathbf{d}, \boldsymbol{\theta})$ is Gaussian.

The q -vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \sigma^2)$ containing the random parameters can be assigned the prior model $p(\boldsymbol{\alpha}, \sigma^2) = p(\boldsymbol{\alpha} | \sigma^2)p(\sigma^2)$. If the first factor is $(q - 1)$ -dimensional Gaussian and the second factor is inverse-gamma, which are conjugate priors for Gaussian models, then the INLA model is fully analytically tractable. The posterior model for the two model parameters is $p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{d}) = p(\boldsymbol{\alpha} | \sigma^2, \mathbf{d})p(\sigma^2 | \mathbf{d})$, where the first factor is Gaussian and the second is inverse-gamma. In this case, the marginal posterior pdf $p(\boldsymbol{\alpha} | \mathbf{d})$ is multivariate T-distributed. The corresponding posterior model for the spatial variable $p(\mathbf{r} | \mathbf{d})$ is a spatially discretised T-distributed RF, as discussed in Røislien and Omre (2006) and Sect. 7.1.3.

The INLA framework is frequently used to evaluate hierarchical Poisson RF models, as described in Illian et al. (2013) and Sect. 10.7. We outline the use in evaluating the spatially discretised log-Gaussian Poisson count RF model, also termed the log-Gaussian Cox RF model. The spatial event intensity $\{\lambda(\mathbf{x}); \mathbf{x} \in D\}; \lambda(\mathbf{x}) \in \mathbb{R}_+$ is the variable of primary interest. This variable is spatially discretised on a grid $L \subset D$ and represented in the n -vector $\boldsymbol{\lambda}$. The prior model for the intensity vector $\boldsymbol{\lambda}$ is a hierarchical log-Gaussian model, with $\lambda_L = \log \lambda$ and

$$[\lambda_L | \boldsymbol{\theta}_p] \sim p(\lambda_L | \boldsymbol{\theta}_p) = \phi_n(\lambda_L; \boldsymbol{\mu}_{\lambda|\boldsymbol{\theta}_p}, \sigma_{\lambda|\boldsymbol{\theta}_p}^2 \boldsymbol{\Sigma}_{\lambda|\boldsymbol{\theta}_p}^\rho).$$

The explanatory spatial variables are included in the expectation n -vector $\boldsymbol{\mu}_{\lambda|\boldsymbol{\theta}_p}$, and the prior variance is $\sigma_{\lambda|\boldsymbol{\theta}_p}^2$. The associated correlation matrix $\boldsymbol{\Sigma}_{\lambda|\boldsymbol{\theta}_p}^\rho$, of dimension $(n \times n)$, captures the prior spatial dependence. The q_p -vector $\boldsymbol{\theta}_p$ contains the random parameters. This prior model corresponds to the $p(\mathbf{r} | \boldsymbol{\theta}_p)$ in the introduction to the INLA framework earlier.

The observation n -vector $\mathbf{k} \in \mathbb{N}_+^n$ contains the exact event counts in the spatially discretised model. The likelihood model can be expressed as

$$[\mathbf{k} | \boldsymbol{\lambda}] \sim p(\mathbf{k} | \boldsymbol{\lambda}) = \prod_{i=1}^n \frac{[\lambda_i \Delta_n]^{k_i}}{k_i!} \exp(-\lambda_i \Delta_n),$$

where the transformation $\boldsymbol{\lambda} = \exp(\lambda_L)$ ensures that $\boldsymbol{\lambda} \in \mathbb{R}_+^n$. The likelihood model is specified without parameters. This likelihood model corresponds to $p(\mathbf{d} | \mathbf{r}, \boldsymbol{\theta}_l)$ in the INLA introduction above. The random model parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}_p$ is assigned the prior model $p(\boldsymbol{\theta})$.

Based on these specifications, the INLA framework can be employed to assess $p^*(\lambda_i | \mathbf{k}) ; i = 1, 2, \dots, n$ and $p^*(\theta_i | \mathbf{k}) ; i = 1, 2, \dots, q$. Observe that in the notation of the book, the model discussed above would be referred to as a doubly

hierarchical Poisson RF model because the hyper-parameter θ is also considered to be random.

10.8 Markov Random Profile Versus Markov Random Chain

Consider a one-dimensional spatially discretised mosaic random profile (RP) $\mathbf{l} = (l_1, l_2, \dots, l_n)$ with $l_i \in \mathbb{L}$ and assume that it is a Markov RP with a clique system consisting of all nearest pairs of nodes. The following results can, however, be generalised to general clique systems at the expense of more complex notation. Define a Markov RP model in Gibbs form

$$p(\mathbf{l}) = \text{const} \times \prod_{i=1}^n \alpha_i(l_i) \times \prod_{i=1}^{n-1} \beta_i(l_i, l_{i+1}) \quad (10.2)$$

with a clique system that consists of all pairs of nearest nodes. The model parameters are

$$\begin{aligned} \alpha_i(l_i) &\in \mathbb{R}_+; l_i \in \mathbb{L}, i = 1, 2, \dots, n \\ \beta_i(l_i, l_{i+1}) &\in \mathbb{R}_+; l_i, l_{i+1} \in \mathbb{L}, i = 1, 2, \dots, n-1. \end{aligned}$$

The corresponding Markov form of the Markov RP model can be defined from the relation valid for $k = 2, 3, \dots, n-1$ as

$$\begin{aligned} p(l_k | \mathbf{l}_{-k}) &= \frac{p(\mathbf{l})}{\sum_{l'_k \in \mathbb{L}} p(l'_k, \mathbf{l}_{-k})} \\ &= \frac{\text{const} \times \prod_{i=1}^{k-1} \alpha_i(l_i) \times \prod_{i=1}^{k-2} \beta_i(l_i, l_{i+1})}{\text{const} \times \prod_{i=1}^{k-1} \alpha_i(l_i) \times \prod_{i=1}^{k-2} \beta_i(l_i, l_{i+1})} \\ &\quad \times \frac{\alpha_k(l_k) \times \beta_{k-1}(l_{k-1}, l_k) \beta_k(l_k, l_{k+1})}{\sum_{l'_k \in \mathbb{L}} \alpha_k(l'_k) \times \beta_{k-1}(l_{k-1}, l'_k) \beta_k(l'_k, l_{k+1})} \\ &\quad \times \frac{\prod_{i=k+1}^n \alpha_i(l_i) \times \prod_{i=k+1}^{n-1} \beta_i(l_i, l_{i+1})}{\prod_{i=k+1}^n \alpha_i(l_i) \times \prod_{i=k+1}^{n-1} \beta_i(l_i, l_{i+1})} \\ &= \left[\sum_{l'_k \in \mathbb{L}} \alpha_k(l'_k) \times \beta_{k-1}(l_{k-1}, l'_k) \beta_k(l'_k, l_{k+1}) \right]^{-1} \\ &\quad \times \alpha_k(l_k) \times \beta_{k-1}(l_{k-1}, l_k) \beta_k(l_k, l_{k+1}) \\ &= p(l_k | l_{k-1}, l_{k+1}). \end{aligned}$$

Therefore, the Markov formulation has a neighbourhood design that contains the two nearest nodes, with suitable boundary adjustments. The Markov formulation is

$$\begin{aligned}
 p(l_n | l_{n-1}) &= \left[\sum_{l'_n \in \mathbb{L}} \alpha_n(l'_n) \times \beta_{n-1}(l_{n-1}, l'_n) \right]^{-1} \alpha_n(l_n) \times \beta_{n-1}(l_{n-1}, l_n) \\
 p(l_k | l_{k-1}, l_{k+1}) &= \left[\sum_{l'_k \in \mathbb{L}} \alpha_k(l'_k) \times \beta_{k-1}(l_{k-1}, l'_k) \beta_k(l'_k, l_{k+1}) \right]^{-1} \\
 &\quad \times \alpha_k(l_k) \times \beta_{k-1}(l_{k-1}, l_k) \beta_k(l_k, l_{k+1}); k = 2, 3, \dots, n-1 \\
 p(l_1 | l_2) &= \left[\sum_{l'_1 \in \mathbb{L}} \alpha_1(l'_1) \times \beta_1(l'_1, l_2) \right]^{-1} \alpha_1(l_1) \times \beta_1(l_1, l_2).
 \end{aligned}$$

To develop the expressions for the corresponding Markov RC, it is convenient first to demonstrate that this Markov RC is also first-order. Consider the general sequential decomposition

$$p(\mathbf{l}) = p(l_1) \times \prod_{i=2}^n p(l_i | l_{i-1}, l_{i-2}, \dots, l_1)$$

and the first-order Markov RC defined by

$$p(l_i | l_{i-1}, l_{i-2}, \dots, l_1) = p(l_i | l_{i-1}); i = 2, 3, \dots, n.$$

Use the Markov RP on Gibbs form to demonstrate this first-order dependence:

$$\begin{aligned}
 p(l_i | l_{i-1}, l_{i-2}, \dots, l_1) &= \frac{\sum_{l'_n \in \mathbb{L}} \cdots \sum_{l'_{i+1} \in \mathbb{L}} p(l'_n, \dots, l'_{i+1}, l_i, l_{i-1}, \dots, l_1)}{\sum_{l'_i \in \mathbb{L}} [\sum_{l'_n \in \mathbb{L}} \cdots \sum_{l'_{i+1} \in \mathbb{L}} p(l'_n, \dots, l'_{i+1}, l'_i, l_{i-1}, \dots, l_1)]} \\
 &= \frac{\sum_{l'_n \in \mathbb{L}} \cdots \sum_{l'_{i+1} \in \mathbb{L}} [\prod_{j=i+1}^n \alpha_j(l'_j) \prod_{j=i+1}^{n-1} \beta_j(l'_j, l'_{j+1}) \beta_i(l_i, l'_{i+1})] \alpha_i(l_i) \beta_{i-1}(l_{i-1}, l_i)}{\sum_{l'_i \in \mathbb{L}} [\sum_{l'_n \in \mathbb{L}} \cdots \sum_{l'_{i+1} \in \mathbb{L}} [\prod_{j=i+1}^n \alpha_j(l'_j) \prod_{j=i+1}^{n-1} \beta_j(l'_j, l'_{j+1}) \beta_i(l'_i, l'_{i+1})] \alpha_i(l'_i) \beta_{i-1}(l_{i-1}, l'_i)]} \\
 &\times \frac{\prod_{j=1}^{i-1} \alpha_j(l_j) \prod_{l=1}^{i-2} \beta_j(l_j, l_{j+1}) \times \text{const}}{\prod_{j=1}^{i-1} \alpha_j(l_j) \prod_{l=1}^{i-2} \beta_j(l_j, l_{j+1}) \times \text{const}}
 \end{aligned}$$

$$\begin{aligned}
&= \left[\sum_{l'_i \in \mathbb{L}} h_i(l'_i) \alpha_i(l'_i) \beta_{i-1}(l_{i-1}, l'_i) \right]^{-1} h(l_i) \alpha_i(l_i) \beta_{i-1}(l_{i-1}, l_i) \\
&= p(l_i \mid l_{i-1}),
\end{aligned}$$

where $h_i(l_i) = \sum_{l'_n \in \mathbb{L}} \cdots \sum_{l'_{i+1} \in \mathbb{L}} [\prod_{j=i+1}^n \alpha_j(l'_j) \prod_{j=i+1}^{n-1} \beta_j(l'_j, l'_{j+1}) \beta_i(l_i, l'_{i+1})]$. Observe that $h_i(l_i)$ includes only $\alpha_j(\cdot); j = i+1, i+2, \dots, n$ and $\beta_j(\cdot, \cdot); j = i, i+1, \dots, n$, i.e. only the parameters at the same node and above. From the expression above, the Markov RC is parametrised by $p(l_1)$ and $p(l_i \mid l_{i-1}); i = 2, 3, \dots, n$.

The relations between the parametrisations of the Markov RP and Markov RC are

$$\begin{aligned}
p(l_1 \mid l_2) &= \text{const} \times p(l_2 \mid l_1) p(l_1) \\
p(l_i \mid l_{i-1}, l_{i+1}) &= \text{const} \times p(l_{i+1} \mid l_i) p(l_i \mid l_{i-1}); i = 2, 3, \dots, n-1 \\
p(l_n \mid l_{n-1}) &= \text{const} \times p(l_n \mid l_{n-1}).
\end{aligned}$$

The relations can also be expressed in recursive form as

$$\begin{aligned}
p(l_n \mid l_{n-1}) &= p(l_n \mid l_{n-1}) \\
p(l_i \mid l_{i-1}) &= \text{const} \times \frac{p(l_i \mid l_{i-1}, l_{i+1})}{p(l_{i+1} \mid l_i)}; i = n-1, n-2, \dots, 2 \\
p(l_1) &= \text{const} \times \frac{p(l_1 \mid l_2)}{p(l_2 \mid l_1)}.
\end{aligned}$$

The normalising constant can be efficiently calculated by summing over $l_i \in \mathbb{L}$. These relations provide the two algorithms:

- Direct: Markov RC to Markov RP Algorithm, see Algorithm 28. Given the Markov RC parametrisation, the computation of the Markov RP parametrisation becomes straightforward.
- Reverse: Markov RP to Markov RC Algorithm, see Algorithm 29. Given the Markov RP parametrisation, initiate by setting $p(l_n \mid l_{n-1})$ identical for the two parametrisations and calculate $p(l_i \mid l_{i-1})$ recursively for $i = n-1, n-2, \dots, 2$. Following this, $p(l_1)$ is easy to assess.

Algorithm 28: Direct: Markov random chain to Markov random profile

```

for all  $l_1, l_2 \in \mathbb{L}$  do
    Calculate  $p(l_1 | l_2) = \text{const} \times p(l_1)p(l_2 | l_1)$ 
    where const =  $\left[ \sum_{l'_1 \in \mathbb{L}} p(l'_1)p(l_2 | l'_1) \right]^{-1}$ 
end
for  $i = 2, 3, \dots, n - 1$  do
    for all  $l_{i-1}, l_i, l_{i+1} \in \mathbb{L}$  do
        Calculate  $p(l_i | l_{i-1}, l_{i+1}) = \text{const} \times p(l_{i+1} | l_i)p(l_i | l_{i-1})$ 
        where const =  $\left[ \sum_{l'_i \in \mathbb{L}} p(l_{i+1} | l'_i)p(l'_i | l_{i-1}) \right]^{-1}$ 
    end
end
for all  $l_n, l_{n-1} \in \mathbb{L}$  do
    Calculate  $p(l_n | l_{n-1}) = p(l_n | l_{n-1})$ 
end

```

Result: A Markov RP specification based on the specification of the Markov RC

Algorithm 29: Reverse: Markov random profile to Markov random chain

```

for all  $l_n, l_{n-1} \in \mathbb{L}$  do
    Set  $p(l_n | l_{n-1}) = p(l_n | l_{n-1})$ 
end
for  $i = n - 1, n - 2, \dots, 2$  do
    for all  $l_i, l_{i-1} \in \mathbb{L}$  and arbitrary  $l_{i+1} \in \mathbb{L}$  do
        Calculate  $p(l_i | l_{i-1}) = \text{const} \times \frac{p(l_i | l_{i-1}, l_{i+1})}{p(l_{i+1} | l_i)}$ 
        where const =  $\left[ \sum_{l'_i \in \mathbb{L}} \frac{p(l'_i | l_{i-1}, l_{i+1})}{p(l_{i+1} | l'_i)} \right]^{-1}$ 
    end
end
for  $l_1 \in \mathbb{L}$  and arbitrary  $l_2 \in \mathbb{L}$  do
    Calculate  $p(l_1) = \text{const} \times \frac{p(l_1 | l_2)}{p(l_2 | l_1)}$ 
    const =  $\left[ \sum_{l'_1 \in \mathbb{L}} \frac{p(l'_1 | l_2)}{p(l_2 | l'_1)} \right]^{-1}$ 
end

```

Result: A Markov RC specification given the specification of a Markov RP

10.9 Model Parameter Inference in Hierarchical RF Models

Model parameter inference in hierarchical RF models is not frequently used, possibly due to their computational complexity. The development of the posterior models for the parameters is not necessarily easily available in the literature. Therefore, the details of their developments are presented in this section.

Hierarchical Gaussian RF

Consider the discretised Gaussian RF prior model represented by the n -vector \mathbf{r} and let the model parameters $\mu_r \in \mathbb{R}$ and $\sigma_r^2 \in \mathbb{R}_+$ be random variables. The likelihood model is specified as Dirac-linear as defined in Sect. 5.1. The corresponding conditional variables can be defined as

$$[\mathbf{r} | \mu_r, \sigma_r^2] \sim p(\mathbf{r} | \mu_r, \sigma_r^2) = \phi_n(\mathbf{r}; \mu_r \mathbf{i}_n, \sigma_r^2 \Sigma_{r|\mu}^\rho)$$

$$[\mathbf{d}_e | \mathbf{r}] \sim p(\mathbf{d}_e | \mathbf{r}) = \delta_m(\mathbf{d}_e; \mathbf{H}\mathbf{r}),$$

where the correlation matrix $\Sigma_{r|\mu}^\rho$, of dimension $(n \times n)$, is defined by the spatial correlation function $\rho_r(\tau)$. The likelihood model is a Dirac pdf which represents observations without observation errors. The conditional model parameter $[\mu_r | \sigma_r^2]$ is assigned a Gaussian prior model. Thus, $p(\mu_r | \sigma_r^2)$ is Gaussian with hyper-parameters $\mu_\mu \in \mathbb{R}$ and $\gamma_\mu \in \mathbb{R}_+$,

$$[\mu_r | \sigma_r^2] \sim p(\mu_r | \sigma_r^2; \mu_\mu, \gamma_\mu) = \phi_1(\mu_r; \mu_\mu, \gamma_\mu \sigma_r^2).$$

The definitions above yield the joint conditional pdf of the random $(1 + n + m)$ -vector $[\mu_r, \mathbf{r}, \mathbf{d}_e | \sigma_r^2]$ as

$$\begin{aligned} & \begin{bmatrix} \mu_r \\ \mathbf{r} | \sigma_r^2 \\ \mathbf{d}_e \end{bmatrix} \sim p(\mu_r, \mathbf{r}, \mathbf{d}_e | \sigma_r^2) \\ &= \phi_{1+n+m} \left(\begin{bmatrix} \mu_r \\ \mathbf{r} \\ \mathbf{d}_e \end{bmatrix}; \begin{bmatrix} \mu_\mu \\ \mu_\mu \mathbf{i}_n \\ \mu_\mu \mathbf{H} \mathbf{i}_n \end{bmatrix}, \begin{bmatrix} \gamma_\mu \sigma_r^2 & \gamma_\mu \sigma_r^2 \mathbf{i}_n^T & \gamma_\mu \sigma_r^2 \mathbf{i}_n^T \mathbf{H}^T \\ \gamma_\mu \sigma_r^2 \mathbf{i}_n & \sigma_r^2 \Sigma_{r_H}^o & \sigma_r^2 \Sigma_{r_H}^o \mathbf{H}^T \\ \gamma_\mu \sigma_r^2 \mathbf{H} \mathbf{i}_n & \sigma_r^2 \mathbf{H} \Sigma_{r_H}^o & \sigma_r^2 \mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T \end{bmatrix} \right) \end{aligned}$$

with the spatial coupling matrix $\Sigma_{r_H}^o = \gamma_\mu \mathbf{i}_n \mathbf{i}_n^T + \Sigma_{r|\mu}^\rho$ of dimension $(n \times n)$.

The prior model for the random model parameter σ_r^2 is an inverse-gamma pdf with model hyper-parameters $v_{\sigma^2} \in \mathbb{R}_+$ and $\xi_{\sigma^2} \in \mathbb{R}_+$

$$\begin{aligned} \sigma_r^2 &\sim p(\sigma_r^2; v_{\sigma^2}, \xi_{\sigma^2}) = \left[\Gamma \left(\frac{v_{\sigma^2}}{2} \right) \right]^{-1} \left[\frac{v_{\sigma^2} \xi_{\sigma^2}^2}{2} \right]^{v_{\sigma^2}/2} \\ &\quad \times [\sigma_r^2]^{-(v_{\sigma^2}+2)/2} \exp \left(-\frac{v_{\sigma^2} \xi_{\sigma^2}^2}{2 \sigma_r^2} \right). \end{aligned}$$

Consider first the posterior model for the variance level σ_r^2 , expressed as

$$\begin{aligned} [\sigma_r^2 | \mathbf{d}_e] &\sim p(\sigma_r^2 | \mathbf{d}_e) = \text{const} \times p(\mathbf{d}_e | \sigma_r^2) p(\sigma_r^2) \\ &= \text{const} \times [2\pi]^{-m/2} [\sigma_r^2]^{-m/2} |\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T|^{-1/2} \end{aligned}$$

$$\begin{aligned}
& \times \exp\left(-\frac{1}{2\sigma_r^2}(\mathbf{d}_e - \mu_\mu \mathbf{H}\mathbf{i}_n)^T [\mathbf{H}\Sigma_{rH}^o \mathbf{H}^T]^{-1}(\mathbf{d}_e - \mu_\mu \mathbf{H}\mathbf{i}_n)\right) \\
& \times \left[\Gamma\left(\frac{\nu_{\sigma^2}}{2}\right)\right]^{-1} \left[\frac{\nu_{\sigma^2}\xi_{\sigma^2}^2}{2}\right]^{\nu_{\sigma^2}/2} [\sigma_r^2]^{-(\nu_{\sigma^2}+2)/2} \exp\left(-\frac{\nu_{\sigma^2}\xi_{\sigma^2}^2}{2\sigma_r^2}\right) \\
& = \text{const} \times [\sigma_r^2]^{-(\nu_{\sigma^2}+m+2)/2} \exp\left(-\frac{1}{2\sigma_r^2}[\nu_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)]\right),
\end{aligned}$$

where

$$\sigma(\mathbf{d}_e) = (\mathbf{d}_e - \mu_\mu \mathbf{H}\mathbf{i}_n)^T [\mathbf{H}\Sigma_{rH}^o \mathbf{H}^T]^{-1}(\mathbf{d}_e - \mu_\mu \mathbf{H}\mathbf{i}_n).$$

This expression has a shape function identical to the inverse-gamma pdf. Thus, the posterior of the hyper-parameters can be identified by setting

$$\begin{aligned}
\nu_{\sigma^2|d} &= \nu_{\sigma^2} + m \\
\nu_{\sigma^2|d}\xi_{\sigma^2|d}^2 &= \nu_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}),
\end{aligned}$$

which entails that

$$\begin{aligned}
\nu_{\sigma^2|d} &= \nu_{\sigma^2} + m \\
\xi_{\sigma^2|d}^2 &= \frac{\nu_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d})}{\nu_{\sigma^2} + m}.
\end{aligned}$$

Consequently, the posterior model for the model parameter $p(\sigma_r^2 | \mathbf{d}_e)$ is an inverse-gamma pdf with hyper-parameters $(\nu_{\sigma^2|d}, \xi_{\sigma^2|d})$. The natural estimator for $[\sigma_r^2|\mathbf{d}_e]$ with associated estimation variance is

$$\begin{aligned}
\hat{\sigma}_r^2 &= E\{\sigma_r^2|\mathbf{d}_e\} = \frac{\nu_{\sigma^2|d}\xi_{\sigma^2|d}^2}{\nu_{\sigma^2|d} - 2} = \frac{\nu_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)}{\nu_{\sigma^2} + m - 2} \\
\hat{\sigma}_{\hat{\sigma}}^2 &= \text{Var}\{\sigma_r^2|\mathbf{d}_e\} = \frac{[\nu_{\sigma^2|d}\xi_{\sigma^2|d}^2]^2}{[\nu_{\sigma^2|d} - 2]^2[\nu_{\sigma^2|d}/2 - 2]} = \frac{[\nu_{\sigma^2}\xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)]^2}{[\nu_{\sigma^2} + m - 2]^2[(\nu_{\sigma^2} + m)/2 - 2]}.
\end{aligned}$$

Note that if the grid increases by extension asymptotic analysis and the number of observations m increases, m and $\sigma(\mathbf{d}_e)$ increase proportionally and dominate the hyper-parameter values. Therefore, in the extension limit, the estimator coincides with the traditional ML estimator for σ_r^2 . Furthermore, the estimation variance approaches zero in the extension limit because m is cubic in the denominator, whereas $\sigma(\mathbf{d}_e)$ is only in quadratic in the numerator.

Next, consider the posterior model for the expectation level μ_r , expressed as

$$[\mu_r | \mathbf{d}_e] \sim p(\mu_r | \mathbf{d}_e) = \int_{\mathbb{R}_{\oplus}} p(\mu_r | \mathbf{d}_e, \sigma_r^2) p(\sigma_r^2 | \mathbf{d}_e) d\sigma_r^2.$$

The first factor is defined by the joint conditional pdf specified above,

$$[\mu_r | \mathbf{d}_e, \sigma_r^2] \sim p(\mu_r | \mathbf{d}_e, \sigma_r^2) = \phi_1(\mu_r; \mu_{\mu|d,\sigma^2}, \sigma_{\mu|d,\sigma^2}^2)$$

with

$$\begin{aligned} \mu_{\mu|d,\sigma^2} &= \mu_\mu + \gamma_\mu \sigma_r^2 \mathbf{i}_n^T \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ &= \mu_\mu + \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ \sigma_{\mu|d,\sigma^2}^2 &= \gamma_\mu \sigma_r^2 - \gamma_\mu \sigma_r^2 \mathbf{i}_n^T \mathbf{H}^T [\sigma_r^2 \mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} \gamma_\mu \sigma_r^2 \mathbf{H} \mathbf{i}_n \\ &= \gamma_\mu \sigma_r^2 [1 - \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} \mathbf{H} \mathbf{i}_n]. \end{aligned}$$

The posterior pdf for μ_r is obtained by multiplying the expressions for $p(\mu_r | \mathbf{d}_e, \sigma_r^2)$ and $p(\sigma_r^2 | \mathbf{d}_e)$ and marginalising with respect to σ_r^2 . It can be demonstrated that the posterior pdf $p(\mu_r | \mathbf{d}_e)$ is T-distributed because the former pdf is Gaussian with variance proportional to σ_r^2 and the latter is inverse-gamma. The estimator for $[\mu_r | \mathbf{d}_e]$ and the associated estimation variance can alternatively be developed by using the expressions for double expectation and total variance based on $[\mu_r | \mathbf{d}_e, \sigma_r^2]$:

$$\begin{aligned} \hat{\mu}_r &= E\{\mu_r | \mathbf{d}_e\} = E\{E\{\mu_r | \mathbf{d}_e, \sigma_r^2\}\} \\ &= \mu_\mu + \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} (\mathbf{d}_e - \mu_\mu \mathbf{H} \mathbf{i}_n) \\ &= [1 - \boldsymbol{\beta}^T \mathbf{H} \mathbf{i}_n] \mu_\mu + \boldsymbol{\beta}^T \mathbf{d}_e \\ \hat{\sigma}_{\mu_r}^2 &= \text{Var}\{\mu_r | \mathbf{d}_e\} = E\{\text{Var}\{\mu_r | \mathbf{d}_e, \sigma_r^2\}\} + \text{Var}\{E\{\mu_r | \mathbf{d}_e, \sigma_r^2\}\} \\ &= \gamma_\mu [1 - \gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1} \mathbf{H} \mathbf{i}_n] \times E\{\sigma_r^2 | \mathbf{d}_e\} \\ &= \gamma_\mu [1 - \boldsymbol{\beta}^T \mathbf{H} \mathbf{i}_n] \times \frac{\nu_{\sigma^2} \xi_{\sigma^2}^2 + \sigma(\mathbf{d}_e)}{\nu_{\sigma^2} + m - 2} \end{aligned}$$

with weight m -vector $\boldsymbol{\beta} = [\gamma_\mu \mathbf{i}_n^T \mathbf{H}^T [\mathbf{H} \Sigma_{r_H}^o \mathbf{H}^T]^{-1}]^T$. Observe that the estimator for the expectation is independent of σ_r^2 , as is typical for Gauss-linear/Gaussian models. Observe also that if the grid increases by extension asymptotic analysis, and the number of observations m increases, the term $\boldsymbol{\beta}^T \mathbf{H} \mathbf{i}_n$ tends towards 1.0. Thus, all weight in the estimator is assigned to the observations \mathbf{d}_e as in the traditional ML estimator for μ_r . Furthermore, the estimation variance tends to zero, as expected.

Hierarchical Poisson RF

Consider the discretised Poisson RF prior model represented by the n -vector \mathbf{k} and let the model parameter $\lambda_k \in \mathbb{R}_+$ be a random variable. Assume that the observations \mathbf{d} are collected according to a misclassification likelihood model. The corresponding conditional variables can be defined as

$$\begin{aligned} [\mathbf{k} | \lambda_k] &\sim p(\mathbf{k} | \lambda_k) = \prod_{i=1}^n \frac{(\lambda_k \Delta_n)^{k_i}}{k_i!} \times \exp(-\lambda_k \Delta_n) \\ [\mathbf{d} | \lambda_k] &\sim p(\mathbf{d} | \lambda_k) = \sum_{k=0}^{\infty} p(\mathbf{d} | \mathbf{k}) p(\mathbf{k} | \lambda_k) \\ &= \prod_{i=1}^n \frac{(\alpha_i \lambda_k \Delta_n)^{d_i}}{d_i!} \exp(-\alpha_i \lambda_k \Delta_n) \\ &= \prod_{i=1}^n \frac{\alpha_i^{d_i}}{d_i!} \times \lambda_k^{n_d} \Delta_n^{n_d} \exp(-\lambda_k \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n). \end{aligned}$$

The prior model for the random model parameter $\lambda_k \in \mathbb{R}_\oplus$ is a gamma pdf with model parameters $v_\lambda \in \mathbb{R}_+$ and $\xi_\lambda \in \mathbb{R}_+$,

$$\lambda_k \sim p(\lambda_k; v_\lambda, \xi_\lambda) = [\Gamma(v_\lambda)]^{-1} \xi_\lambda^{v_\lambda} \lambda_k^{v_\lambda - 1} \exp(-\xi_\lambda \lambda_k).$$

The posterior model for the model parameter λ_k is

$$\begin{aligned} [\lambda_k | \mathbf{d}] &\sim p(\lambda_k | \mathbf{d}) = \text{const} \times p(\mathbf{d} | \lambda_k) p(\lambda_k) \\ &= \text{const} \times \prod_{i=1}^n \frac{\alpha_i^{d_i}}{d_i!} \times \lambda_k^{n_d} \Delta_n^{n_d} \exp(-\lambda_k \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n) \\ &\quad \times [\Gamma(v_\lambda)]^{-1} \xi_\lambda^{v_\lambda} \lambda_k^{v_\lambda - 1} \exp(-\xi_\lambda \lambda_k) \\ &= \text{const} \times \lambda_k^{v_\lambda + n_d - 1} \exp(-(\xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n) \lambda_k) \end{aligned}$$

with $n_d = \sum_{i=1}^n d_i$. This expression has a shape function identical to the gamma pdf. Hence, the posterior of the hyper-parameters can be identified by setting

$$v_{\lambda|d} = v_\lambda + n_d$$

$$\xi_{\lambda|d} = \xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n.$$

Therefore, the posterior model for the parameter $p(\lambda_k \mid \mathbf{d})$ is a gamma pdf with hyper-parameters $(\nu_{\lambda|d}, \xi_{\lambda|d})$. The natural estimator for $[\lambda_k|\mathbf{d}]$ is then

$$\hat{\lambda}_k = E\{\lambda_k|\mathbf{d}\} = \frac{\nu_{\lambda|d}}{\xi_{\lambda|d}} = \frac{\nu_\lambda + n_d}{\xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n}$$

$$\sigma_{\hat{\lambda}}^2 = \text{Var}\{\lambda_k|\mathbf{d}\} = \frac{\nu_{\lambda|d}}{\xi_{\lambda|d}^2} = \frac{\nu_\lambda + n_d}{[\xi_\lambda + \Delta_n \boldsymbol{\alpha}^T \mathbf{i}_n]^2}.$$

Note that if the grid increases by extension asymptotic analysis and the number of observations n_d increases, n_d and $\boldsymbol{\alpha}^T \mathbf{i}_n$ increase proportionally and dominate the hyper-parameter values. Thus, in the extension limit, the estimator coincides with the traditional ML estimator for λ_k . Furthermore, the estimation variance tends to zero, as expected.

Chapter 11

Selected Applications



Statistical methods are valuable tools in various applications. Information is extracted from observation sets and predictions with quantified uncertainties improve decision making. In return, applications also enrich statistics. Each applied study is unique, and statisticians are challenged to tailor their models. These challenges spur statistical creativity. The critical factor in all research is creativity—methodological skills are secondary.

Applications of spatial statistics appear in many disciplines, and a small subset of publications of such studies is presented. The selected publications contain a discussion of the methodology based on observations of real spatial variables in various applications. Thus, they are not focused only on new methods. The objective of the presentation is to cast the methodologies in the terminology and notation of the book, not to discuss the conclusions of the studies. The readers should read the original publications and consider them as selected examples of thorough studies in spatial statistics.

11.1 Continuous Spatial Variables

These publications are considered to be readable case studies in spatial statistical modelling of continuous spatial variables.

Coverage Bias in the HadCRUT4 Temperature Series and Its Impact on Recent Temperature Trends

In Cowtan and Way (2014), the focus is on replacing missing observations in global spatio-temporal temperature records. Records from land-based weather stations and sea-based vessels and buoys are used in the study. Reliable long-term temperature records are a primary source of information in climate change studies.

The continuous spatial variable is the global monthly temperature average for a given year/month. The variable is spatially discretised to a grid with grid unit area ($5^\circ \times 5^\circ$), which entails varying grid unit area across the globe: Units are

larger near the equator than near the poles. The number of grid units is 2592, which entails a relatively rough spatial discretisation. The observations are primarily daily temperature registrations, and all registrations within a (month \times grid unit)-window are averaged and used as a single observation. The spatio-temporal coverage of the observations varies dramatically. The record starts in 1850 and has historically had coverage of about 53%, improving to about 70% after 1980. The study is based on the spatio-temporal record for 1980–2012. The spatial coverage is characterised by sparsity in the Arctic and Antarctic regions, whereas the density is high in developed countries. One spatial explanatory variable is included in one of the models: satellite-based temperature predictions for the Earth's surface. This variable is available after 1980 and has excellent spatial coverage. The precisions of the predicted temperatures are, however, poor, and they are spatio-temporally varying. The spatio-temporal observation set is typically subject to considerable study-specific data cleaning, the impact of which on the results is challenging to evaluate.

The likelihood model for the observed month-averaged temperatures, given the true month-averaged temperature spatial variable, is conditionally independent with a single-site response. The likelihood model in each observed grid unit is of the Dirac-point type. Thus, the observed values are assumed to be exact without observation error.

The prior model for the spatial month-averaged temperatures for a specific year/month is assigned a stationary, isotropic Gaussian RF model. The expectation and variance levels are estimated from the observations, as is the parameter in the exponential spatial correlation function. An alternative prior model, also a Gaussian RF model, is defined as the residuals between the observed and the satellite-predicted temperatures. The model parameters are assessed accordingly.

Given the observations, the posterior model for the spatial month-averaged temperatures for a specific year/month is defined by the Dirac-point likelihood and Gaussian prior models. These likelihood and prior models constitute a conjugate pair of models. Thus, the posterior model is a Gaussian model, which is analytically tractable. The relatively modest size of the grid entails that the required matrix inversion is computationally feasible. Consequently, the posterior non-stationary Gaussian model is fully specified. The posterior model based on the alternative prior model, with satellite-predicted temperatures as the spatial explanatory variable, is also a Gaussian model. Thus, this posterior model is also analytically tractable.

The resulting global spatio-temporal record of month-averaged temperatures is obtained during the period 1980–2012. Quantifications of uncertainties are calculated by cross-validation. This record is graphically displayed and thoroughly discussed. The effect of using a record with infilling is evaluated. Results obtained from the posterior models based on each of the two prior models, with and without the satellite predictions, are compared.

A Geostatistical Spatially Varying Coefficient Model for Mean Annual Runoff That Incorporates Process-Based Simulations and Short Records

In Roksvåg et al. (2022), the prediction of mean annual runoff in Norway is studied. Runoff is the part of precipitation that does not evaporate but flows across

land before it reaches a river. This flow is a physical process, which may be modelled within the catchments based on the snow melt, snow storage, evaporation, soil infiltration, etc. These hydrological predictions are important when designing infrastructure, evaluating water supply and managing hydropower plants.

The continuous spatial variable of interest is the mean annual runoff in an area. The variable is spatially discretised to 1 km^2 grid units covering the study area. The results, reported as the mean annual runoff from a partition of the study area into areal catchments, are obtained through the numerical integration of the grid predictions.

Each observation is the mean of the time series of annual runoffs for each gauged catchment across Norway. However, these time series are of two types: catchments with a complete historical record of annual runoff and catchments without a full set of observations, e.g. missing values for certain years. The latter catchments are pre-processed. Furthermore, one explanatory spatial variable is available from a deterministic hydrological, process-based spatial model capturing precipitation, temperature and land use effects on water runoff.

The likelihood model for the observations is defined with an areal reference for each catchment. The likelihood model is a conditionally independent single-site model for each catchment. Each observation model is of Gauss-point type, with the correct annual runoff as expectation and variance dependent on record length and the observed runoff. The uncertainty of the catchments with missing data is scaled relative to those catchments with a complete set of observations. In addition, the observation error variance is specified with a scale parameter defined to be random. Thus, the likelihood model is hierarchical.

The prior model for the mean annual runoff in grid units is assigned a spatially discretised Gaussian RF model. Mean annual runoff is a non-negative variable, whereas the Gaussian variable is not. However, because the runoff observations in the case study are far from zero, this mismatch does not cause problems. A hydrological process-generated map provides the explanatory spatial variable. This map is included in the expectation function as a spatial regression with spatially varying regression coefficients defined as a Gaussian RF model. Thus, the regression exhibits heterogeneity. The prior model also contains a residual Gaussian RF term. The model parameters are defined as random; hence, the prior model is hierarchical. The prior models for these parameters and the likelihood parameter are specified with no consideration of conjugate properties.

The hierarchical posterior model, defined by the Gauss-linear likelihood, Gaussian prior model and the prior for the model parameters, is not analytically tractable because the parameter prior pdfs are not conjugate. Given the model parameter values, the conditional posterior model is a Gaussian RF model for which an analytical solution is available. Thus, the spatial part of the posterior model, given the model parameters, can be assessed analytically. The study's hierarchical model is assessed using the INLA framework for computational efficiency.

A cross-validation criterion compares the resulting predictions of annual water runoff in the catchments to alternative prediction procedures. The prediction variances are more challenging to interpret because the areal averaging of the runoff

observations is not considered in the model. The current model is claimed to be superior to its alternative in this case study.

Joint AVO Inversion, Wavelet Estimation and Noise-Level Estimation Using a Spatially Coupled Hierarchical Bayesian Model

In Buland and Omre (2003), the focus is on mapping of the material properties, i.e. the pressure-wave velocity, shear-wave velocity and density in a cross-section of a sub-surface petroleum reservoir given amplitude versus offset (AVO) data. These material properties are strongly correlated with porosity and permeability characteristics, which can form the basis for decisions regarding the depletion of the reservoir.

The continuous spatial variable is trivariate in each grid unit, containing the log-transform of the three material properties. The variable is represented in a two-dimensional grid of size $160 \times 210 = 33,600$ in a vertical cross-section of the sub-surface reservoir.

The observations are of two types: well observations and seismic data. Two vertical wells lie in the cross-section, and the logarithm of the material properties is observed along the well traces. These observations are precise, but they have poor spatial coverage. Seismic data are collected in a grid on the surface. Thus, they have good spatial coverage but are very imprecise because a geophysical procedure indirectly acquires it. The seismic data are pre-processed such that one set of seismic data is assigned to each grid unit in the sub-surface.

The likelihood model for the well observations is conditional independent and single-site in the grid nodes penetrated by the wells. The model is of Gauss-point type with expectation equal to the logarithm of the material properties and error according to a material cross-covariance matrix. The scale of the latter matrix is considered to be random. The seismic likelihood model is more complex because the data are indirectly acquired. Given the logarithm of the material properties, the seismic data are modelled based on a vertical convolutional model together with a linearised wave equation and a Gaussian error term. Both the convolutional kernel and the error scale are defined as random. Thus, the seismic likelihood model is of hierarchical, Gauss-linear type.

The prior model for the trivariate spatial material logarithm of the properties is a stationary Gaussian RF model. The expectation function is horizontally constant and slowly vertically varying. The material cross-covariance matrix is estimated from the well observations, whereas the spatial correlation function is stationary and isotropic. Conjugate prior models for the random parameters of the likelihood model are assigned. The parameters of the error term in the seismic likelihood model are considered very influential for the posterior model, and alternative conjugate prior models for them are evaluated.

The posterior model, defined by the likelihood and prior models, is not analytically tractable. However, an iterative McMC/block Gibbs simulation algorithm assesses it. The posterior model for the spatial material logarithm of the properties, given the model parameters, the well observations and the seismic data, is an analytically tractable Gaussian pdf. Thus, this posterior Gaussian pdf can be used as

block updating in the McMC/Gibbs algorithm. The multivariate convolution kernel is also a Gaussian pdf given the remaining parameters, variables and observations. Thus, the convolution kernel can be block updated in the algorithm. The McMC algorithm exhibits favourable convergence rates and good mixing because of these block updates in the algorithm.

The McMC algorithm provides realisations from the posterior model for the joint model parameters and the logarithm of the spatial material property variable. The posterior pdf for the model parameters can be assessed from these realisations. The prediction maps for the material properties are constructed as the average of the realisations of the logarithms of the spatial material property variable after applying an inverse-log transform. The prediction variance maps are constructed accordingly.

A One-Step Bayesian Inversion Framework for 3D Reservoir Characterisation Based on a Gaussian Mixture Model: A Norwegian Sea Demonstration

In Fjeldstad et al. (2021), the focus is on the prediction of reservoir properties in a sub-surface petroleum reservoir. The continuous spatial variables of primary interest are the petrophysical ones, such as porosity, water saturation and clay proportion. Support spatial variables like the continuous seismic elastic attributes, such as acoustic impedance and Poisson ratio, and the mosaic lithology/fluid type, are included in the model to facilitate the use of established physical relations. Consequently, six spatial variables are jointly modelled. Reliable prediction of the spatial petrophysical variables with quantifications of uncertainty is important for efficient reservoir management.

A mixed continuous/mosaic spatial variable is defined for the six-variate reservoir variable. The spatial discretisation is specified to be $(98 \times 75 \times 100) = 735,000$. Thus, with a total of six spatial variables, the total number of grid units is $(6 \times 735,000) = 4,410,000$. The observations consist of a three-dimensional cube of prestack seismic data for two angles. Each angle-dependent dataset is assigned to each grid unit, and the data dimension is $(2 \times 735,000) = 1,470,000$. Consequently, the data have good spatial coverage, but the precision is low because it is indirectly acquired and pre-processed. The study is considered to be of very high dimension.

Given the six-variate spatial reservoir variable, the likelihood model for the seismic data is only dependent on the seismic elastic attributes because they are canonical variables for seismic sub-surface reflections. The likelihood model is a spatially coupled Gauss-linear model. The expectation function is a convolution of a linearised wave equation operating on the elastic attributes. The error covariance also exhibits spatial dependence. Thus, the likelihood model is spatially coupled with a multi-site response. All model parameters are assigned values based on experience.

The mixed continuous/mosaic prior model for the six-variate spatial reservoir variables is decomposed to facilitate the use of established rock physics and petrophysical models. The mosaic lithology/fluid type model has labels {sandstone/brine, sandstone/gas, shale}. Brine is a technical term for sub-surface water. The prior model is specified to be a Markov RF model with a six-nearest-node

neighbourhood design. The Markov RF model is expressed as a profile Markov RF model.

The conditional prior model for the continuous petrophysical variable, given the lithology/fluid type, is of logit-Gaussian type. The logit transform of the spatial petrophysical variable is assigned a spatially coupled Gaussian RF model. A grid unit's expectation and variance levels depend on the corresponding lithology/fluid type. A spatial correlation function defines the spatial coupling. The resulting marginal spatial prior model for the petrophysical variables is a Gaussian mixture RF model.

Last, given the petrophysical variable and lithology/fluid type, the conditional prior model for the continuous elastic attribute is of log-Gaussian type. The logarithmic transform of the elastic attributes is assigned a Gaussian RF model. The expectation function has a level dependent on the lithology/fluid type and a linear term proportional to the petrophysical variable. The variance and spatial coupling depend on the lithology/fluid type.

All parameters in the prior model are estimated from measurements in wells penetrating comparable reservoirs in the Norwegian Sea.

The likelihood and prior models define the joint posterior model for the six-variate spatial reservoir variable. The marginal posterior model for the lithology/fluid type is a Markov RF model, with a neighbourhood design much wider than the design for the prior model because the observations are spatially convolved. The marginal posterior models for the transformed petrophysical variable and elastic attribute are both Gaussian mixture RF models. From the perspective of the petrophysical spatial variable of primary interest, the posterior model is a hierarchical Gaussian RF model, where the lithology/fluid type is a parameter mosaic spatial variable assigned a Markov RF model.

The joint posterior model can only be assessed by simulation. An iterative McMC/Metropolis–Hastings algorithm inspired by the profile McMC/Gibbs algorithm is used. The algorithm's convergence rate is judged to be satisfactory, and a set of approximately independent realisations of the six-variate reservoir variable is generated from the posterior model.

The set of realisations from the posterior model is used to provide petrophysical variable predictions with associated quantifications of uncertainty. The results are compared to measurements made in one well penetrating the reservoir not used in the modelling. Moreover, comparisons between predictions made by the current three-dimensional model, with horizontal spatial coupling, and predictions based on a simple model consisting of several one-dimensional vertical models are performed.

11.2 Event Spatial Variables

These publications are considered to be readable case studies in spatial statistical modelling of event spatial variables.

A Bayesian Marked Spatial Point Processes Model for Basketball Shot Chart

In Jiao et al. (2021), the focus is on characterising the shot patterns of individual basketball players. Patterns, like shots beneath the basket and three-point corner shots, are evaluated for four dominating players during the 2017–2018 season. Moreover, this season’s 50 most influential players are categorised into five shot pattern classes. Understanding the shot patterns of individual players on the opposing team should impact a team’s defensive strategy in a game.

The event spatial variable is defined as the shot locations for a given player during the 2017–2018 basketball season. The mark attached to each shot is a binary label {no-score, score}. Observations are obtained from the shot charts from the 2017–2018 basketball season. These charts contain detailed information from each game during the season.

The likelihood model is, in the notation of the book, conditionally independent with a single-site response. It is a Dirac model because all shot locations with associated marks are exactly registered.

The prior model is a marked non-stationary Poisson event model. The non-stationary shot location intensity contains a baseline intensity and a regression on spatially discretised explanatory variables. Both the baseline intensity and the regression coefficients are specified to be random variables. Given the shot intensity, the binary scoring mark is modelled by a Bernoulli pdf. In this model, the scoring probabilities serve as the parameters. After a logistic transformation, the probability is defined as linear in the shot location intensity and a regression on spatial explanatory variables. All hyper-parameters are specified to be random variables. The spatial explanatory variables, which parametrise the shot location intensity, are defined by ten different shot types, including a shot beneath the basket and a three-point corner shot. These explanatory variables are spatially discretised to a $(50 \times 25) = 1250$ grid across the offensive half of the basketball court. Other explanatory variables like distance to the basket and the remaining time to the end of the period are also included. The hyper-parameters are assigned prior models. The baseline shot intensity is assigned a conjugate gamma pdf, whereas the other parameters are assigned fairly non-informative prior pdfs. With these specifications, the prior model is of hierarchical type.

The likelihood and prior models define the posterior model, which is specified as hierarchical. A general iterative McMC algorithm assesses the posterior model, and convergence and mixing rates are acceptable. The focus is on model parameter inference, and no predictions are performed. The ten shot types parametrise each player’s shot location intensity map, and the associated estimated regression coefficients characterise the individual player. The scoring location map is parametrised accordingly. The four basketball players discussed in detail have very different shot patterns. Lastly, the 50 basketball players with the highest shot intensity during the 2017–2018 season are categorised into five classes based on their shot patterns.

Spatial Modelling and Prediction of Loa loa Risk: Decision Making Under Uncertainty

In Diggle and Ribeiro (2007), the focus is on mapping the prevalence of Loa loa infection. Individuals infected by Loa loa experience severe, sometimes fatal, reactions to medication against parasite worms frequently used in tropical areas. Health authorities have therefore decided that in regions with Loa loa prevalence above 0.20, preventive measures shall be implemented.

The event spatial variable is defined to be individuals infected by Loa loa. The area under study is spatially discretised into a grid of size 550,000. Observations are only available in some of the grid units, detailing the number of inhabitants and the registered number of those infected with Loa loa. Furthermore, three spatial explanatory variables are available in all grid units: elevation above sea level, maximum vegetation index and temporal variability of vegetation index.

In the book's notation, the likelihood model is independent and single-site with a Dirac model in each grid unit. Thus, the number of infected individuals is assumed to be exactly registered in the grid units where the observations are available.

The prior model is double hierarchical. The spatially discretised event count RF model is in factorial form, with the model parameters in each grid unit being the number of inhabitants and the infection probability. The infection probability is defined to be a random spatial variable. Therefore, the number of infected individuals, given the infection probability, is assigned a binomial model. Recall that the binomial pdf tends towards the Poisson pdf as the number of inhabitants increases and the infection probability decreases, while their product remains constant and equal to the intensity. The logit transform of the infection probability, denoted the infection logit probability, is assigned a Gaussian RF model. The spatial expectation function contains regressions on the three explanatory variables, and a stationary, isotropic spatial covariance function specifies the covariance matrix. The model parameters in the regression and the covariance function are defined to be random, and they are assigned suitable prior pdfs. Thus, the prior model is double hierarchical.

The likelihood and prior models define the posterior model. The focus of the study is not on predicting the number of infected individuals but rather on estimating the spatial infection probability map, referred to as the prevalence map. Given the observations, the posterior model for the infection logit-probability map with associated parameters must be assessed by simulation-based inference. An iterative McMC simulation algorithm is used to provide realisations. The posterior model for the infection logit-probability map, given the model parameter values and the observations, is an analytically tractable Gaussian pdf. Thus, this Gaussian pdf is used as a one-block-proposal pdf in the McMC algorithm. The grid size is such that convergence of the algorithm is difficult to obtain. Consequently, the area is segmented into five units evaluated independently and pasted together. The McMC algorithm provides a set of realisations of the spatial infection logit-probability map from the posterior model.

After inverse-logit transformations, the average of these realisations provides the estimated infection probability map. Areas with expected infection probability

above 0.20 are of particular interest. A probability map provides the uncertainty quantification for exceeding the 0.20 infection probability level.

Integration of Presence-Only Data from Several Sources: A Case Study on Dolphins' Spatial Distribution

In Martino et al. (2021), the focus is on the estimation of the spatial abundance intensity of two dolphin species. Information about the number and spatial distribution of each species can guide authorities in their effort to protect these endangered dolphin species.

In theory, the event spatial variable is the location of individual dolphins at a certain point in time. This study focuses on the spatial intensity function of the dolphin locations. Thus, the study aims not at prediction but parameter inference. In practice, the dolphins swim and dive, and sampling is made over time. Therefore, several model simplifications must be assumed. Originally, the spatial event RF model's prior model was defined with a continuous intensity function. However, the conditioning on observations necessitates a spatial discretisation, which, in this study, is triangular.

The observations are of three types: targeted adaptive, distance and social-media-extracted. The two former types are traditional acquisition procedures frequently used in practice. The latter collection design is experimental in this study. Dolphins swim and dive, making observation challenging as they can easily be overlooked or counted twice. All observation schemes are expected to be biased and imprecise. Moreover, several explanatory spatial variables for the species intensities are available, such as distance from the coast, depth/slope of the sea floor and water temperature.

The three likelihood models for the observed dolphin locations are challenging to specify. They are also expected to be critical for the estimates. The models are defined as conditionally independent with single-site response, given the correct dolphin event population. Each observation likelihood is of misclassification type, causing the dolphins to be only partly observed. Targeted adaptive and distance samplings are well-established collection schemes with associated likelihood models. The model for social-media-extracted observations is not established, and several alternatives are explored and evaluated in the study. The sampling takes place within a fairly broad time window of varying length, and a correction factor for this effect is included in the likelihood models. Most parameters in the three different likelihood models are defined to be random. Thus, a hierarchical likelihood model is specified.

The prior model for the dolphin event spatial variable is a Poisson event location RF model, parametrised by a spatial intensity function, which is considered an RF. A log-Gaussian RF model with five explanatory spatial variables is assigned. The associated stationary and isotropic Gaussian RF model is specified with a Matérn order one covariance function and cast in an SPDE format. The associated hyper-parameters are considered random. Thus, the prior model is a hierarchical log-Gaussian Cox model. Last, all model parameters, in both the likelihood and prior models, are assigned fairly complex prior pdfs, none of which is of conjugate type.

The posterior model appears as a double hierarchical Poisson event RF model. The study is not cast in a predictive setting. Instead, the focus is on the estimation of the dolphin abundance spatial intensity function. The prior log-transformed intensity Gaussian RF model is defined within the SPDE format. The INLA framework provides approximate posterior pdfs for the model hyper-parameters and the intensity RF.

The resulting estimates of the dolphin intensity functions are presented as posterior median maps. The estimation uncertainties are quantified as maps of the interquartile range divided by the median.

11.3 Mosaic Spatial Variables

These publications are considered to be readable case studies in spatial statistical modelling of mosaic spatial variables.

Thin Cloud Detection of All-Sky Images Using Markov Random Fields

In Li et al. (2012), the focus is on the prediction of thin clouds in the sky on an otherwise clear sky day. Clouds are important for the atmospheric energy balance, and thin clouds are notoriously difficult to identify and outline visually.

The mosaic spatial variable is the cloud coverage with two labels: thin cloud and no cloud. The all-sky image is spatially discretised to a grid of dimension $(200 \times 200) = 40,000$ grid units.

The observations are acquired by a sky imager unit, which provides an all-sky image. The image is processed to provide three features: physical, visual and statistical. The physical feature is the colour spectrum's normalised blue/red intensities. The visual feature is based on the saturation intensity of the spectrum. In contrast, the statistical feature represents the distance to the diagonal of the spectre's red/green/blue intensity cube. All these features are specified based on experience in thin cloud detection. Moreover, thin clouds cannot be reliably detected by simply thresholding any one of the features.

The likelihood model for the trivariate observations collected in each grid unit is conditionally independent with a single-site response. The model is trivariate Gaussian in each grid unit, assuming independence between the observed features. The model parameters, being the expectation and variance levels, are unknown and are later estimated from the observations.

The prior model for the cloud-type mosaic spatial variable is defined to be a Markov RF model. The Markov formulation of the model is based on a neighbourhood design of the eight nearest nodes. The single model parameter is heuristically determined.

The posterior model appears as a Markov RF model with the same neighbourhood design as the prior model. An iterative McMC algorithm can assess this posterior model. However, this study focuses on the MAP prediction of cloud type. The inference of the likelihood parameters and the MAP prediction are performed

iteratively. The former inference is simple, whereas the prediction is only made approximately with an iterative conditional algorithm.

The resulting MAP prediction is defined as the thin cloud prediction image. No uncertainty quantifications are provided because no simulations are made, only optimisations.

Markov Random Field Segmentation of Brain MR Images

In Held et al. (1997), the focus is on the segmentation of three-dimensional magnetic resonance (MR) images into five different brain tissue classes. The objective is predictive, to provide a reliable brain image prediction for an individual. The medical diagnosis provided by a medical doctor will most likely be improved by having a reliable three-dimensional image of the brain available to an individual.

The mosaic spatial variable is brain tissues with five labels: grey matter, white matter, cerebrospinal fluid, scalp bone and background. The three-dimensional brain image is spatially discretised into a cubic grid volumes of size 1 mm^3 , with $n = (256 \times 160 \times 32) = 1,310,720$ grid nodes. The observations come from an MR image acquired in the same grid design as specified above. The signals are double echo MR signals. Thus, the observations are bivariate in each grid unit. The signal strength is spatially inhomogeneous due to a spatially varying magnetic field, and this feature is considered to be important in the analysis. A nuisance continuous spatial variable representing signal strength is introduced to account for this inhomogeneity.

The likelihood model is specified to be conditionally independent with a single-site response. In each grid unit, the bivariate MR signal, given the brain tissue label and the signal strength, is defined by a non-parametric pdf. A kernel estimator obtains this pdf based on a training set of tissue observations with associated MR signals.

The prior models, for the brain tissue and signal strength spatial variables, are both of Markov RF type. The brain tissues are assigned a mosaic Markov RF model with the five tissue types as labels. The spatial coupling parameter in the model accounts for the medical experiences that certain tissue types do not appear adjacent to each other. The signal strength is assigned a continuous Gaussian Markov RF model. Both Markov RF models are defined with six-nearest-node neighbourhoods.

The posterior model is defined by the likelihood and prior models, which constitute a conjugate pair of models. Thus, the posterior model is also a Markov RF with a six-nearest-node neighbourhood. The focus in the study is only on prediction; no quantifications of uncertainties are provided. Thus, the MAP criterion is used, and the predictions are obtained by maximisation of the posterior model. The optimisation is performed using a simulated annealing technique and an iterative conditional mode algorithm.

The methodology is tested on both simulated MR images and real brain MR images from individuals.

Hierarchical Bayesian Lithology/Fluid Prediction: A North Sea Case Study

In Rimstad et al. (2012), the focus is on mapping the lithology/fluid characteristics of a sub-surface three-dimensional petroleum reservoir. Reliable mapping of the oil

and gas zones in the reservoir is important for designing an efficient depletion plan for the resource.

A mosaic spatial variable is defined with the four labels {sand-gas, sand-oil, sand-brine, shale}. Brine is a technical term for sub-surface water. The variable is represented on a sub-surface three-dimensional grid of dimension $n = (76 \times 128 \times 91) = 885,248$. The observations are of two types: well observations and seismic data. A vertical well is drilled through the sub-surface reservoir, and the lithology/fluid characteristics are monitored along the well trace. These observations are very precise, but they appear with poor spatial coverage. The seismic data are collected in a grid on the surface. Thus, they have good spatial coverage but are highly imprecise because a geophysical procedure indirectly acquires the data. The seismic data are pre-processed to have seismic data for each of the two angles in each three-dimensional sub-surface grid unit.

The likelihood model for the well observations is conditionally independent, with a single-site response in each grid unit penetrated by the well trace. The observations are assumed to be without error. Thus, they are of Dirac type. The seismic likelihood model is much more complicated because the data are indirectly acquired. The seismic reflection coefficients in each grid unit are linked to the lithology/fluid variables by established rock physics relations, with certain model parameters defined to be random. Given the reflection coefficients in all grid units, the seismic gathers are modelled as a vertical convolutional model based on a linearised wave equation with Gaussian error terms. This likelihood model is defined with random parameters. Thus, the seismic likelihood model is hierarchical and convolutional with a multi-site response.

The prior model for the lithology/fluid labels is a Markov RF model. The model is formulated as a profile Markov RF, with a four-nearest-node neighbourhood design horizontally. The vertical dimension, in which the seismic convolution occurs, is phrased as a Markov RC, and the transition probability matrix is defined to control the label proportions and ensure the fluid's gravitational sorting. Thus, the matrix entries for the upward transition from gas to brine or oil and that from oil to brine are set to zero. These hard transition constraints in the Markov RF prior model are crucial for obtaining realistic realisations from the posterior model. Prior models for the random model parameters are also assigned.

The posterior model, defined by the likelihood and prior models, is also a Markov RF model with considerably larger neighbourhood designs than the Markov prior model. The posterior model's realisations are generated by an iterative simulation algorithm inspired by the profile McMC/Gibbs algorithm. The convergence rate and mixing are surprisingly favourable, considering the dimension and complexity of the model.

The lithology/fluid predictions of the sub-surface reservoir are based on a probabilistic criterion. Sand-gas and sand-oil are predicted in a grid unit if their respective probabilities are estimated to be above 0.5. Otherwise, the grid unit is left unpredicted. The prediction uncertainties are represented by the probability maps for the four labels. The reliability of the predictions is evaluated by cross-validation of observations in five wells not used in the modelling. Four alternative likelihood model formulations are used, and the results, which are very different, are compared.

Chapter 12

Projects and Exercises



Whether you want to become an excellent sport athlete, artist or scientific researcher, the way to excel in any field is by practising the requisite skills through targeted drills. Project and exercise work on basic topics in your scientific field provide such targeted drills for future researchers.

This book is based on a revised compendium for a university course in graduate-level spatial statistics. The course is module-based:

- Four weeks dedicated to lectures covering the topic of continuous RFs
- Two weeks of project work on continuous RFs
- Two weeks dedicated to lectures covering the topic of event RFs
- Two weeks of project work on event RFs
- Two weeks dedicated to lectures covering the topic of mosaic RFs
- Two weeks of project work on mosaic RFs

The final grade for students is based on a revised version of one of the project papers (40%) and a written exam (60%). The following sections contain the three project texts and several exercise sets inspired by the exam problems in the course. Supplementary information, including datasets and solution sketches, is available online.

12.1 Projects in Spatial Modelling

The projects are completed by the students in groups of two. A written report is to be delivered after two weeks of work. After one week, the project paper is returned to the students with comments and suggestions for improvements. The student must

Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1007/978-3-031-65418-3_12).

have all three project papers accepted to be eligible for the exam. Each student must deliver a revised version of one of their project papers on the day of the exam. The evaluation is based not only on statistical quality but also on clarity in notation and writing and the aesthetics of the figures. The project texts are presented in the following subsections.

12.1.1 Project Text: Continuous Spatial Variables

Introduction

This assignment contains problems related to continuous spatial variables and Gaussian RFs. We recommend using R to solve these problems, and you can find relevant functions in the R libraries `geoR`, `MASS`, `spatial`, `akima` and `fields`.

The three problems consider the following dataset, which is available as supplementary material online:

- Terrain elevation data, available as `topo` (Problem 2)

Problem 1: Gaussian RF: Model Characteristics

Consider the continuous spatial variable $\{r(x) : x \in D : [1, 50] \subset R^1\}$, and assume that it is modelled as a stationary one-dimensional Gaussian RF with the following model parameters:

$$\begin{aligned} E\{r(x)\} &= \mu_r = 0 \\ \text{Var}\{r(x)\} &= \sigma_r^2 \\ \text{Corr}\{r(x), r(x')\} &= \rho_r(\tau), \end{aligned}$$

where $\rho_r(\tau); \tau = |x - x'|/10$ is the spatial correlation function. Let $D : [1, 50]$ be discretised into $L \in \{1, 2, \dots, 50\}$ and define the discretised Gaussian RF $\{r(x); x \in L\}$.

Let the spatial correlation function, $\rho_r(\tau)$, be either powered exponential with parameter $v_r \in \{1, 1.9\}$ or Matérn with parameter $v_r \in \{1, 3\}$. Let the variance take values $\sigma_r^2 \in \{1, 5\}$.

- (a) The spatial correlation function must be a non-negative definite function. Specify this requirement mathematically and explain why it is necessary.

Display the two spatial correlation functions for $\tau \in \mathbb{R}_+$ with the model parameters specified above. Discuss the features of the spatial correlation function that are important for a realisation of the Gaussian RF. Also, discuss the relationship between the correlation and variogram functions. Display the associated variogram functions $\gamma_r(\tau)$.

The functions `cov.spatial()` and `Matern()` in `geoR` package may be helpful.

- (b) Define the pdf for the corresponding Gaussian model and utilise it as the prior model.

Simulate four realisations of the Gaussian RF on L for each of the eight sets of model parameters defined above and present them in eight displays, with four realisations in each.

Discuss the relationship between the realisations and the model parameters.

Let the spatial variable be observed as $\{d(x); x \in \{10, 25, 30\} \subset L\}$ according to the acquisition model

$$d(x) = r(x) + e(x); x \in \{10, 25, 30\}$$

where the measurement errors $e(\cdot)$ are independent and identically distributed according to centred Gaussian pdfs with variance σ_e^2 . Furthermore, assume that $r(x)$ and $e(x')$ are independent for all x, x' .

(c) Specify the expression for the corresponding likelihood model.

Consider the simulated realisations in (b) with $\sigma_r^2 = 5$ and select one realisation. Use the values at $x \in \{10, 25, 30\}$ in this realisation as the observed values $\{d(x); x \in \{10, 25, 30\} \subset L\}$. Let the observation error variance take on the values $\sigma_e^2 \in \{0, 0.25\}$.

(d) Specify the pdf for the discretised Gaussian RF posterior model given the observations.

Use the prior model, the likelihood model with the two error variances listed above and the observed values. Compute the corresponding predictions for the spatial variable $\{\hat{r}(x); x \in L\}$ along with their associated 0.9 prediction intervals and present the results in two displays.

Discuss these two displays and the relationship between the model parameters and the predictions, including the prediction intervals. Inspect carefully the appearance of the predictions at the observation locations.

(e) Simulate 100 realisations from each of the two discretised Gaussian RF posterior models and empirically estimate the prediction along with their associated 0.9 prediction intervals based on these realisations for each model.

Present the simulated realisations in two displays, one for each model, and overlay the corresponding empirically estimated predictions and prediction intervals in each display.

Discuss the relationship between the model parameters and the realisations, as well as the relationship between the analytically and empirically obtained predictions with prediction intervals.

Consider the non-linear function on $\{r(x); x \in D\}$,

$$A_r = \sum_{x \in L} I(r(x) > 2) (r(x) - 2),$$

which approximates the area below the spatial variable and above the level 2.

- (f) Use the 100 realisations from the posterior model with $\sigma_e^2 = 0$ to provide a prediction \hat{A}_r and its associated prediction variance.

An alternative predictor for this area is based on the predicted spatial variable with $\sigma_e^2 = 0$, denoted as $\{\hat{r}(x); x \in L\}$,

$$\tilde{A}_r = \sum_{x \in L} I(\hat{r}(x) > 2)(\hat{r}(x) - 2).$$

Calculate this prediction.

Consider the two predictions and the prediction variance of the former. Compare the predictions and use Jensen's inequality to explain why one would expect $\hat{A}_r \geq \tilde{A}_r$.

- (g) Present a summary of the experiences you have had while evaluating the model's characteristics.

Problem 2: Gaussian RF: Real Data

Consider observations of terrain elevation [m] in the **topo** dataset. The 52 observations are located in the domain $D = [(0, 315) \times (0, 315)] \subset \mathbb{R}^2$. Let the 52-vector of exact observations by $\mathbf{d} = (r(\mathbf{x}_1^d), \dots, r(\mathbf{x}_{52}^d))^T$.

- (a) Display the observations in various ways. Is a stationary Gaussian RF a suitable model for the terrain elevation in domain D?

The functions `interp()`, `contour()` and `image.plot()` in the R libraries `akima`, `graphics` and `fields` may be useful.

Let the terrain elevation in domain D be modelled by the Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$ with the following parameters:

$$\begin{aligned} E\{r(\mathbf{x})\} &= \mathbf{g}(\mathbf{x})^T \boldsymbol{\beta}_r \\ \text{Var}\{r(\mathbf{x})\} &= \sigma_r^2 \\ \text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} &= \rho_r(\tau; \psi). \end{aligned}$$

Here, $\mathbf{g}(\mathbf{x}) = (1, g_2(\mathbf{x}), \dots, g_{n_g}(\mathbf{x}))^T$ is an n_g -vector of known explanatory spatial variables on $\mathbf{x} \in D$, and $\boldsymbol{\beta}_r = (\beta_1, \dots, \beta_{n_g})^T$ is an n_g -vector of unknown parameters. Moreover, let the variance be $\sigma_r^2 = 2500$ and the spatial correlation function be $\rho_r(\tau) = \exp(-(0.01\tau)^{1.5})$ with $\tau = |\mathbf{x} - \mathbf{x}'|$.

- (b) Develop the expression for the minimisation problem to be solved for the universal Kriging predictor and the associated prediction variance at an arbitrary location $\mathbf{x}_o \in D$. The actual optimisation problem need not be solved.

Is it necessary to adjust the variance value if the parametrisation of the expectation function is modified?

Consider the case with $E\{r(\mathbf{x})\} = \beta_1$, known as the ordinary Kriging model. Discretise the Gaussian RF to $\{r(\mathbf{x}); \mathbf{x} \in L\}$ using the grid $L : [(0, 315) \times (0, 315)] \in D$.

- (c) Calculate the universal Kriging predictor along with its associated prediction variance, $\{\hat{r}(\mathbf{x}); \mathbf{x} \in L\}$ and $\{\sigma_{\hat{r}}^2(\mathbf{x}); \mathbf{x} \in L\}$. Display the results and provide comments on them.

Use the function `krige.conv()` and the arguments `trend.d` and `trend.l` in `krige.control()` to specify the parametric form of the expectation function. The function `expand.grid()` may also be helpful.

Let the reference variable $\mathbf{x} \in D \subset \mathbb{R}^2$ be denoted $\mathbf{x} = (x_v, x_h)$, set $n_g = 6$ and define the set of known polynomial functions $\mathbf{g}(\mathbf{x})$ to be all polynomials $x_v^k x_h^l$ for $(k, l) \in \{(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (0, 2)\}$. Discretise the Gaussian RF to $\{r(\mathbf{x}); \mathbf{x} \in L\}$ with the grid $L : [(0, 315) \times (0, 315)] \in D$.

- (d) Specify the resulting n_g -vector $\mathbf{g}(\mathbf{x})$ and the expected value of $r(\mathbf{x})$.

Calculate the universal Kriging predictor along with its associated prediction variance, $\{\hat{r}(\mathbf{x}); \mathbf{x} \in L\}$ and $\{\sigma_{\hat{r}}^2(\mathbf{x}); \mathbf{x} \in L\}$. Display the results and provide comments on them.

Use the function `krige.conv()` and the arguments `trend.d` and `trend.l` in `krige.control()` to specify the parametric form of the expectation function. The function `expand.grid()` may also be helpful.

Use the ordinary Kriging predictor's associated prediction variance and consider the grid node $\mathbf{x}_0 = (100, 100)$.

- (e) Calculate the probability for the elevation to be higher than 850 m at this location. Further, calculate the elevation at which the probability is 0.90 that the true elevation is below it.
(f) Present a summary of the experiences you had while encountering the evaluation of the real data.

Problem 3: Parameter Estimation

Consider the stationary Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$ with $D : [(1, 30) \times (1, 30)]$, where

$$\begin{aligned} E\{r(\mathbf{x})\} &= \mu_r = 0 \\ \text{Var}\{r(\mathbf{x})\} &= \sigma_r^2 \\ \text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} &= \exp(-\tau/\psi_r) \end{aligned}$$

and $\tau = |\mathbf{x} - \mathbf{x}'|$. Consider the discretised Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in L\}$ on a grid $L : [(1, 30) \times (1, 30)] \in D$. Set the model parameters $\sigma_r^2 = 2$ and $\psi_r = 3$.

- (a) Generate one realisation of the discretised Gaussian RF and display it.
(b) Compute the empirical variogram based on exact observations of the full realisation and display the estimate alongside the correct variogram function.

Comment on the result, particularly regarding the precision of the estimates within the finite domain D.

Use the function `variog()` in `geoR`.

- (c) Repeat points (a) and (b) three times, and then comment on the results.

Generate 36 locations uniformly randomly in the grid L.

- (d) Compute the empirical variogram estimate based on the corresponding 36 exact observations. Display the estimate with the theoretical variogram function and comment on the results.

Consider the model parameters variance σ_r^2 and ψ_r to be unknown.

- (e) Estimate the parameters using a maximum likelihood criterion based on exact observation of the full realisation and based on the 36 exact observations. Display the corresponding estimated variogram functions jointly with the correct variogram function and comment on the result.

Discuss the results.

The function `likfit(.)` in `geoR` may be helpful.

- (f) Repeat the procedure from (d) with 9, 64 and 100 uniformly generated exact observations from the realisation.

Present the estimates jointly with the correct variogram function in separate displays and comment on the results.

- (g) Present a summary of your experiences during parameter estimation.

12.1.2 Project Text: Event Spatial Variables

Introduction

This assignment contains problems related to event spatial variables and Poisson RFs. The programming language R can be used to solve the problems, and you can find relevant functions in the R library `spatial`. The four problems consider the following datasets, which are available as supplementary material online:

- Biological cell data, available as `cells` (Problems 1 and 4)
- Redwood tree data, available as `redwood` (Problems 1 and 3)
- Pine tree data, available as `pines` (Problem 1)

and two datasets related to the pine tree dataset:

- Observation probability map, available as `obspprob` (Problem 2)
- Observed pine tree count map, available as `obspines` (Problem 2)

Problem 1: Analysis of Point Patterns

Consider the above-defined biological cell, redwood tree and pine tree dataset point patterns.

- (a) Display each point pattern and discuss their appearances. Try to relate the point patterns to real processes in nature that could explain their behaviour.

- (b) Compute the empirical L - or J -function for each point pattern, using the `Kfn()` function.
 Display the functions for each point pattern and discuss their appearance.
 Specify the expression for the theoretical L - and J -functions for a stationary Poisson RF.
 Display the empirical L - or J -function for each point pattern along with the corresponding theoretical function for a stationary Poisson RF. Discuss whether a stationary Poisson RF is a suitable model for each point pattern.
- (c) Perform an empirical Monte Carlo test to determine whether the stationary Poisson RF is a suitable model.
 Consider each point pattern under the hypothesis that the point pattern originates from a stationary Poisson RF, conditional on the point count observed. Generate 100 realisations of the Poisson RF for each point pattern. For each realisation, compute the associated L - or J -function and use the set of functions to empirically test whether the actual point pattern could originate from a Poisson RF.
 Display the empirical 0.9-intervals for the functions alongside the corresponding estimated functions for each point pattern. Discuss whether a Poisson RF is a suitable model for each point pattern.

Problem 2: Bayesian Inversion in Poisson RF

Consider an area of size (300×300) m² containing a pine tree forest, where the exact locations of the pine trees are to be assessed. The pine tree locations are observed from a satellite by remote sensing, and due to partly cloudy weather, the observation probability for individual trees varies across the area.

Consider a discretisation of the area into a regular $([1, 30] \times [1, 30])$ -grid L with each grid unit size being 100 m². The true but unknown number of pine trees located in each grid unit is $\{k(\mathbf{x}); \mathbf{x} \in L\}$.

The probability of observing a pine tree occurring in a given grid unit is represented by $\{\alpha(\mathbf{x}); \mathbf{x} \in L\}$, and this probability varies across the area. The probabilities are provided in `obsprob`, together with the x - and y -coordinates of the centroids of the grid units. The number of pine trees observed in each grid unit is represented by $\{d(\mathbf{x}); \mathbf{x} \in L\}$, and these numbers are listed in `obspines` in a format similar to the probabilities above.

- (a) Display the observations and the observation probabilities. Assume that the observations in the grid units, given the true number of pine trees in each grid unit, are spatially uncorrelated. Specify the expression for the corresponding likelihood model for the observations.
- (b) Assume a priori that the distribution of pine trees is according to a stationary Poisson RF with model parameter λ_k . Specify the expression for the corresponding prior model for the discretised Poisson count model.
- (c) Estimate the intensity λ_k based on the observations in the grid units with associated observation probabilities.

Generate six realisations from this prior stationary Poisson event count model and the associated approximate Poisson event-location realisations. Display the approximate Poisson event-location realisations.

- (d) Develop the expression for the posterior discretised event count model and justify that this posterior model is a discretised Poisson RF model. Generate six realisations of the associated approximate event-location model and display these realisations.

Discuss the similarities and differences between the prior and posterior approximate event-location realisations.

- (e) Simulate 100 realisations of the discretised event count model, both for the prior and the posterior models.

Compute the average of these 100 realisations, representing the expected event-count number, for each of the two models. Display these averages graphically, compare the displays and explain the differences.

Problem 3: Clustered Event Spatial Variables

Consider the redwood tree data listed above. Consider the Neyman–Scott (mother-child) cluster model for spatial event variables, with a Poisson mother intensity model and Gaussian child intensity model.

- (a) Describe the model and specify the full set of model parameters. Discuss potential boundary problems caused by using a finite domain D when simulating realisations from the model. Suggest solutions to overcome these problems.

- (b) Make an empirical fit of the model parameters to the redwood tree data. This fit must be done only by inspecting the tree pattern and estimating the model parameter values from your intuitive understanding of their impact on the pattern. Evaluate your parameter values by a Monte Carlo test on the L - or J -interaction function.

- (c) Iterate your guesstimation procedure to improve the fit and make the Monte Carlo test appear significant. List and justify the final model parameter guesstimates by displaying the Monte Carlo test results.

Discuss the results.

- (d) Display the redwood tree dataset next to three realisations from the guesstimated Neyman–Scott model.

Comment on the display.

Problem 4: Repulsive Event Spatial Variables

Consider the biological cell data listed above.

- (a) Specify the expression for the Strauss repulsion model for spatial event variables with exponential interaction function and fixed event count. Describe the full set of model parameters. Discuss potential boundary problems caused by using a finite domain D when simulating realisations from the model. Suggest solutions to overcome these problems.

- (b) Make an empirical fit of the model parameters to the biological cell data. This fit must be done only by inspecting the cell pattern and estimating the

model parameter values from your intuitive understanding of their impact on the pattern.

Evaluate your parameter values by a Monte Carlo test on the L - or J -interaction function.

- (c) Iterate your guesstimation procedure to improve the fit and make the Monte Carlo test appear significant. List and justify the final model parameter guesstimates and justify them by displaying the Monte Carlo test results.
Discuss the results.
- (d) Display the biological cell dataset next to three realisations from the guesstimation Strauss repulsion model.
Comment on the display.

12.1.3 Project Text: Mosaic Spatial Variables

Introduction

This assignment contains a problem set related to mosaic spatial variables or, more specifically, Markov RFs.

The problem set considers the following dataset, which is available as supplementary material online:

- Seismic reflection data, available as **seismic** (Problem 1)
and a dataset related to the seismic dataset:
- Lithology map, available as **complit** (Problem 1)

Problem 1: Markov RF

This problem is based on observations of seismic data over a domain $D \subset \mathbb{R}^2$. The objective is to identify the underlying {sand, shale} lithology distribution over D , represented by $\{0, 1\}$, respectively. The observations are collected on a regular (75×75) grid $L \in D$, and the seismic data are denoted $\{d(\mathbf{x}); \mathbf{x} \in L\}; d(\mathbf{x}) \in \mathbb{R}$, represented by the n -vector \mathbf{d} . The observations can be found in the dataset **seismic**.

Moreover, observations of the lithology distribution {sand, shale} in a geologically comparable domain $D_c \subset \mathbb{R}^2$ are available. The lithology distribution is collected on a regular (66×66) grid L_c , with the same spacing as L , over D . The observations, coded as $\{0, 1\}$ for {sand, shale}, are available in the dataset **complit**.

Assume that the underlying distribution of lithologies can be represented by a Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L\}; l(\mathbf{x}) \in \{0, 1\}$ represented by the n -vector \mathbf{l} .

The seismic data collection procedure defines the likelihood model:

$$[d_i | \mathbf{l}] = \begin{cases} 0.02 + e_i & \text{if } l_i = 0 \text{ (sand)} \\ 0.08 + e_i & \text{if } l_i = 1 \text{ (shale)} \end{cases} ; i = 1, 2, \dots, n$$

with $e_i; i = 1, 2, \dots, n$ being independent centred Gaussian with variance $\sigma_e^2 = 0.06^2$.

- (a) Specify the expression for the likelihood model $p(\mathbf{d} | \mathbf{l})$.

Display the observations $\{d(\mathbf{x}); \mathbf{x} \in L\}$ as a map.

Consider an independent, uniform prior model on \mathbf{l} , i.e. $p(\mathbf{l}) = \text{const}$.

- (b) Develop an expression for the posterior model $p(\mathbf{l} | \mathbf{d})$.

Simulate six realisations of the posterior Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L | \mathbf{d}\}$ and display them as maps.

Develop expressions for the posterior expectation $E\{\mathbf{l} | \mathbf{d}\}$ and the posterior variance $\text{Var}\{\mathbf{l} | \mathbf{d}\}$, and display them as maps.

Develop expressions for the maximum marginal posterior predictor $\text{MMAP}\{\mathbf{l} | \mathbf{d}\}$, and display the result as a map.

Comment on the results.

Consider a Markov RF prior model for $\{l(\mathbf{x}); \mathbf{x} \in L\}$, represented by the n -vector \mathbf{l} , with a clique system \mathbf{c}_L consisting of all two-nearest-nodes on the grid L . The corresponding Gibbs formulation is as follows:

$$p(\mathbf{l}) = \text{const} \times \prod_{\mathbf{c} \in \mathbf{c}_L} v_{\mathbf{l}|I}(l_i; i \in \mathbf{c}) = \text{const} \times \prod_{<i,j> \in L} \beta^{I(l_i=l_j)}.$$

Here, $< i, j > \in L$ defines the set of two-nearest-nodes on the grid L , the parameter $\beta \in \mathbb{R}_{[1, \infty)}$, and $I(A)$ equals 1 if A is true and 0 otherwise.

- (c) Specify the associated Markov formulation for the Markov RF.

Develop expressions for the posterior models $p(\mathbf{l} | \mathbf{d})$ and $p(l_i | \mathbf{d}, \mathbf{l}_{-i}); i = 1, 2, \dots, n$.

Display the observations from the geologically comparable domain D_c as a map. Use these observations to estimate β by a maximum pseudo-likelihood procedure. Explain the procedure, and denote the estimate $\hat{\beta}$.

Next, focus on realisations from $p(\mathbf{l} | \mathbf{d})$ with related prediction $E\{\mathbf{l} | \mathbf{d}\}$, variance $\text{Var}\{\mathbf{l} | \mathbf{d}\}$ and alternative prediction $\text{MMAP}\{\mathbf{l} | \mathbf{d}\}$.

- (d) Set the model parameter $\beta = \hat{\beta}$ and use an McMC/Gibbs algorithm to generate realisations from the posterior model $p(\mathbf{l} | \mathbf{d})$. Specify the McMC procedure in an algorithmic format. Use torus/wrapping boundary conditions to avoid border problems. One simulation sweep corresponds to one visit per node in expectation. Consider carefully the number of sweeps required to obtain approximate convergence. Document that the algorithm has approximately converged by displaying convergence plots of sand proportion as a convergence indicator, and explain how approximately independent realisations can be obtained.

Display six approximately independent realisations as maps.

Comment on the results.

- (e) Estimate the predictor $E\{\mathbf{l} | \mathbf{d}\}$, the prediction variances in the diagonal terms of $\text{Var}\{\mathbf{l} | \mathbf{d}\}$ and the alternative predictor $\text{MMAP}\{\mathbf{l} | \mathbf{d}\}$. Explain the procedure for estimating them.

Display the results from the estimations as maps.

Comment on the results.

- (f) Compare the results in (b) and e) and comment on them.

12.2 Exercises in Spatial Modelling

The written exam is conducted without any printed material or books; only one self-made handwritten A5 peep sheet is permitted. The use of computers and calculators is not allowed. Exercise sets, inspired by exam problems, are presented in the following subsections. Suggested solutions can be found as supplementary material online.

12.2.1 Exercise Sets: Continuous Spatial Variables

This subsection contains exercises for continuous spatial variables in what we consider to be an increasing order of difficulty.

Continuous RF: Exercise 1

Consider a one-dimensional Gaussian RF $\{r(x); x \in D \subset \mathbb{R}\}$ parametrised by

$$\begin{aligned} E\{r(x)\} &= \mu_{r0} + \mu_{r1}g(x) \\ \text{Var}\{r(x)\} &= \sigma_r^2 h(x) \\ \text{Corr}\{r(x), r(x')\} &= \rho_r(x - x'), \end{aligned}$$

where $\{g(x); x \in D\}; g(x) \in \mathbb{R}$ and $\{h(x); x \in D\}; h(x) \in \mathbb{R}_+$ are known functions on D . The remaining model parameters are $\mu_{r0}, \mu_{r1} \in \mathbb{R}$, $\sigma_r^2 \in \mathbb{R}_+$ and $\rho_r(\tau) \in \mathbb{R}_{[-1,1]}$. Assume that the model parameters σ_r^2 and $\rho_r(\tau)$ are known, whereas $[\mu_{r0}, \mu_{r1}]$ are unknown.

- (a) Specify the mathematical requirements for $\rho_r(\tau)$ to be a valid spatial correlation function.

Assume that the Gaussian RF over D is exactly observed by $(r(x_1), r(x_2), \dots, r(x_m))$. Consider an arbitrary location $x_0 \in D$ and define the linear predictor

$$\hat{r}(x_0) = \sum_{i=1}^m \alpha_i r(x_i)$$

with unknown weights $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)$ to be determined.

- (b) Develop the expression for the minimisation system to be solved to determine the weights for the best linear unbiased (BLU) predictor under squared-error loss.

Consider the linear estimators for the unknown model parameters,

$$\hat{\mu}_{r0} = \sum_{i=1}^m \beta_i^0 r(x_i)$$

$$\hat{\mu}_{r1} = \sum_{i=1}^m \beta_i^1 r(x_i)$$

with unknown weights $\boldsymbol{\beta}^0 = (\beta_1^0, \beta_2^0, \dots, \beta_m^0)$ and $\boldsymbol{\beta}^1 = (\beta_1^1, \beta_2^1, \dots, \beta_m^1)$ to be determined.

- (c) Develop the expressions for the two minimisation systems to be solved in order to obtain the BLU estimators under squared-error loss for μ_{r0} and μ_{r1} , respectively.

Continuous RF: Exercise 2

Consider a continuous RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$ with model parameters $E\{r(\mathbf{x})\} = \mu_r \in \mathbb{R}$ and $\text{Var}\{r(\mathbf{x})\} = \sigma_r^2 \in \mathbb{R}_+$. The spatial isotropic correlation function is $\text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} = \rho_r(\tau) \in \mathbb{R}_{[-1,1]}$ with $\tau = |\mathbf{x} - \mathbf{x}'|$. Let $D = [0, 10] \times [0, 10] \subset \mathbb{R}^2$.

Let the expectation μ_r be unknown, whereas the variance σ_r^2 and spatial correlation function $\rho_r(\tau)$ are known.

Define also the spatial average over D

$$r_D = \frac{1}{|D|} \int_D r(\mathbf{u}) d\mathbf{u},$$

where $|D|$ is the area of D .

- (a) Develop expressions for $E\{r_D\}$ and $\text{Var}\{r_D\}$.

Consider an arbitrary location $\mathbf{x}_0 \in D$ and develop an expression for the covariance $\text{Cov}\{r(\mathbf{x}_0), r_D\}$.

Let the RF be observed at locations $\mathbf{x}_1^d, \mathbf{x}_2^d, \dots, \mathbf{x}_m^d \in D$. Consequently, the observations $(r(\mathbf{x}_1^d), r(\mathbf{x}_2^d), \dots, r(\mathbf{x}_m^d))$ are collected.

Define the linear estimator for expectation μ_r

$$\hat{\mu}_r = \sum_{i=1}^m \beta_i r(\mathbf{x}_i^d),$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ are unknown weights to be determined.

- (b) Develop an expression for the BLU estimator under squared-error loss for μ_r , with associated estimation variance. Only the minimisation problem to be solved needs to be specified.

Define the linear predictor for the spatial average r_D

$$\hat{r}_D = \sum_{i=1}^m \alpha_i r(\mathbf{x}_i^d),$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ are unknown weights to be determined.

- (c) Develop an expression for the BLU predictor under square-error loss for r_D , with associated prediction variance. Only the minimisation problem to be solved needs to be specified.

Compare the expressions developed in the current point and point (b).

Continuous RF: Exercise 3

Consider a one-dimensional Gaussian RF $\{r(x); x \in \mathbb{R}\}$ with model parameters

$$E\{r(x)\} = \begin{cases} \alpha & \text{for } x < 0.0 \\ \beta & \text{for } x \geq 0.0 \end{cases}; x \in \mathbb{R}$$

$$\text{Cov}\{r(x), r(x')\} = \sigma_r^2 \rho_r(\tau); x, x' \in \mathbb{R},$$

where $\alpha, \beta \in \mathbb{R}$ are unknown constants and $\tau = |x - x'|$.

- (a) Specify the requirements for the function $\rho_r(\tau)$ to be an eligible spatial correlation function.

Let the spatial variable be observed in the locations $(-2, -1, 2, 4)$, giving the observed values $(r(-2), r(-1), r(2), r(4))$.

- (b) Develop the BLU estimator under squared-error loss for the model parameters α and β and denote them $\hat{\alpha}$ and $\hat{\beta}$. Only the minimisation problem to be solved needs to be developed.

Specify the expressions for the estimation variances for α and β as functions of the BLU estimator weights developed.

The expectation function $E\{r(x)\} = \mu_r(x)$ has a discontinuity at $x = 0.0$. The step size at this discontinuity is $\Delta = \alpha - \beta$, which can be either positive or negative. An unbiased estimator for this step size is $\hat{\Delta} = \hat{\alpha} - \hat{\beta}$.

- (c) Develop the expression for the 0.9 confidence interval for the step size based on this estimator. Let the expression be a function of the BLU estimator weights identified in point (b).

Is the estimator $\hat{\Delta}$ the BLU estimator for Δ ? Justify your answer mathematically.

Continuous RF: Exercise 4

Consider a one-dimensional Gaussian RF $\{r(x); x \in D \subset \mathbb{R}\}$ parametrised by

$$\begin{aligned} E\{r(x)\} &= \mu_r \\ \text{Var}\{r(x)\} &= \sigma_r^2 \\ \text{Corr}\{r(x), r(x')\} &= \rho_r(\tau). \end{aligned}$$

The model parameters are $\mu_r \in \mathbb{R}$, $\sigma_r^2 \in \mathbb{R}_+$ and $\rho_r(\tau) \in \mathbb{R}_{[-1,1]}$; $\tau = x - x' \in \mathbb{R}$. Assume that the model parameters σ_r^2 and $\rho_r(\tau)$ are known, whereas μ_r is unknown.

- (a) Specify the mathematical requirements for $\rho_r(\tau)$ to be a valid non-negative definite spatial correlation function.

Assume that $\rho_r^i(\tau); i = 1, 2, \dots, n_\rho$ are non-negative definite correlation functions. Specify two alternative ways of combining these functions that ensure the non-negative definiteness of their combinations.

Define a supplementary one-dimensional Gaussian RF $\{s(x); x \in D \subset \mathbb{R}\}$ where

$$\{[s(x) \mid r(x)] = \gamma_{sr}r(x) + e(x); x \in D\}$$

with the Gaussian RF $\{e(x); x \in D \subset \mathbb{R}\}$ parametrised by

$$\begin{aligned} E\{e(x)\} &= 0.0 \\ \text{Var}\{e(x)\} &= \sigma_e^2 \\ \text{Corr}\{e(x), e(x')\} &= \rho_e(\tau) \\ \text{Corr}\{e(x), r(x')\} &= 0.0. \end{aligned}$$

The model parameters are $\gamma_{sr} \in \mathbb{R}$, $\sigma_e^2 \in \mathbb{R}_+$ and $\rho_e(\tau) \in \mathbb{R}_{[-1,1]}$; $\tau = x - x' \in \mathbb{R}$. Assume that the model parameters γ_{sr} , σ_e^2 and $\rho_e(\tau)$ are known.

- (b) Develop expressions for the parameters of the joint Gaussian RF $\{[r(x), s(x)]; x \in D \subset \mathbb{R}\}$.

Consider exact observations of the RFs

$$\begin{aligned} \mathbf{r}^d &= \{r(x_1^{dr}), r(x_2^{dr})\} \\ \mathbf{s}^d &= \{s(x_1^{ds}), s(x_2^{ds})\}, \end{aligned}$$

where the corresponding four locations in the domain D may be different.

Consider an arbitrary location $x_0 \in D$ and define the linear predictor

$$\hat{r}(x_0) = \sum_{i=1}^2 \alpha_i r(x_i^{dr}) + \sum_{i=1}^2 \beta_i s(x_i^{ds})$$

with unknown weights $[(\alpha_1, \alpha_2), (\beta_1, \beta_2)]$ to be determined.

- (c) Develop the expression for the minimisation system to be solved to determine the weights for the BLU predictor under squared-error loss. Note that the minimisation does not need to be performed.

Continuous RF: Exercise 4

Consider a one-dimensional continuous RF $\{r(x); x \in \mathbb{R}\}$. Assume the following:

$$E\{r(x)\} = 0.0$$

$$\text{Var}\{r(x)\} = 1.0$$

$$\text{Corr}\{r(x), r(x')\} = \rho_r(\tau) = \exp(-\tau^2),$$

where $\tau = x - x' \in \mathbb{R}_\oplus$.

- (a) Which additional assumptions must be made for $\{r(x); x \in \mathbb{R}\}$ to be a Gaussian RF?

Assume that the RF is Gaussian. Sketch graphically the bivariate pdfs for $[r(0.0), r(0.1)]$, $[r(0.0), r(1.0)]$ and $[r(0.0), r(10.0)]$.

Define the differential RF

$$\left\{ \dot{r}(x) = \frac{dr(x)}{dx}; x \in \mathbb{R} \right\}.$$

- (b) Specify the requirements for $\{\dot{r}(x); x \in \mathbb{R}\}$ to exist.

Develop the expressions for the covariance function $\text{Cov}\{\dot{r}(x), \dot{r}(x')\}$ and the cross-covariance function $\text{Cov}\{r(x), \dot{r}(x')\}$.

Sketch the covariance function for $\{r(x); x \in \mathbb{R}\}$ graphically, and the two covariance functions developed above.

Comment on the covariance between $r(x)$ and $\dot{r}(x)$ in an arbitrary location $x \in \mathbb{R}$.

Assume that the differential RF is observed at $x = 0.0$, with a value $\dot{r}(0.0) = 0.5$.

- (c) Develop the BLU predictor under squared-error loss with associated prediction variance for $\{r(x); x \in \mathbb{R}\}$ based on the observation.

Sketch the BLU predictor and prediction variance graphically and comment on the results.

Continuous RF: Exercise 5

Consider the Gaussian RF $\{r(x); x \in D \subset \mathbb{R}\}$ with model parameters

$$\mathbb{E}\{r(x)\} = \mu_r(x) = 10 - (x - a)^2; x \in \mathbb{R}$$

$$\text{Cov}\{r(x), r(x')\} = c_R(\tau) = \sigma_r^2 \exp\left(-\frac{1}{10}\tau^2\right); x, x' \in \mathbb{R}$$

with $\tau = x - x'$. Consequently, the field has the highest expected value of 10 in the unknown location $x = a$. The covariance function is translation-invariant and assumed known.

Let the RF be observed in the locations $(x_1^d, x_2^d, \dots, x_m^d)$, which entails that the observations are $\mathbf{r}^d = (r(x_1^d), r(x_2^d), \dots, r(x_m^d))$.

Define a linear estimator for the location of the highest expected value a as

$$\hat{a} = \sum_{i=1}^m \beta_i r(x_i^d),$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ are unknown weights to be determined.

- (a)** Develop an expression for the BLU estimator under squared-error loss for a based on the set of observations \mathbf{r}^d . Only the minimisation problem to be solved to identify the weights needs to be developed.

Continuous RF: Exercise 6

Consider a continuous RF $\{r(\mathbf{x}); \mathbf{x} \in \mathbb{R}^q\}$ with model parameters

$$\mathbb{E}\{r(\mathbf{x})\} = \mu_r$$

$$\text{Var}\{r(\mathbf{x}) - r(\mathbf{x}')\} = 2\gamma_r(\tau)$$

with $\tau = |\mathbf{x} - \mathbf{x}'|$. The model parameters are $\mu_r \in \mathbb{R}$ an unknown constant and $\gamma_r(\tau) \in \mathbb{R}_+$ the known spatial variogram function.

Assume first that the RF $\{r(\mathbf{x}); \mathbf{x} \in \mathbb{R}^q\}$ is second-order stationary.

- (a)** Explain what the second-order stationarity assumption entails.

Under this stationarity assumption, develop the relation between the spatial covariance and spatial variogram function.

From here on, assume that $\{r(\mathbf{x}); \mathbf{x} \in \mathbb{R}^q\}$ is an intrinsic RF of order one, denoted IRF-1.

- (b)** Explain what the IRF-1 assumption entails.

List the requirements that the spatial variogram function, $\gamma_r(\tau)$, must satisfy to be considered an eligible function under the IRF-1 assumptions.

Briefly explain the reason for these requirements.

Assume that the spatial variogram function has the following form:

$$\gamma_r(\tau) = \sigma |\tau|^\nu$$

with scale parameter $\sigma \in \mathbb{R}_+$ and shape parameter $\nu \in \mathbb{R}_{(0,2]}$. An RF with this spatial variogram function is known to be a fractal, more specifically affine fractal.

- (c) Explain the characteristics of an affine fractal RF.

Demonstrate that the spatial variogram function specified above defines an affine fractal RF.

Continuous RF: Exercise 7

Consider a continuous RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}; r(\mathbf{x}) \in \mathbb{R}$ with expectation function $\{\mathbb{E}\{r(\mathbf{x})\} = \mu_r(\mathbf{x}); \mathbf{x} \in D\}$, variance function $\{\text{Var}\{r(\mathbf{x})\} = \sigma_r^2(\mathbf{x}); \mathbf{x} \in D\}$ and non-negative definite spatial correlation function $\{\text{Corr}\{r(\mathbf{x}), r(\mathbf{x}')\} = \rho_r(\mathbf{x}, \mathbf{x}'); \mathbf{x}, \mathbf{x}' \in D\}$. Assume that all model parameters are known.

- (a) Specify the exact requirements for the continuous RF $\{r(\mathbf{x}); \mathbf{x} \in D\}$ to be a Gaussian RF.

Specify the additional requirements for $\{r(\mathbf{x}); \mathbf{x} \in D\}$ to be stationary and isotropic.

Consider a related continuous RF

$$\{s(\mathbf{x}) = \sum_{l=1}^{n_g} b_l g_l(\mathbf{x}) + r(\mathbf{x}) = \mathbf{g}^T \mathbf{b} + r(\mathbf{x}); \mathbf{x} \in D\}$$

with

$$\begin{aligned} \mathbf{b} &= (b_1, b_2, \dots, b_{n_g})^T \sim p(\mathbf{b}) = \phi_{n_g}(\mathbf{b}; \mathbf{0} i_{n_g}, \sigma_b^2 \mathbf{I}_{n_g}) \\ \{\mathbf{g}(\mathbf{x}) &= (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_{n_g}(\mathbf{x}))^T; \mathbf{x} \in D\}, \end{aligned}$$

where the model parameter $\sigma_b^2 \in \mathbb{R}_+$ and the functions in $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^{n_g}$ are known. Assume further that $\{r(\mathbf{x}); \mathbf{x} \in D\}$ is a stationary and isotropic Gaussian RF, independent of \mathbf{b} .

- (b) Is $\{s(\mathbf{x}); \mathbf{x} \in D\}$ a Gaussian RF? Justify the answer.

Develop expressions for the expectation function $\mu_s(\mathbf{x})$, the variance function $\sigma_s^2(\mathbf{x})$ and the spatial correlation function $\rho_s(\mathbf{x}, \mathbf{x}')$ for $\{s(\mathbf{x}); \mathbf{x} \in D\}$. The expressions depend on the model parameters of $\{r(\mathbf{x}); \mathbf{x} \in D\}$.

Consider a set of observations of $\{s(\mathbf{x}); \mathbf{x} \in D\}$ in locations $\mathbf{x}^d = (\mathbf{x}_1^d, \mathbf{x}_2^d, \dots, \mathbf{x}_m^d)$ and denote the observed values $\mathbf{s}^d = (s(\mathbf{x}_1^d), s(\mathbf{x}_2^d), \dots, s(\mathbf{x}_m^d))$. Consider an arbitrary location $\mathbf{x}_0 \in D$.

- (c) Develop the expression for

$$\text{Prob}\{s(\mathbf{x}_0) > s_0 \mid \mathbf{s}^d\}.$$

Continuous RF: Exercise 8

Consider a Gaussian RF $\{r(\mathbf{x}); \mathbf{x} \in D \subset \mathbb{R}^2\}$. Assume that the RF is stationary and isotropic with expectation and variance levels $\mu_r \in \mathbb{R}$ and $\sigma_r^2 \in \mathbb{R}_+$, respectively. The spatial correlation function $\rho_r(\tau) \in \mathbb{R}_{[-1,1]}; \tau = |\mathbf{x} - \mathbf{x}'| \in \mathbb{R}_+$ is non-negative definite.

Let the RF be represented in a n -vector \mathbf{r} from the grid system $L \subset D$. The n -vector including the expectations is $\mu_r \mathbf{i}_n$ and the covariance matrix, of dimension $(n \times n)$, is Σ_r . The spatial variable is observed in an m -vector \mathbf{d} according to the likelihood relation $p(\mathbf{d} \mid \mathbf{r})$ defined by

$$[\mathbf{d} \mid \mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e},$$

where the \mathbf{H} is an observation matrix, of dimension $(m \times n)$, and the m -vector \mathbf{e} contains centred Gaussian errors with covariance matrix, of dimension $(m \times m)$, $\Sigma_{d|r}$.

- (a) Specify the $(n+m)$ -dimensional joint pdf for $[\mathbf{r}, \mathbf{d}]$.

Specify the n -dimensional conditional pdf for $[\mathbf{r} \mid \mathbf{d}]$, denoted the posterior pdf $p(\mathbf{r} \mid \mathbf{d})$.

A realisation \mathbf{r}^s from the posterior pdf $p(\mathbf{r} \mid \mathbf{d})$ can be generated by a randomised optimisation approach:

Generate: $\mathbf{r}^* \sim \phi_n(\mathbf{r}^*; \mu_r \mathbf{i}_n, \Sigma_r)$

Generate: $\mathbf{d}^* \sim \phi_m(\mathbf{d}^*; \mathbf{d}, \Sigma_{d|r})$

Determine: $\mathbf{r}^s = \arg \min_{\mathbf{r}} \{(\mathbf{r}^* - \mathbf{r})^T \Sigma_r^{-1} (\mathbf{r}^* - \mathbf{r}) + (\mathbf{d}^* - \mathbf{H}\mathbf{r})^T \Sigma_{d|r}^{-1} (\mathbf{d}^* - \mathbf{H}\mathbf{r})\}$.

- (b) Demonstrate that the approach generates a realisation from the posterior pdf $p(\mathbf{r} \mid \mathbf{d})$.

The following matrix identities may be helpful in the demonstration:

$$\Sigma_r \mathbf{H}^T \Sigma_d^{-1} = \Sigma_{r|d} \mathbf{H}^T \Sigma_{d|r}^{-1}$$

$$[\Sigma_r^{-1} + \mathbf{H}^T \Sigma_{d|r}^{-1} \mathbf{H}]^{-1} = \Sigma_r - \Sigma_r \mathbf{H}^T [\mathbf{H} \Sigma_r \mathbf{H}^T + \Sigma_{d|r}]^{-1} \mathbf{H} \Sigma_r.$$

12.2.2 Exercise Sets: Event Spatial Variables

This subsection contains exercises for event spatial variables in what we consider to be an increasing order of difficulty.

Event RF: Exercise 1

Consider a stationary Poisson event-location RF $\{\mathbf{x}_i; i = 1, 2, \dots, k_D, D \subset \mathbb{R}^2\}$ defined over the area $D : [0, 5] \times [0, 5] \text{ m}^2 \subset \mathbb{R}^2$. Let the intensity parameter be $\lambda \in \mathbb{R}_+ [\text{m}^{-2}]$.

- (a) Specify the probability distribution for the number of events, $k_D \in \mathbb{N}_+$, in D . If an area of size $1 \text{ m}^2 \in D$ is covered, what is the probability that it covers exactly 5 events? Given that there are 5 events in the covered area, what is the probability distribution for the number of events in the remaining area, i.e. the area not covered?

An inspector surveys the entire area D to register the events, but each event is only registered with probability $\alpha \in \mathbb{R}_{[0,1]}$. Consequently, each event may be overlooked with probability $(1 - \alpha)$. The registration process is considered to be independent from one event to another.

- (b) Given that the number of events in D is $k \in \mathbb{N}_+$, develop an expression for the probability distribution for the number of registered events k_o . Without knowledge of the total number k_D , develop an expression for the probability distribution for the registered events k_o .
 - (c) Given that the number of registered events is $k_o \in \mathbb{N}_+$, develop an expression for the probability distribution for the total number of events k_D . Without knowledge of the total number k_D , are the number of registered events k_o and the number of non-registered events ($k_D - k_o$) independent? Justify the answer.
- Comment on the answers given above.

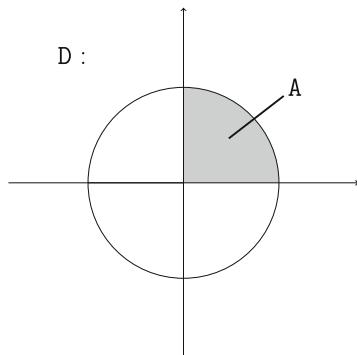
Event RF: Exercise 2

Consider a stationary Poisson event-location RF represented by $\{\mathbf{x}_i; i = 1, 2, \dots, k_D\}; \mathbf{x}_i \in D \subset \mathbb{R}^2; k_D \in \mathbb{N}_+$ with model intensity parameter $\lambda \in \mathbb{R}_+$. Let the domain D be a circular disc centred at the origin $(0, 0)$ with radius r .

Define a sub-domain $A \subset D$, where A is the upper-right quarter sector of the disc D , as shown in the Fig. 12.1.

Let $k_D \in \mathbb{N}_+$ and $k_A \in \mathbb{N}_+$ denote the number of events in the domains D and A , respectively.

- (a) Develop expressions for $E\{k_D\}$, $E\{k_A\}$, $\text{Var}\{k_D\}$, $\text{Var}\{k_A\}$ and $\text{Cov}\{k_D, k_A\}$.
- (b) Assume first that k_D is unknown and that $k_A = k$ has been observed. Specify the expression for $\text{Prob}\{k_D = i | k_A = k\}, i \in \mathbb{N}_+$. Next, assume that k_A is unknown and that $k_D = k$ has been observed. Specify the expression for $\text{Prob}\{k_A = i | k_D = k\}, i \in \mathbb{N}_+$.

Fig. 12.1 Sketch of domains

Assume that k_D is unknown and that $k_A = k \in \mathbb{N}_+$ has been observed. Consider the centre location in D , i.e. the origin, and define d as the distance from this centre location to the closest event location in the event RF.

- (c) Develop the expression for the pdf $p(d | k_A = k); d \in \mathbb{R}_+$.

Event RF: Exercise 3

Consider a two-dimensional stationary Poisson event RF on \mathbb{R}^2 with event intensity $\lambda \in \mathbb{R}_+$. Let $A \subset \mathbb{R}^2$ and $B \subset \mathbb{R}^2$ with $A \cap B = \emptyset$ be two disjoint domains, and let $k_A \in \mathbb{N}_+$ and $k_B \in \mathbb{N}_+$ be the number of events in the domains A and B , respectively.

- (a) Specify the probability for observing exactly two events in domain A .

Assume that $k_A = 2$, and develop the expression for the conditional probability $\text{Prob}\{k_B = k | k_A = 2\}$ for $k \in \mathbb{N}_+$.

Consider an arbitrary location $\mathbf{x}_0 \in \mathbb{R}^2$ and

$t_{(1)}$ – distance from \mathbf{x}_0 to the nearest event location in the Poisson RF

$t_{(2)}$ – distance from \mathbf{x}_0 to the second-nearest event location in the Poisson RF

- (b) Develop the pdf for the variable $t_{(2)}$.

Develop the pdf for the bivariate variable $(t_{(1)}, t_{(2)})$.

Demonstrate that the two pdfs are valid pdfs.

Event RF: Exercise 4

A cheese type contains holes that can be modelled as follows. The hole centres are located according to a three-dimensional stationary Poisson event RF $\{\mathbf{x}_i; i = 1, 2, \dots, k_D; D \subset \mathbb{R}^3\}$ with intensity $\lambda [(\text{dm}^3)^{-1}]$. The holes are circular with constant radius $r [\text{dm}]$. Hence, a cheese unit consists of cheese with density $\rho [\text{g}/\text{dm}^3]$ and possibly overlapping holes that add nothing to the weight.

- (a) Specify an expression for the pdf of the number of hole centres inside a cheese unit of volume 1 dm^3 .

Note that the pdf for the distance d from an arbitrary location in the cheese to the nearest hole centre is

$$d \sim p(d) = 4\lambda d^2 \times \exp(-\lambda \frac{4}{3}\pi d^3); d \in \mathbb{R}_+.$$

- (b) Consider a cheese unit of volume 1 dm^3 and develop an expression for the expected weight of this cheese unit.
- (c) Consider an exactly circular hole in the cheese unit having no overlap with other holes. Develop an expression for the pdf of the minimum thickness of the cheese layer between this hole and another hole.

Event RF: Exercise 5

Consider a stationary Poisson event-location RF on \mathbb{R}^2 with intensity parameter $\lambda \in \mathbb{R}_+$. Let $\mathbf{x}_0 \in \mathbb{R}^2$ be an arbitrary location, and define $t_{(i)} \in \mathbb{R}_+$ for $i = 1, 2, \dots$ as the distances to the i th nearest event location from location \mathbf{x}_0 .

- (a) Specify the definition of a Poisson event-location RF with event intensity λ on \mathbb{R}^2 .
If the Poisson RF is constrained to be in a finite domain $D \subset \mathbb{R}^2$, i.e. $|D| < \infty$, an alternative definition is available. Specify this alternative definition.
- (b) Demonstrate that the joint pdf for $(t_{(1)}, t_{(2)}, \dots, t_{(k)})$ for $k \in \mathbb{N}_+$ is

$$p(t_{(1)}, t_{(2)}, \dots, t_{(k)}) = \begin{cases} [2\lambda\pi]^k \prod_{i=1}^k t_{(i)} \exp\{-\lambda\pi t_{(k)}^2\} & \text{for } 0 < t_{(1)} < \dots < t_{(k)} \\ 0.0 & \text{otherwise.} \end{cases}$$

Assume that only the nearest and the third nearest event locations to the location \mathbf{x}_0 are observed. Thus, only $t_{(1)}$ and $t_{(3)}$ are known.

- (c) Develop the maximum likelihood estimator for the event intensity λ based on these two observations.

Event RF: Exercise 6

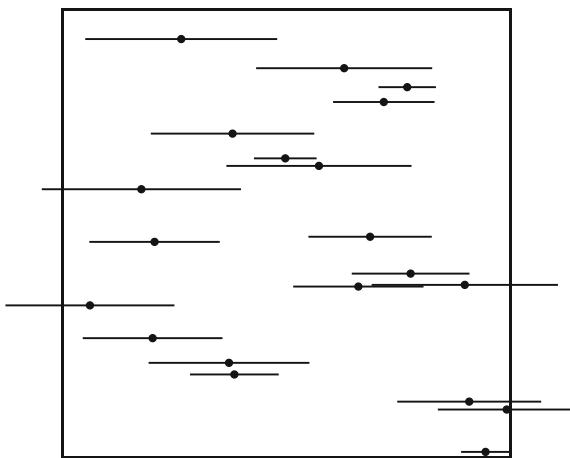
Consider a stationary, marked event-location RF $\{(\mathbf{x}_i, l_i); i = 1, 2, \dots, k_D; D \subset \mathbb{R}^2\}$, defined over area $D : [0, 10] \times [0, 10] [\text{m}^2] \subset \mathbb{R}^2$, representing horizontal (parallel to the first axis) line segments of length l centred at location $\mathbf{x} \in D$, as shown in the Fig. 12.2.

Let $\{\mathbf{x}_i; i = 1, 2, \dots, k_D; D \subset \mathbb{R}^2\}$ be a stationary Poisson event-location RF with intensity parameter $\lambda \in \mathbb{R}_+ [\text{m}^{-2}]$, whereas the line segment length l has pdf $p(l)$. Moreover, let the random variables \mathbf{x} and l be independent.

Note that the line segments may intersect the vertical (parallel to the second axis) boundaries of the domain D .

Assume first that the line segment lengths l are constant to $2 [\text{m}]$, i.e. $p(l)$ is a Dirac pdf at $l = 2$.

Fig. 12.2 Example of line segments



- (a) Specify an expression for the pdf of the number of line centres in the domain D . Specify an expression for the pdf of the number of line segments intersecting the boundary of D . Specify the expected number of intersections.

Consider one specific line segment located centrally in the domain D , such that boundary effects can be ignored.

- (b) Develop an expression for the pdf of the shortest distance between this specific line segment and any other line segment.

Assume now that the line segment lengths are random with pdf

$$l \sim p(l) = \begin{cases} \frac{1}{2}l & \text{for } 0 < l < 2 \\ 0 & \text{otherwise.} \end{cases}$$

- (c) Develop an expression for the pdf of the number of line segments that intersect the boundary of the domain D .

Consider one arbitrary line segment intersecting the boundary of D .

- (d) Develop an expression for the pdf of the length of this line segment that intersects the boundary. Comment on the result.

Event RF: Exercise 7

One evening, a student couple buys a circular pizza P with radius r , i.e. with area $a_P = \pi r^2$. On top of the pizza, pieces of olives are distributed. Denote the locations of the olive pieces on P as $\mathbf{X}_P : \{\mathbf{x}_i ; i = 1, 2, \dots, k_P\}; \mathbf{x}_i \in P \subset \mathbb{R}^2, k_P \in \mathbb{N}_0$, where k_P and \mathbf{x}_i are the number and centre locations of the olive pieces, respectively.

Assume that \mathbf{X}_P is distributed according to a stationary Poisson event-location RF with intensity parameter λ_o , and hence both k_P and \mathbf{x}_i are random variables.

- (a) Specify the expression for the pdf of the olive piece locations on top of the pizza P , $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_P})$.
- (b) Specify an expression for the expected number of olive pieces on the pizza.
Given that one half of the pizza contains exactly k_h olive pieces, specify an expression for the expected total number of olive pieces on the pizza.

The student couple always share the pizza by splitting it into two slices, P_M and P_F , one for the male and one for the female student. Let the respective pizza areas be a_M and a_F , hence $a_P = a_M + a_F$, and the respective proportions of the pizza are $v_M = \frac{a_M}{a_P}$ and $v_F = \frac{a_F}{a_P}$.

Experience from the student couple relationship tells that these proportions are distributed according to the pdf

$$p(v_M, v_F; \alpha, \beta) = \text{const} \times v_M^{\alpha-1} v_F^{\beta-1}$$

with $v_m + v_F = 1$; $v_M, v_F \in \mathbb{R}_{[0,1]}$, and model parameters $\alpha, \beta \in \mathbb{R}_+$.

The expected areas of P_M and P_F are then $\frac{\alpha}{\alpha+\beta} a_P$ and $\frac{\beta}{\alpha+\beta} a_P$, respectively. Since the male student usually eats more than the female one, assume that $\alpha > \beta$.

This evening, after having split the pizza into two slices, P_M and P_F , the students count the number of olive pieces on each pizza slice and observe exactly k_M and k_F .

- (c) Given the olive piece counts, k_M and k_F , and the experience from previous pizza evenings, develop expressions for the expected areas of the pizza slices P_M and P_F this very evening.

Event RF: Exercise 8

Consider a Poisson event count RF spatially discretised to a grid of size n , represented by the n -vector \mathbf{k} . Let the prior model for $\mathbf{k} \in \mathbb{N}_+^n$ be

$$\mathbf{k} \sim p(\mathbf{k}) = \prod_{i=1}^n \frac{[\lambda_k^n]^{k_i}}{k_i!} \exp(-\lambda_k^n),$$

where the parameter λ_k^n is the prior intensity.

Consider an observation $2n$ -vector $\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n)$; $\mathbf{d}_i = (d_{i1}, d_{i2}) \in \mathbb{N}_+^2$. Let the likelihood model be of conditionally independent type with single-site response

$$[\mathbf{d}|\mathbf{k}] \sim p(\mathbf{d} | \mathbf{k}) = \prod_{i=1}^n p(\mathbf{d}_i | k_i).$$

- (a) Assume that the grid unit likelihood model is multinomial with misclassification probabilities $(1 - \alpha_1)$ and $(1 - \alpha_2)$, where $\alpha_i \in \mathbb{R}_{[0,1]}$ and $\alpha_1 + \alpha_2 \leq 1$. Thus,

$$\begin{aligned} [\mathbf{d}_i | k_i] \sim p(\mathbf{d}_i | k_i) &= \frac{k_i!}{(k_i - d_{i1} - d_{i2})! d_{i1}! d_{i2}!} \\ &\times \alpha_1^{d_{i1}} \alpha_2^{d_{i2}} (1 - \alpha_1 - \alpha_2)^{k_i - d_{i1} - d_{i2}}. \end{aligned}$$

Demonstrate that the Poisson event count prior model is conjugate with respect to this likelihood model. Thus, the posterior model is also a Poisson event count RF model.

- (b) Assume that the grid unit observations are collected sequentially with misclassification probabilities $(1 - \alpha_1)$ and $(1 - \alpha_2)$, where $\alpha_i \in \mathbb{R}_{[0,1]}$. Thus,

$$\begin{aligned} [\mathbf{d}_i | k_i] \sim p(\mathbf{d}_i | k_i) &= p(d_{i2} | k_i, d_{i1}) p(d_{i1} | k_i) \\ &= \frac{(k_i - d_{i1})!}{(k_i - d_{i1} - d_{i2})! d_{i2}!} \times \alpha_2^{d_{i2}} (1 - \alpha_2)^{k_i - d_{i1} - d_{i2}} \\ &\times \frac{k_i!}{(k_i - d_{i1})! d_{i1}!} \times \alpha_1^{d_{i1}} (1 - \alpha_1)^{k_i - d_{i1}}. \end{aligned}$$

Demonstrate that the Poisson event count prior model is conjugate with respect to this likelihood model. Thus, the posterior model is also a Poisson event count RF model.

Compare the result with the result in point (a).

12.2.3 Exercise Sets: Mosaic Spatial Variables

This subsection contains exercises for mosaic spatial variables in what we consider to be an increasing order of difficulty.

Mosaic RF: Exercise 1

Consider the two-dimensional spatially discretised mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L \in D \subset \mathbb{R}^2\}$ where L is a grid of size n over the domain D . The sample space for each $\mathbf{x} \in L$ is $l(\mathbf{x}) \in \mathbb{L} : \{1, 2, \dots, L\}$, and assume that all outcomes have positive probability. Let the mosaic RF be represented by the n -vector \mathbf{l} .

Assume that the mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L \in D \subset \mathbb{R}^2\}$ is a Markov RF with a (3×3) -neighbourhood design for all internal grid nodes $\mathbf{x} \in L$.

- (a) Mathematically specify the Markov RF in Markov formulation.

Mathematically specify and comment on the corresponding Gibbs formulation for the Markov RF.

Discuss the major message in the Hammersley–Clifford theorem.

Mosaic RF: Exercise 2

Consider a mosaic RF $\mathbf{l} : \{l(\mathbf{x}); \mathbf{x} \in L\}$ where L is a regular spatial grid over $D \subset \mathbb{R}^2$ and $l(\mathbf{x}) \in \mathbb{L} : \{W, G, B\}$. Consequently, the variables $l(\mathbf{x})$ can belong to one of the labels white (W), grey (G) or black (B) for each $\mathbf{x} \in L$. Let the mosaic RF be represented by the n -vector \mathbf{l} .

Define the following Gibbs formulation for the mosaic RF:

$$\mathbf{l} \sim p(\mathbf{l}; \boldsymbol{\beta} = (\beta_W, \beta_G, \beta_B)) = \text{const}_{\beta} \times \exp \left(\sum_{<i,j>} \sum_{l_i \in \mathbb{L}} \beta_{l_i} I(l_i = l_j) \right),$$

where $< i, j >$ represents all pairs of nearest nodes in the grid L and $I(A)$ is an indicator function taking the value 1 whenever A is true and 0 otherwise. The vector of model parameter $\boldsymbol{\beta} = (\beta_W, \beta_G, \beta_B) \in \mathbb{R}^3$ is associated with the spatial continuity for each of the labels.

- (a) Develop the corresponding Markov formulation for the mosaic RF. Be precise with the notation.

Discuss the most important differences between the Gibbs and the Markov formulations.

Assume that a realisation of the mosaic RF is known $\mathbf{l}^d : \{l^d(\mathbf{x}); \mathbf{x} \in L\}$. This realisation is used to estimate the vector of model parameter $\boldsymbol{\beta} = (\beta_W, \beta_G, \beta_B)$.

- (b) Discuss how to perform this estimation in a reliable manner.

Specify the expression for a pseudo-likelihood that could be maximised to determine $\boldsymbol{\beta}$.

Mosaic RF: Exercise 3

Consider a one-dimensional Markov RF $\{l(x); x \in L\}$, denoted a Markov random profile (RP), where L is a regular grid with n nodes on $D \subset \mathbb{R}$, represented by the n -vector $\mathbf{l} = (l_1, l_2, \dots, l_n)$. Let $l(x) \in \Omega_l : \{W, B\}$. Consequently, the variable $l(x)$ belongs to one of the labels, white (W) or black (B), for each $x \in L$.

Define the following Gibbs formulation for the RP:

$$\begin{aligned} p(\mathbf{l}) &= \text{const} \times \prod_{<i,j>} \beta^{I(l_i = l_j)} \\ &= \text{const} \times \prod_{i=1}^{n-1} \beta^{I(l_i = l_{i+1})}, \end{aligned}$$

where $< i, j >$ represent all pairs of nearest nodes in the grid L and $I(A)$ is an indicator function taking the value 1 whenever A is true and 0 otherwise. Thereby, the RP is an Ising RP. The model parameter $\beta \in \mathbb{R}_{[1, \infty]}$ is assumed known.

- (a) Develop the expression for the Markov formulation of the RP $p(l_i | \mathbf{l}_{-i})$; $i = 1, 2, \dots, n$. Be aware of the boundary expressions.

Any pdf for a multivariate random variable can be sequentially decomposed as

$$p(\mathbf{l}) = p(l_1) \prod_{i=2}^n p(l_i | l_{i-1}, l_{i-2}, \dots, l_1).$$

- (b)** Develop the expressions for $p(l_i | l_{i-1}, l_{i-2}, \dots, l_1); i = 2, 3, \dots, n$ based on the Gibbs formulation of the RP.

Demonstrate that this sequential decomposition defines a first-order Markov chain along the profile.

Mosaic RF: Exercise 4

Consider a spatially discretised two-dimensional mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L \in D \subset \mathbb{R}^2\}$ with L being a regular grid over D . Let the sample space for all $\mathbf{x} \in L$ be $l(\mathbf{x}) \in \mathbb{L} : \{-1, 1\}$, and assume that all outcomes have positive probability. Let the mosaic RF be represented by the n -vector \mathbf{l} . Assume that the mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L\}$ is a Markov RF with a four-nearest-node neighbourhood design. Thus, it is an Ising RF.

- (a)** Specify both the Markov and the Gibbs formulation of the RF model.
(b) Specify a McMC Metropolis–Hastings algorithm on Gibbs-sampler form to generate a realisation from $\{l(\mathbf{x}); \mathbf{x} \in L\}$. The algorithm shall be a two-neighbour-site updating algorithm, different from the single-site updating algorithm usually applied. Start with the general McMC Metropolis–Hastings algorithm when developing the algorithm.

Mosaic RF: Exercise 5

Consider a mosaic RF $\mathbf{l} : \{l(\mathbf{x}); \mathbf{x} \in L\}$, where L is a regular grid with n nodes over $D \subset \mathbb{R}^2$ and $l(\mathbf{x}) \in \Omega_l : \{W, B\}$. Consequently, the variable $l(\mathbf{x})$ takes one of the labels white (W) or black (B) for each $\mathbf{x} \in L$. The mosaic RF is represented by the n -vector \mathbf{l} .

Define the following Gibbs formulation for the mosaic RF:

$$\mathbf{l} \sim p(\mathbf{l}; \boldsymbol{\beta} = (\beta_W, \beta_B)) = \text{const}_{\beta}$$

$$\times \exp \left(\beta_W \times \sum_{i=1}^n I(l_i = W) + \beta_B \times \sum_{i=1}^n I(l_i = B) + \frac{1}{2} \times \sum_{<i,j>} I(l_i = l_j) \right),$$

where $< i, j >$ represents all pairs of nearest nodes in the grid L and $I(A)$ is an indicator function taking the value 1 whenever A is true and 0 otherwise. Here, $\boldsymbol{\beta} = (\beta_W, \beta_B) \in \mathbb{R}^2$ is a vector of unknown model parameters.

- (a)** Present an interpretation of the vector of model parameters $\boldsymbol{\beta} = (\beta_W, \beta_B)$. Demonstrate that the Gibbs formulation is over-parametrised and only depends on $\Delta\beta = (\beta_W - \beta_B)$. Use a concise notation.

- (b) Develop the corresponding Markov formulation as a function of $\Delta\beta = (\beta_W - \beta_B)$.

Discuss the most important differences between the Gibbs and the Markov formulations.

Mosaic RF: Exercise 6

Consider a two-dimensional Markov RF $\{l(\mathbf{x}); \mathbf{x} \in L\}$ where L is a regular lattice with n nodes over the domain $D \subset \mathbb{R}^2$, represented by the n -vector \mathbf{l} . Let $l(\mathbf{x}) \in \Omega_l : \{W, B\}$. Thus, the variable $l(\mathbf{x})$ belongs to one of the labels white (W) or black (B) for each $\mathbf{x} \in L$.

Define the following Gibbs formulation for the RF:

$$p(\mathbf{l}) = \text{const} \times \prod_{<i,j>} \beta^{I(l_i=l_j)},$$

where $< i, j >$ represent all pairs of nearest nodes in the grid L and $I(A)$ is an indicator function taking the value 1 whenever A is true and 0 otherwise. Hence, the RF is an Ising RF. The model parameter $\beta \in \mathbb{R}_{[1, \infty)}$ is assumed known.

Moreover, let the observations be $\{d(\mathbf{x}); \mathbf{x} \in L\}; d(\mathbf{x}) \in \mathbb{R}$, represented by the n -vector \mathbf{d} . Consider a likelihood function that is conditionally independent and with a single-site response

$$p(\mathbf{d} | \mathbf{l}) = \prod_{i \in L} p(d_i | l_i) = \prod_{i \in L} p(d_i | l_i).$$

- (a) Develop the expression for the Gibbs and Markov formulations of the posterior RF.

Consider a likelihood function that is conditionally independent with a five-site response, known as blurring. The response sites of $d(\mathbf{x})$ are the location \mathbf{x} and the four nearest sites, i.e. a cross of five sites, denoted \mathbf{o}_x . The likelihood function is then

$$p(\mathbf{d} | \mathbf{l}) = \prod_{i \in L} p(d_i | l_i) = \prod_{i \in L} p(d_i | l_j; j \in \mathbf{o}_i).$$

- (b) Develop the corresponding expression for the Gibbs and Markov formulations of the posterior RF.

Specify the neighbourhood of the Markov form graphically.

References

- Adler, R., & Taylor, J. (2007). *Random fields and geometry*. Springer.
- Asfaw, Z., & Omre, H. (2016). Localized/Shrinkage Kriging Predictors. *Mathematical Geosciences*, 48, 595–618.
- Aunon, J., & Gomez-Hernandez, J. J. (2000). Dual Kriging with local neighborhoods: Application to the representation of surfaces. *Mathematical Geology*, 32(1), 69–85.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall.
- Banerjee, S., Gelfand, A., Finley, A., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society B*, 70(4), 825–848.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 192–236.
- Borgman, L., Taheri, M., & Hagan, R. (1984). Three-dimensional, frequency-domain simulation of geological variables. In G. V. et al (Ed.), *Geostatistics for natural resources characterization* (pp. 517–541). Springer.
- Buland, A., Kolbjørnsen, O., & Omre, H. (2003). Rapid spatially coupled AVO inversion in the Fourier domain. *Geophysics*, 68(3), 824–836.
- Buland, A., & Omre, H. (2003). Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical Bayesian model. *Geophysical Prospecting*, 51, 531–550.
- Butler, K., & Vance, T. (2022). Spatial statistics for big data analytics in the ocean and atmosphere: Perspectives, challenges, and opportunities. In *Big data analytics in Earth, atmospheric, and ocean sciences* (pp. 161–168).
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Casella, G., & Berger, R. L. (2002). *Statistical inference*. Duxbury.
- Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. Wiley.
- Cootes, T., & Taylor, C. (2001). Statistical models of appearance for medical image and computer vision. *Proceedings SPIE—Medical Imaging 2001*, 4322, 236–248.
- Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, 140, 1935–1944.
- Cressie, N., & Davidson, J. (1998). Image analysis with partially ordered Markov models. *Computational Statistics and Data Analysis*, 29, 1–26.
- Cressie, N., & Johannesson, G. (2008). Fixed rank Kriging for very large spatial data sets. *Journal of the Royal Statistical Society - Series B*, 70, 209–226.

- Cressie, N., Sainsbury-Dale, M., & Zammit-Mangion, A. (2022). Basis-function models in spatial statistics. *Annual Review of Statistics and Its Application*, 9, 373–400.
- Cressie, N., & Wikle, C. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.
- Datta, A., Banerjee, S., Finley, A., & Gelfand, A. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111, 800–812.
- Diggle, P., & Giorgi, E. (2019). *Model-based geostatistics for global public health*. Chapman and Hall/CRC.
- Diggle, P., & Ribeiro, P. (2007). *Model-based geostatistics*. Springer.
- Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association*, 68(341), 117–130.
- Eidsvik, J., Martino, S., & Rue, H. (2009). Approximate Bayesian inference in spatial generalized linear mixed models. *Scandinavian Journal of Statistics*, 36(1), 1–22.
- Eidsvik, J., Shaby, B., Reich, B., Wheeler, M., & Niemi, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23, 295–315.
- Fjeldstad, T., Avseth, P., & Omre, H. (2021). A one-step Bayesian inversion framework for 3D reservoir characterization based on a Gaussian mixture model—A Norwegian Sea demonstration. *Geophysics*, 86(2), R221–R236.
- Fjellveit, M. (2021). *Image analysis - Bayesian inversion in hidden Markov models*. Master's thesis, Dept of Math Sci, NTNU.
- Forberg, O., Kjøsnes, Ø., & Omre, H. (2021). Bayesian seismic amplitude variation with offset inversion for reservoir variables with bimodal spatial histograms. *Geophysics*, 86(3), R331–R350.
- Furrer, R., Genton, M., & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3), 502–523.
- Gaetan, C., & Guyon, X. (2010). *Spatial statistics and modeling* (Vol. 90). Springer series in statistics.
- Gemanian, D. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman & Hall/CRC.
- Gelfand, A., Diggle, P., Fuentes, M., & Guttorp, P. (2010). *Handbook of spatial statistics*. CRC Press.
- Gelfand, A., & Schliep, E. (2016). Spatial statistics and Gaussian processes: A beautiful marriage, *Spatial Statistics*, 18(11), 86–104.
- Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83, 493–508.
- Gneiting, T., Kleiber, W., & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491), 1167–1177.
- Gneiting, T., & Sasvari, Z. (1999). The characterization problem for isotropic covariance functions, *Mathematical Geology*, 31(1), 105–111.
- Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732.
- Guttmann, L. (1946). Enlargement methods for computing the inverse matrix. *Annals of Mathematical Statistics*, 17(3), 336–343.
- Heaton, M., Datta, A., Finley, A., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D., Sun, F., & Zammit-Mangion, A. (2018). A case study competition among methods for analysing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3), 398–425.
- Held, K., Kops, E., Krause, B., Wells, W., Kikinis, R., & Muller-Gartner, H. (1997). Markov random field segmentation of brain MR images, *IEEE Transaction on Medical Imaging*, 16(6), 878–886.
- Hjort, N., & Omre, H. (1994). Topics in spatial statistics—with discussion. *Scandinavian Journal of Statistics*, 21, 289–357.

- Illian, J., Martino, S., Sørbye, S., Gallego-Fernandez, J., Zunzunegui, M., Esquivias, M., & Travis, M. (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution*, 4(4), 305–315.
- Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Wiley and Sons.
- Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1), 253–258.
- Jiao, J., Hu, G., & Yan, J. (2021). A Bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Sports*, 17(2), 77–90.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). *Discrete multivariate distributions* (Vol. 165). Wiley.
- Journal, A., & Huijbregts, C. J. (1978). *Mining geostatistics*. Academic Press.
- Katzfuss, M. (2017). A class of multi-resolution approximations for large spatial datasets. *Journal of the American Statistical Association*, 112, 201–214.
- Kitanidis, P. (1997). *Introduction to geostatistics: Applications in hydrogeology*. Cambridge University Press.
- Kotz, S., & Nadarajah, S. (2004). *Multivariate T-distributions and their applications*. Cambridge University Press.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 1–21.
- Li, Q., Lu, W., Yang, J., & Wang, J. (2012). Thin cloud detection of all-sky images using Markov random fields. *IEEE Geoscience and Remote Sensing Letters*, 9(3), 417–421.
- Lieshout, M. V. (2019). *Theory of spatial statistics: A concise introduction*. Chapman and Hall.
- Lindgren, F., Bolin, D., & Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running, *Spatial Statistics*, 50, 100599.
- Lindgren, F., Rue, H., & Lindstrom, J. (2011). An explicite link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society - Series B*, 73, 423–498.
- Luo, X., & Tjelmeland, H. (2019). Prior specification for binary Markov mesh models. *Statistics and Computing*, 29(2), 367–389.
- Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate analysis*. Academic Press.
- Martino, S., Pace, D. S., Moro, S., Gasoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., & Lasinio, G. J. (2021). Integration of presence-only data from several sources: A case study on dolphins' spatial distribution, *Ecography*, 44(10), 1533–1543.
- Matheron, G. (1971). *The theory of regionalized variables and its applications*. École national supérieure des mines.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.
- Møller, J., Syversveen, A., & Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3), 451–483.
- Møller, J., & Waagepetersen, R. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.
- Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24, 579–599.
- Oliver, D. (1996). On conditional simulation to inaccurate data. *Journal of Mathematical Geology*, 28(6), 811–817.
- Oliver, M. (2010). *Geostatistical applications for precision agriculture*. Springer.
- Omre, H., & Halvorsen, K. (1989). The Bayesian bridge between simple and universal Kriging. *Mathematical Geology*, 21(7), 767–786.
- Omre, H., & Rimstad, K. (2021). Bayesian Spatial Inversion and Conjugate Selection Gaussian Prior Models. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2), 420–445.
- Omre, H., & Spremick, M. (2023). The spatial kernel predictor based on huge observation sets. arXiv 1(2302.00354 [stat.ME]), 1–36.

- Rakitsch, B., Lippert, C., Borgwardt, K., & Stegle, O. (2013). It is all in the noise: Efficient multi-task Gaussian process inference with structured residuals. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26). Curran Associates.
- Rimstad, K., Avseth, P., & Omre, H. (2012). Hierarchical Bayesian lithology/fluid prediction: A North Sea case study. *Geophysics*, 77(2), B69–B85.
- Ripley, B., & Kelly, F. (1977). Markov point processes. *Journal of the London Mathematical Society*, s2–15(1), 188–192.
- Røislien, J., & Omre, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data. *Mathematical Geology*, 38(7), 821–849.
- Roksvåg, T., Steinsland, I., & Engeland, K. (2022). A geostatistical spatially varying coefficient model for mean annual runoff that incorporates process-based simulations and short records. *Hydrology and Earth System Sciences*, 26, 5391–5410.
- Royer, J. J., & Vieira, P. C. (1984). Dual formalism of Kriging. In G. V. et al. (Ed.), *Geostatistics for natural resources characterization* (Vol. 2, pp. 691–702). Reidel.
- Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*, Vol. 104 of Monographs on statistics and applied probability. Chapman & Hall.
- Rue, H., & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference*, 137(10), 3177–3192. Special Issue: Bayesian Inference for Stochastic Processes.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society - Series B*, 71(2), 319–392.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer series in statistics. Springer.
- Stein, M. L. (2013). Statistical properties of covariance tapers. *Journal of Computational and Graphical Statistics*, 22, 866–885.
- Swendsen, R., & Wang, J. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters—American Physical Society*, 58(2), 86–88.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Taylor, B., & Diggle, P. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, 84(10), 2266–2284.
- Tierney, L., Kass, R., & Kadane, J. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84(407), 710–716.
- Tjelmeland, H., & Besag, J. (1998). Markov random fields with higher-order interactions. *Scandinavian Journal of Statistics*, 25, 415–433.
- Trefethen, L., & Bau, D. (1997). *Numerical linear algebra*. SIAM.
- van Niekerk, J., Krainski, E., Rustand, D., & Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 181, 107692.
- van Niekerk, J., & Rue, H. (2023). Low-rank variational Bayes correction to the Laplace method. arXiv (2111.12945 [stat.ME]).
- Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 1, 5–42.
- Vigsnes, M., Kolbjørnsen, O., Hauge, V. L., Dahle, P., & Abrahamsen, P. (2017). Fast and accurate approximation to Kriging using common data neighborhoods. *Mathematical Geosciences*, 49, 619–634.
- Winkler, G. (2006). *Image analysis, random fields and Markov chain Monte Carlo methods: A mathematical introduction*. Springer.
- Yaglom, A. (1987). *Correlation theory of stationary and related random functions*. Springer.