ISBN 978-82-471-4527-2 (printed version)
ISBN 978-82-471-4528-9 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2013:206

Xiangping Hu

Multivariate Gaussian Random Fields

The Stochastic Partial Differential Equation Approach

Xiangping Hu

Doctoral Thesis

Faculty of Information Technology
Mathematics and Electrical Engineering
Department of Mathematical Sciences

NTNU – Trondheim
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2013:206

NTNU

NTNU – Trondheim
Norwegian University of
Science and Technology

Xiangping Hu

Multivariate Gaussian Random Fields

The Stochastic Partial Differential Equation Approach

Thesis for the degree of Philosophiae Doctor

Trondheim, August 2013

Faculty of Information Technology
Mathematics and Electrical Engineering
Department of Mathematical Sciences

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
Mathematics and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-471-4527-2 (printed version)
ISBN 978-82-471-4528-9 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2013:206

Printed by Skipnes Kommunikasjon as

献给我的父母

To my parents

Til foreldrene mine

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) at the Department of Mathematical Sciences at the Norwegian University of Science and Technology (NTNU). The research was carried out at the Department of Mathematical Sciences at NTNU, with funding from the Faculty of Information Technology, Mathematics and Electrical Engineering, NTNU.

I would like to take some time to thank all people who have helped me during these years at NTNU. Without them, it would have been very hard for me to complete my degree smoothly.

First of all, I would like to thank my principle supervisor Professor Håvard Rue for his support and guidance during these 4 years. Many thanks for being patient with my lack of background in spatial statistics. He always encouraged me to overcome the challenges in the research. I consider myself very lucky to have the opportunity to join his group and do research under his supervision.

Second I would like to thank my co-supervisor Professor Ingelin Steinsland for her guidance and fruitful discussions, especially during the last year. She gave me many precious suggestions for how to do a good research and how to write good articles.

I want to express my deep thanks to Dr. Daniel Simpson and Dr. Finn Lindgren for helping me all the time. They have been patient to answer my silly questions. The weekly meetings with them were always fruitful. Whenever I was stuck on some point, they were always available to help me out.

A personal thank goes to PhD candidate Geir-Arne Fuglstad at NTNU. We had many small useful discussions, and I really enjoyed the discussions. He also gave many valuable comments on my papers at the last stage.

Dr. Sara Martino provided the dataset for modeling. She gave much help with the practical issues with the dataset. It took her much time to come to the department and discuss the results. Many thanks for that.

In Trondheim, I have many good colleagues. Thanks to Rikke Ingebrigtsen, Thiago Guerrera Martins, Håvard Wahl Kongsgård and Rupali Akerkar. They have given many useful comments on my drafts. I really enjoyed working within the group. The get-together parties were always very nice and we had so many interesting topics to discuss. I really enjoyed the discussion.

There are many Chinese friends who helped me a lot, which I would like to thank. Thanks to Kangqiao Zhao, Linghua Chen, Long Pei, Yu Zhou, Jie Xu, Liang Wang, Shengjie Lu, Zhenwei Guo, Ye Xu, Peng Wang, Zhiqi Feng, Xin Luo and many others. They helped me a lot during these years.

I also want to thank the computer support group and the administrative staff, especially Per Kristian Hove and Anne M. Kajander, in the Department of Mathematical Sciences at NTNU for their support.

Warm and cordial thanks are given to my parents, my brother and my sisters. They supported and believed in me all the time. Even though I am abroad, I feel I am at home whenever I talk to them on the mobile phone.

Finally, a special thanks to my wife, Taohong Liao, for always being at my side. Thanks for all her support and help over the years. With her love my life becomes more colorful and beautiful.

July 1, 2013, Trondheim, Norway
Xiangping Hu

Introduction and Summary

Introduction and Summary

1 Motivation

Gaussian random fields (GRFs) have a dominant role in spatial modelling and there exists a wealth of application areas, such as geostatistics, atmospheric and environment science (Cressie, 1993; Stein, 1999; Diggle and Ribeiro Jr, 2006). GRFs are practical since the normalizing constant can be computed explicitly. A GRF also has good properties since it can be explicitly specified through its mean function $\boldsymbol{\mu}(\cdot)$ and covariance function $\boldsymbol{C}(\cdot,\cdot)$. In $\mathbb{R}^d$, with $\boldsymbol{s} \in \mathbb{R}^d$, a random field $x(\boldsymbol{s})$ is a continuously indexed GRF if any collection $\{x(\boldsymbol{s}_i); i = 1, 2, \ldots, n\}$ for every $n \geq 1$ jointly has Gaussian distribution.

There are two main challenges in spatial statistics for traditional approaches associated with GRFs. The first challenge is a theoretical challenge since the covariance function of a GRF must be symmetric positive definite (SPD). However, it is difficult to specify a parametric model for the covariance function (Gneiting et al., 2010; Kleiber and Nychka, 2012; Apanasovich et al., 2012). The second challenge is a computational challenge. This challenge is due to the cost of $\mathcal{O}(n^3)$ to factorize a dense $n \times n$ dimensional matrix. Therefore, it is sometimes infeasible to do full Bayesian analysis with large datasets , i.e., when n is large (Banerjee et al., 2004).

Lindgren et al. (2011) proposed a new approach for spatial modelling with GRFs. They showed an explicit link between GRFs and Gaussian Markov random fields (GMRFs) by using stochastic partial differential equations (SPDEs). Using this link it is possible to build the models with GRFs, but do computations with GMRFs. A GMRF is a GRF with Markovian property, and its precision matrix is usually sparse. Due

to the sparse structure of the precision matrix, numerical algorithms for sparse matrix can be applied. Therefore, the computational challenges are partially solved, and hence it is possible to do analysis for large datasets. Lindgren et al. (2011) also claimed that there is no need to consider the positive definite constraint to the covariance matrix since the constructed covariance matrix automatically will be positive definite.

Spatial modelling of multivariate spatial datasets are in demand in many areas, such as in economics (Gelfand et al., 2004; Sain and Cressie, 2007), in the area of air quality (Brown et al., 1994; Schmidt and Gelfand, 2003), weather forecasting (Courtier et al., 1998; Reich and Fuentes, 2007) and quantitative genetics (Mcguigan, 2006; Konigsberg and Ousley, 2009). Many approaches were proposed to model this kind of dataset, such as Linear model of coregionalization (LMC) (Goulard and Voltz, 1992; Wackernagel, 2003; Gel et al., 2004) and covariance-based models (Apanasovich and Genton, 2010; Gneiting et al., 2010; Li and Zhang, 2011; Kleiber and Nychka, 2012; Apanasovich et al., 2012). As pointed out by Diggle and Ribeiro Jr (2006, Chapter 3.12) that without some restrictions models of this kind are usually very poorly identifiable. Covariance-based models have both the theoretical challenge and the computational challenge as discussed above.

In this thesis we extend the approach proposed by Lindgren et al. (2011) to *multivariate* GRFs by using systems of SPDEs. The constructed multivariate GRFs inherit the theoretical and computational advantages. On the theoretical side, the covariance functions of the multivariate GRFs are automatically positive definite. On the computational side, we can use GMRF representations for GRFs when doing computations.

The rest of the introduction of this thesis is organized as follows. First we give some fundamental definitions in Section 2. Then we discuss two main challenges in spatial statistics and review some of the proposed solutions for these problems in Section 3. We give an overview of the state-of-art research of the SPDE approach in Section 4. Some of the methodology for constructing multivariate GRFs are discussed in Section 5. In Section 6 we give some examples for sampling the multivariate GRFs constructed with our proposed systems of SPDEs. Current challenges and possibilities for further research are discussed in Section 7. Finally we summarize the included papers in Section 8.

2 Fundamental definitions

In this section we give some useful definitions of terms which are used extensively in the thesis. Let $\mathcal{D}$ denote a topological space and Ω denote some abstract sample space.

Definition 2.1. *A random field $x(\boldsymbol{s})$ on $\mathcal{D}$ is a random function specified by its finite dimensional joint distributions*

$$F(y_1, \ldots, y_n; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = P(x(\boldsymbol{s}_1) \leq y_1, \ldots, x(\boldsymbol{s}_n) \leq y_n),\ s \in \mathcal{D}$$

for every $n \geq 1$ and every collection $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$ of sites in $\mathcal{D}$.

In Definition 2.1, we have used the notation

$$x(\boldsymbol{s}) = x(\boldsymbol{s}, \omega),\ \ \boldsymbol{s} \in \mathcal{D},\ \ \omega \in \Omega.$$

The focus of the thesis is modelling spatial datasets, and therefore $\mathcal{D} \subset \mathbb{R}^d$ or $\mathbb{S}^2$. The Kolmogorov existence theorem provides the theoretical foundation for determining whether a random field model is valid or not. This theorem states that if the collection of finite-dimensional distributions is consistent under permutations of the sites and marginalization then the random field model is valid. We refer to Billingsley (1986) for a detailed discussion.

Definition 2.2. *A random field $x(\boldsymbol{s})$ on $\mathcal{D}$ is called a Gaussian random field if any subset $\{x(\boldsymbol{s}_1), x(\boldsymbol{s}_2), \ldots, x(\boldsymbol{s}_n)\}$ for $n \geq 1$, where $\boldsymbol{s}_i \in \mathcal{D}$ for $i = 1, 2, \ldots, n$, has a multivariate Gaussian distribution.*

From Definition 2.2, we can notice that any finite dimensional distribution of the field $x(\boldsymbol{s})$ is multivariate Gaussian. A Gaussian random field can be specified by its mean $\mu(\boldsymbol{s}) = \mathrm{E}(x(\boldsymbol{s}))$ and covariance function $C(\boldsymbol{t}, \boldsymbol{u}) = \mathrm{Cov}(x(\boldsymbol{t}), x(\boldsymbol{u}))$. The covariance function must be symmetric positive definite. Let T denote transpose of vectors and matrices.

Definition 2.3. *A matrix $\boldsymbol{A}$ is called positive definite if*

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}. \tag{1}$$

If $\boldsymbol{A}$ is also symmetric, then it is called a symmetric positive definite (SPD) matrix. In this thesis we denote an SPD matrix by $\boldsymbol{A} > \boldsymbol{0}$. Some important properties of this kind of matrix can be found in Rue and Held (2005, Section 2.1.6).

Definition 2.4. *A multivariate Gaussian random field $\boldsymbol{x}(\boldsymbol{s})$ is a collection of Gaussian random fields $\{x_i; i = 1, 2, \ldots, p\}$ such that any collection $\{\boldsymbol{x}(\boldsymbol{s}_1), \boldsymbol{x}(\boldsymbol{s}_2), \ldots, \boldsymbol{x}(\boldsymbol{s}_n)\}$, where $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \ldots, x_p(\boldsymbol{s}))^T$, for all $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ for all $n \geq 1$, has a multivariate Gaussian distribution determined by*

$$\begin{aligned} \mathrm{E}(\boldsymbol{x}(\boldsymbol{s})) &= \boldsymbol{\mu}(\boldsymbol{s}), & \boldsymbol{s} \in \mathcal{D}, \\ \mathrm{Cov}\,(\boldsymbol{x}(\boldsymbol{t}), \boldsymbol{x}(\boldsymbol{u})) &= \boldsymbol{C}(\boldsymbol{x}(\boldsymbol{t}), \boldsymbol{x}(\boldsymbol{u})), & \boldsymbol{t}, \boldsymbol{u} \in \mathcal{D}, \end{aligned}$$

where $\boldsymbol{\mu} : \mathcal{D} \to \mathbb{R}^p$ is the mean function and $\boldsymbol{C} : \mathcal{D} \to \mathbb{R}^{p \times p}$ is a matrix-valued covariance function.

Multivariate GRFs are commonly used in spatial statistics in many different areas. The covariance function of a GRF or a multivariate GRF must fulfill the positive definiteness constraint. Due to the positive definite constraint it is difficult to specify a valid parametric model for cross-covariance function.

Definition 2.5. *A matrix $\boldsymbol{A}$ is called nonnegative definite if and only if*

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}. \tag{2}$$

Nonnegative definite matrix is also called positive semidefinite matrix. We denote the nonnegative definite matrix as $\boldsymbol{A} \geq \boldsymbol{0}$ in this thesis.

Definition 2.6. *A random field $x(\boldsymbol{s})$ is called weakly stationary if the mean function $\mu(\boldsymbol{s})$ is constant and the covariance function satisfies the following relationship*

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2).$$

Weakly stationary is also called second-order stationary or covariance stationary (Lindgren, 2010). In this thesis when we say that a random field is stationary, we mean it is weakly stationary unless otherwise

specified. Isotropic random fields are important weakly stationary random fields. The covariance functions for isotropic random fields fulfill $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|) = C(\|\boldsymbol{h}\|)$, where $\|\cdot\|$ denotes the Euclidean norm. These covariance functions depend only on the distance between the sites but not the direction. The covariance functions constructed with the systems of SPDEs in this thesis belong to this group. One of the important and popular families of covariance functions is the Matérn family (Matérn, 1960, 1986).

Definition 2.7. *The Matérn covariance function has the following form*

$$M(\boldsymbol{h}) = \frac{2^{1-\nu}\sigma^2}{\Gamma(\nu)}(\kappa\|\boldsymbol{h}\|)^\nu K_\nu(\kappa\|\boldsymbol{h}\|), \sigma^2 > 0, \nu > 0, \kappa > 0, \tag{3}$$

where K_ν is the modified Bessel function of second kind.

In this covariance function σ^2 is the marginal variance, κ is a scaling parameter and $\Gamma(\cdot)$ is the gamma function. ν is a smoothness parameter and it defines the Hausdorff dimension and the differentiability of the sample paths. But unfortunately this parameter is poorly identifiable from the data and hence we will fix it during inference (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011). This covariance function contains the popular exponential covariance function since $M(\boldsymbol{h}) = \sigma^2 \exp(-\kappa\|\boldsymbol{h}\|)$ when $\nu = \frac{1}{2}$. Matérn covariance function plays a key role in this thesis. We return to it in Section 4.

The covariance function for a second-order stationary multivariate GRF has the form

$$\boldsymbol{C}(\boldsymbol{h}) = \begin{pmatrix} C_{11}(\boldsymbol{h}) & C_{12}(\boldsymbol{h}) & \cdots & C_{1p}(\boldsymbol{h}) \\ C_{21}(\boldsymbol{h}) & C_{22}(\boldsymbol{h}) & \cdots & C_{2p}(\boldsymbol{h}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{h}) & C_{p2}(\boldsymbol{h}) & \cdots & C_{pp}(\boldsymbol{h}) \end{pmatrix}, \tag{4}$$

where $C_{ii}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s} + \boldsymbol{h})x_i(\boldsymbol{s}))$ is the covariance function within field $x_i(\boldsymbol{s})$ and $C_{ij}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s} + \boldsymbol{h})x_j(\boldsymbol{s}))$ is the cross-covariance function $C_{ij}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s} + \boldsymbol{h})x_j(\boldsymbol{s}))$ between fields $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$.

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ have a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Define a labeled

graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ with vertexes $\mathscr{V} = \{1, 2, \ldots, n\}$ and edges $\mathscr{E}$ for $\boldsymbol{x}$ such that there is no edge between i and j if and only if $x_i \perp x_j | \boldsymbol{x}_{-ij}$, where $\boldsymbol{x}_{-ij}$ denotes $\boldsymbol{x}_{-\{i,j\}}$.

The inverse of the covariance matrix is called precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ and we also need the constraint $\boldsymbol{Q} > 0$. The precision matrix plays a key role when we deal with GMRFs and this is because the precision matrix for a GMRF is usually sparse,

$$x_i \perp x_j | \boldsymbol{x}_{-ij} \Longleftrightarrow Q_{ij} = 0 \text{ for } i \neq j. \tag{5}$$

This simply means that we can know whether x_i and x_j are conditionally independent or not from the graph $\mathscr{G}$ since the nonzero pattern of $\boldsymbol{Q}$ determines $\mathscr{G}$ (Rue and Held, 2005). We now give the formal definition of a GMRF.

Definition 2.8. *A random vector* $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \ldots, x_n(\boldsymbol{s}))^T \in \mathbb{R}^n$ *is called a Gaussian Markov random field with mean value* $\boldsymbol{\mu}$ *and precision matrix* $\boldsymbol{Q}$ *with respect to a labeled graph* $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ *if and only if it has the probability density function*

$$\pi(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{6}$$

and

$$Q_{ij} \neq 0 \iff \{i, j\} \in \mathscr{E} \text{ for all } i \neq j.$$

From Definition 2.8, we can notice that if $\boldsymbol{Q}$ is a completely dense matrix then $\mathscr{G}$ is fully connected.

3 Two main challenges in spatial statistics

Two main challenges in spatial statistics consists of a theoretical challenge and a computational challenge. For the theoretical challenge, the covariance function of a proposed model must be positive definite. For the computational challenge, we need to construct computationally efficient model in order to do full Bayesian analysis for large datasets.

3.1 Positive definite constraint

Let $C(\boldsymbol{h})$ be a covariance function for a GRF, then we have the requirement $C(\boldsymbol{h}) > 0$. This is a well-known challenge since specifying a valid parametric model for the covariance function is difficult. Bochner's theorem has characterized all the continuous covariance functions for random fields on $\mathbb{R}^d$.

Theorem 3.1 (Bochner's theorem)**.** *Denote a continuous function, real or complex, as $C(\boldsymbol{h})$. This function is positive definite if and only if there is some positive and symmetric measure $\boldsymbol{F}$ such that*

$$C(\boldsymbol{h}) = \int_{-\infty}^{+\infty} \exp(i\boldsymbol{h}^T\boldsymbol{k}) d\boldsymbol{F}(\boldsymbol{k}). \tag{7}$$

Detailed description about this theorem can be found in Stein (1999, Chapter 2.5). A proof of Bochber's theorem can be found in Cox and Miller (1977). We call Equation (7) the spectral representation of the covariance function. If the measure $\boldsymbol{F}$ has a Lebesgue density $\boldsymbol{S}$, this is called the spectral density. The spectral density of the Matérn covariance function given in Equation (3) is

$$\boldsymbol{S}(\boldsymbol{k}) = \frac{\sigma^2}{(2\pi)^d(\kappa^2 + \|\boldsymbol{k}\|^2)^{\nu+d/2}}. \tag{8}$$

The spectral representation is widely used in this thesis in order to investigate the properties of the multivariate GRFs constructed from our proposed approach.

In this thesis we show that the theoretical challenge is solved for models constructed by our proposed SPDEs approach since the covariance functions are automatically positive definite. We return to this in Section 5.

3.2 Computational efficiency

As mentioned in Section 1, the second challenge is due to the cost of $\mathcal{O}(n^3)$ to factorize a dense $n \times n$ matrix. When n becomes large, it becomes infeasible to get results in reasonable time (Banerjee et al., 2004). For this challenge, many different approaches have been proposed, such as using

separable covariance functions (Fuentes, 2006; Genton, 2007), covariance tapering (Furrer et al., 2006; Zhang and Du, 2008; Kaufman et al., 2008; Shaby and Ruppert, 2012), fixed rank kriging and fixed rank filtering (Cressie and Johannesson, 2008; Cressie et al., 2010), likelihood approximations (Vecchia, 1988; Stein et al., 2004) and GMRF approximations (Rue and Tjelmeland, 2002; Rue et al., 2004; Rue and Held, 2005; Lindgren et al., 2011). We give a brief overview about these approaches and refer to Sun et al. (2012) for a detailed summary of different approaches for dealing with large datasets in geostatistics.

3.2.1 Separable covariance functions

The main idea of this approach is to take advantage of special structures of the covariance functions, and rewrite the covariance function in a separable form (Fuentes, 2006; Genton, 2007). One of the commonly used separable functions has the form $C(\boldsymbol{s}, t) = C(\boldsymbol{s}, 0)C(\boldsymbol{0}, t)/C(\boldsymbol{0}, 0)$. This kind of function is commonly used for spatio-temporal models. With this form we assume that the space-time covariance function can be treated as a product of a spatial covariance function and a temporal covariance function. The covariance matrix for this kind of model can then be written as a Kronecker products of two smaller matrices. Therefore, nice properties of Kronecker product can be applied in order to achieve computational efficiency. For more information about the Kronecker product, see for example Graham (1982). Another separate form used for spatial or spatial-temporal models is $\boldsymbol{C}(\boldsymbol{k}) = \rho(\boldsymbol{k})\boldsymbol{A}$, where $\rho(\boldsymbol{k})$ is a correlation function and $\boldsymbol{A}$ is a positive definite matrix. This form has the assumption that the covariance function is independent of the spatial locations. Li et al. (2007) provided a nonparametric assessment for evaluating the separability of covariance structures.

3.2.2 Covariance tapering

The main aim of covariance tapering technique is to make a sparser and better structured covariance matrices and hence numerical algorithms for sparse matrices can be used for fast statistical inference (Furrer et al., 2006; Zhang and Du, 2008; Kaufman et al., 2008; Shaby and Ruppert, 2012). Assuming the vector $\boldsymbol{x}$ follows a multivariate Gaussian distribution with

parameter $\boldsymbol{\theta}$, the log-likelihood of the vector has the form

$$\log(\pi(x|\boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{C}(\boldsymbol{\theta})| - \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{\theta})\boldsymbol{x}, \tag{9}$$

where n is the length of the vector $\boldsymbol{x}$. We need to evaluate $\boldsymbol{Q}(\boldsymbol{\theta}) = \boldsymbol{C}^{-1}(\boldsymbol{\theta})$ and $|\boldsymbol{C}(\boldsymbol{\theta})|$. The common approach is to use the Cholesky factorization of precision matrix $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$, where the Cholesky triangle $\boldsymbol{L}$ is a lower triangular matrix. The covariance tapering technique is based on the fact that the correlations between observations far enough apart become negligible, and hence can be set to 0. The commonly used approach for tapering the covariance function is the formula

$$\hat{\boldsymbol{C}}(h;\boldsymbol{\theta},\gamma) = \boldsymbol{C}(h;\boldsymbol{\theta})\boldsymbol{C}_t(h;\gamma), \quad h > 0, \quad \gamma > 0, \tag{10}$$

where γ is a pre-defined threshold and h is the distance between two observations. $\boldsymbol{C}_t(h;\gamma)$ is the tapering function, and it is an isotropic correlation function with compact support. It also has the additional properties that $\boldsymbol{C}_t(h;\gamma) = 0$ for any $h \geq \gamma$. $\hat{\boldsymbol{C}}(h;\boldsymbol{\theta},\gamma)$ is the tapered covariance function. The tapered covariance matrix $\hat{\boldsymbol{C}}$ is a sparse matrix and also a positive definite matrix. Therefore, we save computational resources.

3.2.3 Fixed rank kriging

The fixed rank kriging (FRK) method was proposed by Cressie and Johannesson (2006). Shi and Cressie (2007) applied this approach to analyze the MISR aerosol data. The main idea of the FRK approach is trying to use a set of m basis functions, $\boldsymbol{B}(\boldsymbol{s}) = (B_1(\boldsymbol{s}), B_2(\boldsymbol{s}), \ldots, B_m(\boldsymbol{s}))^{\mathrm{T}}$ to capture the spatial dependence, and hence the covariance matrix of $\boldsymbol{x}$ can be written as

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{K}\boldsymbol{B}^{\mathrm{T}} + \tau^2\boldsymbol{V}, \tag{11}$$

where $\boldsymbol{K}$ is any $m \times m$ positive definite matrix with unknown parameters, and $\tau^2\boldsymbol{V}$ is a diagonal matrix with entries given by the measurement-error variances. The inverse of the covariance matrix then turns into the inverse of $m \times m$ positive definite matrix. Therefore, we can save computational resources. For more information about this approach we refer to Cressie and Johannesson (2008) and Sun et al. (2012). Cressie

et al. (2010) discussed a new approach called fixed rank filtering for spatio-temporal data.

3.2.4 Likelihood approximation

Approximating the likelihood function is another way to achieve computational efficiency. This technique can be applied both in the spatial domain and in the spectral domain. Vecchia (1988) and Stein et al. (2004) discussed approaches to write the conditional distribution based only on some of the 'past' observations which can lessen the computational burden. Fuentes (2007) discussed how to use Whittle's approximation of the Gaussian negative log-likelihood function for irregularly spaced data with a lattice process. They claimed that they can save computational resources by truncating the spectral representation. They pointed out that likelihood approximation in the spectral domain does not require the calculation of the determinants.

3.2.5 GMRF approximations

GMRF models are widely used due to their good Markovian property. Due to the Markovian property, the precision matrices for GMRFs are usually sparse. Rue and Tjelmeland (2002) discussed how to use GMRFs to fit GRFs. They argued that the most commonly used GRFs used in geostatistics can be approximated by GMRFs, and hence GMRFs can be used as a computational replacement for GRFs. This kind of approach was also used by Allcroft and Glasbey (2003) for spatio-temporal models. Hartman and Hössjer (2008) suggested the use of GMRFs for GRFs when doing spatial prediction using Kriging, from the computational point of view.

In this thesis the GMRFs approximation to GRFs discussed by Lindgren et al. (2011) is adopted and widely used. We apply GMRF approximations to the multivariate GRFs constructed from the system of SPDEs, and hence the precision matrices are sparse. Since it is crucial for our computations, we discuss this approach in Section 4.2 in detail.

4 SPDE approach for modelling GRF

In this section we give an overview of the state-of-the-art research for the SPDE approach for spatial statistics. The GMRF approximation is also introduced here. The discussions focus on the univariate GRFs constructed with the SPDE approach.

4.1 SPDE approach

The new approach discussed by Lindgren et al. (2011) was to use SPDEs to construct GRFs, which showed an explicit link between GRFs and GMRFs. The SPDE has the form

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \nu > 0, \tag{12}$$

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator and $\mathcal{W}(\boldsymbol{s})$ is a standard Gaussian white noise process. Δ is the *Laplacian* given by

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

It turns out that the stationary solution $x(\boldsymbol{s})$ to the linear (fractional) SPDE (12) is a GRF with Matérn covariance function given in Equation (3). We can also get the closed form for the marginal variance for the random field,

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)} (4\pi)^{-d/2} \kappa^{-2\nu}.$$

Following the terminology from Lindgren et al. (2011), we call a GRF with Matérn covariance function a Matérn random field. Lindgren et al. (2011) pointed out that the Matérn fields are the only stationary solutions to the SPDE (12). Moreover, even though the solutions to SPDE (12) do not have Matérn covariance function as $\kappa \to 0$ or $\nu \to 0$, the solution for the SPDE with $\kappa = 0$ or $\nu = 0$ still has a well-defined measure (Lindgren et al., 2011). Denote the Fourier transform of a function by $\mathscr{F}(\cdot)$. We apply the Fourier transform to SPDE (12)

$$\left\{ \mathscr{F}(\kappa^2 - \Delta)^{\alpha/2} \phi \right\} (\boldsymbol{k}) = \left(\kappa^2 + \|\boldsymbol{k}\|^2 \right)^{\alpha/2} (\mathscr{F}\phi)(\boldsymbol{k}), \tag{13}$$

where $\boldsymbol{k}$ is the frequency, and ϕ is a smooth, rapidly decaying function in $\mathbb{R}^d$. We can obtain the spectral density of the stationary solution

$$\boldsymbol{S}(\boldsymbol{k}) = \frac{1}{(2\pi)^d(\kappa^2 + \|\boldsymbol{k}\|)^\alpha}. \tag{14}$$

Lindgren et al. (2011) discussed some possible extensions beyond the classical Matérn models, such as Matérn fields on manifolds, random fields with oscillating covariance functions, non-stationary fields, and non-separable space-time models. Some extensions of the SPDE approach given by other researchers is discussed in Section 4.3.

4.2 GMRF approximation

In this paper we follow the approach presented in Lindgren et al. (2011) for GMRF approximations to GRFs in order to partially resolve the "*big n problem*" (Banerjee et al., 2004) with our models. Using SPDE (12) the model is built theoretically from a discretized GRF on a set of locations $\boldsymbol{s}_i$ with a covariance matrix $\boldsymbol{\Sigma}$, and then we use a GMRF to represent the GRF. This means the precision matrix of the GMRF fulfills the condition $\boldsymbol{Q^{-1}} \simeq \boldsymbol{\Sigma}$ on some predefined norm.

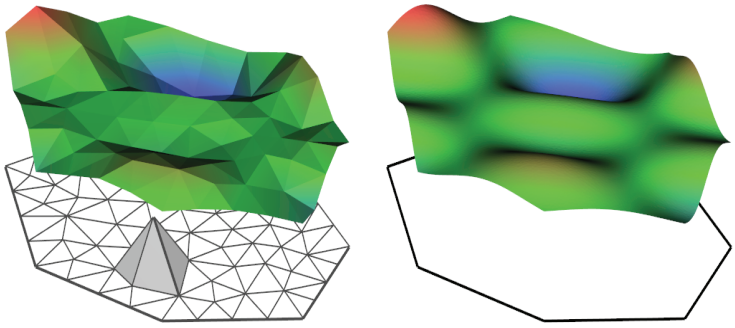
The first step in constructing the GMRF representation for $x(\boldsymbol{s})$ on the triangulated lattice is to find the stochastic weak formulation of Equation (12) (Kloeden and Platen, 1999). In this paper Delaunay triangulation is chosen, and we refer to Hjelle and Dæhlen (2006) for detailed discussion about Delaunay triangulations. Denote the inner product of functions h and g as

$$\langle h, g\rangle = \int h(\boldsymbol{s})g(\boldsymbol{s})d(\boldsymbol{s}), \tag{15}$$

where the integral is taken over the region of interest. We can find the stochastic weak solution of SPDE (12) by requiring that

$$\left\{\langle \phi_k, (\kappa^2 - \Delta)^{\alpha/2}x\rangle\right\}_{i=1}^{M} \stackrel{d}{=} \{\langle \phi_k, \mathcal{W}_i\rangle\}_{i=1}^{M}. \tag{16}$$

In the second step we need to construct a finite element representation of the solution to the SPDE. We refer to Zienkiewicz et al. (2005) and



(a) (b)

Figure 1: Approximation to a surface with piecewise linear basis function. The approximated surface is given in (a) and the true surface is given in (b). The pyramid in (a) is one basis function.

Bathe (2008) for more information about finite element methods. The finite element representation of the solution to SPDE (12) is

$$x(\boldsymbol{s}) = \sum_{k=1}^{N} \psi_k(\boldsymbol{s})\omega_k, \tag{17}$$

where $\psi_k(\boldsymbol{s})$ is some chosen basis function, ω_k is some Gaussian distributed weight, and N is the number of the vertexes in the triangulation. The basis-function $\psi_k(\boldsymbol{s})$ is chosen to be piecewise linear on each triangle with value 1 at vertex k and 0 at all other vertexes. This means that a continuously indexed solution is approximated by a piecewise linear function defined through the joint distribution of $\omega_k, k = 1, 2, \ldots, N$ (Lindgren et al., 2011). The chosen basis-functions ensure that the local interpolation on a triangle is piecewise linear. Figure 1 illustrated one example of a surface approximated by a piecewise linear approximation.

The third step is to choose the test functions. In this paper we follow the setting used by Lindgren et al. (2011). With $M = N$ the test functions are chosen as $\phi_k = (\kappa^2 - \Delta)^{1/2}\psi_k$ for $\alpha = 1$ and $\phi_k = \psi_k$ for $\alpha = 2$, which are denoted as a *least squares* and a *Galerkin* solution, respectively. When $\alpha \geq 3$, the approximation can be obtained by setting $\alpha = 2$ at the left-hand

side and replacing the right hand side with a field generated with $\alpha - 2$ and let $\phi_k = \psi_k$. This iteration procedure terminates when $\alpha = 1$ or 2. This is the essence of the recursive Galerkin formulation. More detailed description can be found in Lindgren et al. (2011). α can currently only be integer-valued. When α is not an integer, different approximation methods must be used and this is beyond of scope of our discussion. The response to the discussion in Lindgren et al. (2011) discussed fractional α. We define the required $N \times N$ matrices $\boldsymbol{C}$, $\boldsymbol{G}$ and $\boldsymbol{K}_{\kappa^2}$ with entries

$$\begin{aligned} C_{mn} &= \langle \psi_m, \psi_n \rangle, \\ G_{mn} &= \langle \nabla\psi_m, \nabla\psi_n \rangle, \\ (\boldsymbol{K}_{\kappa^2})_{mn} &= \kappa^2 C_{mn} + G_{mn}, \end{aligned} \tag{18}$$

for $m, n = 1, 2, \ldots, N$.

Finally, by using these matrices together with a Neumann boundary condition (which means that at the boundary the normal derivative is zero), the precision matrices $\boldsymbol{Q}$ for $x(\boldsymbol{s})$ can be obtained,

$$\begin{cases} \mathbf{Q}_{1,\kappa^2} & = \mathbf{K}_{\kappa^2} & \text{for } \alpha = 1, \\ \mathbf{Q}_{2,\kappa^2} & = \mathbf{K}_{\kappa^2}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{K}_{\kappa^2} & \text{for } \alpha = 2, \\ \mathbf{Q}_{\alpha,\kappa^2} & = \mathbf{K}_{\kappa^2}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{Q}_{\alpha-2,\kappa^2} \mathbf{C}^{-1} \mathbf{K}_{\kappa^2}, & \text{for } \alpha = 3, 4, \cdots . \end{cases} \tag{19}$$

As pointed out by Lindgren et al. (2011) the inverse of $\boldsymbol{C}$, i.e., $\boldsymbol{C}^{-1}$, is usually a dense matrix. This causes the precision matrix to be dense and ruins all the effort we have made. So this matrix is replaced with a diagonal matrix $\tilde{\boldsymbol{C}}$, where $\tilde{\boldsymbol{C}}_{ii} = \langle \psi_i, 1 \rangle$. Using $\tilde{\boldsymbol{C}}_{ii}$ instead of $\boldsymbol{C}$ yields a Markov approximation to the FEM solution. The effects of the Markov approximation have been studied by Bolin and Lindgren (2009). They claimed that the difference between the exact representation through the finite element method and the Markov approximation is negligible.

4.3 Current research on the SPDE approach

We give an overview of the state-of-art of the current research win the SPDE approach.

4.3.1 Nested SPDEs

Bolin and Lindgren (2011) extended the ideas from Lindgren et al. (2011) to build a richer class of random fields using nested SPDEs. The nested SPDEs have the form

$$\mathcal{L}_1 x_0(\boldsymbol{s}) = \mathcal{L}_2 \mathcal{W}(\boldsymbol{s}), \tag{20}$$

where $\mathcal{L}_1$ and $\mathcal{L}_2$ are linear differential operators. If these two operators are commutative, then we can rewrite Equation (20) as

$$\begin{aligned} \mathcal{L}_1 x_0(\boldsymbol{s}) &= \mathcal{W}(\boldsymbol{s}), \\ x(\boldsymbol{s}) &= \mathcal{L}_2 x_0(\boldsymbol{s}). \end{aligned} \tag{21}$$

Bolin and Lindgren (2011) pointed out that system of equations (21) can give us an interpretation of the effects of operator $\mathcal{L}_2$. Furthermore, they claimed that a larger class of models can be constructed by increasing the order of operators $\mathcal{L}_1$ and $\mathcal{L}_2$. The operators they discussed have the form

$$\begin{aligned} \mathcal{L}_1 &= \prod_{i=1}^{n_1} \left((\kappa_i^2 - \Delta)^{\alpha_i/2} \right), \\ \mathcal{L}_2 &= \prod_{i=1}^{n_2} \left(c_i + D_i^{\mathrm{T}} \nabla \right), \end{aligned}$$

with $\alpha_i \in \mathbb{N}$, $\kappa_i^2 > 0$, $c_i \in \mathbb{R}, D_i \in \mathbb{R}^d$. ∇ is the gradient and Δ is the Laplacian operator. Similarly as in Lindgren et al. (2011, Section 3.2), nonstationary models can be obtained with spatially varying parameters. We refer to their original paper for more properties of their models. Bolin and Lindgren (2011) pointed out that this approach is related to the approaches discussed by Jun and Stein (2007) and Jun and Stein (2008).

4.3.2 Space-time models

Fuglstad (2010) discussed how to extend the SPDE in order to build space-time model. He discussed how to find the approximate solutions of the SPDE

$$\frac{\partial}{\partial t} x(s,t) - \nabla \cdot \nabla x(s,t) = \tau \mathcal{W}(s,t), \qquad (s,t) \in [0,L] \times [0,T], \tag{22}$$

where $\nabla = \frac{\partial}{\partial x}, \tau > 0$ is a constant, and $\mathcal{W}(s,t)$ is a space-time standard Gaussian white noise. A finite volume method (FVM) has been chosen to solve Equation (22) with Neumann boundary conditions. The main idea for the FVM approach is to partition the spatial domain into cells, for a fixed t, such that the cells in some sense "respect" a conservation law. For more information about FVMs, see, for example Eymard et al. (2000). The properties of the approximation are discussed in Fuglstad (2010, Section 4.3).

4.3.3 Modelling nonstationarity

Modelling nonstationarity is of great interest in spatial statistics and Fuglstad (2011) discussed one way to build this kind of model by introducing a diffusion matrix in the SPDE. This results in the SPDE

$$\kappa^2(\boldsymbol{s})x(\boldsymbol{s}) - \nabla \cdot \boldsymbol{H}(\boldsymbol{s})\nabla x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{23}$$

where $\kappa^2(\boldsymbol{s})$ is a scalar function, H is a 2×2 matrix-valued diffusion function and $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ is a differential operator. Further, $\mathcal{W}(\boldsymbol{s})$ is a standard Gaussian white noise process, $\boldsymbol{s} \in [0, A] \times [0, B]$ with strictly positive constants A and B. The periodic boundary conditions are used in this model. With different parametrization of the matrix-valued function $\boldsymbol{H}(\boldsymbol{s})$ some interesting nonstationary structure can be achieved.

4.3.4 Applications with SPDE approach

There exist many applications using the SPDE approach. Ingebrigtsen et al. (2013) used the SPDE approach for spatial reconstructing of annual precipitation in Norway. They have slightly modified the SPDE (12) given by Lindgren et al. (2011),

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \nu > 0, \tag{24}$$

where κ is a range parameter and τ is a variance parameter to rescale the variance of the field $x(\boldsymbol{s})$. They get nonstationary random fields by allowing τ and κ varying with spatial location, i.e., $\tau(\boldsymbol{s})$ and $\kappa(\boldsymbol{s})$. For more information about the parametrization about $\tau(\boldsymbol{s})$ and $\kappa(\boldsymbol{s})$, we refer to Ingebrigtsen et al. (2013, Section 3). Simpson et al. (2011) discussed

the use of the SPDE approach for log-Gaussian Cox processes, and did the inference using the integrated nested Laplace approximations (INLA) method discussed by Rue et al. (2009). Aune and Simpson (2012) discussed how to use a systems of SPDEs to construct priors for multivariate models. Cameletti et al. (2011) used the SPDE approach to model particulate matter concentration.

4.3.5 Other directions

There exist more extensions of the SPDE approach. Bolin (2011) discussed how to construct spatial Matérn fields driven by non-Gaussian noise. Sigrist et al. (2012) discussed the use of an transport-diffusion SPDE

$$\frac{\partial}{\partial(t)}x(t,\boldsymbol{s}) = -\boldsymbol{v}\cdot\nabla x(t,\boldsymbol{s}) + \nabla\cdot\boldsymbol{A}\nabla x(t,\boldsymbol{s}) - \eta x(t,\boldsymbol{s}) + \varepsilon(t,\boldsymbol{s}) \qquad (25)$$

to construct a continuous space-time model, where $\boldsymbol{s}\in\mathbb{R}^2$,$\nabla$ is the gradient operator, $\boldsymbol{A}$ is a diffusion matrix and $\varepsilon(t,\boldsymbol{s})$ is a Gaussian process that is temporally white and spatially colored. We refer to their paper for more information for SPDE (25). They pointed out that by solving SPDE (25) in the spectral domain the computational cost of Bayesian or frequentist inference is dominated by the cost of fast Fourier transform.

There exists more literature discussing research on the SPDE approach, but we do not continue the discussion here.

5 Multivariate GRFs

Multivariate GRFs play an important role in spatial statistics since they can model not only the covariances within the random fields, but also the cross-covariances between the random fields. Gneiting et al. (2010) discussed a parametric family of matrix-valued covariance functions for multivariate random fields. Each component in the matrix-valued covari-

ance function is a Matérn covariance function,

$$\boldsymbol{C}(\boldsymbol{h}) = \begin{pmatrix} C_{11}(\boldsymbol{h}) & C_{12}(\boldsymbol{h}) & \cdots & C_{1p}(\boldsymbol{h}) \\ C_{21}(\boldsymbol{h}) & C_{22}(\boldsymbol{h}) & \cdots & C_{2p}(\boldsymbol{h}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{h}) & C_{p2}(\boldsymbol{h}) & \cdots & C_{pp}(\boldsymbol{h}) \end{pmatrix}, \tag{26}$$

where

$$C_{ii}(\boldsymbol{h}) = \sigma_i^2 M(\boldsymbol{h}|\nu_{ii}, a_{ii}), \quad \text{for } i = 1, 2, \ldots, p, \tag{27}$$

is the marginal covariance function for field $x_i(\boldsymbol{s})$, $C_{ii}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s} + \boldsymbol{h})x_i(\boldsymbol{s}))$, and

$$C_{ij}(\boldsymbol{h}) = \rho_{ij}\sigma_i\sigma_j M(\boldsymbol{h}|\nu_{ij}, a_{ij}), \quad \text{for } i, j = 1, 2, \ldots, p, \ \ i \neq j, \tag{28}$$

is the cross-covariance function between fields $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$, $C_{ij}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s} + \boldsymbol{h})x_j(\boldsymbol{s}))$. Theories are given in order to ensure the positive definiteness of covariance function for multivariate GRFs. Gneiting et al. (2010) pointed out that they have the symmetry constraint $C_{ij}(\boldsymbol{h}) = C_{ji}(\boldsymbol{h})$ on the multivariate GRFs constructed with this approach. LMC also has this symmetry constraint for the cross-covariance function.

In Paper A we extended the SPDE approach to construct multivariate GRFs with systems of SPDEs. The system of equations has the following form,

$$\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \ldots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \ldots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \ldots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{29}$$

where $\{\mathcal{L}_{ij} = b_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2}; 1 \leq i, j \leq p\}$ are linear (fractional) differential operators with $\{\alpha_{ij} = 0 \text{ or } 2; 1 \leq i, j \leq p\}$. $\{\varepsilon_i(\boldsymbol{s}); i = 1, 2, \ldots, p\}$ are Gaussian noise processes which are independent, but not necessarily identically distributed. We show that the solution of system (29) defines a multivariate GRF $\boldsymbol{x} = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \cdots, x_p(\boldsymbol{s}))^{\mathrm{T}}$. In Section 2 of Paper A we discuss the connection between the system of SPDEs approach and the covariance-based approach discussed by Gneiting et al. (2010). The GMRF approximation can be extended to multivariate GRFs. So we can

make the precision matrices for multivariate GRFs to be sparse, and hence save computational resources.

In additional to the computational efficiency, there are three other main advantages for our SPDEs approach. The first advantage is that our new approach does not explicitly depend on the theory of positive definite matrices. We do not have to worry about the notorious requirement of positive definite covariance matrices. The second advantage is that we can remove the symmetry property which is shared by the covariance-based approach (Gneiting et al., 2010) and the LMC approach (Gelfand et al., 2004; Gneiting et al., 2010). The third advantage, which has not yet been discussed, is that we can construct multivariate GRFs on manifolds, such as the 2-sphere $\mathbb{S}^2$, by simply interpreting the systems of SPDEs to be defined on the manifold. Lindgren et al. (2011, Section 3.1) discussed the theoretical background for the univariate setting, which is basically the same as for our multivariate setting.

Following the ideas from Lindgren et al. (2011) for constructing univariate random fields with oscillating covariance functions, we discuss the construction of multivariate GRFs with oscillating covariance functions in Paper B. The main idea is to introduce noise processes with oscillating covariance functions in the system of SPDEs given in Equation (29). With a *triangular* system of SPDEs,

$$\begin{pmatrix} \mathcal{L}_{11} & & & \\ \mathcal{L}_{21} & \mathcal{L}_{22} & & \\ \vdots & \vdots & \ddots & \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{30}$$

and noise processes $\{\varepsilon_i; i = 1, 2, \dots, p\}$ with oscillating covariance functions and other noise processes $\{\varepsilon_j; j = 1, 2, \dots, p, j \neq i\}$ that are white noise processes or noise processes with Matérn covariance functions, it is possible to determine which random fields have oscillating covariance functions, have non-oscillating covariance functions and have covariance functions which might be oscillating.

Assume only $\varepsilon_i(\boldsymbol{s})$ has an oscillating covariance function, then all the random fields $\{x_j(\boldsymbol{s}); j < i\}$ constructed through triangular system of SPDEs (30) have non-oscillating covariance functions. The covariance function of $x_i(\boldsymbol{s})$ is oscillating. The covariance functions of $\{x_j(\boldsymbol{s}); j >$

$i\}$ are possibly oscillating. In Paper B we construct bivariate random fields explicitly and have applied the methodology to some datasets of simulated data and to a dataset with measurements of global temperatures and pressures.

6 Examples with SPDE approach

Some examples are chosen to illustrate how the SPDE approach works and how to use the approach. We show how to sample the multivariate random fields constructed by the systems of SPDEs in $\mathbb{R}^d$ and in $\mathbb{S}^2$. We also show some examples of applications with the multivariate random fields discussed in Paper A, Paper B and Paper C.

6.1 Sampling a multivariate GMRF

In order to get a sample from a GMRF, the Cholesky triangle $\boldsymbol{L}$ of the sparse precision matrix $\boldsymbol{Q}$ is usually used. The following steps are the most commonly used for sampling GMRF.

1. Find the Cholesky factor $\boldsymbol{L}$ from the Cholesky factorization, $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$.

2. Sample $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix with the same dimensions as the precision matrix $\boldsymbol{Q}$.

3. Solve for $\boldsymbol{v}$ in the linear system of equations $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{z}$. Then $\boldsymbol{v}$ has the correct precision matrix $\boldsymbol{Q}$, and $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$.

4. Finally correct the mean by $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{v}$.

Wist and Rue (2006) discussed a way to specify a GMRF by a sparse Cholesky triangle from a incomplete Cholesky factorization. They claimed that this sparse Cholesky triangle provided computationally efficient representation for simulating from the GMRF. However, they pointed out that the sparse Cholesky representation is fragile when conditioning a GMRF on a subset of the variables or observed data. This means the computational cost increases in these situations. We refer to Wist and Rue (2006) for more information about their approach.

Table 1: Parameters for simulating the bivariate GRF in the non-oscillating case.

parameters		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$
$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$
$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$	
$\alpha_{n_2} = 0$	$\kappa_{n_2} = 0.3$	

In Paper D we discuss one way to solve the problem from Wist and Rue (2006). We propose to use an incomplete orthogonal factorization with Givens rotations to find a sparse Cholesky triangle. We show that this algorithm is robust and always returns a sparse Cholesky factor even when the GMRF is conditioned on a subset of the variables or observed data. The drawback of this approach is that the incomplete orthogonal factorization is usually slower than the standard Cholesky factorization. Implementing a fast incomplete orthogonal factorization is beyond the scope of discussion.

I. Bivariate random fields with non-oscillating covariance functions

With the parameters given in Table 1, one sample from the bivariate random field is given in Figure 2 together with the corresponding correlation functions. As discussed in Paper A, with a triangular system of SPDEs, the parameter b_{21} is related to the sign of the correlation between the random fields if we have the constraints $b_{11} > 0$ and $b_{22} > 0$. Random fields are positively correlated if $b_{21} < 0$ and negatively correlated if $b_{21} > 0$, and become independent when $b_{21} = 0$. We can notice that the two random fields are positively correlated in our example since $b_{21} < 0$.

II. Bivariate random fields with oscillating covariance functions

In Paper B we show that if one or more noise processes have oscillating covariance functions in the system of SPDEs, then some or all random fields

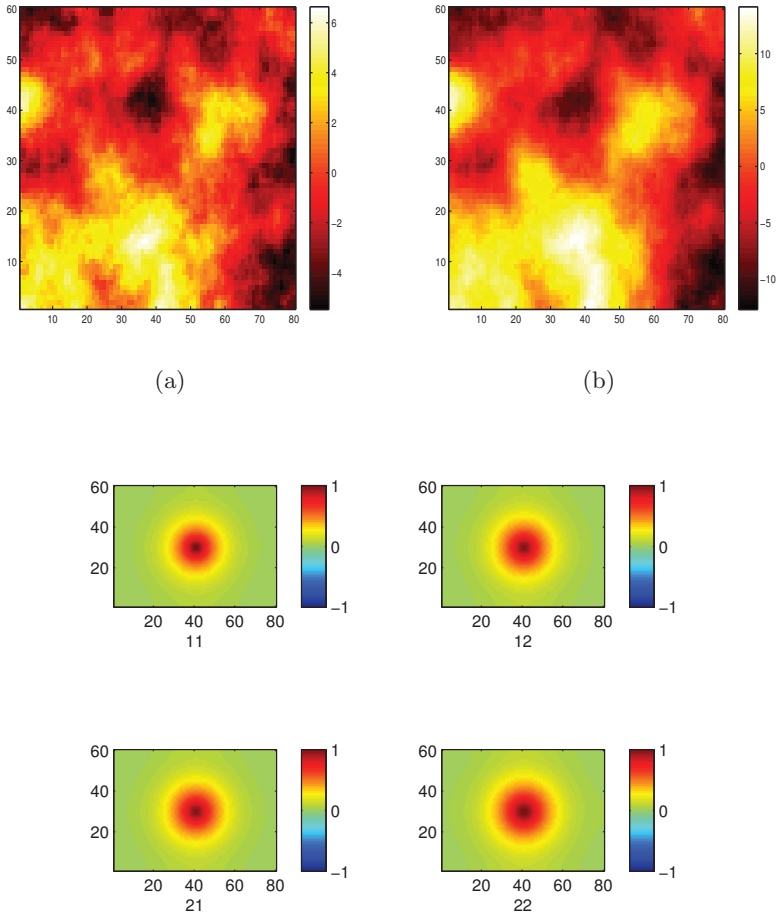


Figure 2: Sample of the bivariate GRF with the correlation matrices with the parameters given in Table 1. The two random fields are positively correlated.

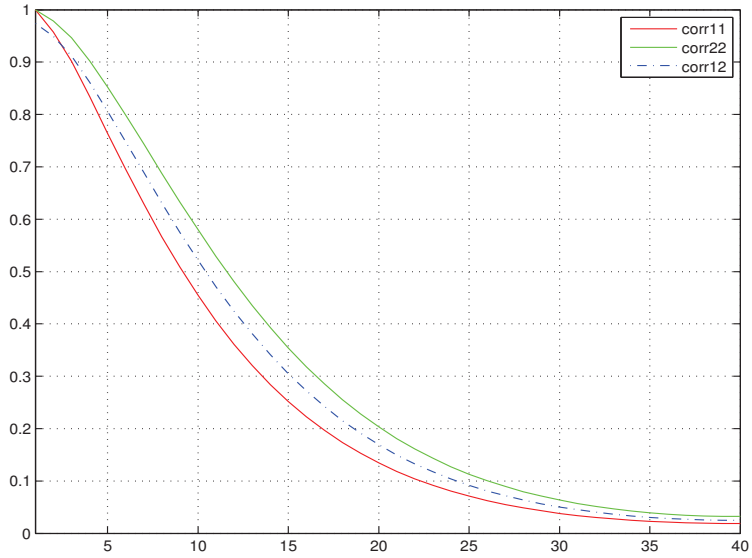


Figure 3: Marginal correlations and cross-correlations with the reference points in the middle of the two GRFs for the positively correlated random fields with the parameters given in Table 1. 'corr11' means the marginal correlation within random field 1. 'corr22' means the marginal correlation within random field 2. 'corr12' means the cross-correlation between random fields 1 and 2.

Table 2: Parameters for sampling the bivariate GRFs in the oscillating case.

parameters		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$h_{11} = 0.25$	$b_{11} = 0.5$
$\alpha_{12} = 0$	$h_{22} = 0.36$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{n_1} = 0.5$	$b_{21} = 0.25$
$\alpha_{22} = 2$	$\kappa_{n_2} = 0.6$	$b_{22} = 1$
$\alpha_{n_1} = 2$	$\omega = 0.95$	
$\alpha_{n_2} = 2$		

have oscillating covariance functions. With the parameters given in Table 2 and an oscillating covariance function only for the second noise process $\varepsilon_2(\boldsymbol{s})$, we get one sample from the bivariate GMRF shown in Figure 4. We notice that only the second random field $x_2(\boldsymbol{s})$ has an oscillating covariance function. If we switch the order of noise processes, i.e, we give the first noise process $\varepsilon_1(\boldsymbol{s})$ an oscillating covariance function and the second noise process $\varepsilon_2(\boldsymbol{s})$ a non-oscillating covariance function, both the random fields have oscillating covariance functions. We refer to Paper B for a detailed discussion.

III. Bivariate random fields on $\mathbb{S}^2$

Following the path discussed by Lindgren et al. (2011), we can interpret the system of SPDEs to be defined on a manifold, such as $\mathbb{S}^2$, in order to construct a multivariate GRF on the manifold. With the parameters given in Table 3, we get one sample of the bivaraite GRF shown in Figure 6 on the sphere with radius $R = 6378.1$km. In this example we only give $\varepsilon_2(\boldsymbol{s})$ an oscillating covariance function. Therefore, the second random field $x_2(\boldsymbol{s})$ has an oscillating covariance function, but not the first random field $x_1(\boldsymbol{s})$. We used similar setting in Paper C in order to model global temperature and pressure.

6.2 Applications with multivariate GRFs

The multivariate GRFs constructed with the systems of SPDEs can be used widely. We have applied the methods to several different datasets.

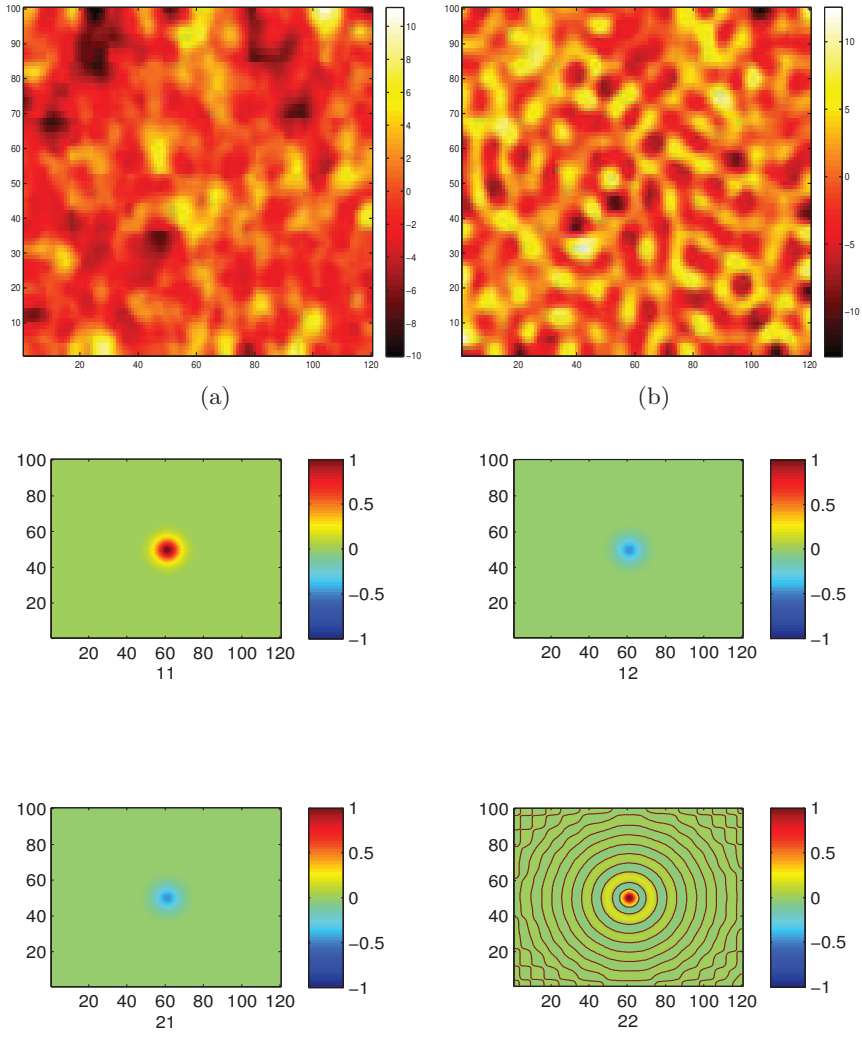


Figure 4: Sample of the bivariate random field together with the correlation matrices with the parameters given in Table 2. The second noise process has an oscillating covariance function. Only the second random field has an oscillating covariance function in this case.

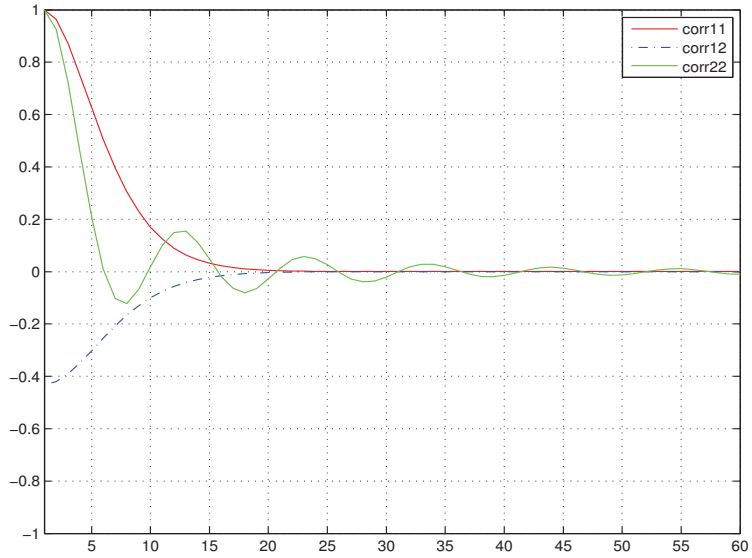


Figure 5: Correlations and cross-correlation functions for the bivariate random field with parameters given in Table 2. In this example the first noise process has a non-oscillating covariance function and the second noise process has an oscillating covariance function. We see that only the second random field has an oscillating covariance function.

Table 3: Parameters for sampling the bivariate GRFs on $\mathbb{S}^2$ in the oscillating case ($R = 6378.1$km)

parameters		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$h_{11} = 360/R^2$	$b_{11} = 5R$
$\alpha_{12} = 0$	$h_{22} = 500/R^2$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{n_1} = \sqrt{h_{11}}$	$b_{21} = -1$
$\alpha_{22} = 2$	$\kappa_{n_2} = 15/R$	$b_{22} = 20R$
$\alpha_{n_1} = 2$	$\omega = 0.95$	
$\alpha_{n_2} = 2$		

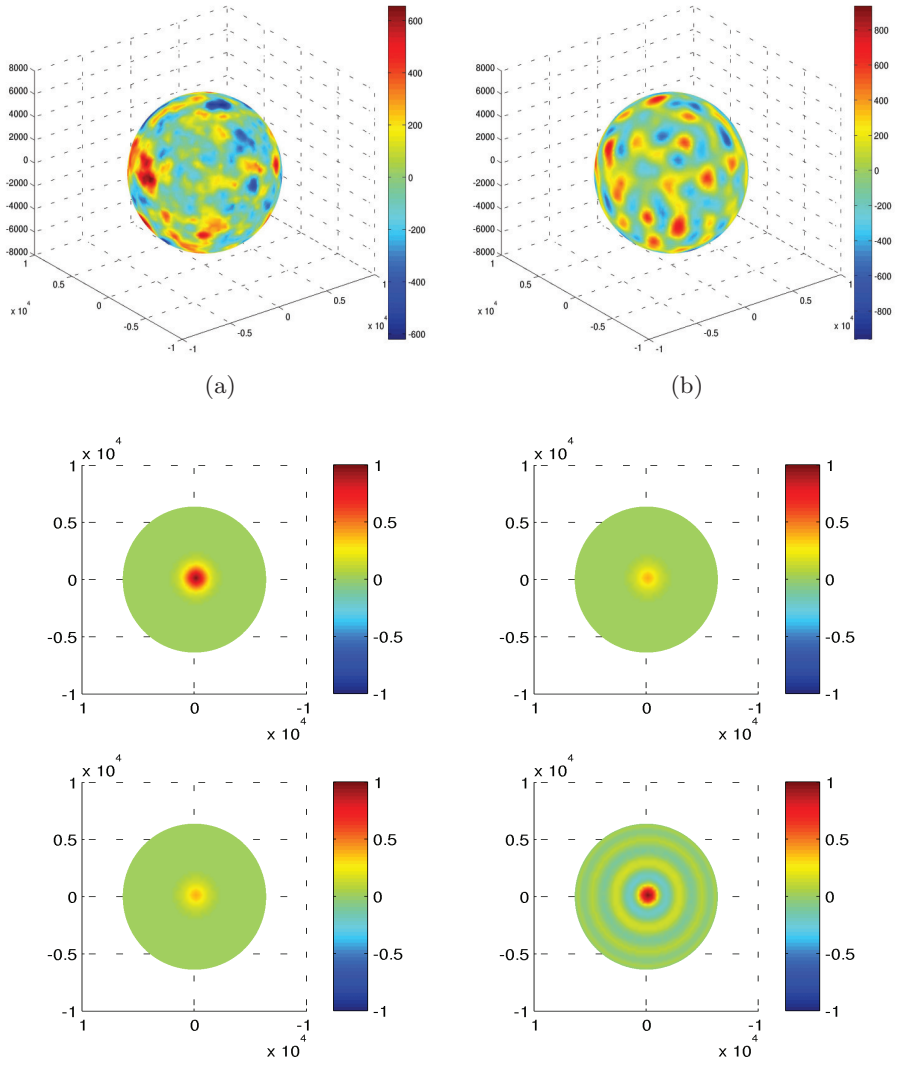


Figure 6: Sample of the bivariate random field on $\mathbb{S}^2$ together with the correlation functions with the parameters given in Table 3. The second noise process has an oscillating covariance function. Only the second random field has an oscillating covariance function in this case.

Some selected results from the included papers in this thesis are given in what follows.

In Paper A we choose the same dataset as Gneiting et al. (2010). This dataset contains one realization with 157 observations for both temperature and pressure at different locations in the north American Pacific Northwest. The reconstructed linear predictors for pressure and temperature are given in Figure 7. We have compared the predictive performance between the system of SPDEs approach and the covariance-based approach proposed by Gneiting et al. (2010), and results show that our method gives slightly better prediction.

In Paper C we apply our approach to model humidity and temperature in southern Norway. The conditional mean of humidity and temperature in 2008 are shown in Figure 8. For other information, such as dependence structures between the humidity and temperature and prediction accuracy of the bivariate model, we refer to Paper C.

7 Possibilities and challenges

The multivariate GRFs constructed by systems of SPDEs are still under study, but we can already handle many real applications as discussed in Paper A, Paper B and Paper C. We are still working on theoretical modelling and implementation of computationally efficient models. We are also looking for new datasets for practical applications of the proposed methods. One current on-going project is trying to integrate all the above mentioned models into the R-INLA package. Some other models, such as non-separable space-time models and non-stationary models can also be possible directions for our further research.

The proposed methods have some challenges. The first one is the interpretation of the parameters. This influences our choice of prior distributions for the hyper-parameters. The second challenge is the inference. It is common that parts of the parameter space for the multivariate models is usually rather flat. Therefore, extending the multivariate model naively to be more flexible might lead to a poorly identifiable model.

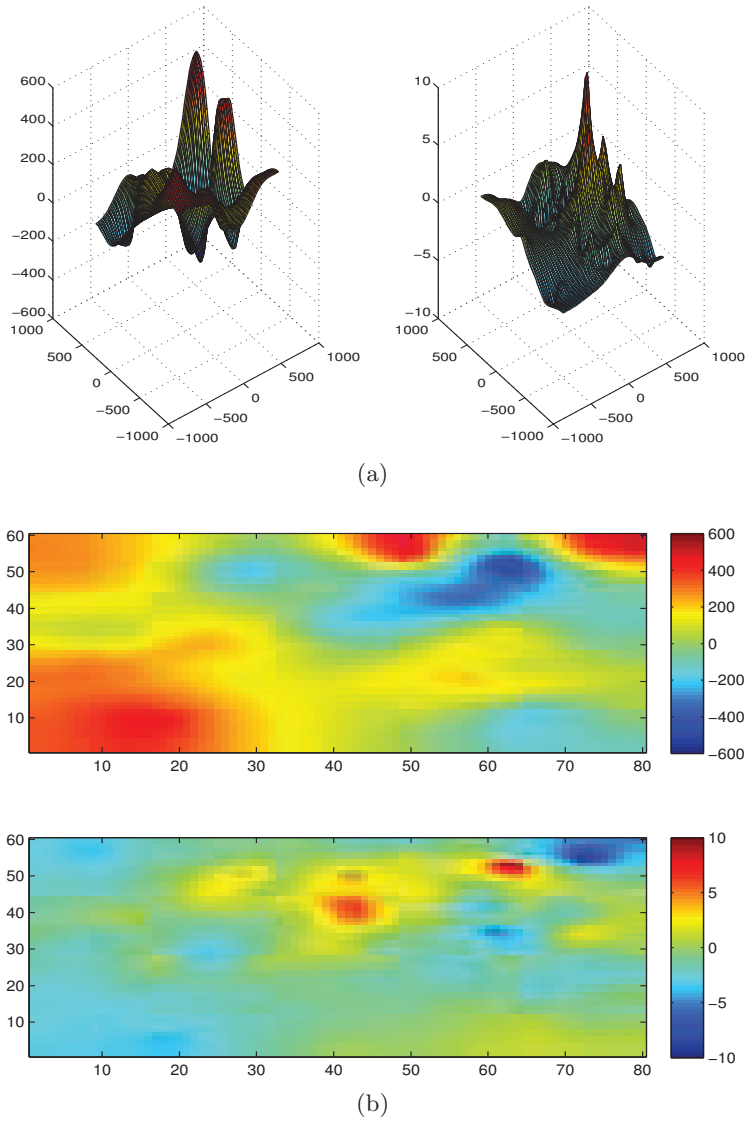
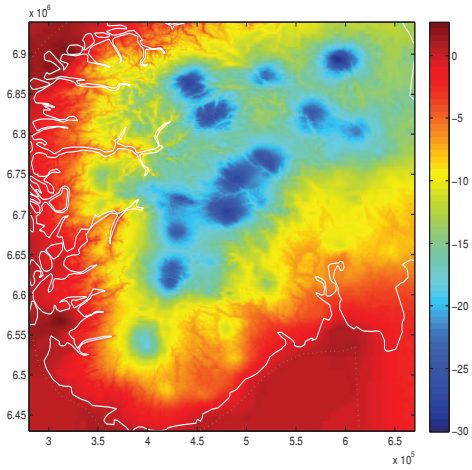
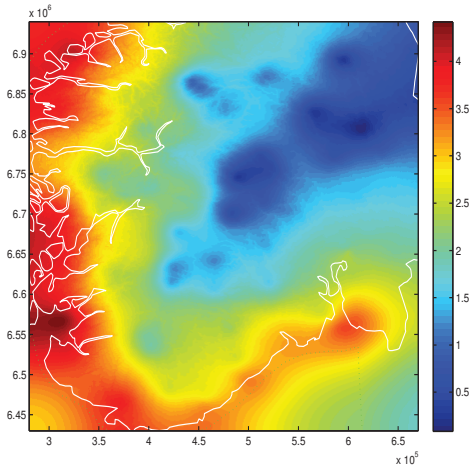


Figure 7: Bivariate GRFs from SPDEs approach in 3D plot (a) and 2D plot (b).



(a)



(b)

Figure 8: Reconstructed linear predictors for temperature (a) and humidity (b) in southern Norway in 2008 by the bivariate model.

8 Summary of papers

Paper A
Multivariate Gaussian Random Fields Using Systems of Stochastic Partial Differential Equations

Xiangping Hu, Daniel Simpson, Finn Lindgren and Håvard Rue

In this paper we discussed how to construct multivariate GRFs with systems of SPDEs. We compared the models constructed with the systems of SPDEs to the covariance based approach discussed by Gneiting et al. (2010) and found that we could construct a subset of their models. But we could also construct other models which were not covered by their approach. Examples are given with simulated data and real data in order to illustrate how to use the proposed methods.

The method was developed in collaboration with all the co-authors. One part of the code was written by Finn Lindgren. I took the lead in all other work, such as implementations and writing the paper.

Paper B
Multivariate Gaussian Random Fields with Oscillating Covariance Functions Using Systems of Stochastic Partial Differential Equations

Xiangping Hu, Finn Lindgren, Daniel Simpson, and Håvard Rue

In this paper we discussed how to constructed multivariate GRFs with oscillating covariance functions. The main idea was to use noise processes with oscillating covariance functions to introduce oscillation features to covariance functions of the multivariate GRFs. The triangular system of SPDEs was recommended due to some advantages discussed in the paper. Four simulated data examples and one real data example are used to illustrate the use of the models in different applications.

The method was developed in collaboration with the all the co-authors. One part of the code was written by Finn Lindgren. I took the lead in all other work, such as implementations and writing the paper.

Paper C
Spatial Modelling of Temperature and Humidity using Systems of Stochastic Partial Differential Equations

Xiangping Hu, Ingelin Steinsland, Daniel Simpson, Sara Martino, and Håvard Rue

In this paper we modelled the dataset with the temperature and humidity in southern Norway with a bivariate GRF from a system of SPDEs. The results showed that we can capture the positive correlation between temperature and the humidity. Furthermore, the results agreed with physical and empirical knowledge. At the end, we concluded that using a bivariate GRF to model this dataset is superior to the approach with univariate GRFs, both when evaluating point predictions and for quantifying prediction uncertainty.

The idea for modelling the temperature and humidity was done in collaboration with the rest of co-authors. The dataset was from Sara Martino. Ingelin Steinsland read through the paper and made several improvements to the paper. I took the lead in all other work, such as implementations and writing the paper.

Paper D
Specifying Gaussian Markov Random Fields with Incomplete Orthogonal Factorization using Givens Rotations

Xiangping Hu, Daniel Simpson and Håvard Rue

In this paper we discussed an approach for finding the incomplete Cholesky factor from an incomplete orthogonal factorization with Givens Rotations. This leaded to a computationally more efficient representation for GMRFs since the Cholesky triangle obtained from the incomplete orthogonal factorization was usually sparser than the Cholesky factor from the standard Cholesky factorization. We showed that the errors for using this sparser triangle can become negligible with proper chosen thresholds, and hence have potential to be used in practical applications. However, the current implementation is slower than the standard Cholesky factorization.

The initial idea for this paper is from the co-authors. I took the lead

in all other work, such as implementations and writing the paper.

References

D.J. Allcroft and C.A. Glasbey. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):487–498, 2003. ISSN 1467-9876.

T.V. Apanasovich and M.G. Genton. Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97(1):15–30, 2010.

T.V. Apanasovich, M.G. Genton, and Y. Sun. A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193, 2012.

E. Aune and D. Simpson. The use of systems of stochastic pdes as priors for multivariate models with discrete structures. *arXiv preprint arXiv:1208.1717*, 2012.

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman & Hall, 2004. ISBN 158488410X.

K.J. Bathe. *Finite element method*. Wiley Online Library, 2008.

P. Billingsley. *Probability and measure*. London: Chapman & Hall, 2nd edition, 1986.

D. Bolin. Spatial matérn fields driven by non gaussian noise. *Preprints in Mathematical Sciences 2011:4, Lund university*, 2011.

D. Bolin and F. Lindgren. Wavelet markov models as efficient alternatives to tapering and convolution fields. Technical report, Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University, 2009.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550, 2011.

P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994.

M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, pages 1–23, 2011.

P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

D.R. Cox and H.D. Miller. *The theory of stochastic processes*, volume 134. Chapman & Hall/CRC, 1977.

N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Tehnical Report No. 780, Department of Statistics, The Ohio State University, Columbus, OH*, 2006.

N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70 (1):209–226, 2008.

N. Cressie, T. Shi, and E.L. Kang. Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19(3):724–745, 2010.

N.A.C. Cressie. *Statistics for spatial data*, volume 298. Wiley-Interscience, 1993.

P.J. Diggle and P.J. Ribeiro Jr. *Model-based Geostatistics*. Springer, 2006.

R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000.

M. Fuentes. Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447–466, 2006.

M. Fuentes. Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331, 2007.

G.A. Fuglstad. Approximating solutions of stochastic differential equations with gussian markov random fields. Technical report, Department of Mathematical Science, Norwegian University of Science and Technology, 2010.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011.

R. Furrer, M.G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, 2006.

Y. Gel, A.E. Raftery, and T. Gneiting. Calibrated probabilistic mesoscale weather field forecasting. *Journal of the American Statistical Association*, 99(467):575–583, 2004.

A.E. Gelfand, A.M. Schmidt, S. Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

M.G. Genton. Separable approximations of space-time covariance matrices. *Environmetrics*, 18(7):681–695, 2007.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

M. Goulard and M. Voltz. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286, 1992. ISSN 0882-8121.

A. Graham. Kronecker products and matrix calculus: With applications. *John Wiley & Sons, Inc., 605 Third Ave., New York, NY 10158, 130*, 1982.

L. Hartman and O. Hössjer. Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 52(5):2331–2349, 2008. ISSN 0167-9473.

Ø. Hjelle and M. Dæhlen. *Triangulations and applications*. Springer Verlag, 2006.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Using stochastic partial differential equation models for spatial reconstruction of annual precipitation. *submitted*, 2013.

M. Jun and M.L. Stein. An approach to producing space–time covariance functions on spheres. *Technometrics*, 49(4):468–479, 2007.

M. Jun and M.L. Stein. Nonstationary covariance models for global data. *The Annals of Applied Statistics*, 2(4):1271–1289, 2008.

C.G. Kaufman, M.J. Schervish, and D.W. Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.

W. Kleiber and D. Nychka. Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 2012.

P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 3rd edition, 1999. ISBN 3540540628.

L.W. Konigsberg and S.D. Ousley. Multivariate quantitative genetics of anthropometric traits from the boas data. *Human biology*, 81(5/6):579–594, 2009.

B. Li and H. Zhang. An approach to modeling asymmetric multivariate spatial covariance structures. *Journal of Multivariate Analysis*, 102(10):1445–1453, 2011.

B. Li, M.G. Genton, and M. Sherman. A nonparametric assessment of properties of space–time covariance functions. *Journal of the American Statistical Association*, 102(478):736–744, 2007.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

G. Lindgren. *A second course on stationary stochastic processes*. Center for Mathematical Sciences, Lund University, December 2010.

B. Matérn. Spatial variation. *Meddelanden fran statens Skogsforskningsinstitut*, 49(5), 1960.

B. Matérn. *Spatial variation*. Springer-Verlag Berlin, 1986.

K. Mcguigan. Studying phenotypic evolution using multivariate quantitative genetics. *Molecular ecology*, 15(4):883–896, 2006.

B.J. Reich and M. Fuentes. A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264, 2007.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue and H. Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49, 2002. ISSN 1467-9469.

H. Rue, I. Steinsland, and S. Erland. Approximating hidden Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):877–892, 2004. ISSN 1467-9868.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868.

S.R. Sain and N. Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007.

A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108(D24):8783, 2003.

B. Shaby and D. Ruppert. Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452, 2012.

T. Shi and N. Cressie. Global statistical analysis of misr aerosol data: a massive data product from nasa's terra satellite. *Environmetrics*, 18(7):665–680, 2007.

F Sigrist, HR Künsch, and WA Stahel. Spde based modeling of large space-time data sets. *Preprint (http://arxiv. org/abs/1204.6118)*, 2012.

D. Simpson, J. Illian, F. Lindgren, S. Sørbye, and H. Rue. Going off grid: Computationally efficient inference for log-gaussian cox processes. *arXiv preprint arXiv:1111.0641*, 2011.

M.L. Stein. *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag, 1999. ISBN 0387986294.

M.L. Stein, Z. Chi, and L.J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296, 2004.

Y. Sun, B. Li, and M.G. Genton. Geostatistics for large datasets. *Advances and challenges in space-time modelling of natural events*, pages 55–77, 2012.

A.V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 297–312, 1988.

H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer Verlag, 2003. ISBN 3540441425.

H.T. Wist and H. Rue. Specifying a Gaussian Markov random field by a sparse Cholesky triangle. *Communications in Statistics-Simulation and Computation*, 35(1):161–176, 2006. ISSN 0361-0918.

H. Zhang and J. Du. Covariance tapering in spatial statistics. *Positive definite functions: From Schoenberg to space-time challenges*, pages 181–196, 2008.

O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, and JZ Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-heinemann, 2005.

Paper I

Multivariate Gaussian Random Fields using Systems of Stochastic Partial Differential Equations

Xiangping Hu, Daniel Simpson, Finn Lindgren and Håvard Rue

Technical Report, 2013

Multivariate Gaussian Random Fields Using Systems of Stochastic Partial Differential Equations

Xiangping Hu*[1], Daniel Simpson[1], Finn Lindgren[2], and Håvard Rue[1]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[2]Department of Mathematical Sciences, University of Bath, BA2 7AY, United Kingdom

Abstract

In this paper a new approach for constructing *multivariate* Gaussian random fields (GRFs) using systems of stochastic partial differential equations (SPDEs) has been introduced and applied to simulated data and real data. By solving a system of SPDEs, we can construct multivariate GRFs. On the theoretical side, the notorious requirement of non-negative definiteness for the covariance matrix of the GRF is satisfied since the constructed covariance matrices with this approach are automatically symmetric positive definite. Using the approximate stochastic weak solutions to the systems of SPDEs, multivariate GRFs are represented by multivariate Gaussian *Markov* random fields (GMRFs) with sparse precision matrices. Therefore, on the computational side, the sparse structures make it possible to use numerical algorithms for sparse matrices to do fast sampling from the random fields and statistical inference. Therefore, the *big-n* problem can also be partially resolved for these models. These models out-preform existing multivariate GRF models on a commonly used real dataset.

*Corresponding author. Email: `Xiangping.Hu@math.ntnu.no`

Keywords: Multivariate Gaussian random fields; Gaussian Markov random fields; covariance matrix; stochastic partial differential equations; sparse matrices

1 Introduction

Gaussian random fields (GRFs) have a dominant role in spatial modelling and there exists a wealth of applications in areas such as geostatistics, atmospheric and environmental science and many other fields (Cressie, 1993; Stein, 1999; Diggle and Ribeiro Jr, 2006). GRFs are practical since the normalizing constant can be computed explicitly. They also process good properties since they can be explicitly specified through the mean function $\boldsymbol{\mu}(\cdot)$ and covariance function $\boldsymbol{C}(\cdot,\cdot)$. In $\mathbb{R}^d$, with $\boldsymbol{s} \in \mathbb{R}^d$, $x(\boldsymbol{s})$ is a continuously indexed GRF if all the finite subsets $(x(\boldsymbol{s}_i))_{i=1}^n$ jointly have Gaussian distributions. A GRF is said be stationary or homogeneous if the mean function $\mu(\cdot)$ is constant and the covariance function $\boldsymbol{C}(\cdot,\cdot)$ is a function only of the distance between the coordinates (Adler and Taylor, 2007). It is common to use covariance functions that are isotropic. The isotropic covariance functions are functions of just the Euclidean distance between two locations.

1.1 Matérn covariance functions and Multivariate GRFs

The Matérn family of covariance functions is a class of commonly used isotropic covariance functions introduced by Matérn (Matérn, 1986). This family of covariance functions is usually parametrized by $\sigma^2 M(\boldsymbol{m}, \boldsymbol{n}|\nu, a)$, where, $M(\boldsymbol{m}, \boldsymbol{n}|\nu, a)$ is the Matérn correlation function between locations $\boldsymbol{m}$, $\boldsymbol{n} \in \mathbb{R}^d$ and has the form

$$M(\boldsymbol{h}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a\|\boldsymbol{h}\|)^\nu K_\nu(a\|\boldsymbol{h}\|), \tag{1}$$

where σ^2 is the marginal variance. $\|\boldsymbol{h}\|$ denotes the Euclidean distance between $\boldsymbol{m}$ and $\boldsymbol{n}$. K_ν is the modified Bessel function of second kind and $a > 0$ is a scaling parameter. The order ν is the smoothness parameter. It defines the critical concerns in spatial statistical modelling and simulation (Stein, 1999), such as the differentiability of the sample paths and

the Hausdorff dimension (Handcock and Stein, 1993; Goff and Jordan, 1988). We follow the common practice to fix the smoothness parameter ν because it is poorly identifiable from data (Ribeiro Jr and Diggle, 2006; Lindgren et al., 2011). The range ρ connects the scaling parameter a and the smoothness parameter ν. Throughout the paper, the simple relationship $\rho = \sqrt{8\nu}/a$ is assumed from the empirically derived definition (Lindgren et al., 2011). We refer to Matérn (1986), Diggle et al. (1998), Guttorp and Gneiting (2006) and Lindgren et al. (2011) for detailed information about Matérn covariance functions. One can show that the Matérn covariance function can be reduced to the product of a polynomial and an exponential function when the smoothness parameter ν is equal to an integer m plus $\frac{1}{2}$ (Gneiting et al., 2010). The Matérn covariance function nests the popular exponential model since $M(\boldsymbol{h}|\frac{1}{2}, a) = \exp(-a\|\boldsymbol{h}\|)$.

Analogously to the univariate case, we need to specify a mean vector function $\boldsymbol{\mu}(\boldsymbol{s})$ and a covariance function $\boldsymbol{C}(\|\boldsymbol{h}\|)$ that assigns to each distance $\|\boldsymbol{h}\|$ a $p \times p$ symmetric non-negative definite covariance matrix in order to specify an isotropic p-dimensional multivariate Gaussian random field $\boldsymbol{x}(\boldsymbol{s})$. It is known that it is quite difficult to construct flexible covariance functions $\boldsymbol{C}(\cdot)$ that satisfy this requirement.

Gneiting et al. (2010) presented an approach for constructing multivariate random fields using matrix-valued covariance functions, where each constituent component in the matrix-valued covariance function is a Matérn covariance function. Define $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \cdots, x_p(\boldsymbol{s}))^{\mathrm{T}}$, with $\boldsymbol{s} \in \mathbb{R}^d$, so that each location consists p components. T denotes the transpose of a vector or a matrix. Assume the process is second-order stationary with mean vector zero and matrix-valued covariance function of the form

$$\boldsymbol{C}(\boldsymbol{h}) = \begin{pmatrix} C_{11}(\boldsymbol{h}) & C_{12}(\boldsymbol{h}) & \cdots & C_{1p}(\boldsymbol{h}) \\ C_{21}(\boldsymbol{h}) & C_{22}(\boldsymbol{h}) & \cdots & C_{2p}(\boldsymbol{h}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{h}) & C_{p2}(\boldsymbol{h}) & \cdots & C_{pp}(\boldsymbol{h}) \end{pmatrix}, \tag{2}$$

where,

$$C_{ii}(\boldsymbol{h}) = \sigma_{ii} M(\boldsymbol{h}|\nu_{ii}, a_{ii}), \tag{3}$$

is the covariance function $C_{ii}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s}+\boldsymbol{h})x_i(\boldsymbol{s}))$ within the field $x_i(\boldsymbol{s})$

and

$$C_{ij}(\boldsymbol{h}) = \rho_{ij}\sigma_{ij}M(\boldsymbol{h}|\nu_{ij}, a_{ij}) \tag{4}$$

is the cross-covariance function $C_{ij}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s}+\boldsymbol{h})x_j(\boldsymbol{s}))$ between the fields $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$. ρ_{ij} is the co-located correlation coefficient, $\sigma_{ii} \geq 0$ is the marginal variance, and σ_i and σ_j are the corresponding standard deviations, $1 \leq i \neq j \leq p$. We use the following notations through out the paper

$$\begin{aligned} \sigma_{ii} &= \sigma_i\sigma_i, \\ \sigma_{ij} &= \sigma_i\sigma_j. \end{aligned} \tag{5}$$

Gneiting et al. (2010) presented some conditions to ensure the matrix-valued covariance function in Equation (2) is symmetric and non-negative definite, with focus on the bivariate case. They claimed that the parameters in the parametric family of matrix-valued covariance function given in Equation (2) for multivariate random fields are interpretable in terms of smoothness, correlation length, variances of the processes and co-located coefficients. It was shown that a parsimonious bivariate Matérn model, with fewer parameters than the full bivariate Matérn model, is preferred to the traditional linear model of coregionalization (LMC) (Gelfand et al., 2004). Even though the GRFs have convenient analytical properties, the covariance-based models for constructing GRFs are hindered by computational issues, or the so-called "*big-n* problem" (Banerjee et al., 2004). This is because inference with these models requires the factorization of dense $n \times n$ covariance matrices, which requires $\mathcal{O}(n^2)$ storage and $\mathcal{O}(n^3)$ floating point operations. It follows that this kind of model is not suitable for problems with realistically large amounts of data.

Statistical inference for large datasets within feasible time is still a challenge in modern spatial statistics. The size of modern datasets typically overwhelm the traditional models in spatial statistics and a great deal of effort has been expended trying to construct good models that scale well computationally. In this paper, we extend the models proposed by Lindgren et al. (2011), which exploited an explicit link between GRFs and GMRFs through stochastic partial differential equations (SPDEs). They showed that for some univariate Gaussian fields with Matérn covariance functions, it is possible to do the modelling with GRFs, but do the computations using GMRFs.

A GMRF is a discretely indexed Gaussian random field $\boldsymbol{x}(\boldsymbol{s})$ with the property that its full conditional distributions $\{\pi(x_i|\boldsymbol{x}_{-i}); i = 1, \ldots, n\}$, only depend on the set of neighbors ∂i to each node i. It is obvious that if $i \in \partial j$ then $j \in \partial i$ from consistency. Rather than parameterizing GMRFs through their covariance matrix, they are usually parametrized through their precision matrix (the inverse of the covariance matrix). The Markov property implies that the precision matrix $\boldsymbol{Q}$ is sparse due to the fact that $Q_{ij} \neq 0 \iff i \in \partial j \cup j$ (Rue and Held, 2005). This allows for the use of numerical methods for sparse matrices for fast sampling and also for fast statistical inference (Rue, 2001; Rue and Held, 2005; Lindgren et al., 2011). The general cost of factorizing the precision matrix $\boldsymbol{Q}$ for a spatial Markovian structure is $\mathcal{O}(n)$ in one dimension, $\mathcal{O}(n^{3/2})$ in two dimensions and $\mathcal{O}(n^2)$ in three dimensions (Rue and Held, 2005).

Lindgren et al. (2011) pointed out that using the link between the GRFs and GMRFs can open new doors to modelling difficult problems with simple models. The SPDE approach can be extended to Gaussian random fields on manifolds, non-stationary random fields, random fields with oscillating covariance functions and non-separable space-time models.

1.2 Matérn covariance models through SPDEs

In this paper we use the following characterization of Matérn random fields, originally due to Whittle (1954, 1963), that formed the basis for the methods of Lindgren et al. (2011). A GRF $x(\mathbf{s})$ with the Matérn covariance function can be described as a solution to the linear fractional SPDE (Whittle, 1954, 1963; Lindgren et al., 2011)

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{6}$$

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator and $\alpha = \nu + d/2$, $\kappa > 0$, $\nu > 0$. The innovation process $\mathcal{W}(\boldsymbol{s})$ is a spatial standard Gaussian white noise. $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian on $\mathbb{R}^d$. Applying the Fourier transform to the (fractional) SPDE given in $\mathbb{R}^d$ in (6) yields

$$\left\{\mathscr{F}(\kappa^2 - \Delta)^{\alpha/2}\phi\right\}(\boldsymbol{k}) = \left(\kappa^2 + \|\boldsymbol{k}\|^2\right)^{\alpha/2} (\mathscr{F}\phi)(\boldsymbol{k}), \tag{7}$$

where $\mathscr{F}$ denotes the Fourier transform, ϕ is a smooth, rapidly decaying function in $\mathbb{R}^d$. See for example, Lindgren et al. (2011) for detailed de-

scription. Equation (7) is used in Section 3 for model comparison. One might think that the Matérn covariance function seems rather restrictive in statistical modelling, but it covers the most commonly used models in spatial statistics (Lindgren et al., 2011). Stein (1999, Page 14) has a practical suggestion "use the Matérn model". For more information about the Matérn family, see, for example Diggle et al. (1998, Section 3.4.1) and Stein (1999, Section 2.6).

1.3 Outline of the paper

The rest of the paper is organized as follows. In Section 2 we discuss the construction of multivariate GRFs using systems of SPDEs, we call this approach the SPDEs approach. Additionally, a brief introduction of how to construct the multivariate GRFs using the covariance-based models is also included in this section. Section 3 contains a detailed model comparisons between the SPDEs approach and the covariance-based models presented by Gneiting et al. (2010). Section 3.4 discusses how to sample from these models, and statistical inference for simulated data and real dataset is presented in Section 4. The paper ends with a general discussion in Section 5.

2 Model construction

In this section we discuss how to use the SPDEs approach to construct multivariate GRFs. One of the appealing properties of this approach is that the SPDE specification *automatically* constructs valid covariance functions. This means that if the solution to a system of SPDEs exists, then it will construct a matrix-valued covariance function which fulfills the symmetric non-negative definiteness property. The parameters in the parametric model from the SPDEs approach are interpretable in terms of co-located correlation coefficients, smoothness, marginal variances and correlations. And there is a correspondence to the parameters in the covariance matrix based models.

2.1 Multivariate GRFs and the SPDEs approach

A zero mean d–dimensional multivariate (or d–dimensional p-variate) GRF is a collection of continuously indexed multivariate normal random vectors

$$\boldsymbol{x}(\boldsymbol{s}) \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{s})),$$

where $\boldsymbol{\Sigma}(\boldsymbol{s})$ is a non-negative definite matrix that depends on the point $\boldsymbol{s} \in \mathbb{R}^d$. Define the system of SPDEs

$$\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} f_1(\boldsymbol{s}) \\ f_2(\boldsymbol{s}) \\ \vdots \\ f_p(\boldsymbol{s}) \end{pmatrix}, \tag{8}$$

where $\{\mathcal{L}_{ij} = b_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2};\ \alpha_{ij} = 0 \text{ or } 2, 1 \leq i, j \leq p\}$, are differential operators and $\{f_i(\boldsymbol{s}); 1 \leq i, j \leq p\}$ are independent but not necessarily identically distributed noise processes. It turns out that the solution to (8) defines a multivariate GRF $\boldsymbol{x} = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \cdots, x_p(\boldsymbol{s}))^{\mathrm{T}}$. Define the operator matrix

$$\mathscr{L} = \begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \tag{9}$$

and

$$\boldsymbol{f}(\boldsymbol{s}) = (f_1(\boldsymbol{s}), f_2(\boldsymbol{s}), \cdots, f_p(\boldsymbol{s}))^{\mathrm{T}},$$

and then the system of SPDEs (8) can be written in compact matrix form as

$$\mathscr{L}(\boldsymbol{\theta})\boldsymbol{x}(\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{s}), \tag{10}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{b}\}$ denotes the collection of parameters. Similarly as Equation (7) we apply Fourier transforms to Equation (10), and it yields

$$\mathscr{H}(\boldsymbol{\theta})\hat{\boldsymbol{x}}(\boldsymbol{k}) = \hat{\boldsymbol{f}}(\boldsymbol{k}), \tag{11}$$

where $\boldsymbol{k}$ is the frequency, $\hat{\boldsymbol{x}}(\boldsymbol{k}) = \mathscr{F}(\boldsymbol{x}(\boldsymbol{s}))$ and $\hat{\boldsymbol{f}}(\boldsymbol{k}) = \mathscr{F}(\boldsymbol{f}(\boldsymbol{s}))$ are the Fourier transforms of the random fields and the noise processes, respectively, and $\mathscr{H}(\boldsymbol{\theta})$ is the matrix formed from the Fourier transforms

of the operator matrix $\mathscr{L}(\boldsymbol{\theta})$. Each element has the form $\mathcal{H}_{ij}(\theta_{ij}) = b_{ij}(\kappa_{ij}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{ij}/2}$ with $\theta_{ij} = \{\alpha_{ij}, \kappa_{ij}, b_{ij}\}$, $1 \leq i, j \leq p$. From Equation (11) one can find the spectral process as $\hat{\boldsymbol{x}}(\boldsymbol{k}) = \mathscr{H}^{-1}(\boldsymbol{\theta})\hat{\boldsymbol{f}}(\boldsymbol{k})$. The corresponding power spectrum is defined as $\boldsymbol{S_x}(\boldsymbol{k}) = \mathbb{E}\left(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\mathrm{H}}\right)$, where H denotes the Hermitian transpose of a matrix. Simple calculations using the above mentioned formulas yield

$$\boldsymbol{S_x}(\boldsymbol{k}) = \mathscr{H}^{-1}(\boldsymbol{\theta})\boldsymbol{S_f}(\boldsymbol{k})\mathscr{H}^{-H}(\boldsymbol{\theta}), \tag{12}$$

where $\mathscr{H}^{-H}$ denotes the inverse of the Hermitian transpose of Fourier transform of the operator matrix $\mathscr{L}$. The Equation (12) can be written explicitly as

$$\boldsymbol{S_x}(\boldsymbol{k}) = \begin{pmatrix} S_{x_{11}}(\boldsymbol{k}) & S_{x_{12}}(\boldsymbol{k}) & \cdots & S_{x_{1p}}(\boldsymbol{k}) \\ S_{x_{21}}(\boldsymbol{k}) & S_{x_{22}}(\boldsymbol{k}) & \cdots & S_{x_{2p}}(\boldsymbol{k}) \\ \vdots & \vdots & \ddots & \vdots \\ S_{x_{p1}}(\boldsymbol{k}) & S_{x_{p2}}(\boldsymbol{k}) & \cdots & S_{x_{pp}}(\boldsymbol{k}) \end{pmatrix}. \tag{13}$$

Let $\boldsymbol{S_f}(\boldsymbol{k})$ denote the power spectrum matrix of the noise processes, $\boldsymbol{S_f}(\boldsymbol{k}) = \mathbb{E}\left(\hat{\boldsymbol{f}}\hat{\boldsymbol{f}}^{\mathrm{H}}\right)$. If the noise processes are mutually independent, $\boldsymbol{S_f}$ is a diagonal matrix and we write it as

$$\boldsymbol{S_f}(\boldsymbol{k}) = \mathrm{diag}(S_{f_{11}}, S_{f_{22}}, \cdots, S_{f_{pp}}),$$

where $\{S_{f_i} = \mathbb{E}\left(\hat{f}_i\hat{f}_i^{\mathrm{H}}\right); i = 1, \cdots, p\}$. This means it is easy to obtain this matrix. In general, with Equations (8) to (12) we can compute all the components in (13). We show how to do this for bivariate random fields in Section 3.1.

2.2 Covariance-based model for multivariate GRFs

It is also possible to construct a multivariate GRF using the covariance-based model, but the notorious non-negative definiteness restriction makes it hard. Gneiting et al. (2010) discussed one such approach in detail. The main aim of their approach is to find the proper constrains for $\nu_{ij}, a_{ij}, \rho_{ij}$ which result in valid matrix-valued covariance functions for second-order

stationary multivariate GRFs with a covariance function of the form in Equation (2), that is, a symmetric non-negative definite covariance function. Some useful theorems for constructing the covariance functions for bivariate case are presented in Gneiting et al. (2010, Section 2.2).

In this covariance-based model, the components $C_{ij}(\boldsymbol{h})$ in the matrix-valued covariance function $\boldsymbol{C}(\boldsymbol{h})$ given in Equation (2) are modelled directly. Gneiting et al. (2010) try to find conditions on the parameter space which result in valid multivariate Matérn models. In the bivariate case when $p = 2$, a full characterization of the parameter space is achieved. For $p \geq 3$ Gneiting et al. (2010) suggested that a parsimonious model should be used in practice. This kind of model has more restrictions on the smoothness parameter and the scale parameters, such that $a_{ij} = a$, where $a > 0$ is the common scale parameter and $\nu_{ij} = \frac{1}{2}(\nu_i + \nu_j)$ for $1 \leq i \neq j \leq p$. We refer to Gneiting et al. (2010) for more information about different kinds of conditions that yield the valid matrix-valued covariance functions.

Assume that the components of the covariance matrix are known. The power spectrum can also be obtained from in this covariance-based approach. This can in turn be used to compare parameters with the SPDEs approach. The covariance matrix of a second-order stationary multivariate Matérn GRF was given in Equation (2). The power spectrum in $\mathbb{R}^d$ of the cross-covariance function for $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$ is defined as $R_{x_{ij}}(\boldsymbol{k}) = \mathscr{F}\left(\mathrm{Cov}(x_i(\boldsymbol{s}), x_j(\boldsymbol{s}))\right)$, $1 \leq i, j \leq p$, which is given by

$$R_{x_{ij}}(\boldsymbol{k}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\boldsymbol{k}\boldsymbol{h}} C_{ij}(\|\boldsymbol{h}\|) d\boldsymbol{h}. \tag{14}$$

Applying the Fourier transform to the covariance matrix in Equation (2) yields

$$\boldsymbol{R}(\boldsymbol{k}) = \begin{pmatrix} R_{x_{11}}(\boldsymbol{k}) & R_{x_{12}}(\boldsymbol{k}) & \cdots & R_{x_{1p}}(\boldsymbol{k}) \\ R_{x_{21}}(\boldsymbol{k}) & R_{x_{22}}(\boldsymbol{k}) & \cdots & R_{x_{2p}}(\boldsymbol{k}) \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_{p1}}(\boldsymbol{k}) & R_{x_{p2}}(\boldsymbol{k}) & \cdots & R_{x_{pp}}(\boldsymbol{k}) \end{pmatrix}. \tag{15}$$

The comparison between the covariance-based approach and the SPDEs approach in Section 3 is based on the following fact: if $S_{x_{ij}}(\boldsymbol{k}) = R_{x_{ij}}(\boldsymbol{k})$, for each $1 \leq i, j \leq p$, then the multivariate GRFs constructed through

the SPDEs approach (8) and through the covariance-based model (2) are equivalent.

2.3 GMRF approximations to GRFs

It is known that spatial Markovian GMRFs have good computational properties since the precision matrix $\boldsymbol{Q}$ is sparse. The Markov structure here means that $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$ are independent conditioned on $x_{-ij}(\boldsymbol{s})$, when i and j are not neighbors, where $x_{-ij}(\boldsymbol{s})$ means for $x_{-\{i,j\}}(\boldsymbol{s})$. In other words, $Q_{ij} = 0$ if and only if $x_i(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$ are independent conditioned $x_{-ij}(\boldsymbol{s})$. So it is possible to read off whether x_i and x_j are conditionally independent or not from the elements of $\boldsymbol{Q}$. Note that the precision matrix $\boldsymbol{Q}$ should also be non-negative definite. The sparse structure of the precision matrix $\boldsymbol{Q}$ is crucial with our models in Bayesian inference methods on large datasets. We refer to Rue and Held (2005, Chapter 2) for more information on the theory for GMRFs.

Rue and Tjelmeland (2002) demonstrated that most of the GRFs in geostatistics can be approximated with GMRFs. Hartman and Hössjer (2008) proposed to use GMRFs instead of GRFs, from a computational point of view, when doing spatial prediction using Kriging. This approach can also be used for spatio-temporal models (Allcroft and Glasbey, 2003). In this paper we follow the approach presented in Lindgren et al. (2011) for GMRF approximations to GRFs in order to partially resolve the "*big n problem*" with our models. Lindgren et al. (2011) mainly considered univariate GRFs with Matérn covariance functions. In this case the GMRF representation was constructed explicitly through the SPDE, and the solution to the SPDE driven by the Gaussian white noise is a GRFs with Matérn covariance function. They showed how to build a GRF model with a covariance matrix $\boldsymbol{\Sigma}$ theoretically from SPDE, and then use the GMRF to represent the GRF, which means the precision matrix of the GMRF fulfills the condition $\boldsymbol{Q^{-1}} \simeq \boldsymbol{\Sigma}$ on some predefined norm. Fuglstad (2011) considered a modification of the SPDE with a diffusion matrices to control the covariance structure of the GRF, and create inhomogeneous GRFs, but the GMRF representation is still the main ingredient for computations.

With our SPDEs approach we use two steps to construct the multivariate GRFs. The first step is to construct the precision matrices for the noise processes in Equation (8), and the second step is to solve the system

SPDEs (8) with the constructed noise processes. We focus on the noise processes, if they are not white noise processes, generated by the SPDE given as follows

$$(\kappa_{n_i}^2 - \Delta)^{\alpha_{n_i}/2} f_i(\boldsymbol{s}) = \mathcal{W}_i(\boldsymbol{s}),\ \alpha_{n_i} = \nu_{n_i} + d/2,\ i = 1, 2, \ldots, p, \tag{16}$$

with the requirements that $\kappa_{n_i} > 0$ and $\nu_{n_i} > 0$. κ_{n_i} and ν_{n_i} are the scaling parameter and smoothness parameter for noise process $f_i(\boldsymbol{s})$. $\{\mathcal{W}_i(\boldsymbol{s}); i = 1, 2, \ldots, p\}$ are standard Gaussian white noise processes. The generated noise processes in this way are independent but not necessarily identically distributed.

The first step in constructing the GMRF representation for the noise process $f_i(\boldsymbol{s})$ on the triangulated lattice is to find the stochastic weak formulation of Equation (16) (Kloeden and Platen, 1999). In this paper Delaunay triangulation is chosen, and we refer to Hjelle and Dæhlen (2006) for detailed discussion about Delaunay triangulations. Denote the inner product of functions h and g as

$$\langle h, g \rangle = \int h(\boldsymbol{s}) g(\boldsymbol{s}) d(\boldsymbol{s}), \tag{17}$$

where the integral is over the region of interest. We can find the stochastic weak solution of SPDE (16) by requiring that

$$\left\{ \langle \phi_k, (\kappa_{n_i}^2 - \Delta)^{\alpha_{n_i}/2} f_i \rangle \right\}_{i=1}^{M} \stackrel{d}{=} \{ \langle \phi_k, \mathcal{W}_i \rangle \}_{i=1}^{M}. \tag{18}$$

In the second step we need to construct a finite element representation of the solution to the SPDE. We refer to Zienkiewicz et al. (2005) and Bathe (2008) for more information about finite element methods. The finite element representation of the solution to SPDE (16) is

$$f_i(\boldsymbol{s}) = \sum_{k=1}^{N} \psi_k(\boldsymbol{s}) \omega_k, \tag{19}$$

where $\psi_k(\boldsymbol{s})$ is some chosen basis function, ω_k is some Gaussian distributed weight, and N is the number of the vertexes in the triangulation. The basis function $\psi_k(\boldsymbol{s})$ is chosen to be piecewise linear with value 1 at vertex k and 0 at all other vertexes. This means that a continuously indexed solution

is approximated with a piecewise linear function defined through the joint distribution of $\{\omega_k; k = 1, 2, \ldots, N\}$ (Lindgren et al., 2011). The chosen basis functions ensure that the local interpolation on a triangle is piecewise linear.

The third step is to choose the test functions. In this paper we follow the setting used by Lindgren et al. (2011). With $M = N$ the test functions are chosen as $\phi_k = (\kappa_{n_i}^2 - \Delta)^{1/2}\psi_k$ for $\alpha_{n_i} = 1$ and $\phi_k = \psi_k$ for $\alpha_{n_i} = 2$, which are denoted as a *least squares* and a *Galerkin* solution, respectively. When $\alpha_{n_i} \geq 3$, the approximation can be obtained by setting $\alpha_{n_i} = 2$ at the left-hand side of Equation (16) and replacing the right hand side of Equation (16) with a field generated with $\alpha_{n_i} - 2$ and let $\phi_k = \psi_k$. This iteration procedure terminates when $\alpha_{n_i} = 1$ or 2. This is the essence of the recursive Galerkin formulation. More detailed description can be found in Lindgren et al. (2011). α_{n_i} can only be integer-valued currently. When α_{n_i} is not an integer, different approximation methods must be used and this is beyond the scope of our discussion. Response to the discussion to Lindgren et al. (2011) discussed fractional α. We define the required $N \times N$ matrices $\boldsymbol{C}$, and $\boldsymbol{G}$ with entries

$$C_{mn} = \langle \psi_m, \psi_n \rangle, \;\; G_{mn} = \langle \nabla\psi_m, \nabla\psi_n \rangle, \;\; m, n = 1, 2, \ldots, N, \tag{20}$$

and

$$(\boldsymbol{K}_{\kappa_{n_i}^2}) = \left(\kappa_{n_i}^2 \boldsymbol{C} + \boldsymbol{G}\right). \tag{21}$$

Finally, by using these matrices given in (20) and (21) together with the Neumann boundary conditions (zero normal derivatives at the boundaries), the precision matrices $\boldsymbol{Q}_{f_i}$ for noise process $f_i(\boldsymbol{s})$ can be obtained,

$$\begin{cases} \mathbf{Q}_{1,\kappa_{n_i}^2} & = \mathbf{K}_{\kappa_{n_i}^2}, & \text{for } \alpha_{n_i} = 1, \\ \mathbf{Q}_{2,\kappa_{n_i}^2} & = \mathbf{K}_{\kappa_{n_i}^2}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{K}_{\kappa_{n_i}^2}, & \text{for } \alpha_{n_i} = 2, \\ \mathbf{Q}_{\alpha_{n_i},\kappa_{n_i}^2} & = \mathbf{K}_{\kappa_{n_i}^2}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{Q}_{\alpha_{n_i}-2,\kappa_{n_i}^2} \mathbf{C}^{-1} \mathbf{K}_{\kappa_{n_i}^2}, & \text{for } \alpha_{n_i} = 3, 4, \cdots. \end{cases} \tag{22}$$

As pointed out by Lindgren et al. (2011) the inverse of $\boldsymbol{C}$, i.e., $\boldsymbol{C}^{-1}$, is usually a dense matrix. This causes the precision matrix $\boldsymbol{Q}_{f_i}$ to be dense and ruins all the effort we have made. So we actually used a diagonal matrix $\tilde{\boldsymbol{C}}$, where $\tilde{\boldsymbol{C}}_{ii} = \langle \psi_i, 1 \rangle$, instead of $\boldsymbol{C}$. This diagonal matrix results in a sparse precision matrix and hence sparse GMRFs models can be

obtained. Using the diagonal matrix $\tilde{\boldsymbol{C}}_{ii}$ yields a Markov approximation to the FEM solution. The effects of the Markov approximation have been studied by Bolin and Lindgren (2009). They claimed that the difference between the exact representation by the finite element method and the Markov approximation is negligible.

Move to our multivariate GRFs constructed by systems of SPDEs (8). In this case we take the same set of basis functions $\{\psi_k; k = 1, 2, \ldots, M\}$ and construct a basis for the solution space for $(x_1, x_2, \ldots, x_p)^{\mathrm{T}}$ as

$$p\left\{\begin{pmatrix}\psi_1\\0\\\vdots\\0\end{pmatrix},\ldots,\begin{pmatrix}\psi_M\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}0\\\psi_1\\\vdots\\0\end{pmatrix},\ldots,\begin{pmatrix}0\\\psi_M\\\vdots\\0\end{pmatrix},\ldots,\begin{pmatrix}0\\0\\\vdots\\\psi_1\end{pmatrix},\ldots,\begin{pmatrix}0\\0\\\vdots\\\psi_M\end{pmatrix}\right., \tag{23}$$

where there are a total of Mp basis functions $\{\boldsymbol{\psi}_i\}$ which are numbered in the order listed above. The weak solution of the system of SPDEs (8) requires

$$[\langle \boldsymbol{\psi}_i, \mathscr{L}\boldsymbol{x}\rangle]_{i=1}^{Mp} \overset{d}{=} [\langle \boldsymbol{\psi}_i, \boldsymbol{f}\rangle]_{i=1}^{Mp}, \tag{24}$$

with $\mathscr{L}$ defined in Equation (9). The finite element representation to the solution of the system of SPDEs is

$$\boldsymbol{x}(\boldsymbol{s}) = \sum_{i=1}^{Mp} \boldsymbol{\psi}_i(\boldsymbol{s})\omega_i \tag{25}$$

In order to find the precision matrix for the solution we need to define the following matrices,

$$\begin{aligned}
\boldsymbol{D} &= \begin{pmatrix}\tilde{\boldsymbol{C}} & & & \\ & \tilde{\boldsymbol{C}} & & \\ & & \ddots & \\ & & & \tilde{\boldsymbol{C}}\end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix}\boldsymbol{K}_{11} & \boldsymbol{K}_{12} & \cdots & \boldsymbol{K}_{1p}\\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} & \cdots & \boldsymbol{K}_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{K}_{p1} & \boldsymbol{K}_{p2} & \cdots & \boldsymbol{K}_{pp}\end{pmatrix},\\
\boldsymbol{Q}_{\boldsymbol{f}} &= \begin{pmatrix}\boldsymbol{Q}_{f_1} & & & \\ & \boldsymbol{Q}_{f_2} & & \\ & & \ddots & \\ & & & \boldsymbol{Q}_{f_p}\end{pmatrix},
\end{aligned} \tag{26}$$

where $\boldsymbol{D}$ and $\boldsymbol{Q_f}$ are block diagonal matrices with p blocks on the diagonal. $\{\boldsymbol{K}_{ij} = b_{ij}(\kappa_{ij}^2\boldsymbol{C} + \boldsymbol{G}); i, j = 1, 2, \ldots, p\}$ for $\alpha_{ij} = 2$ and $\{\boldsymbol{K}_{ij} = b_{ij}\boldsymbol{I}_{M\times M}; i, j = 1, 2, \ldots, p\}$ for $\alpha_{ij} = 0$.

We summarize the results for multivariate GRFs with our systems of SPDEs approach in Main result **1**.

Main result 1. *Let $\boldsymbol{Q}$ be the precision matrix for the Gaussian weights ω_i in the system of SPDEs* (8)*, then the precision matrix of the multivariate GMRF is*

$$\boldsymbol{Q} = \boldsymbol{K}\boldsymbol{D}^{-1}\boldsymbol{Q_f}\boldsymbol{D}^{-1}\boldsymbol{K} \tag{27}$$

with $\boldsymbol{D}$ and $\boldsymbol{K}$ defined in (26)*.*

The form of the Main result **1** is similar to the discussion in Lindgren et al. (2011, Appendix C.4). With our approach the precision matrices for the multivariate GRFs are sparse and the smoothness of the fields is mainly controlled by the noise processes. This sparse precision matrix is used for sampling the multivariate GRFs in Section 3.4 and statistical inference with simulated data and real data in Section 4.

3 Model comparison and sampling the GRFs

In this section we focus on the construction of the bivariate GRF, i.e., $p = 2$ in $\mathbb{R}^d$ using the SPDEs approach and then compare with the covariance-based approach presented by Gneiting et al. (2010). As discussed in Section 2.2, the SPDEs approach can construct the same multivariate GRFs as the covariance-based models if $S_{x_{ij}}(\boldsymbol{k}) = R_{x_{ij}}(\boldsymbol{k})$ for all $1 \leq i, j \leq q$. The comparison for the univariate GRF is trivial and the multivariate GRF when $p > 2$ can be done in a similar way.

3.1 Bivariate GRF with SPDEs

When $p = 2$ in the system of SPDEs (8), we can construct bivariate GRFs when $b_{12} \neq 0$ or $b_{21} \neq 0$. In this case, the system of SPDEs has the following form

$$\begin{aligned} b_{11}(\kappa_{11}^2 - \Delta)^{\alpha_{11}/2}x_1(\boldsymbol{s}) + b_{12}(\kappa_{12}^2 - \Delta)^{\alpha_{12}/2}x_2(\boldsymbol{s}) &= f_1(\boldsymbol{s}), \\ b_{22}(\kappa_{22}^2 - \Delta)^{\alpha_{22}/2}x_2(\boldsymbol{s}) + b_{21}(\kappa_{21}^2 - \Delta)^{\alpha_{21}/2}x_1(\boldsymbol{s}) &= f_2(\boldsymbol{s}), \end{aligned} \tag{28}$$

and the solution $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}))^{\mathrm{T}}$ to the system of equations (28) is a bivariate random field. Since it is convenient to study the properties of the bivariate GRFs in the spectral domain, the Fourier transform is applied,

$$\begin{aligned} b_{11}(\kappa_{11}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{11}/2}\hat{x}_1(\boldsymbol{k}) + b_{12}(\kappa_{12}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{12}/2}\hat{x}_2(\boldsymbol{k}) &= \hat{f}_1(\boldsymbol{k}), \\ b_{22}(\kappa_{22}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{22}/2}\hat{x}_2(\boldsymbol{k}) + b_{21}(\kappa_{21}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{21}/2}\hat{x}_1(\boldsymbol{k}) &= \hat{f}_2(\boldsymbol{k}). \end{aligned} \tag{29}$$

The matrix form of the differential operator in the spectral domain can be written as

$$\mathscr{H}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{H}_{11}(\theta_{11}) & \mathcal{H}_{12}(\theta_{12}) \\ \mathcal{H}_{21}(\theta_{21}) & \mathcal{H}_{22}(\theta_{22}) \end{pmatrix}. \tag{30}$$

If the noise processes are mutually independent (but not necessarily identically distributed), the power spectrum matrix of the noise processes is a block diagonal matrix with the form

$$S_{\boldsymbol{f}}(\boldsymbol{k}) = \begin{pmatrix} S_{f_1}(\boldsymbol{k}) & 0 \\ 0 & S_{f_2}(\boldsymbol{k}) \end{pmatrix}, \tag{31}$$

where $S_{f_1}(\boldsymbol{k})$ and $S_{f_2}(\boldsymbol{k})$ are the power spectra in $\mathbb{R}^d$ for the noise processes $f_1(\boldsymbol{s})$ and $f_2(\boldsymbol{s})$, respectively. If the noise processes are white, then the problems is simplified and the corresponding power spectra have the forms $S_{\mathcal{W}_1}(\boldsymbol{k}) = (2\pi)^{-d}\sigma_{n_1}^2$ and $S_{\mathcal{W}_2}(\boldsymbol{k}) = (2\pi)^{-d}\sigma_{n_2}^2$, where σ_{n_1} and σ_{n_2} are the standard deviations for the white noise processes $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively. However, these two parameters are confounded with $\{b_{i,j}; i, j = 1, 2\}$ and we fix $\sigma_{n_1} = 1$ and $\sigma_{n_2} = 1$. The conclusion is also valid with other types of noise processes. Notice that we have used the new notation $\{S_{\mathcal{W}_i}; i = 1, 2\}$, because it is also possible to use more interesting noise processes which are not only the simple white noise. For instance, we can use the noise processes with Matérn covariance functions. These kinds of noise processes can be generated (independently) from the SPDEs

$$\begin{aligned} (\kappa_{n_1}^2 - \Delta)^{\alpha_{n_1}/2} f_1 &= \mathcal{W}_1, \\ (\kappa_{n_2}^2 - \Delta)^{\alpha_{n_2}/2} f_2 &= \mathcal{W}_2, \end{aligned} \tag{32}$$

where $\mathcal{W}_1$ and $\mathcal{W}_2$ are standard Gaussian white noise processes and κ_{n_1} and κ_{n_2} are scaling parameters. α_{n_1} and α_{n_2} are related to smoothness parameters ν_{n_1} and ν_{n_2} for f_1 and f_2 and $\alpha_{n_i} = \nu_{n_i} + d/2$. Apply the Fourier

transform to (32) and use a similar procedure as defined in Equation (10) - Equation (13), and then the power spectra for the noise processes generated from SPDEs (32) can be obtained and they have the forms

$$\begin{aligned} S_{f_1}(\boldsymbol{k}) &= \frac{1}{(2\pi)^d}\frac{1}{(\kappa_{n_1}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{n_1}}}, \\ S_{f_2}(\boldsymbol{k}) &= \frac{1}{(2\pi)^d}\frac{1}{(\kappa_{n_2}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{n_2}}}. \end{aligned} \tag{33}$$

The power spectrum matrix of the GRFs presented in (13) can also be simplified further and has the form

$$\boldsymbol{S_x}(\boldsymbol{k}) = \begin{pmatrix} S_{x_{11}}(\boldsymbol{k}) & S_{x_{12}}(\boldsymbol{k}) \\ S_{x_{21}}(\boldsymbol{k}) & S_{x_{22}}(\boldsymbol{k}) \end{pmatrix}. \tag{34}$$

By using (12) together with the exact formula for the differential operators (30) and the spectra for the noise processes as defined in (33), we can get an exact symbolic expressions for all the elements in the power spectrum matrix (34)

$$\begin{aligned} S_{x_{11}} &= \frac{S_{f_1}|\mathcal{H}_{22}^2| + S_{f_2}|\mathcal{H}_{12}^2|}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|}, \\ S_{x_{12}} &= -\frac{\mathcal{H}_{22}S_{f_1}|\mathcal{H}_{21}^2|\mathcal{H}_{11} + \mathcal{H}_{12}S_{f_2}|\mathcal{H}_{11}^2|\mathcal{H}_{21}}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|\mathcal{H}_{21}\mathcal{H}_{11}}, \\ S_{x_{21}} &= -\frac{\mathcal{H}_{21}S_{f_1}|\mathcal{H}_{22}^2|\mathcal{H}_{12} + \mathcal{H}_{11}S_{f_2}|\mathcal{H}_{12}^2|\mathcal{H}_{22}}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|\mathcal{H}_{22}H_{12}}, \\ S_{x_{22}} &= \frac{S_{f_1}|\mathcal{H}_{21}^2| + S_{f_2}|\mathcal{H}_{11}^2|}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|}. \end{aligned} \tag{35}$$

In order to simply the problem, we make the operator matrix in the spectral domain a lower triangular matrix by setting $H_{12}(\theta_{12}, \boldsymbol{k}) = 0$, or equivalently, by setting $b_{12} = 0$. The system of SPDEs in the spectral domain becomes

$$b_{11}(\kappa_{11}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{11}/2}\hat{x}_1(\boldsymbol{k}) = \hat{f}_1(\boldsymbol{k}), \tag{36}$$

$$b_{22}(\kappa_{22}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{22}/2}\hat{x}_2(\boldsymbol{k}) + b_{21}(\kappa_{21}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{21}/2}\hat{x}_1(\boldsymbol{k}) = \hat{f}_2(\boldsymbol{k}). \tag{37}$$

This means expression (35) becomes

$$\begin{aligned} S_{x_{11}} &= \frac{S_{f_1}}{|\mathcal{H}_{11}^2|}, & S_{x_{21}} &= -\frac{\bar{\mathcal{H}}_{21} S_{f_1}}{\bar{\mathcal{H}}_{22}|\mathcal{H}_{11}|^2}, \\ S_{x_{12}} &= -\frac{S_{f_1}\mathcal{H}_{21}}{\mathcal{H}_{22}|\mathcal{H}_{11}|^2}, & S_{x_{22}} &= \frac{|\mathcal{H}_{21}|^2 S_{f_1} + |\mathcal{H}_{11}|^2 S_{f_2}}{|\mathcal{H}_{11}|^2|\mathcal{H}_{22}|^2}, \end{aligned} \tag{38}$$

where $\bar{\mathcal{H}}_{ij}$ denotes the conjugate of $\mathcal{H}_{ij}$. This is called the *triangular* version of SPDEs in this paper. If the operators $\mathcal{H}_{12}$ and $\mathcal{H}_{21}$ are real, we obtain $S_{x_{12}}(\boldsymbol{k}) = S_{x_{21}}(\boldsymbol{k})$. In other words, we have an imposed symmetry property on the cross-covariance in this case. With the triangular version of the SPDEs, under some extra conditions, the properties of the multivariate GRFs is easy to interpret. We will see more about this in Section 4. In this paper we focus on the triangular version of the SPDEs. However, the full version of the SPDEs could be handled analogously to the triangular version of the SPDEs. If there is no constraint on the operator matrix, i.e., $b_{ij} \neq 0$ for all $i, j = 1, 2$, it is called the *full* version of the SPDEs. In general, $S_{x_{12}}(\boldsymbol{k})$ and $S_{x_{21}}(\boldsymbol{k})$ are not necessarily equal. This can release the constraint shared by the Matérn model proposed by Gneiting et al. (2010) and the Linear model of coregionalization (LMC), namely that the cross-covariance has an imposed symmetry property, i.e., $C_{ij}(\boldsymbol{h}) = C_{ji}(\boldsymbol{h})$. This does in general not hold as discussion by Wackernagel (2003, Chapter 20). We refer to Goulard and Voltz (1992), Wackernagel (2003, Chapter 14) and Gelfand et al. (2004) for more information about LMC. In this paper we assume that all the operators are real which simplifies calculations and discussion. We reorganize (38) and find

$$\begin{aligned} \frac{S_{x_{12}}}{S_{x_{11}}} &= -\frac{\mathcal{H}_{21}}{\mathcal{H}_{22}}, \\ \frac{S_{x_{22}}}{S_{x_{11}}} &= \frac{|\mathcal{H}_{21}^2|}{|\mathcal{H}_{22}^2|} + \frac{|\mathcal{H}_{11}^2|}{\mathcal{H}_{22}^2}\frac{S_{f_2}}{S_{f_1}}, \\ \frac{S_{x_{22}}}{S_{x_{12}}} &= -\left(\frac{|\mathcal{H}_{11}^2|}{\mathcal{H}_{21}\mathcal{H}_{22}}\right)\frac{S_{f_2}}{S_{f_1}} - \frac{\mathcal{H}_{21}}{\mathcal{H}_{22}}. \end{aligned} \tag{39}$$

From the results given in (35) and (39), we can notice that $S_{x_{12}}$ and $S_{x_{11}}$ only depend on the power spectrum of the noise process f_1 and $S_{x_{22}}(\boldsymbol{k})$ depends on the noise process power spectra $S_{f_1}(\boldsymbol{k})$ and $S_{f_2}(\boldsymbol{k})$. The ratio

between the power spectra $S_{x_{12}}$ and $S_{x_{11}}$ is independent of both the noise processes.

When $\{f_i(\boldsymbol{s}); i = 1, 2\}$ are mutually independent and generated from (32), the elements in the power spectrum matrix (34) can be written down explicitly by using (33) and (38),

$$
\begin{aligned}
S_{x_{11}}(\boldsymbol{k}) &= \frac{1}{(2\pi)^d(\kappa_{n_1}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{n_1}} b_{11}^2 \left(\kappa_{11}^2 + \|\boldsymbol{k}\|^2\right)^{\alpha_{11}}}, \\
S_{x_{12}}(\boldsymbol{k}) &= -\frac{b_{21}(\kappa_{n_1}^2 + \|\boldsymbol{k}\|^2)^{-\alpha_{n_1}} (\kappa_{21}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{21}/2}}{(2\pi)^d \left(b_{22}(\kappa_{22}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{22}/2} b_{11}^2 (\kappa_{11}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{11}}\right)}, \\
S_{x_{22}}(\boldsymbol{k}) &= \frac{\frac{b_{21}^2(\kappa_{21}^2+\|\boldsymbol{k}\|^2)^{\alpha_{21}}}{(2\pi)^d(\kappa_{n_1}^2+\|\boldsymbol{k}\|^2)^{\alpha_{n_1}}} + \frac{b_{11}^2(\kappa_{11}^2+\|\boldsymbol{k}\|^2)^{\alpha_{11}}}{(2\pi)^d(\kappa_{n_2}^2+\|\boldsymbol{k}\|^2)^{\alpha_{n_2}}}}{\left(b_{11}^2(\kappa_{11}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{11}}\right) \left(b_{22}^2(\kappa_{22}^2 + \|\boldsymbol{k}\|^2)^{\alpha_{22}}\right)}.
\end{aligned}
\tag{40}
$$

The asymptotic behavior of the power spectra for the bivariate GRF can be obtained from (40). For some selected parameter values defined in Table 1, the power spectra are shown in Figure 1. From these figure, we can notice that the parameters $\{b_{ij}; i, j = 1, 2, i > j\}$ controls the correlation between the two fields. When a smaller absolute value of b_{21} is chosen, the correlation between these two fields decreases rapidly. We can also show that the sign of $b_{21} \cdot b_{22}$ is related the sign of the correlation between the two GRFs. Obviously, b_{ij} are also related to the variance of the GRFs. The parameters κ_{ij} are related to the range of the two fields. α_{ij} and α_{n_i} are related to the smoothness of the two GRFs. The asymptotic behavior is curial when dealing with the real-world applications.

3.2 Bivariate GRF with the covariance-based model

As presented by Gneiting et al. (2010), a multivariate GRFs can be constructed using a covariance-based model. In the bivariate setting, the matrix-valued covariance function for the bivariate GRF contains 4 elements and the corresponding power spectrum matrix becomes

$$
\boldsymbol{R}(\boldsymbol{k}) = \begin{pmatrix} R_{x_{11}}(\boldsymbol{k}) & R_{x_{12}}(\boldsymbol{k}) \\ R_{x_{21}}(\boldsymbol{k}) & R_{x_{22}}(\boldsymbol{k}) \end{pmatrix}. \tag{41}
$$

Table 1: Parameter values for asymptotic behaviors of the power spectra for bivariate GRFs

case 1			case 2		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$	$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$	$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$	$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$
$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$	$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$
$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$		$\alpha_{n_1} = 1$	$\kappa_{n_1} = 0.15$	
$\alpha_{n_1} = 0$	$\kappa_{n_2} = 0.3$		$\alpha_{n_2} = 1$	$\kappa_{n_2} = 0.3$	
case 3			case 4		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$	$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$	$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.15$	$b_{21} = -0.5$	$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$
$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$	$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$
$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$		$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$	
$\alpha_{n_1} = 0$	$\kappa_{n_2} = 0.3$		$\alpha_{n_2} = 0$	$\kappa_{n_2} = 0.3$	

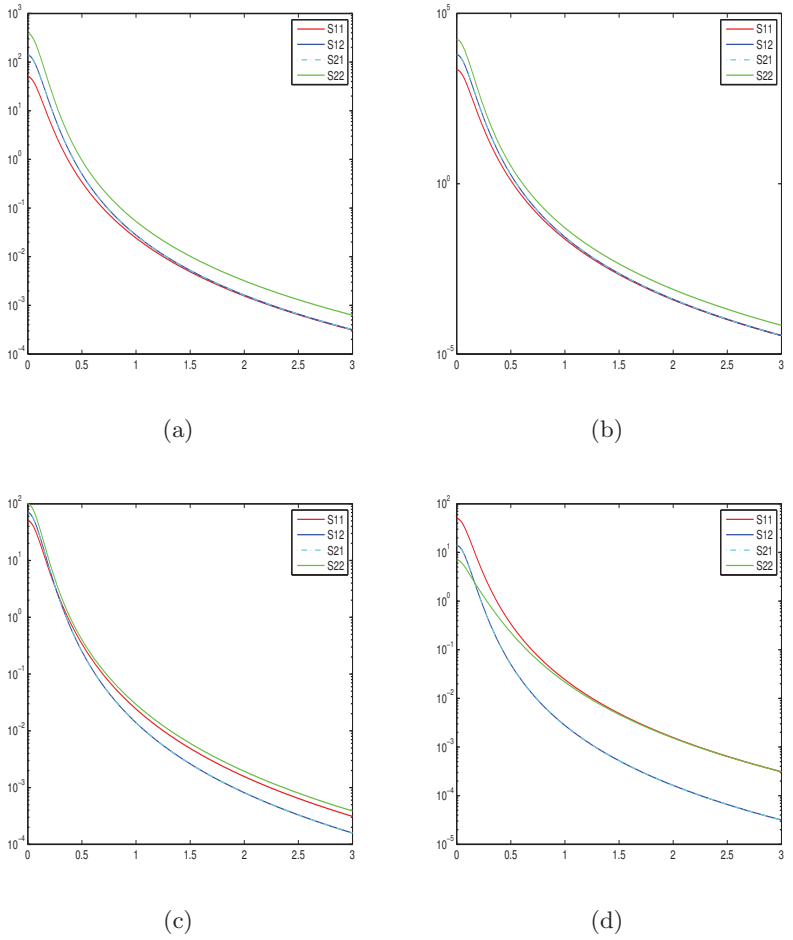


Figure 1: Asymptotic behaviors of the power spectra corresponding to case 1 (a), case 2 (b), case 3 (c) and case 4 (d) with parameter values given in Table 1.

Applying integral given in Equation (14), together with the expression for the marginal variances in Equation (3) and cross-covariance in Equation (4), we can find closed form for the elements in the power spectrum matrix,

$$\begin{aligned} R_{x_{ii}}(\boldsymbol{k}) &= \frac{1}{(2\pi)^d}\left[a_{ii}^{2\nu_{ii}}(\sqrt{4\pi})^d \frac{\sigma_{ii}\Gamma(\nu_{ii}+d/2)}{(a_{ii}^2+\|\boldsymbol{k}\|^2)^{\nu_{ii}+d/2}\Gamma(\nu_{ii})}\right], \\ R_{x_{ij}}(\boldsymbol{k}) &= \frac{1}{(2\pi)^d}\left[a_{ij}^{2\nu_{ij}}(\sqrt{4\pi})^d \frac{\rho_{ij}\sigma_i\sigma_j\Gamma(\nu_{ij}+d/2)}{(a_{ij}^2+\|\boldsymbol{k}\|^2)^{\nu_{ij}+d/2}\Gamma(\nu_{ij})}\right]. \end{aligned} \tag{42}$$

Notice that if $d = \{2, 4, ...\}$, then the expressions in (42) can be simplified even further by (recursively) using the well-known formula

$$\Gamma(\nu+1) = \nu\Gamma(\nu).$$

Since all the components $\{C_{ij}; i, j = 1, 2\}$ in the matrix-valued covariance matrix $\boldsymbol{C}$ are from Matérn family, as defined in Equation (3) and Equation (4) (Gneiting et al., 2010), the power spectra of the marginal covariance functions $\{C_{ii}; i = 1, 2\}$ and the cross covariance functions $\{C_{ij}; i, j = 1, 2, i \neq j\}$ have similar forms as indicated in (42).

3.3 Parameter matching

As mentioned in Section 2.2, the model comparison in this paper is based on the fact that if $S_{ij}(\boldsymbol{k}) = R_{ij}(\boldsymbol{k})$, for each $1 \leq i, j \leq p$, and then the multivariate GRFs constructed from the SPDEs approach (8) and from the covariance-based model (2) will be equivalent.

By comparing each element in (40) and (42), we can get the results given as follows,

$$\begin{aligned} 0 &= \rho_{12} - \rho_{21}, \\ \kappa_{11} = \kappa_{21} = \kappa_{22} = \kappa_{n_1} = \kappa_{n_2} &= a_{11} = a_{12} = a_{21} = a_{22}, \\ \alpha_{11} + \alpha_{n_1} &= \nu_{11} + \frac{d}{2}, \\ \alpha_{11} + \frac{\alpha_{22}}{2} + \alpha_{n_1} - \frac{\alpha_{21}}{2} &= \nu_{12} + \frac{d}{2}, \end{aligned} \tag{43}$$

and

$$\begin{aligned}
\frac{1}{b_{11}^2} &= \frac{(4\pi)^{d/2} a_{11}^{2\nu_{n1}} \sigma_{11} \Gamma(\nu_{11} + d/2)}{\Gamma(\nu_{11})}, \\
-\frac{b_{21}}{b_{22} b_{11}^2} &= \frac{(4\pi)^{d/2} a_{12}^{2\nu_{12}} \Gamma(\nu_{12} + d/2) \rho_{12} \sigma_1 \sigma_2}{\Gamma(\nu_{12})}, \\
\frac{b_{21}^2 + b_{11}^2}{b_{11}^2 b_{22}^2} &= \frac{(4\pi)^{d/2} \sigma_{22} a_{22}^{2\nu_{22}} \Gamma(\nu_{22} + d/2)}{\Gamma(\nu_{22})}.
\end{aligned} \tag{44}$$

More equalities can be obtained based on the relationship between $\alpha_{21} + \alpha_{n_2}$ and $\alpha_{11} + \alpha_{n_1}$. When $\alpha_{21} + \alpha_{n_2} \leq \alpha_{11} + \alpha_{n_1}$, we can get

$$\alpha_{n_2} + \alpha_{22} = \nu_{22} + \frac{d}{2}. \tag{45}$$

When $\alpha_{21} + \alpha_{n_2} \geq \alpha_{11} + \alpha_{n_1}$, we can get

$$\alpha_{11} + \alpha_{22} + \alpha_{n_1} - \alpha_{21} = \nu_{22} + \frac{d}{2}. \tag{46}$$

The Equation (45) and Equation (46) are critical when discussing about the asymptotic behaviors of bivariate GRF and the corresponding power spectra. From (43), we can notice that in order to construct equivalent models to the ones as discussed by Gneiting et al. (2010), we need to have the constraint that all the scaling parameters should be equal in addition to some constraints for the smoothness parameters given in the third and the fourth equations in (43) and one more conditional equality given in (45) or (46). From the first equation in (43), we notice the extra symmetry restriction discussed in Gneiting et al. (2010, Section 4) also must be fulfilled. Additionally, from the second equation in (44), we notice that there is one important relationship

$$\rho_{12} b_{22} b_{21} \leq 0, \tag{47}$$

between the co-located correlation coefficient ρ_{12} and b_{21} and b_{22}. It shows that the correlation ρ_{12} must have opposite sign to the product of b_{21} and b_{22}. This is an important information not only for sampling from the bivariate GRFs but also for interpreting the results from inference.

From the results given in (44), we can obtain the following results

$$\begin{aligned} -\frac{b_{22}}{b_{21}} &= \frac{a_{11}^{\nu_{11}}\sigma_1\Gamma(\nu_{11}+d/2)/\Gamma(\nu_{11})}{a_{12}^{\nu_{12}}\rho\sigma_2\Gamma(\nu_{12}+d/2)/\Gamma(\nu_{12})}, \\ \frac{b_{22}^2}{b_{11}^2+b_{21}^2} &= \frac{a_{11}^{\nu_{11}}\sigma_{11}\Gamma(\nu_{11}+d/2)/\Gamma(\nu_{11})}{a_{22}^{\nu_{22}}\sigma_{22}\Gamma(\nu_{22}+d/2)/\Gamma(\nu_{22})}. \end{aligned} \tag{48}$$

Notice that when $a_{11} = a_{21} = a_{22}$ and $\nu_{11} = \nu_{21} = \nu_{22}$, the above results in (48) can be simplified to the following form

$$\begin{aligned} -\frac{b_{22}}{b_{21}} &= \frac{\sigma_1}{\rho\sigma_2}, \\ \frac{b_{22}^2}{b_{11}^2+b_{21}^2} &= \frac{\sigma_{11}}{\sigma_{22}}. \end{aligned} \tag{49}$$

These results show that the parameters $\{b_{ij}; i, j = 1, 2\}$ are not only connected to the correlation between these two fields, but also connected with the marginal variance of the GRF. From the results given in (43) to (49), we can notice that the parameters in the system of SPDEs have similar interpretations as the parameters from the covariance-based model.

From these results we see that it is possible to construct *multivariate* GRFs through the SPDEs approach as the covariance-based approach proposed by Gneiting et al. (2010). Additionally, there are three main advantages for our SPDEs approach. The first advantage is that our new approach does not explicitly depend on the theory of positive definite matrix. We do not need to worry about the notorious requirement of positive definite covariance matrices. The second advantage is that we can remove the symmetry property which is shared by the covariance-based approach and the LMC approach (Gelfand et al., 2004; Gneiting et al., 2010). The third advantage, which has not yet been discussed, is that we can construct multivariate GRFs on manifolds, such as the sphere $\mathbb{S}^2$, by simply reinterpret the systems of SPDEs to be defined on the manifold. Lindgren et al. (2011, Section 3.1) discussed the theoretical background in the univariate setting, which is basically the same as for our multivariate setting.

Furthermore, we can actually go even future for the multivariate GRFs, such as multivariate GRFs with oscillating covariance functions and non-stationary multivariate GRFs. These kinds of multivariate GRFs are under development but are beyond the scope of this paper.

3.4 Sampling bivariate GRFs and trivariate GRFs

As presented in Section 2.3 we can construct multivariate GRFs by the SPDEs approach theoretically, but do the computations using the GMRF representation. The precision matrix $\boldsymbol{Q}$ is usually sparse and the sparseness of the precision matrices enables us to apply numerical linear algebra for sparse matrices for sampling from the GRFs and for fast statistical inference. We now assume that the multivariate GMRF has mean vector $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$, i.e., $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$. When sampling from the GMRF, the forward substitution and backward substitution is applied by using the Cholesky triangle $\boldsymbol{L}$, where by definition $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$. The commonly used steps for sampling a GRF are as follows.

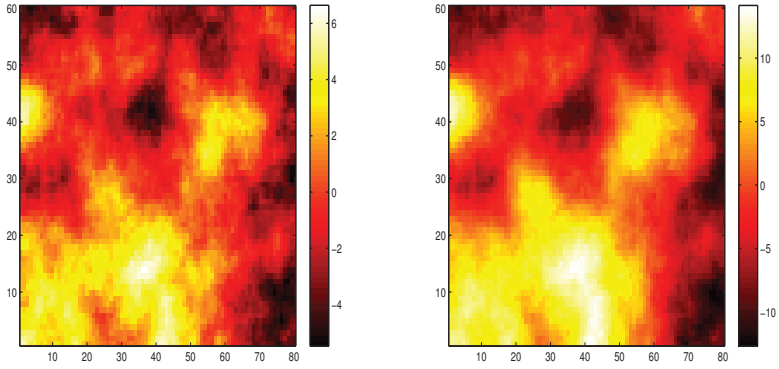
1. Calculate the Cholesky triangle $\boldsymbol{L}$ from the Cholesky factorization;
2. Sample $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix with the same dimensions as the precision matrix $\boldsymbol{Q}$;
3. Solve the linear system of equations $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{z}$ for $\boldsymbol{v}$. Then $\boldsymbol{v}$ has the correct precision matrix $\boldsymbol{Q}$, and $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$;
4. Finally correct the mean by computing $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{v}$.

Then $\boldsymbol{x}$ is a sample of the GMRF with correct mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$. If $\boldsymbol{Q}$ is a band matrix, a band-Cholesky factorization can be used with the algorithm given by Rue and Held (2005, Algorithm 2.9, Page 45). Different types of parametrization of GMRFs and their corresponding sampling procedures can be found in Rue and Held (2005). If the mean of the field is $\boldsymbol{0}$, the Step 4 is not needed and $\boldsymbol{v}$ is a sample from the GMRF.

With white noise at the right hand of SPDEs (28), and the parameters from the differential operators given in Table 2, we can get samples for positively correlated random fields with $b_{21} < 0$ and negatively correlated random fields with $b_{21} > 0$. These two samples are shown in Figure 2 and Figure 3. Choosing the reference points in the middle of the two GRFs for these two cases, we can get the corresponding marginal correlations within each of the GRFs and the cross-correlations between the two GRFs. The results are shown in Figure 4 and Figure 6. We can see that the random fields are isotropic and have the same correlation range. This is because

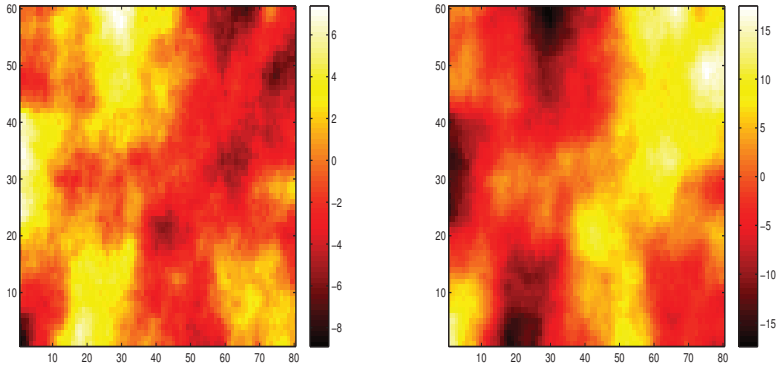
Table 2: Parameters for sampling bivariate GRFs

positively correlated GRFs			negatively correlated GRFs		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$	$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$	$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$	$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = 1$
$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$	$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$
$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$		$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$	
$\alpha_{n_2} = 0$	$\kappa_{n_2} = 0.3$		$\alpha_{n_2} = 0$	$\kappa_{n_2} = 0.3$	



(a) (b)

Figure 2: One simulated realization from the bivariate Gaussian Random Field with positive correlation with parameters $\alpha_{11} = 2, \alpha_{12} = 0, \alpha_{21} = 2, \alpha_{22} = 2, \alpha_{n_1} = 0, \alpha_{n_2} = 0$, $\kappa_{11} = 0.15, \kappa_{22} = 0.3, \kappa_{21} = 0.5, \kappa_{n_1} = 0.15, \kappa_{n_2} = 0.3$, $b_{11} = 1, b_{12} = 0, b_{22} = 1$ and $b_{21} = -1$.



(a) (b)

Figure 3: One simulated realization from the bivariate Gaussian Random Field with negative correlation with parameters $\alpha_{11} = 2, \alpha_{12} = 0, \alpha_{21} = 2, \alpha_{22} = 2, \alpha_{n_1} = 0, \alpha_{n_2} = 0,\ \kappa_{11} = 0.15, \kappa_{22} = 0.3, \kappa_{21} = 0.5, \kappa_{n_1} = 0.15, \kappa_{n_2} = 0.3,\ b_{11} = 1, b_{12} = 0, b_{22} = 1$ and $b_{21} = 1$.

we have chosen almost the same parameters for our cases, except the signs of b_{21}.

From Figure 2 - Figure 6 and many more samples from the bivariate GRFs which we have simulated, we can notice that the sign of $b_{21} \cdot b_{22}$ is related to the sign of the correlation between these two GRFs, which corresponds to the comparison result in (47). The smoothness of the GRFs are related to the values α_{ij}. These results also verify the conclusion given in (43).

We can also notice that the first GRF is a Matérn random field when we use the white noise process as the driving process or under the condition that $\kappa_{n_1} = \kappa_{11}$ with the triangular systems of SPDEs. The second field, in general, is not a Matérn random field, but it can be relatively close to a Matérn random field. With additional conditions, the second random fields could also be a Matérn random field, but this is not focused in this paper. With the triangular system of SPDEs, together with some other conditions, the correlation range which was mentioned in Section 1 for the first random field can be calculated by the empirically derived formula

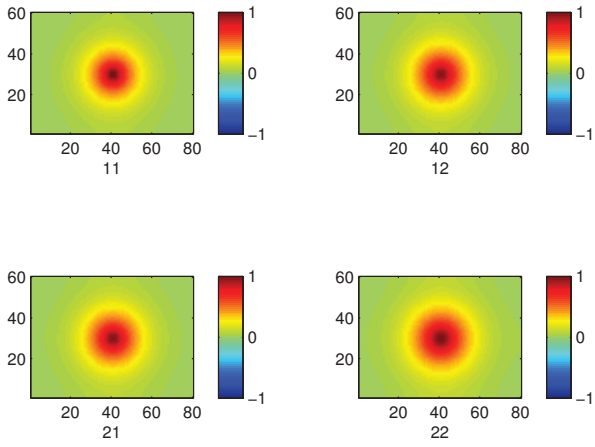


Figure 4: Marginal correlations (diagonal direction) and cross-correlations (anti-diagonal direction) with the reference points in the middle of the two GRFs for the positively correlated random fields.

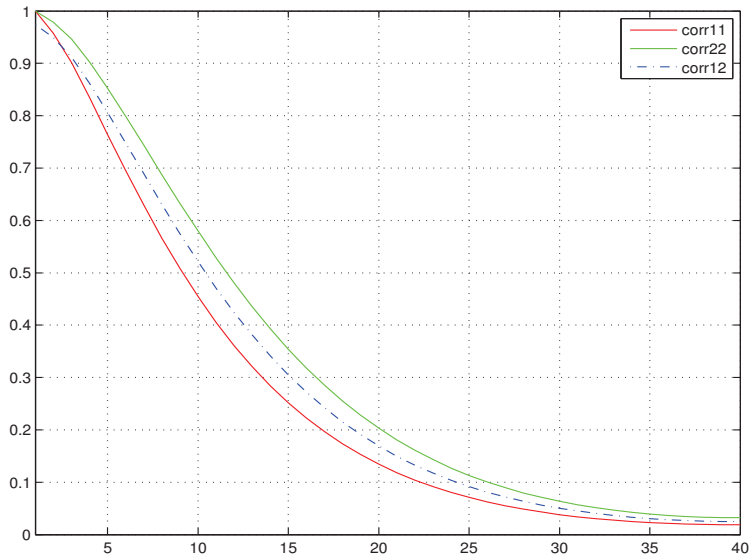


Figure 5: Marginal correlations and cross-correlations with the reference points in the middle of the two GRFs for the positively correlated random fields. 'corr11' means the marginal correlation within random field 1. 'corr22' means the marginal correlation within random field 2. 'corr12' means the cross-correlation between random fields 1 and 2.

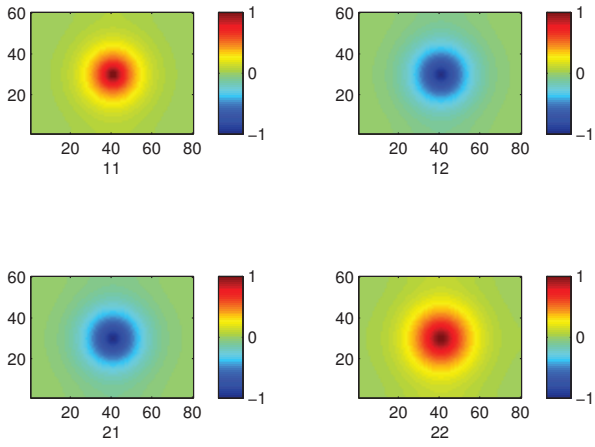


Figure 6: Marginal correlations (diagonal direction) and cross-correlations (anti-diagonal direction) with the reference points in the middle of the two GRFs for the negatively correlated random fields.

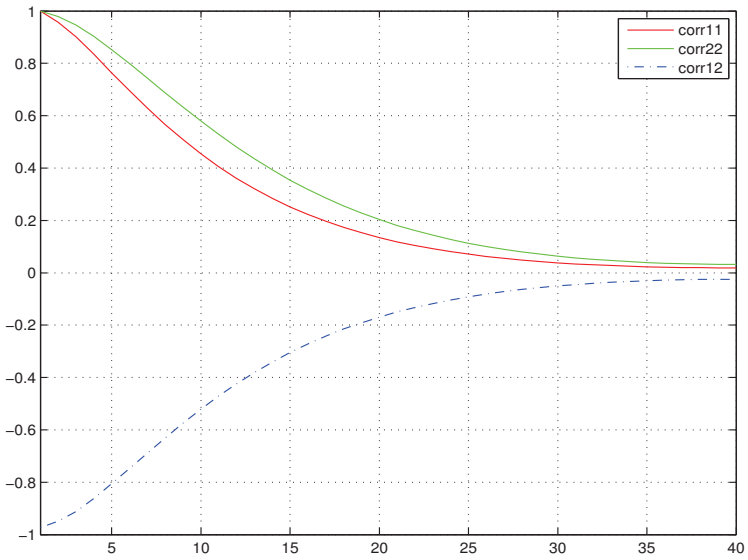


Figure 7: Marginal correlations and cross-correlations with the reference points in the middle of the two GRFs for the negatively correlated random fields. 'corr11' means the marginal correlation within random field 1. 'corr22' means the marginal correlation within random field 2. 'corr12' means the cross-correlation between random fields 1 and 2.

Table 3: Parameters for sampling trivariate GRFs

$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.5$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{13} = 0$	$\kappa_{13} = 0$	$b_{13} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.6$	$b_{21} = 0.8$
$\alpha_{22} = 2$	$\kappa_{22} = 0.4$	$b_{22} = 1$
$\alpha_{23} = 0$	$\kappa_{23} = 0$	$b_{23} = 0$
$\alpha_{31} = 2$	$\kappa_{31} = 0.5$	$b_{31} = 1$
$\alpha_{32} = 2$	$\kappa_{32} = 1$	$b_{32} = 0.9$
$\alpha_{33} = 2$	$\kappa_{33} = 0.3$	$b_{33} = 1$
$\alpha_{n_1} = 1$	$\kappa_{n_1} = 0.5$	
$\alpha_{n_2} = 1$	$\kappa_{n_2} = 0.4$	
$\alpha_{n_3} = 1$	$\kappa_{n_3} = 0.3$	

$\rho = \sqrt{8\nu}/\kappa$ but not for the second random field. In general we need to find the correlation ranges for the random fields numerically.

For trivariate GRFs the sampling procedure is exactly the same as for bivariate GRFs. We use the triangular system of the SPDEs. The true values given in Table 7 are used. One sample from the trivariate GRFs is shown in Figure 8 with the corresponding correlation functions given in Figure 9. From these figures, we notice that the trivariate random fields have similar interpretation as the bivariate random fields, but are more complicated. Since the triangular version of the system of SPDEs has been used, the sign of $b_{21} \cdot b_{22}$ is related to the sign of the correlation between the first two GRFs. But for the third fields, it is not only related to the sign of $b_{32} \cdot b_{33}$ but also the influence from the sign of $b_{21} \cdot b_{22}$. By choosing a different parametrization, we can end up with more interesting models such as the random field 1 and random field 3 are positively correlated in some locations and negatively correlated in other locations. We are not going to discuss these issues here since this is ongoing research.

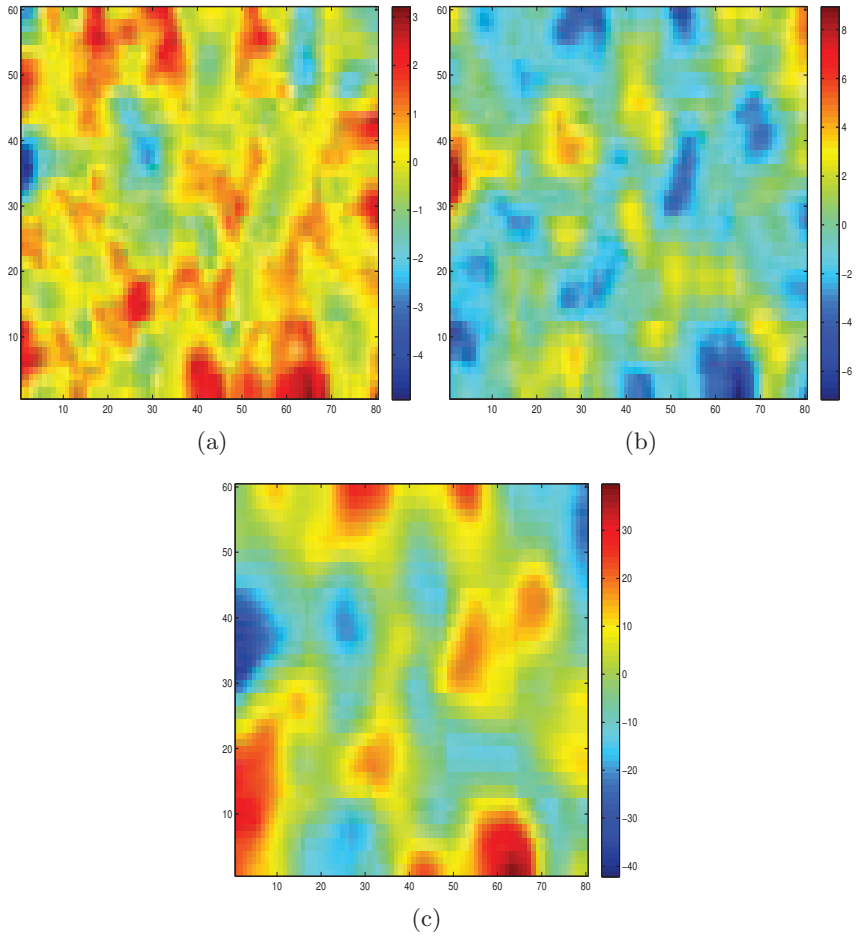


Figure 8: One simulated realization from the trivariate Gaussian Random Fields with parameters given in Table 3.

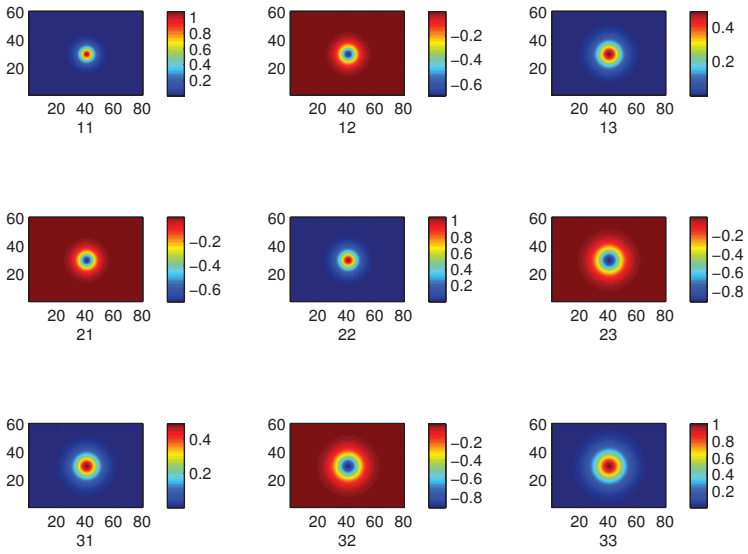


Figure 9: Marginal correlation matrices (diagonal) and cross-correlation matrices (off-diagonal) with each of the reference point in the middle of the three GRFs.

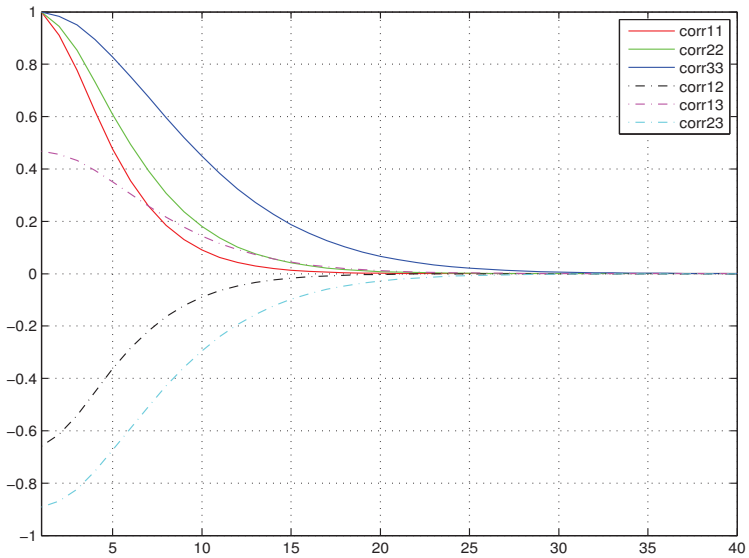


Figure 10: Marginal correlation matrices and cross-correlation matrices with each of the reference point in the middle of the three GRFs. 'corr11' means the marginal correlation within random field 1. 'corr22' means the marginal correlation within random field 2. 'corr33' means the marginal correlation within random field 3. 'corr12' means the cross-correlation between random fields 1 and 2. 'corr13' means the cross-correlation between random fields 1 and 3. 'corr23' means the cross-correlation between random fields 2 and 3.

4 Examples and applications

In order to illustrate how to use the SPDEs approach for constructing multivariate GRFs and the usefulness of our approach, some examples both with simulated data and real data are chosen. First, some basic theory on inference with multivariate GMRFs is given. We focus on the triangular system of SPDEs in this paper.

We use the bivariate GMRFs as an example. The multivariate GMRFs can be done analogously. Assume that we have used N triangles in the discretization for each random field $x_i(\boldsymbol{s})|\theta_i, i = 1, 2$. For the bivariate case $\boldsymbol{x}(\boldsymbol{s}) = (x(\boldsymbol{s})_1, x(\boldsymbol{s})_2)^{\mathrm{T}}$ has a $2N$-dimensional multivariate Gaussian distribution with probability density function

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{2N} |\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x}\right), \tag{50}$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix of the bivariate GMRF with the parameters $\boldsymbol{\theta}$. Furthermore, we assume the length of the data $\boldsymbol{y} = (y_1, y_2)^{\mathrm{T}}$ is $t = k_1 + k_2$ where y_1 is the observation of $x_1(\boldsymbol{s})$ with length k_1 and y_2 is the observation of $x_2(\boldsymbol{s})$ with length k_2. Then $\boldsymbol{y}$ is a t-dimensional random variable with probability density function

$$\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{t} |\boldsymbol{Q}_n|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{Q}_n(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right), \tag{51}$$

where $\boldsymbol{Q}_n$ is defined in Section 2.3 with size $t \times t$, and $\boldsymbol{A}$ is a $t \times 2N$ matrix which links the sparse observations to our bivariate GMRFs. Notice that the density function $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$. Hence we can write the probability density function as $\pi(\boldsymbol{y}|\boldsymbol{x})$. We first find the probability density function of $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}$ from (50) - (51)

$$\begin{aligned}
\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) \\
&= \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) \\
&\propto \exp\left(-\frac{1}{2}\left(x^{\mathrm{T}}(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{A})\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{y}\right)\right).
\end{aligned} \tag{52}$$

Let $\boldsymbol{\mu}_c(\boldsymbol{\theta}) = \boldsymbol{Q}_c^{-1}(\boldsymbol{\theta})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{y}$, and $\boldsymbol{Q}_c(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{A}$, and then $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ can be denoted as

$$\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_c(\boldsymbol{\theta}), \boldsymbol{Q}_c^{-1}(\boldsymbol{\theta})\right),$$

or in the canonical parametrization as

$$\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}_c\left(\boldsymbol{b}_c, \boldsymbol{Q}_c(\boldsymbol{\theta})\right),$$

with $\boldsymbol{b}_c = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{y}$. Thus $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}$ is a $2N$-dimensional multivariate Gaussian distribution. The canonical parametrization for a GMRF is useful with successive conditioning (Rue and Held, 2005). For more information about the canonical parametrization for the GMRFs, we refer to Rue and Held (2005, Chapter 2.2.3).

The probability density function (52) can be used to integrate out $\boldsymbol{x}$ from the joint density of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\theta}$,

$$\begin{aligned}\pi(\boldsymbol{y},\boldsymbol{\theta}) &= \frac{\pi(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \\ &= \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})},\end{aligned} \tag{53}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. With (53), the posterior distribution of $\boldsymbol{\theta}$ can be obtained as

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})&\frac{|\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2}|\boldsymbol{Q_n}|^{1/2}}{|\boldsymbol{Q}_c(\boldsymbol{\theta})|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x}\right) \\ &\times \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{Q}_n(\boldsymbol{\theta})(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\right) \\ &\times \exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c(\boldsymbol{\theta}))^{\mathrm{T}}\boldsymbol{Q}_c(\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\mu}_c(\boldsymbol{\theta}))\right).\end{aligned} \tag{54}$$

The quadratic terms in the exponential functions in (54) can be simplified by using $\boldsymbol{\mu}_c(\boldsymbol{\theta})$ and $\boldsymbol{Q}_c(\boldsymbol{\theta})$. It is also convenient to use the logarithm of the posterior distribution of $\boldsymbol{\theta}$. Reorganize (54) to get the formula which will be used for inference

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \text{Const} &+ \log(\pi(\boldsymbol{\theta})) + \frac{1}{2}\log(|\boldsymbol{Q}(\boldsymbol{\theta})|) \\ &- \frac{1}{2}\log(|\boldsymbol{Q}_c(\boldsymbol{\theta})|) + \frac{1}{2}\boldsymbol{\mu}_c(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{Q}_c(\boldsymbol{\theta})\boldsymbol{\mu}_c(\boldsymbol{\theta}).\end{aligned} \tag{55}$$

From (55) we can see that it is difficult to handle the posterior distribution of $\boldsymbol{\theta}$ analytically since both the determinants and the quadratic terms

are hard to handle. Thus numerical methods should be applied for the statistical inference in this paper.

Furthermore, even though it is not the topic for our paper, we point out that it is also possible to obtain the probability density function $\pi(\boldsymbol{x}|\boldsymbol{y})$ by integrate out $\boldsymbol{\theta}$. But this is difficult and needs to be obtained numerically using the following expression (56)

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \int_{\mathbb{R}^m} \pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \tag{56}$$

where m is the dimension of $\boldsymbol{\theta}$. When the dimension of $\boldsymbol{\theta}$ is large, then this integration might be infeasible in practice.

4.1 Statistical inference with simulated data

First, we illustrate how to do the statistical inference for simulated data with the SPDEs approach with known true parameter values. These datasets both contain one realization with 2000 observations at different locations. The parameters used for generating the simulated data are presented in Table 4. To make it simpler, the nugget effects are assumed to be known, $\tau_1 = \tau_2 = 0.001$. As discussed in Section 1 it is hard to estimate the smoothness parameters, so we fix $\{\alpha_{ij}; i, j = 1, 2\}$ in the system of SPDEs to the known values. The smoothness parameters α_{n1} and α_{n_2} for the noise processes are also fixed to the known values due to the same reason. The scaling parameters κ_{n_1} and κ_{n_2} for the noise processes are restricted with $\kappa_{n_1} = \kappa_{11}$ and $\kappa_{n_2} = \kappa_{22}$ for the simulated data. Thus in the simulated data examples, only $\boldsymbol{\theta} = \{\kappa_{11}, \kappa_{21}, \kappa_{22}, b_{11}, b_{21}, b_{22}\}$ needed to be estimated. Since κ_{11}, κ_{21} and κ_{22} have to take positive values, we a priori assign *log-normal* distributions with mean zero and large variances for each of the parameters. b_{11}, b_{21} and b_{22} are given *normal* priors with mean zero and large variances.

In the first example, the two GRFs are negatively correlated and the realizations are shown in Figure 11(a) and Figure 11(b). The corresponding estimated conditional mean for the negatively correlated GRFs are given in Figure 11(c) and Figure 11(d). We can notice that there are no large differences between the estimated conditional mean for the GRFs and the true bivariate GRFs. The estimates for the parameters are given in Table 5 with their standard derivations. From this table, we notice that

the estimates for all the parameters are quite accurate with accuracy to 2 digits. The true values of the parameters are within 1 standard deviation from the estimates.

The second example with the simulated data uses two GRFs that are positively correlated, and the realizations are shown in Figure 12(a) and Figure 12(b). The corresponding estimated conditional mean for the bivariate GRFs are illustrated in Figure 12(c) and Figure 12(d). From Figure 12(a) to Figure 12(d), we can again notice that the estimated conditional mean for the positively correlated bivariate GRFs are almost the same as the true bivariate GRFs. The estimates are accurate and the true values of the parameters are within 1 standard deviation from the estimates.

The third example uses a trivariate GRF. We reuse the parameters given in Table 7 as true parameters for the simulated data. One realization from the trivariate GRF is given in Figure 8 in Section 3.4. The nugget effects are assumed to be known as $\tau_1 = \tau_2 = \tau_3 = 0.01$. In this example, we also fix $\{\alpha_{ij}; i, j = 1, 2\}$ and $\{\alpha_{n_i}, i = 1, 2, 3\}$. The scaling parameters for the noise processes are fixed with similar settings as the bivariate GRFs $\{\kappa_{n_i} = \kappa_{ii}; i = 1, 2, 3\}$. Therefore we have 12 parameters to be estimated. They are b_{ij} and $\{\kappa_{ij}; i, j = 1, 2, 3, i \geq j\}$. The estimates for these parameters are given in Table 7. We can notice that the estimates are accurate in this example since all the estimated are accurate to 2 digits, and the true values of the parameters are within 2 standard deviations from the corresponding estimates. The estimated conditional mean for the trivariate GRF are given in Figure 13(a) - Figure 13(c). Comparing the results given in Figure 13 with the corresponding fields given in Figure 8, we can see that there are no large differences between the true random fields and the estimated conditional mean. From the examples we can notice that our approach works well not only for bivariate GRFs but also for multivariate GMRFs.

4.2 Inference with real data

We illustrate how to use the SPDEs approach for multivariate data in real application in this section. The meteorological dataset used by Gneiting et al. (2010) is analyzed in this paper. This meteorological dataset contains one realization consisting of 157 observations of both temperature and pressure in the north American Pacific Northwest, and the tempera-

Table 4: Parameters for simulating the bivariate GRFs

dataset 1			dataset 2		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$	$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$\kappa_{11} = 0.3$	$b_{11} = 1$	$\alpha_{11} = 2$	$\kappa_{11} = 0.15$	$b_{11} = 1$
$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$	$\alpha_{12} = 0$	$\kappa_{12} = 0$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = 1$	$\alpha_{21} = 2$	$\kappa_{21} = 0.5$	$b_{21} = -1$
$\alpha_{22} = 2$	$\kappa_{22} = 0.4$	$b_{22} = 1$	$\alpha_{22} = 2$	$\kappa_{22} = 0.3$	$b_{22} = 1$
$\alpha_{n_1} = 1$	$\kappa_{n_1} = 0.3$		$\alpha_{n_1} = 0$	$\kappa_{n_1} = 0.15$	
$\alpha_{n_1} = 0$	$\kappa_{n_2} = 0.4$		$\alpha_{n_2} = 0$	$\kappa_{n_2} = 0.3$	

Table 5: Inference with simulated dataset 1

Parameters	True value	Estimated	Standard deviations
κ_{11}	0.3	0.295	0.019
κ_{21}	0.5	0.471	0.044
κ_{22}	0.4	0.380	0.020
b_{11}	1	1.009	0.069
b_{21}	1	1.032	0.064
b_{22}	1	0.997	0.059

Table 6: Inference with simulated dataset 2

Parameters	True value	Estimated	Standard deviations
κ_{11}	0.15	0.139	0.124
κ_{21}	0.5	0.487	0.059
κ_{22}	0.3	0.293	0.061
b_{11}	1	0.991	0.017
b_{21}	-1	-1.002	0.033
b_{22}	1	1.011	0.018

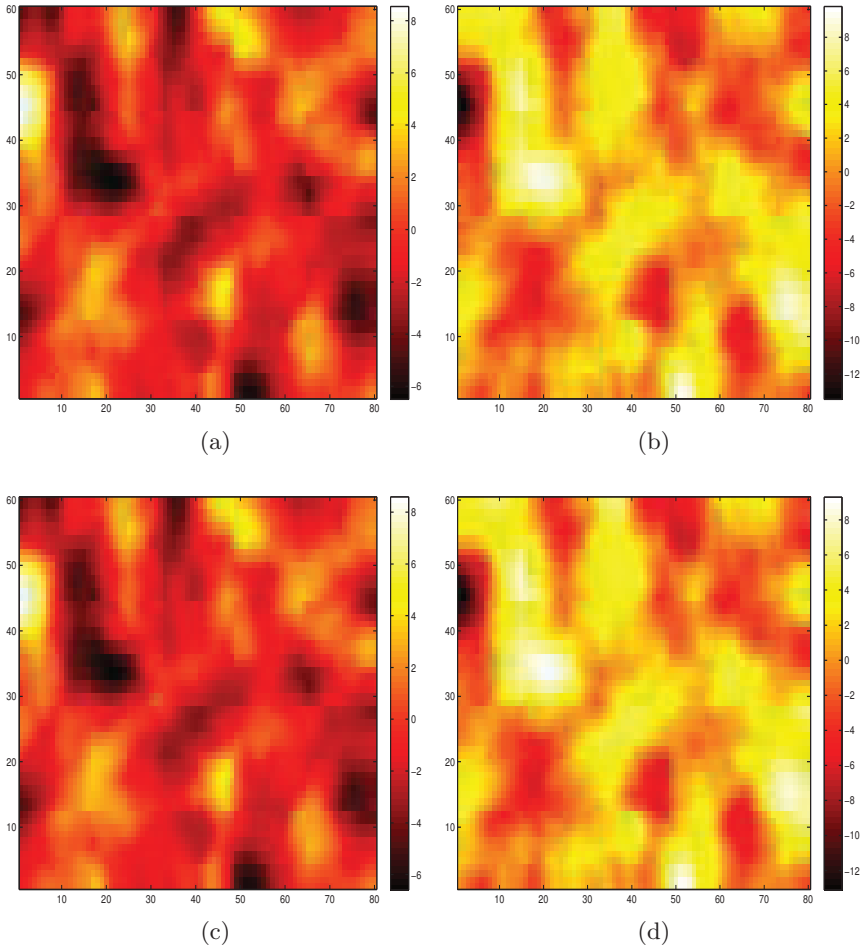


Figure 11: The true bivariate GRFs (a) - (b) and the estimated conditional mean for the bivariate GRFs (c) - (d) for simulated dataset 1.

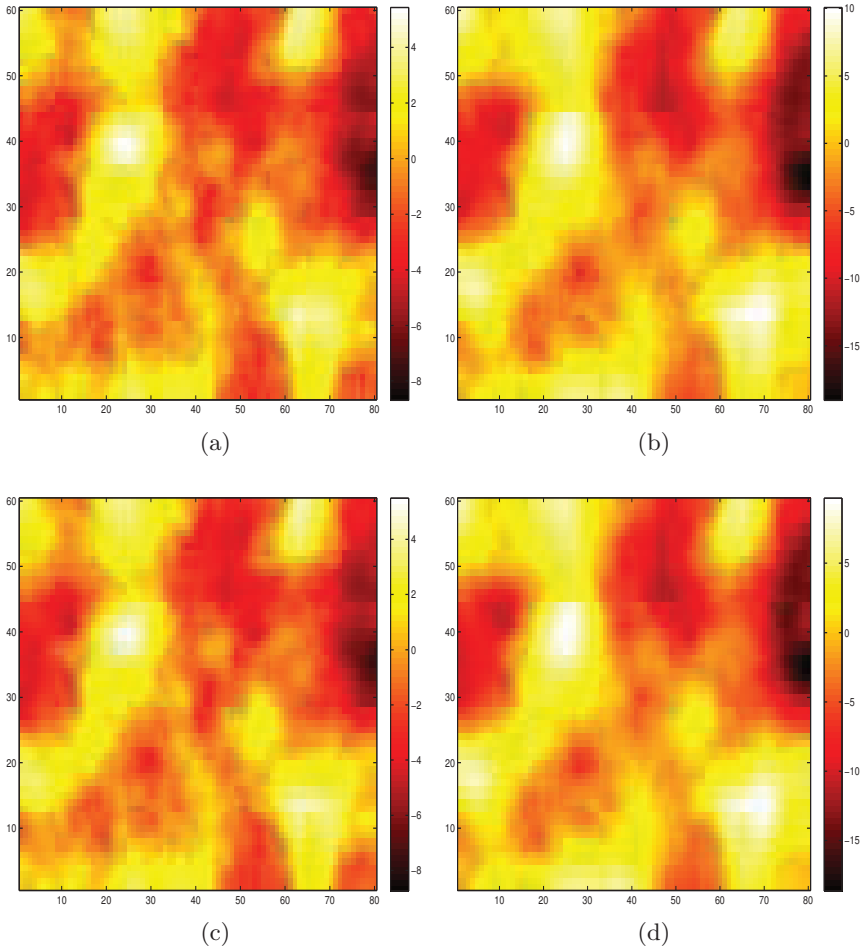


Figure 12: The true bivariate GRFs (a) - (b) and the estimated conditional mean for the bivariate GRFs (c) - (d) for simulated dataset 2.

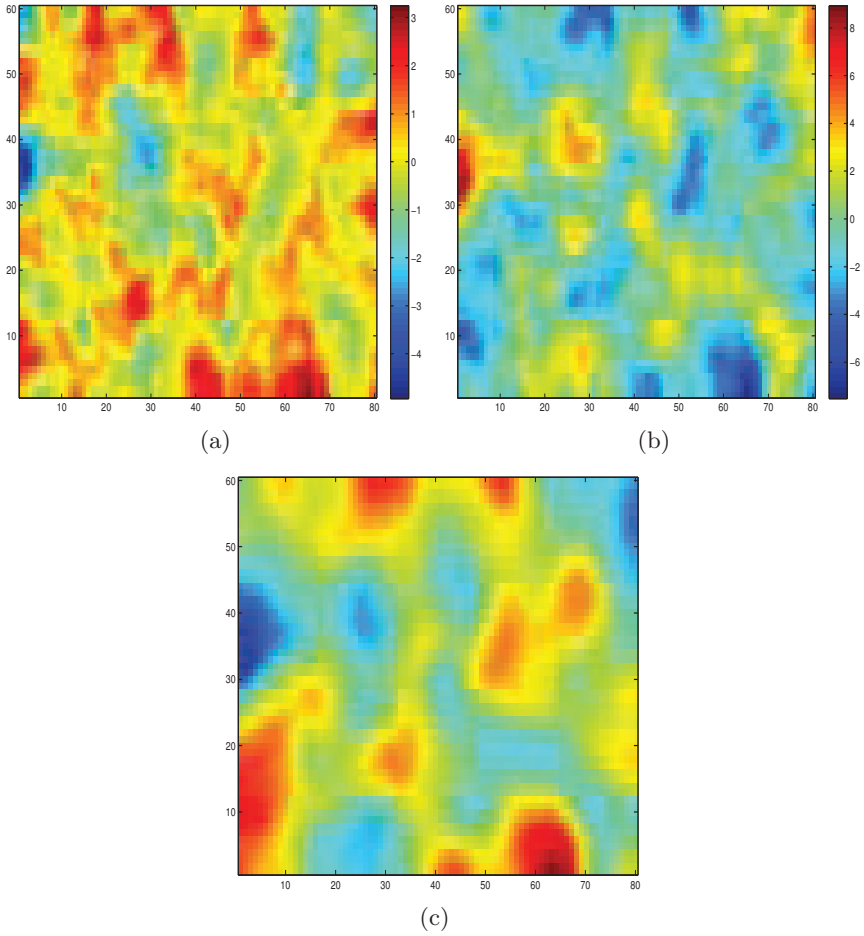


Figure 13: The estimated conditional mean for the trivariate GRFs.

Table 7: Inference for the trivariate GRF

Parameters	True value	Estimated	Standard deviations
b_{11}	1	1.002	0.014
b_{21}	0.8	0.807	0.024
b_{22}	1	0.984	0.013
b_{31}	1	0.958	0.031
b_{32}	0.9	0.878	0.026
b_{33}	1	0.986	0.012
κ_{11}	0.5	0.481	0.022
κ_{21}	0.6	0.582	0.044
κ_{22}	0.4	0.393	0.021
κ_{31}	0.5	0.564	0.050
κ_{32}	1	1.008	0.034
κ_{33}	0.3	0.286	0.012

ture and pressure are always observed at the same locations. It contains the following data: pressure errors (in Pascal) and temperature errors (in Kelvin), measured against longitude and latitude. The errors are defined as the forecasts minus the observations. The data are valid on 18th, December of 2003 at 16:00 local time, at a forecast horizon of 48 hours. For information about the data, see, for example, Eckel and Mass (2005) and Gneiting et al. (2010). Gneiting et al. (2010) have chosen this dataset with the aim of doing probabilistic weather field forecasting, which is a big research area (Gel et al., 2004; Berrocal et al., 2007; Sloughter et al., 2007; Berrocal et al., 2008). One of the main focuses in this area is to fit and sample spatially correlated error fields. This aim fits our motivation well since the SPDEs approach can be applied to construct multivariate random fields in order to capture the dependence structures not only within the fields but also between the fields.

The main aim of this example is to illustrate how to fit a bivariate random field with the SPDEs approach for pressure and temperature errors data. This bivariate random field can be used to explain the features of the two random fields as in Gneiting et al. (2010). The constructed bivariate spatial random fields are used to represent the error fields for temperature and pressure. It is known that the temperature and the pressure are neg-

atively correlated (Courtier et al., 1998; Ingleby, 2001). As pointed out by Gneiting et al. (2010), the forecast fields are usually smooth fields. The observation field for the pressure is smooth. However, the observation field for temperature is rough. So the model should set up to give the same type of behavior. Without any confusion, we will call the temperature error field the temperature and the pressure error field the pressure. In general, we need to choose the order of the fields at the modelling stage. The simple way is fitting the data with both the orders and selecting the best one using some predefined scoring rules. For more information about the scoring rules, see for example, Gneiting et al. (2005). In this paper we will set the first field $x_1(\boldsymbol{s})$ as the pressure and the second field $x_2(\boldsymbol{s})$ as the temperature.

The bivariate random field with our approach is constructed with a system of SPDEs,

$$\begin{aligned}
&(\kappa_{n_1}^2 - \Delta)^{\alpha_{n_1}/2} f_1(\boldsymbol{s}) = \mathcal{W}_1(\boldsymbol{s}),\\
&(\kappa_{n_2}^2 - \Delta)^{\alpha_{n_2}/2} f_2(\boldsymbol{s}) = \mathcal{W}_2(\boldsymbol{s}),\\
&b_{11}(\kappa_{11}^2 - \Delta)^{\alpha_{11}/2} x_1(\boldsymbol{s}) = f_1(\boldsymbol{s}),\\
&b_{22}(\kappa_{22}^2 - \Delta)^{\alpha_{22}/2} x_2(\boldsymbol{s}) + b_{21}(\kappa_{21}^2 - \Delta)^{\alpha_{21}/2} x_1(\boldsymbol{s}) = f_2(\boldsymbol{s}).
\end{aligned} \tag{57}$$

Since the main purpose of this example is to illustrate that we can construct the same (or similar) model as the covariance-based approach presented by Gneiting et al. (2010), only the models which can result in similar models as theirs have been chosen from the parameter matching results given in Section 3.3. We assume that $\kappa_{n_1} = \kappa_{11}$ and $\kappa_{n_2} = \kappa_{22}$. This particular setting corresponds to the first random field $x_1(\boldsymbol{s})$ being a Matérn random field and the second random field $x_2(\boldsymbol{s})$ being close to a Matérn random field. These settings make the model constructed through our SPDEs approach closer to the Gneiting et al. (2010) approach.

The results for the estimates with the SPDEs approach are given in Table 8. From the table we can notice that we can capture the negative dependence structure between temperature and pressure since $b_{21} > 0$. The estimated co-located correlation coefficient $\rho_{pt} = -0.53$ which is quite similar as the result from Gneiting et al. (2010). The standard deviations for pressure and temperature are $\sigma_P = 202.1$ Pascal and $\sigma_T = 2.76$ Celsius degrees. In order to compare the predictive perform between our approach

and approach proposed by Gneiting et al. (2010), we randomly leave out 25 observations from each field and use only 132 observations for parameters estimation. The relative error $\mathscr{E}$ has been chosen to compare the predictions and is defined as

$$\mathscr{E} = \frac{\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|_2}{\|\hat{\boldsymbol{y}}\|_2}.$$

where $\|\cdot\|_2$ denotes the 2-norm. $\hat{\boldsymbol{y}}$ denotes the vector of predictions for the observations $\boldsymbol{y}$. The predictive performances for these two approaches are given in Figure 14. We can notice that the results from our SPDEs approach and the approach in Gneiting et al. (2010) are quite similar. Table 9 shows the corresponding predictive errors for these two approaches. From this table, we can notice that our model gives slightly better results than the covariance-based model but the difference is not statistically significant. Another merit with our approach which has not been discussed until now is that the SPDEs approach in general is much more computationally efficient since the precision matrix $\boldsymbol{Q}$ is sparse. The conditional mean in 3D and 2D for the bivariate GRF are shown in Figure 15 and Figure 16 for our approach, respectively. The corresponding results for the covariance-based approach are shown in Figure 17 and Figure 18. As we can see from the 3D and 2D figures, the bivariate GRFs from our approach and from the covariance-based approach give quite similar results.

For the nugget effects (measurement error variances) of pressure and temperature $\boldsymbol{\tau} = (\tau_1, \tau_2)^{\mathrm{T}}$, we are going to use an iterative bias correction to estimate them. This is based on formula

$$\sigma_{ij}^2 = \tau_i^2 + V_{ij}, \quad i = 1, 2, \quad j = 1, \ldots, n, \tag{58}$$

where i indicates the pressure with $i = 1$ and the temperature with $i = 2$. $\sigma_{ij}^2 = \mathrm{Var}(y_{ij} - \hat{y}_{ij})$ is the variance which contains the nugget effects and the kriging variances. $\{\tau_i^2 = \mathrm{Var}(y_{ij} - x_{ij}); i = 1, 2, j = 1, \ldots, n\}$ are the nugget effects for the pressure when $i = 1$ or temperature when $i = 2$. $\{V_{ij} = \mathrm{Var}(x_{ij} - \hat{y}_{ij}); i = 1, 2, j = 1, \ldots, n\}$ are the kriging variances and they are from the model bias. n denotes the total number of data points in each field. y_{ij} denotes the observed value at data point for each field. $\hat{y}_{ij}$ is the predicated values from a given model and x_{ij} denotes the true value which is unknown. See Appendix for a simple proof of (58).

Table 8: Estimated parameters for the SPDEs approach

κ_{11}	κ_{21}	κ_{22}	b_{11}	b_{21}	b_{22}
6.353×10^{-3}	0.413	2.243×10^{-2}	0.2165	1.298×10^{-4}	5.458

Table 9: Predictive relative errors for the SPDEs approach and the covariance-based models

relative errors		
Models	random field 1	random field 2
SPDEs approach	0.777	0.690
Covariance-based model	0.821	0.716

One simple way is to use Equation (59) where we just subtract the kriging variance V_{ij} from the empirical variance $\hat{\sigma}_{ij}^2$ to get the estimate of nugget effect $\hat{\tau}_i^2$ for each field

$$\hat{\tau}_i^2 = \frac{1}{n}\sum_j (\hat{\sigma}_{ij}^2 - V_{ij}), \tag{59}$$

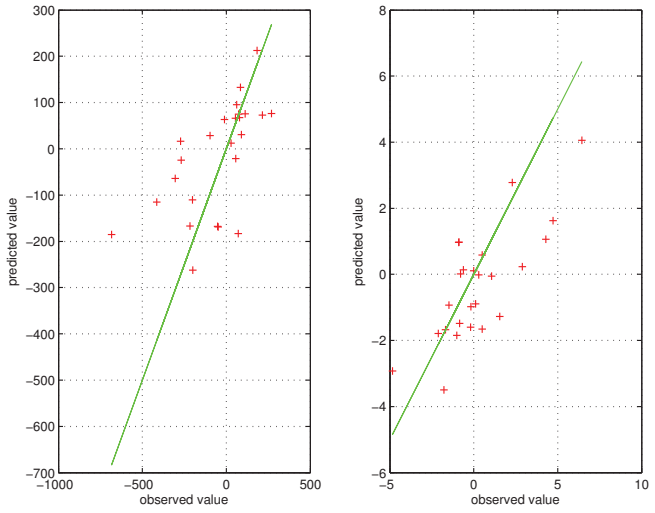
where $\hat{\sigma}_{ij}^2 = (y_{ij} - \hat{y}_{ij})^2$. But another preferable way which we have used in this paper is with the formula

$$\hat{\tau}_i^2 = \sum_j \left\{ w_{ij} \left(\hat{\sigma}_{ij}^2 - V_{ij} \right) \right\}, \tag{60}$$

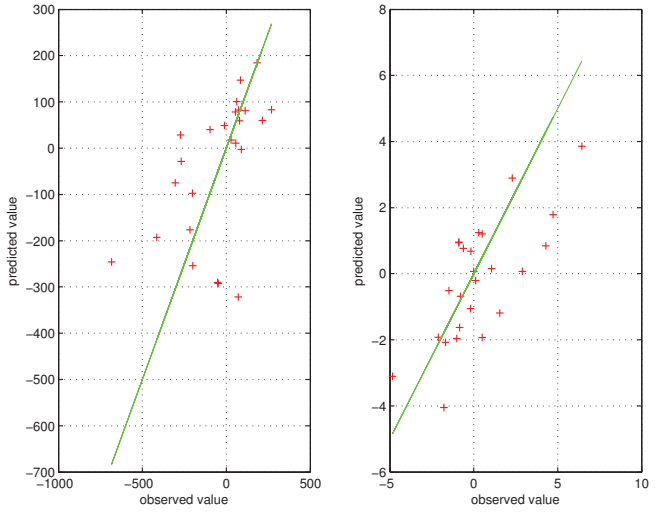
where $w_{ij} = \frac{1/(\tau_i^2 + V_{ij})}{\sum_j \left(1/(\tau_i^2 + V_{ij})\right)}$. This can give an unbiased estimator of nugget effects. See Appendix for the proof. The initial values are chosen to be similar to the results given in Gneiting et al. (2010), and the results from the bias correction approach are shown in Table 10. The convergence curves of the nugget effects for the fields are illustrated in Figure 19.

5 Discussion

Spatial modelling of multivariate data are of demand in many areas, such as weather forecasting (Courtier et al., 1998; Reich and Fuentes, 2007), air quality (Brown et al., 1994; Schmidt and Gelfand, 2003), economics



(a)



(b)

Figure 14: The predictive performances of SPDEs approach (a) and the covariance-based approach (b)

Table 10: The nugget effects for pressure and temperature

	initial values	estimated values
τ_1	68.4	81.1
τ_2	0.01	0.53

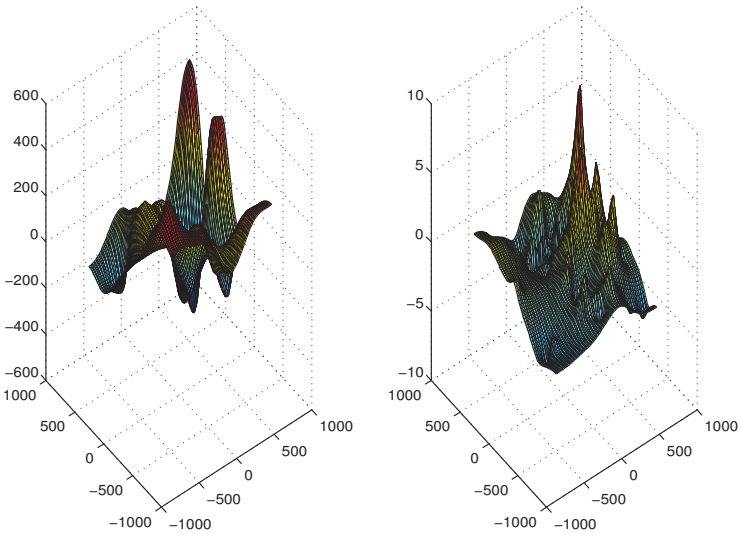


Figure 15: Estimated conditional mean for bivariate GRFs for SPDEs approach in 3D.

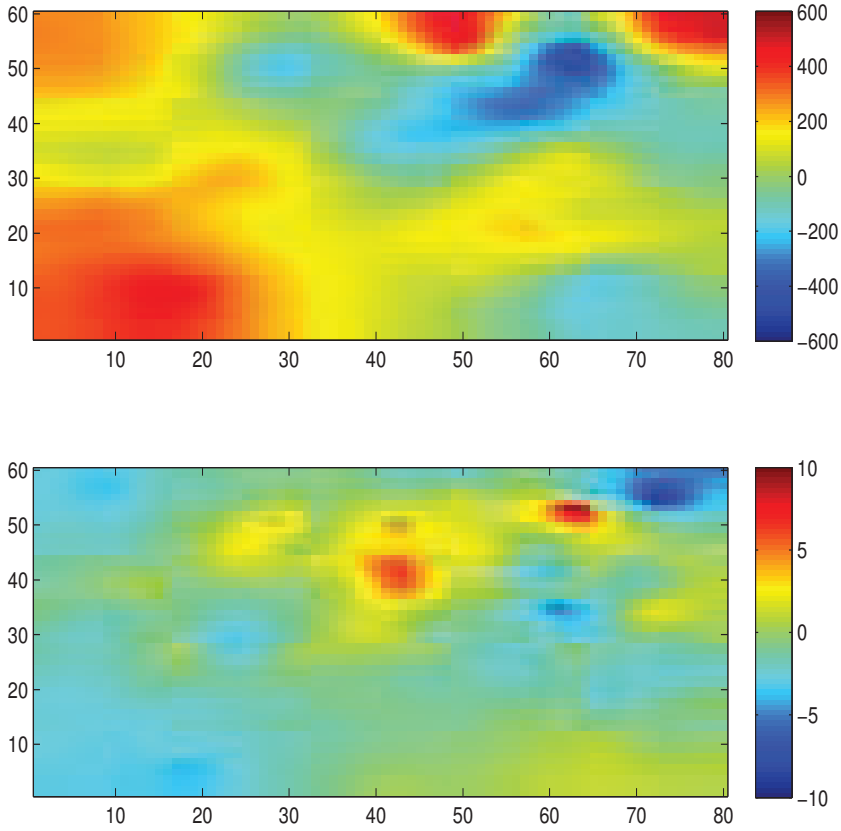


Figure 16: Estimated conditional mean for bivariate GRFs for SPDEs approach in 2D.

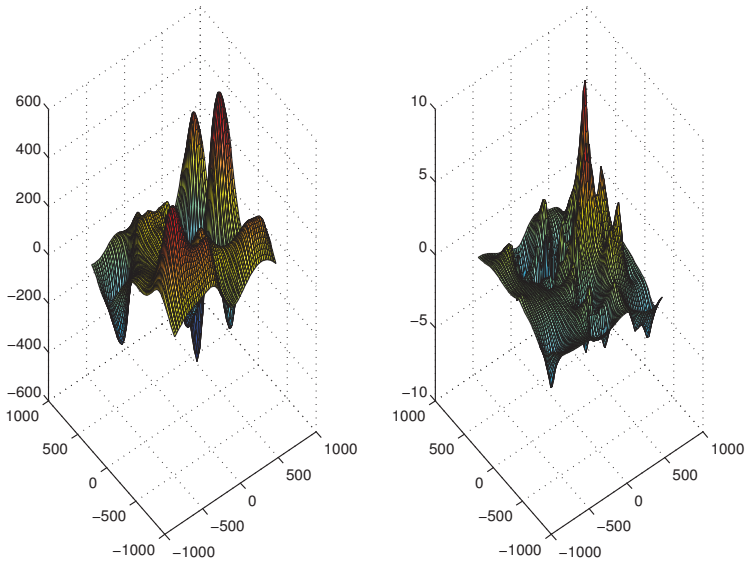


Figure 17: Estimated conditional mean for bivariate GRFs for covariance-based model in 3D.

(Gelfand et al., 2004; Sain and Cressie, 2007). The important issue for modelling spatial multivariate data is that the approach can not only model the marginal covariances within each field, but also has the ability to model the cross-covariances between the random fields. In addition, we need to solve the theoretical challenge for the positive definite constraint of the covariance functions and the computational challenge for large dataset.

The main aim of this work is to illustrated the possibility of constructing *multivariate* GRFs through the SPDEs approach. We notice that the parameters in the SPDEs approach is interpretable and can link our approach to the covariance-based approach. By using the approximate weak solution of the corresponding system of SPDEs, we can represent multivariate GRFs by GMRFs. Since the precision matrices of the GMRFs are sparse, numerical algorithms for sparse matrices can be applied, and therefore fast sampling and inference are feasible. Our approach inherited the properties from the approach discussed by Lindgren et al. (2011). There are three main advantages for our newly proposed SPDEs approach. The first advantage is that our new SPDEs approach does not depend on

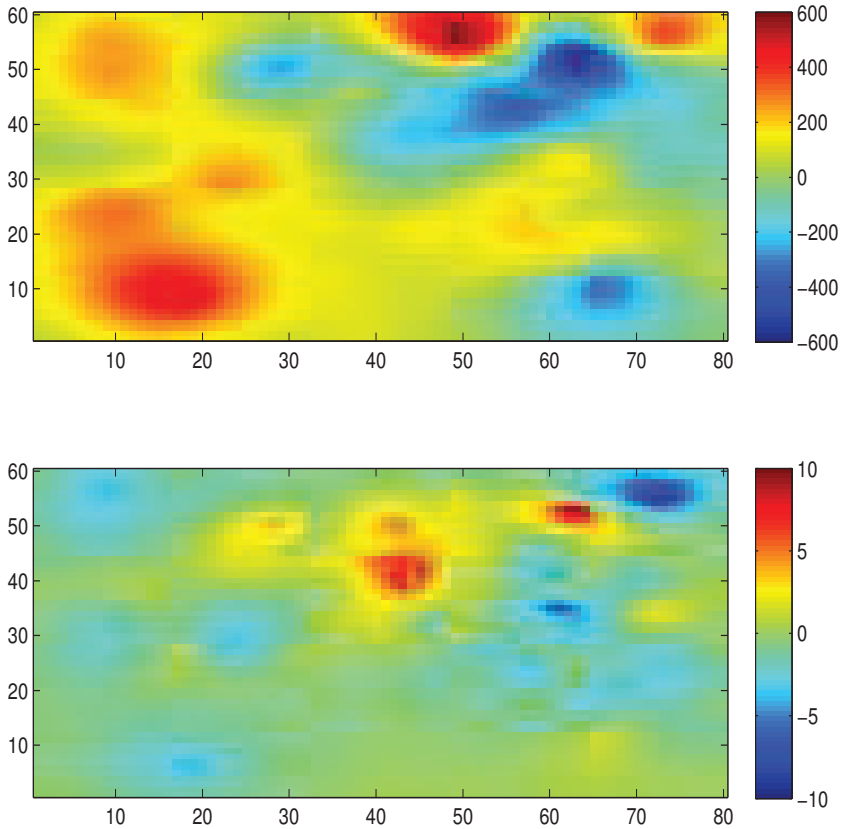


Figure 18: Estimated conditional mean for bivariate GRFs for covariance-based model in 2D.

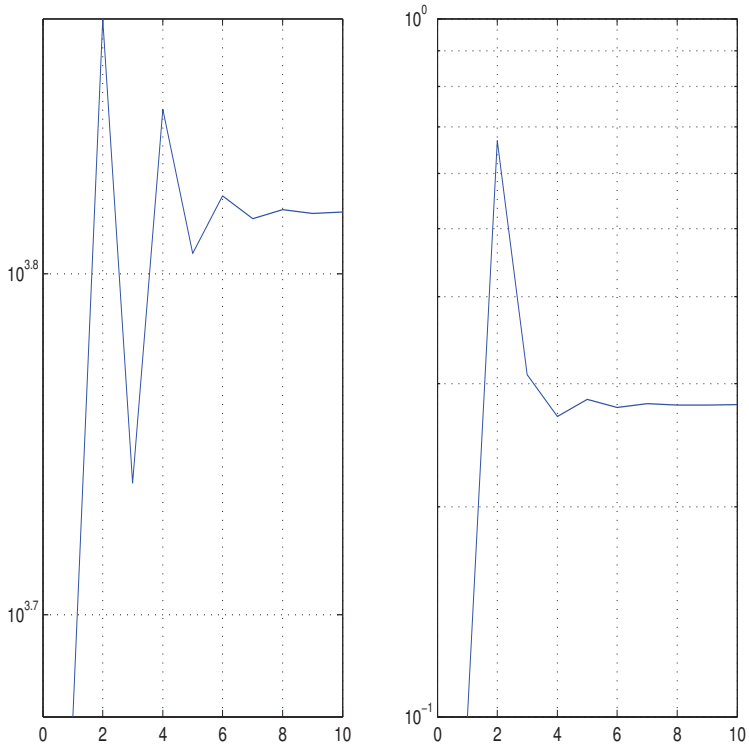


Figure 19: Bias correction iteration. The curves are for τ_1^2 and τ_2^2 respectively

direct construction of positive definite matrix. The notorious requirement of positive definite covariance matrix is fulfilled automatically. The second advantage is that we can remove the symmetry property constraint shared by both the covariance-based approach Gneiting et al. (2010) and the LMC approach (Gelfand et al., 2004; Gneiting et al., 2010). The third advantage is that we can construct the multivariate GRFs on manifolds, such as on the sphere $\mathbb{S}^2$. The extension follows the discussion given by Lindgren et al. (2011). We just need to reinterpret the systems of SPDEs to be defined on the manifold.

One issue that needs to be pointed out is that we have chosen $\kappa_{n_1} = \kappa_{11}$ and $\kappa_{n_2} = \kappa_{22}$ in the model. This restriction used in this paper is to make the model closer to the models presented by Gneiting et al. (2010) and also make the inference easier. This may not be needed in other applications and κ_{n_1} and κ_{n_2} can be estimated from the data. However, κ_{n_1} and κ_{11} might be not distinguishable with the triangular systems of SPDEs. So one suggestion is that if we are use the triangular systems of SPDEs, then we fix κ_{n_1} based on some other information or set $\kappa_{n_1} = \kappa_{11}$ when doing inference.

Another issue that needed to be pointed out is that we have prosecuted the *full* version of the SPDEs approach but have not used in all the examples. This version could give us more flexibility to construct multivariate GRFs. The modelling procedures and the inference are the same as for the *triangular* version of the SPDEs which were used and discussed extensively in this paper.

We also want to point out that the proposed approach costs effort during the implementation and pre-processing stages since we need to build the system of SPDEs, discretize the fields and do the approximation to obtain a GMRF representation. But we believe, as pointed out by Lindgren et al. (2011), that " *such costs are unavoidable when efficient computations are required*".

Similarly, as pointed out by Lindgren et al. (2011), although the SPDEs approach presented here is not generally applicable for all covariance functions, it covers many useful model in spatial statistics. And it is possible to extend our approach to construct even richer class of models by following the discussion given by, for instance, Bolin and Lindgren (2011) and Fuglstad (2011). Our approach extends the applicability of multivari-

ate GRFs in practical applications since we can build and interpret the model using GRFs but do computation with GMRFs. It is further possible to include this approach in the integrated nested Laplace approximation (INLA) framework (Rue et al., 2009). These extensions are under study.

Appendix

In order to prove Equation (58), we need to write down the expression explicitly. Let y_{ij} denote the observed value for each field, $\hat{y}_{ij}$ the predicated value from a given model, and x_{ij} the true value. Then we have

$$\begin{aligned}
\mathrm{Var}(y_{ij} - \hat{y}_{ij}) &= \mathrm{Var}\left((y_{ij} - x_{ij}) + (x_{ij} - \hat{y}_{ij})\right) \\
&= \mathrm{Var}\left(y_{ij} - x_{ij}\right) + \mathrm{Var}\left(x_{ij} - \hat{y}_{ij}\right) \\
&\quad + 2\mathrm{Cov}\left(y_{ij} - x_{ij}, x_{ij} - \hat{y}_{ij}\right) \\
&= \mathrm{Var}\left(y_{ij} - x_{ij}\right) + \mathrm{Var}\left(x_{ij} - \hat{y}_{ij}\right),
\end{aligned} \tag{61}$$

since

$$\begin{aligned}
\mathrm{Cov}\left(y_{ij} - x_{ij}, x_{ij} - \hat{y}_{ij}\right) &= \mathbb{E}\left((y_{ij} - x_{ij})(x_{ij} - \hat{y}_{ij})\right) \\
&= \mathbb{E}\left[\mathbb{E}\left((y_{ij} - x_{ij})(x_{ij} - \hat{y}_{ij})|\boldsymbol{y}\right)\right] \\
&= \mathbb{E}\left[(y_{ij} - \hat{y}_{ij}) \times 0\right] \\
&= 0.
\end{aligned} \tag{62}$$

So Equation (58) is established.

Now we show that with the weights $w_{ij} = \frac{1/(\tau_i^2 + V_{ij})}{\sum_j (\tau_i^2 + V_{ij})}$, the estimator in Equation (60) is an unbiased estimator for the nugget effects. Using Equation (61), we know that $(y_{ij} - \hat{y}_{ij}) \sim \mathcal{N}(0, \tau_i^2 + V_{ij})$. Then

$$\begin{aligned}
\mathrm{Var}\left((y_{ij} - \hat{y}_{ij})^2\right) &= \mathbb{E}\left((y_{ij} - \hat{y}_{ij})^4\right) - \left(\mathbb{E}\left((y_{ij} - \hat{y}_{ij})^2\right)\right)^2 \\
&= 3(\tau_i^2 + V_{ij})^2 - (\tau_i^2 + V_{ij})^2 \\
&= 2(\tau_i^2 + V_{ij})^2
\end{aligned} \tag{63}$$

and thus the estimator is unbiased since the scaling constant cancels in the weight coefficient.

References

R.J. Adler and J.E. Taylor. *Random fields and geometry*. Springer Verlag, 2007. ISBN 0387481125.

D.J. Allcroft and C.A. Glasbey. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):487–498, 2003. ISSN 1467-9876.

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman & Hall, 2004. ISBN 158488410X.

K.J. Bathe. *Finite element method*. Wiley Online Library, 2008.

V.J. Berrocal, A.E. Raftery, and T. Gneiting. Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Monthly Weather Review*, 135(4):1386–1402, 2007.

V.J. Berrocal, A.E. Raftery, and T. Gneiting. Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *The Annals of Applied Statistics*, 2(4):1170–1193, 2008.

D. Bolin and F. Lindgren. Wavelet markov models as efficient alternatives to tapering and convolution fields. Technical report, Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University, 2009.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550, 2011.

P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994.

P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

N.A.C. Cressie. *Statistics for spatial data*, volume 298. Wiley-Interscience, 1993.

P.J. Diggle and P.J. Ribeiro Jr. *Model-based Geostatistics*. Springer, 2006.

P.J. Diggle, JA Tawn, and RA Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998. ISSN 1467-9876.

F.A. Eckel and C.F. Mass. Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, 20(3):328–350, 2005.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011.

Y. Gel, A.E. Raftery, and T. Gneiting. Calibrated probabilistic mesoscale weather field forecasting. *Journal of the American Statistical Association*, 99(467):575–583, 2004.

A.E. Gelfand, A.M. Schmidt, S. Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

T. Gneiting, A.E. Raftery, A.H. Westveld III, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

J.A. Goff and T.H. Jordan. Stochastic modeling of seafloor morphology: Inversion of sea beam data for second-order statistics. *Journal of Geophysical Research*, 93(B11):13589–13, 1988. ISSN 0148-0227.

M. Goulard and M. Voltz. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286, 1992. ISSN 0882-8121.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX On the Matern correlation family. *Biometrika*, 93(4):989, 2006. ISSN 0006-3444.

M.S. Handcock and M.L. Stein. A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410, 1993. ISSN 0040-1706.

L. Hartman and O. Hössjer. Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 52(5):2331–2349, 2008. ISSN 0167-9473.

Ø. Hjelle and M. Dæhlen. *Triangulations and applications*. Springer Verlag, 2006.

N.B. Ingleby. The statistical structure of forecast errors and its representation in the met. office global 3-d variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127(571):209–231, 2001.

P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 3rd edition, 1999. ISBN 3540540628.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

B. Matérn. *Spatial variation*. Springer-Verlag Berlin, 1986.

B.J. Reich and M. Fuentes. A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264, 2007.

P.J. Ribeiro Jr and P.J. Diggle. *MODEL BASED GEOSTATISTICS*. Springer Series in Statistics. Springer, 2006.

H. Rue. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338, 2001. ISSN 1467-9868.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue and H. Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49, 2002. ISSN 1467-9469.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868.

S.R. Sain and N. Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007.

A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108(D24):8783, 2003.

J.M.L. Sloughter, A.E. Raftery, T. Gneiting, and C. Fraley. Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Monthly Weather Review*, 135(9):3209–3220, 2007.

M.L. Stein. *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag, 1999. ISBN 0387986294.

H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer Verlag, 2003. ISBN 3540441425.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3-4):434–449, 1954. ISSN 0006-3444.

P. Whittle. Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, 40:974–994, 1963.

O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, and JZ Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-heinemann, 2005.

Paper II

Multivariate Gaussian Random Fields with Oscillating Covariance Functions using Systems of Stochastic Partial Differential Equations

Xiangping Hu, Finn Lindgren, Daniel Simpson, and Håvard Rue

Technical Report, 2013

Multivariate Gaussian Random Fields with Oscillating Covariance Functions using Systems of Stochastic Partial Differential Equations

Xiangping Hu*[1], Finn Lindgren[2], Daniel Simpson[1], and Håvard Rue[1]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[2]Department of Mathematical Sciences, University of Bath, BA2 7AY, United Kingdom

Abstract

In this paper we propose a new approach for constructing *multivariate* Gaussian random fields (GRFs) with oscillating covariance functions through systems of stochastic partial differential equations (SPDEs). We discuss how to build systems of SPDEs that introduces oscillation characteristics in the covariance functions of the multivariate GRFs. By choosing different parametrization of the equations, some GRFs can be made with oscillating covariance functions but other fields can have Matérn covariance functions or close to Matérn covariance functions. The multivariate GRFs constructed by solving the systems of SPDEs automatically fulfill the hard requirement of nonnegative definiteness for the covariance functions. The approximate weak solutions to the systems of SPDEs are used to represent the multivariate GRFs by multivariate Gaussian *Markov* random fields (GMRFs). Since the multivariate GMRFs have sparse precision matrices (inverse of the covariance matrices), numerical algorithms for sparse matrices can be applied to the precision matrices for sampling and inference. Thus from a computational point

*Corresponding author. Email: Xiangping.Hu@math.ntnu.no

of view, the *big-n* problem can be partially solved with these types of models. Another advantage of the method is that the oscillation in the covariance function can be controlled directly by the parameters in the system of SPDEs. We show how to use this proposed approach with simulated data and real data examples.

Keywords: Multivariate Gaussian random fields, Oscillating covariance functions, Multivariate Gaussian Markov random fields, Sparse matrix, Stochastic partial differential equations

1 Introduction

Statistics for spatial data appeared from hundreds of years ago, but spatial models for this type of data appeared much later (Cressie, 1993). The spatial models have been widely used to model the spatial data in many areas. Gaussian random fields (GRFs) are some of the most commonly used models in spatial statistics. Since the normalizing constant can be computed explicitly, the GRFs are convenient to be used in many applications, such as geo-statistical data, environmental and atmospheric data, longitudinal and survival data (Cressie, 1993; Stein, 1999; Rue and Held, 2005; Gelfand et al., 2010). GRFs also have other good properties, such as the fact that a GRF can be explicitly specified through a mean $\boldsymbol{\mu}(\cdot)$ and a covariance function $\boldsymbol{C}(\cdot,\cdot)$. Let the coordinates members of $\mathbb{R}^d$. If $(x(\boldsymbol{s}_1), x(\boldsymbol{s}_2), \ldots, x(\boldsymbol{s}_n))$ is Gaussian for every selection of points $(\boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \ldots, \boldsymbol{s}_n^T)$ for every $n \geq 1$, then we call $x(\boldsymbol{s})$ a continuously indexed GRF. The covariance matrix of a collection $(x(\boldsymbol{s}_1), x(\boldsymbol{s}_2), \ldots, x(\boldsymbol{s}_n))$ is given by $\boldsymbol{\Sigma} = [C(\boldsymbol{s}_i, \boldsymbol{s}_j)]$. However, there is a hard nonnegative definite requirement that must be fulfilled for the function $\boldsymbol{C}(\cdot,\cdot)$. This is one of the main concerns when we build models with GRFs using covariance function based approaches.

A widely used class of covariance functions is the Matérn family which was introduced by Matérn (Matérn, 1986). This family of covariance functions captures the most common form of the empirical behavior of stationary covariance functions, namely that the correlation between the locations $\boldsymbol{m}$ and $\boldsymbol{n}$ should decrease when the Euclidean distance $\|\boldsymbol{h}\| = \|\boldsymbol{m} - \boldsymbol{n}\|$ increases (Diggle and Ribeiro Jr, 2006). This family of covariance functions is isotropic and usually written as $\sigma^2 M(\boldsymbol{m}, \boldsymbol{n}|\nu, \kappa)$, where $M(\boldsymbol{m}, \boldsymbol{n}|\nu, \kappa)$ is the Matérn correlation function between the spatial lo-

cations $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{R}^d$. The Matérn correlation function is a two-parameter family with the form

$$M(\boldsymbol{h}|\nu,\kappa) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa\|\boldsymbol{h}\|)^{\nu}K_{\nu}(\kappa\|\boldsymbol{h}\|), \tag{1}$$

where σ^2 is the marginal variance, K_ν is the modified Bessel function of second kind and $\kappa > 0$ is the scaling parameter. The order ν is a smoothness parameter and must be positive. The smoothness parameter ν is of critical concern in spatial statistics since it defines the differentiability of the sample paths and the Hausdorff dimension (Goff and Jordan, 1988; Handcock and Stein, 1993; Gneiting et al., 2010). It is known that the smoothness parameter ν is poorly identifiable from the data and hence it is usually fixed (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011; Hu et al., 2012a). The Matérn family contains the commonly used models with exponential covariance function $M(\boldsymbol{h}|\frac{1}{2},\kappa) = \exp(-\kappa\|\boldsymbol{h}\|)$. The Matérn covariance function is the key factor in the explicit link between the GRFs and GMRFs through the stochastic partial differential equations (SPDEs) discussed by Lindgren et al. (2011). Gneiting et al. (2010) presented one direct approach for constructing the multivariate GRFs by using matrix-valued covariance functions, and all components in the matrix-valued covariance function are Matérn covariance functions. Hu et al. (2012a) discussed the important role of Matérn covariance function for constructing stationary and isotropic multivariate GRFs with systems of SPDEs.

In spatial statistics oscillating models usually deal with ocean waves, and we usually work entirely with their spectra only rarely with covariance functions (Lindgren, 2010). For time series there are plenty of applications with oscillating models in discrete time since the p-order auto-regressive (AR(p)) processes can result in oscillating models (Wei, 2006). However, it is less common for continuous time models. Lindgren et al. (2011) have discussed an approach for constructing univariate GRFs with oscillating covariance functions. In their approach they have chosen a coupled system of SPDEs to construct two independent random fields with the same precision matrix. Their discussion was focused on the case $\alpha = 2$. Since it is important for our approach for constructing multivariate GRFs with oscillating covariance functions, we give an overview of their approach in Section 3.1.

In this paper we focus on the methodology for constructing multivariate GRFs with oscillating covariance functions through systems of SPDEs. This work is an extension of the discussion given by Lindgren et al. (2011) for the univariate case. It is also an extension of the approach discussed by Hu et al. (2012a) to construct larger class of useful models. In our approach the GRFs are constructed by solving systems of stochastic partial differential equations. One of the main advantages of this approach is that we do not need to consider the notorious nonnegative definite requirement for the covariance functions. This requirement is fulfilled automatically because we are working on the processes directly. During the computational stage the GRFs are represented by the Gaussian *Markov* random fields (GMRFs) by GMRF approximations. A GMRF $x(\boldsymbol{s})$ is a discretely indexed GRF with some Markov property. The full conditionals $\{\pi(x_i|\boldsymbol{x}_{-i}); i = 1, \ldots, n\}$ of a GMRF depend only on a set of neighbors of each site i. Denote the neighbors of the node i by ∂i. The Markov property implies that $Q_{ij} \neq 0$ if and only if $j \in \partial i \cup i$, where $\boldsymbol{Q}$ denotes the precision matrix of the GMRF. Consistency requirement implies that if $i \in \partial j$, then $j \in \partial i$. The precision matrix $\boldsymbol{Q}$ for the GMRF is sparse which enables us to use numerical algorithms for sparse matrices. Thus the *big-n* problem (Banerjee et al., 2004) can be partially solved in our case with these types of models. We refer to Rue et al. (2009, Section 2.1) for a condensed overview of the theory of GMRFs. Detailed discussions about GMRFs are given in Rue and Held (2005, Chapter 2).

2 State-of-the-art and preliminaries

We review the state-of-the-art research on GRFs with SPDE approach in Section 2.1 and the methodologies for constructing multivariate GRFs in Section 2.2. The GMRF approximation for representing a GRF with a GMRF is introduced in Section 2.3 since it is crucial for computations.

2.1 GRFs through the SPDE approach

As mentioned in Section 1, Lindgren et al. (2011) proposed a novel approach for constructing GRFs by using the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{2}$$

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator, κ is the scaling parameter, α is related to the smoothness parameters $\nu > 0$ and $\alpha = \nu + d/2$. Δ is the Laplacian with definition

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

A range parameter ρ connects the scaling parameter κ and the smoothness parameter ν. The simple and empirically derived relationship $\rho = \sqrt{8\nu}/\kappa$ is commonly used (Lindgren et al., 2011; Hu et al., 2012a). It corresponds to correlation near 0.1 at distance ρ, with parameters κ and ν. $\mathcal{W}(\boldsymbol{s})$ is the innovation process which is a standard spatial Gaussian white noise. The most important result given by Whittle (1954, 1963), and used by Lindgren et al. (2011) extensively, is that the solution $x(\boldsymbol{s})$ to SPDE (2) is a GRF with the Matérn covariance function given in Equation (1). We follow the terminology used by Lindgren et al. (2011) and call GRFs with Matérn covariance functions Matérn random fields. Lindgren et al. (2011) commented that their approach can be extended in many directions.

Fuglstad (2011) has extended the SPDE approach to include a diffusion matrix in Equation (2) which provides a way of controlling the covariance structures of the GRF. Using the diffusion matrix $\boldsymbol{H}$ in SPDE (2), it is not only possible to construct homogeneous isotropic fields, but also anisotropic fields. Fuglstad (2011) showed that it is possible to construct inhomogeneous fields. The SPDE discussed by Fuglstad (2011) has the form

$$\kappa^2 x(\boldsymbol{s}) - \nabla \cdot \boldsymbol{H}(\boldsymbol{s}) \nabla x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{3}$$

where $\boldsymbol{H}$ is a 2×2 matrix-valued function, ∇ is the gradient operator and $\mathcal{W}(\boldsymbol{s})$ is a standard spatial Gaussian white noise process. The main contribution of Fuglstad (2011) is that he introduced the matrix $\boldsymbol{H}$ to (2) to control the structure of covariance matrix. However, he focused on discussing the univariate GRFs with $\alpha = 2$.

Bolin and Lindgren (2011) discussed how to use nested SPDEs to construct stationary and non-stationary GRFs. The equation chosen by Bolin and Lindgren (2011) has the form

$$\mathcal{L}_1 x(\boldsymbol{s}) = \mathcal{L}_2 \mathcal{W}(\boldsymbol{s}) \tag{4}$$

for some linear differential operators $\mathcal{L}_1$ and $\mathcal{L}_2$. $\mathcal{W}(\boldsymbol{s})$ is a standard spatial Gaussian white noise process. The SPDE (4) may not exist in the common sense since the operator $\mathcal{L}_2$ may contain some differentiation of the noise process $\mathcal{W}(\boldsymbol{s})$. In this case SPDE (4) can then be interpreted as the following nested system of SPDEs

$$\begin{aligned} \mathcal{L}_1 x_0(\boldsymbol{s}) &= \mathcal{W}(\boldsymbol{s}), \\ x(\boldsymbol{s}) &= \mathcal{L}_2 x_0(\boldsymbol{s}), \end{aligned} \tag{5}$$

when $\mathcal{L}_1$ and $\mathcal{L}_2$ are commutative operators. As pointed out by Bolin and Lindgren (2011) this interpretation not only avoids the apparent problem with the differentiation of the Gaussian white noise process, but also gives an interpretation of the consequence of the additional differential operator $\mathcal{L}_2$. The random field $x(\boldsymbol{s})$ is obtained by applying $\mathcal{L}_2$ to the solution of (4) with $\mathcal{L}_2 = \boldsymbol{I}$.

Fuglstad (2010) discussed the approximated solutions to a SPDE for constructing space-time GMRFs. The equation discussed by him has the form

$$\frac{\partial}{\partial t} x(s,t) - \nabla \cdot \nabla x(s,t) = \tau \mathcal{W}(s,t), \qquad (s,t) \in [0,L] \times [0,T], \tag{6}$$

where, $\nabla = \frac{\partial}{\partial x}$, $\tau > 0$ is a constant and $\mathcal{W}(s,t)$ is Gaussian space-time white noise. Fuglstad (2010) only discussed $s \in \mathbb{R}$ and argued that this equation has real physical meaning since the prototype of (6) is the stochastic heat equation

$$\frac{\partial}{\partial x} q(s,t) + \nabla F(s,t) = f(s,t). \tag{7}$$

The heat equation relates the change of the size q in time to the spatial divergences of the flux $F(s,t)$ and the source term $f(s,t)$. Fuglstad (2010) chose a finite volume method for solving Equation (6), and claimed that the finite volume method gave the correct distribution of the total energy of the solution to SPDE (6).

Lindgren et al. (2011) discussed the methodology for constructing non-stationary GRFs and non-separable space-time models. They claimed that if the parameters κ and τ depend on the coordinate $\boldsymbol{s}$, then we can construct a non-stationary GRF. The SPDE then has the form

$$(\kappa^2(\boldsymbol{s}) - \Delta)^{\alpha/2} \{\tau(\boldsymbol{s}) x(\boldsymbol{s})\} = \mathcal{W}(\boldsymbol{s}). \tag{8}$$

The non-separable space-time models have interaction between space and time in the covariance structure. In general it is difficult to construct this kind of model through a covariance function based approach. However, the SPDE approach can be used. One of the non-separable SPDEs which can result in this kind of model is

$$\left\{\frac{\partial}{\partial t} + (\kappa^2 + \boldsymbol{m}\cdot\nabla - \nabla\cdot\boldsymbol{H}\nabla)\right\} x(\boldsymbol{s},t) = \varepsilon(\boldsymbol{s},t), \tag{9}$$

where $\boldsymbol{m}$ is a transport vector, $\boldsymbol{H}$ is a positive definite diffusion matrix and $\varepsilon(\boldsymbol{s},t)$ is a stochastic space-time noise process. We refer to Lindgren et al. (2011, Section 3.5) for detailed discussion on this topic.

2.2 Multivariate GRFs

A multivariate GRF with p components $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \cdots, x_p(\boldsymbol{s}))^T$, $\boldsymbol{s}\in\mathbb{R}^d$, is a collection of continuously indexed multivariate normal random vectors such that

$$\boldsymbol{x}(\boldsymbol{s}) \sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ is the mean of the random field and $\boldsymbol{\Sigma}$ is the covariance matrix. Assume, at the current stage, that the process is second-order stationary with mean zero. One approach for constructing stationary and isotropic multivariate GRFs using covariance-based models was proposed by Gneiting et al. (2010). The covariance function $\boldsymbol{C}$ in their approach is given by

$$\boldsymbol{C}(\boldsymbol{h}) = \begin{pmatrix} C_{11}(\boldsymbol{h}) & C_{12}(\boldsymbol{h}) & \cdots & C_{1p}(\boldsymbol{h}) \\ C_{21}(\boldsymbol{h}) & C_{22}(\boldsymbol{h}) & \cdots & C_{2p}(\boldsymbol{h}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{h}) & C_{p2}(\boldsymbol{h}) & \cdots & C_{pp}(\boldsymbol{h}) \end{pmatrix}, \tag{10}$$

where $\{C_{ii}(\boldsymbol{h}) = \sigma_{ii}M(\boldsymbol{h}|\nu_{ii},\kappa_{ii}); i = 1,\ldots,p\}$ are the marginal covariance functions and $\{C_{ij}(\boldsymbol{h}) = \rho_{ij}\sigma_i\sigma_j M(\boldsymbol{h}|\nu_{ij},\kappa_{ij}); i,j = 1,\ldots,p, i\neq j\}$ are the cross-covariance functions. $\{C_{ii}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s}+\boldsymbol{h})x_i(\boldsymbol{s})); i = 1,2,\ldots,p\}$ give information about the covariance structures within the fields $\{x_i(\boldsymbol{s})\}$. $\{C_{ij}(\boldsymbol{h}) = \mathbb{E}(x_i(\boldsymbol{s}+\boldsymbol{h})x_j(\boldsymbol{s})); i,j = 1,2,\ldots,p, i\neq j\}$ describes the covariance structure between fields $\{x_i(\boldsymbol{s})\}$ and $\{x_j(\boldsymbol{s})\}$. $\{\rho_{ij}\}$ are the co-located

correlation coefficients. $\{\sigma_{ii} \geq 0\}$ are the marginal variances, and $\{\sigma_i\}$ and $\{\sigma_j\}$ are the corresponding standard deviations. They satisfy the relationships throughout this paper $\sigma_{ii} = \sigma_i^2, \sigma_{ij} = \sigma_i \sigma_j$. The main difficulty in constructing useful multivariate models using this kind of approach is the nonnegative definiteness requirement for the covariance functions. Gneiting et al. (2010) proposed a way to specify valid parametric models through the covariance functions given in Equation (10) directly. Several theorems were presented in order to ensure the matrix-valued covariance function to be symmetric and nonnegative definite.

Hu et al. (2012a) proposed to use a system of SPDEs to construct a multivariate GRF. They claimed that the notorious requirement nonnegative definiteness for the covariance function is automatically fulfilled with their approach. Hu et al. (2012a) also discussed the link between the system of SPDEs approach and the covariance function based approach discussed by Gneiting et al. (2010). The system of SPDEs which has been used for constructing the multivariate GRFs by Hu et al. (2012a) is

$$\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{11}$$

where $\{\mathcal{L}_{ij} = b_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2}; i = j = 1, 2, \dots, p\}$ are differential operators with $\{\alpha_{ij} = 0 \text{ or } 2\}$, $\{\kappa_{ij}\}$ and $\{\nu_{ij}\}$ are scaling parameters and smoothness parameters. $\{\varepsilon_i(\boldsymbol{s}); i = 1, 2, \dots, p\}$ are independent but not necessarily identically distributed noise processes, and $\{b_{ij}\}$ are the parameters related to the marginal variances of the fields and the cross-covariances between the fields. Hu et al. (2012a) pointed out that the GMRF approximation can be applied to the GRFs. Hence they can use computationally efficient GMRFs for sampling and inference. However, the constructed GRFs are always stationary and isotropic, and the covariance functions are not oscillating.

2.3 GMRFs approximation to GRFs

Generally speaking, GRFs are commonly used in statistical modelling because of their good theoretical properties. However, the GRFs have a

bottle-neck on the computational side. The computational cost for factorizing a dense covariance matrix $\boldsymbol{\Sigma}$ with dimension $n \times n$ is $\mathcal{O}(n^3)$. Even though the computational power is at an all time high, it seems that in many situations it is infeasible to do the computations in reasonable time. Banerjee et al. (2004, Appendix A.5) informally call this situation "the big n problem".

There are many different approaches trying to avoid or overcome "the big n problem", such as covariance tapering (Furrer et al., 2006; Zhang and Du, 2008; Kaufman et al., 2008; Shaby and Ruppert, 2012), likelihood approximations (Vecchia, 1988; Stein et al., 2004), and fixed rank kriging and fixed rank filtering (Cressie and Johannesson, 2008; Cressie et al., 2010) . The approach which has been chosen in this paper is based on the GMRF approximation to GRFs. The sparsity of the precision matrix $\boldsymbol{Q}$ enables the numerical algorithms for sparse matrix for fast inference with large datasets (Rue, 2001; Rue and Held, 2005; Lindgren et al., 2011). The general cost for factorizing the sparse matrix $\boldsymbol{Q}$ is $\mathcal{O}(n)$, $\mathcal{O}(n^{3/2})$ and $\mathcal{O}(n^2)$ in one dimension, two dimensions and three dimensions, respectively (Rue and Held, 2005). Hartman and Hössjer (2008) proposed to use the GMRFs for GRFs for spatial prediction with Kriging, due to the pleasant computational properties of GMRFs.

In this paper we only give an overview of GMRF approximation to univariate GRF and refer to Hu et al. (2012a, Section 2.3) for detailed discussion on GMRF approximation to multivariate GRF. In order to find a GMRF approximation of a GRF on a triangulated lattice, we at first need to find the stochastic weak formulation of SPDE (2) (Kloeden and Platen, 1999). In this paper we use Delaunay triangulation. We refer to Hjelle and Dæhlen (2006) for more information about triangulations. Denote the inner product of functions h and g as

$$\langle h, g \rangle = \int h(\boldsymbol{s}) g(\boldsymbol{s}) d(\boldsymbol{s}), \tag{12}$$

where the integration is within the region of interest. The stochastic weak solution of SPDE (2) is found by requiring

$$\left\{ \langle \phi_i, (\kappa^2 - \Delta)^{\alpha/2} x \rangle; i = 1, \ldots, M \right\} \stackrel{d}{=} \left\{ \langle \phi_i, \mathcal{W} \rangle; i = 1, \ldots, M \right\}, \tag{13}$$

where M is the number of test functions $\{\phi_i(\boldsymbol{s})\}$ and "$\stackrel{d}{=}$" denotes equality

in distribution.

Then we need to find the finite element representation of the solution to the SPDE. The finite element representation of the solution is

$$x(\boldsymbol{s}) = \sum_{i=1}^{N} \psi_i(\boldsymbol{s})\omega_i \tag{14}$$

with basis functions $\{\psi_i(\boldsymbol{s}); i = 1, 2, \ldots, N\}$ and Gaussian distributed weights $\{\omega_i; i = 1, 2, \ldots, N\}$. N is the number of vertexes in the triangulation. We refer to Zienkiewicz et al. (2005) and Brenner and Scott (2008) for more information and theoretical background of finite element methods. The approach given by Lindgren et al. (2011) for choosing the basis functions is used in this paper. With $M = N$ we choose each basis function $\psi_i(\boldsymbol{s})$ to be piecewise linear on each triangle with $\psi_i(\boldsymbol{s}) = 1$ at vertex i and $\psi_i(\boldsymbol{s}) = 0$ at other vertexes. This choice of basis functions means that the local interpolation on a triangle is linear. Lindgren et al. (2011) pointed out that other methods, such as kernel method, are useful in theory but not necessary in practice. When $\alpha_{ij} = 1$ the *least squares* approximation is chosen, $\phi_k = (\kappa^2 - \Delta)^{\frac{1}{2}}\psi_k$. When $\alpha_{ij} = 2$ the *Galerkin* solution is chosen, $\phi_k = \psi_k$. When $\alpha_{ij} \geq 3$ the *recursive Galerkin* formulation is used. We refer to Lindgren et al. (2011, Section 2.3) for more information about the recursive Galerkin formation.

2.4 Outline of the paper

The structure of the rest of the paper is organized as follows. Section 3 gives the detailed discussion on how to construct multivariate GRFs with oscillating covariance functions through the systems of SPDEs approach. Examples with simulated data and real data are given in Section 4. Discussion and future work in Section 5 ends this paper.

3 Model formulation

GRFs with oscillating covariance functions can be used in many situations, for example, for modelling global pressure (Lindgren et al., 2011) and ocean waves (Lindgren, 2010). First, an overview for constructing

the *univariate* GRFs with oscillating covariance function is given since it is needed for constructing *multivariate* GRFs with oscillating covariance functions. Next, we introduce a general approach for constructing the multivariate GRFs with oscillating covariance functions. Then explicit approach for constructing the *bivariate* GRFs is discussed. At last we discuss the procedure for sampling the multivariate GRFs with oscillating covariance functions.

3.1 Univariate GRFs with oscillating covariance functions

Lindgren et al. (2011, Section 3.3) discussed how to construct an univariate GRF with oscillating covariance function using a SPDE with complex number. For the case $\alpha = 2$ the SPDE has the form

$$\left\{\kappa^2 \exp(i\pi\omega) - \Delta\right\} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{15}$$

where $\omega \in [0, 1]$ is the oscillation parameter, $x(\boldsymbol{s}) = x_1(\boldsymbol{s}) + ix_2(\boldsymbol{s})$, and $\mathcal{W}(\boldsymbol{s}) = \mathcal{W}_1(\boldsymbol{s}) + i\mathcal{W}_2(\boldsymbol{s})$. The innovation processes $\mathcal{W}_1(\boldsymbol{s})$ and $\mathcal{W}_2(\boldsymbol{s})$ are independent standard Gaussian white noise processes. Lindgren et al. (2011) pointed out that the real and imaginary parts, $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$, of the stationary solution $x(\boldsymbol{s})$ are independent with identical spectrum densities

$$R(\boldsymbol{k}) = \frac{1}{(2\pi)^d \left(\kappa^4 + 2\cos(\pi\omega)\kappa^2\|\boldsymbol{k}\|^2 + \|\boldsymbol{k}\|^4\right)}, \quad \boldsymbol{k} \in \mathbb{R}^d. \tag{16}$$

With this approach the common isotropic stationary Matérn random fields can be obtained by setting $\omega = 0$. We can notice that $\omega = 1$ generates intrinsic stationary random fields. We refer to Rue and Held (2005, Chapter 3) for more information on the intrinsic random fields. When $\omega \in (0, 1)$, the constructed GRFs have covariance functions with oscillation. The oscillation is increasing with larger value of ω. The closed form of the precision matrix $\boldsymbol{Q}$ for the stationary GRFs with oscillation can be obtained from (16),

$$\boldsymbol{Q}(\kappa^2, \omega) = \kappa^4 \boldsymbol{C} + 2\cos(\pi\omega)\kappa^2 \boldsymbol{G} + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}. \tag{17}$$

The matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ in Equation (17) are defined through

$$\begin{aligned} C_{ij} &= \langle \psi_i, 1 \rangle, \\ G_{ij} &= \langle \nabla\psi_i, \nabla\psi_j \rangle, \end{aligned} \tag{18}$$

with basis functions $\{\psi_i; i = 1, 2, \ldots, n\}$. We use $C_{ij} = \langle \psi_i, 1 \rangle$, instead of $C_{ij} = \langle \psi_i, \psi_j \rangle$, in order to make the precision matrix sparse. This setting yields a Markov approximation to the FEM solution. Bolin and Lindgren (2009) studied the effects of the Markov approximation and claimed that the difference between the Markov approximation and exact FEM representation is negligible.

Lindgren et al. (2011) pointed out that this complex-valued version of SPDE (15) can be rewritten as a special case of the coupled systems of SPDEs

$$\begin{pmatrix} h_1 - \Delta & -h_2 \\ h_2 & h_1 - \Delta \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \mathcal{W}_1(\boldsymbol{s}) \\ \mathcal{W}_2(\boldsymbol{s}) \end{pmatrix}, \tag{19}$$

where $h_1 = \kappa^2 \cos(\pi\omega)$ and $h_2 = \kappa^2 \sin(\pi\omega)$. The random fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ from Equation (19) have the same precision matrix $\boldsymbol{Q}$ given in Equation (17). Lindgren et al. (2011) commented that it is surprising that these two fields from the coupled system of SPDEs (19) are always independent regardless of the choices of parameters. Additionally, the univariate GRFs with oscillating covariance functions from Equation (15) are always isotropic. However, it is possible to construct non-isotropic GRFs by slightly modifying the coupled system of SPDEs (19). We are not going to discuss this issue here and we focus only on the isotropic GRFs. We refer to Lindgren et al. (2011, Appendix C.4) for more information about the oscillating and non-isotropic cases.

3.2 Multivariate GRFs with oscillating covariance functions

The multivariate GRFs with oscillating covariance functions, in this paper, all have the assumption that the mean is zero, i.e., $\boldsymbol{x}(\boldsymbol{s}) \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma})$. Hu et al. (2012a) proposed to construct the multivariate GRFs using the system of SPDEs given in (11). In their approach the multivariate GRFs are always isotropic and stationary. The covariance functions cannot be oscillating. They argued that, under some conditions, it is possible to construct the multivariate GRFs with Matérn covariance functions as discussed by Gneiting et al. (2010). In this section we are going to discuss how to construct multivariate GRFs where some components of the random fields have oscillating covariance functions. The main idea is to re-

place the noise processes by noise processes with oscillating covariance functions. With this approach the systems of SPDEs have the same form as given in (11), but the noise processes are different. Even though the system of SPDEs in (11) is theoretically general, we recommend to use the triangular system of SPDEs in applications. The triangular system of SPDEs is

$$\begin{pmatrix} \mathcal{L}_{11} & & & \\ \mathcal{L}_{21} & \mathcal{L}_{22} & & \\ \vdots & \vdots & \ddots & \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{20}$$

where $\{\mathcal{L}_{ij}; i, j = 1, 2, \dots, p, i \geq j\}$ are differential operators as defined in (11), $\{\mathcal{L}_{ij} = 0; i, j = 1, 2, \dots, p, i < j\}$ and $\{\varepsilon_i(\boldsymbol{s}); i = 1, 2, \dots, p\}$ are noise processes where some of them have oscillating covariance functions. We recommend to use as fewer noise processes with oscillating covariance functions as possible. This system has many advantages, such as interpretation of the properties of the fields. For example, we know which components of the random field must have non-oscillating covariance functions and have oscillating covariance functions. However, there are some components of the random field which might have oscillating covariance functions. We divide the random fields into three categories $\boldsymbol{x}_{\mathrm{m}}$, $\boldsymbol{x}_{\mathrm{o}}$ and $\boldsymbol{x}_{\mathrm{p}}$, where $\boldsymbol{x}_{\mathrm{m}}$ denotes the random fields with non-oscillating covariance functions, $\boldsymbol{x}_{\mathrm{o}}$ denotes the random fields with oscillating covariance functions and $\boldsymbol{x}_{\mathrm{p}}$ denotes the random fields with covariance functions which might be oscillating. Assume that only $\{\varepsilon_i(\boldsymbol{s}); i = 1, 2, \dots, p\}$ is the noise process with oscillating covariance function and other noise processes $\{\varepsilon_j(\boldsymbol{s}); j = 1, 2, \dots, p, j \neq i\}$ are noise processes with non-oscillating covariance functions, and then we can obtain the following results.

- If only the covariance function for the noise process $\varepsilon_i(\boldsymbol{s})$ is oscillating, the covariances functions for all the random fields $x_j(\boldsymbol{s})(j < i)$ are non-oscillating, $x_j(\boldsymbol{s}) \in \boldsymbol{x}_{\mathrm{n}}(\boldsymbol{s})$;
- If only the covariance function for the noise process $\varepsilon_i(\boldsymbol{s})$ is oscillating, the random field $x_j(\boldsymbol{s})(j = i)$ has an oscillating covariance function, $x_j(\boldsymbol{s}) \in \boldsymbol{x}_{\mathrm{o}}(\boldsymbol{s})$

- If only the covariance function for the noise process $\varepsilon_i(\boldsymbol{s})$ is oscillating, the random fields $x_j(\boldsymbol{s})(j > i)$ belong to $\boldsymbol{x}_{\mathrm{p}}(\boldsymbol{s})$, which means that the covariance functions for these random fields might be oscillating.

This result is rather intuitive since we can obtain it by checking the system of SPDEs (20) directly. However, these results are important in the real-world application since it gives information for how to build models in a reasonable way. For instance, we can get information about how to choose the order of the random fields.

3.3 Bivariate GRFs with oscillating covariance functions

The methodology for constructing non-oscillating and isotropic bivariate GRFs explicitly has been studied by Gneiting et al. (2010) and Hu et al. (2012a). In this section we discuss the approach for constructing bivariate GRFs explicitly with oscillating covariance functions using systems of SPDEs. We start the investigation with random fields constructed by the full system of SPDEs

$$\begin{aligned} b_{11}(\kappa_{11}^2 - \Delta)^{\alpha_{11}/2} x_1(\boldsymbol{s}) + b_{12}(\kappa_{12}^2 - \Delta)^{\alpha_{12}/2} x_2(\boldsymbol{s}) &= \varepsilon_1(\boldsymbol{s}), \\ b_{22}(\kappa_{22}^2 - \Delta)^{\alpha_{22}/2} x_2(\boldsymbol{s}) + b_{21}(\kappa_{21}^2 - \Delta)^{\alpha_{21}/2} x_1(\boldsymbol{s}) &= \varepsilon_2(\boldsymbol{s}), \end{aligned} \tag{21}$$

where b_{ij} and $\{\mathcal{L}_{ij}(\boldsymbol{s}); i, j = 1, 2\}$ are the same as in (11), and $\{\varepsilon_1(\boldsymbol{s}); i = 1, 2\}$ are noise processes which can have oscillating covariance functions. By changing the properties of the noise processes we can construct more interesting random fields. Use the matrix notion and define the operator matrix as

$$\mathscr{L}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{L}_{11}(\theta_{11}) & \mathcal{L}_{12}(\theta_{12}) \\ \mathcal{L}_{21}(\theta_{21}) & \mathcal{L}_{22}(\theta_{22}) \end{pmatrix}, \tag{22}$$

and let $\boldsymbol{\varepsilon}(\boldsymbol{s}) = (\varepsilon_1(\boldsymbol{s}), \varepsilon_2(\boldsymbol{s}))^{\mathrm{H}}$, where H denotes the Hermitian transpose of a vector or a matrix. $\theta_{ij} = \{\alpha_{ij}, \kappa_{ij}, b_{ij}\}$ is defined as the collection of parameters for $\mathcal{L}_{ij}$. The system of equations (21) can then be written in a compact matrix form as

$$\mathscr{L}(\boldsymbol{\theta})\boldsymbol{x}(\boldsymbol{s}) = \boldsymbol{\varepsilon}(\boldsymbol{s}), \tag{23}$$

where $\boldsymbol{\theta} = \{\theta_{ij}, i, j = 1, 2\}$. With (23) we can obtain the power spectrum $\boldsymbol{S_x} = \mathbb{E}(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}^{\mathrm{H}})$ by

$$\boldsymbol{S_x} = \mathscr{H}^{\mathrm{H}} \boldsymbol{S_\varepsilon} \mathscr{H}^{-\mathrm{H}}, \tag{24}$$

where $-\mathrm{H}$ denotes the inverse of the complex conjugate of the matrix. $\hat{x}_{ij}$ is the Fourier transform of x_{ij}, $\hat{x}_{ij} = \mathscr{F}(x_{ij})$, and $\mathscr{H}$ is the Fourier transform of the operator matrix $\mathscr{L}$,

$$\mathscr{H}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{H}_{11}(\theta_{11}) & \mathcal{H}_{12}(\theta_{12}) \\ \mathcal{H}_{21}(\theta_{21}) & \mathcal{H}_{22}(\theta_{22}) \end{pmatrix}. \tag{25}$$

$\boldsymbol{S_\varepsilon}(\boldsymbol{k}) = \mathbb{E}\left(\hat{\varepsilon}\hat{\varepsilon}^{\mathrm{H}}\right)$ is the power spectrum matrix for the independent noise processes

$$\boldsymbol{S_\varepsilon}(\boldsymbol{k}) = \begin{pmatrix} S_{\varepsilon_1}(\boldsymbol{k}) & 0 \\ 0 & S_{\varepsilon_2}(\boldsymbol{k}) \end{pmatrix}, \tag{26}$$

where $\boldsymbol{k}$ is the frequency. Since the noise processes are mutually independent, the power spectrum matrix of noise processes is a diagonal matrix. Using Equation (24) - Equation (25), the elements in the power spectrum matrix of the bivariate fields from the full system of SPDEs in (21) can be obtained,

$$\begin{aligned}
S_{x_{11}}(\boldsymbol{k}) &= \frac{S_{\varepsilon_1}|\mathcal{H}_{22}^2| + S_{\varepsilon_2}|\mathcal{H}_{12}^2|}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|}, \\
S_{x_{12}}(\boldsymbol{k}) &= -\frac{\mathcal{H}_{22}S_{\varepsilon_1}|\mathcal{H}_{21}^2|\mathcal{H}_{11} + \mathcal{H}_{12}S_{\varepsilon_2}|\mathcal{H}_{11}^2|\mathcal{H}_{21}}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|\mathcal{H}_{21}\mathcal{H}_{11}}, \\
S_{x_{21}}(\boldsymbol{k}) &= -\frac{\mathcal{H}_{21}S_{\varepsilon_1}|\mathcal{H}_{22}^2|\mathcal{H}_{12} + \mathcal{H}_{11}S_{\varepsilon_2}|\mathcal{H}_{12}^2|\mathcal{H}_{22}}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|\mathcal{H}_{22}\mathcal{H}_{12}}, \\
S_{x_{22}}(\boldsymbol{k}) &= \frac{S_{\varepsilon_1}|\mathcal{H}_{21}^2| + S_{\varepsilon_2}|\mathcal{H}_{11}^2|}{|(\mathcal{H}_{11}\mathcal{H}_{22} - \mathcal{H}_{12}\mathcal{H}_{21})^2|}.
\end{aligned} \tag{27}$$

Define poles and zeros as the roots of the denominators and numerators of the power spectrum elements $\{S_{x_{ij}}; i, j = 1, 2\}$, respectively. From (27), we can see that the poles of the power spectrum, in general, are the same for both the fields, but zeros of the power spectrum will be different. It

gives us a possibility to construct bivariate GRFs with oscillating covariance functions by carefully re-parametrization of system of SPDEs (21). However, we will not discuss this approach in this paper, but leave it for future research. The approach we have chosen here is to change the noise process at the right hand of system of SPDEs (21).

Theoretically, we could choose the full version of the system of SPDEs given in (21) and give more flexibility for constructing bivariate random fields. However, we choose to simplify the model. Hu et al. (2012a) used the triangular version of the SPDEs system to construct bivaraite GRFs and this suggestion is followed in this paper. In the following sections, we focus on a special form of the triangular system of the SPDEs discussed in Section 3.2. In the special form the operator matrix is

$$\mathscr{L}_1(\boldsymbol{\theta}) = \begin{pmatrix} b_{11}\left(h_{11} - \Delta\right) & 0 \\ b_{21} & b_{22}\left(h_{22} - \Delta\right) \end{pmatrix}, \tag{28}$$

where the subscript "1" in $\mathscr{L}_1$ is used to denote the first operator matrix we use for constructions. Some other operator matrices are discussed in Appendix A. We can rewrite the system of SPDEs with matrix notation as

$$\mathscr{L}_1(\boldsymbol{\theta})\boldsymbol{x}(\boldsymbol{s}) = \boldsymbol{\varepsilon}(\boldsymbol{s}), \tag{29}$$

and Equation (29) can be written down explicitly as

$$\begin{aligned} b_{11}(h_{11} - \Delta)x_1(\boldsymbol{s}) &= \varepsilon_1(\boldsymbol{s}), \\ b_{21}x_1(\boldsymbol{s}) + b_{22}(h_{22} - \Delta)x_2(\boldsymbol{s}) &= \varepsilon_2(\boldsymbol{s}). \end{aligned} \tag{30}$$

In this form both random fields can have oscillating covariance function. The following discussion are based on the system of SPDEs (30). Let $\varepsilon_1(\boldsymbol{s})$ be a noise process with non-oscillating covariance function, such as a white noise process or noise process with Matérn covariance function, and $\varepsilon_2(\boldsymbol{s})$ be a noise process with oscillating covariance function generated from the complex-valued SPDEs (15). We can then conclude that the first field $x_1(\boldsymbol{s})$ is a stationary and isotropic random field with non-oscillating covariance function, and that $x_2(\boldsymbol{s})$ is a random field with oscillating covariance function. On the other hand, if $\varepsilon_1(\boldsymbol{s})$ has an oscillating covariance

function and $\varepsilon_2(\boldsymbol{s})$ has a non-oscillating covariance function, then the covariance functions for both random fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are oscillating given that $b_{21} \neq 0$.

We use the power spectra $\{S_{x_{ii}}(\boldsymbol{k}); i = 1, 2\}$ of the random fields together with the cross spectrum $S_{x_{21}}(\boldsymbol{k})$ to investigate the properties of the random fields,

$$\begin{aligned} S_{x_{11}}(\boldsymbol{k}) &= \frac{S_{\varepsilon_1}}{b_{11}^2(h_{11} + \|\boldsymbol{k}\|^2)^2}, \\ S_{x_{21}}(\boldsymbol{k}) &= -\frac{b_{21}S_{\varepsilon_1}}{b_{22}(h_{22} + \|\boldsymbol{k}\|^2)b_{11}^2(h_{11} + \|\boldsymbol{k}\|^2)^2}, \\ S_{x_{22}}(\boldsymbol{k}) &= \frac{b_{21}^2 S_{\varepsilon_1} + b_{11}^2(h_{11} + \|\boldsymbol{k}\|^2)^2 S_{\varepsilon_2}}{b_{11}^2(h_{11} + \|\boldsymbol{k}\|^2)^2 b_{22}^2(h_{22} + \|\boldsymbol{k}\|^2)^2}. \end{aligned} \tag{31}$$

The following results can be obtained from (31).

- The marginal variance of the first random field $x_1(\boldsymbol{s})$ is related only to the parameters b_{11} and h_{11}.
- The marginal variance of the second random field $x_2(\boldsymbol{s})$ is related only to $\{b_{ij}, h_{ij}; i, j = 1, 2\}$.
- The sign of b_{11} is irrelevant to the sign of the cross-correlation between the two fields. Since the marginal variance of the first random field $x_1(\boldsymbol{s})$ is only related to b_{11} and h_{11}, and there is a requirement $h_{11} > 0$, we can set $b_{11} > 0$;
- The sign of the correlation between the two fields only depends on the sign of the product of b_{22} and b_{21}. We recommend to set $b_{22} > 0$, and then the sign of the correlation between the fields will be related only to the sign of b_{21}. If $b_{21} > 0$, the two random fields are negatively correlated. If $b_{21} < 0$, the random fields are positively correlated.

3.4 Sampling the bivariate GRFs

The common approach for sampling GRFs uses the covariance matrix $\boldsymbol{\Sigma}$ or precision matrix $\boldsymbol{Q}$. Since the bivariate GRFs from the systems of SPDEs

are represented by the GMRFs, the precision matrices $\boldsymbol{Q}$ are sparse. Therefore, the direct approach for sampling a (multivariate) GMRF is usually through the Cholesky triangle $\boldsymbol{L}$, where $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^T$. The commonly used procedure for getting a sample from the GMRF $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is through the following steps

I. Use the Cholesky factorization to find the Cholesky triangle $\boldsymbol{L}$ of the precision matrix $\boldsymbol{Q}$. We usually do the Cholesky factorization with standard libraries.

II. Get a sample $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. $\boldsymbol{I}$ is an identity matrix and has the same dimensions as the precision matrix $\boldsymbol{Q}$.

III. Solve a linear system of equations with Cholesky triangle $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{z}$. Thus $\boldsymbol{v}$ has the correct covariance matrix $\boldsymbol{Q}^{-1}$ since $\text{Cov}(\boldsymbol{v}) = \text{Cov}(\boldsymbol{L}^{-T}\boldsymbol{z}) = (\boldsymbol{L}\boldsymbol{L}^T)^{-1} = \boldsymbol{Q}^{-1}$.

IV. Correct the mean by $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{v}$, and then $\boldsymbol{x}$ has the correct mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{Q}^{-1}$, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$.

If the precision matrix $\boldsymbol{Q}$ is a band matrix, the Cholesky triangle $\boldsymbol{L}$ will be also a band matrix. The corresponding algorithm for finding the Cholesky triangle when $\boldsymbol{Q}$ is a band matrix can be found in Rue and Held (2005, Algorithm 2.9). We also refer to Rue and Held (2005, Chapter 2) for detailed discussion about different sampling algorithms for GMRFs with different kinds of parametrization. Hu et al. (2012b) showed that it is possible to find a sparser triangular matrix $\tilde{\boldsymbol{L}}$ with incomplete orthogonal factorization for sampling the GMRF, but they pointed out that it needs longer computation time for finding the sparse triangular matrix.

In the following two examples, we choose all values of parameters to be equal. However, we set $\varepsilon_1(\boldsymbol{s})$ to be a noise process with a non-oscillating covariance function and $\varepsilon_2(\boldsymbol{s})$ to be a noise process with an oscillating covariance function in the first example. In the second example, we simply switch the order of the noise processes, i.e., we set $\varepsilon_1(\boldsymbol{s})$ to be a noise process with an oscillating covariance function and $\varepsilon_2(\boldsymbol{s})$ to be a noise process with a non-oscillating covariance function.

One sample from the GMRF in the first example is shown in Fig. 1 and one sample in the second example is shown in Fig. 4. The corresponding

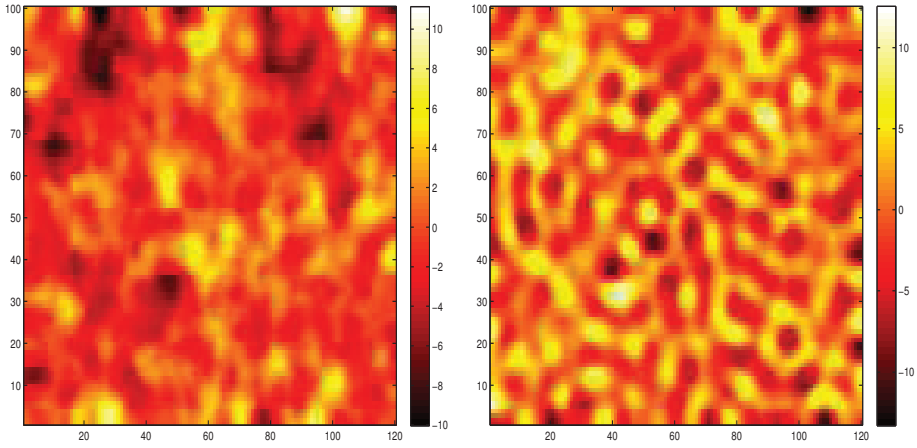
Table 1: Parameters for sampling the bivariate GRFs

Parameters		
$\boldsymbol{\alpha}$	$\boldsymbol{\kappa}$	$\boldsymbol{b}$
$\alpha_{11} = 2$	$h_{11} = 0.25$	$b_{11} = 0.5$
$\alpha_{12} = 0$	$h_{22} = 0.36$	$b_{12} = 0$
$\alpha_{21} = 2$	$\kappa_{n_1} = 0.5$	$b_{21} = 0.25$
$\alpha_{22} = 2$	$\kappa_{n_2} = 0.6$	$b_{22} = 1$
$\alpha_{n_1} = 2$	$\omega = 0.95$	
$\alpha_{n_2} = 2$		

correlation matrices are shown in Fig. 2 and Fig. 5. In these figures, the red lines indicate that the correlation is 0. In the first example the random field $x_1(\boldsymbol{s})$ has a non-oscillating covariance function and the second random field $x_2(\boldsymbol{s})$ has an oscillating covariance function. In the second example both the fields have oscillating covariance functions. These two examples verify the conclusion in Section 3.3.

4 Inference with simulated data and real data

In this section we illustrate how to use our approach with some simulated data examples and one real data example. In the first example, the covariance function of the first random field is non-oscillating, but the second random field has an oscillating covariance function. In the second example, both the random fields have oscillating covariance functions. The third and the forth examples show that if the two fields are independent, the inferences give indications about this, no matter which field is oscillating. One real data example in the end shows that our approach can be applied in practical applications. As pointed out by Diggle et al. (1998, Chapter 5) and Lindgren et al. (2011, Section 2), the smoothness parameter $\{\nu_{ij}; i, j = 1, 2\}$ are poorly identifiable. Therefore, we fix the values of $\{\alpha_{ij}\}$ in the simulated data examples and in the real data example.



(a) (b)

Figure 1: A realization of the bivariate random field with parameters given in Table 1. In this example the first noise process has a non-oscillating covariance function and the second noise process has an oscillating covariance function.

4.1 Posterior for the hyper-parameters

The first step of the inference is usually to derive the (log-) posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of $\boldsymbol{\theta}$. The well known Bayesian formula (32) is at the core of Bayesian inference,

$$\begin{aligned} \pi(\boldsymbol{y}, \boldsymbol{\theta}) &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \\ &= \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})}. \end{aligned} \tag{32}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of the hyper-parameters, and we return to this topic in Section 4.2, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ is the density for the bivariate random fields, $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is the density for the observations given the random field and the parameters and $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is the full conditional of the random fields given the observations and parameters.

Assume that there are N triangles in the domain for each of the random

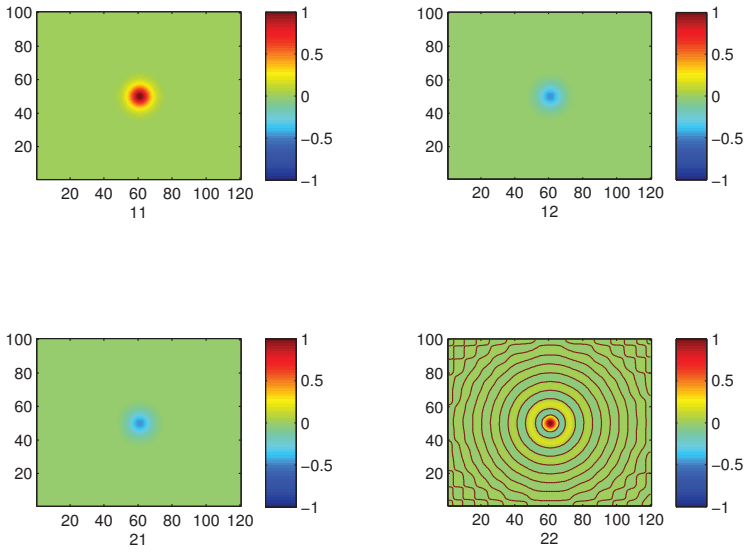


Figure 2: Correlations and cross-correlation functions for the bivariate random field with parameters given in Table 1. In this example the first noise process has a non-oscillating covariance function and the second noise process has an oscillating covariance function. We see that only the second random field has an oscillating covariance function.

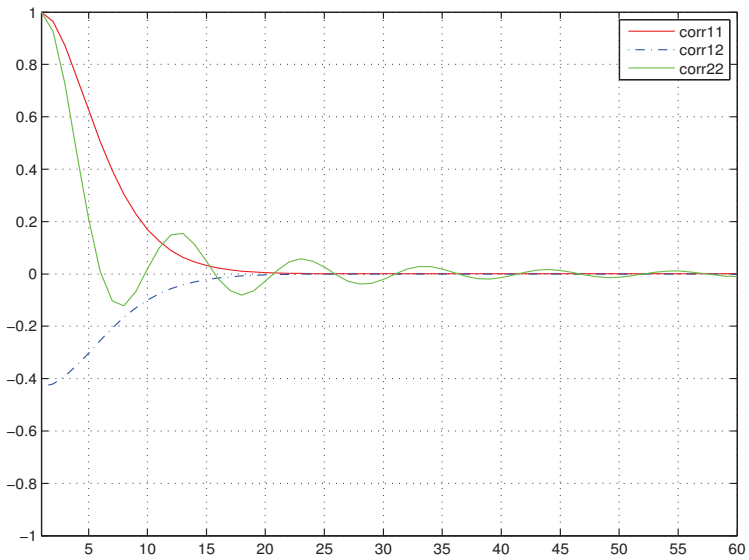
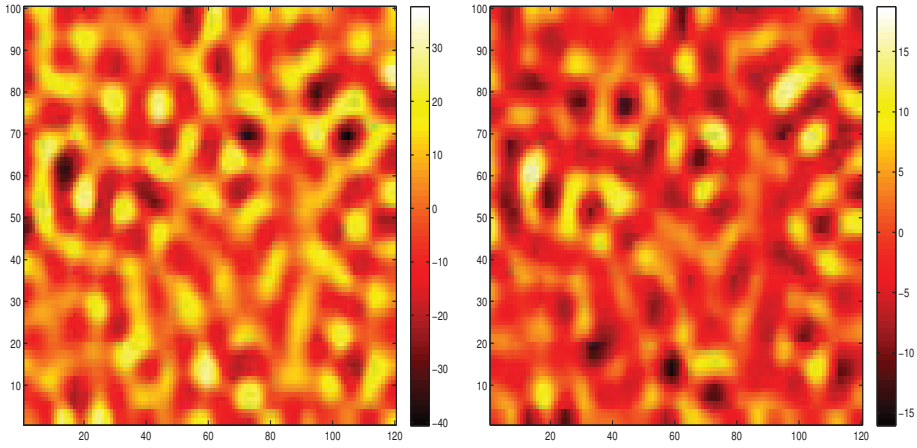


Figure 3: Correlations and cross-correlation functions for the bivariate random field with parameters given in Table 1. In this example the first noise process has a non-oscillating covariance function and the second noise process has an oscillating covariance function. We see that only the second random field has an oscillating covariance function.



(a) (b)

Figure 4: A realization of the bivariate random field with the parameters given in Table 1. In this example the first noise process has an oscillating covariance function and the second noise process has a non-oscillating covariance function.

field $\{x_i(\boldsymbol{\theta}); i = 1, 2\}$. With the bivariate random fields $\boldsymbol{x} = (x_1, x_2)^T$, $2N$ triangles are used and hence the probability density of the bivariate random field has the form

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{2N} |\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x}\right), \tag{33}$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix for the bivariate field. We assume that the length of the data is $t = k_1 + k_2$, where $\{k_i; i = 1, 2\}$ are the length of the observations for each field. Then $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ has the form

$$\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{t} |\boldsymbol{Q}_n|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T\boldsymbol{Q}_n(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right), \tag{34}$$

where $\boldsymbol{Q}_n$ is the precision matrix for the measurement errors with dimension $t \times 2k$, and $\boldsymbol{A}$ is a matrix with dimension $t \times 2N$ that links the sparse observations to the dense random fields. One thing we want to point out is

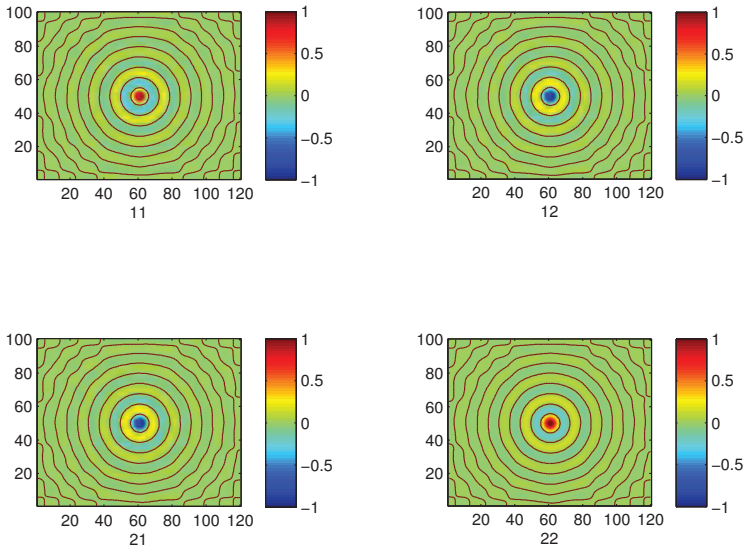


Figure 5: Correlations and cross-correlation functions for the bivariate random field with parameters given in Table 1. In this example the first noise process has an oscillating covariance function and the second noise process has a non-oscillating covariance function. We see that both the random fields have oscillating covariance functions.

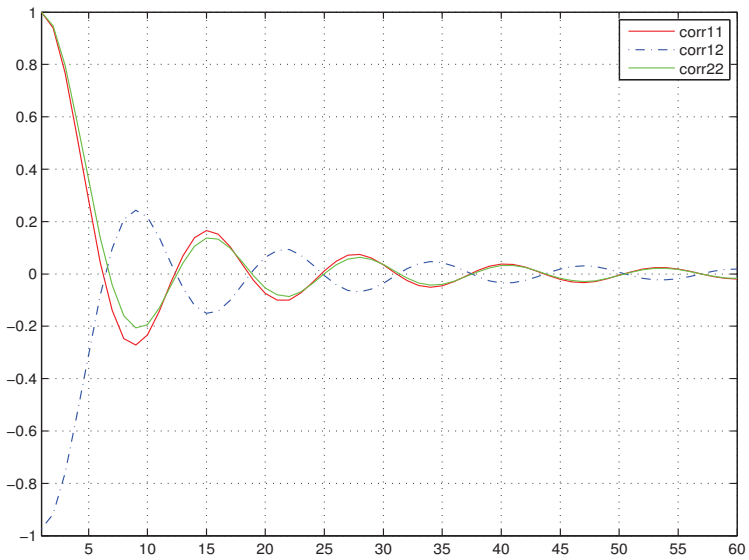


Figure 6: Correlations and cross-correlation functions for the bivariate random field with parameters given in Table 1. In this example the first noise process has an oscillating covariance function and the second noise process has a non-oscillating covariance function. We see that both the random fields have oscillating covariance functions.

that the length of the observations for each field can be different and they are not necessarily observed at the same locations. We used the notation $\pi(\boldsymbol{y}|\boldsymbol{x})$ instead of $\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ since this function is independent of $\boldsymbol{\theta}$. The full conditional $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ can be obtained,

$$\begin{aligned}\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \pi(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2}\left[x^T(\boldsymbol{Q}(\boldsymbol{\theta})+\boldsymbol{A}^T\boldsymbol{Q}_n\boldsymbol{A})\boldsymbol{x}-2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Q}_n\boldsymbol{y}\right]\right).\end{aligned} \tag{35}$$

Denote $\boldsymbol{\mu}_c(\boldsymbol{\theta}) = \boldsymbol{Q}_c^{-1}(\boldsymbol{\theta})\boldsymbol{A}^T\boldsymbol{Q}_n\boldsymbol{y}$, and $\boldsymbol{Q}_c(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{A}^T\boldsymbol{Q}_n\boldsymbol{A}$. Then we have

$$\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_c(\boldsymbol{\theta}), \boldsymbol{Q}_c^{-1}(\boldsymbol{\theta})\right). \tag{36}$$

Thus $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}$ is a $2N$-dimensional multivariate Gaussian distribution. We can write (36) in the canonical form $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}_c\left(\boldsymbol{A}^T\boldsymbol{Q}_n\boldsymbol{y}, \boldsymbol{Q}_c(\boldsymbol{\theta})\right)$. For more information about canonical form of the GMRFs, we refer to, for example, Rue and Held (2005, Chapter 2).

With a given prior $\pi(\boldsymbol{\theta})$, together with (33) to (36), the posterior distribution of $\boldsymbol{\theta}$ becomes

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})&\frac{|\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2}|\boldsymbol{Q_n}|^{1/2}}{|\boldsymbol{Q}_c(\boldsymbol{\theta})|^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x}\right) \\ &\times\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{Q}_n(\boldsymbol{\theta})(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\right) \\ &\times\exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c(\boldsymbol{\theta}))^T\boldsymbol{Q}_c(\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\mu}_c(\boldsymbol{\theta}))\right).\end{aligned} \tag{37}$$

And hence the logarithm of the posterior distribution is

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \text{Const} &+ \log(\pi(\boldsymbol{\theta})) + \frac{1}{2}\log(|\boldsymbol{Q}(\boldsymbol{\theta})|) \\ &- \frac{1}{2}\log(|\boldsymbol{Q}_c(\boldsymbol{\theta})|) + \frac{1}{2}\boldsymbol{\mu}_c(\boldsymbol{\theta})^T\boldsymbol{Q}_c(\boldsymbol{\theta})\boldsymbol{\mu}_c(\boldsymbol{\theta}).\end{aligned} \tag{38}$$

4.2 Priors for the parameters

The prior distribution is important in Bayesian inference, and choosing the priors is an important part of inference. Two common approaches for choosing the prior distribution are the conjugate prior approach and the non-informative prior approach. There is no unique way for choosing priors. We refer to Robert (2007, Chapter 3) for detailed discussion about the prior information and prior distribution.

General speaking, it is hard to specify an informative prior for the hyper-parameters in our system of SPDEs approach. Therefore, the non-informative approach has been chosen. The following choice for the priors of the parameters are recommended with the bivariate random fields.

- b_{11} and b_{22} should be positive values. So log-normal distributions are used for these two parameters. Gamma distribution can also be considered;
- Because of the requirement on the systems of SPDEs that $\{h_{ij}; i, j = 1, 2\}$ and $\{\kappa_{n_i}; i = 1, 2\}$ should be positive values, we can use log-normal or gamma distributions;
- b_{21} is related to the sign of the correlation of the two random fields and it can be either positive or negative. Therefore, a Gaussian distribution can be used;
- The oscillation parameter ω should fulfill the requirement $\omega \in [0, 1]$ and hence a beta distribution can be used.

4.3 Inference with simulated data

Four simulated data examples are presented in this section to illustrate how to use our proposed approach. The datasets are divided into 2 groups. In the first group we use the correlated random fields given in Section 3.4. In the second group the fields are independent. We want our model to capture these features, and to return whether $b_{21} = 0$ or not. However, if the first noise process is generated from the univariate SPDE given in Equation (2), $\kappa_{n_1}^2$ and h_{11} are not identifiable. See Appendix B for more information. We use the setting $\kappa_{n_1}^2 = h_{11}$ in this situation. It is our experience that ω is likely to be in the range of $(0.5, 1)$ if we have empirical

Table 2: Inference for the simulated dataset 1

Parameters	True values	Estimates	Standard deviations
b_{11}	0.5	0.495	0.013
b_{21}	0.25	0.248	0.017
b_{22}	1	1.027	0.032
h_{11}	0.25	0.248	0.010
h_{22}	0.36	0.355	0.029
κ_{n_2}	0.6	0.601	0.004
ω	0.95	0.953	0.092

knowledge that the random field has an oscillating covariance function, and hence we recommend to use a beta distribution with negative skew. In all of our simulated data examples, we use the following priors for the parameters (if they are needed to be estimated) following the discussion given in Section 4.2.

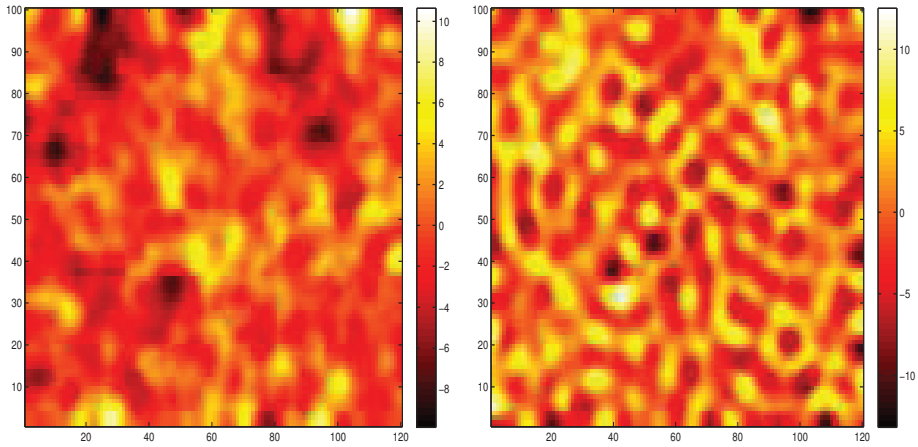
- $b_{11}, b_{22}, h_{11}, h_{22}, \kappa_{n_1}$ and κ_{n_2} have the log-normal distributions with $\mu = 0$ and $\sigma^2 = 100$;
- b_{21} has a normal distribution with $\mu = 0$ and $\sigma^2 = 100$, $b_{21} \sim \mathcal{N}(0, 100)$
- ω has a beta distribution with $\alpha = 1$ and $\beta = 1$, $\omega \sim \text{Beta}(1, 1)$, i.e., it is a uniform distribution.

The results for the first and second simulated datasets are given in Table 2 and Table 3, respectively. We can notice that the estimates are quite precise. Most of the true values are within 1 standard derivation away from the estimates. None of the true values are 2 standard deviations away from the estimates. The estimated conditional mean of the bivariate fields for these two datasets are given in Fig. 7 and Fig. 8. Compare with the true random fields given in Fig. 1(a) - Fig. 1(b) and Fig. 4(a) - Fig. 4(b). There is no large difference between them.

Similarly, the results for the second group are given in Table 4 and Table 5. In both examples, the estimates are precise and they are within 2 standard derivations from the true values. We can notice that if the

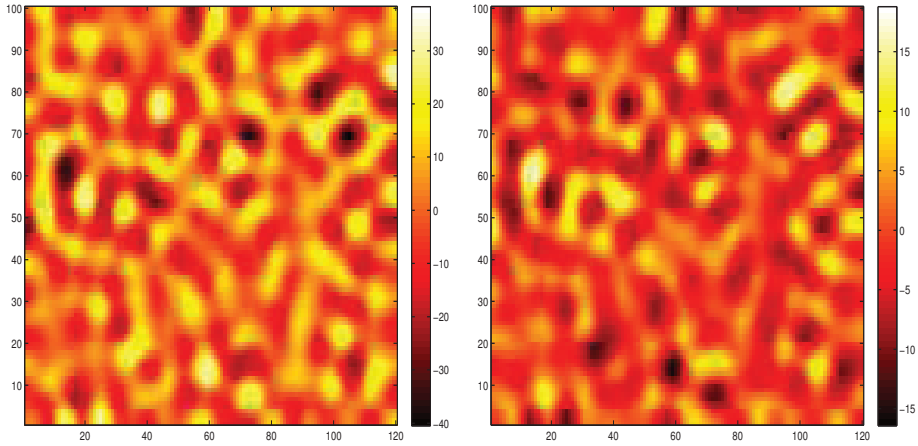
Table 3: Inference for the simulated dataset 2

Parameters	True values	Estimates	Standard deviations
b_{11}	0.5	0.497	0.014
b_{21}	0.25	0.234	0.012
b_{22}	1	0.964	0.029
h_{11}	0.25	0.269	0.024
h_{22}	0.36	0.339	0.022
κ_{n_1}	0.5	0.496	0.005
κ_{n_2}	0.6	0.636	0.049
ω	0.95	0.956	0.113



(a) (b)

Figure 7: Estimated conditional mean of the bivariate random field for the dataset 1.



(a) (b)

Figure 8: Estimated conditional mean of the bivariate random field for the dataset 2.

fields are independent, i.e., $b_{21} = 0$, our model captures this characteristic since b_{21} is small and 0 is within the 95% credible interval.

4.4 Inference with real data

In this section a dataset has been chosen in order to illustrate how to use our approach in real-world applications. This dataset is from the ERA 40 database and can be downloaded from the ERA 40 project homepage. The dataset contains the temperature and pressure on the whole globe on 4th of September, 2002. The main objective for this section is to illustrate how to use our model for a big dataset. All the results are only from the prediction point of view. The dataset contains 10368 observations both for temperature and pressure, and the observations are on the grid. The grid is constructed with the latitude and longitude. The latitudes are from 90^o to -90^o and longitudes are from 0^o to 357.5^o, with increments of 2.5^o for both axes. The dataset contains the temperatures in Kelvin and the mean sea level pressure in Pascal. We have subtracted the monthly mean for the temperature and pressure, respectively.

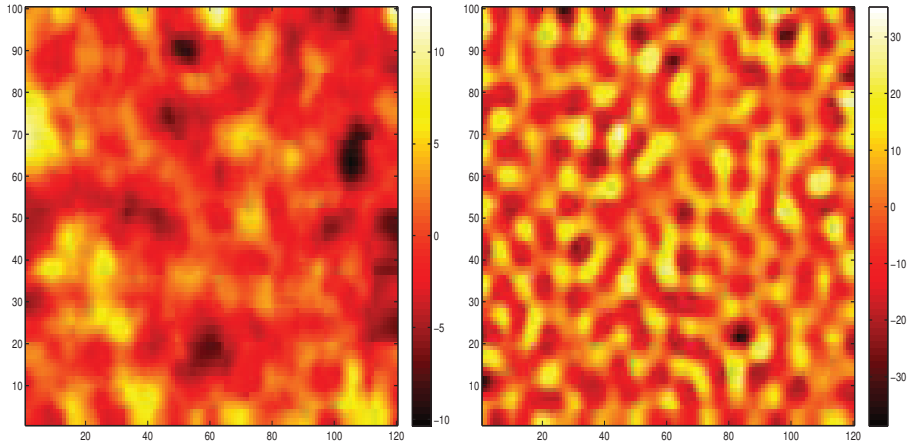
Since the dataset is on the entire globe, we need to construct our

Table 4: Inference for the simulated dataset 3

Parameters	True value	Estimated	Standard deviations
b_{11}	0.5	0.491	0.012
b_{21}	0	0.012	0.010
b_{22}	0.3	0.301	0.010
h_{11}	0.25	0.247	0.009
h_{22}	0.36	0.374	0.033
κ_{n_2}	0.6	0.596	0.004
ω	0.95	0.951	0.092

Table 5: Inference for the simulated dataset 4

Parameters	True value	Estimated	Standard deviations
b_{11}	0.5	0.487	0.015
b_{21}	0	0.001	0.002
b_{22}	0.3	0.308	0.009
h_{11}	0.25	0.284	0.026
h_{22}	0.36	0.359	0.122
κ_{n_1}	0.5	0.502	0.004
κ_{n_2}	0.6	0.599	0.102
ω	0.95	0.949	0.107



(a) (b)

Figure 9: One realization of the bivariate random field with parameters given in Table 4. The first field random field has a non-oscillating covariance function and the second field has an oscillating covariance. The two random fields are independent.

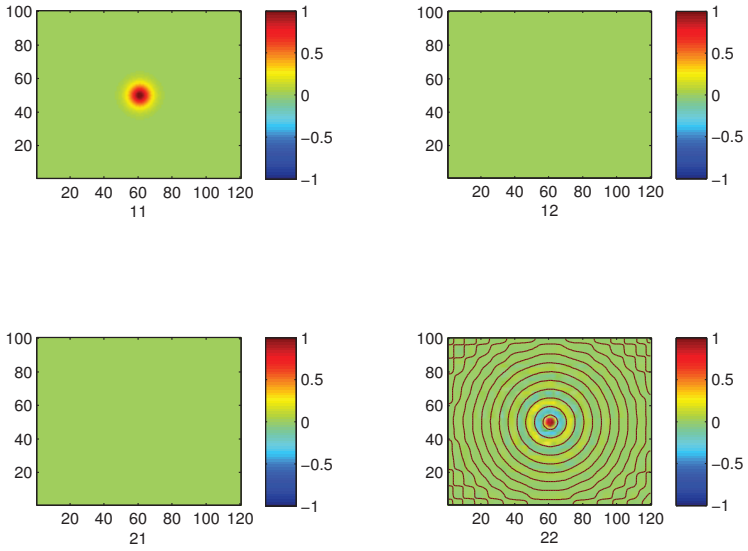


Figure 10: The correlation and cross correlation functions with parameters given in Table 4. The first field random field has a non-oscillating covariance function and the second field has an oscillating covariance. The two random fields are independent.

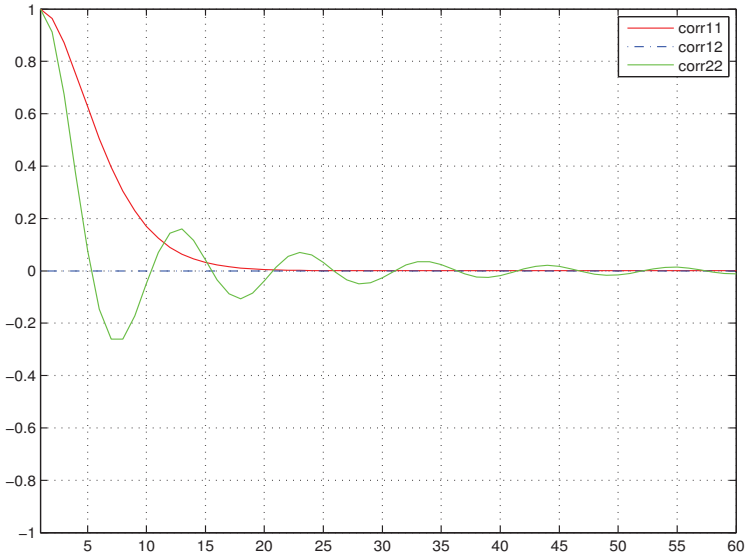
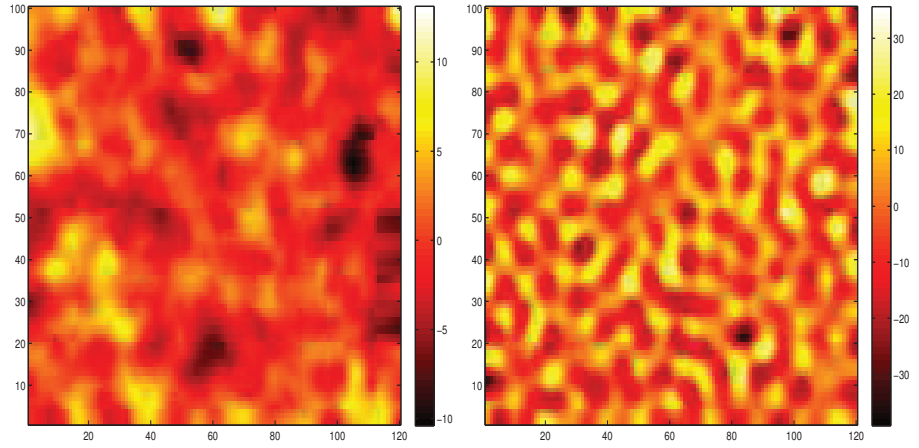


Figure 11: The correlation and cross correlation functions with parameters given in Table 4. The first field random field has a non-oscillating covariance function and the second field has an oscillating covariance. The two random fields are independent.

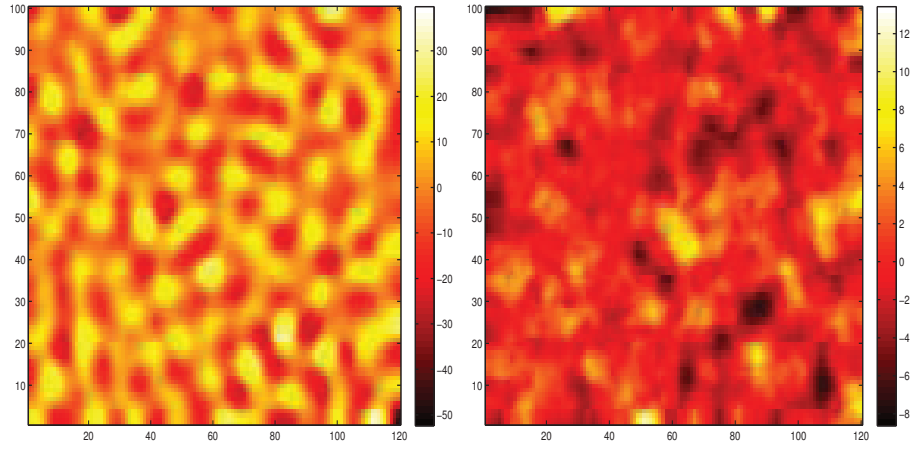


(a) (b)

Figure 12: Estimated conditional mean of the bivariate random fields for dataset 3.

model on the sphere. Jones (1963) discussed how to construct stochastic processes on a sphere using the spectral representations for spherically symmetric and the axially symmetric cases. Another approach is to consider the sphere as a surface in $\mathbb{R}^3$. However, this has the disadvantage that the correlation between points are determined by the chordal distances (Lindgren et al., 2011). Gneiting (1998) pointed out that the random fields constructed on the plane were not suitable for this kind of dataset since the great circle distances in the original covariance function would not work in general. Jun and Stein (2007) discussed an approach for constructing space-time covariance functions on spheres using a sum of independent processes. The main idea for their approach is to sum independent processes where each process is obtained by applying the first-order differential operations to a fully symmetric processes on sphere $\times$ time. We refer to Jun and Stein (2007) for more information on the fully symmetric processes on sphere $\times$ time.

In this paper we follow the approach discussed by Lindgren et al. (2011) to construct the GRFs on the sphere. They claimed that using the SPDE approach for constructing GRFs on the sphere is similar to constructing the GRFs on $\mathbb{R}^d$. By reinterpreting the SPDE defined on $\mathbb{S}^2$, the solutions



(a) (b)

Figure 13: One realization of the bivariate random fields with parameters given in Table 5. The field random field has an oscillating covariance function and the second field has a non-oscillating covariance. The two random fields are independent.

of the SPDE are GRFs defined on $\mathbb{S}^2$. Our proposed system of SPDEs approach inherits this property. The only place has been changed is that the system of SPDEs is directly defined on $\mathbb{S}^2$. Another advantage of our approach is that the GMRF approximation can still be used. In other words, we can use GMRFs to represent GRFs for computation. For more information about GRFs on manifolds, we refer to Lindgren et al. (2011, Section 3.1).

Since it is known that the pressure on the globe has an oscillating covariance function, it is reasonable to set $x_1(\boldsymbol{s})$ as the temperature and $x_2(\boldsymbol{s})$ as the pressure, and let the second noise process $\varepsilon_2(\boldsymbol{s})$ have an oscillating covariance function, but not the first noise process $\varepsilon_1(\boldsymbol{s})$. The original dataset is shown in Fig. 17(a) and Fig. 17(b), and the reconstructed temperature and pressure are shown in Fig. 18(a) and Fig. 18(b). We also give $3D$ images for the true datasets on the sphere in Fig. 19(a) and Fig. 19(b), and the reconstructed temperature and pressure on the sphere in Fig. 20(a) and Fig. 20(b). One thing we want to point out is that we

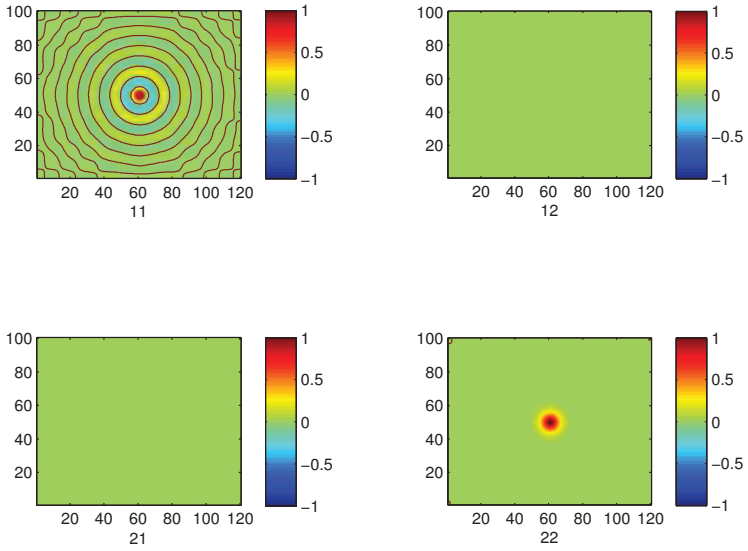


Figure 14: The correlation functions and cross correlation functions with parameters given in Table 5. The field random field has an oscillating covariance function and the second field has a non-oscillating covariance. The two random fields are independent.

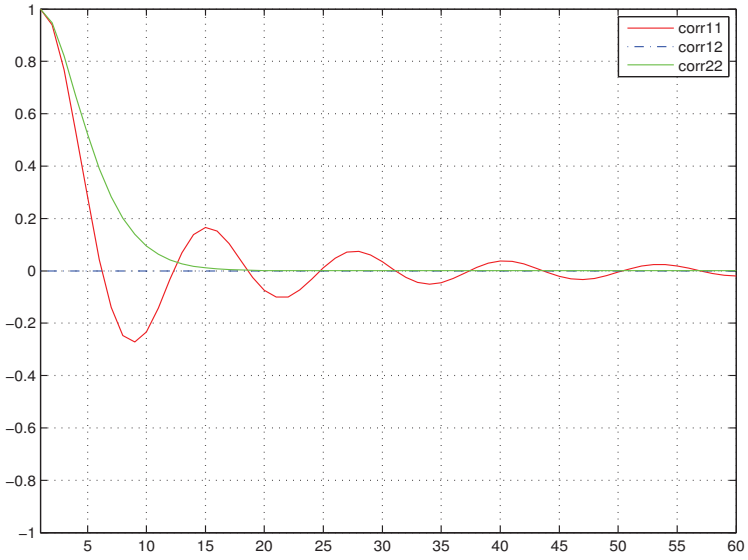
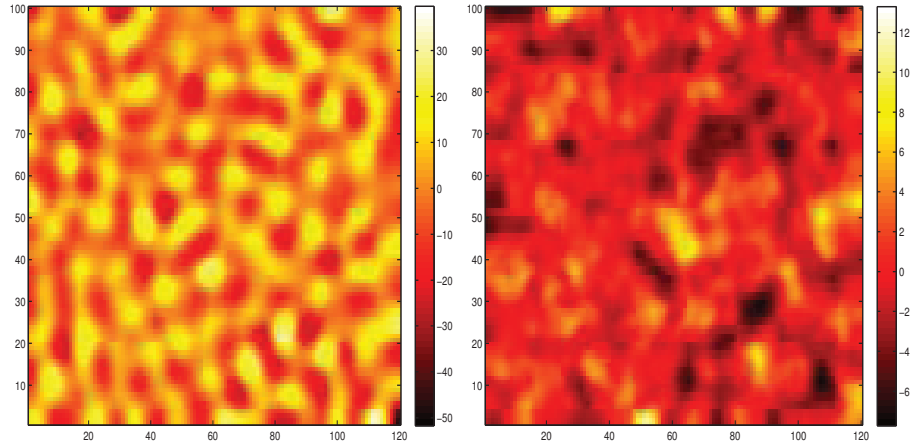


Figure 15: The correlation functions and cross correlation functions with parameters given in Table 5. The field random field has an oscillating covariance function and the second field has a non-oscillating covariance. The two random fields are independent.

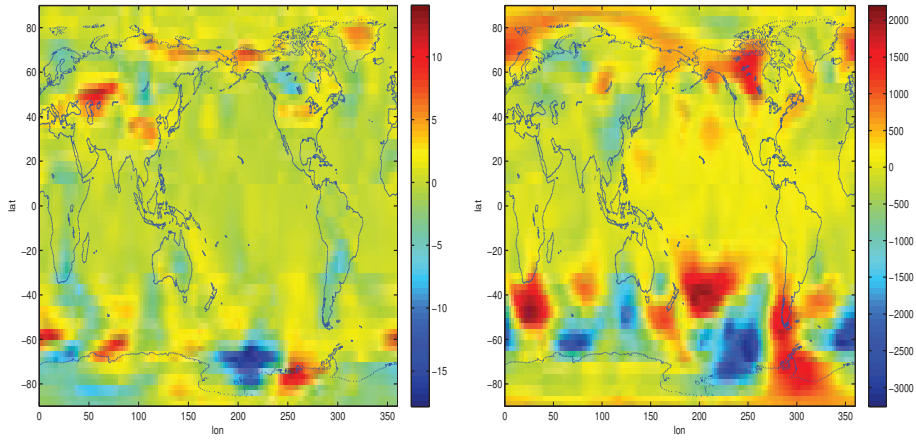


(a) (b)

Figure 16: Estimated conditional mean of the bivariate random fields for dataset 4.

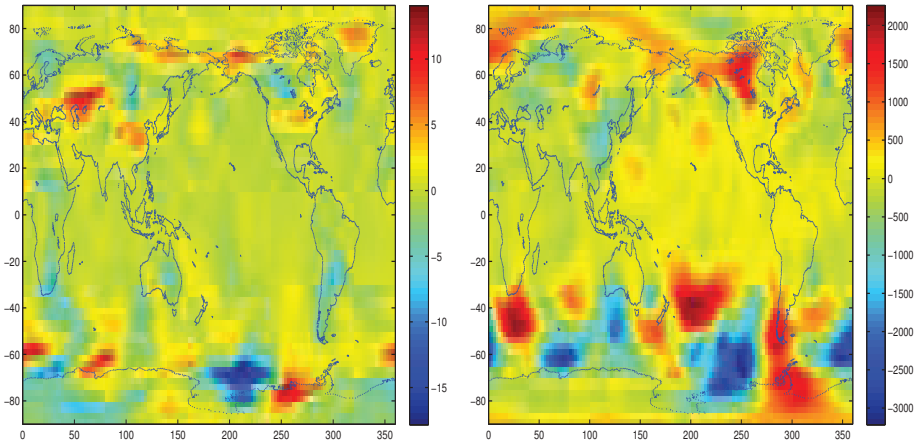
follow the methodology given in Lindgren et al. (2011) and construct the GRF on the unit radius sphere $\mathbb{S}^2$. Another important point is that we set $\kappa_{n_1}^2 = h_{11}$ to simplify the model and inference.

In order to check the predictive performance of our approach, we have divided the dataset into two subsets. We used a subset containing 5368 observations for both temperature and pressure for estimating the parameters and predict the remaining 5000 observations. The estimates are given in Table 6. From the results we notice that the model captures the empirical knowledge that the temperature and pressure are negative correlated since $b_{21} > 0$. The prediction for the 5000 observations are given in Fig. 21. From a prediction point of view, the model works well since most of the prediction are close to the true observed values. The correlation functions are given in Fig. 22, and we notice that the covariance function of pressure indeed has oscillation behavior.



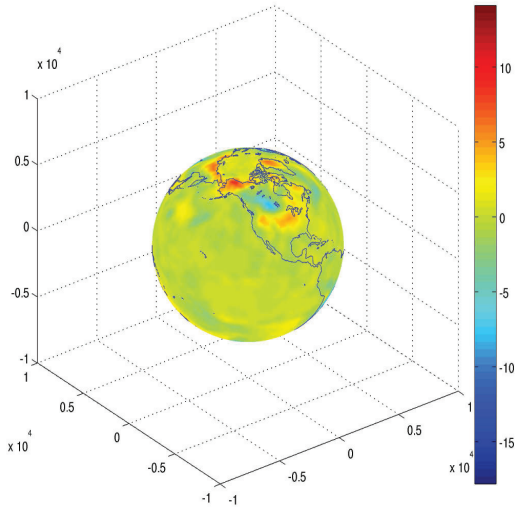
(a) (b)

Figure 17: Dataset with temperature (a) and pressure (b) from ERA 40 database.

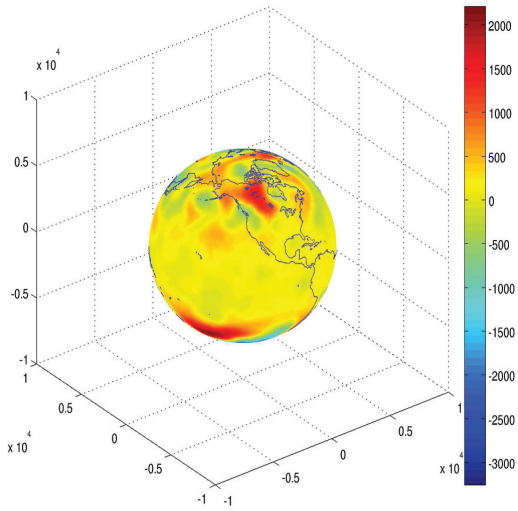


(a) (b)

Figure 18: Reconstructed bivariate random fields for temperature (a) and pressure (b).

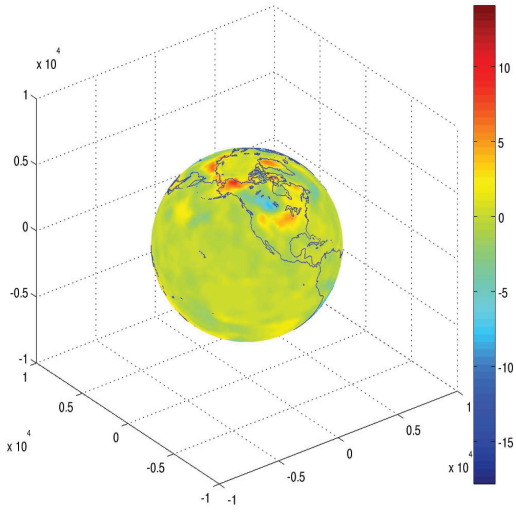


(a)

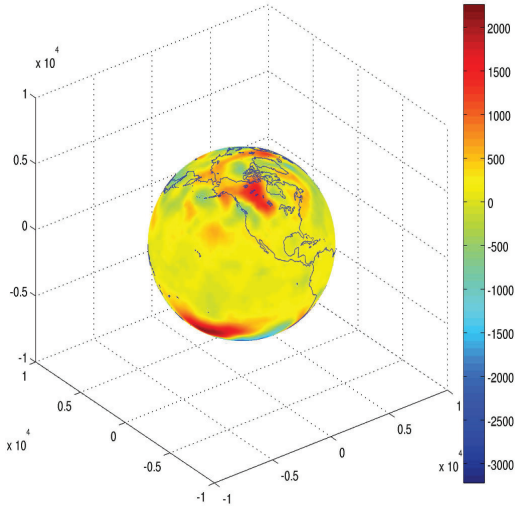


(b)

Figure 19: Dataset with temperature (a) and pressure (b) on the sphere from ERA 40 database.



(a)



(b)

Figure 20: Reconstructed bivariate random fields for temperature (a) and pressure (b) on the sphere.

Table 6: Inference for real dataset

Parameters	Estimated
b_{11}	8.952×10^5
b_{21}	1.546
b_{22}	4.714×10^1
h_{11}	1.224×10^{-6}
h_{22}	8.089×10^{-6}
κ_{n_2}	1.013×10^{-3}
ω	0.819

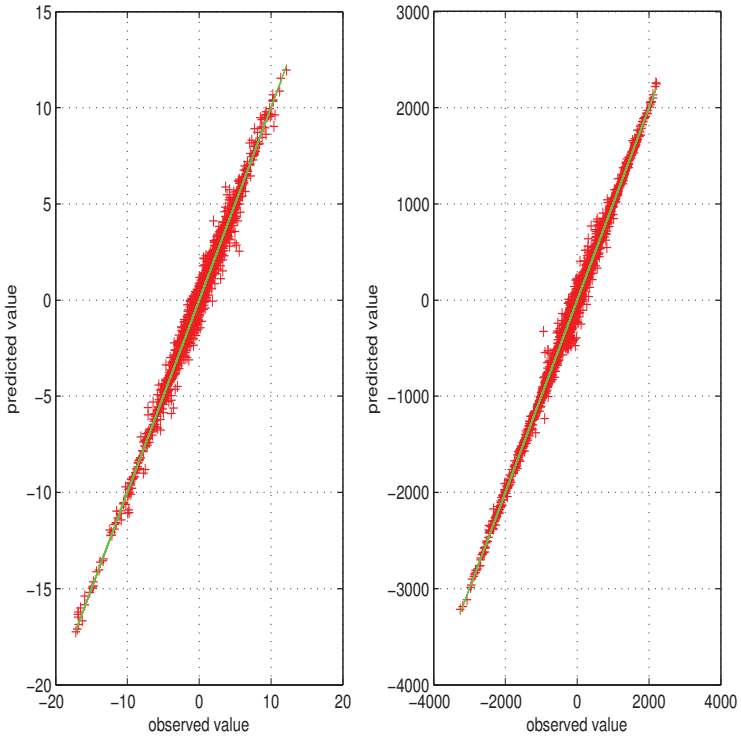


Figure 21: Prediction for the leaving out 5000 observations for temperature (left) and pressure (right)

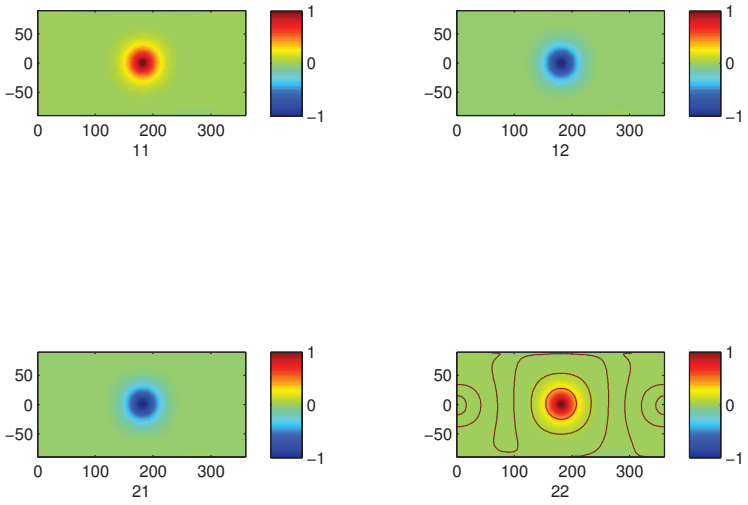


Figure 22: Correlation functions for the bivariate random fields. 11 indicates the marginal correlation of the temperature. 12 and 21 indicate the cross-correlation between the temperature and pressure. 22 indicates the marginal correlation of the pressure.

5 Discussion and future work

Due to the increasing importance of spatial statistics in applications, new approaches for handling different complex datasets are demanded. The methodologies for dealing with multivariate datasets appear in many areas, such as in air quality (Brown et al., 1994; Schmidt and Gelfand, 2003), weather forecasting (Courtier et al., 1998; Reich and Fuentes, 2007), and economics (Gelfand et al., 2004; Sain and Cressie, 2007). Two of the most important issues with these methodologies are how to handle large datasets and how to ensure the nonnegative definiteness constraint for the covariance function. Gneiting et al. (2010) gave some theorems in order to construct valid covariance functions for multivariate random fields. In their approach every component in the matrix-valued covariance function was a Matérn covariance function. Hu et al. (2012a) proposed to use the systems of SPDEs to construct multivariate GRFs with isotropic and non-oscillating covariance functions. The summary paper by Sun et al. (2012) discussed the approaches for how to handle large datasets. They discussed several approaches such as separable covariance structures (Genton, 2007; Fuentes, 2006), covariance tapering (Furrer et al., 2006; Zhang and Du, 2008), likelihood approximations (Vecchia, 1988; Stein et al., 2004), fixed rank kriging and fixed rank filtering (Cressie and Johannesson, 2008; Cressie et al., 2010) and Gaussian Markov random fields approximation (Rue and Tjelmeland, 2002; Rue et al., 2004; Rue and Held, 2005; Rue et al., 2009; Lindgren et al., 2011).

This paper is an extension of Lindgren et al. (2011) and Hu et al. (2012a). The main contribution of this paper is the proposed approach for constructing multivariate random fields with oscillating covariance functions using systems of SPDEs. The main idea is to use noise processes with oscillating covariance functions in order to introduce oscillation in the covariance functions of the random fields. We recommend to use the triangular systems of SPDEs since these models have some advantages. For instance, we have fewer hyper-parameters and we can locate which random fields have non-oscillating, oscillating, possibly oscillating covariance functions. This approach can construct many models discussed by Hu et al. (2012a) if we set the oscillation parameter $\omega = 0$. It also inherit most of the advantages of the SPDE approach discussed by Lindgren et al. (2011) and systems of SPDEs approach discussed by Hu et al. (2012a).

The two main challenges in multivariate random fields mentioned above can be partially solved with our model. On the theoretical side, the covariance functions of the multivariate random fields fulfill the nonnegative definite constraint automatically. On the computational side, the GMRF representation makes the precision matrices to be sparse. Thus numerical algorithms for sparse matrices can be used for fast sampling and inference. Four simulated datasets and one real dataset have been used to illustrate how to use our approach in different situations. The results have illustrated the effectiveness of the proposed approach.

There are several possible extensions for further research, such as constructing non-stationary multivariate GRFs from the systems of SPDEs, and spatio-temporal models both in $\mathbb{R}^d$ and on manifolds. It should also be possible to use the integrated nested Laplace approximation (INLA) framework (Rue et al., 2009) for doing inference for the multivariate GRFs. More applied work using the proposed approach is under development.

Appendix A

There are different kinds of parametrization for the system of SPDEs. The main idea is to change the operators in (22) and use a different parametrization as discussed in Section 3.3. Two intuitive operator matrices are

$$\mathscr{L}_2(\boldsymbol{\theta}) = \begin{pmatrix} b_{11}\,(h_{11} - \Delta) & 0 \\ b_{21}\,(h_{21} - \Delta) & b_{22}\,(h_{22} - \Delta) \end{pmatrix}, \tag{39}$$

$$\mathscr{L}_3(\boldsymbol{\theta}) = \begin{pmatrix} b_{11}\,(h_{11} - \Delta) & 0 \\ b_{21} & b_{22} \end{pmatrix}. \tag{40}$$

With (39) and (40) the correlations between the fields will be changed. The operator matrix given in (39) introduces more flexibility since we have one more parameter h_{21} to control the range of cross-correlation. However, it might be hard to estimate all the parameters in this case. With operator matrix given in (40), we have fewer parameters in the model, but the correlation structure between the fields is simplified. The second random field has the same correlation range as the first field. These two systems can be use in different applications.

Appendix B

When we use the triangular systems of SPDEs, we need to set the constraint $\kappa_{n_1}^2 = h_{11}$ since when the first noise process is generated from Equation (2), they are not identifiable together. Write the system of equations for the bivariate random field given in (30) explicitly as

$$\begin{aligned} b_{11}(h_{11} - \Delta)x_1(\boldsymbol{s}) &= \varepsilon_1(\boldsymbol{s}), \\ b_{21}x_1(\boldsymbol{s}) + b_{22}(h_{22} - \Delta)x_2(\boldsymbol{s}) &= \varepsilon_2(\boldsymbol{s}). \end{aligned}$$

Assume the first noise process $\varepsilon_1(\boldsymbol{s})$ is generated by

$$(\kappa_{n_i}^2 - \Delta)\varepsilon_i(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}). \tag{41}$$

We can now rewrite the first equation in system (30) as a system of equations

$$\begin{aligned} b_{11}(h_{11} - \Delta)x_1(\boldsymbol{s}) &= \varepsilon_1(\boldsymbol{s}), \\ (\kappa_{n_1}^2 - \Delta)\varepsilon_1(\boldsymbol{s}) &= \mathcal{W}(\boldsymbol{s}). \end{aligned} \tag{42}$$

This system of equations can be rewritten into one equation with white noise as the driving process,

$$(h_{11} - \Delta)(\kappa_{n_1}^2 - \Delta)x_1(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}). \tag{43}$$

It is obvious that $\kappa_{n_1}^2$ and h_{11} are not identifiable from each other since $(h_{11} - \Delta)$ and $(\kappa_{n_1}^2 - \Delta)$ commute. Therefore, we suggest the constraint $\kappa_{n_1}^2 = h_{11}$. However, if the first noise process is oscillating and is generated from Equation (15), we don't need this constraint because they are identifiable. However, we still recommend to use this setting in order to simplify the inference.

References

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman & Hall, 2004. ISBN 158488410X.

D. Bolin and F. Lindgren. Wavelet markov models as efficient alternatives to tapering and convolution fields. Technical report, Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University, 2009.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550, 2011.

S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*, volume 15. Springer Verlag, 2008.

P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994.

P Courtier, E Andersson, W Heckley, J Pailleux, D Vasiljevic, M Hamrud, A Hollingsworth, F Rabier, and M Fisher. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124:1783–1807, 1998.

N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70 (1):209–226, 2008.

N. Cressie, T. Shi, and E.L. Kang. Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19(3):724–745, 2010.

N.A.C. Cressie. *Statistics for spatial data*, volume 298. Wiley-Interscience, 1993.

P.J. Diggle and P.J. Ribeiro Jr. *Model-based Geostatistics*. Springer, 2006.

P.J. Diggle, JA Tawn, and RA Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998. ISSN 1467-9876.

M. Fuentes. Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447–466, 2006.

G.A. Fuglstad. Approximating solutions of stochastic differential equations with gussian markov random fields. Technical report, Department of Mathematical Science, Norwegian University of Science and Technology, 2010.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011.

R. Furrer, M.G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, 2006.

A.E. Gelfand, A.M. Schmidt, S. Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp. *Handbook of spatial statistics*. CRC Press, 2010.

M.G. Genton. Separable approximations of space-time covariance matrices. *Environmetrics*, 18(7):681–695, 2007.

T. Gneiting. Simple tests for the validity of correlation function models on the circle. *Statistics & probability letters*, 39(2):119–122, 1998.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

J.A. Goff and T.H. Jordan. Stochastic modeling of seafloor morphology: Inversion of sea beam data for second-order statistics. *Journal of Geophysical Research*, 93(B11):13589–13, 1988. ISSN 0148-0227.

M.S. Handcock and M.L. Stein. A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410, 1993. ISSN 0040-1706.

L. Hartman and O. Hössjer. Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 52(5):2331–2349, 2008. ISSN 0167-9473.

Ø. Hjelle and M. Dæhlen. *Triangulations and applications*. Springer Verlag, 2006.

X. Hu, D.P. Simpson, F. Lindgren, and H. Rue. Multivariate gaussian random fields using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012a.

X. Hu, D.P. Simpson, and H. Rue. Specifying gaussian markov random fields with incomplete orthogonal factorization using givens rotations. Technical report, Department of Mathematical Science, norwegian University of Science and Technology, 2012b.

R.H. Jones. Stochastic processes on a sphere. *The Annals of mathematical statistics*, 34(1):213–218, 1963.

M. Jun and M.L. Stein. An approach to producing space–time covariance functions on spheres. *Technometrics*, 49(4):468–479, 2007.

C.G. Kaufman, M.J. Schervish, and D.W. Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.

P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 3rd edition, 1999. ISBN 3540540628.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

G. Lindgren. *A second course on stationary stochastic processes*. Center for Mathematical Sciences, Lund University, December 2010.

B. Matérn. *Spatial variation*. Springer-Verlag Berlin, 1986.

B.J. Reich and M. Fuentes. A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264, 2007.

C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Verlag, 2007.

H. Rue. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338, 2001. ISSN 1467-9868.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue and H. Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49, 2002. ISSN 1467-9469.

H. Rue, I. Steinsland, and S. Erland. Approximating hidden Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):877–892, 2004. ISSN 1467-9868.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868.

S.R. Sain and N. Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007.

A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108(D24):8783, 2003.

B. Shaby and D. Ruppert. Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452, 2012.

M.L. Stein. *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag, 1999. ISBN 0387986294.

M.L. Stein, Z. Chi, and L.J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296, 2004.

Y. Sun, B. Li, and M.G. Genton. Geostatistics for large datasets. *Advances and challenges in space-time modelling of natural events*, pages 55–77, 2012.

A.V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 297–312, 1988.

W.W.S. Wei. *Time series analysis: univariate and multivariate methods*. Addison-Wesley, 2006.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3-4):434–449, 1954. ISSN 0006-3444.

P. Whittle. Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, 40:974–994, 1963.

H. Zhang and J. Du. Covariance tapering in spatial statistics. *Positive definite functions: From Schoenberg to space-time challenges*, pages 181–196, 2008.

O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, and JZ Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-heinemann, 2005.

Paper III

Spatial Modelling of Temperature and Humidity with System of Stochastic Partial Differential Equations

Xiangping Hu, Ingelin Steinsland, Daniel Simpson, Sara Martino, and Håvard Rue

Technical Report, 2013

Spatial Modelling of Temperature and Humidity using Systems of Stochastic Partial Differential Equations

Xiangping Hu*[1], Ingelin Steinsland[1], Daniel Simpson[1], Sara Martino[2] and Håvard Rue[1]

[1]Department of Mathematical Science, Norwegian University of Science and Technology, Trondheim, Norway
[2]SINTEF, Trondheim, Norway

Abstract

This work is motivated by constructing a weather simulator for precipitation. Temperature and humidity are two of the most important driving forces of precipitation, and the strategy is to have a stochastic model for temperature and humidity, and use a deterministic model to go from these variables to precipitation. Temperature and humidity are empirically positively correlated. Generally speaking, if variables are empirically dependent, then multivariate models should be considered. With the increasing rich methodologies introduced in spatial and computational statistics, using multivariate random fields for modeling multivariate spatial dataset is now available for full Bayesian analysis. In this work we model humidity and temperature in southern Norway. We want to construct bivariate Gaussian random fields (GRFs) based on this dataset. The aim of our work is to use the bivariate GRFs to capture both the dependence structure between humidity and temperature as well as their spatial dependencies. One important feature for the dataset is that the humidity and temperature are not necessarily observed at the

*Corresponding author. Email: Xiangping.Hu@math.ntnu.no

same locations. Both univariate and bivariate spatial models are fitted and compared. For modeling and inference the SPDE approach proposed by Lindgren et al. (2011) for univariate models and the systems of SPDEs approach introduced by Hu et al. (2012) for multivariate models have been used.

To evaluate the performance of the difference between the univariate and bivariate models, we compare predictive performance using some commonly used scoring rules: mean absolute error (MAE), mean-square error (MSE) and continuous ranked probability score (CRPS). The results illustrate that we can capture strong positive correlation between the temperature and the humidity. Furthermore, the results also agree with the physical or empirical knowledge. At the end, we conclude that using the bivariate GRFs to model this dataset is superior to the approach with independent univariate GRFs both when evaluating point predictions and for quantifying prediction uncertainty.

Keywords: Spatial statistics, SPDEs, bivariate random fields, covariance matrix, Gaussian random fields, Gaussian Markov random fields

1 Introduction

Using spatial statistical models for spatial datasets is of great importance in real-world applications. There are many different approaches for modelling spatial datasets. For instance, Cressie (1993) discussed many of the commonly used statistical methods in spatial statistics. Some theories for kriging were discussed by Stein (1999). Diggle and Ribeiro Jr (2006) gave a detailed discussion for geostatistical datasets from a model-based perspective. A handbook for spatial statistics (Gelfand et al., 2010) gave a comprehensive discussion on different methodologies in spatial statistics. A book written by Cressie and Wikle (2011) was emphasized on discussing on statistics methods for spatio-temporal data.

With the increasing requirement of prediction accuracy in spatial statistics, using *multivariate* models to capture the dependence between the components in the dataset is one of the common approaches to fulfill this requirement if the components are actually or empirically dependent. Multivariate models have been under research for a long time. For instance, Gneiting et al. (2010) and Hu et al. (2012b,a) proposed some methods

to build stationary and isotropic models, and Gelfand et al. (2004) and Apanasovich et al. (2012) presented approaches to deal with nonstationarity in multivariate settings. There are many applications for multivariate random fields, such as in economics (Gelfand et al., 2004; Sain and Cressie, 2007), in the area of air quality (Brown et al., 1994; Schmidt and Gelfand, 2003), weather forecasting (Courtier et al., 1998; Reich and Fuentes, 2007) and quantitative genetics (Mcguigan, 2006; Konigsberg and Ousley, 2009).

Generally speaking, if the components in a dataset are empirically dependent, a multivariate model should be taken into consideration. In this paper we model the dataset with humidity and temperature from southern Norway. Since it is known that temperature and humidity are empirically positively correlated, and they are two of the most important driving forces of precipitation, we want to construct bivariate Gaussian random fields (GRFs) for the dataset. The strategy is to have a stochastic model for temperature and humidity, and use a deterministic model to go from these variables to precipitation. The aim of paper is to use the bivariate GRFs to model not only the marginal covariance functions for temperature and humidity but also the cross-covariance function between the temperature and humidity. The posterior mean (given the observations) surfaces of temperature and humidity are reconstructed which can then be used for simulating the precipitation. One important feature of the dataset is that the observations for humidity and temperature are not necessarily measured at the same locations. The dataset is fully discussed in Section 2.

We use the approach proposed by Hu et al. (2012b) to construct bivariate Gaussian random fields for humidity and temperature. With this approach systems of stochastic partial differential equations (SPDEs) are used to build Gaussian random fields for the dataset. There are two main advantages by using the proposed systems of SPDEs approach (Hu et al., 2012b). The first advantage is that the notorious nonnegative definiteness requirement for the covariance matrix is satisfied automatically since the constructed covariance matrix of the GRF from this approach is symmetric positive definite. The second advantage is that we can apply Gaussian Markov random field (GMRF) approximation to the constructed GRF. Since GMRF models are usually computationally efficient, this approach can be applied to large datasets. A brief introducton to GMRF is given

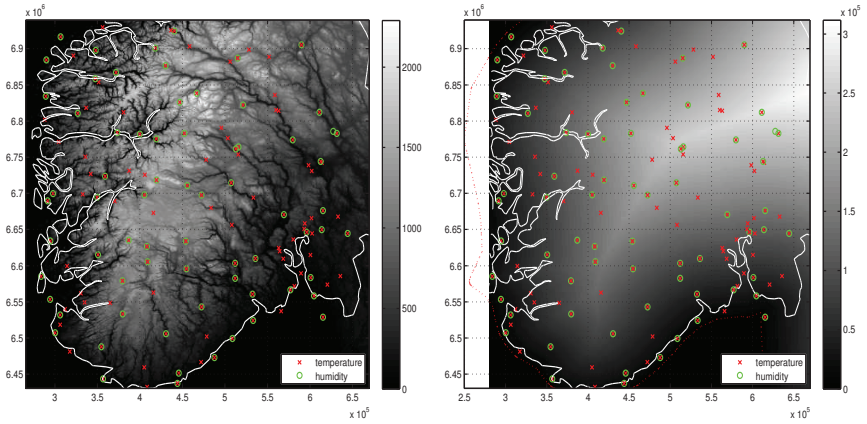
in Section 3.1, and a detailed discussion about the models with SPDE approach is given in Section 3.2 - Section 3.4.

The rest of this paper is organized as followings. Section 2 describes the data. We review the knowledge about the SPDE approach for spatial statistics and introduce the spatial model for our dataset in Section 3. Section 4 discusses the evaluation procedure. Results are given in Section 5. Section 6 ends the paper with discussion and conclusion.

2 Data

The dataset contains daily mean temperature in Celsius degree and humidity recorded measured in mixing ratio for locations in southern Norway. The mixing ratio of humidity is defined as the mass of water vapor contained in a unit mass of dry air, and hence has a unit kg/kg. Two covariates are also included in the model: elevation at the measurement location and the distance to the ocean from the location. Both covariates are in meters. The dataset contains observations for temperature and humidity on 7th of December from year 2007 to year 2011, i.e. for 5 years. It is important to point it out that the observations are not necessarily at the same locations for all the 5 years. Most of humidity observations are measured at a subset of locations of temperature. Figure 1(a) and Figure 1(b) give an overview of locations for temperature and humidity. In addition, the elevation and distance to ocean are available at all location on a 1km by 1km grid. The elevation map of southern Norway and the distances to ocean are given in Figure 1(a) and Figure 1(b), respectively.

The dotted line is the base line for calculating the distance to ocean and the solid line is the coast line of southern Norway. We can clearly see that the distance to ocean is not the same as the distance to the coast. The cross marks ($\times$) and the circle marks ($\circ$) in Figure 1 are locations for temperature and humidity observations on 7th of December in 2011, respectively. The number of observations of temperature and humidity in different years is given in Table 1. Necessary pre-processing of the dataset has been done before modelling. More information about the pre-processing of the dataset together with some empirical data analysis can be found in Section 5.1.



(a) (b)

Figure 1: Locations of temperature and humidity observations on 7th of December in 2011 with elevation (a) and distances to ocean (b) on a 1km by 1km grid. The base line for calculating the distance to ocean (dotted-line) and the coast line (solid line) of southern Norway are also given. The cross marks ($\times$) and the circle marks ($\circ$) are locations for temperature and humidity observations, repectively.

Table 1: Number of observations for temperature and humidity

Year	2007	2008	2009	2010	2011
temperature	97	104	111	122	128
humidity	56	63	62	62	70

3 Model using the SPDE approach

In this section we discuss the construction of spatial models for temperature and humidity using the SPDE approach and system of SPDEs approach. Three models are used and fitted to the data. The first model is a univariate GRF model. In this model we construct independent spatial random fields for temperature and humidity with the approach proposed by Lindgren et al. (2011). The second and the third models are bivariate models constructed with the approach proposed by Hu et al. (2012b), where we model temperature and humidity jointly. Since GMRFs are the main tool for achieving computational efficiency with models built by the SPDE approach, a brief introduction to GMRF is given in Section 3.1. The SPDE approach for spatial modelling univariate and multivariate GRFs are described in Section 3.2 and in Section 3.3, respectively. The spatial models used to model temperature and humidity in this paper are presented in Section 3.4.

3.1 Gaussian Markov random fields

A random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is a Gaussian random field with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$ ($\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$) if and only if its density is

$$\pi(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} |\boldsymbol{Q}|^{1/2} \exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right). \tag{1}$$

where $\boldsymbol{x}_{-ij}$ denotes for $\boldsymbol{x}_{-\{i,j\}}$. $\boldsymbol{Q} > 0$ denotes the precision matrix $\boldsymbol{Q}$ is positive definite. A Gaussian *Markov* random fields is a GRF with Markov property

$$Q_{ij} = 0 \Longleftrightarrow x_i \perp x_j | \boldsymbol{x}_{-ij}, \tag{2}$$

and hence the precision matrix $\boldsymbol{Q}$ for a GMRF is usually sparse. Therefore, numerical algorithms for sparse matrices can be applied when doing computations. Rue and Held (2005) gives a more detailed discussion on the theories for GMRFs. A condensed version about GMRFs can be found in Gelfand et al. (2010, Chapter 12).

3.2 SPDE approach for univariate GRFs

The main idea of the newly proposed approach by Lindgren et al. (2011) is to use SPDEs to construct GRFs for modelling spatial datasets. The SPDE used in this paper has the form

$$b(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \nu > 0, \tag{3}$$

where b is a parameter related to the variance of the random field $x(\boldsymbol{s})$, $\mathcal{W}(\boldsymbol{s})$ is a standard Gaussian white noise process, $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo (fractional) differential operator and α must be a non-negative integer. Δ is the standard *Laplacian* with definition

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

With this approach the most important relationship is that the stationary solution $x(\boldsymbol{s})$ to the SPDE (3) is a GRF with a Matérn covariance function. The Matérn covariance function has the form

$$M(\boldsymbol{h}|\nu, \kappa) = \frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\boldsymbol{h}\|)^{\nu} K_{\nu}(\kappa \|\boldsymbol{h}\|), \tag{4}$$

where ν is the smoothness parameter, κ is the scaling parameter and K_{ν} is the modified Bessel function of second kind of order ν, $\|h\|$ denotes the Euclidean distance in $\mathbb{R}^d$ and σ^2 is the marginal variance. The closed form for σ^2 of random field $x(\boldsymbol{s})$ constructed from Equation (3) is

$$\sigma^2 = \frac{\Gamma(\nu)/\Gamma(\nu + d/2)}{(4\pi)^{d/2} b^2 \kappa^{2\nu}}.$$

The Matérn covariance function is isotropic and it is widely used in spatial statistics (Stein, 1999; Diggle and Ribeiro Jr, 2006; Simpson et al., 2010; Lindgren et al., 2011; Bolin and Lindgren, 2011; Ingebrigtsen et al., 2013; Hu et al., 2012b,a). We call a GRF with a Matérn covariance function a Matérn random field. Lindgren et al. (2011) pointed out that there was an explicit link between GRFs and GMRFs. They showed that GRFs can be represented by GMRFs. By using this technique, we can build the models theoretically with GRFs but doing computations with GMRFs. We use

the finite element methods (FEMs) to solve the SPDE (3), and then apply the GMRF approximation to the solution in order to obtain computationally efficient GMRF models for fast inference. Bolin and Lindgren (2009) showed that the differences between the exact FEM representation and the GMRFs approximation are negligible. We refer to Zienkiewicz et al. (2005) and Bathe (2008) for more information on FEMs. Fuglstad (2011) and Ingebrigtsen et al. (2013) extended the SPDE approach to nonstationary GRFs. Nested SPDEs were proposed by Bolin and Lindgren (2011) for constructing a larger class of models for spatial datasets. Hu et al. (2012b) have extended the approaches from Lindgren et al. (2011) to construct multivariate GRFs. Hu et al. (2012a) proposed to use systems of SPDEs to construct multivariate GRFs with oscillating covariances functions.

Since the smoothness parameter ν is poorly identifiable (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011), we fix $\alpha_{11} = 2$ and $\alpha_{22} = 2$ when we do inference. With this univariate model for modelling humidity and temperature independently we have 4 parameters $\boldsymbol{\theta} = \{\kappa_{11}, \kappa_{22}, b_{11}, b_{22}\}$. The results for this model are given in Section 5.

3.3 Multivariate GRFs with systems of SPDEs

Hu et al. (2012b) extended the approach given by Lindgren et al. (2011) and proposed a new approach for constructing a multivariate GRF using a system of SPDEs. Hu et al. (2012b) claimed that this approach for constructing multivariate GRFs inherits both theoretical and computational advantages from the approach for univariate GRFs given by Lindgren et al. (2011). The system of SPDEs for constructing a multivariate GRF has the form

$$\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{5}$$

where $\mathcal{L}_{ij} = b_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2}$ are similar differential operators as given in Equation (3) with $\{\alpha_{ij} = 0 \text{ or } 2; 1 \leq i, j \leq p\}$, $\{\varepsilon_i(\boldsymbol{s}); i, j = 1, \dots, p\}$ are Gaussian noise processes which are independent but not necessarily

identically distributed. It was shown by Hu et al. (2012b) that the solution $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \ldots, x_p(\boldsymbol{s}))$ to the system of SPDE (5) is a multivariate GRF. The parameters $\{\kappa_{ij}; i, j = 1, \ldots, p\}$ and $\{\nu_{ij}; i, j = 1, \ldots, p\}$ are scaling parameters and smoothness parameters, respectively. $\{b_{ij}; i, j = 1, \ldots, p\}$ are related to both the marginal covariance functions of the fields and the cross-covariance functions among the GRFs. On the theoretical side, similarly as discussed by Lindgren et al. (2011), the precision matrix $\boldsymbol{Q}$ for the multivariate GRF constructed from the system of SPDEs (5) satisfies the positive definite constraint automatically. Hu et al. (2012b) demonstrated that the link between the GMRFs and GRFs could be used, and hence we can construct models with GRFs but use GMRFs for computations. The precision matrix $\boldsymbol{Q}$ of the multivariate GMRF $\boldsymbol{x}(\boldsymbol{s})$ is sparse. Therefore, on the computational side, numerical algorithms for sparse matrices can be applied for sampling and fast inference. These are the main reasons why we have selected this approach for modelling the dataset with temperature and humidity in this paper.

The system of SPDEs we have actually chosen has the form

$$\begin{pmatrix} \mathcal{L}_{11} & \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \mathcal{W}_1(\boldsymbol{s}) \\ \mathcal{W}_2(\boldsymbol{s}) \end{pmatrix}, \tag{6}$$

where $\{\mathcal{W}_i(\boldsymbol{s}); i = 1, 2\}$ are standard Gaussian white noise processes. This is a special case of (5) with $\mathcal{L}_{12} = 0$ and $\{\varepsilon_i(\boldsymbol{s}) = \mathcal{W}_i(\boldsymbol{s}); i = 1, 2\}$ when $p = 2$. This system of SPDEs is called triangular system of SPDEs by Hu et al. (2012b,a). The advantage with a triangular systems of SPDEs is that this simplification makes both computations and interpretation easier. We refer to Hu et al. (2012b,a) for a detailed discussion on the triangular system of SPDEs. The smoothness parameters $\{\nu_{ij}; i, j = 1, 2\}$ are poorly identifiable, and we fix $\{\alpha_{ij} = 2; i, j = 1, 2, i > j\}$ (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011; Hu et al., 2012b,a).

With this setting we know that $x_1(\boldsymbol{s})$ is a Matérn random field and $x_2(\boldsymbol{s})$ is generally not a Matérn random field, but close to a Matérn random field (Hu et al., 2012b). This implies that the order of the random fields matters. Generally speaking, we need to choose the order of the random fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ for temperature and humidity, and this is usually done by a model selection test. Fit models with both orders and pick the one that minimizes some criterion, such as minimizing prediction error. In

this paper we first set the first field $x_1(\boldsymbol{s})$ as temperature and the second field $x_2(\boldsymbol{s})$ as humidity. Then we switch the order of the fields, i.e., we set the first field $x_1(\boldsymbol{s})$ as humidity and the second field $x_2(\boldsymbol{s})$ as temperature.

Using the triangular system of SPDEs (6) for constructing a bivariate GRF, we have 6 parameters to estimate $\boldsymbol{\theta} = \{\kappa_{11}, \kappa_{21}, \kappa_{22}, b_{11}, b_{21}, b_{22}\}$ from the system of SPDEs when we model the temperature and humidity jointly.

Hu et al. (2012b) showed that the sign of cross-correlation between the humidity and temperature is only related to the product $b_{21}b_{22}$ with a triangular system of SPDE. In the extreme case, if b_{21} is zero, i.e., $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are independent, then b_{22} can only be positive value. So restricting b_{22} to be only positive value is a natural choice. Therefore, the sign of the cross-correlation between the two random fields is only related to the sign of b_{21}. When $b_{21} < 0$, $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are positively correlated, and when $b_{21} > 0$, $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are negatively correlated. This setting is chosen in this paper. All results and corresponding discussion are given in Section 5.

3.4 Spatial model for temperature and humidity

As mentioned in Section 2, the dataset contains observations on 7th of December from year 2007 to year 2011 for both temperature and humidity in southern Norway. These observations are not necessary measured at the same locations in each year. We use a model of the following form

$$y_{ijk} = v_{ijk} + \xi_{ijk} + \varepsilon_{ijk}, \tag{7}$$

where i denotes the index of the observation, j denotes the index of the year, k denotes the index of field, v_{ijk} is the fix effect, ξ_{ijk} is the spatial effect and ε_{ijk} is the noise or the measurement error. The noise terms $\{\varepsilon_{ijk}; i = 1, 2, \ldots, n_k, j = 1, 2, \ldots, 5, k = 1, 2\}$ are independent and identically distributed (iid) with Gaussian distribution $\mathcal{N}(\mathbf{0}, \tau^2_{\epsilon k})$, and are independent of the fix effect and spatial effect. $\{n_k; k = 1, 2\}$ denote the number of observations in all years for temperature and humidity, and $\{\tau^2_{\epsilon k}; k = 1, 2\}$ are the measurement error variances for temperature and humidity. Since we assume that the noise processes for temperature and humidity are independent, the precision matrix $\boldsymbol{Q}_\epsilon$ for the noise processes

is a diagonal matrix. Model (7) can be written in vector form

$$\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\boldsymbol{v} = \boldsymbol{X}\boldsymbol{\beta}$ is the fixed effect with coefficients $\boldsymbol{\beta}$ and design matrix $\boldsymbol{X}$, and it consists of effect from covariates and from the year, i.e the year effect. The matrix $\boldsymbol{A}$ links the dense spatial fields to the observations. $\boldsymbol{\xi}$ is a spatial process with mean zero and precision matrix $\boldsymbol{Q}$. The precision matrix $\boldsymbol{Q}$ is constructed using the system of SPDEs (6) for bivariate model. For the univariate model the precision matrices for temperature and humidity are constructed by Equation (3) independently. $\boldsymbol{\varepsilon}$ is the unexplained random effects for humidity and temperature. This model can be formulated as a Bayesian hierarchical model, and it can be stated explicitly as

- Data model: $y_{ijk}|\eta_{ijk} \sim \mathcal{N}\left(\eta_{ijk}, \tau^2_{\epsilon k}\right)$. We assume that $\{\tau^2_{\epsilon k}; k = 1, 2\}$ are known;

- Process model: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}^{-1}\right)$. As discussed above, the precision matrix $\boldsymbol{Q}$ is constructed by the system of SPDEs (6) and the SPDE (3) for bivariate model and univariate model, respectively. Here we denote $\boldsymbol{\xi} \sim$ BSPDE$(\boldsymbol{b}, \boldsymbol{\kappa})$ which means that the spatial effects are construct by (6) for bivariate model, and correspondingly, we use $\boldsymbol{\xi} \sim$ USPDE$(\boldsymbol{b}, \boldsymbol{\kappa})$ for univariate model constructed from (3). We assume that the fixed effects from the covariates, i.e., elevation and distance to ocean, are the same for all 5 years with coefficients $\{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$. However, each year has different yearly effect $\left\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\right\}$ for temperature and humidity in order to capture the multi-year effect.

- Parameter model: Specify the prior distributions $p(\boldsymbol{\theta})$ for parameters $\boldsymbol{\theta} = \{b_{11}, b_{21}, b_{22}, \kappa_{11}, \kappa_{21}, \kappa_{22}\}$ from the spatial effects for bivariate model, and correspondingly, $\boldsymbol{\theta} = \{b_{11}, b_{22}, \kappa_{11}, \kappa_{22}\}$ for univariate model. We also need to specify the prior distributions for the coefficients of the covariates $\{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$ and for the yearly effects $\left\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\right\}$ for both the bivariate model and the univariate model.

The prior distributions of parameters are assumed to be independent and have the following distributions (if the parameter is included in the model),

- $\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\}$: Gaussian distributions
- $\{\beta_{11}, \beta_{21}, \beta_{21}, \beta_{22}\}$: Gaussian distributions
- $\{b_{ii}, i = 1, 2\}$: Log-Gaussian distributions
- b_{21}: Gaussian distribution
- $\{\kappa_{11}, \kappa_{21}, \kappa_{22}\}$: Log-Gaussian distributions

This model formulation is similar to the ones in Hu et al. (2012b) and Hu et al. (2012a).

3.5 Statistical inference

We point out that since the coefficient parameters for the covariates can be modelled with Gaussian distributions, we can treat the coefficients $\boldsymbol{\beta}_j$ as the latent field together the spatial process $\boldsymbol{x}(\boldsymbol{s})$ and model them jointly instead of treating the coefficient parameters as hyper-parameters. Thus the hyper-parameters only contains the parameters from the systems of SPDEs, $\boldsymbol{\theta} = \{b_{11}, b_{21}, b_{22}, \kappa_{11}, \kappa_{21}, \kappa_{22}\}$ for bivariate model and $\boldsymbol{\theta} = \{b_{11}, b_{22}, \kappa_{11}, \kappa_{22}\}$ for univariate model, since we fix the values of $\{\alpha_{ij}; i, j = 1, 2\}$ for both the models. The latent field in this case is $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{\beta})^{\mathrm{T}}$, where T denotes the transpose of a vector or a matrix. This can speed up the optimization considerably since there are much few parameters in the numerical optimization. This is the commonly used setting in Rue et al. (2009).

Let $\boldsymbol{Q}(\boldsymbol{\theta})$ denote the precision matrix for the random fields constructed by the system of SPDEs (6) for the bivariate GRFs or the precision matrix for the univariate random fields with SPDE (3) with hyper-parameters $\boldsymbol{\theta}$. With the univariate model we construct the precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ as a block diagonal precision matrix, then inference for this two univariate random fields can be done simultaneously. In this case we can use the same program for the bivariate model, and the univariate model has only

one more constraint $b_{21} = 0$. Hu et al. (2012b) have shown that from the well known Bayesian formula

$$\pi(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y})}{\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\theta})}. \tag{9}$$

We can derive the posterior distribution for hyper-parameters

$$\begin{aligned} \log(\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \text{Const.} + \log(\pi(\boldsymbol{\theta})) + \frac{1}{2}\log(|\boldsymbol{Q}(\boldsymbol{\theta})|) \\ - \frac{1}{2}\log(|\boldsymbol{Q}_c(\boldsymbol{\theta})|) + \frac{1}{2}\boldsymbol{\mu}_c^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{Q}_c(\boldsymbol{\theta})\boldsymbol{\mu}_c(\boldsymbol{\theta}), \end{aligned} \tag{10}$$

with $\boldsymbol{\mu}_c = \boldsymbol{Q}_c^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_\epsilon\boldsymbol{y}$, $\boldsymbol{Q}_c(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_\epsilon\boldsymbol{C}$, and $\boldsymbol{C} = (\boldsymbol{A}, \boldsymbol{X})$. $\boldsymbol{A}$ is a sparse matrix which links the sparse observations of temperature and humidity to our bivariate GRF or univariate GRFs. $\boldsymbol{X}$ is the design matrix.

4 Evaluation

In this section we explain the evaluation schemes for comparing the results for three different settings given in Section 4.2 with three different models. We have two bivariate models. In the first one we set the first random field as temperature and the second random field as humidity. In the second one we switch the order of the random fields, i.e., we set the first random field as humidity and set the second random field as temperature. We also have one univariate model for comparison. Some commonly used scoring rules described in Section 4.1 are chosen in order to compare the predictive performances.

4.1 Scoring rules

In order to evaluate the results, scoring rules are used to compare the predictive performance between the univariate model and the bivariate model for temperature and humidity with different settings. In this paper the commonly used scoring rules mean absolute error (MAE), mean-square error (MSE) and the average of the continuous ranked probability score

(CRPS) are used. Let $\hat{y}_{ijk}$ denote the prediction for the observations y_{ijk} for the observation i in year j for the kth field, and then the MAE and MSE for the kth field have the following definitions

$$\text{MAE}_k = \frac{1}{n_k} \sum_j \sum_i |y_{ijk} - \hat{y}_{ijk}|,$$
$$\text{MSE}_k = \frac{1}{n_k} \sum_j \sum_i (y_{ijk} - \hat{y}_{ijk})^2,$$

The CRPS is also a commonly used scoring rule to evaluate the probabilistic forecasts, and it is the integral of the Brier scores for a continuous predictand at all possible threshold values p (Hersbach, 2000; Gneiting et al., 2005). We refer to Toth et al. (2003, Section 7.3.2) for detailed discussion about the Brier scores. Let F denote the predictive cumulative distribution function (CDF) and $H(p-y)$ be the Heaviside function with value 1 whenever $p - y > 0$ and value 0 otherwise. Then the continuous ranked probability score is defined as

$$\text{crps}(F, y) = \int_{-\infty}^{\infty} (F(p) - H(p-y))^2 dp. \tag{11}$$

Gneiting et al. (2005) pointed out that if F is the CDF of a Gaussian distribution, then a closed form of the continuous ranked probability score can be obtained, and this form is usually used in applications. The average of continuous ranked probability score which is denoted as CRPS then has the form

$$\text{CRPS}_k = \frac{1}{n_k} \sum_j \sum_i \text{crps}(F_{ijk}, y_{ijk}). \tag{12}$$

We refer to Gneiting et al. (2005) for more information on scoring rules.

4.2 Cross validation scheme

In this section we explain the cross validation scheme for comparing the results from bivariate models with the results from univariate model. Three different settings have been chosen. We only consider the locations where both observations for temperature and humidity are presented in all settings.

Setting 1

In this setting we only predict the second field $x_2(\boldsymbol{s})$ at 20 locations removed from the dataset for each year from 2007 to 2011. The remaining observations from the second fields $x_2(\boldsymbol{s})$ together with all observations in the first field $x_1(\boldsymbol{s})$ are used to estimate parameters.

Setting 2

This setting is similar as the Setting 1. With this setting, however, 20 locations in each year in the first field $x_1(\boldsymbol{s})$ are left out for prediction and all observations in the second field $x_2(\boldsymbol{s})$ together with the rest of observations in the first field $x_1(\boldsymbol{s})$ are used to estimate the parameters.

Setting 3

In this setting we have left out observations at 20 locations from both fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ in each year for prediction. The rest of the observations are used to estimate the parameters.

The scoring rules defined in Section 4.1 are calculated for different models for each of these settings for both univariate and bivariate models. The results are presented in Section 5.

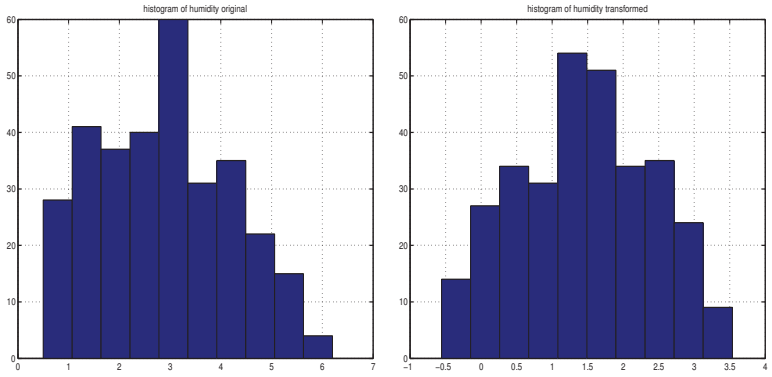
5 Results

Results from different models with the three different settings are discussed in this section. Before modelling the dataset, some empirical data analysis have been conducted in Section 5.1. Inference results of the parameters are given in Section 5.2. Reconstruction of fields for temperature and humidity are shown in Section 5.3. The results for predictive performance are given in Section 5.4.

5.1 Empirical data analysis

Since the observations of humidity are positive, they are preprocessed with the widely used Box-Cox family of transformations given by Box and Cox (1964), in order to transform them to be approximately Gaussian distributed. We use the original observations of temperature. The Box-Cox family of transformations has the form

$$\hat{Y} = \begin{cases} \left(Y^{\lambda} - 1\right)/\lambda & \text{if } \lambda \neq 0 \\ \log(Y) & \text{if } \lambda = 0 \end{cases}. \tag{13}$$

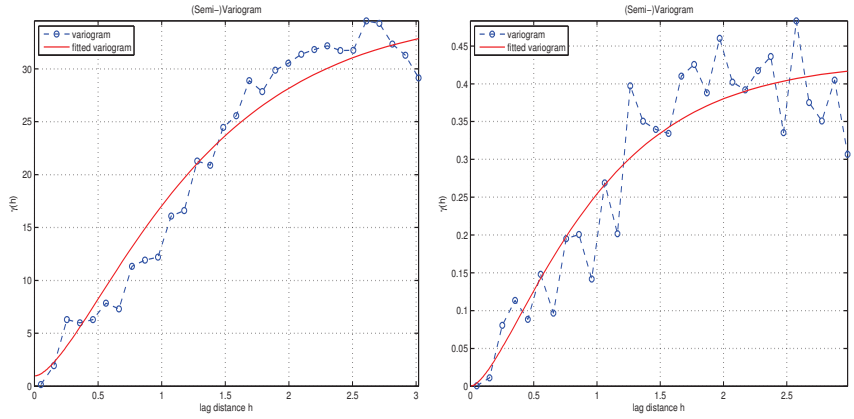


(a) (b)

Figure 2: Histograms of original humidity observations(a) and transformed humidity observations (b).

The estimated values of λ for humidity for the Box-Cox transform is $\lambda = 0.66$, and the histograms for the humidity before the transformation and after the transformation are given in Figure 2(a) and Figure 2(b), respectively. From these two histograms, we can notice that the transformed humidity seems more reasonable to be modelled with Gaussian distribution. For more information about the Box-Cox transformation and other transformation methods, see, for example, Sakia (1992) and Diggle and Ribeiro Jr (2006).

The empirical variograms of both temperature and humidity have been calculated and fitted to theoretical variograms. In the theoretical variograms, we choose to fit the Matérn model. We only include the results for the dataset in 2011 and omit the others since they are similar. The empirical variograms and the corresponding fitted theoretical variograms for this year are shown in Figure 3. This analysis suggest the smoothness parameters for both the fields with $\nu = 1$ are reasonable, and hence fixing $\alpha = 2$ in our analysis is also reasonable. For a detailed discussion on variograms, we refer to Diggle and Ribeiro Jr (2006) and Banerjee et al. (2004). The nugget effects or the measurement error variances for temperature and humidity are assumed known and are fixed to $\tau_1 = 0.1$ and $\tau_2 = 0.01$,respectively, for both the bivariate model and univariate model.



(a) (b)

Figure 3: Empirical variograms for temperature and humidity (broken lines with circles) with fitted theoretical variograms (solid lines) for temperature (a) and humidity (b).

5.2 Inference results of parameters

With Equation (10) together with the priors discussed in Section 3.4, the estimates, with the standard deviations given in the brackets, for the parameters for univariate model and for bivariate model are presented in Table 2 when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. When we switch the order of the fields, i.e., we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, the corresponding results are given in Table 3. From Table 2 and Table 3 we can notice that temperature and humidity are positively correlated since $b_{21} < 0$ for both models.

With $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity, the correlations within temperature and humidity together with the cross-correlation between temperature and humidity can be calculated from the precision matrix $\boldsymbol{Q}$ for the bivariate model, and are given in Figure 4 (solid lines). The correlations within temperature and humidity for the univariate model can be obtained similarly, and are included in Figure 4 (dash-dot lines). From Figure 4 we can notice that the cross-correlation between temperature and humidity at the same location is $\gamma_{th} = 0.6351$. When we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, the corresponding results for the

Table 2: Estimated hyper-parameters for bivariate model and for univariate model with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity

	Bivariate	Univariate
b_{11}	0.0104 (8.106×10^{-4})	0.0104 (8.134×10^{-4})
b_{21}	-0.0219 (2.512×10^{-3})	
b_{22}	0.3132 (2.035×10^{-2})	0.2149 (1.460×10^{-2})
κ_{11}	7.6867 (0.6428)	7.6721 (0.6427)
κ_{21}	3.2291 (0.6037)	
κ_{22}	2.7981 (0.401)	3.1969 (0.2659)

correlations within humidity and temperature and cross-correlation between them are given in Figure 5 (solid lines). From Figure 5 we can get that the cross-correlation between the temperature and humidity at the same location is $\gamma_{ht} = 0.6613$, which is slightly higher than the previous model. From these results we notice that the cross-correlation between temperature and humidity are relatively high and indeed needed to be considered.

We define the correlation range as the correlation near 0.1 at distance ρ. The results of the correlation ranges for the above mentioned two models are given in Table 4 and Table 5. From these two tables we can notice that the correlation range for the temperature has been increased when we set $\boldsymbol{x}_1(\boldsymbol{s})$ as the humidity. The same conclusion can be drawn for the range of the cross-correlation between the humidity and the temperature with $\boldsymbol{x}_1(\boldsymbol{s})$ as humidity. The correlation range for the humidity, however, has been decreased when we set $\boldsymbol{x}_1(\boldsymbol{s})$ as the humidity. These differences indicates that we might need to investigate how to set the order of the fields in real-world applications when we used the triangular system of SPDEs (6). We return to this issue in Section 5.4.

As pointed out in Section 3.4, we treat the coefficients for the covariates and the yearly effects of temperature and humidity as parts of the latent

Table 3: Estimated hyper-parameters for bivariate model and for univariate model with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature

	Bivariate	Univariate
b_{11}	$0.1711\ (1.787 \times 10^{-2})$	$0.2149\ (1.460 \times 10^{-2})$
b_{21}	$-0.2230\ (2.875 \times 10^{-2})$	
b_{22}	$0.0198\ (2.843 \times 10^{-3})$	$0.0104\ (8.134 \times 10^{-4})$
κ_{11}	3.9537 (0.4118)	3.1969 (0.2659)
κ_{21}	2.5362 (0.5917)	
κ_{22}	5.6358 (0.7176)	7.6721 (0.6427)

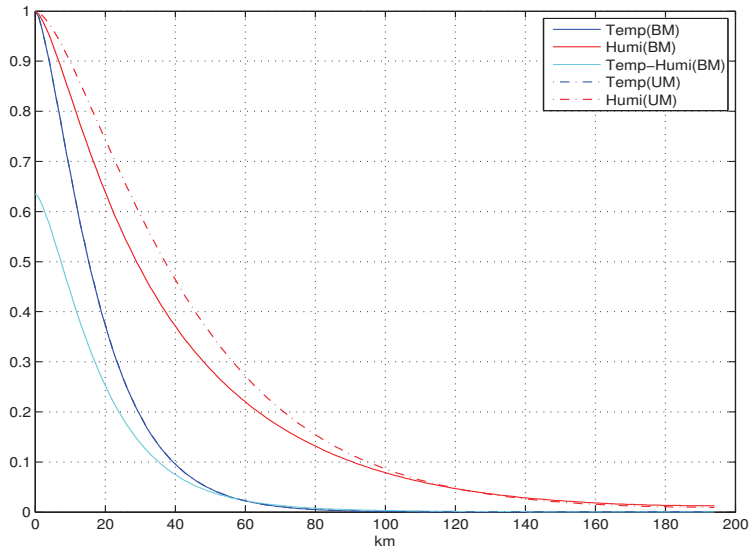


Figure 4: Correlations within temperature and humidity and cross-correlation between temperature and humidity for bivariate model together with the correlations within temperature and humidity for the univariate model when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. "BM" means Bivariate model. "UM" means univariate model. "Temp" means correlation within temperature. "Humi" means correlation within humidity. "Temp-Humi" means the cross-correlation between the temperature and humidity.

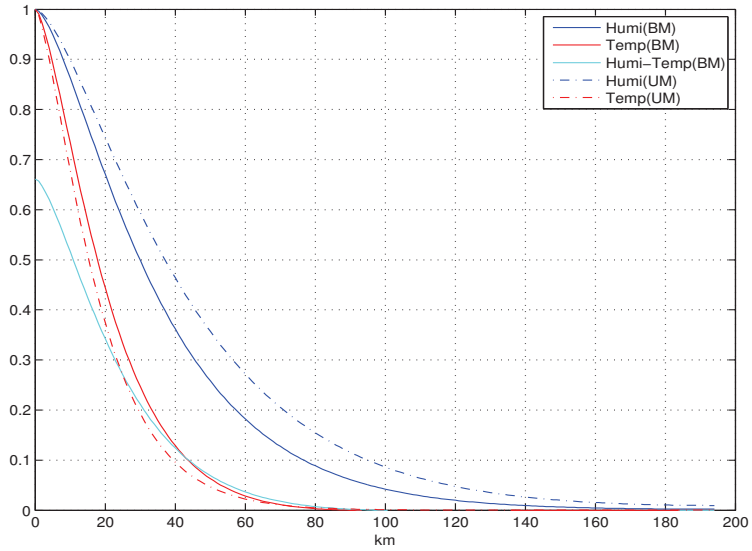


Figure 5: Correlations within temperature and humidity and cross-correlation between temperature and humidity for bivariate model together with the correlations within temperature and humidity for the univariate model when we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.. "BM" means Bivariate model. "UM" means univariate model. "Temp" means correlation within temperature. "Humi" means correlation within humidity. "Humi-Temp" means the cross-correlation between the temperature and humidity.

Table 4: Correlation ranges for bivariate model and univariate model with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

	ρ_t	ρ_h	ρ_{th}
Bivariate Model	39.4km	90.7km	35.2km
Univariate Model	39.4km	94.9km	

Table 5: Correlation ranges for bivariate model and univariate model with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

	ρ_t	ρ_h	ρ_{ht}
Bivariate Model	43.7km	76.7km	43.8km
Univariate Model	39.4km	94.9km	

field $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{\beta})^{\mathrm{T}}$. It can be shown that

$$\begin{aligned} \pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{z}, \boldsymbol{y}|\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{C})\boldsymbol{z} - 2\boldsymbol{z}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{y}\right]\right), \end{aligned} \tag{14}$$

and

$$\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_c(\boldsymbol{\theta}), \boldsymbol{Q}_c(\boldsymbol{\theta})\right), \tag{15}$$

with $\boldsymbol{\mu}_c$, $\boldsymbol{Q}_c$ and $\boldsymbol{C}$ given in Section 3.4. From Equation (14) we can get the estimates for the yearly effects and for the coefficients of the covariates. When we set $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity, the estimates for the yearly effects are given in Table 6 with the standard deviations given in the brackets. Table 6 shows that the yearly effects are quite different. This explains the high temperature in 2007 but low temperature in 2010. The estimates of the coefficients of the covariates are given in Table 7. We can notice that the two covariates give negative contribution to both fields. When we set $\boldsymbol{x}_1$ as the humidity and $\boldsymbol{x}_2$ as the temperature, the corresponding results for the yearly effects and the coefficients for the covariates are given in Table 8 and Table 9. Similar conclusions can be drawn from Table 8 and Table 9 as from Table 6 and Table 7. These results agree with the empirical results, and we summarize as follows.

- The higher the elevation, the lower the temperature;
- The higher the elevation, the lower the humidity.
- The longer the distance to ocean, the lower the temperature;
- The longer the distance to ocean, the lower the humidity;

Table 6: Estimated yearly effects for different years with bivariate model and univariate model with $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity

Model	Parameter	2007	2008	2009	2010	2011
Bivariate	β_{10}	8.808	−0.025	8.566	−5.981	0.894
		(0.546)	(0.541)	(0.532)	(0.525)	(0.519)
	β_{20}	3.037	1.817	2.859	1.153	1.665
		(0.112)	(0.112)	(0.111)	(0.111)	(0.110)
Univariate	β_{10}	8.794	−0.045	8.537	−6.004	0.874
		(0.547)	(0.542)	(0.532)	(0.526)	(0.519)
	β_{20}	3.081	1.844	2.899	1.171	1.698
		(0.118)	(0.118)	(0.117)	(0.117)	(0.116)

Table 7: Estimated coefficients for covariates and standard deviations with $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity

Model	Parameter	Estimate	Std. dev.
Bivariate	β_{11}	−6.825	0.683
	β_{12}	−9.859	0.721
	β_{21}	−0.199	0.092
	β_{22}	−1.460	0.128
Univariate	β_{11}	−6.832	0.683
	β_{12}	−9.815	0.721
	β_{21}	−0.493	0.107
	β_{22}	−1.438	0.138

Table 8: Estimated yearly effects for different years with bivariate model and univariate model with $\boldsymbol{x}_1$ as humidity and $\boldsymbol{x}_2$ as temperature

Model	Parameter	2007	2008	2009	2010	2011
Bivariate	β_{10}	3.031	1.819	2.852	1.158	1.663
		(0.101)	(0.100)	(0.099)	(0.099)	(0.098)
	β_{20}	8.843	0.048	8.634	−5.891	0.948
		(0.586)	(0.581)	(0.572)	(0.565)	(0.558)
Univariate	β_{10}	3.081	1.844	2.899	1.171	1.698
		(0.118)	(0.118)	(0.117)	(0.117)	(0.116)
	β_{20}	8.794	−0.045	8.537	−6.004	0.874
		(0.547)	(0.542)	(0.532)	(0.526)	(0.519)

Table 9: Estimated coefficients for the covariates and standard deviations with $\boldsymbol{x}_1$ as humidity and $\boldsymbol{x}_2$ as temperature

Model	Parameter	Estimate	Std. dev.
Bivariate	β_{11}	−0.212	0.093
	β_{12}	−1.438	0.138
	β_{21}	−6.467	0.656
	β_{22}	−10.108	0.764
Univariate	β_{11}	−0.493	0.107
	β_{12}	−1.438	0.138
	β_{21}	−6.832	0.683
	β_{22}	−9.815	0.721

5.3 Reconstruction of fields

With the estimates given in Section 5.2, we can reconstruct the fixed effects $\{\boldsymbol{v}_j; j = 1, 2, \ldots, 5\}$, the spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ and the posterior mean of temperature and humidity $\boldsymbol{v}_j + \boldsymbol{\xi}_j$ for different years j. We show the fixed effect $\boldsymbol{v}_j$ for year 2008 in Figure 6(a) and Figure 6(b) for the bivariate model with 1km by 1km resolution when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. Figure 6(c) and Figure 6(d) give the corresponding results for the univariate model. The way for reconstructing the fixed effects with 1km by 1km resolution is that we first estimate the relevant parameters with the lower resolution model, and then plug in the estimates into the 1km by 1km resolution model. The fixed effects for other years are omitted since they are just shifted versions of each other because they just have different yearly effects but share the same coefficients from the two covariates. The corresponding results for the fixed effects with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature are given in Figure 9(a) and Figure 9(b).

Using the same approach as the fixed effects, we can get the spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ on high resolution with the estimates given in Table 2. Figure 7(a) and Figure 7(b) illustrate the spatial effects with 1km by 1km resolution for the bivariate model and Figure 7(c) and Figure 7(d) for the univariate model, respectively. The corresponding results for spatial effects with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature are given in Figure 9(c) and Figure 9(d). As discussed in Section 3.4, the spatial effects are constructed by using the SPDEs or the systems of SPDEs with the same hyper-parameters. In other words, they are just different realizations of the same latent fields. We only show the results for year 2008 but emphasize that they are different from year to year.

With the fixed effects $\{\boldsymbol{v}_j; j = 1, 2, \ldots, 5\}$ and spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ given in Figure 6 and Figure 7, we can get the posterior mean of temperature and humidity $\{\boldsymbol{\eta}_j = \boldsymbol{v}_j + \boldsymbol{\xi}_j; j = 1, 2 \ldots, 5\}$, and Figure 8(a) and Figure 8(b) illustrate the results for year 2008 for the bivariate model and Figure 8(c) and Figure 8(d) for the univariate model, respectively, with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. When we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, we can get the posterior mean of temperature and humidity shown in Figure 10(a) and Figure 10(b) for bivariate model. We notice that the fixed effects, the spatial effects and

the posterior mean for temperature and humidity in these two different bivariate models are not the same but quite similar.

5.4 Predictive performance

In this section the predictive performance for both the bivariate models and the univariate model with different settings, using the cross validation schemes given in Section 4.2, are discussed for both temperature and humidity. Using the left-out observations in different settings together with the scoring rules discussed in Section 4.1, we compare the values of scoring rules MAE, MSE and CRPS.

Table 10 illustrates the results for the scoring rules when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. In this table and the tables thereafter, "UM" means the results are from the univariate model. "BM (Setting 3)" means the results are from the bivariate model with Setting 3, i.e., there are 20 locations left-out both for temperature and humidity in each year. " BM" means the results are from bivariate model with Setting 1 or Setting 2, i.e., there are 20 locations left-out from only temperature or humidity in each year. "T" and "H" denote the temperature and humidity, respectively. From Table 10 we can notice that the bivariate model performances uniformly better than the univariate model for humidity. We can also notice that "BM" outperforms the "BM(Setting 3)" uniformly, which means the observations from another field at the same locations improve the prediction accuracy. The scoring rules with "BM" is also uniformly better than the"UM" for temperature. However, the MSE and CRPS of temperature for bivariate model with Setting 3 is slightly higher than the corresponding results from univariate model.

Table 11 shows the predictive performance with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature. We can notice that the bivariate model with "BM" performs uniformly better than the univariate model for temperature. But the bivariate model with Setting 3 perform a little worse than the univariate model for temperature. We can also notice that the bivariate model with Setting 3 performs uniformly better than the univariate model for the humidity. However, the bivariate model with "BM" performs the worst for the humidity.

Deeper analysis releases the reasons. We find that there is due to some "outliers" in the temperature observations in year 2009. Figure 11

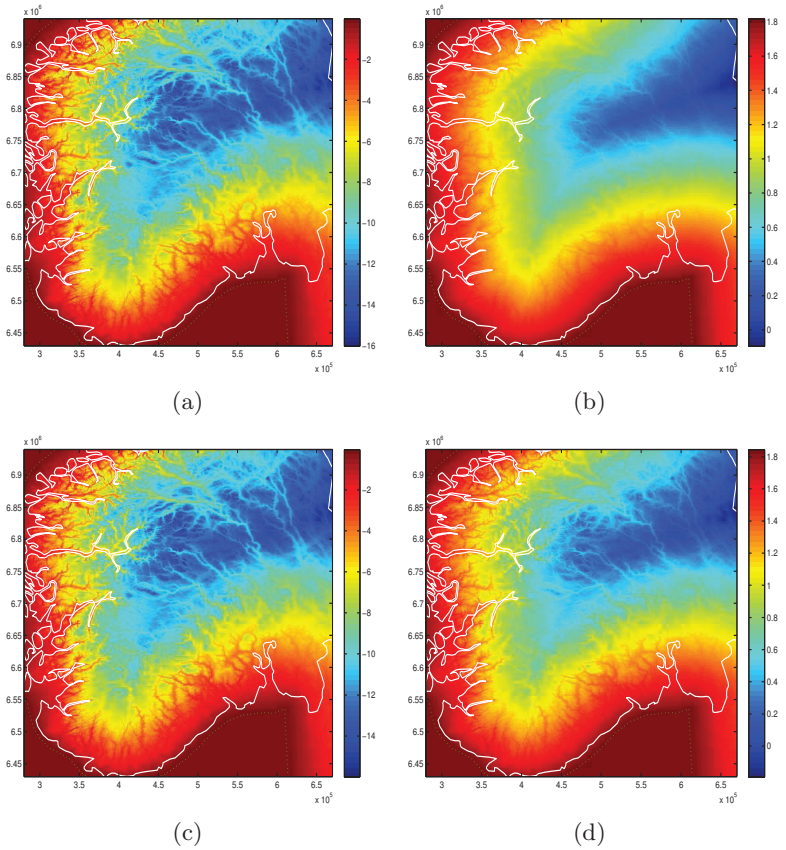


Figure 6: Fixed effects for bivariate model (a) - (b) and for the univariate model (c) - (d) of temperature and humidity in 2008 with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

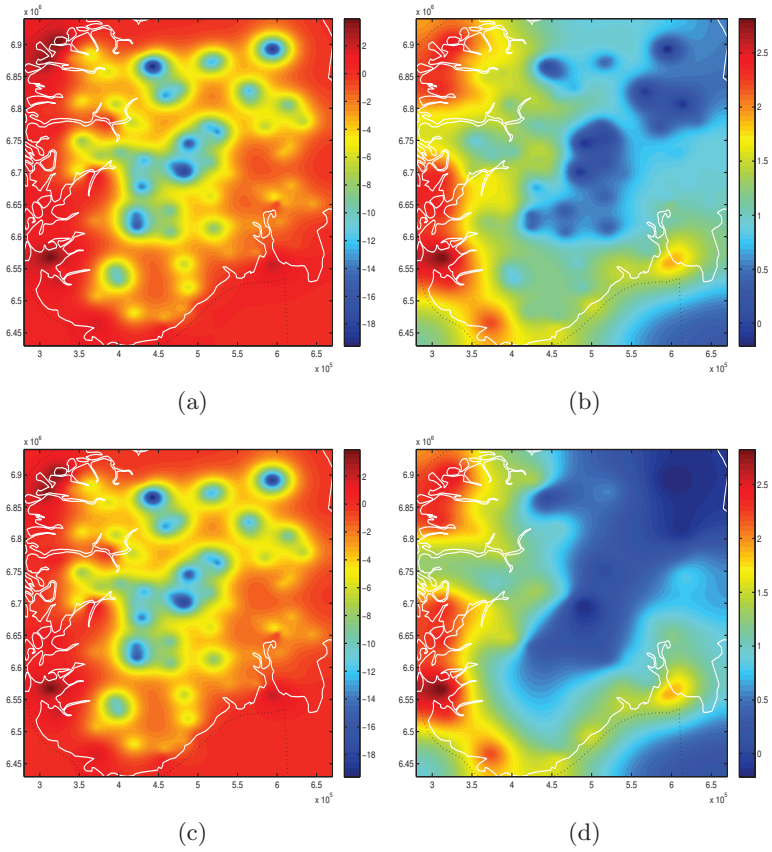


Figure 7: Spatial effects for bivariate model (a) - (b) and for univariate model (c) - (d) of temperature and humidity in 2008 with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

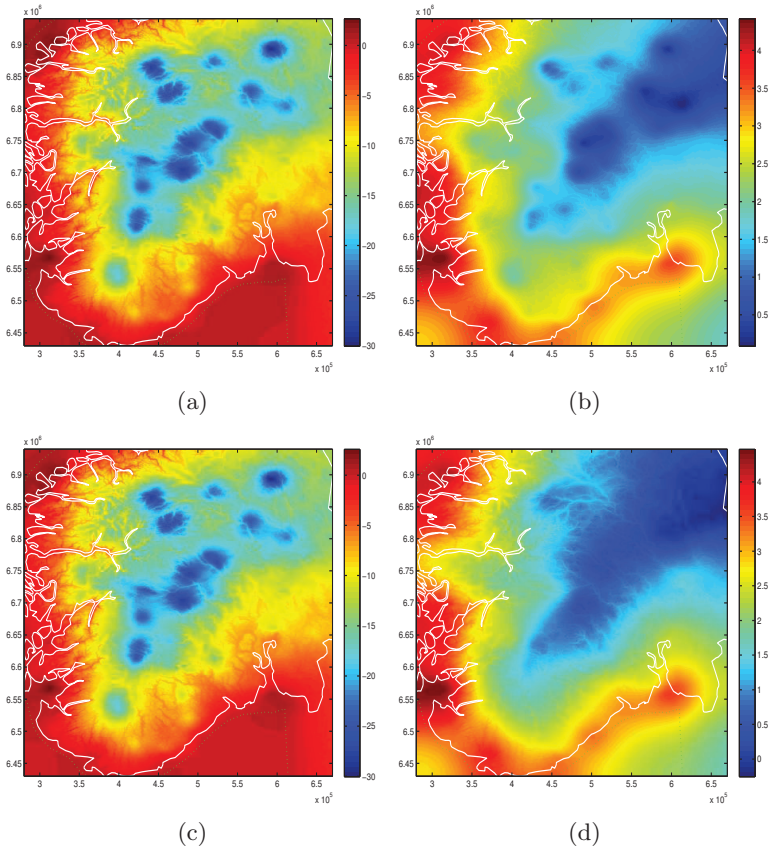


Figure 8: Posterior mean of temperature and humidity in 2008 for bivariate model (a) - (b) ad for univariate model (c) - (d) with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

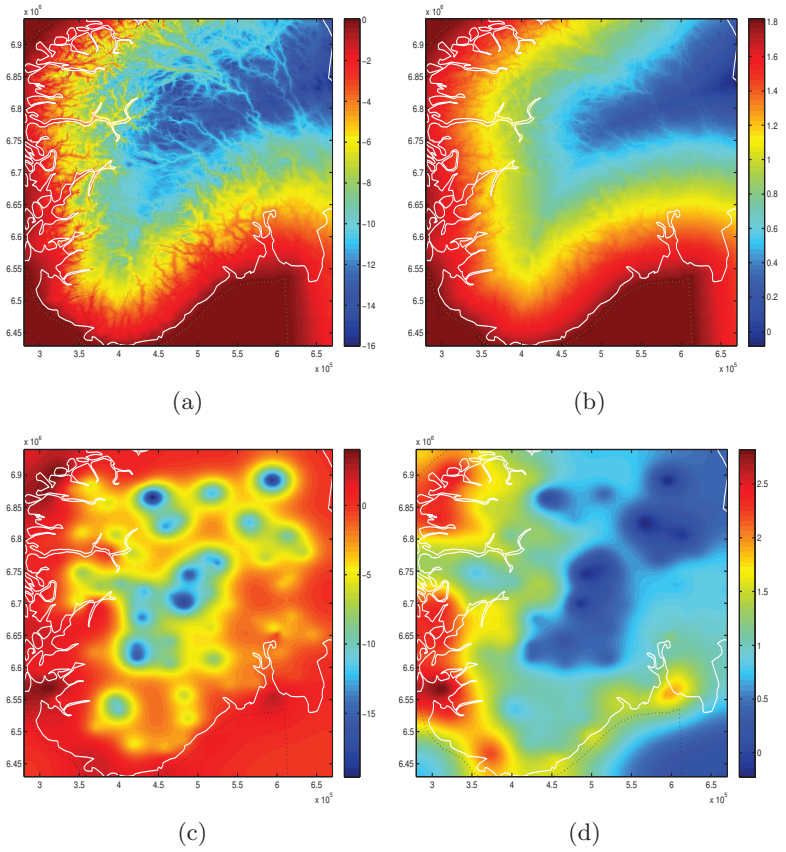
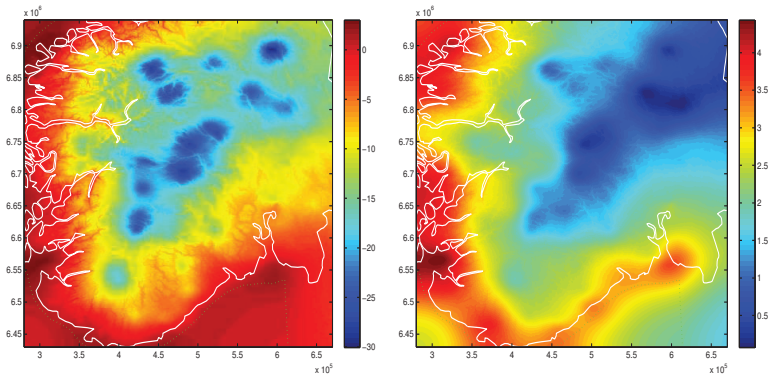


Figure 9: Fixed effects (a) - (b) and spatial effects (c) - (d) for bivariate model in 2008 with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.



(a) (b)

Figure 10: Posterior mean of temperature (a) and humidity (b) in 2008 for bivariate model with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

has illustrated the left-out observations in year 2009. We can see that there are 5 locations with very high temperature but rather low humidity. The bivariate model has difficulties at these locations not only for predicting the temperature itself but also for humidity since the information from temperature leads the prediction of humidity in the wrong direction. The results of the scoring rules without year 2009 are given in Table 12 and 13. From these two tables, we can notice that the bivariate model performs uniformly better than the univariate model. Furthermore the bivariate model with "BM" performs uniformly better than the bivariate model with Setting 3. In addition, from Table 12 and Table 13, we can notice that we get better predictive performances when we setting the temperature as the second field when we need to predict temperature. This is also true with humidity.

The posterior standard deviations for temperature and humidity in 2011 with the bivariate model and the univariate model are presented in Figure 12(a) - Figure 12(b) and in Figure 12(c) - Figure 12(d), respectively, when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. With the bivariate model, we can notice that the posterior standard derivations for locations where we have the temperature observations but not the hu-

Table 10: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7485	0.2039	6.1102	0.0722	1.3524	0.1487
BM(Setting 3)	1.7463	0.1918	6.1164	0.0639	1.3539	0.1396
BM	1.5862	0.1501	4.3089	0.0352	1.1464	0.1060

Table 11: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7485	0.2039	6.1102	0.0722	1.3524	0.1487
BM(Setting 3)	1.7949	0.1929	6.2014	0.0638	1.3665	0.1402
BM	1.5564	0.2215	4.4149	0.0845	1.1579	0.1579

midity observations are lower than the corresponding univariate models. Same conclusion can be drawn from Figure 13(a) and Figure 13(b), when we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature. The results for other years are similar and omitted here.

From all these results, we can notice that the bivariate models give better prediction accuracy than the univariate model. When the observations in one field is presented, it does not only improve the prediction accuracy but also have lower posterior standard deviations. We can also conclude that the order of the field has some influence for the prediction. The generally suggestion is that if we want to predict humidity or temperature, we should set it as the second field when we have enough time and computational resources. If time or computational resources is limited, then we do not need to consider about the order of fields, since the bivariate model can provide satisfiable results with both the orders.

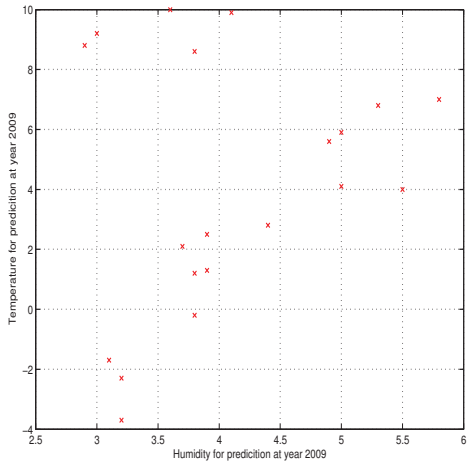


Figure 11: The observations of humidity and temperature for predictions at year 2009.

Table 12: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity without year 2009

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7842	0.2095	6.7240	0.0791	1.4031	0.1553
BM(Setting 3)	1.7671	0.2055	6.6736	0.0737	1.3985	0.1481
BM	1.4693	0.1426	3.7366	0.0318	1.0677	0.1006

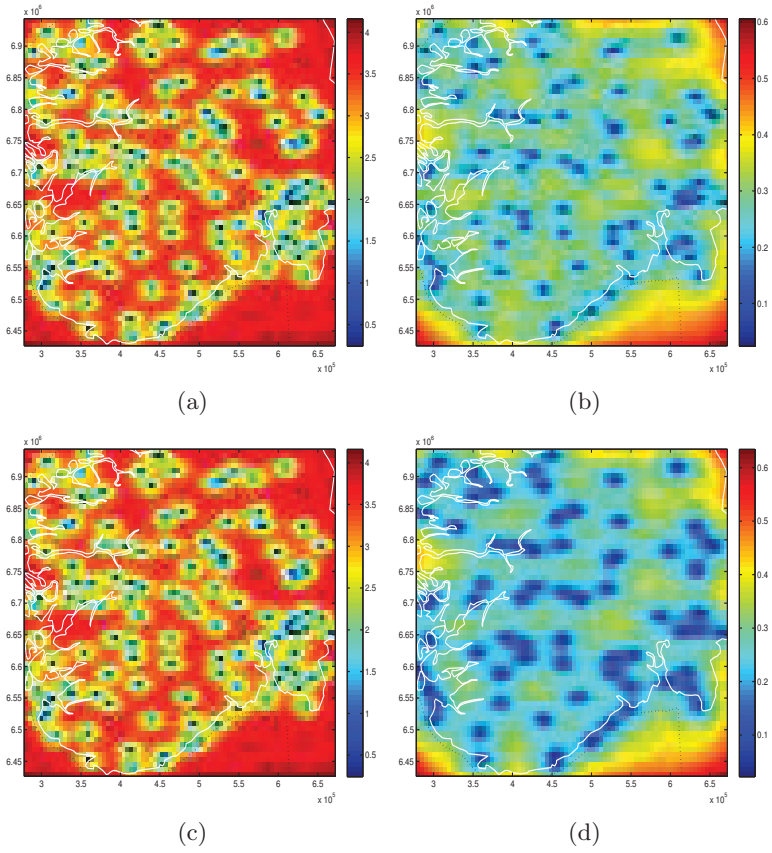
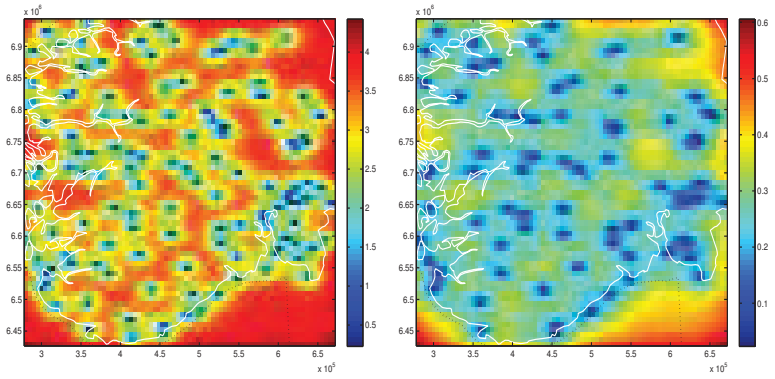


Figure 12: Posterior standard deviation for temperature and humidity in 2011 by bivariate model (a) - (b) and by univariate model (c) - (d). We set the first field $x_1(\boldsymbol{s})$ as temperature and the second field $x_2(\boldsymbol{s})$ as humidity.

Table 13: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature without year 2009

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7842	0.2095	6.7240	0.0791	1.4031	0.1553
BM(Setting 3)	1.7961	0.2010	6.5859	0.0713	1.3940	0.1467
BM	1.3837	0.1940	3.4967	0.0596	1.0513	0.1365



(a) (b)

Figure 13: The posterior standard deviation for temperature (a) and humidity (b) in 2011 by bivariate model. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

6 Discussion and conclusion

We have modelled temperature and humidity in southern Norway based on the observations on 7th of December from 2007 to 2011. Three different models are used in this paper: two bivariate models for modelling them jointly and one univariate model for modelling them independently. The system of SPDEs approach proposed by Hu et al. (2012b) is chosen for constructing bivariate GRFs and the SPDE approach given by Lindgren et al. (2011) is chosen for constructing univariate GRFs. Three different settings are chosen in order to compare the predictive performance between the bivariate model and the univariate model with different settings. Computational efficiency is obtained by using the link between the GRFs and GMRFs. All our models are constructed with GRFs theoretically and all computations are conducted with GMRFs. The results illustrate that there is a strong positive correlation between temperature and humidity. The other results agree with the physical and empirical knowledge. We conclude that using a bivariate GRF to model temperature and humidity jointly is superior to model them independently using univariate GRFs, not only in term of prediction accuracy, but also in term of quantifying prediction uncertainty.

The results also illustrate that the order of fields seems relevant from the prediction point of view when we use a triangular system of SPDEs for constructing a bivariate field. However, since the results from both orders are satisfiable, we do not need to consider it if the computational resources or time is limited. There might be some other covariates, such as wind speed and solar radiation which needs to be included in our analysis. However, this is beyond the scope of this paper and we leave it for future research.

References

T.V. Apanasovich, M.G. Genton, and Y. Sun. A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193, 2012.

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman & Hall, 2004. ISBN 158488410X.

K.J. Bathe. *Finite element method*. Wiley Online Library, 2008.

D. Bolin and F. Lindgren. Wavelet markov models as efficient alternatives to tapering and convolution fields. Technical report, Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University, 2009.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550, 2011.

G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994.

P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

N. Cressie and C.K. Wikle. *Statistics for spatio-temporal data*, volume 465. Wiley, 2011.

N.A.C. Cressie. *Statistics for spatial data*, volume 298. Wiley-Interscience, 1993.

P.J. Diggle and P.J. Ribeiro Jr. *Model-based Geostatistics*. Springer, 2006.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011.

A.E. Gelfand, A.M. Schmidt, S. Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp. *Handbook of spatial statistics*. CRC Press, 2010.

T. Gneiting, A.E. Raftery, A.H. Westveld III, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

X. Hu, F. Lindgren, D.P. Simpson, and H. Rue. Multivariate gaussian random fields with oscillating covariance functions using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012a.

X. Hu, D.P. Simpson, F. Lindgren, and H. Rue. Multivariate gaussian random fields using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012b.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Using stochastic partial differential equation models for spatial reconstruction of annual precipitation. *submitted*, 2013.

L.W. Konigsberg and S.D. Ousley. Multivariate quantitative genetics of anthropometric traits from the boas data. *Human biology*, 81(5/6):579–594, 2009.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

K. Mcguigan. Studying phenotypic evolution using multivariate quantitative genetics. *Molecular ecology*, 15(4):883–896, 2006.

B.J. Reich and M. Fuentes. A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264, 2007.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868.

S.R. Sain and N. Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007.

RM Sakia. The box-cox transformation technique: a review. *The statistician*, pages 169–178, 1992.

A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108(D24):8783, 2003.

D. Simpson, F. Lindgren, and H. Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 2010.

M.L. Stein. *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag, 1999. ISBN 0387986294.

Z. Toth, O. Talagrand, G. Candille, and Y. Zhu. 7 probability and ensemble forecasts. 2003.

O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, and JZ Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-heinemann, 2005.

Paper IV

Specifying Gaussian Markov Random Fields with Incomplete Orthogonal Factorization using Givens Rotations

Xiangping Hu, Daniel Simpson and Håvard Rue

Technical Report, 2013

Specifying Gaussian Markov Random Fields with Incomplete Orthogonal Factorization using Givens Rotations

Xiangping Hu*[1], Daniel Simpson[1], and Håvard Rue[1]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Abstract

In this paper an approach for finding a sparse incomplete Cholesky factor through an incomplete orthogonal factorization with Givens rotations is discussed and applied to Gaussian Markov random fields (GMRFs). The incomplete Cholesky factor obtained from the incomplete orthogonal factorization is usually sparser than the commonly used Cholesky factor obtained through the standard Cholesky factorization. On the computational side, this approach can provide a sparser Cholesky factor, which gives a computationally more efficient representation of GMRFs. On the theoretical side, this approach is stable and robust and always returns a sparse Cholesky factor. Since this approach applies both to square matrices and to rectangle matrices, it works well not only on precision matrices for GMRFs but also when the GMRFs are conditioned on a subset of the variables or on observed data. Some common structures for precision matrices are tested in order to illustrate the usefulness of the approach. One drawback to this approach is that the incomplete orthogonal factorization is usually slower than the standard Cholesky factorization implemented in standard libraries and currently it can be slower to build the sparse Cholesky factor.

*Corresponding author. Email: `Xiangping.Hu@math.ntnu.no`

Keywords: Gaussian Markov random field; Incomplete orthogonal factorization; Upper triangular matrix, Givens Rotation; Sparse matrix; Precision matrix

1 Introduction

Gaussian Markov random fields(GMRFs) are useful models in spatial statistics due to the Gaussian properties together with Markovian structures. They can also be formulated as conditional auto-regressions (CARs) models (Rue and Held, 2005). GMRFs have applications in many areas, such as spatial statistics, time-series models, analysis of longitudinal survival data, image analysis and geostatistics. See Rue and Held (2005, Chapter 1) for more information and literature on how the GMRFs can be applied in different areas. From an analytical point of view GMRFs have good properties and can be specified through mean values $\boldsymbol{\mu}$ and covariance matrices $\boldsymbol{\Sigma}$. While from a computational point of view GMRFs can conveniently specified through precision matrices $\boldsymbol{Q}$ (the inverse of the covariance matrices $\boldsymbol{\Sigma}$), which are usually sparse matrices. The numerical algorithms for sparse matrices can be exploited for calculations with the sparse precision matrices and hence fast statistical inference is possible (Rue, 2001). The numerical algorithms for sparse matrices can be applied to achieve fast simulation of the fields and evaluation the densities (mostly, log-densities) of GMRFs and GMRFs with conditioning on subset of variables or linear constraints. See Rue and Held (2005, Chapter 2) for further details. These algorithms can also be used to calculate the marginal variances (Rue, 2005), and they can be extended to non-Gaussian cases (Rue et al., 2004).

Precision matrices $\boldsymbol{Q}$ are commonly used to specify GMRFs. This approach is natural due to the sparsity patterns of the precision matrices in Markovian models. In many situations the Cholesky factors are required and are crucial for simulation and inferences with GMRFs, and the Cholesky factors are normally obtained with Cholesky factorization routines in standard libraries. See Rue and Held (2005, Chapter 2) for different simulation algorithms for GMRFs using Cholesky factors. In order to get an even sparser Cholesky factor, with the purposes of saving computational resources, Wist and Rue (2006) showed that the Cholesky

factor from an incomplete Cholesky factorization can be much sparser than the Cholesky factor from the regular Cholesky factorization. However, they provided theoretical and empirical evidence showing that the representation of sparser Cholesky factor was fragile when conditioning the GMRF on a subset of the variables or on observed data. It means that the sparsity patterns of the sparser Cholesky factors are destroyed when some constraints or observed data are introduced and the computational cost increases. Additionally, the sparser Cholesky factor from the incomplete Cholesky factorization is only valid for a specific precision matrix. Their approach is illustrated in Figure 1 with Routine 1.

In this paper a different approach is chosen to solve the problem presented by Wist and Rue (2006) . The main idea is given in the Figure 1 with Routine 2. In this approach one rectangular matrix $\boldsymbol{A}$ is formulated,

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{L}_1^{\mathrm{T}} \\ \boldsymbol{L}_2^{\mathrm{T}} \end{pmatrix}. \tag{1}$$

It consists of the Cholesky factor $\boldsymbol{L}_1^{\mathrm{T}}$ from the precision matrix $\boldsymbol{Q}_1$ of a given GMRF and the Cholesky factor $\boldsymbol{L}_2^{\mathrm{T}}$ of the matrix $\boldsymbol{Q}_2$. The matrix $\boldsymbol{Q}_2$ can be the additional effect when the GMRF is conditioned on observed data or on a subset of the variables. Both $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are lower triangular matrices.

An incomplete orthogonal factorization is then used to factorize the matrix $\boldsymbol{A}$ in Equation (1) to find the sparse Cholesky factor for specifying the GMRF. It is shown that by using this approach an upper triangular matrix $\boldsymbol{R}$ which is sparser than the standard Cholesky factor is obtained. Furthermore, this approach is applicable when the GMRF is conditioned on a subset of the variables or on observed data. Since the upper triangular matrix $\boldsymbol{R}$ is sparser in structure than the common Cholesky factor, it is better for applications.

The rest of this paper is organized as follows. In Section 2 some basic theory on GMRFs, sparsity patterns for precision matrices and Cholesky factors of GMRFs are presented. Some basic theories on the orthogonal factorization and the incomplete orthogonal factorization are also introduced in this section. In Section 3 the algorithm for obtaining the sparse Cholesky factor from the incomplete orthogonal factorization is in-

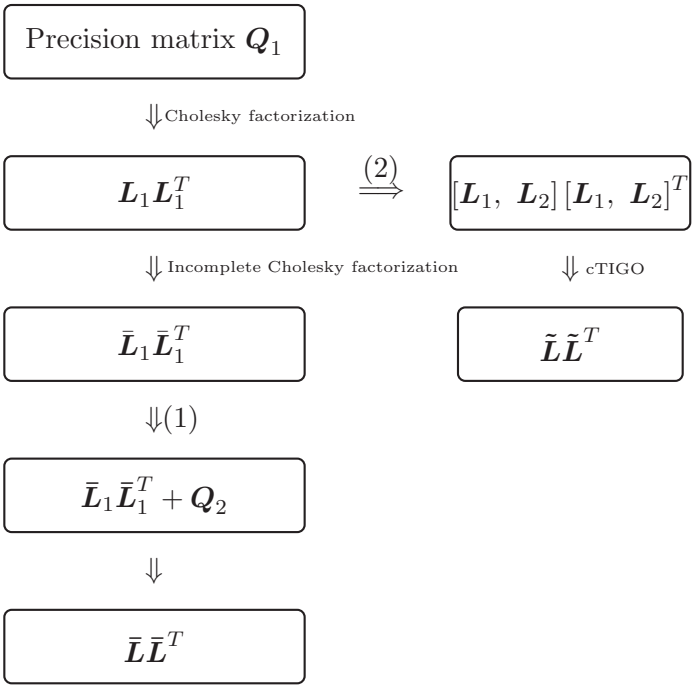


Figure 1: Diagram for the algorithm for finding sparser Cholesky factor by incomplete Cholesky factorization used by Wist and Rue (2006) (Routine 1) and the algorithm used in this paper (Routine 2).

troduced. A small example is given in order to illustrate how the algorithm works when the GMRFs are conditioned on a subset of the variables or on observed data. Results for different structures on the precision matrices are given in Section 4. Conclusion and general discussion in Section 5 ends the paper.

2 Background and Preliminaries

2.1 Basic theory on GMRFs

A random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ is called a GMRF if it is Gaussian distributed and processes a Markov property. The structure of a GMRF is usually presented by a labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertexes $\{1, 2, \ldots, n\}$ and $\mathcal{E}$ is the set of edges. The graph $\mathcal{G}$ satisfies the properties that no edge between node i and node j if and only if $x_i \perp x_j | \boldsymbol{x}_{-ij}$ (Rue and Held, 2005), where $\{\boldsymbol{x}_{-ij}; i, j = 1, 2, \ldots, n\}$ denotes $\boldsymbol{x}_{-\{i,j\}}$. If the random vector $\boldsymbol{x}$ has a mean $\boldsymbol{\mu}$ and a precision matrix $\boldsymbol{Q}_1 > 0$, the probability density of the vector $\boldsymbol{x}$ is

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{Q}_1(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}_1(\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\mu})\right), \quad (2)$$

with the property

$$Q_{ij} \neq 0 \Longleftrightarrow \{i, j\} \in \mathcal{E} \text{ for all } i \neq j.$$

The notation $\boldsymbol{Q}_1 > 0$ means that $\boldsymbol{Q}_1$ is a symmetric positive definite matrix. $\boldsymbol{\theta}$ denotes the parameters in the precision matrix. This implies that any vector with a Gaussian distribution and a symmetric positive definite covariance matrix is a Gaussian random field (GRF), and GMRFs are GRFs with Markov properties. The graph $\mathcal{G}$ determines the nonzero pattern of $\boldsymbol{Q}_1$. If $\mathcal{G}$ is fully connected, then $\boldsymbol{Q}_1$ is a complete dense matrix. A useful property of GMRF is that we can know whether x_i and x_j are conditionally independently or not directly from the precision matrix $\boldsymbol{Q}_1$ and the graph $\mathcal{G}$. Values of mean $\boldsymbol{\mu}$ do not have any influence on the pairwise conditional independence properties of the GMRFs, and hence we set $\boldsymbol{\mu} = \boldsymbol{0}$ in the following sections unless otherwise specified. The diagonal

elements in the precision matrix $Q_{ii}\,(i = 1, 2, \ldots, n)$ are the conditional precisions of x_i given all the other nodes $\boldsymbol{x}_{-i}$. The off-diagonal elements $Q_{ij}\,(i, j = 1, 2, \ldots, n, i \neq j)$ can provide information about the correlations between x_i and x_j given on the nodes $\boldsymbol{x}_{-ij}$. These are the main differences in the interpretation between the precision matrix $\boldsymbol{Q}_1$ and the covariance matrix $\boldsymbol{\Sigma}_1$. The covariance matrix $\boldsymbol{\Sigma}_1$ contains the marginal variance of x_i and the marginal correlation between x_i and x_j. However, with the precision matrix the marginal properties are not directly available (Rue and Held, 2005).

Since $\boldsymbol{Q}_1$ is symmetric positive definite, there is a unique Cholesky factor $\boldsymbol{L}_1$ where $\boldsymbol{L}_1$ is a lower triangle matrix satisfying $\boldsymbol{Q}_1 = \boldsymbol{L}_1\boldsymbol{L}_1^{\mathrm{T}}$. If we want to sample from the GMRF $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}_1^{-1})$, the Cholesky factor $\boldsymbol{L}_1$ is commonly used. One algorithm for sampling GMRFs is given in Section 4.3. More algorithms for sampling GMRFs with different specifications are also available. See Rue and Held (2005, Chapter 2) for a detailed discussion on these algorithms. Rue and Held (2005) showed how to check the sparsity pattern of the Cholesky factor of a GMRF. Define

$$F(i, j) = i + 1, i + 2, \ldots, j - 1, j + 1, \ldots, n, \tag{3}$$

which is the future of i except j. Then

$$x_i \perp x_j \mid \boldsymbol{x}_{F(i,j)} \Longleftrightarrow L_{ji} = 0, \tag{4}$$

and $F(i, j)$ is called a separating subset of i and j. If $i \sim j$ denotes that i and j are neighbors, then $F(i, j)$ cannot be a separating subset for i and j whenever $i \sim j$. Further, the Cholesky factor of the precision matrix of a GMRF is always equally dense or denser than the lower triangle part of $\boldsymbol{Q}_1$.

In many situations there are more nonzero elements in $\boldsymbol{L}_1$ than in the lower triangular part of $\boldsymbol{Q}_1$. Denote $n_{\boldsymbol{L}_1}$ and $n_{\boldsymbol{Q}_1}$ the numbers of nonzero elements in the Cholesky factor $\boldsymbol{L}_1$ and the lower triangular part of precision matrix $\boldsymbol{Q}_1$, respectively. The difference $n_f = n_{\boldsymbol{L}_1} - n_{\boldsymbol{Q}_1}$ is called the fill-in. The ideal case is $n_f = 0$ or $n_{\boldsymbol{L}_1} = n_{\boldsymbol{Q}_1}$, but commonly $n_{\boldsymbol{L}_1} > n_{\boldsymbol{Q}_1}$ or even $n_{\boldsymbol{L}_1} \gg n_{\boldsymbol{Q}_1}$. It is known that the fill-in n_f not only depends on the graph, but also on the order of the nodes in the graph (Rue and Held, 2005). Thus a re-ordering is usually needed before doing a Cholesky factorization. It is desirable to find an optimal or approximately

optimal ordering of the graph in order to make the Cholesky factor of $\boldsymbol{Q}_1$ sparser and to save computational resources, but this is not the focus in this paper. We refer to Rue and Held (2005) for more information on why it is desirable to do re-ordering of the graph of a GMRF.

2.2 Orthogonal factorization

With an $m \times n$ matrix $\boldsymbol{A}$, the orthogonal factorization of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{S} \cdot \boldsymbol{R}, \tag{5}$$

where $\boldsymbol{S} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix and $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ is an upper triangular matrix. We assume without loss of the generality that $m \geq n$. There exist many algorithms for orthogonal factorization, such as the standard Gram - Schmidt algorithm or the modified Gram–Schmidt (MGS) algorithm and the Householder orthogonal factorization. We refer to Saad (2003) and Björck (1996) for more algorithms. If $\boldsymbol{A}$ has full column rank, then the first n columns of $\boldsymbol{S}$ forms an orthonormal basis of $\text{ran}(\boldsymbol{A})$, where $\text{ran}(\boldsymbol{A})$ denotes the range of $\boldsymbol{A}$

$$\text{ran}(\boldsymbol{A}) = \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \text{ for some } \boldsymbol{x} \in \mathbb{R}^n\}.$$

The orthogonal factorization is usually used to find an orthonormal basis for a matrix. The orthogonal factorization has many advantages and some of them are given in what follows.

1. It is numerically stable and robust both with a Householder orthogonal factorization and with a orthogonal factorization using Givens rotations. If the matrix $\boldsymbol{A}$ is non-singular, it always produces an orthogonal matrix $\boldsymbol{S}$ and an upper triangular matrix $\boldsymbol{R}$ which satisfy Equation (5);

2. It is easy to solve the linear system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ using the upper triangular matrix $\boldsymbol{R}$ since $\boldsymbol{S}$ is an orthogonal matrix;

3. The normal equation has the form $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ and the normal equation matrix is $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, where $\boldsymbol{A}^{\mathrm{T}}$ denotes the transpose of $\boldsymbol{A}$. Then the triangular matrix $\boldsymbol{R}$ is the Cholesky factor of the normal equations matrix.

2.3 Givens rotations

A Givens rotation $G(i, j, \theta) \in \mathbb{R}^{m \times n}$ is an identity matrix $\boldsymbol{I}$ except that

$$\begin{aligned} G_{ii} &= c, \quad & G_{ij} &= s, \\ G_{ji} &= -s, \quad & G_{jj} &= c. \end{aligned}$$

If $c = \cos(\theta)$ and $s = \sin(\theta)$, then $\boldsymbol{y} = G(i, j, \theta) \cdot \boldsymbol{x}$ rotates $\boldsymbol{x}$ clockwise in the (i, j)-plane with θ radians, which gives

$$y_l = \begin{cases} x_l, & \text{when } l \neq i, j, \\ cx_i + sx_j, & \text{when } l = i, \\ -sx_i + cx_j, & \text{when } l = j. \end{cases} \qquad (1 \leq l \leq m), \tag{6}$$

If we want to rotate $\boldsymbol{x}$ counterclockwise in the (i, j)-plane with θ radians, then we can set $c = \cos(\theta)$ and $s = -\sin(\theta)$. It is obvious from Equation (6) that if

$$s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}} \text{ and } c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}$$

then $y_j = 0$. So the Givens rotations can set the elements in $\boldsymbol{A}$ to zeros one at a time. This is useful when dealing with sparse matrices. At the same time, c and s are the only two values which we need for this algorithm. Givens rotations are suitable for structured least squares problems such as the problems at the heart of GMRFs.

2.4 Incomplete factorization algorithms

There are many algorithms for incomplete factorizations of matrices, such as the incomplete triangular factorization and the incomplete orthogonal factorization. These algorithms are commonly used in practical applications (Axelsson, 1996; Meijerink and van der Vorst, 1981; Saad, 1988). The incomplete factorizations usually have the form

$$\boldsymbol{A} = \boldsymbol{M}_1 \cdot \boldsymbol{M}_2 + \boldsymbol{E}, \tag{7}$$

where $\boldsymbol{E}$ is the error matrix, and $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are some well-structured matrices. The incomplete factorization algorithms are usually associated

with dropping strategies. A dropping strategy for an incomplete factorization specifies rules for when elements of the factors should be dropped. We returns to a detailed discussion on the dropping strategies in Section 3.5.

One of the commonly used incomplete factorization algorithms is the incomplete triangular factorization, and it is also called incomplete LU (ILU) factorization since $\boldsymbol{M}_1$ is a *lower* triangular matrix and $\boldsymbol{M}_2$ is an *upper* triangular matrix. This algorithm is usually applied to the square matrices, and it uses Gaussian elimination together with a predefined dropping strategy. Many incomplete orthogonal factorizations can be used both for square matrices and for rectangular matrices, and these algorithms usually use the modified Gram-Schmidt procedure together with some dropping strategies in order to return a sparse and generally non-orthogonal matrix $\boldsymbol{S}$ and a sparse upper triangular matrix $\boldsymbol{R}$. Wang et al. (1997) proved the existence and stability of the associated incomplete orthogonal factorization. Incomplete orthogonal factorization using Givens rotations was proposed by Bai et al. (2001). The main idea of the incomplete orthogonal factorization is to use the Givens rotations to zero-out the elements in the matrix one at a time. Some predefined dropping strategies are needed in order to achieve the sparsity pattern for the upper triangular matrix $\boldsymbol{R}$. This algorithm computes a sparse matrix $\boldsymbol{S}$, which is always an orthogonal matrix, together with a sparse upper triangular matrix $\boldsymbol{R}$. Since the matrix $\boldsymbol{S}$ is the product of the Givens rotations matrices, it is always an orthogonal matrix. The incomplete orthogonal factorization has the form

$$\boldsymbol{A} = \boldsymbol{S} \cdot \boldsymbol{R} + \boldsymbol{E}. \tag{8}$$

This method was originally described and implemented by Jennings and Ajiz (1984). Saad (1988) described this incomplete orthogonal factorization with the modified Gram–Schmidt process using some numerical dropping strategy. Another version of the incomplete orthogonal factorization is given by Bai et al. (2001) with Givens rotations. Bai et al. (2001) claimed that this incomplete algorithm inherited the good properties of the orthogonal factorization.

1. $\boldsymbol{R}$ is a sparse triangular matrix and $\boldsymbol{S}$ is an orthogonal matrix. Bai

et al. (2009) pointed out that the sparsity pattern of the upper-triangular part of $\boldsymbol{A}$ is inherited by the incomplete upper triangular matrix $\boldsymbol{R}$. They also pointed out that the number of nonzero elements in the upper triangular matrix $\boldsymbol{R}$ is less than the number of nonzero elements in the upper-triangular part of $\boldsymbol{A}$.

2. The error matrix $\boldsymbol{E} = \boldsymbol{A} - \boldsymbol{S} \cdot \boldsymbol{R}$ is "small" in some sense and the size of the errors can be controlled by the pre-defined threshold.

3. The triangular matrix $\boldsymbol{R}$ is non-singular whenever $\boldsymbol{A}$ is not singular. We can always obtain this triangular matrix in the same way as the orthogonal factorization and $\boldsymbol{R}$ will always be an incomplete Cholesky factor for the normal equation matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

4. Another merit of the incomplete orthogonal factorization with Givens rotations is that we do not need to form the corresponding normal matrices $\boldsymbol{S}$ since only the (c, s)-pair is needed in order to find the upper triangular matrix $\boldsymbol{R}$. More information about the Givens rotations and the (c, s)-pairs are given in Section 2.3

Papadopoulos et al. (2005) implemented different versions of the algorithm proposed by Bai et al. (2001). There are two main differences between these versions. The first one is the order in which elements in the matrix $\boldsymbol{A}$ are zeroed out, and the second one is the rules for dropping strategies. We refer to Bai et al. (2001) and Papadopoulos et al. (2005) for more information about this algorithm and implementations. There are also more variations for incomplete orthogonal factorization using Givens rotations, such as Bai and Yin (2009) and Bai et al. (2009). Bai and Yin (2009) proposed some modified incomplete orthogonal factorization methods and these algorithms have special storage and sparsity-preserving techniques. Bai and Yin (2009) showed a way to adopt a diagonal compensation strategy by reusing the dropped elements. These dropped elements are added to the main diagonal elements of the same rows in the incomplete upper-triangular matrix $\boldsymbol{R}$. Bai et al. (2009) proposed practical incomplete Givens orthogonalization (IGO) methods for solving large sparse systems of linear equations. They claimed that these incomplete IGO methods took the storage requirements, the accuracy of the solutions and the coding of the pre-conditioners into consideration.

In this report, we have chosen the column-wise threshold incomplete Givens orthogonal (cTIGO) factorization algorithm for finding the sparse upper triangular matrix $\boldsymbol{R}$. This sparse upper triangular matrix $\boldsymbol{R}$ has sparse structure and can be used for specifying the GMRFs. The matrix $\boldsymbol{S}$ does not need to be stored in our setting since we only need the upper-triangular matrix $\boldsymbol{R}$. The matrix $\boldsymbol{S}$ only needs computed whenever it is explicitly needed.

3 Specifying GMRFs using sparse Cholesky factors

In this section we begin by introducing the background of GMRFs conditioned on a subset of the variables or on observed data. A small example is used to illustrate how the cTIGO algorithm works when applied to GMRFs.

3.1 GMRFs conditioned on a subset of the variables

I. GMRFs with soft constraint
Let $\boldsymbol{x}$ be a GMRF and assume that we have observed some linear transformation $\boldsymbol{Ax}$ with additional Gaussian distributed noise

$$\boldsymbol{e}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{Ax}, \boldsymbol{Q}_\epsilon^{-1}),$$

where k is the dimension of the vector $\boldsymbol{e}$, $\boldsymbol{A}$ is a $k \times n$ matrix with rank k and $k < n$, and $\boldsymbol{Q}_{\boldsymbol{\epsilon}} > 0$ is the precision matrix of $\boldsymbol{e}$. This is called "soft constraint" by Rue and Held (2005) and the log-density for the model is

$$\log \pi(\boldsymbol{x}|\boldsymbol{e}) = -\frac{1}{2}(\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{\mu})\boldsymbol{Q}_1(\boldsymbol{x} - \boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{e} - \boldsymbol{Ax})^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}(\boldsymbol{e} - \boldsymbol{Ax}) + \text{const}, \quad (9)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{Q}_1$ are the mean and the precision matrix of the GRMF, respectively, and "const" is constant. If $\boldsymbol{x}$ has mean $\boldsymbol{\mu} = \boldsymbol{0}$ then

$$\boldsymbol{x}|\boldsymbol{e} \sim \mathcal{N}_c(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{e}, \boldsymbol{Q}_1 + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A}). \quad (10)$$

Here we use the canonical form $\mathcal{N}_c(\cdot, \cdot)$ for $\boldsymbol{x}|\boldsymbol{e}$. We refer to Rue and Held (2005, Chapter 2.3.2) for more information about the canonical form

for GMRF. We can notice that for specifying the GMRFs with "soft constraint", the Routine (2) as shown in Figure 1 can be applied since $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A}$ with $\boldsymbol{Q_2} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A}$.

II. Models with auxiliary variables

Auxiliary variables are crucial in some models to retrieve GMRF full conditionals. We look at binary regression models with auxiliary variables.

Assume that we have Bernoulli observational model for binary responses. The binary responses have latent parameters which is a GMRF $\boldsymbol{x}$, and the GMRF usually depends on some hyperparameters $\boldsymbol{\theta}$. We usually choose the logit or probit models in this case, where

$$y_i \sim \mathcal{B}\left(\eta^{-1}(z_i^{\mathrm{T}}\boldsymbol{x})\right), \quad i = 1, 2, \ldots, m \tag{11}$$

where $\mathcal{B}(p)$ denotes a Bernoulli distribution with probability p for 1 and $1-p$ for 0. $\boldsymbol{z}_i$ is a vector of covariates and we assume it is fixed. $\eta(\cdot)$ is a link function

$$\eta(p) = \begin{cases} \log\left(p/(1-p)\right) & \text{for logit link} \\ \Phi(p) & \text{for probit link} \end{cases} \tag{12}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) for standard Gaussian distribution. We can use models with auxiliary variables $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_m)$ to represent these models,

$$\begin{aligned} \epsilon_i &\overset{iid}{\sim} G(\epsilon_i), \\ \omega_i &= \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{x} + \epsilon_i, \\ y_i &= \begin{cases} 1, & \text{if } \omega_i > 0, \\ 0, & \text{otherwise}, \end{cases} \end{aligned}$$

where $G(\cdot)$ is the CDF of standard logistic distribution in the logit case and $G(\cdot) = \Phi(\cdot)$ in the probit case. We refer to Forbes et al. (2011, Chapter 28) for more information about the standard logistic distribution and its CDF. Let $\boldsymbol{x}|\boldsymbol{\theta}$ be a GMRF of dimension n with mean $\boldsymbol{\mu} = \boldsymbol{0}$, and assume that we have $\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{x} = x_i$ and $m = n$. With the probit link the posterior distribution is

$$\pi(\boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\omega}|\boldsymbol{x})\pi(\boldsymbol{y}|\boldsymbol{\omega}). \tag{13}$$

The conditional distribution of $\boldsymbol{x}$ given the auxiliary variables can then be obtained

$$\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\omega}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}_1(\boldsymbol{\theta})\boldsymbol{x} - \frac{1}{2}\sum_i (x_i - \omega_i)\right)^2,$$

and this can be written in the canonical form

$$\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\omega} \sim \mathcal{N}_c(\omega, \boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{I}).$$

A general form for the conditional distribution of $\boldsymbol{x}$ given the auxiliary variables, for this binomial model with a probit link function, is given as

$$\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\omega} \sim \mathcal{N}_c(\boldsymbol{Z}^{\mathrm{T}}\omega, \boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}),$$

where $\boldsymbol{Z}$ is an $m \times n$ matrix. Similarly, the conditional distribution of $\boldsymbol{x}$ given the auxiliary variables for the logistic regression model can be written as

$$\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\omega} \sim \mathcal{N}_c(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Lambda}\omega, \boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{Z}),$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$, and λ_i is from the model specification. See more discussions on these models in Rue and Held (2005, Chapter 4.3).

In all the examples in this section, the models are suitable for use Routine (2) in Figure 1 to find the sparse Cholesky factors of the precision matrices of GMRFs.

3.2 GMRFs conditioned on data

As mentioned in Section 2.1, if a vector $\boldsymbol{x}$ is a GMRF with precision matrix $\boldsymbol{Q}_1$ and mean vector $\boldsymbol{\mu}$, then the density of the vector is given by Equation (2). In practical applications it is common to set $\boldsymbol{\mu} = \boldsymbol{0}$ (Rue and Held, 2005; Gneiting et al., 2010), which gives the probability density function

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{n/2} |\boldsymbol{Q}_1(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}_1(\boldsymbol{\theta})\boldsymbol{x}\right). \tag{14}$$

Assume that the data are of dimension k and defined as a k-dimensional random vector

$$\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{Q}_{\boldsymbol{\epsilon}}^{-1}\right)$$

and has the probability density function

$$\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{k} |\boldsymbol{Q}_{\boldsymbol{\epsilon}}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\right), \quad (15)$$

where $\boldsymbol{A}$ is a $k \times n$ matrix used to select the data location. The precision matrix $\boldsymbol{Q}_{\boldsymbol{\epsilon}}$ for the noise process is a positive definite matrix with dimension $k \times k$. Notice that the density function $\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ is not dependent on the $\boldsymbol{\theta}$, and hence the probability density function $\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ can be written as $\pi(\boldsymbol{y}|\boldsymbol{x})$. The probability density function of $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}$ can be found from Equations (14) and (15) through

$$\begin{aligned}\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \pi(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2}\left[x^{\mathrm{T}}(\boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A})\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{y}\right]\right).\end{aligned} \quad (16)$$

Similarly, the density function (16) can be written in the canonical form as

$$\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_c(\boldsymbol{\theta}), \boldsymbol{Q}_c(\boldsymbol{\theta})\right). \quad (17)$$

where $\boldsymbol{\mu}_c(\boldsymbol{\theta}) = \boldsymbol{Q}_c(\boldsymbol{\theta})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{y}$, and $\boldsymbol{Q}_c(\boldsymbol{\theta}) = \boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A}$. Now we can notice that the precision matrix for the GMRF conditional on data has the form $\boldsymbol{Q} = \boldsymbol{Q}_1(\boldsymbol{\theta}) + \boldsymbol{Q}_2$ with $\boldsymbol{Q}_2 = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{A}$, where $\boldsymbol{Q}_2$ does not depend on $\boldsymbol{\theta}$. Since $\boldsymbol{Q}$ has the same form as given in Routine (2) in Figure 1, it is possible to use the proposed routine to find the sparse Cholesky factor of the precision matrix of the GMRF conditioned on data.

Even though it is not the focus of this paper, it might be useful to point out that using Equations (14) - (16), we can find the analytical formula for the posterior density function of $(\boldsymbol{\theta}|\boldsymbol{y})$ through Bayes' formula. It is given by

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \text{const.} &+ \log(\pi(\boldsymbol{\theta})) + \frac{1}{2}\log(|\boldsymbol{Q}_1(\boldsymbol{\theta})|) \\ &- \frac{1}{2}\log(|\boldsymbol{Q}_c(\boldsymbol{\theta})|) + \frac{1}{2}\boldsymbol{\mu}_c(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{Q}_c(\boldsymbol{\theta})\boldsymbol{\mu}_c(\boldsymbol{\theta}).\end{aligned} \quad (18)$$

We refer to Hu et al. (2012) for detailed information about this log-posterior density function. The log-posterior density function $\log(\boldsymbol{\theta}|\boldsymbol{y})$ is crucial when doing statistical inference in Bayesian statistics.

The sparse structure of $\boldsymbol{Q}_2$ depends both on the structures of $\boldsymbol{A}$ and of $\boldsymbol{Q}_{\boldsymbol{\epsilon}}$. In most cases, the $\boldsymbol{Q}_{\boldsymbol{\epsilon}}$ is a diagonal matrix and the matrix $\boldsymbol{A}$ has sparse structure. Therefore $\boldsymbol{Q}_2$ should also have a sparse structure. When the observations are conditional independent, but have a non-Gaussian distribution, then we can use a GMRF approximation to obtain a sparse structure of $\boldsymbol{Q}_2$ as presented in Section 3.3.

3.3 GMRF approximation

Suppose there are n conditionally independent observations $y_1, y_2, \ldots, y_n$ from a non-Gaussian distribution and that y_i is an indirect observation of x_i. $\boldsymbol{x}$ is a GMRF with mean $\boldsymbol{\mu} = \boldsymbol{0}$ and precision matrix $\boldsymbol{Q}_1$. The full conditional $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ then has the form

$$\pi\left(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}_1\boldsymbol{x} + \sum_{i=1}^{n}\log\pi(y_i|x_i)\right). \tag{19}$$

Apply a second-order Taylor expansion of $\sum_{i=1}^{n}\log\pi(y_i|x_i)$ around $\boldsymbol{\mu}_0$. In other words, construct a suitable GMRF proposal density $\widetilde{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$

$$\begin{aligned}\widetilde{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}_1(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i=1}(a_i + b_i x_i - \frac{1}{2}c_i x_i^2)\right) \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{Q}_1(\boldsymbol{\theta}) + \mathrm{diag}(\boldsymbol{c})\right)\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}\right).\end{aligned} \tag{20}$$

c_i should set to zero when $c_i < 0$. $\boldsymbol{b}$ and $\boldsymbol{c}$ depend on $\boldsymbol{\mu}_0$. The canonical parametrization of $\widetilde{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ has the form

$$\mathcal{N}_c\left(\boldsymbol{b}, \boldsymbol{Q}_1(\boldsymbol{\theta}) + \mathrm{diag}(\boldsymbol{c})\right).$$

In this case $\boldsymbol{Q}_2$ has a diagonal structure. An important feature of (20) is that it inherits the *Markov* property of the prior on $\boldsymbol{x}$, which is useful for sampling GMRF. When $\boldsymbol{\mu} \neq \boldsymbol{0}$, the canonical parametrization of the

$(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is changed to

$$\mathcal{N}_c\left(\boldsymbol{Q}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{Q}_1(\boldsymbol{\theta}) + \mathrm{diag}(\boldsymbol{c})\right),$$

and does not change the matrix $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$.

As it was pointed out in Section 2.1, to sample from the GMRFs, the Cholesky factor $\boldsymbol{L}$ is one of most important factors. In order to save computational resources, a sparse Cholesky factor is preferable if the approximated precision matrix is "close" to the original precision matrix, where "close" means both in structure and the elements.

3.4 Theoretical background

It has been mentioned in Section 2.1 that the sparsity pattern of the Cholesky factor is determined by the graph $\mathcal{G}$, and it is unnecessary to calculate the zero elements in the Cholesky factor. In this section, we are going to introduce the theoretical background for finding the Cholesky factor from the orthogonal factorization when the GMRF is conditioned on observed data or a subset of the variables.

Let $\boldsymbol{y}$ be the observed data and assume $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ has the Gaussian distribution, then the density of $\boldsymbol{x}$ conditioned on $\boldsymbol{y}$ has the form in (16). In the discussed situations in Section 3.1 - Section 3.3, the precision matrix $\boldsymbol{Q}$ can be split into two parts, the precision matrix $\boldsymbol{Q}_1$ of the GMRF $(\boldsymbol{x}|\boldsymbol{\theta})$ and the matrix $\boldsymbol{Q}_2$ which is the additional effect. The matrix $\boldsymbol{Q}_2$ is usually a diagonal matrix or another type of sparse matrix. If the data is not Gaussian distributed, then we can apply the GMRFs approximation given in (20) and it returns the precision matrix $\boldsymbol{Q}_1$ with a diagonal matrix $\boldsymbol{Q}_2$ added. This structure satisfies the Routine (2) in Figure 1.

Let $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$ and assume that the Cholesky factors for $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$ and $\boldsymbol{Q}$ are $\boldsymbol{L}_1$, $\boldsymbol{L}_2$ and $\boldsymbol{L}$, respectively. The Cholesky factors $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are assumed to be known. We have the following results.

Observation 1. *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a zero mean GMRF with precision matrix $\boldsymbol{Q}_1$. Assume that the precision matrix has the form $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2 \in \mathbb{R}^{n \times n}$ when conditioned on observed data or a subset of the variables. Let the*

Cholesky factors for $\boldsymbol{Q}_1$ *and* $\boldsymbol{Q}_2$ *be* $\boldsymbol{L}_1$ *and* $\boldsymbol{L}_2$*, respectively. Form*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{pmatrix}.$$

Then $\boldsymbol{A}^T\boldsymbol{A}$ *is the precision matrix* $\boldsymbol{Q}$*.*

Proof. $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{L}_1^{\mathrm{T}} \\ \boldsymbol{L}_2^{\mathrm{T}} \end{pmatrix} = \boldsymbol{L}_1\boldsymbol{L}_1^{\mathrm{T}} + \boldsymbol{L}_2\boldsymbol{L}_2^{\mathrm{T}} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2 = \boldsymbol{Q}.$ □

From Observation 1 the following corollaries are established. Sketched proofs for these corollaries are given. We refer to Simpson (2008) for numerical examples with Corollary 1.

Corollary 1. *Let* $\boldsymbol{X}$ *be a zero mean GMRF with precision matrix* $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2 \in \mathbb{R}^{n\times n}$*, and let* $\boldsymbol{A}$ *have the form given in Observation 1. Let* $\boldsymbol{z} \in \mathbb{R}^{2n}$ *be a vector of independent and identically distributed (i.i.d.) standard Gaussian random variables. Then the solution of the least squares problem*

$$\boldsymbol{x} = \arg\min_{y} \|\boldsymbol{A}\boldsymbol{y} - \boldsymbol{z}\|_2 \tag{21}$$

is a sample from the GMRF $\boldsymbol{X}$*.*

Proof. $\boldsymbol{Q} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ from Observation 1 is the starting point to prove this Corollary. Denote $\boldsymbol{A}^\dagger$ the Moore-Penrose pseudo-inverse of $\boldsymbol{A}$, and then the solution to the least squares problem is $\boldsymbol{x} = \boldsymbol{A}^\dagger\boldsymbol{z}$ (Björck, 1996). From the definition of the pseudo-inverse, $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{S}^\dagger\boldsymbol{U}^{\mathrm{T}}\boldsymbol{z}$, where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{W}^{\mathrm{T}}$ is a singular value decomposition of $\boldsymbol{A}$ and $\boldsymbol{S}^\dagger \in \mathbb{R}^{d\times 2d}$ is the matrix with the reciprocals of the non-zero singular values on the diagonal. We can verify that $\boldsymbol{x}$ has the required distribution, and it is sufficient to check the first two moments since $\boldsymbol{x}$ has a Gaussian distribution, being linear in $\boldsymbol{z}$. It is clear that $\mathbb{E}(\boldsymbol{x}) = 0$. Furthermore,

$$\begin{aligned}
\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}) &= \boldsymbol{A}^\dagger\mathbb{E}(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}})(\boldsymbol{A}^\dagger)^{\mathrm{T}} \\
&= \boldsymbol{W}\boldsymbol{S}^\dagger\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{S}^\dagger)^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}} \\
&= \boldsymbol{W}\boldsymbol{S}^\dagger(\boldsymbol{S}^\dagger)^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}} \\
&= \boldsymbol{W}(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S})^\dagger\boldsymbol{W}^{\mathrm{T}}.
\end{aligned}$$

Calculations yield $\boldsymbol{Q} = \boldsymbol{W}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{W}^{\mathrm{T}}$ and, hence, $\boldsymbol{x} \sim MVN(0, \boldsymbol{Q}^{\dagger})$. □

Therefore, it is possible to sample from a GMRF by solving the sparse least squares problem given in (21) with some conditions on GMRFs.

Corollary 2. *The upper triangular matrix $\boldsymbol{R}$ from the orthogonal factorization of the rectangular matrix*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{pmatrix} \tag{22}$$

is the Cholesky factor of the precision matrix $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$.

Proof. Since the upper triangular matrix from the orthogonal factorization is the Cholesky factor for the normal equations matrix, this is obvious from Observation 1. □

By using the orthogonal factorization of the rectangular matrix $\boldsymbol{A}$, it is possible to get samples from the GMRFs when they are conditioned on data or a subset of the variables by using Corollary 1 or the Cholesky factor from Observation 1 together with the sampling algorithms discussed in Rue and Held (2005, Chapter 2).

3.5 Dropping strategies

In this section the dropping strategy for the incomplete orthogonal factorization is introduced in order to find the incomplete Cholesky factor for matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. Together with some dropping strategy for the incomplete orthogonal factorization of the rectangular matrix $\boldsymbol{A}$, a sparse upper triangular matrix $\boldsymbol{R}$ can be obtained. From Corollary 2 and the discussion in Section 2.2, we know that $\boldsymbol{R}$ is an incomplete Cholesky factor or sparse Cholesky factor for the precision matrix $\boldsymbol{Q}$. This sparse Cholesky factor can then be used to specify the GMRF. The dropping strategies are important when doing the incomplete orthogonal factorization. Generally speaking, there are two kinds of dropping strategies.

1. Drop fill-ins based on sparsity patterns. Before doing the incomplete orthogonal factorization, the sparsity pattern of the upper triangular matrix is predefined and fixed. If the factorization based only on

the sparsity pattern of the original matrix, we drop all the elements which are pre-defined to be zeros. The algorithm does not consider the actual numerical values of the elements during the factorizations.

2. Drop fill-ins by using a numerical threshold. This strategy only includes the elements in $\boldsymbol{R}$ if they are bigger than a predefined threshold value. Munksgaard (1980) presented one way to select the value of the threshold parameter. His strategy drops the elements which are smaller than the diagonal elements of their rows and columns, multiplied by some predefined small value (called dropping tolerance). In this report, a slightly different dropping strategy is chosen. During the incomplete orthogonal factorization using Givens rotations, or the column-wise threshold incomplete Givens orthogonal (cTIGO) factorization (Papadopoulos et al., 2005), we drop the elements according to their magnitudes with some predefined dropping tolerance. The nonzero pattern of $\boldsymbol{R}$ is determined dynamically.

Both the fixed sparsity pattern strategy and the dynamic strategy are useful in applications. The fixed sparsity pattern strategy is the candidate when the computation resources are low. It is usually faster but sometimes returns unsatisfactory results. The dynamic strategy will in most cases return satisfactory results by choosing proper dropping tolerances but it is usually more expensive both in time and computations.

There are different versions of orthogonal factorizations. We refer to Saad (2003), Golub and Van Loan (1996) and Trefethen and Bau (1997) for more information. Based on the research of Bai et al. (2001), Papadopoulos et al. (2005) and Bai et al. (2009), we choose the incomplete orthogonal factorization using Givens rotations to find the sparse Cholesky factor. This algorithm is stable and robust and always returns a sparse matrix. This algorithm inherits the advantages of orthogonal factorization. Bai et al. (2001) commented that there is little attention given to incomplete orthogonal factorization with Givens rotations, which is actually useful in many numerical problems.

In order to use Givens rotations for incomplete orthogonal factorization, the following nonzero patterns needs to be defined,

$$
\begin{aligned}
N_{\boldsymbol{Q}} &= \{(i,j) \mid Q_{ij} \neq 0, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{Q},l} &= \{(i,j) \mid Q_{ij} \neq 0, i \geq j, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{Q},u} &= \{(i,j) \mid Q_{ij} \neq 0, i \leq j, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{L}_1} &= \{(i,j) \mid L_{1_{ij}} \neq 0, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{L}_2} &= \{(i,j) \mid L_{2_{ij}} \neq 0, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{L}} &= \{(i,j) \mid L_{ij} \neq 0, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{A}} &= \{(i,j) \mid A_{ij} \neq 0, 1 \leq i,j \leq n\},\\
N_{\boldsymbol{R}} &= \{(i,j) \mid R_{ij} \neq 0, 1 \leq i,j \leq n\},
\end{aligned}
$$

where $N_{\boldsymbol{Q}}$ is the nonzero pattern of the matrix $\boldsymbol{Q}$, and $N_{\boldsymbol{Q},u}$, $N_{\boldsymbol{Q},l}$ are the nonzero patterns of the upper and lower triangular parts of the matrix $\boldsymbol{Q}$, respectively. $N_{\boldsymbol{L}_1}$, $N_{\boldsymbol{L}_2}$, $N_{\boldsymbol{L}}$ $N_{\boldsymbol{A}}$ and $N_{\boldsymbol{R}}$ are the nonzero patterns of the lower triangular matrix $\boldsymbol{L}_1$, the lower triangular matrix $\boldsymbol{L}_2$, the lower triangular matrix $\boldsymbol{L}$, the matrix $\boldsymbol{A}$ and the matrix $\boldsymbol{R}$, respectively. These matrices are already formulated in previous sections.

In order to use the cTIGO algorithm, the rectangular matrix $\boldsymbol{A}$ in (22) is formed. The sparsity pattern of matrix $\boldsymbol{A}$ is already known beforehand. However, since the dynamic strategy is chosen, there will be some fill-in during Givens rotations process, and the sparsity pattern of the sparse Cholesky factor $\boldsymbol{R}$ will depend on the dropping tolerance and usually $N_{\boldsymbol{R}} < N_{\boldsymbol{A}}$. For more information about cTIGO algorithm, see Bai et al. (2001) for theoretical issues and Papadopoulos et al. (2005) for implementations.

3.6 A small example

A small example is explored in this section to illustrate how to use the cTIGO algorithm to find the sparse Cholesky factor $\boldsymbol{R}$. For simplicity and without loss of generality, we assume that $\boldsymbol{Q}_1$ is the precision matrix for a zero mean GMRF $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_1)$, and that the data are normally distributed, i.e., $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and hence $\boldsymbol{Q}_2 = \boldsymbol{I}$. Assume that these matrices are given as follows

$$\boldsymbol{Q_1} = \begin{pmatrix} 5 & -1 & 0 & \dots & 0 & -1 \\ -1 & 5 & -1 & 0 & \dots & 0 \\ 0 & -1 & 5 & -1 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ -1 & 0 & \dots & & -1 & 5 \end{pmatrix}_{9\times 9}$$

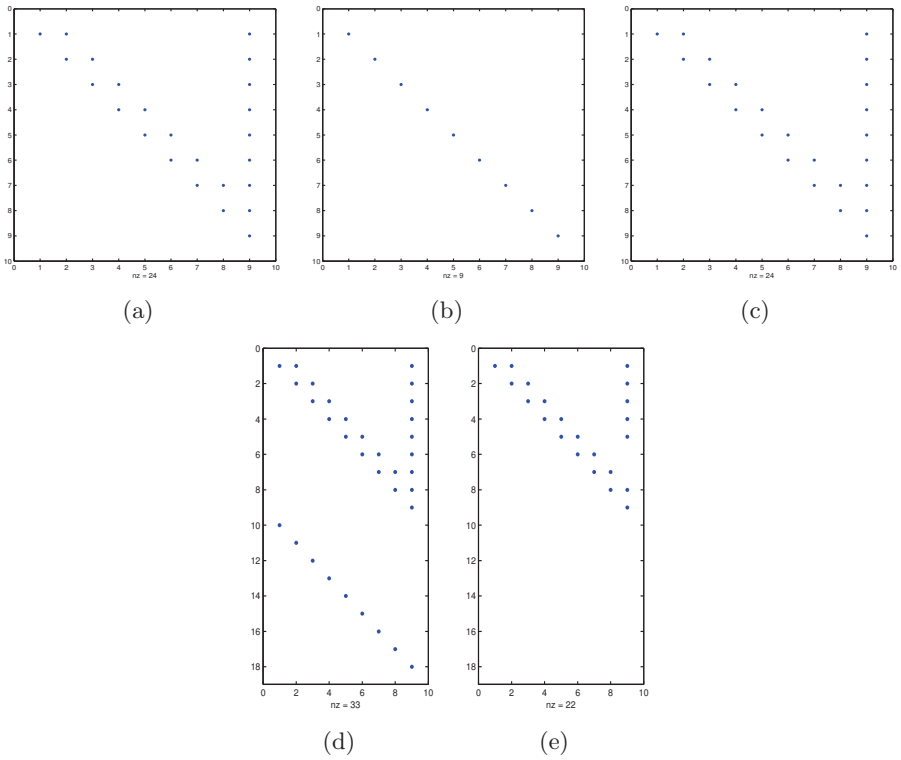
and

$$\boldsymbol{Q}_2 = \boldsymbol{I}_{9\times 9} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & & & 1 \end{pmatrix}_{9\times 9}.$$

Let $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ denote the Cholesky factor of the two matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, respectively, with the sparsity patterns given in Figure 2(a) and Figure 2(b). The rectangular matrix $\boldsymbol{A}$ can then be formed as given in (22) with the sparsity pattern given in Figure 2(d). Apply the cTIGO algorithm to the rectangular matrix $\boldsymbol{A}$ with a dropping tolerance of 0.0001 to find the sparse incomplete Cholesky factor $\boldsymbol{R}$. The sparsity pattern of $\boldsymbol{R}$ is given in Figure 2(e). The sparsity pattern of the Cholesky factor $\boldsymbol{L}$ from the standard Cholesky factorization of the precision matrix $\boldsymbol{Q}$ is given in Figure 2(c).

We notice that the precision matrix $\boldsymbol{Q}_1$ is quite similar to the tridiagonal matrix except the values at two of the corners. However, there is a lot of fill-in in the Cholesky factor $\boldsymbol{L}_1$. This is a common structure for the precision matrix of a GMRF, for instance, a GMRF on a torus. The same comments can be given for $\boldsymbol{Q}$ and $\boldsymbol{L}$. Note that the upper triangular matrix $\boldsymbol{R}$ has less nonzero elements than $\boldsymbol{L}_1$, $\boldsymbol{L}$ and $\boldsymbol{A}$, $N_{\boldsymbol{R}} < N_{\boldsymbol{L}_1}, N_{\boldsymbol{R}} < N_{\boldsymbol{L}}$ and $N_{\boldsymbol{R}} < N_{\boldsymbol{A}}$. The sparsity pattern of $\boldsymbol{R}$ depends on the dropping tolerance and also the elements of the matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, but we are not going deeper here.

As discussed in Section 2.2 and Section 3.5, the sparse upper triangular matrix $\boldsymbol{R}$ is an incomplete Cholesky factor for the precision matrix $\boldsymbol{Q}$ of the GMRF when it is conditioned on data. The error matrix $\boldsymbol{E}$ between the true precision matrix $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$ and the approximated precision



(a) (b) (c)

(d) (e)

Figure 2: The sparsity patterns of $\boldsymbol{L}_1$ (a), $\boldsymbol{L}_2$ (b), $\boldsymbol{L}$ (c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$ (e)

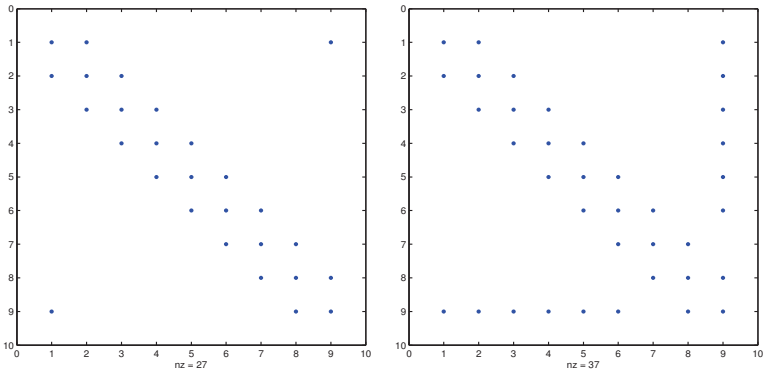
matrix $\widetilde{\boldsymbol{Q}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$ is given by

$$\boldsymbol{E} = (\boldsymbol{Q}_1 + \boldsymbol{Q}_2) - \widetilde{\boldsymbol{Q}}. \tag{23}$$

The sparsity patterns of the precision matrix $\boldsymbol{Q}$ and its approximation $\widetilde{\boldsymbol{Q}}$ are shown in Figure 3(a) and Figure 3(b), respectively. In order to compare the difference between the approximated covariance matrices (inverse of the approximated precision matrix) $\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{Q}}^{-1}$ and the true covariance matrix (inverse of the true precision matrix) $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$, we calculate the error matrix $\widetilde{\boldsymbol{E}}$,

$$\widetilde{\boldsymbol{E}} = \boldsymbol{\Sigma} - \widetilde{\boldsymbol{\Sigma}}. \tag{24}$$

The images of $\boldsymbol{\Sigma}$, $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{E}}$ are given in Figure 4, and they show that the difference between $\boldsymbol{\Sigma}$ and $\widetilde{\boldsymbol{\Sigma}}$ is quite small. By chosen different dropping tolerance, the error can be made smaller and become negligible.

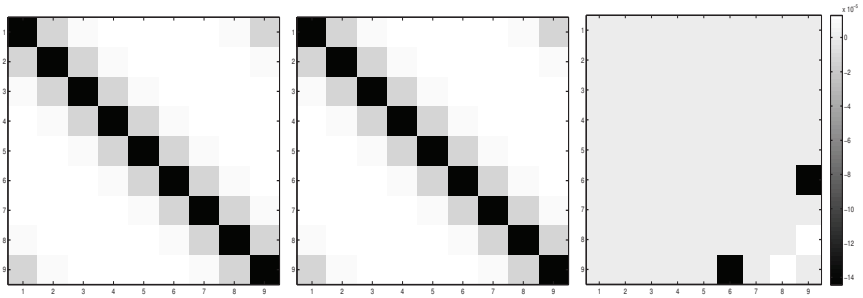


(a) (b)

Figure 3: Sparsity patterns of the true precision matrix $\boldsymbol{Q}$ (a) and the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ (b)

4 Simulation Results with cTIGO algorithm

Using the incomplete orthogonal factorization with Givens rotations, it leads to a sparse upper triangular matrix $\boldsymbol{R}$, which is a sparse incomplete Cholesky factor for the precision matrix $\boldsymbol{Q}$ and can be used to specify the GMRF. Hence it has the potential possibility to reduce the computational cost. We first apply the cTIGO algorithm to some commonly used structures of the precision matrices in Section 4.1. In Section 4.2, we apply the cTIGO algorithm to precision matrices which are generated from the stochastic partial differential equations (SPDEs) discussed in Lindgren et al. (2011) and Fuglstad (2011).



(a) (b) (c)

Figure 4: Images of the true covariance matrix $\boldsymbol{\Sigma}$ (a), approximately covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (b) and the error matrix $\widetilde{\boldsymbol{E}}$ (c)

4.1 Simulation results for precision matrices with commonly used structures

It is known that if the precision matrix $\boldsymbol{Q} > 0$ is a band matrix with bandwidth p, then its Cholesky factor $\boldsymbol{L}$ (lower triangular matrix) has the same bandwidth p. See Golub and Van Loan (1996) (Theorem 4.3.1) for a direct proof and Rue and Held (2005, Chapter 2.4.1) for more information on how to finding Cholesky factor efficiently in this case with Algorithm 2.9. Wist and Rue (2006) pointed out that if the original precision matrix $\boldsymbol{Q}$ is a band matrix, then the incomplete Cholesky factor $\widetilde{\boldsymbol{L}}$ from the incomplete Cholesky factorization will also be a band matrix with the same bandwidth p.

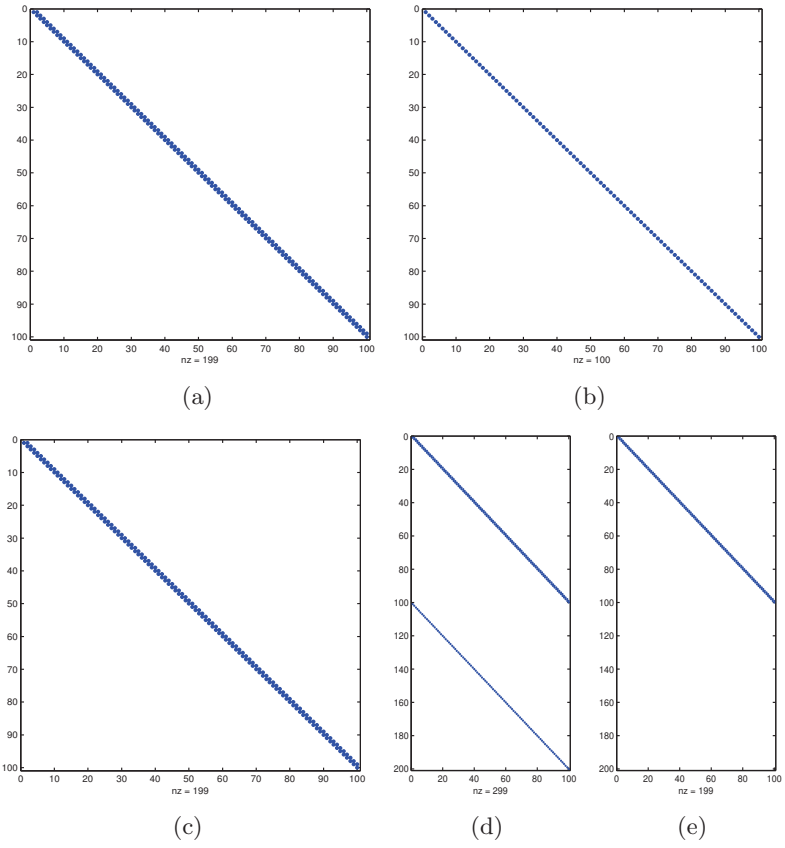
In this section we consider some commonly used structures for the precision matrices. The first two examples are band matrices with different bandwidths. Let x be Gaussian auto-regressive processes of order 1 or 2, and then the precision matrix for the process will be a band matrix with bandwidth $p = 2$ or $p = 3$, respectively. The precision matrices for the first-order Random Walk (RW1) and the second-order Random Walk (RW2) models have bandwidths $p = 2$ and $p = 3$. Since these models are intrinsic GMRFs, the precision matrices are not of full rank. We fix this by slightly modifying the elements in the precision matrices for the RW1 and RW2 models but we still called them as the precision matrices for the RW1 and the RW2 models. For more information about intrinsic GMRFs

and the RW1 and RW2 models, see, for example, Rue and Held (2005, Chapter 3).

Assume that the data are Gaussian distributed. Then from Section 3.2 the matrix $\boldsymbol{Q}_2$ is a diagonal matrix when $\boldsymbol{A} = \boldsymbol{I}$. For simplicity and without lost of generality, assume the data $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then the matrix $\boldsymbol{Q}_2$ and its Cholesky factor $\boldsymbol{L}_2$ are identity matrices. Since we know exactly what the sparsity patterns of the precision matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ and the Cholesky factors $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are, the sparsity pattern of $\boldsymbol{A}$ is known beforehand and can be taken advantage of in the implementation. By applying the cTIGO algorithm to the matrix $\boldsymbol{A}$ with dropping tolerance $\tau = 0.0001$, the sparse upper triangular matrix $\boldsymbol{R}$ can be obtained. The sparsity patterns of the matrices $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{L}$, $\boldsymbol{A}$ and $\boldsymbol{R}$ are given in Figure 5. The sparsity patterns of the true precision matrix $\boldsymbol{Q}$ and the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ are given in Figure 6. The image of the true covariance matrices $\boldsymbol{\Sigma}$, the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ and the error matrix $\widetilde{\boldsymbol{E}}$ for the RW1 model are shown in Figure 7. Note that the order of the numerical values in the error matrix $\widetilde{\boldsymbol{E}}$ is 10^{-8}, which is essentially zero in practice applications.

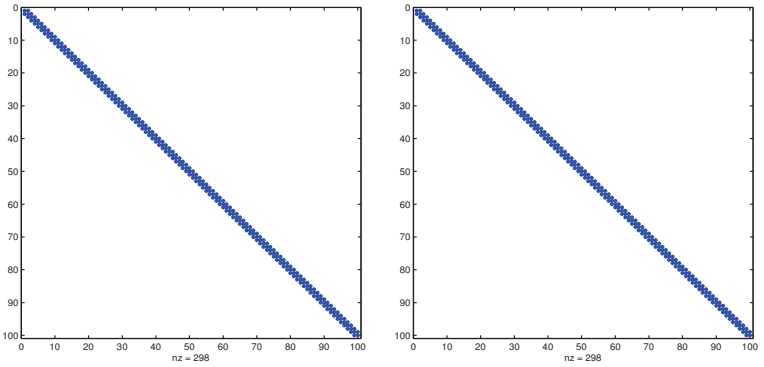
Similarly for the RW2 model we apply the cTIGO algorithm to the matrix $\boldsymbol{A}$ with the dropping tolerance $\tau = 0.0001$. The results in this case are quite similar to the results for the RW1 model. We only show the images of the true covariance matrix $\boldsymbol{\Sigma}$, the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$, and the error matrix $\widetilde{\boldsymbol{E}}$. The results are given in Figure 8. Note that the order of the numerical values in the error matrix $\widetilde{\boldsymbol{E}}$ is 10^{-8} as for the RW1 model. See Section 4.4 from more simulation results for the RW1 and RW2 models and discussions. We can notice that the sparseness of $\boldsymbol{R}$ is the same as $\boldsymbol{L}$. Hence in these two cases, we do not save computational resources. However, this approach is still have the potential to be used in applications since it is robust.

The next example we have chosen is a block tridiagonal matrix of order n^2 resulting from discretizing Poisson's equation with the 5-point operator on an n-by-n mesh. Thus it is called Poisson matrix in this paper. The sparsity pattern of this matrix is given in Figure 10(a). With the Poisson matrix and $\boldsymbol{Q}_2$ as before, we find the Cholesky factors $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ and form the rectangular matrix $\boldsymbol{A}$. We apply the cTIGO algorithm to the matrix $\boldsymbol{A}$ with dropping tolerance $\tau = 0.0001$ to find the sparse upper triangular



(a) (b) (c) (d) (e)

Figure 5: Sparsity patterns of $\boldsymbol{L}_1$ (a), $\boldsymbol{L}_2$ (b), $\boldsymbol{L}$ (c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$ (e) for the RW1 model

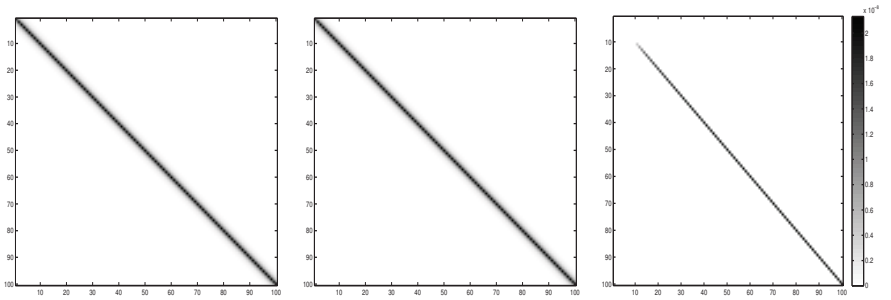


(a) (b)

Figure 6: Sparsity patterns for the true precision matrix $\boldsymbol{Q}$ (a) and the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ (b) for the RW1 model

matrix $\boldsymbol{R}$. The sparsity patterns of the matrices $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{L}$, $\boldsymbol{A}$ and $\boldsymbol{R}$ are given in Figure 9(a) - Figure 9(e), respectively. The sparsity patterns of the true precision matrix $\boldsymbol{Q}$ and the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ are given in Figure 10. We can notice that the upper triangular matrix $\boldsymbol{R}$ is sparser than the Cholesky factor $\boldsymbol{L}$ from the original precision matrix $\boldsymbol{Q}$. It can be shown that the sparseness depends on the dropping tolerance τ. The images of the true covariance matrices $\boldsymbol{\Sigma}$, the approximated precision matrix $\widetilde{\boldsymbol{\Sigma}}$, and the error matrix $\widetilde{\boldsymbol{E}}$ in this case are shown in Figure 11. Note that the order of the numerical values in the error matrix $\widetilde{\boldsymbol{E}}$ is 10^{-5}. This is small for practical use. More results for this band matrix are given in Section 4.4.

The next example is a precision matrix with a nearly band matrix. Assume that $\boldsymbol{Q}_1$ is a nearly banded matrix but with the values $\boldsymbol{Q}_1(1,n) = 1$ and $\boldsymbol{Q}_1(n,1) = 1$. We call this matrix as Toeplitz matrix in this paper. The sparsity pattern of this matrix is given in Figure 13(a). With the dropping tolerance $\tau = 0.0001$, we apply the cTIGO algorithm to the rectangular matrix $\boldsymbol{A}$. The sparsity patterns of $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{L}$, $\boldsymbol{A}$ and $\boldsymbol{R}$ are given in Figure 12(a) - Figure 12(e), respectively. We notice that the upper triangular matrix $\boldsymbol{R}$ is sparser than the matrix $\boldsymbol{L}$. We can also notice that the sparseness of $\boldsymbol{R}$ depends on the tolerance τ. The sparsity pattern of



(a) (b) (c)

Figure 7: Images of true covariance matrix $\boldsymbol{\Sigma}$ (a), the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (b) and and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for RW1 model

the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ is given in Fig 13(b). The image of the true covariance matrices $\boldsymbol{\Sigma}$, the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$, and the error matrix $\widetilde{\boldsymbol{E}}$ are shown in Figure 14. Note that the order of the numerical values in the error matrix is 10^{-5} with the given tolerance. More simulation results for this matrix are found in Section 4.4.

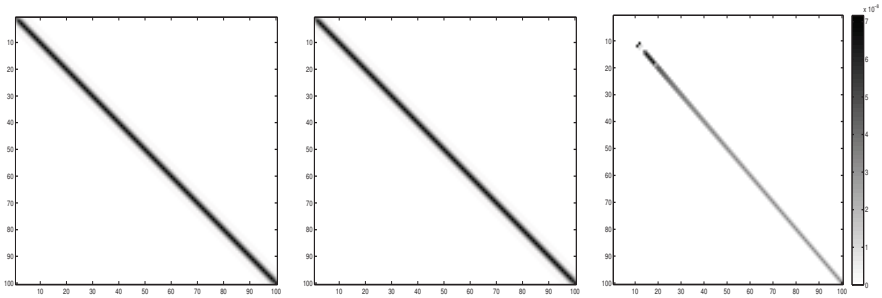
4.2 Simulation results for particular precision matrices

In this section we emphasize on some particular precision matrices, namely the precision matrices from the stochastic partial differential equations (SPDEs) approach discussed by Lindgren et al. (2011) and Fuglstad (2011). As pointed out by Lindgren et al. (2011) there is an explicit link between GRFs and GMRFs through SPDEs. The important relationship which was initially used by Lindgren et al. (2011) is that the solution $\boldsymbol{x}(\boldsymbol{u})$ to the following SPDE is a GRF with Matérn covariance function,

$$(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \;\; \boldsymbol{s} \in \mathbb{R}^d, \;\; \alpha = \nu + d/2, \;\; \kappa > 0, \;\; \nu > 0, \quad (25)$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial}{\partial x_i^2}$ is the Laplacian, $(\kappa^2 - \Delta)^{\alpha/2}$ is a differential operator and d is the dimension of the field $x(\boldsymbol{s})$. Fuglstad (2011) extended this approach to construct anisotropic and inhomogeneous fields with the SPDE

$$\kappa^2(\boldsymbol{u}) x(\boldsymbol{u}) - \nabla \cdot \boldsymbol{H}(\boldsymbol{u}) \nabla x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), \quad (26)$$



(a) (b) (c)

Figure 8: Images of the true covariance matrix $\boldsymbol{\Sigma}$ (a), the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (b) and and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for RW2 model

where κ and $\boldsymbol{H}$ control the local range and anisotropy, and $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$. One important difference between Lindgren et al. (2011) and Fuglstad (2011) is that Lindgren et al. (2011) have chosen the Neumann boundary condition but Fuglstad (2011) has chosen the periodic boundary condition. With Neumann boundary condition the precision matrix $\boldsymbol{Q}_1$ is a band matrix. However, the periodic boundary condition gives elements " in the corners" of the precision matrix. Hu et al. (2012) extended the approach to multivariate settings by using systems of SPDEs. For more information about the SPDE approach, We refer to Lindgren et al. (2011), Fuglstad (2011) and Hu et al. (2012).

First, choose the precision matrix for $\boldsymbol{Q}_1$ that results from the discretization of the SPDE (25) with $\alpha = 2$, $d = 2$ and $\kappa = 0.3$. The sparsity pattern of $\boldsymbol{Q}_1$ is given in Figure 16(a). We still assume $\boldsymbol{Q}_2 = \boldsymbol{I}$. The sparsity patterns of $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{L}$, $\boldsymbol{A}$ and $\boldsymbol{R}$ are given in Figure 15(a) -15(e), respectively. We notice that the upper triangular matrix $\boldsymbol{R}$ is sparser than the matrix $\boldsymbol{L}$. The sparsity pattern of the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ is given in Figure 16(b). The images of the true covariance matrix $\boldsymbol{\Sigma}$, the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ and the error matrix $\widetilde{\boldsymbol{E}}$ are shown in Figure 17. We can notice that the elements of in the error matrix $\widetilde{\boldsymbol{E}}$ are reasonably small.

The second precision matrix for $\boldsymbol{Q}_1$ is generated from the SPDE (26)

with $\kappa = 0.1$ and

$$\boldsymbol{H} = 0.1 \times \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The sparsity pattern of the precision matrix $\boldsymbol{Q}_1$ is given in Figure 19(a). We use the same $\boldsymbol{Q}_2$ as previous examples. The sparsity patterns of $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{L}$, $\boldsymbol{A}$ and $\boldsymbol{R}$ are given in Figure 18(a) - Figure 18(e), respectively. We can notice that the upper triangular matrix $\boldsymbol{R}$ is sparser than the matrix $\boldsymbol{L}$. The sparsity pattern of the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ is given in Figure 19(b). The images of the true covariance matrix $\boldsymbol{\Sigma}$, the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ and the error matrix $\widetilde{\boldsymbol{E}}$ are illustrated in Figure 20. We could notice that the order of the numerical values in the error matrix $\widetilde{\boldsymbol{E}}$ are also reasonably small in this case.

4.3 Sampling from GMRFs

In this section samples from a GMRF are obtained using the sparse upper triangular matrices $\boldsymbol{R}$ and the Cholesky factors $\boldsymbol{L}$ for the precision matrices $\boldsymbol{Q}$. Let the precision matrix $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$, where $\boldsymbol{Q}_1$ is from the SPDE (25) or (26), and $\boldsymbol{Q}_2$ is a diagonal matrix. The sampling is done as follows.

- Compute the Cholesky factor $\boldsymbol{L}$ with a Cholesky factorization or compute the sparse upper triangular matrix $\boldsymbol{R}$ from the cTIGO algorithm;
- Sample $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$;
- Solve the equation $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{z}$ or $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{z}$;
- $\boldsymbol{x}$ is the sample of the GMRF with precision matrix $\boldsymbol{Q}$ or $\widetilde{\boldsymbol{Q}}$.

If the mean $\boldsymbol{\mu}$ of the field is not zero, then we just need a last step $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{x}$ to correct the mean. With $\boldsymbol{L}$ the field $\boldsymbol{x}$ has the true covariance matrix $\boldsymbol{Q}$ because

$$\mathrm{Cov}(\boldsymbol{x}) = \mathrm{Cov}(\boldsymbol{L}^{-\boldsymbol{T}}\boldsymbol{z}) = (\boldsymbol{L}\boldsymbol{L}^{T})^{-1} = \boldsymbol{Q}^{-1}.$$

Similarly, with $\boldsymbol{R}$ the field $\boldsymbol{x}$ has the approximated covariance matrix $\tilde{\boldsymbol{Q}}$. Many other sampling algorithms are provided by Rue and Held (2005, Chapter 2) for different parametrization of the GMRF. We cannot notice

any large differences between the samples using the Cholesky factor $\boldsymbol{L}$ and the samples using the sparse matrix $\boldsymbol{R}$ based on Figure 21 and Figure 22.

4.4 Effect of dropping tolerance

In this section we choose different values of τ in order to know the effect of dropping tolerance. We use the same kinds of structures for the precision matrices as discussed in Section 4.1 with dropping tolerances $\tau = \{0, 0.000001, 0.00001, 0.0001, 0.001, 0.01\}$. The 1-norm for the error matrix $\boldsymbol{E} = \boldsymbol{Q} - \widetilde{\boldsymbol{Q}}$ is used for the comparisons. The results are given in Table 1 and Table 2. From these tables, we can see that as the dropping tolerance τ becomes smaller and smaller, the error becomes smaller and smaller. Notice that by choosing 10^{-5} as the dropping tolerance, the error reaches a level acceptable in many applications. If the dropping tolerance is equal to 0, then the error is also equal to zero which means no element has been zeroed out during the Givens rotations and it returns the common Cholesky factor.

Table 1: Comparisons for RW1 Model and RW2 Model

Random Walk 1		Random Walk 2	
tolerance	error	tolerance	error
0.01	2.55E-04	0.01	0.33
0.001	1.66E-06	0.001	1.80E-05
0.0001	2.13E-08	0.0001	1.48E-07
0.00001	2.50E-10	0.00001	1.07E-08
1.00E-6	2.34E-12	1.00E-6	2.15E-09
0	4.00E-15	0	1.73E-14

5 Discussion and Conclusion

In this paper we use the cTIGO algorithm to find sparse Cholesky factors for specifying GMRFs. Some commonly used structures of the precision matrices and two precision matrices generated from SPDEs have been tested. By using the incomplete orthogonal factorization with Givens rotations, a sparse incomplete Cholesky factor can be found and it is usually

Table 2: Comparisons for Poisson matrix and Toeplitz matrix

Poisson matrix		Toeplitz matrix	
tolerance	error	tolerance	error
0.01	0.11	0.01	9.15E-03
0.001	8.51E-03	0.001	9.59E-04
0.0001	6.33E-04	0.0001	7.91E-05
0.00001	5.91E-05	0.00001	8.70E-06
1.00E-6	3.44E-06	1.00E-6	7.19E-07
0	1.49E-14	0	5.66E-15

sparser than the Cholesky factor from the standard Cholesky factorization. The sparsity of the incomplete Cholesky factor depends on the value of the tolerance. With a good choice for the dropping tolerance, the error between the true covariance matrix and the approximated covariance matrix becomes negligible.

One advantage of this approach is that it is robust. It always produces a sparse incomplete Cholesky factor. Since the algorithm works both for square matrices and for rectangular matrices, this approach can be applied to GMRFs conditioned on observed data or a subset of the variable. On the negative side, it seems that our current implementation of the approach is slow when the dimension of the matrix becomes large. We believe that this is due to the nature of the incomplete orthogonal factorization with dynamic dropping strategy. The orthogonal factorization is usually slower than the Cholesky factorization. Further, Givens rotations only zero out values to zeros one at a time. This leads to the slowness of the algorithm. When the computation resources are limited, we might need to use the fixed pattern dropping strategy. However, to implement a fast cTIGO algorithm is out the scope of this paper and it is for further research.

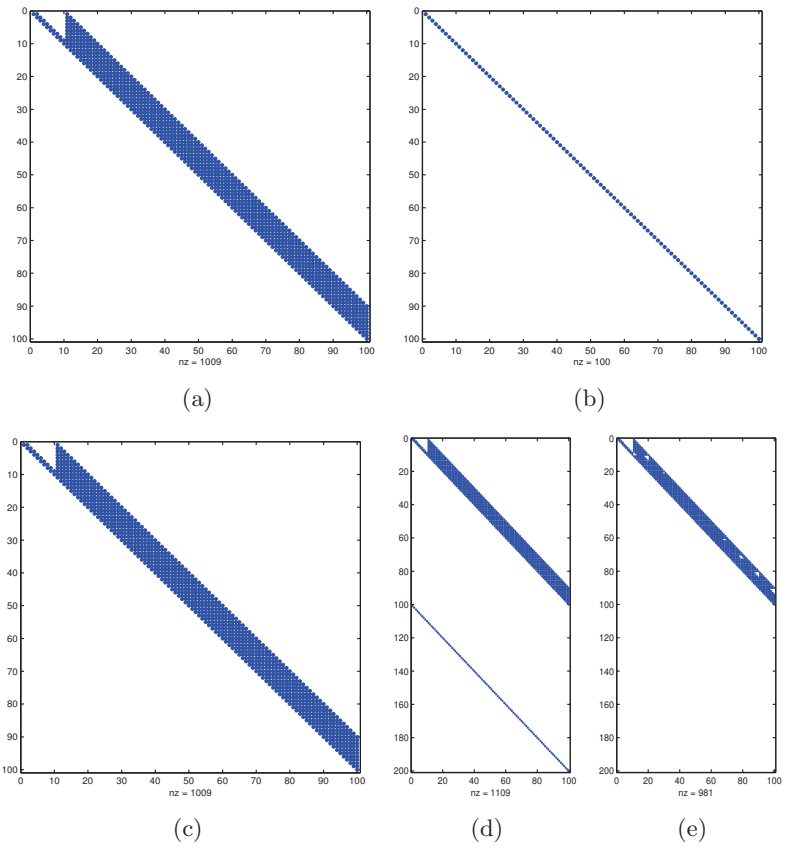
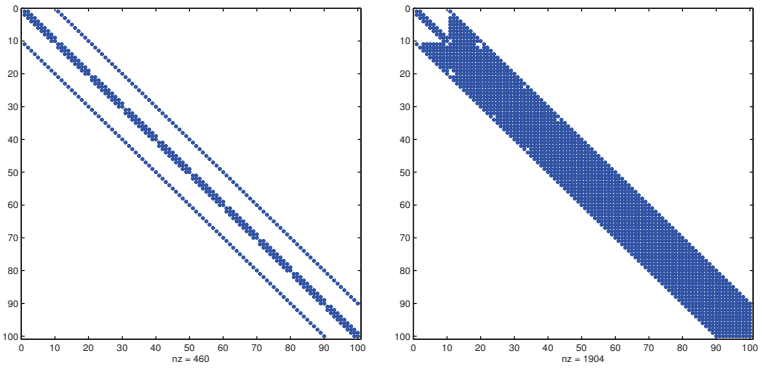
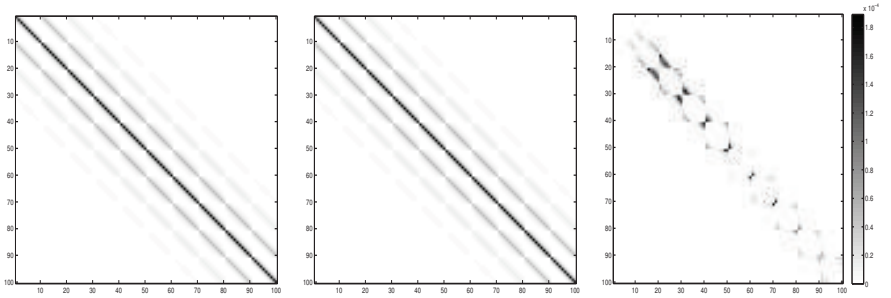


Figure 9: Sparsity patterns of $\boldsymbol{L}_1$ (a), $\boldsymbol{L}_2$ (b), $\boldsymbol{L}$ (c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$ (e) with Poisson matrix



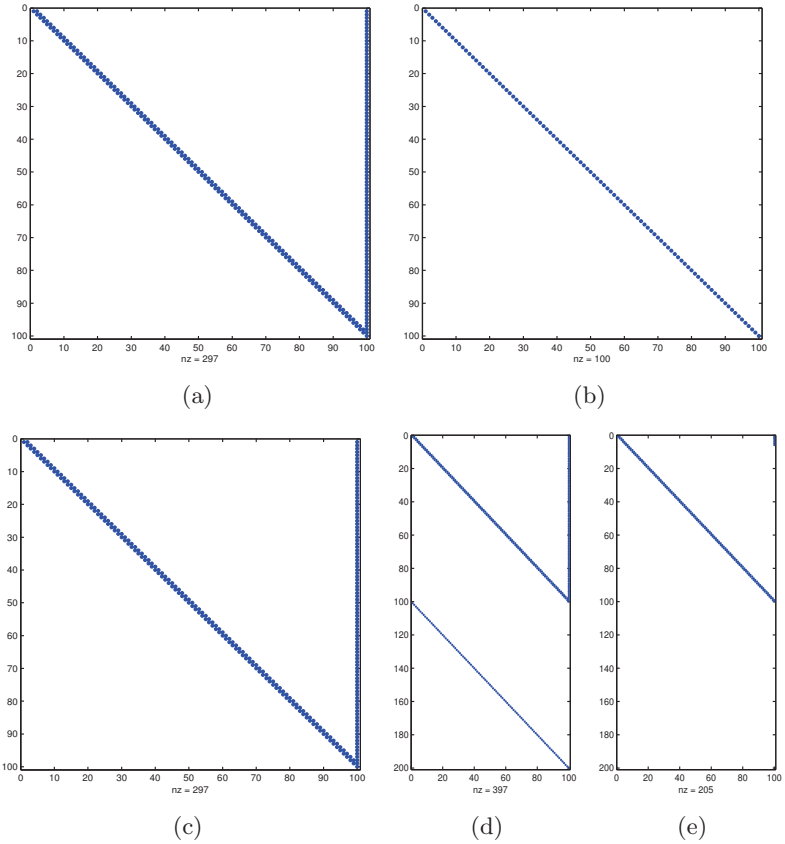
(a) (b)

Figure 10: Sparsity patterns for the true precision matrix $\boldsymbol{Q}$ (a) and the approximated precision matrix $\widetilde{\boldsymbol{Q}}$ (b) with Poisson matrix



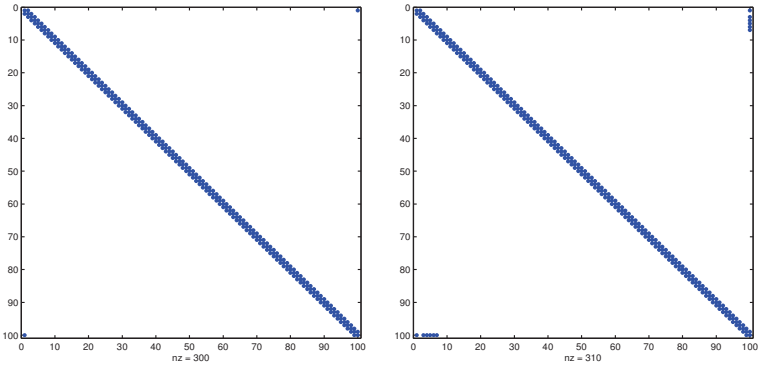
(a) (b) (c)

Figure 11: Images of true covariance matrix $\boldsymbol{\Sigma}$ (a), the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (b), and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for Poisson matrix



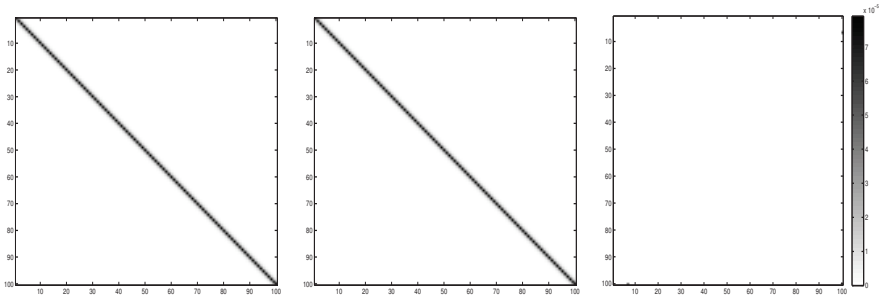
(a) (b) (c) (d) (e)

Figure 12: Sparsity patterns for $\boldsymbol{L}_1$ (a), $\boldsymbol{L}_2$ (b), $\boldsymbol{L}$ (c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$ for Toeplitz matrix.



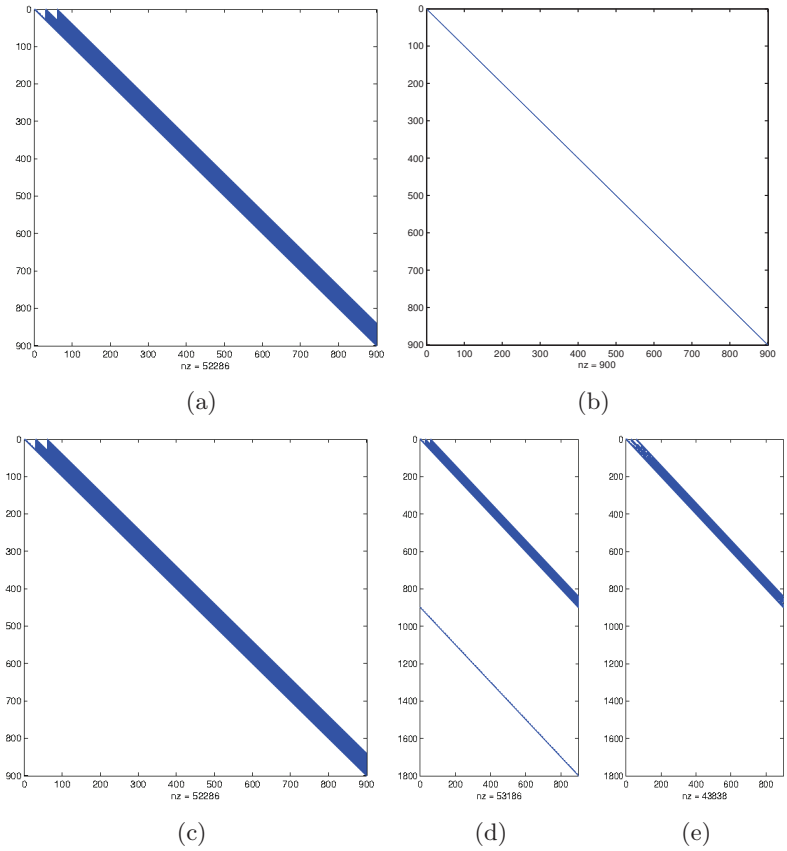
(a) (b)

Figure 13: Sparsity patterns of $\boldsymbol{Q}$ (a) and $\boldsymbol{Qq}$ (b) for Toeplitz matrix.



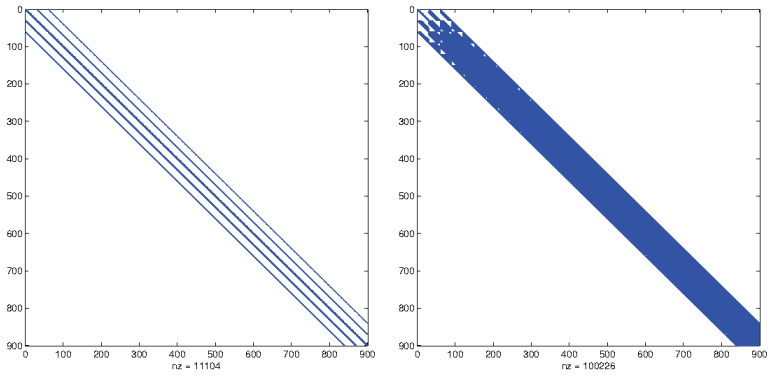
(a) (b) (c)

Figure 14: Images of the true covariance matrix $\boldsymbol{\Sigma}$ (a), the approximated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (b) and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for Toeplitz matrix.



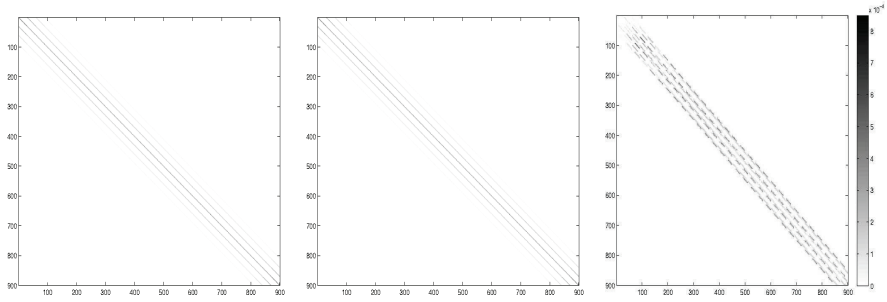
(a) (b) (c) (d) (e)

Figure 15: Sparsity pattern for $\boldsymbol{L}_1$(a), $\boldsymbol{L}_1$(b), $\boldsymbol{L}_1$(c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$(e) for the random field from the SPDE (25).



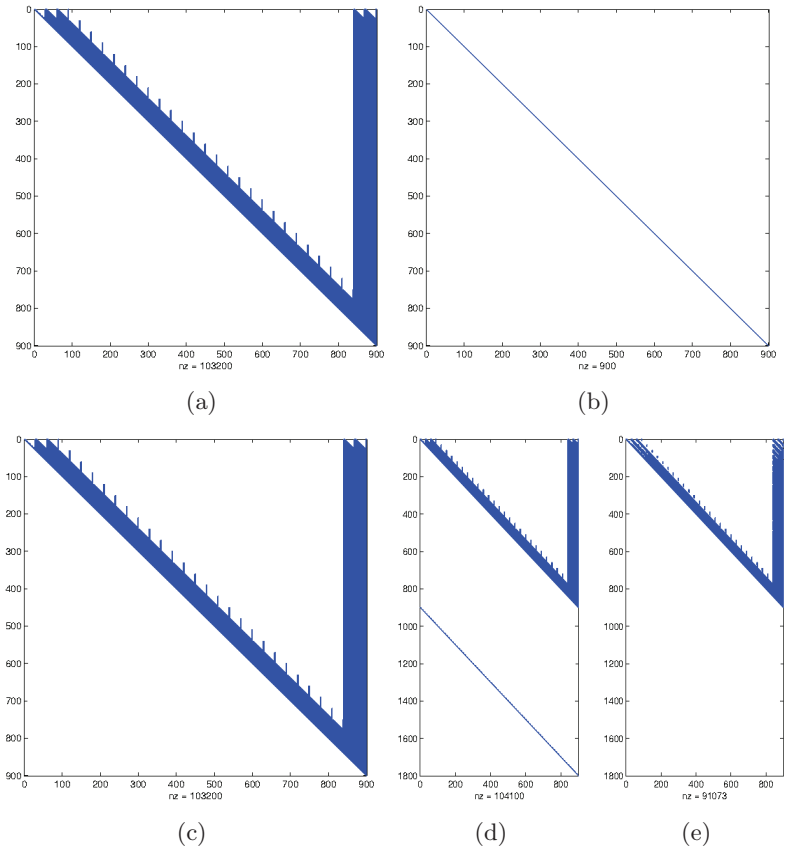
(a) (b)

Figure 16: Sparsity patterns of $\boldsymbol{Q}$ (a) and $\boldsymbol{Qq}$ (b) for the random field generated from the SPDE (25).



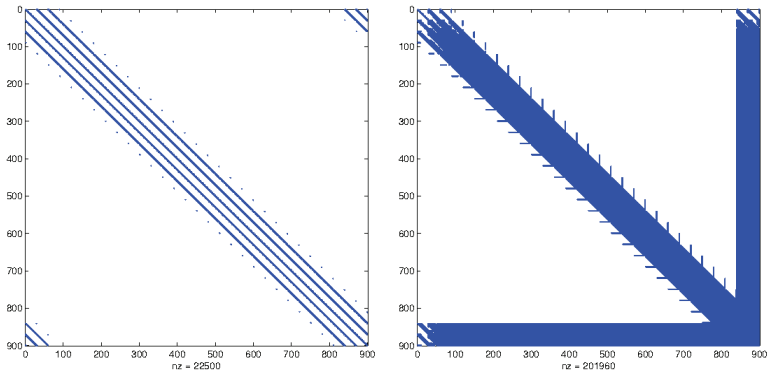
(a) (b) (c)

Figure 17: Images of the true covariance matrix $\boldsymbol{\Sigma}$ (a), the approxiamted covariance matrix$\widetilde{\boldsymbol{\Sigma}}$ (b) and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for the random field generated from the SPDE (25).



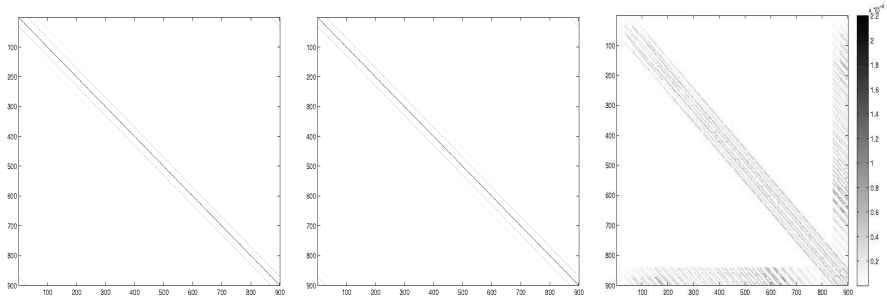
(a) (b) (c) (d) (e)

Figure 18: Sparsity pattern for $\boldsymbol{L}_1$(a), $\boldsymbol{L}_1$(b), $\boldsymbol{L}_1$(c), $\boldsymbol{A}$ (d) and $\boldsymbol{R}$(e) for the random field generated from SPDE given in (26).



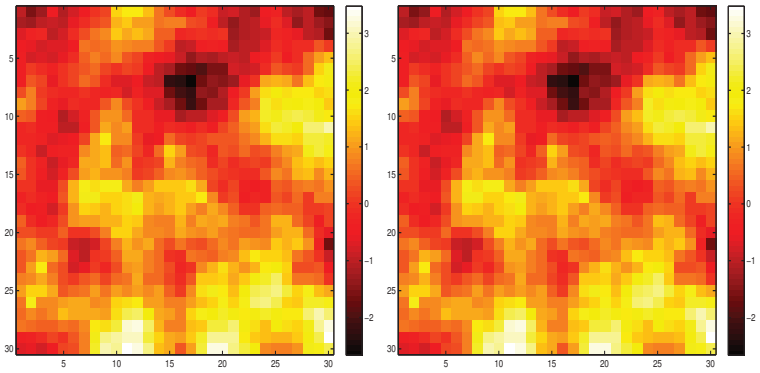
(a) (b)

Figure 19: Sparsity patterns of $\boldsymbol{Q}$ (a) and $\boldsymbol{Qq}$ (b) for the random field generated from SPDE given in (26).



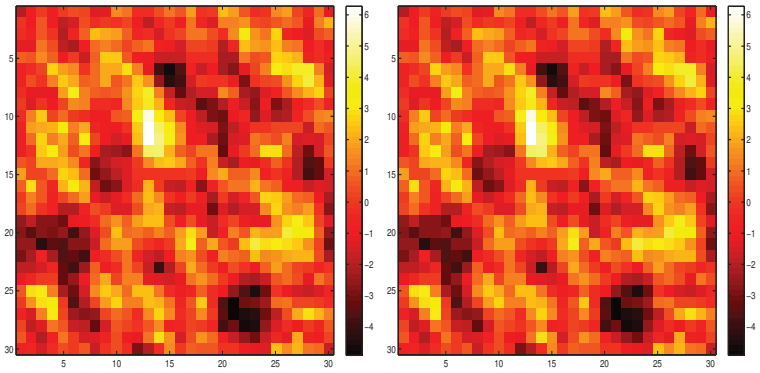
(a) (b) (c)

Figure 20: Images of the true covariance matrix $\boldsymbol{\Sigma}$ (a), the approxiamted covariance matrix$\widetilde{\boldsymbol{\Sigma}}$ (b) and the error matrix $\widetilde{\boldsymbol{E}}$ (c) for the random field generated from SPDE given in (26).



(a) (b)

Figure 21: Samples from the GMRF with the common Cholesky factor $\boldsymbol{L}$ with (25) (a) and the upper triangular matrix $\boldsymbol{R}$ (b) from the cTIGO algorithm.



(a) (b)

Figure 22: Samples from the GMRF with the common Cholesky factor $\boldsymbol{L}$ (a) with (26) and the upper triangular matrix $\boldsymbol{R}$ (b) from the cTIGO algorithm.

References

O. Axelsson. *Iterative solution methods*. Cambridge Univ Pr, 1996.

Z.Z. Bai and J.F. Yin. Modified incomplete orthogonal factorization methods using Givens rotations. *Computing*, 86(1):53–69, 2009. ISSN 0010-485X.

Z.Z. Bai, I.S. Duff, and A.J. Wathen. A class of incomplete orthogonal factorization methods. I: Methods and theories. *BIT Numerical Mathematics*, 41(1): 53–70, 2001. ISSN 0006-3835.

Z.Z. Bai, I.S. Duff, and J.F. Yin. Numerical study on incomplete orthogonal factorization preconditioners. *Journal of Computational and Applied Mathematics*, 226(1):22–41, 2009. ISSN 0377-0427.

Å Björck. *Numerical methods for least squares problems*. Society for Industrial Mathematics, 1996. ISBN 0898713609.

Catherine Forbes, Merran Evans, Nicholas Hastings, and Brian Peacock. *Statistical distributions*. Wiley, 4 edition, 2011.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, NTNU, 2011.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins Univ Pr, 1996. ISBN 0801854148.

X. Hu, D.P. Simpson, F. Lindgren, and H. Rue. Multivariate gaussian random fields using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012.

A. Jennings and MA Ajiz. Incomplete Methods for Solving $A^T Ax = b$. *SIAM Journal on Scientific and Statistical Computing*, 5:978, 1984.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

J.A. Meijerink and H.A. van der Vorst. Guidelines for the usage of incomplete decompositions in solving sets of linear equations as they occur in practical problems. *Journal of computational physics*, 44(1):134–155, 1981.

N. Munksgaard. Solving sparse symmetric sets of linear equations by preconditioned conjugate gradients. *ACM Transactions on Mathematical Software (TOMS)*, 6(2):206–219, 1980. ISSN 0098-3500.

A.T. Papadopoulos, I.S. Duff, and A.J. Wathen. A class of incomplete orthogonal factorization methods. II: Implementation and results. *BIT Numerical Mathematics*, 45(1):159–179, 2005. ISSN 0006-3835.

H. Rue. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338, 2001. ISSN 1467-9868.

H. Rue. Marginal variances for Gaussian Markov random fields. *Statistics Report*, 2005.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue, I. Steinsland, and S. Erland. Approximating hidden Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):877–892, 2004. ISSN 1467-9868.

Y. Saad. Preconditioning techniques for nonsymmetric and indefinite linear systems* 1. *Journal of Computational and Applied Mathematics*, 24(1-2):89–105, 1988. ISSN 0377-0427.

Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial Mathematics, 2003. ISBN 0898715342.

D.P. Simpson. *Krylov subspace methods for approximating functions of symmetric positive definite matrices with applications to applied statistics and anomalous diffusion*. PhD thesis, Queensland University of Technology, 2008.

L.N. Trefethen and D. Bau. *Numerical linear algebra*. Society for Industrial Mathematics, 1997. ISBN 0898713617.

X. Wang, K.A. Gallivan, and R. Bramley. Cimgs: An incomplete orthogonal factorization preconditioner. *SIAM Journal on Scientific Computing*, 18(2): 516–536, 1997.

H.T. Wist and H. Rue. Specifying a Gaussian Markov random field by a sparse Cholesky triangle. *Communications in Statistics-Simulation and Computation*, 35(1):161–176, 2006. ISSN 0361-0918.