ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)
82, *Part* 1, *pp.* 215–239

Multivariate type G Matérn stochastic partial differential equation random fields

David Bolin

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, and University of Gothenburg, Sweden

and Jonas Wallin

Lund University, Sweden

[Received February 2018. Revised October 2019]

Summary. For many applications with multivariate data, random-field models capturing departures from Gaussianity within realizations are appropriate. For this reason, we formulate a new class of multivariate non-Gaussian models based on systems of stochastic partial differential equations with additive type G noise whose marginal covariance functions are of Matérn type. We consider four increasingly flexible constructions of the noise, where the first two are similar to existing copula-based models. In contrast with these, the last two constructions can model non-Gaussian spatial data without replicates. Computationally efficient methods for likelihood-based parameter estimation and probabilistic prediction are proposed, and the flexibility of the models suggested is illustrated by numerical examples and two statistical applications.

Keywords: Matérn covariances; Multivariate random fields; Non-Gaussian models; Spatial statistics; Stochastic partial differential equations

1. Introduction

Motivated by an increasing number of spatial data sets with multiple measured variables, such as different climate variables from weather stations, various pollutants monitored in urban areas or climate model outputs, the literature on models for multivariate random fields is growing rapidly. The majority of research in this area has focused on Gaussian random fields, and how to construct valid multivariate cross-covariance functions.

Of particular interest has been multivariate extensions of the Matérn correlation function (Matérn, 1960), $M(\mathbf{h}|\kappa,\nu)=2^{1-\nu}\Gamma(\nu)^{-1}(\kappa\|\mathbf{h}\|)^{\nu}K_{\nu}(\kappa\|\mathbf{h}\|)$, $\mathbf{h}\in\mathbb{R}^d$. Here K_ν is a modified Bessel function of the second kind and the positive parameters κ and ν determine the practical correlation range and smoothness of the process respectively. Gneiting *et al.* (2010) extended it to the multivariate setting by proposing a model with cross-correlation functions $\rho_{ij}M(\mathbf{h}|\kappa_{ij},\nu_{ij})$, where ρ_{ij} are parameters determining the cross-correlations between the ith and jth component of the multivariate field. The parameters in this construction must be restricted to assure that it is a valid multivariate covariance function, and Gneiting *et al.* (2010) proposed two models that satisfied this requirement: a parsimonious model, where $\kappa_{ij}\equiv\kappa$ and $\nu_{ij}=(\nu_{ii}+\nu_{jj})/2$, and a more general bivariate model that was later extended by Apanasovich *et al.* (2012).

Address for correspondence: David Bolin, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia.
E-mail: david.bolin@kaust.edu.sa

 1369–7412/20/82215

Even though most research has focused on Gaussian random fields, many data sets have features that cannot be captured by Gaussian models, such as exponential tails, non-Gaussian dependence, or asymmetric marginal distributions. There is thus a need for multivariate random fields that are more general than the Gaussian random field. Examples of such models in the literature are multivariate max-stable processes for spatial extremes (Genton *et al.*, 2015) and Mittag-Leffler random fields (Ma, 2013a). A common approach for constructing non-Gaussian fields is to multiply a Gaussian random field with a random scalar. Multivariate versions of this approach were explored by Ma (2013b) and Du *et al.* (2012). Copula-based modelling is another popular method for non-Gaussian data, which has been used for creating both univariate (Gräler, 2014; Bárdossy, 2006) and multivariate (Krupskii *et al.*, 2018) random fields.

However, creating non-Gaussian multivariate random-field models that allow for likelihood-based parameter estimation and probabilistic prediction is difficult, especially if they should be able to capture interesting departures from normality within realizations, and not just have non-Gaussian marginal distributions. This requirement excludes fields that are non-Gaussian only in the presence of repeated measurements, such as the factor copula models (Krupskii *et al.*, 2018) and the constructions based on multiplying Gaussian fields with random scalars. Many other copula-based approaches in geostatistics use Gaussian copulas. The resulting models are then equivalent to transformed Gaussian models (Kazianka and Pilz, 2010), which have many disadvantages (Wallin and Bolin, 2015). Thus, most existing approaches are either too limited, in the sense that they cannot capture essential features such as sample path asymmetry, or they lack methods for practical applications. For this reason, the recent review on multivariate random fields by Genton and Kleiber (2015) listed the creation of practically useful non-Gaussian multivariate random fields as an open problem.

The main contribution of this work is to present a class of models that remedies this problem. The model class is constructed by using systems of stochastic partial differential equations (SPDEs) driven by non-Gaussian noise. To facilitate computationally efficient likelihood-based inference, we use noise with normal variance mixture distributions (Barndorff-Nielsen *et al.*, 1982), which we refer to as type G noise. The restriction to normal variance mixtures is not a big limitation, since several common distributions can be formulated in this way. Four increasingly flexible constructions are considered, where the simplest is closely related to factor copula models and the approach where a Gaussian field is multiplied with a random scalar. The more flexible constructions allow the fields to capture more complex dependence structures and departures from Gaussianity within realizations, while still allowing for likelihood-based inference. As an additional motivation for the more flexible constructions, we investigate the properties of spatial prediction based on type G models, and in particular we prove that distributions of spatial predictions for the simplest construction are asymptotically Gaussian. This means that, if the goal is to use the model for spatial prediction, we might as well use a Gaussian model instead of the simple non-Gaussian constructions.

In the seminal work of Lindgren *et al.* (2011), Gaussian random fields were formulated as solutions to SPDEs, which were approximated by using a finite element (FE) discretization to allow for computationally efficient inference. Hu *et al.* (2013) and Hu and Steinsland (2016) extended the work to multivariate Gaussian random-field models based on systems of SPDEs. However, their models in general do not have explicit covariance functions, which can complicate the understanding of the effect that each model parameter has, in particular in the non-Gaussian case. To avoid this problem, we formulate systems of SPDEs that result in models with marginal Matérn covariance functions, having the parsimonious Matérn model as a special case. We further discover a set of parameters in the model formulation that do not affect the covariance

function, and therefore are unidentifiable for Gaussian models. These parameters, however, control the more complex dependence for non-Gaussian models.

As always with more general models than the Gaussian model, there is an added computational cost for inference. However, using FE discretizations of our non-Gaussian models allows for the same computational complexity with respect to the size of the discretized random field as for the corresponding Gaussian models. This makes the models applicable in scenarios where the data sets are so large that it prohibits the use of standard covariance-based models. As in the Gaussian case, the SPDE approach also facilitates extensions to non-stationary models by using spatially varying parameters. An important fact related to this is that the construction of the FE approximation is identical to that for Gaussian models and thus as easy to compute.

The paper is structured as follows. In Section 2, the link between systems of SPDEs and cross-covariances is studied. Section 3 contains the definitions of the non-Gaussian models, as well as derivations of basic model properties. More details and examples of multivariate normal–inverse Gaussian (NIG) fields, which are a special case of the type G models, are given in Section 4. In Section 5, the type G fields are included in a geostatistical model for which we derive computationally efficient methods for likelihood-based parameter estimation and probabilistic prediction. Section 6 presents two applications, and the paper concludes with a discussion in Section 7. The on-line supplementary materials contain five appendices that present

(a) details on the FE discretizations (section A),
(b) gradients needed for the parameter estimation (section B),
(c) sampling methods for the models (section C),
(d) details on the applications (section D) and
(e) all proofs (section E).

The methods that are developed in this work have been implemented in the R package `ngme`. The programs that were used to analyse the data can be obtained from

`https://rss.onlinelibrary.wiley.com/hub/journal/14679868/series-b-datasets`.

2. Multivariate Matérn fields and systems of stochastic partial differential equations

A Gaussian random field $x(\mathbf{s})$ on $\mathbb{R}^d$ with a Matérn covariance function can be represented as a stationary solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} x = \dot{\mathcal{W}}, \qquad \text{in } \mathcal{D} := \mathbb{R}^d, \tag{1}$$

where $\alpha = \nu + d/2$, Δ is the Laplacian and $\dot{\mathcal{W}}$ is Gaussian white noise (Whittle, 1963). Extending equation (1) to a system of SPDEs can be used to define more general covariance models (Bolin and Lindgren, 2011) and to define multivariate random fields. Hu *et al.* (2013) and later Hu and Steinsland (2016) proposed the use of systems of the form

$$\begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} & \dots & \mathcal{K}_{1p} \\ \mathcal{K}_{21} & \mathcal{K}_{22} & \dots & \mathcal{K}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}_{p1} & \mathcal{K}_{p2} & \dots & \mathcal{K}_{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} \dot{\mathcal{W}}_1 \\ \dot{\mathcal{W}}_2 \\ \vdots \\ \dot{\mathcal{W}}_p \end{pmatrix}, \tag{2}$$

to construct multivariate random fields, $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^{\mathrm{T}}$, where $\mathcal{K}_{ij}$ are pseudo-

differential operators such as $(\kappa^2-\Delta)^{\alpha/2}$ and $\dot{\mathcal{W}}_1,\ldots,\dot{\mathcal{W}}_p$ are mutually independent Gaussian white noise processes. Hu *et al.* (2013) focused on the bivariate triangular system

$$\begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} \\ & \mathcal{K}_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \dot{\mathcal{W}}_1 \\ \dot{\mathcal{W}}_2 \end{pmatrix}, \tag{3}$$

where $\mathcal{K}_{ij}=(\kappa_{ij}^2-\Delta)^{\alpha_{ij}/2}$. To understand the cross-covariance function for this model better, we can informally invert the operator matrix to obtain

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mathcal{K}_{11}^{-1} & -\mathcal{K}_{11}^{-1}\mathcal{K}_{12}\mathcal{K}_{22}^{-1} \\ & \mathcal{K}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \dot{\mathcal{W}}_1 \\ \dot{\mathcal{W}}_2 \end{pmatrix}. \tag{4}$$

From this representation we can see that x_2 is marginally a Gaussian Matérn field whereas x_1 is a sum of two Gaussian fields $\mathcal{K}_{11}^{-1}\dot{\mathcal{W}}_1$ and $-\mathcal{K}_{11}^{-1}\mathcal{K}_{12}\mathcal{K}_{22}^{-1}\dot{\mathcal{W}}_2$ and thus has a more complicated covariance function.

Although the full system (2) may be of interest, the generality comes at the cost of a large number of parameters that are difficult to identify in practice, and equally difficult to estimate. We therefore focus on the case when all marginal covariances are Matérn, and on characterizing systems of SPDEs that result in models with this property.

2.1. Multivariate Matérn stochastic partial differential equation fields

To make the results in this section applicable beyond Gaussian models, we replace the right-hand side of equation (2) by $\dot{\boldsymbol{\mathcal{M}}}=(\dot{\mathcal{M}}_1,\ldots,\dot{\mathcal{M}}_p)^{\mathrm{T}}$, where the components are mutually uncorrelated, but not necessarily independent, L_2-valued independently scattered random measures (see Section 4 and Rajput and Rosinski (1989) for details). This includes Gaussian noise but also the non-Gaussian processes that we shall study in the next section. We introduce the operator matrix $\boldsymbol{\mathcal{K}}$ with entries $\boldsymbol{\mathcal{K}}_{ij}=\mathcal{K}_{ij}$ and write equation (2) more compactly as $\boldsymbol{\mathcal{K}}\mathbf{x}=\dot{\boldsymbol{\mathcal{M}}}$. Investigating equation (4), we can note that x_1 has a Matérn covariance function if $\mathcal{K}_{12}=\mathcal{K}_{22}$. This motivates the following definition of p-variate Matérn SPDE fields.

Definition 1. A multivariate Matérn SPDE field on $\mathbb{R}^d$ is a solution to $\boldsymbol{\mathcal{K}}\mathbf{x}=\dot{\boldsymbol{\mathcal{M}}}$ where the operator matrix is of the form $\boldsymbol{\mathcal{K}}=\mathbf{D}\,\mathrm{diag}(\mathcal{L}_1,\ldots,\mathcal{L}_p)$. Here $\mathbf{D}$ is a real invertible $p\times p$ matrix and $\mathcal{L}_i=(\kappa_i^2-\Delta)^{\alpha_i/2}$ with $\kappa_i>0$ and $\alpha_i>d/2$ for $i=1,\ldots,p$.

Since $\mathbf{D}$ defines the dependence structure of the process, we refer to it as a dependence matrix. That the multivariate Matérn SPDE model indeed has marginal Matérn covariance functions is clarified in the following proposition.

Proposition 1. Given that the driving noise in $\boldsymbol{\mathcal{K}}\mathbf{x}=\dot{\boldsymbol{\mathcal{M}}}$ has unit variance, the multivariate Matérn SPDE field $\mathbf{x}(\mathbf{s})$ on $\mathbb{R}^d$ has covariance function

$$\mathrm{cov}\{x_i(\mathbf{s}),x_j(\mathbf{t})\}=\begin{cases} \dfrac{\Gamma(\nu_i)\sum_{j=1}^p R_{ii}^2}{\Gamma(\alpha_i)(4\pi)^{d/2}\kappa_i^{2\nu_i}} M(\|\mathbf{s}-\mathbf{t}\||\kappa_i,\nu_i) & i=j, \\ \mathcal{F}^{-1}(S_{ij})(\|\mathbf{s}-\mathbf{t}\|) & i\neq j, \end{cases}$$

where R_{ij} are the elements of $\mathbf{R}=\mathbf{D}^{-1}$, $\mathcal{F}^{-1}$ denotes the inverse Fourier transform and

$$S_{ij}(\mathbf{k})=\frac{\sum_{l=1}^p R_{il}R_{jl}}{(2\pi)^d}\frac{1}{(\kappa_i^2+\|\mathbf{k}\|^2)^{\alpha_i/2}(\kappa_j^2+\|\mathbf{k}\|^2)^{\alpha_j/2}}. \tag{5}$$

Note that $\mathbf{D}$ determines the strength of the cross-correlations, and that $\text{cov}\{x_i(\mathbf{s}), x_j(\mathbf{t})\}$ for $i \neq j$ is a Matérn covariance function only if $\kappa_j = \kappa_j$. In the case when $\kappa_i = \kappa$ for all i, the model coincides with the parsimonious Matérn model by Gneiting *et al.* (2010). Also note that the shapes of the cross-correlation functions are determined by the parameters of the marginal correlation functions. This is slightly more restrictive than the general covariance-based multivariate Matérn models but has the advantage that there are no difficult-to-check restrictions on the model parameters. Furthermore, both Gneiting *et al.* (2010) and Apanasovich *et al.* (2012) argued that the most important aspect of multivariate models is to allow for flexibility in the marginal covariances while still allowing for some degree of cross-covariance. Thus, the Matérn SPDE model should be a sufficiently flexible alternative to multivariate Matérn fields for most applications.

Remark 1. An immediate consequence of definition 1 is that the field alternatively can be obtained as a solution to a diagonal system of SPDEs driven by correlated noise: $\text{diag}(\mathcal{L}_1, \ldots, \mathcal{L}_p)\mathbf{x}(\mathbf{s}) = \dot{\mathcal{M}}_R$, where $\dot{\mathcal{M}}_R = \mathbf{R}\dot{\mathcal{M}}$ and $\mathbf{R} = \mathbf{D}^{-1}$. This means that the model can be viewed as a linear model of coregionalization.

2.2. Parameterizing the model

An important question for practical applications of the multivariate Matérn SPDE fields is if the model parameters (the dependence matrix and the parameters of the operators) are identifiable. The following proposition shows that this is not so in general.

Proposition 2. Two multivariate Matérn SPDE fields, with the same operators $\mathcal{L}_1, \ldots, \mathcal{L}_p$ and with dependence matrices $\mathbf{D}$ and $\hat{\mathbf{D}}$, have equal covariance functions if and only if $\mathbf{D} = \mathbf{Q}\hat{\mathbf{D}}$ for an orthogonal matrix $\mathbf{Q}$. For any choice of $\mathbf{D}$, one can find a triangular matrix $\hat{\mathbf{D}}$ that gives the same covariance functions. In particular, $\hat{\mathbf{D}} = \text{chol}(\mathbf{D}^{\mathrm{T}}\mathbf{D})$ is the unique upper triangular choice with positive diagonal elements.

We shall refer to models with triangular dependence matrices as triangular Matérn SPDE fields. Since Gaussian fields are uniquely specified by the first two moments, the proposition implies that the matrix $\mathbf{D}$ is not completely identifiable from data for Gaussian models, so there is no point in considering non-triangular Gaussian models. This is, however, not so for non-Gaussian models, where non-triangular dependence matrices can be used to define more general dependence structures.

Since the dependence matrix is not completely identifiable for Gaussian models, a different model parameterization that separates the control of marginal variances, cross-correlations and higher moments is preferable. To derive such a parameterization, we use proposition 2 to write $\mathbf{D} = \mathbf{Q}_p\mathbf{D}_l$, where $\mathbf{D}_l$ is a triangular matrix and $\mathbf{Q}_p$ is an orthogonal matrix. Then $\mathbf{D}_l$ and $\mathbf{Q}_p$ respectively determine the cross-covariances and the higher moments. To separate the control of the variances and cross-correlations, we rescale the operators $\mathcal{L}_i$ by constants $c_i = \sqrt{\{\sigma_i^{-2}(4\pi)^{-d/2}\kappa_i^{-2\nu_i}\Gamma(\nu_i)/\Gamma(\alpha_i)\}}$ and parameterize $\mathbf{D}_l$ as

$$\mathbf{D}_l(\boldsymbol{\rho}) = \begin{pmatrix} 1 & & & & \\ \rho_{1,1} & 1 & & & \\ \rho_{2,1} & \rho_{2,2} & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \rho_{p,1} & \rho_{p,2} & \cdots & \rho_{p,p-1} & 1 \end{pmatrix}^{-1} \text{diag}\{1, k_2(\boldsymbol{\rho}), k_3(\boldsymbol{\rho}), \ldots, k_p(\boldsymbol{\rho})\},$$

where $k_j(\boldsymbol{\rho}) = \sqrt{(1 + \Sigma_{i<j}\rho_{j,i}^2)}$. With this parameterization, $\boldsymbol{\rho} \in \mathbb{R}^{p(p-1)/2}$ controls the cross-

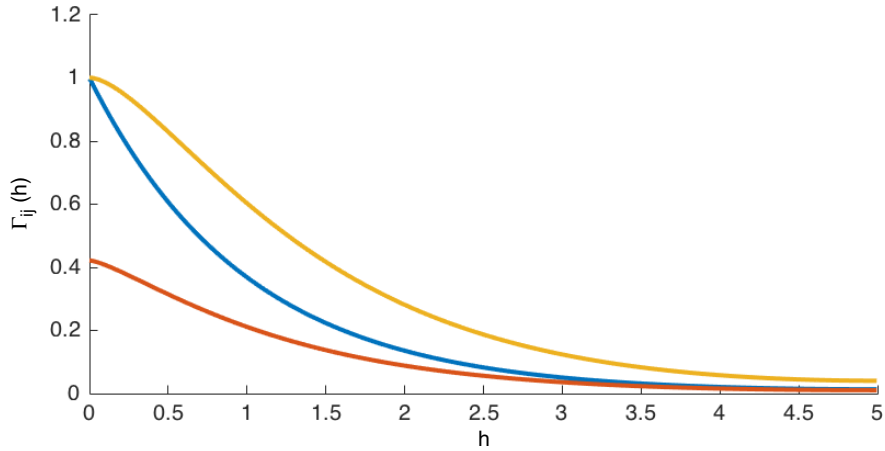

Fig. 1. Example of covariance functions $\Gamma_{ij}(h) = \Gamma_{ij}(\|\mathbf{s}-\mathbf{t}\|) = \text{cov}\{X_i(\mathbf{s}), X_j(\mathbf{t})\}$ for the solution to the triangular Matérn SPDE with $\sigma_1 = \sigma_2 = 1$, $\rho = 0.5$, $\kappa_1 = \kappa_2 = 1$, $\alpha_1 = 1.5$ and $\alpha_2 = 2$: ——, $\Gamma_{11}(h)$; ——, $\Gamma_{12}(h)$; ——, $\Gamma_{22}(h)$

correlations and $\sigma_i^2 = V\{X_i(\mathbf{s})\}$. Fig. 1 shows an example of the resulting covariance function for a bivariate model with $\rho = \rho_{1,1} = 0.5$.

What remains is to find a parameterization of $\mathbf{Q}_p$. The determinant of an orthogonal matrix is ± 1, where the sign is not identifiable in general. It is therefore enough to consider the subclass of special orthogonal matrices, which have determinant 1. For a general p, it is difficult to parameterize such matrices. However, for $p = 2$ and $p = 3$ we can use the fact that they are equivalent to rotation matrices. We can therefore write

$$\mathbf{Q}_2(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$
$$\mathbf{Q}_3(\theta_1, \theta_2, \theta_3) = \mathbf{Q}_{3x}(\theta_1)\mathbf{Q}_{3y}(\theta_2)\mathbf{Q}_{3z}(\theta_3),$$

where $\theta \in [0, 2\pi]$, $\mathbf{Q}_{3x}(\theta) = \text{diag}\{\mathbf{Q}_2(\theta), 1\}$, $\mathbf{Q}_{3z}(\theta) = \text{diag}\{1, \mathbf{Q}_2(\theta)\}$ and

$$\mathbf{Q}_{3y}(\theta) = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}.$$

To summarize, we use the parameterization

$$\mathbf{D}(\boldsymbol{\theta}, \boldsymbol{\rho})\text{diag}(c_1\mathcal{L}_1, \ldots, c_p\mathcal{L}_p)\mathbf{x}(\mathbf{s}) = \dot{\boldsymbol{\mathcal{M}}}, \tag{6}$$

where $\mathbf{D}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \mathbf{Q}_p(\boldsymbol{\theta})\mathbf{D}_l(\boldsymbol{\rho})$ and $\boldsymbol{\theta} \in [0, 2\pi]^{p(p-1)/2}$ will control higher moments for non-Gaussian models. In the bivariate case, the dependence matrix simplifies to

$$\mathbf{D}(\theta, \rho) = \begin{pmatrix} \cos(\theta) + \rho\sin(\theta) & -\sin(\theta)\sqrt{(1+\rho^2)} \\ \sin(\theta) - \rho\cos(\theta) & \cos(\theta)\sqrt{(1+\rho^2)} \end{pmatrix}. \tag{7}$$

3. Type G Matérn stochastic partial differential equation fields

In this section, the multivariate Matérn SPDE model is extended beyond Gaussianity by replacing the Gaussian noise with non-Gaussian noise. In Section 3.1, four different constructions of noise for this approach are introduced and the resulting Matérn SPDE fields are discussed.

The differences between the four constructions are illustrated by using FE discretizations of the models in Section 3.2 and further properties of the models are stated in Section 3.3. Finally, asymptotic properties of spatial prediction based on the simplest type G models are derived in Section 3.4.

3.1. Four increasingly flexible constructions

The four constructions are based on using different types of normal variance mixtures

$$\boldsymbol{\gamma} + v\boldsymbol{\mu} + \sqrt{v}z, \tag{8}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^p$ and $\boldsymbol{\mu} \in \mathbb{R}^p$ are parameters, $z \sim N(0,1)$ and v is a non-negative random variable. Inspired by Lévy processes, which are said to be of type G if their increments are normal variance mixtures, we shall refer to these models as type G Matérn SPDE fields.

The first two constructions are related to the approach where non-Gaussian fields are obtained by multiplying Gaussian fields with random scalars.

Definition 2. Let v and $v_1, \ldots, v_p$ be independent non-negative infinitely divisible random variables and set $\mathbf{v}_1 = v\mathbf{1}_p$ and $\mathbf{v}_2 = (v_1, \ldots, v_p)^{\mathrm{T}}$, where $\mathbf{1}_p$ denotes a vector with p 1s. Further, let $\mathcal{W}(\mathbf{s}) = (\mathcal{W}_1(\mathbf{s}), \ldots, \mathcal{W}_p(\mathbf{s}))^{\mathrm{T}}$ be a vector of independent copies of Brownian sheets on $\mathbb{R}^d$. For $i \in \{1,2\}$, a type G_i Matérn SPDE field is obtained by using $\dot{\boldsymbol{\mathcal{M}}}_i$ in expression (6) where $\dot{\boldsymbol{\mathcal{M}}}_i(\mathbf{s}) = \boldsymbol{\gamma} + \mathrm{diag}(\mathbf{v}_i)\boldsymbol{\mu} + \mathrm{diag}(\sqrt{\mathbf{v}_i})\boldsymbol{\mathcal{W}}(\mathbf{s})$.

It should be noted that $\dot{\boldsymbol{\mathcal{M}}}_1$ and $\dot{\boldsymbol{\mathcal{M}}}_2$ are not independently scattered measures since they have common random scaling $\mathbf{v}_i$ for different spatial locations. Because of this, we cannot directly use the results from the previous section. However, the results can easily be extended by allowing for measures that are independently scattered conditionally on a random variable $\mathbf{v}$. In particular, if we restrict the distribution of $\mathbf{v}_i$ such that $E(\mathbf{v}_i) = \mathbf{1}$, then the result in proposition 1 still holds in the symmetric case with $\boldsymbol{\mu} = \mathbf{0}$. In the non-symmetric case, the resulting fields have covariance functions given by the covariance function in proposition 1 plus a constant factor depending on the variance of $\mathbf{v}$. With the restriction $E(\mathbf{v}_i) = \mathbf{1}$, the mean of the process is given by $\boldsymbol{\mu} + \boldsymbol{\gamma}$, and we therefore set $\boldsymbol{\gamma} = -\boldsymbol{\mu}$ to ensure that the process has zero mean as default. In the type G_1 model, we can interpret v as a random scaling of the variance of the entire process, whereas we scale the variance of each $x_i(\mathbf{s})$ separately with v_i in the type G_2 case. When $\boldsymbol{\mu} \neq \mathbf{0}$, v also decides the skewness of the marginal distributions of $\mathbf{x}(\mathbf{s})$ in the type G_1 case, whereas v_i controls the skewness of $x_i(\mathbf{s})$ in the type G_2 case. Hence, the type G_2 model gives more control of the marginal distributions of the process.

From a Bayesian point of view, one could interpret $\pi(\mathbf{v})$ as a prior distribution on the mean and variance of a multivariate Gaussian random field. Thus, these models can be used in the same way as the Gaussian models in a Bayesian setting, but where we have a specific prior that links the mean and variance of the field. From this point of view, we would probably not refer to these models as non-Gaussian.

The next two constructions are based on type G Lévy noise. A random variable x is said to be of type G if it can be written as $x =^{\mathrm{d}} \sqrt{v}z$, where z is a Gaussian variable and v is an infinitely divisible non-negative random variable. A univariate type G Lévy process is a Lévy process whose increments are of type G. Rosiński (1991) showed that a type G process $\mathcal{M}(s)$, $s \in [0,1]$, with $\mathcal{M}(1) =^{\mathrm{d}} \sqrt{v}z$ can be represented as $\mathcal{M}(s) = \Sigma_{k=1}^{\infty} z_k g(e_k)^{1/2} \mathbb{I}(s \geqslant u_k)$, where e_k are the points of a unit rate Poisson process on $\mathbb{R}^+$, z_k are independent and identically distributed $N(0,1)$ random variables, and u_k are independent and identically distributed uniform random variables on $(0,1)$. The function g is the generalized inverse of the tail Lévy measure for v, defined as $g(u) = \inf\{x > 0 : M(x, \infty) \leqslant u\}$ where M is the Lévy measure of v. The non-decreasing Lévy

process $v(\mathbf{s}) = \Sigma_{k=1}^{\infty} g(e_k)\mathbb{I}(s \geqslant u_k)$ has the same Lévy measure as v and can be used to represent $\mathcal{M}$ as a subordinated Wiener process. We refer to Rosiński (1991) for further technical details on the construction. In the spatial case, a type G process $\mathcal{M}(\mathbf{s})$ on the unit square $D = [0,1] \times [0,1]$ with $M(\mathbf{1}) =^{\mathrm{d}} \sqrt{v}z$ can similarly be represented as $\mathcal{M}(\mathbf{s}) = \Sigma_{k=1}^{\infty} z_k g(e_k)^{1/2}\mathbb{I}(\mathbf{s} \geqslant \mathbf{u}_k)$, where $\mathbf{u}_k$ now are uniform random variables on D and $\mathbb{I}(\mathbf{s} \geqslant \mathbf{u}_k) = \mathbb{I}(s_1 \geqslant u_{k,1})\mathbb{I}(s_2 \geqslant u_{k,2})$ is a two-dimensional indicator function. In this case, the associated process $v(\mathbf{s}) = \Sigma_{k=1}^{\infty} g(e_k)\mathbb{I}(\mathbf{s} \geqslant \mathbf{u}_k)$ can no longer be seen as a subordinator, but it could informally be thought of as a process that determines the variance of the noise. For multivariate processes, there are two natural extensions to vector-valued noise that we use to define type G_3 and type G_4 fields.

Definition 3. Let $\mathcal{M}(\mathbf{s})$ be a type G Lévy process with $v(\mathbf{s}) = \Sigma_{k=1}^{\infty} g(e_k)\mathbb{I}(\mathbf{s} \geqslant \mathbf{u}_k)$ and let $\boldsymbol{\mathcal{M}}(\mathbf{s}) = (\mathcal{M}_1(\mathbf{s}), \ldots, \mathcal{M}_p(\mathbf{s}))^{\mathrm{T}}$ be a vector of independent type G Lévy processes with corresponding variance processes $\mathbf{v}(\mathbf{s}) = (v_1(\mathbf{s}), \ldots, v_p(\mathbf{s}))^{\mathrm{T}}$. For $i \in \{3,4\}$ a type G_i Matérn SPDE field is obtained by using $\dot{\boldsymbol{\mathcal{M}}}_i$ in expression (6) where

$$\boldsymbol{\mathcal{M}}_3(\mathbf{s}) = \boldsymbol{\gamma} + \boldsymbol{\mu} v(\mathbf{s}) + \sum_{k=1}^{\infty} g(e_k)^{1/2}\mathbb{I}(\mathbf{s} \geqslant \mathbf{u}_k)\mathbf{z}_k,$$

$$\boldsymbol{\mathcal{M}}_4(\mathbf{s}) = \boldsymbol{\gamma} + \mathrm{diag}(\boldsymbol{\mu})\mathbf{v}(\mathbf{s}) + \boldsymbol{\mathcal{M}}(\mathbf{s}).$$

Remark 2. In this section we have assumed a multivariate setting, i.e. $p > 1$. However, in the univariate case, the type G_1 and type G_2 Matérn SPDE models are equivalent. Further, if $\boldsymbol{\mu} = 0$, the type G_1 model is a Gaussian Matérn field multiplied by a univariate positive random variable. Thus, models such as the t-distributed random fields by Røislien and Omre (2006) belong to the class of type G_1 fields. Also, when $p = 1$ the type G_3 and type G_4 Matérn SPDE models are also equivalent and coincide with the models in Wallin and Bolin (2015).

3.2. Understanding the four constructions through their discretizations

Although the Matérn SPDE models were formulated on the entire $\mathbb{R}^d$ in Section 2, we consider the system of SPDEs on a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ when implementing them numerically. The operators are then equipped with suitable boundary conditions and the solution is approximated by using an FE discretization derived in the on-line appendix A. To understand the differences between the four type G constructions, we now examine the properties of the discretized models in comparison with the corresponding Gaussian Matérn SPDE model. In the FE approximation, the solution of equation (6) is represented as a basis expansion $\mathbf{x}(\mathbf{s}) = \Sigma_{j=1}^{n} \Sigma_{k=1}^{p} w_{jk} \boldsymbol{\varphi}_j^k(\mathbf{s})$ using piecewise linear basis functions $\boldsymbol{\varphi}_j^k(\mathbf{s})$ obtained from a mesh over $\mathcal{D}$. The value of $x_k(\mathbf{s}_j)$ at the jth node in the mesh, $\mathbf{s}_j$, is then given by the stochastic weight w_{jk}. Assuming that $p = 2$ and $\alpha_1 = \alpha_2 = 2$, the distribution of $\mathbf{w} = (\mathbf{w}_1^{\mathrm{T}}, \mathbf{w}_2^{\mathrm{T}})^{\mathrm{T}} = (w_{11}, \ldots, w_{n1}, w_{12}, \ldots, w_{n2})^{\mathrm{T}}$ for the case of Gaussian noise is

$$\mathbf{w} \sim N\{\mathbf{0}, \mathbf{K}^{-1}\mathrm{diag}(\mathbf{h}, \mathbf{h})\mathbf{K}^{-\mathrm{T}}\}, \tag{9}$$

where $\mathbf{K}$ is a discretization of the operator matrix and $\mathbf{h}$ is a vector with elements h_i depending on the mesh.

For the corresponding type G_3 model, the distribution of the weights is

$$\mathbf{w}|\mathbf{v} \sim N\left\{\mathbf{K}^{-1}\begin{pmatrix}\gamma_1\mathbf{h} + \mu_1\mathbf{v} \\ \gamma_2\mathbf{h} + \mu_2\mathbf{v}\end{pmatrix}, \mathbf{K}^{-1}\mathrm{diag}(\mathbf{v}, \mathbf{v})\mathbf{K}^{-\mathrm{T}}\right\}, \qquad \mathbf{v} \sim \pi(\mathbf{v}), \tag{10}$$

where the elements of $\mathbf{v} \in \mathbb{R}_+^n$ are independent variables relating to the discretization of the variance process $v(\mathbf{s})$. For the type G_4 model, we have

$$\mathbf{w}|\mathbf{v}_1,\mathbf{v}_2 \sim N\left\{\mathbf{K}^{-1}\begin{pmatrix}\gamma_1\mathbf{h}+\mu_1\mathbf{v}_1\\ \gamma_2\mathbf{h}+\mu_2\mathbf{v}_2\end{pmatrix}, \mathbf{K}^{-1}\mathrm{diag}(\mathbf{v}_1,\mathbf{v}_2)\mathbf{K}^{-\mathrm{T}}\right\}, \qquad \mathbf{v}_1,\mathbf{v}_2 \sim \pi(\mathbf{v}), \tag{11}$$

where $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n_+$ have independent components relating to the discretizations of $v_1(\mathbf{s})$ and $v_2(\mathbf{s})$ repectively. Similarly, the discretization in the type G_1 and type G_2 cases can be written as expressions (10) and (11) respectively, if we define $\mathbf{v} = v\mathbf{h}$ and $\mathbf{v}_i = v_i\mathbf{h}$. As we discussed for the first two cases, we set $\boldsymbol{\mu} = -\boldsymbol{\gamma}$ to ensure that the process has zero mean, and we restrict the distribution of the variances to have mean 1. We then have for all cases that $E(\mathbf{v}) = E(\mathbf{v}_1) = E(\mathbf{v}_2) = \mathbf{h}$. Thus, comparing expressions (9), (10) and (11), we see that a difference between the type G processes and the Gaussian process is that we have replaced the deterministic vector $\mathbf{h}$ in the covariance matrix by a stochastic vector that has $\mathbf{h}$ as expected value. Furthermore, the difference between the four constructions lies in the flexibility of this stochastic vector. In the type G_1 case, we scale the entire field by a single stochastic variable, whereas we scaled each dimension separately in the type G_2 case. For the type G_3 case we have replaced the fixed scaling h_i of the distribution of the weights $w_{i1}, \ldots, w_{ip}$ for a given spatial location $\mathbf{s}_i$ by a common stochastic scaling v_i, which thus affect the sample path behaviour of the process. The type G_4 case is even more general where we have individual stochastic scalings v_{ip} for each weight, and thus more control over the sample path behaviour.

3.3. Properties of the four constructions

The four type G constructions provide random fields with increasing flexibility. All contain several interesting special cases depending on which distribution is used for the variance components, such as generalized asymmetric Laplace distributions, normal–inverse gamma distributions, and Student's t-distributions. As an example, we shall in the next section use NIG noise to highlight some properties of the constructions.

Let $\boldsymbol{\Sigma}$ be the covariance matrix of the solution $\mathbf{x}(\mathbf{s})$ in equation (6) in the case of Gaussian driving noise, for a fixed location $\mathbf{s}$. This matrix has diagonal elements $\Sigma_{ii} = \sigma_i^2$ and off-diagonal elements Σ_{ij} depending on σ_i, σ_j and ρ_{ij}. For the type G_1 construction, we can then write the joint cumulative distribution function (CDF) $F^{(1)}$ of $\mathbf{x}(\mathbf{s})$ and the marginal CDFs $F_k^{(1)}$ of $x_k(\mathbf{s})$ for $k = 1, \ldots, p$ as

$$F^{(1)}(\mathbf{u}) = \int \Phi_{\boldsymbol{\Sigma}}\left(\frac{\mathbf{u}-\boldsymbol{\gamma}-\boldsymbol{\mu}v}{\sqrt{v}}\right)\mathrm{d}F_v(v),$$
$$F_k^{(1)}(u) = \int \Phi\left(\frac{u-\gamma_k-\mu_k v}{\sigma_k\sqrt{v}}\right)\mathrm{d}F_v(v),$$

where $\Phi_{\boldsymbol{\Sigma}}$ denotes the CDF of an $N(\mathbf{0}, \boldsymbol{\Sigma})$ random variable and F_v denotes the CDF of v. There are several choices of F_v that result in fields with known marginal distributions. If for example $\boldsymbol{\mu} = \mathbf{0}$, the field has multivariate Student t-marginals if v is inverse gamma distributed, and multivariate Laplace marginals if v is gamma distributed. The copula of $\mathbf{x}(\mathbf{s})$ is

$$C^{(1)}(\mathbf{u}) = F^{(1)}\{(F_1^{(1)})^{-1}(u_1), \ldots, (F_p^{(1)})^{-1}(u_p)\},$$

which could be viewed as a generalization of the one-factor copulas in Krupskii and Joe (2015) and Krupskii *et al.* (2018). However, despite the flexibility of the marginal distributions, the model is limited since it is non-ergodic for any non-singular distribution of v, and the sample paths are indistinguishable from sample paths of a Gaussian random field. If repeated realizations are available, we can estimate the distribution of v, but not the parameter $\boldsymbol{\theta}$ in the dependence matrix.

For the type G_2 construction, the joint CDF of $\mathbf{x}(\mathbf{s})$ is

$$F^{(2)}(\mathbf{u}) = \int \Phi_{\boldsymbol{\Sigma}}\left\{\text{diag}\left(\frac{1}{\sqrt{v_1}}, \ldots, \frac{1}{\sqrt{v_p}}\right)(\mathbf{u} - \boldsymbol{\gamma} - \boldsymbol{\mu} v)\right\} \mathrm{d}F_{v_1}(v_1) \ldots \mathrm{d}F_{v_p}(v_p),$$

and the marginal CDF for $k = 1, \ldots, p$ is

$$F_k^{(2)}(u) = \int \Phi\left(\frac{u - \gamma_k - \mu_k v_k}{\sigma_k \sqrt{v_k}}\right) \mathrm{d}F_{v_k}(v_k).$$

The copula of $\mathbf{x}(\mathbf{s})$ is $C^{(2)}(\mathbf{u}) = F^{(2)}\{(F_1^{(2)})^{-1}(u_1), \ldots, (F_p^{(2)})^{-1}(u_p)\}$, which is similar to the p-factor copulas in Krupskii and Joe (2015). Also fields that are obtained by using the type G_2 construction are non-ergodic and have sample paths that are indistinguishable from Gaussian sample paths. However, it is possible to estimate all parameters of the model given multiple realizations.

Since the type G_1 and the type G_2 constructions have copulas that are similar to factor copulas, we can compute their so-called tail dependence coefficients and derive conditions on the distribution of v to study their asymptotic tail dependence similarly to Krupskii *et al.* (2018). We leave this for future research as our main interest is in the more flexible type G_3 and type G_4 constructions. The reason for this is, as we shall show in the next subsection, that the type G_1 and type G_2 models have asymptotically Gaussian conditional distributions. This greatly limits their flexibility for spatial data.

For the type G_3 and type G_4 constructions, we in general cannot derive closed form expressions for the marginal distributions and copulas (we shall discuss this further in the next section). However, if we use the representation of the process in remark 1, and let F_k^M and $\tilde{F}_k^M$ denote the distribution functions of the laws of $\mathcal{M}_k$ and $(\mathbf{R}\boldsymbol{\mathcal{M}})_k$ respectively, the copula for the law of $\boldsymbol{\mathcal{M}}_R$ can be written as

$$C(\mathbf{u}) = \prod_{k=1}^{p} F_k^M\{\mathbf{D}_k^{\mathrm{T}}((\tilde{F}_1^M)^{-1}(x_1), \ldots, (\tilde{F}_p^M)^{-1}(x_p))^{\mathrm{T}}\},$$

where $\mathbf{D}_k$ is the kth row of $\mathbf{D}$. This is a Gaussian copula only in the case when $\boldsymbol{\mathcal{M}}$ is Gaussian. Thus, also for these constructions, the dependence structure that is induced by the model can be made more flexible than simply using Gaussian copulas to model the dependence. The type G_4 construction is the most general but the type G_3 construction could be of interest for applications where we wish to capture dependence of the extreme values on different variables. It also has the following interesting feature.

Proposition 3. Let $\mathbf{x}$ be a type G_3 Matérn SPDE field with $\boldsymbol{\rho} = \mathbf{0}$. Then, for $\mathbf{s}, \mathbf{t} \in \mathcal{D}$ and $i \neq j$, $x_i(\mathbf{s})$ and $x_j(\mathbf{t})$ are dependent but uncorrelated random variables.

3.4. Increasing domain asymptotics for the type G_1 model

In this section we explore the distributions of spatial predictions for the type G_1 models and show that they converge to Gaussian distributions as the number of observations goes to ∞. This implies that we might as well use the simpler Gaussian model for prediction if the data set is sufficiently large. Similar issues with related non-Gaussian models have been noted in the mixed effects literature (Rubio and Steel, 2018). To simplify the notation, we focus on the mean 0 univariate case, but the results are easily extended to the general multivariate setting for both type G_1 and type G_2 models.

Let $x_i = x(\mathbf{s}_i)$, $i = 1, \ldots, n$, be observations of a mean 0 random field $x(\mathbf{s})$, for which we want to

predict $x_0 = x(\mathbf{s}_0)$. Let $\mathbf{x}_{k:n}$ denote the vector $(x_k, x_{k+1}, \ldots, x_n)^{\mathrm{T}}$ and assume that the covariance function $C(\mathbf{s}, \mathbf{t})$ of x and the locations $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n$ are such that the covariance matrix of $\mathbf{x}_{0:n}$ is positive definite. Assuming that a mean 0 type G_1 model, with the same covariance function as x, is used for the prediction, the distribution of x_0 given $\mathbf{x}_{1:n}$ is

$$\pi_{G_1,x_0}(x_0|\mathbf{x}_{1:n}) = \int N(x_0; \mathbf{c}_{0,1:n}\mathbf{C}_n^{-1}\mathbf{x}_{1:n}, vc_0 - v\mathbf{c}_{0,1:n}^{\mathrm{T}}\mathbf{C}_n^{-1}\mathbf{c}_{0,1:n})\pi(v)\mathrm{d}v,$$

where $c_0 = V(x_0)$, $\mathbf{c}_{0,1:n}$ is the cross-covariance between $\mathbf{x}_{1:n}$ and x_0, and $\mathbf{C}_n$ is the covariance matrix of $\mathbf{x}_{1:n}$. To show that this distribution converges to a Gaussian distribution we need the following weak assumptions on the observed data.

Assumption 1. The random field $x(\mathbf{s})$ and the observations satisfy, as $n \to \infty$,

$$(\mathbf{x}_{1:n}^{\mathrm{T}}\mathbf{C}_n^{-1}\mathbf{x}_{1:n})/n \xrightarrow{\mathrm{p}} K_0, \tag{12}$$

$$V\{(\mathbf{x}_{1:n}^{\mathrm{T}}\mathbf{C}_n^{-1}\mathbf{x}_{1:n})/\sqrt{n}\} \to k_v, \tag{13}$$

$$\mathbf{c}_{0,1:n}\mathbf{C}_n^{-1}\mathbf{x}_{1:n} \xrightarrow{\mathrm{p}} K_1, \tag{14}$$

$$c_0 - \mathbf{c}_{0,1:n}^{\mathrm{T}}\mathbf{C}_n^{-1}\mathbf{c}_{0,1:n} \to k_2, \tag{15}$$

where $K_0 \geqslant 0$ and K_1 are random variables, $k_2 \in [0, c_0]$ and $k_v > 0$.

The first two assumptions are satisfied for all models that are considered in this paper given that they have finite moments, and given that the sequence $\{\mathbf{s}_i\}$ does not result in a singular covariance matrix (which for example is the case if $\mathbf{s}_i = \mathbf{s}_j$ for $i \neq j$). The last two assumptions assure that the linear predictor converges to a constant given the data. Assuming that x has a Matérn covariance function with $\nu < \infty$, this is also fulfilled as long as the sequence $\{\mathbf{s}_i\}$ is not chosen so that the covariance is degenerate. Given these assumptions, we have the following result.

Theorem 1. Let assumption 1 hold and assume that $\pi(v)$ is a bounded function which is absolutely continuous with respect to the Lebesgue measure, such that $E[v] = 1$. Then

$$\pi_{G_1,x_0}(\cdot|\mathbf{x}_{1:n}) \xrightarrow{\mathrm{p}} N(\cdot; k_1, k_0k_2) \qquad \text{as } n \to \infty.$$

Here k_0 and k_1 are the realizations of K_0 and K_1 in expressions (12) and (14) respectively.

Theorem 1 shows that the predictive distribution for a type G_1 model converges to a Gaussian distribution, and thus the predictor (the mean of the distribution) converges to the corresponding predictor for a Gaussian model, under quite general assumptions on the distribution for the data. In particular, it holds if the data come from a type G_1 model.

Corollary 1. Let $x(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$, be a univariate type G_1 Matérn SPDE field and let $\mathbf{s}_0, \ldots, \mathbf{s}_n$ be locations in $\mathbb{R}^d$ such that $i < \|\mathbf{s}_0 - \mathbf{s}_i\| < i + 1$ for $i = 1, \ldots, n$. Assume that $\pi(v)$ is a bounded function and absolutely continuous with respect to the Lebesgue measure, such that $E[v] = 1$. Then the predictive distribution for $x(s_0)$, $\pi_{x(s_0)}(\cdot|x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n))$, converges in probability to a Gaussian distribution as $n \to \infty$.

4. Normal–inverse Gaussian fields

The NIG distribution (Barndorff-Nielsen, 1997) is obtained by choosing $p = 1$ and v as an inverse gamma random variable in equation (8). The inverse gamma distribution has density

$$\mathrm{IG}(v; \eta_1, \eta_2) = \frac{\sqrt{\eta_2}}{\sqrt{(2\pi v^3)}} \exp\left\{ -\frac{\eta_1}{2} v - \frac{\eta_2}{2v} + \sqrt{(\eta_1 \eta_2)} \right\}, \qquad \eta_1, \eta_2 > 0.$$

The resulting density for the NIG variable is

$$\mathrm{NIG}(x; \gamma, \mu, \eta_1, \eta_2) = \frac{\exp\{\sqrt{(\eta_1\eta_2)} + \mu(x-\gamma)\}\sqrt{(\eta_2\mu^2 + \eta_1\eta_2)}}{\pi\sqrt{\{\eta_2 + (x-\gamma)^2\}}} K_1(\sqrt{[\{\eta_2 + (x-\gamma)^2\}(\mu^2 + \eta_1)]}).$$

In this form the NIG density is overparameterized, and we therefore typically set $\eta_1 = \eta_2 = \eta$ which results in $E[v] = 1$. If $\mu = 0$, we have that the random variable defined in equation (8) has variance 1 but, for $\mu \neq 0$, the variance depends on η. An important property of the NIG distribution is that its variance mixture distribution, the inverse gamma distribution, is closed under convolution. This simplifies inference as explained in later sections.

The simplest multivariate NIG Matérn SPDE field is obtained by using the type G_1 construction with $v \sim \mathrm{IG}(\eta, \eta)$, resulting in a field with multivariate NIG marginal distributions. To construct the more flexible type G_3 and type G_4 models, we use inverse gamma random variables in the univariate type G Lévy processes, which results in NIG processes. Since the NIG distribution has both the Gaussian and the Cauchy distributions as limiting cases (as $\eta \to \infty$ and $\eta \to 0$ with suitable scalings of the other parameters), the NIG Matérn SPDE processes have both a Gaussian process and a Lévy flight process as limiting cases. When using NIG noise in equation (6), it is convenient to note that the noise can be represented by an independently scattered random measure (Rajput and Rosinski, 1989). Specifically, for any Borel set A in the domain, the measure is a univariate NIG random variable with probability density function $f_{\mathcal{N}(A)}(x) = \mathrm{NIG}\{x; m(A)\gamma, \mu, \eta, m(A)^2\eta\}$, where $m(A)$ denotes the Lebesgue measure of A. Note that a random variable with density $f_{\mathcal{N}(A)}(x)$ can be obtained from equation (8) where $v \sim \mathrm{IG}\{\eta, m(A)^2\eta\}$ and thus $E(v) = m(A)$.

We let $\boldsymbol{\mathcal{N}}_3$ and $\boldsymbol{\mathcal{N}}_4$ denote the vector-valued processes in definition 3 when univariate NIG processes are used. The density of $\mathbf{x}(\mathbf{s})$ in equation (6) does not have an explicit form in this case but we can derive the characteristic function of $\mathbf{x}(\mathbf{s})$. The following proposition provides the characteristic function for the type G_4 case.

Proposition 4. The characteristic function of a stationary solution $\mathbf{x}$ to equation (6), evaluated at $\mathbf{s}$, where the driving noise is $\boldsymbol{\mathcal{N}}_4$, is $\phi_{\mathbf{x}(\mathbf{s})}(\mathbf{u}) = \Pi_{k=1}^{p} \phi_k(\mathbf{u})$ where

$$\phi_k(\mathbf{u}) = \exp\left[-\mathrm{i}\gamma_k \int \mathbf{u}^{\mathrm{T}} \mathbf{v}_{k,\mathbf{t}} \mathrm{d}\mathbf{t} + \sqrt{\eta_k} \int \eta_k - \sqrt{\{\eta_k - 2\mathrm{i}\mu_k^2 \mathbf{u}^{\mathrm{T}} \mathbf{v}_{k,\mathbf{t}} + (\mathbf{u}^{\mathrm{T}} \mathbf{v}_{k,\mathbf{t}})^2\}} \mathrm{d}\mathbf{t} \right].$$

Here $\mathbf{v}_{k,\mathbf{t}} = (R_{1k} G_1(\mathbf{s}, \mathbf{t}), R_{2k} G_2(\mathbf{s}, \mathbf{t}), \ldots, R_{pk} G_p(\mathbf{s}, \mathbf{t}))^{\mathrm{T}}$, $\mathbf{R} = \mathbf{D}^{-1}$ and

$$G_k(\mathbf{s}, \mathbf{t}) = \frac{\Gamma\left(\dfrac{\alpha_k - d}{2}\right)}{c_k (4\pi)^{d/4} \Gamma\left(\dfrac{\alpha_k}{2}\right) \kappa_k^{\alpha_k - d}} M\left(\|\mathbf{s} - \mathbf{t}\| \kappa_k, \frac{\alpha_k - d}{2} \right), \qquad k = 1, \ldots, p.$$

The following example illustrates the effect of the shape parameter θ on the multivariate marginal distributions of the type G_4 model.

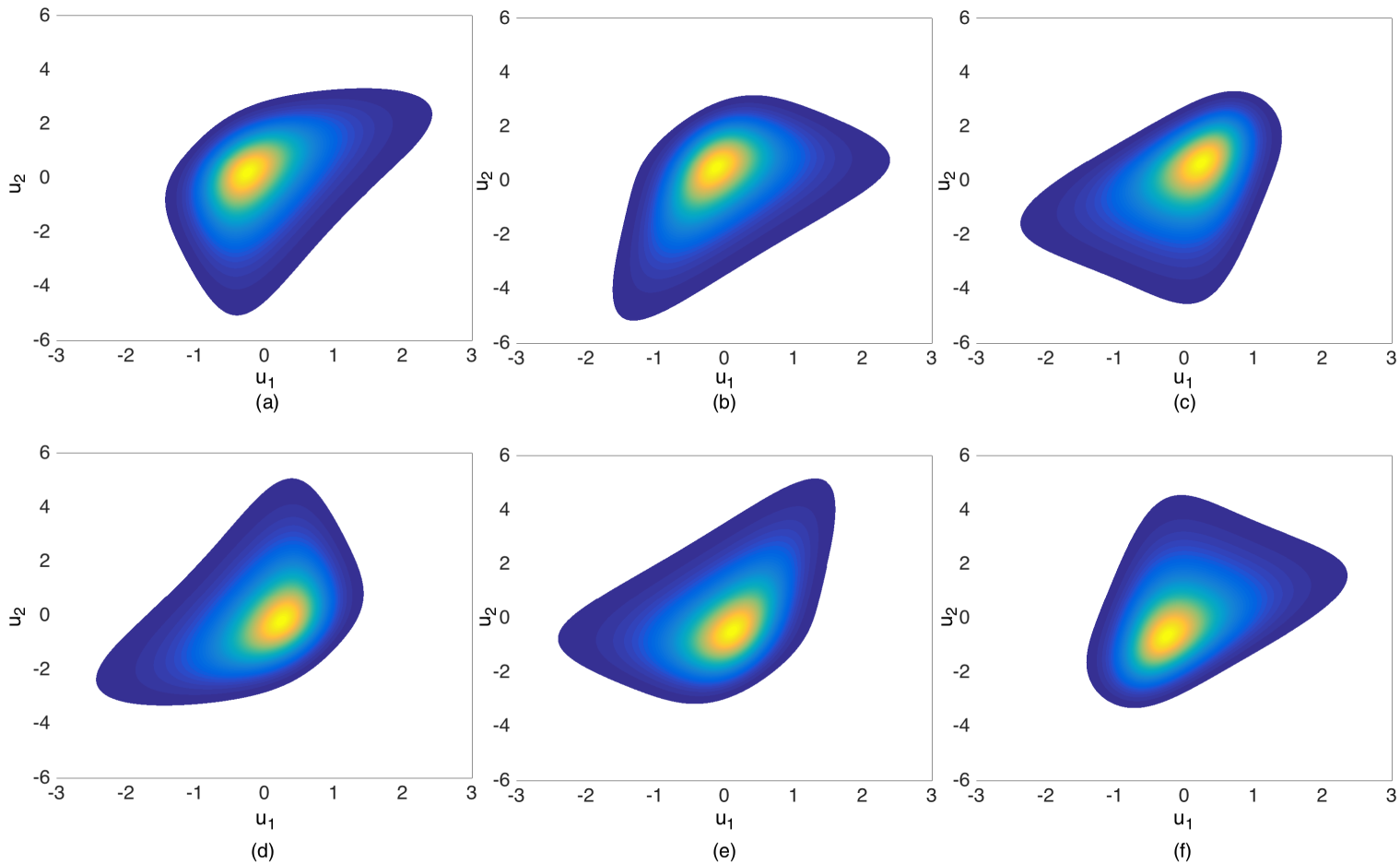

Fig. 2. Marginal distributions $\pi(u_1, u_2)$ of a bivariate NIG Matérn SPDE field $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), x_2(\mathbf{s}))$ for various values of θ (all six cases have the correlation function shown in Fig. 1): (a) $\theta = 0$; (b) $\theta = \tan^{-1}(\rho)$; (c) $\theta = \pi/2$; (d) $\theta = \pi$; (e) $\theta = \pi + \tan^{-1}(\rho)$; (f) $\theta = 3\pi/2$

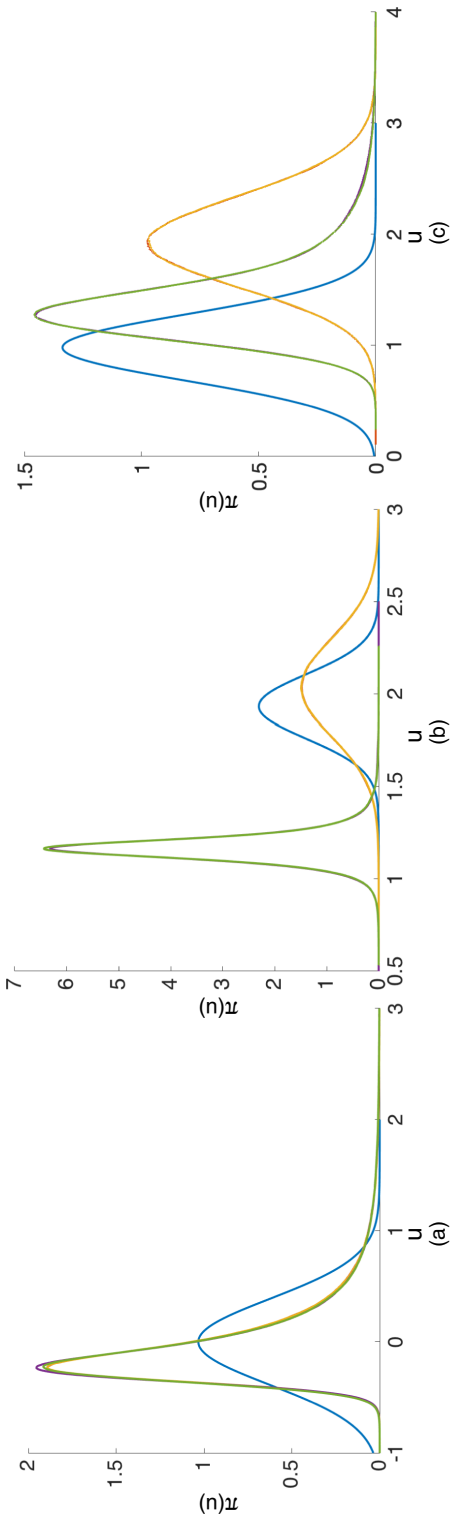

Fig. 3. Marginal distributions for $\mathbf{x}(0)$ based on the five bivariate models in example 2 (here $\mathbf{y} = \{x_1(-1) = 0, x_1(1) = 4\}$ and σ_e denotes the measurement noise standard deviation) (——, Gaussian; ——, type G_1; ——, type G_2; ——, type G_3; ——, type G_4): (a) $\pi\{x_1(0)\}$; (b) $\pi\{x_1(0)|\mathbf{y}\}, \sigma_e = 10^{-3}$; (c) $\pi\{x_1(0)|\mathbf{y}\}, \sigma_e = 0.5$; (d) $\pi\{x_2(0)\}$; (e) $\pi\{x_2(0)|\mathbf{y}\}, \sigma_e = 10^{-3}$; (f) $\pi\{x_2(0)|\mathbf{y}\}, \sigma_e = 0.5$

4.1. Example 1

Let $\mathbf{x}(\mathbf{s})$ be a type G_4 bivariate NIG Matérn SPDE field with the same parameters as in Fig. 1. For the driving noise, we let $\mu_1 = \gamma_2 = 1$, $\mu_2 = \gamma_1 = -1$ and $\eta = 0.9$. Fig. 2 shows bivariate marginal distributions of the resulting field for various values of θ in the dependence matrix (7), computed by using proposition 4. Recall that ρ determines the cross-correlations between $x_1(\mathbf{s})$ and $x_2(\mathbf{s})$ whereas θ determines the shape of the bivariate marginal distributions but does not affect the covariance function. Thus, all six examples have the same correlation function, which is shown in Fig. 1. The case $\theta = 0$ corresponds to a lower triangular operator matrix, and $\theta = \tan^{-1}(\rho)$ corresponds to an upper triangular operator matrix.

As discussed in Section 3, the simpler type G constructions have similar flexibility of the marginal distributions, but lower flexibility in terms of conditional distributions. The following example illustrates how different the predictive distributions can be.

4.2. Example 2

Let $\mathbf{x}_i(t)$, $i = 1, \ldots, 4$, be bivariate type G_i NIG Matérn SPDE processes on $\mathbb{R}$ with $\alpha = 2$, $\kappa = 1$, $\sigma = 0.1$, $\rho = 0.9$ and $\theta = 0$. The processes have the same operator matrix $\boldsymbol{\mathcal{K}}$ and we choose the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ so that they have similar (univariate) marginal distributions; see Figs 3(a) and 3(d), where the marginal distributions of a corresponding Gaussian process also is shown for reference. We predict the value of the four processes at $t = 0$ on the basis of two observations of the first dimension $y_1 = x_{i,1}(-1) + \varepsilon_1 = 0$ and $y_2 = x_{i,1}(1) + \varepsilon_2 = 4$, where ε_{-1} and ε_1 are independent $N(0, 0.001^2)$ variables representing measurement noise. How the prediction is done is presented in Section 5.2. The predictive distributions are shown in Figs 3(b) and 3(e). Even though the four processes have similar marginal distributions for $\mathbf{x}_i(0)$, their predictive distributions are very different. For the prediction of the first dimension, the type G_1 and type G_2 processes have similar distributions, which is expected since they have the same marginal structures. The type G_3 and type G_4 processes also have equal marginal structures and therefore similar predictions, which are very different from the first two. For the prediction of the second dimension, we obtain different predictions for all models since they have different cross-dependence structures. In particular we can note the counterintuitive type G_1 prediction, where the prediction of the second dimension is larger than that of the first, even though there are no observations for this dimension. The same predictive distributions in the case when ε_{-1} and ε_1 instead have variance 0.5^2 are shown in Figs 3(c) and 3(f), and one can note the same behaviour of the predictions for this case.

5. Geostatistical modelling and estimation

To use the multivariate type G fields for geostatistical applications, we need to be able to include them in hierarchical models that include covariates and measurement noise. In this section, we formulate such a model and describe how to perform likelihood-based estimation of the model parameters and how to use the model for spatial prediction.

We consider a standard geostatistical model where a latent field is specified by using covariates for the mean, and the data consist of noisy observations of this latent field at some locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Let y_{ki} be the ith observation of the kth dimension, defined as

$$y_{ki} = \sum_{j=1}^{K} B_{kj}(\mathbf{s}_i)\beta_j + x_k(\mathbf{s}_i) + \varepsilon_{ki}$$

for $k = 1, \ldots, p$, where the independent variables $\varepsilon_{ki} \sim N(0, \sigma^2_{e,k})$ represent the measurement noise. The functions $B_j(\mathbf{s})$ are covariates for the mean and $x_k(\mathbf{s})$ is the kth variable of a mean 0

multivariate type G Matérn SPDE field $\mathbf{x}(\mathbf{s})$. Since the mean of $\mathbf{y}$ is modelled by using covariates, we assume that the mixing variables in the type G construction are scaled so that they have unit expectation (if the expected value exists) and we set $\gamma_k = -\mu_k$ to guarantee that $x_k(\mathbf{s})$ has mean 0 in the case that it has an expected value.

Assuming that the smoothness parameters satisfy $\alpha_i/2 \in \mathbb{N}$ for $i = 1, \ldots, p$ and using the finite dimensional representation of $\mathbf{x}(\mathbf{s})$ that is derived in the on-line appendix A, we have $\mathbf{x}(\mathbf{s}) = \Sigma_{j=1}^{n} \Sigma_{k=1}^{p} w_{jk} \boldsymbol{\varphi}_j^k(\mathbf{s})$. Here $\boldsymbol{\varphi}_j^k(\mathbf{s}) = \varphi_j(s)\mathbf{e}_k$ are p-dimensional basis functions, where $\mathbf{e}_k$ is the kth column in a $p \times p$ identity matrix, $\{\varphi_i\}$ are piecewise linear FE basis functions defined by a mesh on $\mathcal{D}$ and $\{w_{jk}\}$ are stochastic weights. The model is then

$$\begin{aligned} \mathbf{v} &\sim \pi(\mathbf{v}), \\ \mathbf{w}|\mathbf{v} &\sim N\{\mathbf{K}^{-1}(\boldsymbol{\mu} \otimes \mathbf{I}_n)(\mathbf{v} - \mathbf{h}), \mathbf{K}^{-1}\mathrm{diag}(\mathbf{v})\mathbf{K}^{-\mathrm{T}}\}, \\ \mathbf{y}_k|\mathbf{w} &\sim N(\mathbf{B}\boldsymbol{\beta} + \mathbf{A}_k\mathbf{w}, \sigma_{e,k}^2\mathbf{I}), \qquad k = 1, \ldots, p, \end{aligned} \tag{16}$$

where $\mathbf{y}_k$ denotes the vector of all n observations of the kth dimension of the data, $\mathbf{w}$ is a vector with all stochastic weights and $\mathbf{K}$ is a discretization of the operator matrix. The matrix $\mathbf{B}$ contains the covariates that are evaluated at the measurement locations and $\mathbf{A}_k = \mathrm{diag}(\mathbf{e}_k) \otimes \mathbf{A}$ where $\mathbf{A}$ is an observation matrix with elements $\mathbf{A}_{ij} = \varphi_j(\mathbf{s}_i)$. Finally, the distribution of the variance components $\pi(\mathbf{v})$ depends on which model is used, as described in the on-line appendix A.

5.1. Parameter estimation

As is standard in the SPDE approach, we assume that the smoothness parameters are fixed and known. It should be noted that models with general smoothness parameters probably could be estimated from data by using the rational SPDE approach (Bolin and Kirchner, 2019). However, we leave the adaptation of this approach to the multivariate type G setting for future research.

Let $\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_p^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of all observations in expression (16), and let $\boldsymbol{\Psi}$ be the model parameters to be estimated. There is no explicit expression for the likelihood distribution $\pi(\mathbf{y}|\boldsymbol{\Psi})$. However, it is possible to compute maximum likelihood parameter estimates by using Monte Carlo (MC) methods. This is computationally feasible because of two important properties of the model: firstly, $\mathbf{w}|\mathbf{y}, \mathbf{v}, \boldsymbol{\Psi}$ is a Gaussian Markov random field and can thus be sampled efficiently. Secondly, $\mathbf{v}|\mathbf{w}, \mathbf{y}, \boldsymbol{\Psi}$ is a vector of independent variables and can thus be sampled in parallel.

We use a stochastic gradient method (Kushner and Yin, 2003) to estimate the parameters. The idea of the stochastic gradient method is that we need only an asymptotically unbiased estimator (as the number of MC samples goes to ∞) $\mathbf{G}(\boldsymbol{\Psi})$ of the gradient of the likelihood to utilize an iterative procedure where we at iteration i update the parameters as $\boldsymbol{\Psi}^{(i)} = \lambda_i \mathbf{G}(\boldsymbol{\Psi}^{(i-1)}) + \boldsymbol{\Psi}^{(i-1)}$. Here $\{\lambda_i\}$ is a sequence satisfying $\Sigma\lambda_i \to \infty$ and $\Sigma\lambda_i^2 < \infty$, which ensures that the method converges to a stationary point of the likelihood (Kushner and Yin, 2003; Andrieu *et al.*, 2005). To derive the estimator of the gradient, we use Fisher's identity (Dempster *et al.*, 1977) to obtain

$$\nabla_{\boldsymbol{\Psi}} \log\{\pi(\mathbf{y}|\boldsymbol{\Psi})\} = E_{\mathbf{v},\mathbf{w}}[\nabla_{\boldsymbol{\Psi}} \log\{\pi(\mathbf{v}, \mathbf{w}|\mathbf{y}, \boldsymbol{\Psi})\}|\mathbf{y}, \boldsymbol{\Psi}] = E_{\mathbf{v}}[\nabla_{\boldsymbol{\Psi}} \log\{\pi(\mathbf{v}|\mathbf{y}, \boldsymbol{\Psi})\}|\mathbf{y}, \boldsymbol{\Psi}]. \tag{17}$$

Since $\pi(\mathbf{w}|\mathbf{v}, \mathbf{y}, \boldsymbol{\Psi})$ is Gaussian, we have a closed form expression for $\nabla_{\boldsymbol{\Psi}} \log\{\pi(\mathbf{v}|\mathbf{y}, \boldsymbol{\Psi})\}$ (see the on-line appendix B), but there is no closed form expression for its expected value. We therefore use

$$\mathbf{G}(\boldsymbol{\Psi}) = \frac{1}{N} \sum_{i=1}^{N} \nabla_{\boldsymbol{\Psi}} \log\{\pi_{\boldsymbol{\Psi}}(\mathbf{v}^{(i)}|\mathbf{y}, \boldsymbol{\Psi})\}$$

as an MC estimate of the expectation, where $\mathbf{v}^{(i)}$ are samples from distribution $\pi(\mathbf{v}|\mathbf{y}, \boldsymbol{\Psi})$. These samples are obtained by using a Gibbs sampler (algorithm 1 in the on-line appendix C) which

samples $\pi(\mathbf{w}|\mathbf{y}, \mathbf{v}, \boldsymbol{\Psi})$ and $\pi(\mathbf{v}|\mathbf{w}, \mathbf{y}, \boldsymbol{\Psi})$. The sampling of $\pi(\mathbf{v}|\mathbf{w}, \mathbf{y}, \boldsymbol{\Psi})$ typically needs to be done with a general sampling method, such as a Metropolis–Hastings algorithm. However, if $\pi(v)$ is a generalized inverse Gaussian (GIG) distribution, then the conditional distribution remains in the GIG family which can be sampled uniformly fast over the entire parameter space; see Hörmann and Leydold (2014). The GIG distribution has density

$$\mathrm{GIG}(v; c, a, b) = \left(\frac{a}{b}\right)^{c/2} [2K_c\{\sqrt{(ab)}\}]^{-1} v^{c-1} \exp\left\{-\frac{1}{2}(av + bv^{-1})\right\}.$$

For further details, including parameter ranges, see Jørgensen (1982). The GIG distribution contains several known distributions as special cases, such as the gamma distribution, the inverse gamma distribution, and the inverse Gaussian distribution. Because of this, we can sample the variance components of the NIG distribution explicitly. The following example provides the conditional distributions for the NIG Matérn SPDE fields from Section 4.

5.1.1. Example 3

For the NIG processes in Section 4, the distribution of the variance components v, v_i and v_k is $\mathrm{IG}(v; \eta_1, \eta_2) = \mathrm{GIG}(v; -\frac{1}{2}, \eta_1, \eta_2)$. It can therefore be shown that the different type G constructions result in the following posterior distributions: type G_1,

$$\pi(v|\mathbf{E}, \boldsymbol{\Psi}) = \mathrm{GIG}\left\{v; -\frac{np+1}{2}, \eta + \sum_{k=1}^{p} \mu_k^2 \mathbf{1}_n^{\mathrm{T}} \mathbf{h}_k, \eta + \sum_{k=1}^{p} \left(\frac{\boldsymbol{\xi}_k}{\mathbf{h_k}}\right)^{\mathrm{T}} \boldsymbol{\xi}_k\right\};$$

type G_2,

$$\pi(v_k|\mathbf{E}, \boldsymbol{\Psi}) = \mathrm{GIG}\left\{v_k; -\frac{n+1}{2}, \eta_k + \mu_k^2 \mathbf{1}_n^{\mathrm{T}} \mathbf{h}_k, \eta_k + \left(\frac{\boldsymbol{\xi}_k}{\mathbf{h_k}}\right)^{\mathrm{T}} \boldsymbol{\xi}_k\right\};$$

type G_3,

$$\pi(\mathbf{v}|\mathbf{E}, \boldsymbol{\Psi}) = \mathrm{GIG}\left(\mathbf{v}; -\frac{p+1}{2}, \eta + \sum_{k=1}^{p} \mu_k^2, \mathbf{h}_k^2 \eta + \sum_{k=1}^{p} \boldsymbol{\xi}_k^2\right);$$

type G_4,

$$\pi(\mathbf{v}_k|\mathbf{E}, \boldsymbol{\Psi}) = \mathrm{GIG}(\mathbf{v}; -1, \mu_k^2 + \eta_k, \boldsymbol{\xi}_k^2 + \mathbf{h}_k^2 \eta_k),$$

where $\mathbf{E} = (\mathbf{E}_1^{\mathrm{T}}, \ldots, \mathbf{E}_p^{\mathrm{T}})^{\mathrm{T}} = \mathbf{K}\mathbf{w}$ and $\boldsymbol{\xi}_k = \mathbf{E}_k + \mathbf{h}_k \mu_k$. For the last two densities it is explicitly understood that the GIG distribution in vector form denotes a product of independent GIG distributions with parameter values given by the values in the vectors.

5.2. Spatial prediction and evaluation of predictive performance

In applications one is often interested in predictions of the latent field given data. The predictive distribution for the kth variable of the latent field, at a location $\mathbf{s}_0$, is $\pi\{x_k(\mathbf{s}_0)|\mathbf{y}, \boldsymbol{\Psi}\}$. This distribution is often summarized by using the mean as a point estimate, and the variance as a measure of uncertainty. To estimate these two quantities, let $\mathbf{A}_p = (\varphi_1(\mathbf{s}_0), \ldots, \varphi_n(\mathbf{s}_0))$ and use the Gibbs sampler in algorithm 1, on-line appendix C, to obtain samples, $\{\mathbf{v}^i\}_{i=1}^N$, from $\pi(\mathbf{v}|\mathbf{y}, \boldsymbol{\Psi})$. On the basis of these samples, we compute MC estimates $E\{x_k(\mathbf{s}_0)|\mathbf{y}\} \approx (1/N)\Sigma_{i=1}^N \mathbf{A}_p \hat{\boldsymbol{\xi}}^{(i)}$ and $V\{x_k(\mathbf{s}_0)|\mathbf{y}\} \approx (1/N)\Sigma_{i=1}^N \mathbf{A}_p^{\mathrm{T}} (\hat{\mathbf{Q}}^{(i)})^{-1} \mathbf{A}_p$, where $\hat{\boldsymbol{\xi}}^{(i)}$ is the expected value of $\mathbf{w}|y, \mathbf{v}^{(i)}$ and $\hat{\mathbf{Q}}^{(i)}$ is the corresponding precision matrix (see the on-line appendix B for analytic formulae of these quantities). The posterior median, which may be a more appropriate point estimator if the distribution is asymmetric, can similarly be estimated by the sample median of $\{\mathbf{A}_p \hat{\boldsymbol{\xi}}^{(i)}\}_{i=1}^N$.

To evaluate a proposed model we need also to compute various goodness-of-fit measures,

such as the continuous ranked probability score (CRPS) (Matheson and Winkler, 1976). Let y_k be an observation in the kth dimension at $\mathbf{s}_0$, and let F denote the marginal CDF of $\pi(y_k|\mathbf{y}_{-0}, \boldsymbol{\Psi})$, where $\mathbf{y}_{-0}$ denotes all observations except y_k; then the (negatively oriented) CRPS value for this location can be computed as (Gneiting and Raftery, 2007)

$$\mathrm{CRPS}(F, y_k) = E(|Y_k^{(1)} - y_k|) - \tfrac{1}{2}E(|Y_k^{(1)} - Y_k^{(2)}|) \tag{18}$$

where $Y_k^{(1)}$ and $Y_k^{(2)}$ are independent random variables with distribution F. For a Gaussian distribution this expression can be used to derive the CRPS value analytically (see for example Gneiting and Raftery (2007)). For the multivariate type G SPDE Matérn fields, one option is to approximate the expected values in equation (18) by MC integration. Basing such an estimate on N draws of $Y_k^{(1)}$ and $Y_k^{(2)}$ yields an estimate $\mathrm{CRPS}_N(F, y)$. Unfortunately, N often needs to be quite large to obtain good approximations with this estimator. The following proposition provides a more efficient way of approximating the CRPS value in the case of a general normal variance mixture distribution.

Proposition 5. Assume that the random variable X is a normal variance mixture with CDF

$$F(x) = \int \Phi\left\{\frac{x - \mu(v)}{\sigma(v)}\right\} \mathrm{d}F_v(v).$$

Let $V_j^{(i)}$, $j = 1, 2, i = 1, \ldots, N$, be independent draws from the mixing distribution F_v, and define $\mu_V = E(X|V)$, $\sigma_V^2 = V(X|V)$ and

$$M(\mu, \sigma^2) = 2\sigma\varphi\left(\frac{\mu}{\sigma}\right) + \mu\left\{2\Phi\left(\frac{\mu}{\sigma}\right) - 1\right\}, \tag{19}$$

where φ denotes the density function of a standard Gaussian distribution. Then the Rao–Blackwell estimator

$$\mathrm{CRPS}_N^{\mathrm{RB}}(F, y) = \frac{1}{N}\sum_{i=1}^{N}\{M(\mu_{V_1^{(i)}} - y, \sigma^2_{V_1^{(i)}}) - \frac{1}{2}M(\mu_{V_1^{(i)}} - \mu_{V_2^{(i)}}, \sigma^2_{V_1^{(i)}} + \sigma^2_{V_2^{(i)}})\}$$

satisfies $E\{\mathrm{CRPS}_N^{\mathrm{RB}}(F, y)\} = \mathrm{CRPS}(F, y)$ and $V\{\mathrm{CRPS}_N^{\mathrm{RB}}(F, y)\} \leqslant V\{\mathrm{CRPS}_N(F, y)\}$.

The $\mathrm{CRPS}_N^{\mathrm{RB}}$-estimator can be used for the type G fields since $\pi\{x_k(\mathbf{s}_0)|\mathbf{y}_{-0}, \mathbf{v}, \boldsymbol{\Psi}\}$ is Gaussian and since we can easily sample the variances $\mathbf{v}$ by using the Gibbs sampler.

To give an idea of the improvement that can be obtained by using the RB estimator, both estimators were used to compute the CRPS value for the final fold of the most general NIG model in the cross-validation study in Section 6. Based on $N = 10000$ samples, the MC variances of the two estimators were $V\{\sqrt{N}\mathrm{CRPS}_N^{\mathrm{RB}}(F, y)\} \approx 187$ and $V\{\sqrt{N}\mathrm{CRPS}_N(F, y)\} \approx 2225$.

6. Applications

In this section we illustrate for two different data sets how the multivariate type G SPDE fields can be used for spatial modelling. The first data set consists of temperature and pressure measurements from the North American Pacific Northwest and was previously studied in Gneiting *et al.* (2010) and Apanasovich *et al.* (2012). The second data set consists of seawater temperature measurements taken at two different depths in the ocean.

For both data sets we assume the model $\mathbf{y}_i = \boldsymbol{\beta} + \mathbf{x}(\mathbf{s}_i) + \boldsymbol{\varepsilon}_i$ for the bivariate observations $\mathbf{y}_i$, where $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), x_2(\mathbf{s}))^{\mathrm{T}}$ is a mean 0 random field, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$ is the expected value and $\boldsymbol{\varepsilon}_i$ are independent $N\{\mathbf{0}, \mathrm{diag}(\sigma_1^2, \sigma_2^2)\}$ variables representing measurement noise. As reference

models, we shall for each data set use four Gaussian models for $\mathbf{x}(\mathbf{s})$. The first of these assumes that $x_1(\mathbf{s})$ and $x_2(\mathbf{s})$ are independent Gaussian Matérn fields with covariance functions $C_{11}(\mathbf{h}) = \sigma_1^2 M(\mathbf{h} \mid \kappa_1, \nu_1)$ and $C_{22}(\mathbf{h}) = \sigma_2^2 M(\mathbf{h} \mid \kappa_2, \nu_2)$ respectively. We also use the parsimonious Gaussian Matérn field of Gneiting *et al.* (2010) as well as two Gaussian Matérn SPDE models specified by using equation (6): one lower triangular and one independent model with $\rho = 0$. For the applications, we focus on comparing the reference models with the type G_4 models and do not evaluate the simpler type G constructions. We do not consider the type G_1 and type G_2 models since the first data set does not have repeated measurements, and since we do not expect these models to improve the predictive performance compared with the Gaussian models because of theorem 1. We do not consider the type G_3 model since there is no specific reason why a shared variance component would be beneficial for the data sets considered.

6.1. Temperature and pressure in the North American Pacific Northwest

The data, shown in Fig. 4, consist of temperature and pressure observations, $\mathbf{y}_i = (y_P, y_T)_i^{\mathrm{T}}$ where y_P denotes pressure and y_T temperature, at 157 locations in the North American Pacific Northwest. Besides the four baseline models, we test four type G Matérn SPDE models for the data. A Gaussian model for temperature seems adequate whereas the pressure data have short-range variations that are inflating the measurement noise variance (see the parameter estimates in the on-line appendix D), which possibly could be captured by the latent field if a non-Gaussian model was used. We therefore consider type G_4 models where the driving noise for the pressure is NIG distributed, whereas the driving noise for temperature is Gaussian. To investigate the effects of the operator matrix, we use one independent model, with $\rho = 0$, and two dependent models. The first of these is triangular with $\theta = 0$, and the second has a general operator where θ is estimated jointly with the other parameters.

The mesh that is used for the discretization of the SPDE models is shown in Fig. 4. It consists of 981 nodes and was built by using `R-INLA` (Lindgren and Rue, 2015). We fix the α-parameters to 2 for all SPDE models, which corresponds to $\nu = 1$ for the Matérn covariances. The parameters of the Gaussian models are estimated by using numerical maximization of the log-likelihood function, whereas the gradient-based method from Section 5.1 is used for the non-Gaussian models. The gradient method is run for 1000 iterations, using starting values that were obtained from the corresponding Gaussian model. For the lower triangular models, the estimation took

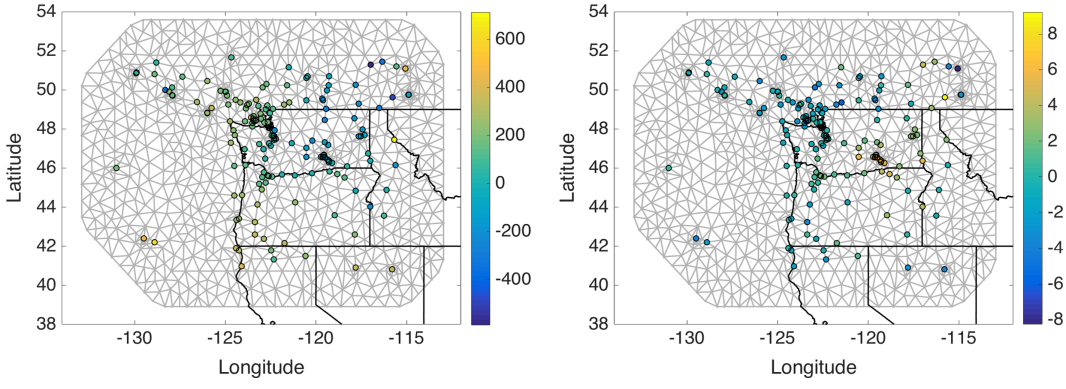

Fig. 4. Measurements of (a) pressure and (b) temperature in the North American Pacific Northwest together with the mesh used for the SPDE models: the sample mean has been subtracted from the data in both cases

Table 1. Cross-validation results comparing the median absolute error MAE and median CRPS for the various models†

Model	*Operator matrix*	*Number of parameters*	*Pressure (Pa)*		*Temperature (°C)*	
			MAE	*CRPS*	*MAE*	*CRPS*
Independent Matérn	—	10	41.632	28.994	0.956	0.598
Parsimonious Matérn	—	10	39.068	27.682	0.921	0.576
Gaussian SPDE	Diagonal	8	38.624	31.711	0.917	0.594
Gaussian SPDE	Triangular	9	38.856	31.829	0.915	0.580
NIG SPDE	Diagonal	10	39.101	25.993	0.847	0.525
NIG SPDE	Triangular	11	39.302	25.776	*0.841*	*0.512*
NIG SPDE	General	12	*38.523*	*25.591*	0.876	0.514

†The best (lowest) value in each column is in italics.

44 s for the Gaussian model and 156 s for the NIG model. These values were obtained by using a MATLAB (2015) implementation of the algorithm on a Macbook Pro computer with a 2.6-GHz Intel Core i7 processor. The parameter estimates for the various models are shown in the on-line appendix D.

To compare the models, we perform a leave-one-out pseudo-cross-validation study. For each observation location, the pressure and temperature values are predicted by using the data from all 156 other locations using the models with parameters given in appendix D. For all models, the point estimates are computed by using the expected values of the values at the held-out location conditionally on the data at all other locations. Using the posterior median as a predictor did not improve the predictive performance for these data, and we therefore omit those results. The predictive performance of the models is assessed by using the median absolute error of the 157 predicted values, as well as the median CRPS. The resulting values are shown in Table 1. We note that the dependent NIG models have better predictive performance than the Gaussian models. Spatial predictions by using the parsimonious Matérn model and the general NIG model can be seen in Fig. 5.

6.2. Seawater temperatures

We now consider Argo floats measurements of seawater temperature at two depths. Since the measurements are sparse in space (and time) an important statistical task is to 'fill in the gaps' through spatial interpolation. The data have been thoroughly analysed from a statistical perspective by Kuusela and Stein (2018), who noted that the data seem to be non-Gaussian at higher depths in certain areas. To investigate whether the type G models could be useful for interpolation of these data, we choose two depths, 300 dbar and 1500 dbar, and investigate whether one could improve the joint prediction of those depths by using the type G models.

We extract data from the month of February for three years (2014–2016). Since Kuusela and Stein (2018) showed that an analysis of the complete data set requires a non-stationary model, we focus on a limited spatial region south of New Zealand to be able to use a stationary model. However, as for the Gaussian SPDE-based models, one could model non-stationarity by allowing the parameters in the operator to be spatially varying. The restriction results in a data set consisting of 312 observations in total. For each location we study the residuals after removing a seasonally varying mean field (the Roemmich–Gilson mean field; see Kuusela and Stein (2018)). Thus, we let $\mathbf{y}_i = (y_{300}(\mathbf{s}_i), y_{1500}(\mathbf{s}_i))^{\mathrm{T}}$ for $i \in \{1, \ldots, n_t\}$ and $t \in \{2014, 2015, 2016\}$.

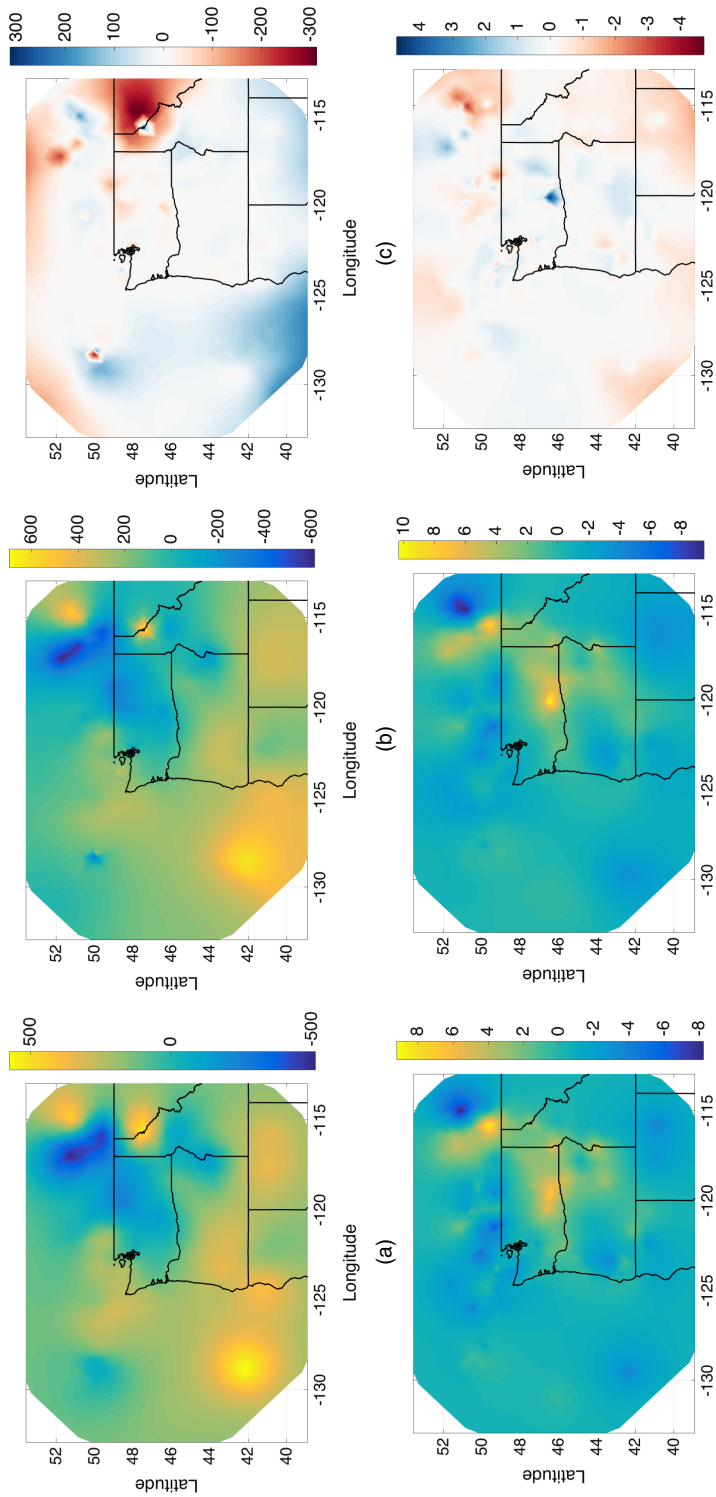

Fig. 5. Estimates of (a)–(c) pressure and (d)–(f) temperature by using (a), (d) the parsimonious Matérn model and (b), (e) the NIG model with the general operator matrix: (c), (f) difference between the estimates

Here y_{300} is the residual at the depth of 300 dbar and y_{1500} is the residual at the depth of 1500 dbar. We assume that data for the different years are independent.

Besides the four baseline models, we test two non-Gaussian type G SPDE models. In the first, we assume that $x_1(\mathbf{s})$ and $x_2(\mathbf{s})$ are independent univariate NIG SPDE fields. In the second, we use the type G_4 construction with a general operator matrix where both noise processes are NIG distributed. For all SPDE-based models, we again fix the smoothness parameters α to 2 and use the mesh that is shown in Fig. 6, which also shows the available data for the year 2016. The parameters are estimated by using an R implementation of the methods proposed, available in the package `ngme`. The parameter estimates for the various models are shown in the on-line appendix D.

To evaluate which of the models tested performs best in terms of prediction, we again use leave-one-out pseudo-cross-validation and compare the median absolute error and median CRPS. However, contrary to the previous application we now do the cross-validation by removing individual univariate observations instead of removing the bivariate observation pairs for each spatial location. The reason for this is that we here do not necessarily have observations of the fields at different depth in the same spatial locations. The results are shown in Table 2, where we see that the multivariate non-Gaussian model is clearly outperforming the other models. Thus, it seems as if we could increase the accuracy of the spatial interpolation of the Argo data by

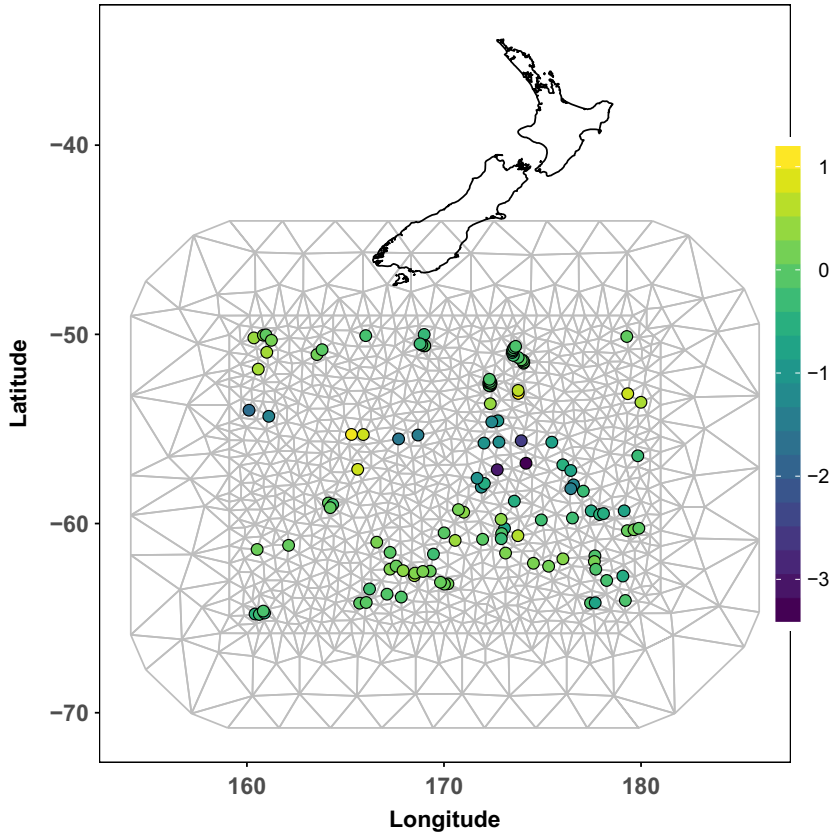

Fig. 6. Measurements of seawater temperature at depth 300 db in 2016 together with the mesh used for the SPDE models: a mean field has been removed from the measurements

Table 2. Cross-validation results comparing the median absolute error MAE and median CRPS for various models for the Argo data†

Model	*Operator matrix*	*Number of parameters*	*Results for the following depths:*			
			300 db		*1500 db*	
			MAE	*CRPS*	*MAE*	*CRPS*
Independent Matérn	—	10	0.183	0.154	0.040	0.029
Parsimonious Matérn	—	10	0.239	0.162	0.033	*0.024*
Gaussian SPDE	Diagonal	8	*0.187*	0.182	0.043	0.032
Gaussian SPDE	Triangular	9	0.216	0.169	0.034	0.025
NIG SPDE	Diagonal	12	0.200	0.159	0.040	0.028
NIG SPDE	General	14	0.201	*0.130*	*0.032*	*0.024*

†The best (lowest) value in each column is in italics.

using type G models. For future work it is therefore interesting to study the entire data set, where a more detailed analysis would require a space–time model; see Kuusela and Stein (2018).

7. Discussion

There is a need for practically useful random-field models with more general distributions than the Gaussian distribution. Especially for multivariate data, finding good alternatives to Gaussian fields has been considered an open problem in the literature. We have introduced one such alternative by formulating a new class of multivariate random fields with flexible multivariate marginal distributions and covariance functions of Matérn type. The fields are constructed as solutions to SPDEs and can be used in a geostatistical setting where likelihood-based parameter estimation can be performed by using a computationally efficient stochastic gradient algorithm. In fact, the models have the same computational advantages as their Gaussian counterparts, which facilitates applications to large data sets, although with additional cost due to MC sampling.

Four constructions of the non-Gaussian noise were considered, where the first two are closely related to existing approaches, such as factor copula models and Student t-fields. We showed that these constructions have significant disadvantages when used for spatial prediction, or on data without replicates. The more sophisticated constructions based on type G Lévy noise do not have these disadvantages, and their combination of flexibility and computational efficiency should therefore make them attractive alternatives to Gaussian models for geostatistical applications.

The computational benefits of the finite dimensional approximations that are presented in the on-line appendix A are available only for fields with $\alpha/2 \in \mathbb{N}$. This restriction of the smoothness parameters is often viewed as one of the main drawbacks of the SPDE approach, since the smoothness of the covariance function is important for the predictive performance. However, in many cases the distributional assumptions can be equally important. This was clearly shown in the application where the covariance-based models, which allow for arbitrary smoothness parameters, were outperformed by the non-Gaussian models with fixed smoothness parameters. Nevertheless, extending the approach to fields with general smoothness would increase the flexibility. As previously mentioned, this could probably be done by using the rational SPDE approach (Bolin and Kirchner, 2019), and extending that method to multivariate type G fields is thus an interesting topic for future research.

Acknowledgements

This work has been supported by the Swedish Research Council under grant 2016-04187 and the Knut and Alice Wallenberg Foundation (grant KAW 20012.0067). The authors thank Holger Rootzén, the Joint Editor, Associate Editor and the reviewers for valuable comments on the manuscript. We also thank Mikael Kuusela for helping with the Argo data.

References

Andrieu, C., Moulines, É. and Priouret, P. (2005) Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optimzn*, **44**, 283–312.

Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012) A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Am. Statist. Ass.*, **107**, 180–193.

Bárdossy, A. (2006) Copula-based geostatistical models for groundwater quality parameters. *Wat. Resour. Res.*, **42**, no. 11, W11416.

Barndorff-Nielsen, O. E. (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.*, **24**, 1–13.

Barndorff-Nielsen, O., Kent, J. and Sørensen, M. (1982) Normal variance-mean mixtures and *z* distributions. *Int. Statist. Rev.*, **50**, 145–159.

Bolin, D. and Kirchner, K. (2019) The rational SPDE approach for Gaussian random fields with general smoothness. *J. Computnl Graph. Statist.*, to be published.

Bolin, D. and Lindgren, F. (2011) Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Statist.*, **5**, 523–550.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Du, J., Leonenko, N., Ma, C. and Shu, H. (2012) Hyperbolic vector random fields with hyperbolic direct and cross covariance functions. *Stoch. Anal. Appl.*, **30**, 662–674.

Genton, M. G. and Kleiber, W. (2015) Cross-covariance functions for multivariate geostatistics. *Statist. Sci.*, **30**, 147–163.

Genton, M. G., Padoan, S. A. and Sang, H. (2015) Multivariate max-stable spatial processes. *Biometrika*, **102**, 215–230.

Gneiting, T., Kleiber, W. and Schlather, M. (2010) Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Ass.*, **105**, 1167–1177.

Gneiting, T. and Raftery, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *J. Am. Statist. Ass.*, **102**, 359–378.

Gräler, B. (2014) Modelling skewed spatial random fields through the spatial vine copula. *Spatl Statist.*, **10**, 87–102.

Hörmann, W. and Leydold, J. (2014) Generating generalized inverse Gaussian random variates. *Statist. Comput.*, **24**, 547–557.

Hu, X., Simpson, D., Lindgren, F. and Rue, H. (2013) Multivariate Gaussian random fields using systems of stochastic partial differential equations. *Preprint arXiv:1307.1379*. Norwegian University of Science and Technology, Trondheim.

Hu, X. and Steinsland, I. (2016) Spatial modeling with system of stochastic partial differential equations. *Comput. Statist.*, **8**, no. 2, 112–125.

Jørgensen, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Berlin: Springer.

Kazianka, H. and Pilz, J. (2010) Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stoch. Environ. Res. Risk Assess.*, **24**, 661–673.

Krupskii, P., Huser, R. and Genton, M. G. (2018) Factor copula models for replicated spatial data. *J. Am. Statist. Ass.*, **113**, 467–479.

Krupskii, P. and Joe, H. (2015) Structured factor copula models: theory, inference and computation. *J. Multiv. Anal.*, **138**, 53–73.

Kushner, H. J. and Yin, G. (2003) *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer.

Kuusela, M. and Stein, M. L. (2018) Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. R. Soc.* A, **474**, article 20180400.

Lindgren, F. and Rue, H. (2015) Bayesian spatial modelling with R-INLA. *J. Statist. Softwr.*, **63**, no. 19, 1–25.

Lindgren, F., Rue, H. and Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. R. Statist. Soc.* B, **73**, 423–498.

Ma, C. (2013a) Mittag-Leffler vector random fields with Mittag-Leffler direct and cross covariance functions. *Ann. Inst. Statist. Math.*, **65**, 941–958.

Ma, C. (2013b) Student's t vector random fields with power-law and log-law decaying direct and cross covariances. *Stoch. Anal. Appl.*, **31**, 167–182.
Matérn, B. (1960) Spatial variation. *Meddel. Från Statens Skogsforskinst.*, **49**, no. 5.
Matheson, J. E. and Winkler, R. L. (1976) Scoring rules for continuous probability distributions. *Mangmnt Sci.*, **22**, 1087–1096.
MATLAB (2015) *MATLAB Version 8.6.0.267246 (R2015b)*. Natick: MathWorks.
Rajput, B. S. and Rosinski, J. (1989) Spectral representations of infinitely divisible processes. *Probab. Theory Reltd Flds*, **82**, 451–487.
Røislien, J. and Omre, H. (2006) T-distributed random fields: a parametric model for heavy-tailed well-log data. **38**, 821–849.
Rosinski, J. (1991) On a class of infinitely divisible processes represented as mixtures of Gaussian processes. In *Stable Processes and Related Topics* (eds S. Cambanis, G. Samorodnitsky and M. S. Taqqu), pp. 27–41. Boston: Birkhäuser.
Rubio, F. and Steel, M. (2018) Flexible linear mixed models with improper priors for longitudinal and survival data. *Electron. J. Statist.*, **12**, 572–598.
Wallin, J. and Bolin, D. (2015) Geostatistical modelling using non-Gaussian Matérn fields. *Scand. J. Statist.*, **42**, 872–890.
Whittle, P. (1963) Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, **40**, 974–994.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article:
'Appendix to Multivariate Type-G random fields'.