Probabilistic Engineering Mechanics 66 (2021) 103167

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Bayesian analysis of hierarchical random fields for material modeling

Sebastian Geyer *, Iason Papaioannou, Daniel Straub

Engineering Risk Analysis Group, Technische Universität München, Arcisstraße 21, 80333 München, Germany

ARTICLE INFO

Keywords:
Gaussian random fields
Student's *t*-distribution
Analytical update
Bayesian analysis
Spatial variability
Conjugate prior

ABSTRACT

In probabilistic assessments, spatially variable material properties are modeled with random fields. These random fields can be learned from spatial data by means of Bayesian analysis. This paper presents analytical expressions for the Bayesian analysis of hierarchical Gaussian random fields. We model the prior spatial distribution by a Gaussian random field with normal-gamma distributed mean and precision and make use of the conjugacy of prior distribution and likelihood function to find the posterior distribution of the random field parameters. We present closed-form expressions for the spatial mean and precision function of the posterior predictive Student's *t*-random field. Furthermore, we discuss the application of the hierarchical model to non-Gaussian random fields (translation random fields) and show the connection of the methodology to the Bayesian approachof EN 1990 for estimating characteristic values for material parameters. The method is illustrated on two spatial data sets of concrete and soil strength parameters.

1. Introduction

Setting up an engineering model requires definition of material properties. To correctly account for their inherent randomness, such material properties are commonly modeled probabilistically. A probabilistic representation with random variables is sufficient for modeling materials without or with negligible spatial variability. However, in many applications the effects of the spatial variability of materials should not be neglected in the modeling process. This is the case, e.g., with soil parameters in geotechnical assessments [1], and material parameters in assessments of large concrete structures [e.g.2].

Spatially variable uncertain quantities can be modeled by random fields (RFs). An RF represents a random variable at each point of a spatial domain [3]. A complete definition of the RF requires specification of the joint distribution of the variables corresponding to any collection of points of the spatial domain. This is nontrivial in general with the exception of Gaussian and a special case of non-Gaussian RFs, termed translation RFs. Translation RFs are RFs that can be expressed as functions of Gaussian RFs [4], e.g., a lognormal RF can be expressed as the exponential of a Gaussian RF. A Gaussian RF implies that the joint distribution for any collection of points is multivariate Gaussian and can be completely defined by the first- and second-moment functions [5]. Gaussian RFs have well established statistical properties and a variety of methods are available for simulating them [e.g. [6]].

RFs can be learned from data through Bayesian analysis [7]. In the general case, such an update needs to be done numerically with methods usually based on Monte Carlo sampling, including Markov chain Monte Carlo methods [8], sequential Monte Carlo methods [9,10] and subset simulation [11–13]. However, Gaussian RFs enable the use of conjugate priors to learn the RF parameters via a closed-form update in a Bayesian analysis [2,14].

RFs have been used for a long time in the field of geostatistics for the interpolation of spatial data by means of kriging, which includes Bayesian inference of Gaussian RF parameters [e.g. [15,16]]. A comprehensive review of hierarchical Bayesian analysis with spatial data from the viewpoint of geostatistics can be found in [17]. More recently, these approaches have gained importance in the field of machine learning. They are used for Gaussian process regression, which is a versatile surrogate model for random functions with noisy observations [18]. In the engineering community, the potential of accounting for spatial variability within Bayesian analysis has been recognized especially in the field of geotechnical engineering [e.g. [19–22]]. Therein, it is often essential to identify site-specific trend functions of soil properties in addition to the inherent spatial variability. Recently, attempts have been made to simultaneously learn the trend function and autocovariance function with sparse measurements in a Bayesian analysis. The approach of [23–25] applies sparse Bayesian learning to learn the trend function of the RF and subsequently draws samples from the posterior distribution of the RF parameters through Markov chain Monte Carlo methods. The authors of [26] applied Bayesian compressive sampling to represent non-homogeneous RFs. This approach does not require the explicit choice of a prior RF model. It expresses the RF as a superposition of a set of basis functions and evaluates the posterior distribution of the coefficients of these functions using sparse measurements. The method has been combined with the Karhunen–Loève expansion to

* Corresponding author.
E-mail address: s.geyer@tum.de (S. Geyer).

https://doi.org/10.1016/j.probengmech.2021.103167
Received 14 January 2021; Received in revised form 25 June 2021; Accepted 13 September 2021
Available online 2 October 2021

obtain realizations of the RF [27–29], and has been recently extended to treat multi-dimensional and cross-correlated RFs [30,31].

Bayesian approaches have also found their way into other engineering fields, e.g., for estimating design values of structural material properties when samples are available [32,33], which is also included in the current European standards for constructions (EN 1990) [34].

The aim of this paper is to present a hierarchical Bayesian model for material properties modeled with Gaussian or translation RFs. Hierarchical Gaussian Bayesian models have been well developed in the context of Bayesian linear regression [e.g.7] and hierarchical spatial modeling [e.g.35]. This work applies existing results from these fields to derive a comprehensive hierarchical RF model that can be used in the context of stochastic material modeling. We make use of the fact that the normal-gamma distribution is the conjugate prior for the mean and precision of a Gaussian RF to obtain the posterior distribution of the RF parameters. The posterior predictive RF is a non-homogeneous RF with Student's t-marginal distribution. Importantly, given a prior distribution for the RF parameters and a chosen autocorrelation function, all steps of the Bayesian analysis can be performed in closed form, providing marginal and multivariate solutions for the posterior predictive RF model. This property should simplify application in practice, especially in engineering domains where accounting for spatial variability is currently not common practice. Moreover, we discuss how existing approaches for simulation of Gaussian RFs can be applied to generate realizations of the derived RF model. The application to situations with non-Gaussian translation prior RFs is investigated and for the specific case of lognormal prior distribution, the equations for the required transformation are given. Furthermore, we discuss the influence of the prior correlation function and a posterior point estimate of its parameters. Finally, we show that the presented updating approach is a generalization of the Bayesian approach for evaluation of characteristic values of EN 1990.

The structure of the paper is as follows. Section 2 presents the structure of the hierarchical RF, followed by a short review of Bayesian analysis and a step-by-step presentation of the proposed Bayesian updating procedure. Section 3 applies the method to two examples from different engineering fields (geotechnical engineering and structural engineering). A summary and main conclusions are given in Section 4. The analytical expressions for updating the RF are derived in Appendices A to C and Appendix D describes properties of the log-Student's t-distribution.

2. Methodology

In a Gaussian RF $X(\boldsymbol{z})$, the joint distribution of $\{X(\boldsymbol{z}_i), i = 1, \dots, n\}$ for any $\boldsymbol{z}_i \in \Omega \subset \mathbb{R}^d$ and $n \in \mathbb{N}$ is jointly Gaussian, with Ω denoting the domain of definition of the RF and d the spatial dimension of Ω [5]. This RF is fully described by the spatial functions for the mean value, the variance and the autocorrelation. Closed-form solutions are available for the posterior distribution of the RF given data $\mathbf{M}$ of X [7,36]. We consider a prior RF for $X(\boldsymbol{z})$ with homogeneous point statistics, i.e., a-priori the RF has constant mean and variance. The vector of uncertain hyperparameters is $\boldsymbol{\theta} = \left[\mu_X, \lambda_X\right]^\intercal$, where μ_X is the mean value and λ_X is the precision (inverse of the variance). The assumption of prior homogeneity is a simplification and limits the application to cases without a spatial trend of the RF or cases where a homogeneous RF $X(\boldsymbol{z})$ can be obtained from the actual RF by a normalization operation [e.g., 3] or by de-trending methods [e.g., 37,38].

Fig. 1 summarizes the investigated problem setting, where the nodes represent uncertain quantities (the random variables and the RF) and the arrows denote the direct dependencies among them [e.g. [39]]. $\boldsymbol{\tau}$ is the vector of correlation parameters, i.e., the parameters of the autocorrelation function of the RF. These are initially considered as deterministic; the estimation of $\boldsymbol{\tau}$ from the data $\mathbf{M}$ is discussed in Section 2.7. It is worth noting that the method can handle arbitrary autocorrelation functions, i.e., we do not require the autocorrelation function to depend on the difference in location, although in most applications this is a standard choice. The aim of the analysis is to learn $X(\boldsymbol{z})$ conditional on $\mathbf{M}$. The individual steps of the analysis are derived in the following, preceded by a short introduction to the basics of Bayesian analysis.

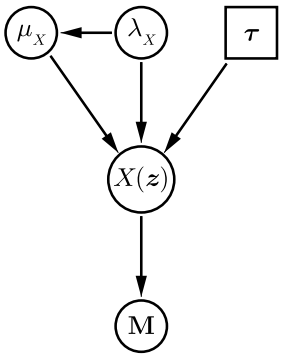

Fig. 1. The hierarchical RF model to learn $X(\boldsymbol{z})$ from $\mathbf{M}$. μ_X and λ_X are the mean and precision of the RF $X(\boldsymbol{z})$ and $\mathbf{M}$ is the measurement data. $\boldsymbol{\tau}$ is the vector of parameters of the autocorrelation function.

2.1. BayesIan analysis

When performing a Bayesian analysis, the first step is setting up a prior joint probability density function (PDF) of the parameters $\boldsymbol{\theta}$. The prior PDF $f(\boldsymbol{\theta})$ is then updated to the posterior PDF $f(\boldsymbol{\theta}|\mathbf{M})$ with data $\mathbf{M}$, by application of Bayes' rule [7]:

$$f(\boldsymbol{\theta}|\mathbf{M}) \propto f(\boldsymbol{\theta}) \cdot \mathrm{L}(\boldsymbol{\theta}|\mathbf{M}), \tag{1}$$

where $\mathrm{L}(\boldsymbol{\theta}|\mathbf{M})$ is the likelihood function, summarizing the information from the data $\mathbf{M}$. Note that a single data point $\mathbf{M}_i$ may contain various types of information, including the measurement outcome, the measurement location or time, the used measurement device and the environmental conditions at the time of the measurement. In this paper, we focus on the case where $\mathbf{M}$ contains spatially distributed measurements of an RF $X(\boldsymbol{z})$. Hence, each $\mathbf{M}_i$ includes the measurement outcome $x_{m,i}$ and the corresponding measurement location $\boldsymbol{z}_{m,i}$, i.e., $\mathbf{M}_i = [x_{m,i}, \boldsymbol{z}_{m,i}]$. Given a set of n direct measurements of the RF $X(\boldsymbol{z})$, $\mathbf{M} = \left[\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n\right]^\intercal$, with measurement outcomes $\boldsymbol{x}_m = \left[x_{m,1}, x_{m,2}, \dots, x_{m,n}\right]^\intercal \subseteq \mathbb{R}^n$ and corresponding measurement locations $\mathbf{Z}_m = \left[\boldsymbol{z}_{m,1}, \boldsymbol{z}_{m,2}, \dots, \boldsymbol{z}_{m,n}\right]^\intercal \subseteq \mathbb{R}^{n \times d}$, the joint likelihood is the PDF of $X(\boldsymbol{z})$ at locations $\mathbf{Z}_m$ conditional on $\boldsymbol{\theta}$:

$$\mathrm{L}(\boldsymbol{\theta}|\mathbf{M}) = f\left(\boldsymbol{x}_m; \mathbf{Z}_m|\boldsymbol{\theta}\right). \tag{2}$$

2.2. Prior model

We consider a Gaussian RF $X(\boldsymbol{z})$ whose parameter vector $\boldsymbol{\theta}$ has a normal-gamma ($\mathcal{NG}$) prior, with PDF [e.g. [33,35,40]]

$$\begin{aligned} f(\boldsymbol{\theta}) &= \mathcal{NG}\left(\mu_X, \lambda_X|\mu_0, \kappa_0, \alpha_0, \beta_0\right) = \mathcal{N}\left(\mu_X|\mu_0, \kappa_0\lambda_X\right) \cdot \mathcal{G}\left(\lambda_X|\alpha_0, \beta_0\right) \\ &= C_0 \lambda_X^{\alpha_0 - \frac{1}{2}} \exp\left(-\lambda_X\left(\frac{\kappa_0}{2}\left(\mu_X - \mu_0\right)^2 + \beta_0\right)\right). \end{aligned} \tag{3}$$

$\Gamma(\cdot)$ is the gamma function and C_0 is a normalizing constant, given by

$$C_0 = \frac{\beta_0^{\alpha_0} \kappa_0^{\frac{1}{2}}}{\Gamma\left(\alpha_0\right)(2\pi)^{\frac{1}{2}}}. \tag{4}$$

The spatial variability of the prior RF is determined by its autocorrelation function $\rho\left(\boldsymbol{z}_1, \boldsymbol{z}_2\right)$ [3,5]. A classical choice for the autocorrelation function is the Matérn model, which includes the exponential model and the square-exponential model [5,18,41].

2.3. Likelihood function

The likelihood function for learning the RF $X(\boldsymbol{z})$ with spatially distributed measurements $\mathbf{M}$ is given by Eq. (2). For the Gaussian RF this translates to:

$$\mathrm{L}(\boldsymbol{\theta}|\mathbf{M}) = \frac{\lambda_X^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}\left(\det\left(\mathbf{R}_m\right)\right)^{\frac{1}{2}}} \exp\left(-\frac{\lambda_X}{2}\left(\boldsymbol{x}_m - \mu_X\mathbf{1}_n\right)\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_X\mathbf{1}_n\right)^{\mathsf{T}}\right), \tag{5}$$

where $\mathbf{R}_m$ is the correlation matrix of the measurement locations with entry $R_{m,i,j}$ calculated as $\rho\left(\boldsymbol{z}_{m,i}, \boldsymbol{z}_{m,j}\right)$. $\mathbf{1}_n$ denotes a $1\times n$-vector of ones.

Uncertainty in the measurement procedure can be accounted for by including a measurement error ε_i. Assuming an additive measurement error yields the following relation between the actual value x at location $\boldsymbol{z}_{m,i}$ and the measured value $x_{m,i}$:

$$x_{m,i} = x\left(\boldsymbol{z}_{m,i}\right) + \varepsilon_i. \tag{6}$$

The error ε_i is often modeled by a zero-mean Gaussian random variable with standard deviation σ_ε and statistical independence between the measurement errors at different locations is assumed. In such case, the joint likelihood function retains the form of Eq. (5) and the measurement error only affects the calculation of the entries in $\mathbf{R}_m$:

$$R_{m,i,j} = \rho\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right) + \delta(i,j)\cdot\sigma_\varepsilon^2, \tag{7}$$

where $\delta(i,j)$ is the Dirac delta function returning 1 if $i = j$ and 0 otherwise.

2.4. Posterior distribution of the parameters

In the general case, Eq. (1) needs to be solved numerically, e.g. through sampling techniques, due to the intractability of the normalizing constant. However, analytical solutions for the posterior distribution are available in some special cases, when using conjugate priors [7,40]. The chosen $\mathcal{NG}$ prior distribution and the multivariate Gaussian likelihood of Eq. (5) are conjugate. Hence, the posterior distribution of $\boldsymbol{\theta}$ can be derived analytically and has the same parametric form as the prior, i.e., it is a $\mathcal{NG}$ distribution. The Bayesian updating simplifies to an update of the parameters of the $\mathcal{NG}$ distribution [35,42]:

$$f(\boldsymbol{\theta}|\mathbf{M}) = \mathcal{NG}\left(\mu_X, \lambda_X|\mu_n, \kappa_n, \alpha_n, \beta_n\right) = C_n\lambda_X^{\alpha_n-\frac{1}{2}}\exp\left(-\lambda_X\left(\frac{\kappa_n}{2}\left(\mu_X-\mu_n\right)^2+\beta_n\right)\right), \tag{8}$$

where the normalizing constant C_n is given by

$$C_n = \frac{\beta_n^{\alpha_n}\kappa_n^{\frac{1}{2}}}{\Gamma\left(\alpha_n\right)(2\pi)^{\frac{1}{2}}}. \tag{9}$$

The parameters of the posterior distribution can be obtained with the following set of equations:

$$\mu_n = \frac{\kappa_0\mu_0 + \mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^{\mathsf{T}}}{\kappa_0 + \mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}}, \tag{10}$$

$$\kappa_n = \kappa_0 + \mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}, \tag{11}$$

$$\alpha_n = \alpha_0 + \frac{n}{2}, \tag{12}$$

$$\beta_n = \beta_0 + \frac{1}{2}\left(\boldsymbol{x}_m\mathbf{R}_m^{-1}\boldsymbol{x}_m^{\mathsf{T}} + \frac{\kappa_0\mu_0^2\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}} - 2\kappa_0\mu_0\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^{\mathsf{T}} - \left(\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^{\mathsf{T}}\right)^2}{\kappa_0 + \mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}}\right). \tag{13}$$

A derivation of the parameters in Eqs. (10) to (13) can be found in [7] in the context of Bayesian linear regression. For easier accessibility, we provide the derivations in Appendix A.

2.5. Marginal posterior predictive distribution

Typically, the goal is to make predictions about the quantity of interest X. To this end, one needs the posterior predictive distribution of X, which is obtained by marginalization of the joint PDF of X conditional on $\boldsymbol{\theta}$ and the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{M}$. When X is modeled by a single random variable and the measurement points are uncorrelated, the posterior predictive distribution is given as [7,19,33]

$$f\left(x|\boldsymbol{x}_m\right) = \int_{\boldsymbol{\Theta}} f\left(x|\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\boldsymbol{x}_m\right)\mathrm{d}\boldsymbol{\theta}, \tag{14}$$

where $\boldsymbol{\Theta}$ denotes the domain of definition of $\boldsymbol{\theta}$. The conditional independence between X given $\boldsymbol{\theta}$ and $\mathbf{M}$ does no longer hold when X is modeled as an RF. Instead, the posterior and the posterior predictive distribution of X will depend on the spatial location $\boldsymbol{z}$. In RF theory, the PDF of the RF $X(\boldsymbol{z})$ at location $\boldsymbol{z}$ is termed marginal (or first order) PDF of $X(\boldsymbol{z})$. The marginal posterior predictive PDF of $X(\boldsymbol{z})$, denoted $f(x;\boldsymbol{z}|\mathbf{M})$, is given as

$$f(x;\boldsymbol{z}|\mathbf{M}) = \int_{\boldsymbol{\Theta}} f(x;\boldsymbol{z}|\boldsymbol{\theta},\mathbf{M})\, f(\boldsymbol{\theta}|\mathbf{M})\,\mathrm{d}\boldsymbol{\theta}. \tag{15}$$

Here, $f(x;\boldsymbol{z}|\boldsymbol{\theta},\mathbf{M})$ is the marginal PDF of $X(\boldsymbol{z})$ given $\boldsymbol{\theta}$ and $\mathbf{M}$, which requires an additional updating step. In this step, the prior is the marginal PDF of $X(\boldsymbol{z})$ given $\boldsymbol{\theta}$, $f(x;\boldsymbol{z}|\boldsymbol{\theta})$, which is a Gaussian PDF with parameters μ_X and λ_X. The posterior PDF $f(x;\boldsymbol{z}|\boldsymbol{\theta},\mathbf{M})$ is again a Gaussian PDF with parameters $\mu_{\boldsymbol{z}}''$ and $\lambda_{\boldsymbol{z}}''$, which can be calculated by application of the following updating rules for the conditional Gaussian distribution [3,15,36]:

$$\mu_{\boldsymbol{z}}'' = \mu_X + \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_X\mathbf{1}_n\right)^{\mathsf{T}}, \tag{16}$$

$$\lambda_{\boldsymbol{z}}'' = \lambda_X\left(1 - \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\boldsymbol{R}_{\boldsymbol{z},m}^{\mathsf{T}}\right)^{-1}, \tag{17}$$

where $\boldsymbol{R}_{\boldsymbol{z},m}: \mathbb{R}^d \to \mathbb{R}^{1\times n}$ is a row vector function with element i defined as $\rho\left(\boldsymbol{z}, \boldsymbol{z}_{m,i}\right)$ with n being the number of measurements and $\mathbf{R}_m$ is given by Eq. (7).

The integral in Eq. (15) can be written as follows:

$$f(x;\boldsymbol{z}|\mathbf{M}) = \int_{\lambda_X=0}^{+\infty}\int_{\mu_X=-\infty}^{+\infty}\mathcal{N}\left(x|\mu_{\boldsymbol{z}}'', \lambda_{\boldsymbol{z}}''\right)\mathcal{NG}\left(\mu_X,\lambda_X|\mu_n,\kappa_n\lambda_X\right)\mathcal{G}\left(\lambda_X|\alpha_n,\beta_n\right)\mathrm{d}\mu_X\mathrm{d}\lambda_X. \tag{18}$$

Solution of the integral in Eq. (18) results in the following marginal posterior predictive PDF:

$$f(x;\boldsymbol{z}|\mathbf{M}) = f_t\left(x|\mu_{\boldsymbol{z},t}, \lambda_{\boldsymbol{z},t}, \nu_t\right) = \frac{\Gamma\left(\frac{\nu_t}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\left(\frac{\lambda_{\boldsymbol{z},t}}{\pi\nu_t}\right)^{\frac{1}{2}}\left(1+\frac{\lambda_{\boldsymbol{z},t}\left(x-\mu_{\boldsymbol{z},t}\right)^2}{\nu_t}\right)^{-\frac{\nu_t}{2}-\frac{1}{2}}, \tag{19}$$

where $f_t\left(x|\mu_t,\lambda_t,\nu_t\right)$ denotes the PDF of the Student's t-distribution with location parameter μ_t, scale parameter λ_t and degrees of freedom ν_t [36].

The spatial functions for the parameters of the posterior predictive Student's t-distribution are given in closed form by the following expressions:

$$\mu_{\boldsymbol{z},t} = \mu_n + \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_n\mathbf{1}_n\right)^{\mathsf{T}}, \tag{20}$$

$$\lambda_{\boldsymbol{z},t} = \frac{\alpha_n}{\beta_n\left(1 - \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\boldsymbol{R}_{\boldsymbol{z},m}^{\mathsf{T}} + \left(1 - \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)^2\kappa_n^{-1}\right)}, \tag{21}$$

$$\nu_t = 2\alpha_n. \tag{22}$$

The parameters μ_n, κ_n, α_n and β_n are obtained following the updating rules in Eqs. (10) to (13). A detailed derivation of the parameter update can be found in Appendix B.

2.6. Posterior predictive random field

The approach presented in Section 2.5 enables predicting the marginal distribution of quantity X at any location $\boldsymbol{z} \in \Omega$ given spatial data $\mathbf{M}$. This is useful in cases where the correlation among values of X at different locations needs not be accounted for in further predictions [33]. However, in many cases the spatial dependence of X is required for predictions. In such cases, the joint distribution of X at k different locations is given by the kth order posterior predictive PDF of $X(\boldsymbol{z})$:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta},\mathbf{M})\, f(\boldsymbol{\theta}|\mathbf{M})\, \mathrm{d}\boldsymbol{\theta}. \tag{23}$$

The posterior distribution for the parameter vector $\boldsymbol{\theta}$ is the same as the one appearing in Eq. (15). The prior distribution of the RF $X(\boldsymbol{z})$ given $\boldsymbol{\theta}$ is Gaussian and, hence, $f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta})$ is k-variate Gaussian. Since the updating rules for a conditional Gaussian distribution of Eqs. (16) and (17) can be extended to the multivariate case, $f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta},\mathbf{M})$ is also k-variate Gaussian with mean vector $\boldsymbol{\mu}''_{\mathbf{Z}}$ and precision matrix $\boldsymbol{\Lambda}''_{\mathbf{Z}}$, which can be calculated by the following equations [36]:

$$\boldsymbol{\mu}''_{\mathbf{Z}} = \mu_X \mathbf{1}_k^{\mathsf{T}} + \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_X \mathbf{1}_n\right)^{\mathsf{T}}, \tag{24}$$

$$\boldsymbol{\Lambda}''_{\mathbf{Z}} = \lambda_X \left(\mathbf{R}_{\mathbf{Z}} - \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{R}_{\mathbf{Z},m}^{\mathsf{T}}\right)^{-1}, \tag{25}$$

where $\mathbf{R}_{\mathbf{Z},m} : \mathbb{R}^{k\times d} \to \mathbb{R}^{k\times n}$ is a matrix function with element i,j defined as $\rho(\boldsymbol{z}_i,\boldsymbol{z}_{m,j})$. $\mathbf{R}_{\mathbf{Z}} : \mathbb{R}^{k\times d} \to \mathbb{R}^{k\times k}$ is a matrix function with element i,j defined as $\rho(\boldsymbol{z}_i,\boldsymbol{z}_j)$. $\mathbf{R}_m$ is the matrix containing the correlation of the measurement locations and a potential measurement error, as introduced in Section 2.3. $\mathbf{1}_k$ is a $1\times k$ vector of ones.

Eq. (23) takes the following form:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\lambda_X=0}^{+\infty}\int_{\mu_X=-\infty}^{+\infty} \mathcal{N}\left(\boldsymbol{x}_z|\boldsymbol{\mu}''_{\mathbf{Z}},\boldsymbol{\Lambda}''_{\mathbf{Z}}\right)\mathcal{N}\left(\mu_X|\mu_n,\kappa_n\lambda_X\right)\mathcal{G}\left(\lambda_X|\alpha_n,\beta_n\right)\mathrm{d}\mu_X\,\mathrm{d}\lambda_X. \tag{26}$$

The integral in Eq. (26) results in the following kth order posterior predictive PDF

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = f_t\left(\boldsymbol{x}|\boldsymbol{\mu}_{\mathbf{Z},t},\boldsymbol{\Lambda}_{\mathbf{Z},t},\nu_t\right) = \frac{\Gamma\left(\frac{\nu_t}{2}+\frac{k}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\frac{\left(\det\left(\boldsymbol{\Lambda}_{\mathbf{Z},t}\right)\right)^{\frac{1}{2}}}{\left(\pi\nu_t\right)^{\frac{k}{2}}} \times\left(1+\frac{\left(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{Z},t}\right)\boldsymbol{\Lambda}_{\mathbf{Z},t}\left(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{Z},t}\right)^{\mathsf{T}}}{\nu_t}\right)^{-\frac{\nu_t}{2}-\frac{k}{2}}. \tag{27}$$

where $f_t\left(\boldsymbol{x}|\boldsymbol{\mu}_{\mathbf{Z},t},\boldsymbol{\Lambda}_{\mathbf{Z},t},\nu_t\right)$ is the k-variate Student's t-distribution [36, 43]. As in the univariate case, ν_t is a scalar parameter denoting the degrees of freedom. ν_t is given by Eq. (22) and the parameters $\boldsymbol{\mu}_{\mathbf{Z},t}$ and $\boldsymbol{\Lambda}_{\mathbf{Z},t}$ are given in closed form:

$$\boldsymbol{\mu}_{\mathbf{Z},t} = \mu_n \mathbf{1}_k^{\mathsf{T}} + \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_n \mathbf{1}_n\right)^{\mathsf{T}}, \tag{28}$$

$$\boldsymbol{\Lambda}_{\mathbf{Z},t} = \frac{\alpha_n}{\beta_n}\left(\mathbf{R}_{\mathbf{Z}} - \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{R}_{\mathbf{Z},m}^{\mathsf{T}} + \left(\mathbf{1}_k^{\mathsf{T}} - \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)\kappa_n^{-1}\left(\mathbf{1}_k^{\mathsf{T}} - \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)^{\mathsf{T}}\right)^{-1}, \tag{29}$$

$\mathbf{R}_{\mathbf{Z}}$, $\mathbf{R}_{\mathbf{Z},m}$ and $\mathbf{R}_m$ follow the definitions for Eqs. (24) and (25) and the parameters μ_n, κ_n, α_n and β_n are obtained following the updating rules in Eqs. (10) to (13). The analytical expressions for the parameters of the multivariate posterior predictive Student's t-distribution are derived in detail in Appendix C.

The multivariate Student's t-distribution as predictive distribution for the multivariate Gaussian distribution also appears in Bayesian regression for the normal linear model [7,35,44]. In fact, the presented model forms a special case of weighted linear Bayesian regression with a single explanatory variable.

Eq. (27) can be used for multivariate predictions of X accounting for the information in $\mathbf{M}$. It is noted that for $k = 1$, Eq. (27) reduces to the expression for the marginal posterior predictive Student's t-distribution given in Eq. (19), accordingly Eqs. (28) and (29) reduce to Eqs. (20) and (21).

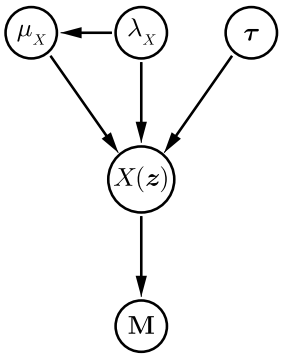

Fig. 2. Adapted hierarchical Bayesian model to consider $\boldsymbol{\tau}$ as uncertain parameter. The dependence of the resulting RF model on $\boldsymbol{\tau}$ can be integrated out when $\boldsymbol{\tau}$ is modeled as random vector.

Eqs. (28) and (29) can be directly transformed to the spatial parameter functions of the posterior predictive RF, i.e., the mean function $\mu_t(\boldsymbol{z})$ and the precision function $\lambda_t(\boldsymbol{z}_1,\boldsymbol{z}_2)$:

$$\mu_t(\boldsymbol{z}) = \mu_n + \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m - \mu_n\mathbf{1}_n\right)^{\mathsf{T}}, \tag{30}$$

$$\lambda_t(\boldsymbol{z}_1,\boldsymbol{z}_2) = \frac{\alpha_n}{\beta_n}\left(\rho(\boldsymbol{z}_1,\boldsymbol{z}_2) - \boldsymbol{R}_{\boldsymbol{z}_1,m}\mathbf{R}_m^{-1}\boldsymbol{R}_{\boldsymbol{z}_2,m}^{\mathsf{T}} + \left(1-\boldsymbol{R}_{\boldsymbol{z}_1,m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)\kappa_n^{-1}\left(1-\boldsymbol{R}_{\boldsymbol{z}_2,m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)\right)^{-1}, \tag{31}$$

where $\rho(\boldsymbol{z}_1,\boldsymbol{z}_2)$ is the prior correlation of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. $\boldsymbol{R}_{\boldsymbol{z},m}$ and $\mathbf{R}_m$ are utilized as in Eqs. (16) and (17). The posterior predictive RF is fully defined by the parameters specified by Eqs. (22), (30) and (31).

2.7. Choice of correlation parameters

The choice of the prior autocorrelation function $\rho(\boldsymbol{z}_1,\boldsymbol{z}_2)$ has significant influence on the predictive distribution of the proposed RF model; it controls the spatial variability of the prior RF and the correlation of the measurement locations in $\mathbf{R}_m$. Hence, the autocorrelation function and its parameters need to be chosen carefully. Although literature is available on different parametric correlation models, their advantages and disadvantages [e.g. [5,41,45]], the specific parameter choice for a problem at hand remains challenging if little information about the modeled quantity is available. This problem can be addressed by treating the vector $\boldsymbol{\tau}$ of correlation parameters as a random vector with associated prior distribution $f(\boldsymbol{\tau})$. The dependency between $\boldsymbol{\tau}$ and the multivariate posterior predictive distribution can be expressed explicitly by extending Eq. (23) as follows:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M},\boldsymbol{\tau}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta},\mathbf{M},\boldsymbol{\tau})\, f(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})\, \mathrm{d}\boldsymbol{\theta}. \tag{32}$$

$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M})$ can then be determined by marginalization of $f(\boldsymbol{x};\mathbf{Z},\boldsymbol{\tau}|\mathbf{M})$:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\mathbf{T}} f(\boldsymbol{x};\mathbf{Z},\boldsymbol{\tau}|\mathbf{M})\,\mathrm{d}\boldsymbol{\tau} = \int_{\mathbf{T}} f(\boldsymbol{x};\mathbf{Z}|\mathbf{M},\boldsymbol{\tau})\, f(\boldsymbol{\tau}|\mathbf{M})\,\mathrm{d}\boldsymbol{\tau}, \tag{33}$$

with $\mathbf{T}$ denoting the domain of definition of $\boldsymbol{\tau}$. Fig. 2 illustrates the adapted hierarchical Bayesian model where $\boldsymbol{\tau}$ is considered as additional uncertain parameter, in contrast to the deterministic choice illustrated in Fig. 1. The closed-form updating procedure for the posterior predictive expressions can only be used to find $f(\boldsymbol{x};\mathbf{Z}|\mathbf{M},\boldsymbol{\tau})$. Direct evaluation of Eq. (33) can be cumbersome or even impossible, as it requires evaluation of $f(\boldsymbol{\tau}|\mathbf{M})$, which depends on the choice of the correlation model and most likely cannot be evaluated in closed form.

Through application of Bayes' theorem, $f(\boldsymbol{\tau}|\mathbf{M})$ is given by the following expression:

$$f(\boldsymbol{\tau}|\mathbf{M}) \propto f(\boldsymbol{\tau})\cdot f(\mathbf{M}|\boldsymbol{\tau}). \tag{34}$$

Including the dependency on $\boldsymbol{\tau}$ in the definition of the likelihood function of Eq. (2) gives

$$\mathrm{L}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau}) = \frac{\lambda_X^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}\det\left(\mathbf{R}_m(\boldsymbol{\tau})\right)}\exp\left(-\frac{\lambda_X}{2}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)\left(\mathbf{R}_m(\boldsymbol{\tau})\right)^{-1}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)^{\mathrm{T}}\right). \tag{35}$$

$f(\mathbf{M}|\boldsymbol{\tau})$ is the proportionality constant in $f(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau}) \propto f(\boldsymbol{\theta})\cdot\mathrm{L}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})$, hence

$$f(\mathbf{M}|\boldsymbol{\tau}) = \frac{f(\boldsymbol{\theta})\cdot\mathrm{L}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})}{f(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})}. \tag{36}$$

Note that $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are independent and thus, $f(\boldsymbol{\theta}|\boldsymbol{\tau}) = f(\boldsymbol{\theta})$. $f(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})$ is the posterior PDF of $\boldsymbol{\theta}$ for a given $\boldsymbol{\tau}$, which is a $\mathcal{NG}$ PDF with parameters given in Section 2.4. Splitting the densities and their respective normalizing constants in Eq. (36) gives

$$f(\mathbf{M}|\boldsymbol{\tau}) = \frac{C_0}{C_n(\boldsymbol{\tau})}\cdot(2\pi)^{-\frac{n}{2}}\det\left(\mathbf{R}_m(\boldsymbol{\tau})\right)^{-\frac{1}{2}}\frac{\hat{f}(\boldsymbol{\theta})\cdot\hat{\mathrm{L}}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})}{\hat{f}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})}, \tag{37}$$

where C_0 and C_n are defined in Eqs. (4) and (9). $\hat{f}(\boldsymbol{\theta})$ and $\hat{f}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})$ are the unnormalized prior and posterior $\mathcal{NG}$ distributions. $\hat{\mathrm{L}}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})$ is the exponential term of the likelihood function and is equal to the ratio of $\hat{f}(\boldsymbol{\theta}|\mathbf{M},\boldsymbol{\tau})$ and $\hat{f}(\boldsymbol{\theta})$ (cf. Appendix A). Thus, the fraction disappears in Eq. (37). Inserting the expressions for C_0 and C_n into Eq. (37) yields

$$f(\mathbf{M}|\boldsymbol{\tau}) = \left(\frac{\kappa_0}{\kappa_n(\boldsymbol{\tau})}\right)^{\frac{1}{2}}\frac{\Gamma(\alpha_n)\,\beta_0^{\alpha_0}}{\Gamma(\alpha_0)\left(\beta_n(\boldsymbol{\tau})\right)^{\alpha_n}}(2\pi)^{-\frac{n}{2}}\det\left(\mathbf{R}_m(\boldsymbol{\tau})\right)^{-\frac{1}{2}}. \tag{38}$$

Using Eq. (38), sampling from $f(\boldsymbol{\tau}|\mathbf{M})$ can be achieved, e.g. by using Markov chain Monte Carlo methods [8]. These samples $\boldsymbol{\tau}_i$, $i = 1,\dots,N_{MCMC}$ can then be used to approximate $f(\boldsymbol{x};\mathbf{Z}|\mathbf{M})$:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) \approx \frac{1}{N_{MCMC}}\sum_{i=1}^{N_{MCMC}} f(\boldsymbol{x};\mathbf{Z}|\mathbf{M},\boldsymbol{\tau}_i). \tag{39}$$

Alternatively, the posterior distribution of $\boldsymbol{\tau}$ can be approximated by its maximum a-posteriori (MAP) estimate [46]. That is, Eq. (33) is approximated by

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) \approx f\left(\boldsymbol{x};\mathbf{Z}|\mathbf{M},\boldsymbol{\tau}^*\right), \tag{40}$$

where $\boldsymbol{\tau}^*$ is the MAP estimate of $\boldsymbol{\tau}$. It is found by maximizing Eq. (34) with respect to $\boldsymbol{\tau}$. Using Eq. (38), this is equivalent to solving the following optimization problem:

$$\arg\max_{\boldsymbol{\tau}\in\mathbf{T}} f(\boldsymbol{\tau}|\mathbf{M}) = \arg\min_{\boldsymbol{\tau}\in\mathbf{T}} \ln\left(\kappa_n(\boldsymbol{\tau})\right)+2\alpha_n\ln\left(\beta_n(\boldsymbol{\tau})\right)+\ln\left(\det\left(\mathbf{R}_m(\boldsymbol{\tau})\right)\right)-2\ln\left(f(\boldsymbol{\tau})\right), \tag{41}$$

where κ_n, α_n and β_n follow the definitions in Section 2.4 conditional on $\boldsymbol{\tau}$.

The parametric form of the correlation model can be chosen among a set of models by means of Bayesian model selection. To this end, the marginal likelihood, i.e., the normalizing constant of Eq. (34), must be evaluated for the different parametric model choices and multiplied with the prior beliefs in the models [47].

2.8. Extension to non-Gaussian prior random fields

The presented Bayesian approach is applicable to Gaussian prior RFs and data assigned with additive Gaussian measurement error. Its applicability can be extended to the class of so-called translation RFs, defined as [4,48]

$$Y(\boldsymbol{z}) = T(U(\boldsymbol{z})), \tag{42}$$

where $U(\boldsymbol{z})$ is a zero-mean and unit-variance Gaussian RF. If the marginal cumulative distribution function (CDF) of the non-Gaussian RF $F_{Y;\boldsymbol{z}}(y(\boldsymbol{z}))$ is given and it is strictly increasing, one can define the transformation of Eq. (42) as $T(\cdot) = F_{Y;\boldsymbol{z}}^{-1}(\Phi(\cdot))$, with $F_{Y;\boldsymbol{z}}^{-1}(\cdot)$ denoting the inverse of $F_{Y;\boldsymbol{z}}(\cdot)$ and $\Phi(\cdot)$ the standard normal CDF [49]. $U(\boldsymbol{z})$ is obtained from $Y(\boldsymbol{z})$ by inversion of Eq. (42):

$$U(\boldsymbol{z}) = T^{-1}(Y(\boldsymbol{z})). \tag{43}$$

To apply the proposed hierarchical Bayesian approach to the non-Gaussian RF $Y(\boldsymbol{z})$, each measurement outcome $y_{m,i}$ transformed to the Gaussian space through Eq. (43) should be associated with an additive Gaussian error. This can be equivalently stated as follows:

$$y_{m,i} = T\left(u\left(\boldsymbol{z}_{m,i}\right)+\varepsilon_i\right), \tag{44}$$

where ε_i is a zero-mean Gaussian measurement error. A special case is a lognormal RF $Y(\boldsymbol{z})$ with parameters $\mu_{\ln Y}$ and $\lambda_{\ln Y}$ and a multiplicative lognormal measurement error, i.e., $y_{m,i} = y\left(\boldsymbol{z}_{m,i}\right)\cdot\varepsilon_{y,i}$. In such case, Eqs. (42) and (44) can be rewritten as functions of a Gaussian RF $X(\boldsymbol{z})$:

$$Y(\boldsymbol{z}) = \exp(X(\boldsymbol{z})), \tag{45}$$

$$\begin{aligned} y_{m,i} &= \exp\left(x\left(\boldsymbol{z}_{m,i}\right)+\lambda_X^{-\frac{1}{2}}\cdot\varepsilon_i\right) = \exp\left(x\left(\boldsymbol{z}_{m,i}\right)\right)\cdot\exp\left(\lambda_X^{-\frac{1}{2}}\cdot\varepsilon_i\right)\\ &= \exp\left(x\left(\boldsymbol{z}_{m,i}\right)\right)\cdot\varepsilon_{y,i}. \end{aligned} \tag{46}$$

$\mu_{\ln Y}$ and $\lambda_{\ln Y}$ are the mean value and precision respectively of the underlying Gaussian RF $X(\boldsymbol{z})$, i.e., $\mu_X = \mu_{\ln Y}$ and $\lambda_X = \lambda_{\ln Y}$. The error term $\varepsilon_{y,i}$ follows a lognormal distribution with median 1. Its parameters are $\mu_{\ln\varepsilon} = 0$ and $\lambda_{\ln\varepsilon} = \lambda_\varepsilon\cdot\lambda_X$, which are mean value and precision respectively of the underlying Gaussian measurement error. $\lambda_{\ln\varepsilon}$ has to be chosen accordingly. That is, the hierarchical Bayesian approach is directly applicable by a simple logarithmic transformation of the data and the measurement error. After the updating procedure, the posterior predictive RF can be transformed back to the original space by applying Eq. (45). The transformed marginal distribution of the posterior predictive RF has the form of a log-Student's t-distribution. This distribution model is used in finance for the pricing of options [50,51] and belongs to the family of log-symmetric distributions [52]. The marginal PDF of the posterior predictive RF is defined as follows:

$$\begin{aligned} f(y;\boldsymbol{z}|\mathbf{M}) = f_{t,\ln}\left(y|\mu_{\boldsymbol{z},t},\lambda_{\boldsymbol{z},t},\nu_t\right) &= y^{-1}\frac{\Gamma\left(\frac{\nu_t}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\left(\frac{\lambda_{\boldsymbol{z},t}}{\pi\nu_t}\right)^{\frac{1}{2}}\\ &\times\left(1+\frac{\lambda_{\boldsymbol{z},t}\left(\ln(y)-\mu_{\boldsymbol{z},t}\right)^2}{\nu_t}\right)^{-\frac{\nu_t}{2}-\frac{1}{2}}. \end{aligned} \tag{47}$$

The finite-dimensional PDF can be derived in a similar manner. It is noted that the log-Student's t-distribution has divergent integer moments of any order. A short proof of this can be found in Appendix D. The parametrization of $Y(\boldsymbol{z})$ conditional on $\mathbf{M}$ is done by means of $\mu_{\boldsymbol{z},t}$, $\lambda_{\boldsymbol{z},t}$ and ν_t, i.e., in terms of the parameters of the underlying Student's t-RF $X(\boldsymbol{z})$. For $\nu_t\to\infty$, $f_{t,\ln}\left(y_{\boldsymbol{z}}|\mathbf{M}\right)$ converges to a lognormal distribution with location parameter $\mu_{\boldsymbol{z},t}$ and scale parameter $\lambda_{\boldsymbol{z},t}^{-\frac{1}{2}}$.

2.9. Sampling the posterior predictive random field

The finite-dimensional distribution of the posterior predictive RF is the multivariate Student's t-distribution with parameters $\boldsymbol{\mu}_{\mathbf{Z},t}, \boldsymbol{\Lambda}_{\mathbf{Z},t}$ and ν_t. The posterior predictive random vector $\boldsymbol{X}(\mathbf{Z})$ corresponding to locations $\mathbf{Z}$ can be expressed as follows [43]:

$$\boldsymbol{X}(\mathbf{Z}) = \frac{\boldsymbol{U}(\mathbf{Z})}{\sqrt{\frac{V}{\nu_t}}}+\boldsymbol{\mu}_{\mathbf{Z},t}, \tag{48}$$

where $\boldsymbol{U}(\mathbf{Z})$ is a zero-mean Gaussian random vector with precision matrix $\boldsymbol{\Lambda}_{\mathbf{Z},t}$. V is a random variable that follows the chi-square distribution with ν_t degrees of freedom and is independent of $\boldsymbol{U}(\mathbf{Z})$. Replacing $\boldsymbol{U}(\mathbf{Z})$ in Eq. (48) by $U(\boldsymbol{z})$, a zero-mean Gaussian RF with spatial precision function $\lambda_t\left(\boldsymbol{z}_1,\boldsymbol{z}_2\right)$ as given by Eq. (31), and furthermore replacing $\boldsymbol{\mu}_{\mathbf{Z},t}$ by $\mu_t(\boldsymbol{z})$, the spatial function for the mean value defined in Eq. (30),

yields the corresponding expression for the posterior predictive Student's t-RF. Hence, the Student's t-RF $X(\boldsymbol{z})$ can be expressed as a function of a Gaussian RF and one additional independent chi-square random variable. In case of sampling from a translation RF $Y(\boldsymbol{z})$, the transformation of Eq. (42) has to be adapted accordingly. Samples from $U(\boldsymbol{z})$ can be generated by a variety of existing methods [e.g. [6]].

2.10. Connection to the Bayesian approach of EN 1990

Annex D.7 of EN 1990 (Eurocode 0) on the basis of structural design offers a method to determine design values for material properties when samples are available [34]. The samples are used to estimate a quantile value of the underlying probability distribution, the so-called characteristic value. This approach distinguishes between the cases where (a) mean and variance of the material property are unknown and (b) only its variance is unknown. In case (a), the characteristic value can be estimated based on the sample mean, sample standard deviation and the number of samples n. The underlying theory is a Bayesian approach and the calculated value is the 5% quantile value of the posterior predictive distribution [2,33]. We show in the following that the hierarchical approach presented in this paper is a generalization of case (a) in Annex D.7 of EN 1990.

We consider a material property X that follows a normal distribution with unknown parameters $\boldsymbol{\theta}$ and that a set of samples $\boldsymbol{x}_m = \left[x_{m,1},\ x_{m,2},\ \dots,\ x_{m,n}\right]^{\mathsf{T}}$ are available. If no prior information about $f(\boldsymbol{\theta})$ is available, a non-informative choice can be made by choosing a $\mathcal{NG}$ distribution with the following parameters [42]:

$$\left[\mu_0, \kappa_0, \alpha_0, \beta_0\right] = \left[/, 0, -\frac{1}{2}, 0\right], \tag{49}$$

resulting in $f(\boldsymbol{\theta}) = \lambda_X^{-1}$.

Furthermore, we assume independence of the random variables corresponding to the measurement locations and neglect the measurement error, i.e., $\mathbf{R}_m = \mathbf{I}$. This leads to a simplification of Eqs. (10) to (13):

$$\mu_n = \frac{\mathbf{1}_n \boldsymbol{x}_m^{\mathsf{T}}}{n} = \frac{1}{n}\sum_{i=1}^{n} x_{m,i}, \tag{50}$$

$$\kappa_n = n, \tag{51}$$

$$\alpha_n = \frac{n-1}{2}, \tag{52}$$

$$\beta_n = \frac{1}{2}\left(\boldsymbol{x}_m \boldsymbol{x}_m^{\mathsf{T}} - \frac{\left(\mathbf{1}_n \boldsymbol{x}_m^{\mathsf{T}}\right)^2}{n}\right) = \frac{1}{2}\sum_{i=1}^{n}\left(x_{m,i} - \mu_n\right)^2. \tag{53}$$

If one neglects the dependence between the measurements and the RF at the predictive locations, the posterior predictive distribution $f\left(x|\boldsymbol{x}_m\right)$ is obtained following Eq. (14) and is space-invariant. It is a Student's t-distribution with the following parameters:

$$\mu_t = \mu_n = \frac{1}{n}\sum_{i=1}^{n} x_{m,i}, \tag{54}$$

$$\lambda_t = \frac{\alpha_n}{\beta_n\left(1+\kappa_n^{-1}\right)} = \frac{n(n-1)}{(n+1)\sum_{i=1}^{n}\left(x_{m,i}-\mu_n\right)^2}, \tag{55}$$

$$\nu_t = 2\alpha_n = n-1. \tag{56}$$

The characteristic values in the method in EN 1990 are defined as 5% quantile values of a Student's t-distribution with parameters given by Eqs. (54) to (56). An additional transformation step is added for ease of use, in which the Student's t-distributed random variable X is normalized:

$$U_t = \left(X - \mu_t\right)\lambda^{\frac{1}{2}}, \tag{57}$$

where U_t follows the standard Student's t-distribution with ν_t degrees of freedom, i.e., $\mu_t = 0$ and $\lambda_t = 1$. This normalization allows the use of standardized coefficients (k_n values), which only depend on n:

$$k_n = -F_{U_t}^{-1}(p)\sqrt{\frac{n+1}{n}}, \tag{58}$$

where $F_{U_t}^{-1}(\cdot)$ is the inverse CDF of U_t and $p = 0.05$, since the characteristic value x_k is defined as the 5% quantile value. Using the k_n value, x_k is obtained as follows:

$$x_k = \bar{\mu}_X\left(1 - k_n \bar{\delta}_X\right), \tag{59}$$

where $\bar{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_{m,i}$ is the sample mean and $\bar{\delta}_X = \frac{\bar{\mu}_X}{\bar{\sigma}_X}$ is the sample coefficient of variation with $\bar{\sigma}_X = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{m,i} - \bar{\mu}_X\right)^2$. EN 1990 provides tabulated values of k_n for varying n.

The method in EN 1990 also covers the case when the material property Y follows a lognormal distribution and $\boldsymbol{y}_m = \left[y_{m,1},\ y_{m,2}, \dots,\ y_{m,n}\right]^{\mathsf{T}}$ are the available samples. In this case, the Bayesian analysis underlying the method is conducted as described above for the Gaussian random variable $X = \ln(Y)$ with the logarithmic samples $x_{m,i} = \ln(y_{m,i})$, $i = 1, \dots, n$. The posterior predictive distribution $f\left(y|\boldsymbol{y}_m\right)$ is a log-Student's t-distribution parameterized in terms of the parameters of the underlying Student's t-distribution given by Eqs. (54) to (56). The characteristic value y_k is the 5% quantile value of $f\left(y|\boldsymbol{y}_m\right)$, which is equivalent to the exponential of the 5% quantile value of the underlying Student's t-distribution. Thus, y_k can be calculated as

$$y_k = \exp\left(\bar{\mu}_X\left(1 - k_n \bar{\delta}_X\right)\right), \tag{60}$$

where $\bar{\mu}_X$ and $\bar{\delta}_X$ are the sample mean and sample coefficient of variation of the logarithmic samples and k_n is given by Eq. (58).

In a nutshell, the method in Annex D.7 of EN 1990 to determine characteristic values for the design of structures is a special case of the presented RF analysis, which assumes a non-informative prior distribution, independent measurements without measurement error and independence between measurement locations and the material parameter at the predictive locations.

3. Numerical examples

In this section, the proposed approach is applied to two numerical examples. The first one involves a one-dimensional RF of a geotechnical material, while the second one models the concrete compressive strength of a ship lock wall with a two-dimensional anisotropic RF.

3.1. Tip resistance of cohesive soil

Soil parameters are often determined based on measurements from cone penetration testing (CPT). In CPT, the tip resistance q_T measures the force required to push the cone through the soil and can be used to infer further soil parameters. In this example, data from a CPT is used, where the tip resistance of a cohesive soil layer was measured in depths from $z = 3.900\,\text{m}$ to $z = 10.275\,\text{m}$ resulting in 256 equidistant measurements of the tip resistance. The data is taken from [53] and was also used by Wang and Zhao to illustrate the performance of Bayesian compressive sampling when sparse data is available [26]. The tip resistance is modeled by the one-dimensional RF $q_T(z)$ in vertical direction with lognormal prior marginal distribution. Hence, the transformation of Eq. (45) is applied:

$$q_T(z) = \exp\left(X(\boldsymbol{z})\right). \tag{61}$$

The underlying prior RF $X(\boldsymbol{z})$ is a homogeneous Gaussian RF with unknown mean value μ_X and unknown precision λ_X. The prior autocorrelation function is modeled by the exponential model with unknown correlation length l_c:

$$\rho\left(z_i, z_j\right) = \exp\left(-\frac{\left|z_j - z_i\right|}{l_c}\right) \tag{62}$$

Furthermore, no prior information on μ_X or λ_X are available and thus a non-informative prior $\mathcal{NG}$ distribution is chosen with the parameters from Eq. (49).

It is assumed that knowledge of the full data set is not available but only a subset of 13 measurement values taken at equidistant locations,

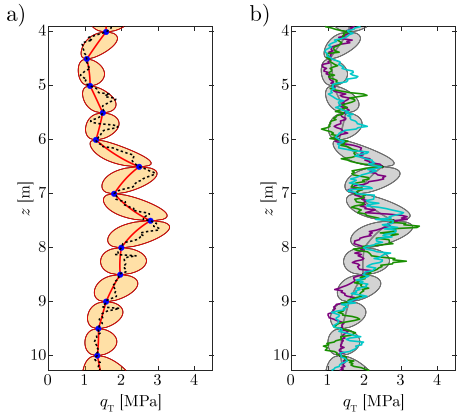

Fig. 3. Posterior predictive RF of the tip resistance q_T. Panel a) shows the median (red line) and the two-sided 90% credible interval, i.e., the area between the 5% and 95% quantile value (orange area) of the marginal log-Student's t-distributions. The 13 blue dots mark the used measurement locations and values while the full data set is illustrated by the dotted black line. Panel b) shows three independent realizations of the posterior predictive RF in comparison to the two-sided 90% credible interval in gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

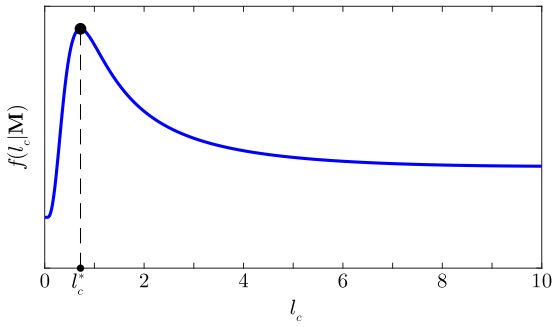

Fig. 5. Posterior distribution of the correlation length $f(l_c|\mathbf{M})$ as function of the correlation length l_c and the corresponding MAP estimate l_c^*.

as illustrated by the blue dots in panel a) of Figure 3. It is assumed that the measurements are associated with a multiplicative lognormal measurement error with median 1 and coefficient of variation $\mathrm{CV}_\varepsilon = 0.05$. In a first step, the MAP estimate for l_c is obtained by solving the minimization problem of Eq. (41), where the vector $\boldsymbol{\tau}$ only consists of l_c. A uniform prior on the positive numbers is employed for l_c and hence the term $\ln(f(\boldsymbol{\tau}))$ in the optimization problem can be dropped and the MAP estimate reduces to a maximum likelihood estimate [54]. The resulting estimate for l_c is obtained as $l_c^* = 0.72\,\mathrm{m}$.

Consequently, the posterior parameters of the $\mathcal{NG}$ distribution are obtained by application of Eqs. (10) to (13) in combination with Eq. (46) to account for the log-transformation of the measurements. The spatial parameter functions of the posterior predictive Student's t-RF are calculated by means of Eqs. (30) and (31). From Eq. (22) the degrees of freedom are calculated as $\nu_t = 12$. These are the parameters of the RF $q_T(z)$ given $\mathbf{M}$, which has log-Student's t-marginal distribution with PDF given by Eq. (47). As the moments are not defined, the illustration in panel a) of Figure 3 shows the median of the posterior predictive tip resistance and the corresponding 5% and 95% quantile values along the depth of the soil layer. The increasing width of the orange area shows that the uncertainty is very small close to the measurement locations and increases away from the measurements. The full data set of 256 measurements is indicated by a black dotted line. Panel b) of Fig. 3 shows three independent realizations of the posterior predictive RF. Comparison of the random realizations with the full data set shows good accordance regarding the number and amplitude of strong local deviations from the posterior predictive median. Hence, the proposed approach can sufficiently approximate both the overall trend of the RF and the associated uncertainty.

To illustrate the influence of the number of measurements on the posterior prediction, the above calculations are repeated for $n = 6$ and $n = 64$ equidistant measurements. Fig. 4 illustrates the measurement values and locations by blue dots in panel a) and panel c), respectively. For $n = 6$, the MAP optimization results in $l_{c,6}^* = 4.60\,\mathrm{m}$ and for $n = 64$ it gives $l_{c,64}^* = 1.51\,\mathrm{m}$. This large difference in the MAP estimates is due to the assumed uninformative prior distribution for the correlation length, in which case, the MAP estimate only depends on the data. Large differences in the data can lead to significant variation in the estimated correlation length. The median and corresponding 5% and 95% quantile values of $q_T(z)$ are illustrated in panel a) and c), respectively of Fig. 4. Comparison to Fig. 3 shows that with increasing amount of data, the uncertainty, i.e., the variability of $q_T(z)$ is reduced. However, even with a small amount of data ($n = 6$), the global trend of the tip resistance can be predicted and the location-specific information can be used efficiently to set up an RF model. The large variability in the areas between the measurements is illustrated by three independent realizations in panel b) of Fig. 4. When the amount of data is relatively large ($n = 64$), the remaining uncertainty in the tip resistance becomes

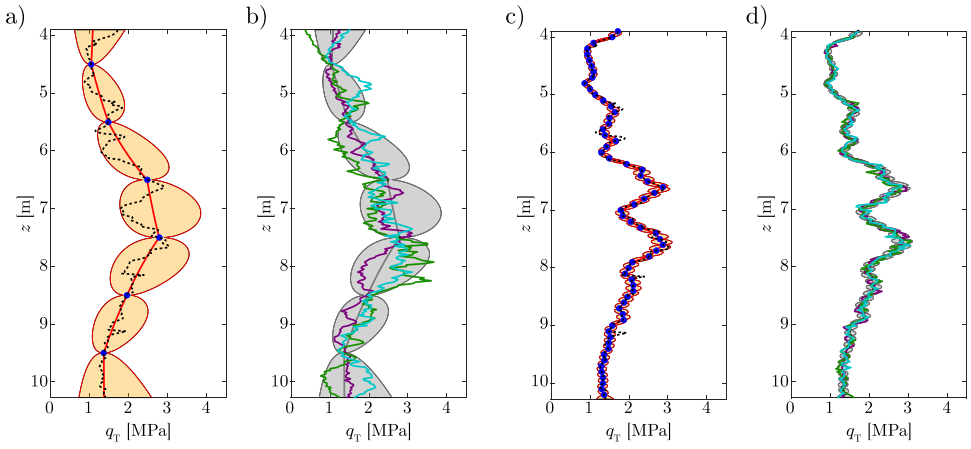

Fig. 4. Posterior predictive RF of the tip resistance q_T for $n = 6$ (panel a) and b)) and $n = 64$ (panel c) and d)). Panel a) and c) show the median (red line) and the two-sided 90% credible intervals of the marginal log-Student's t-distributions. The blue dots mark the used measurement locations and values while the full data set is illustrated by the dotted black line. Panel b) and c) each show three independent realizations of the posterior predictive RF in comparison to the two-sided 90% credible intervals in gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

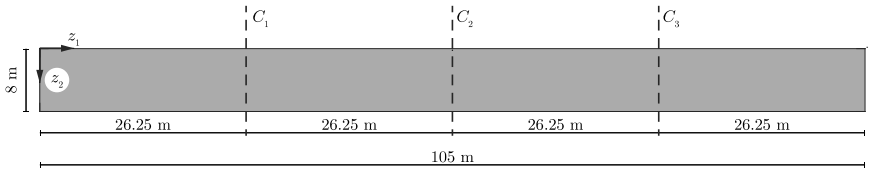

Fig. 6. Ship lock wall with a total length of 105 m and a total height of 8 m made of tamped concrete from the 1920s. Three vertical core samples (C_1, C_2 and C_3) were taken at the quarter points of the wall indicated by the three dashed lines.

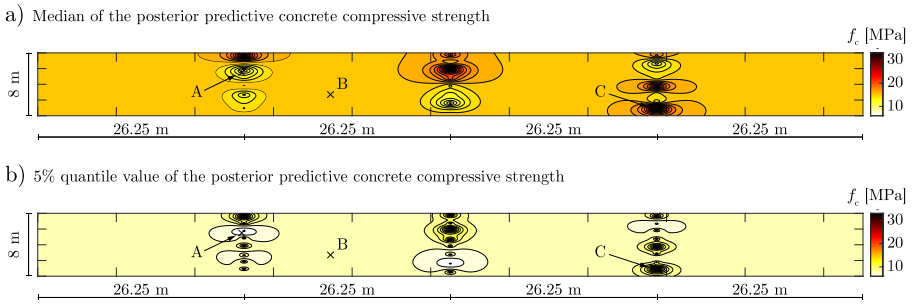

Fig. 7. Posterior predictive median (panel a)) and 5% quantile value (panel b)) of the concrete compressive strength f_c of a ship lock wall obtained with data from three vertical core samples ($n = 24$ measurements of the concrete compressive strength). The median and 5% quantile value at points A, B and C are listed in Table 2.

Table 1
Measurements of the concrete compressive strength f_c and the corresponding measurement locations of 24 specimens from 3 vertical core samples (C_1, C_2 and C_3) in the quarter points of the ship lock wall.

Core sample C_1			Core sample C_2			Core sample C_3		
z_1 [m]	z_2 [m]	$f_{c,m}$ [MPa]	z_1 [m]	z_2 [m]	$f_{c,m}$ [MPa]	z_1 [m]	z_2 [m]	$f_{c,m}$ [MPa]
26.25	0.40	29.2	52.5	0.21	21.2	78.75	0.34	18.5
	1.24	15.5		1.25	16.0		1.34	10.3
	2.25	8.7		2.05	32.0		2.17	13.2
	3.15	12.3		3.33	20.7		3.24	14.5
	4.12	16.2		4.15	13.8		4.27	25.4
	5.33	11.6		5.25	12.1		5.12	14.5
	6.15	13.4		6.40	8.6		6.23	13.2
	7.05	13.9		7.45	14.8		7.08	33.0

comparatively small and random realizations of the RF do not differ significantly from the full data set, as can be seen in panel d) of Fig. 4.

Fig. 5 plots $f\left(l_c|\mathbf{M}\right)$ with the MAP estimate $l_c^* = 0.72$ m located at the mode of $f\left(l_c|\mathbf{M}\right)$. It appears that, although the posterior distribution has a distinct mode, it covers a broad range by remaining relatively flat for increasing values of l_c. This is caused by the uniform prior distribution for l_c and shows that such a uniform prior can lead to an improper posterior distribution of the correlation length. While this is not a problem when using MAP, it is an issue when the full posterior distribution of l_c is to be used. In such cases, a different prior distribution should be chosen.

3.2. Concrete compressive strength of a ship lock wall

In this example, we investigate the concrete compressive strength f_c of a ship lock wall made of tamped concrete in the 1920s. The length of the wall is 105 m and the height of the tamped concrete layer is 8 m, the third dimension is not taken into account for this study. 24 measurements of f_c are available from three vertical core samples taken at the quarter points of the wall [2]. The situation is illustrated in Fig. 6 and the measurement data and corresponding locations are shown in Table 1. We assume that the measurements are associated with a multiplicative lognormal measurement error with coefficient of variation $\mathrm{CV}_\epsilon = 0.025$.

Applying the transformation of Eq. (45), the logarithm of f_c is modeled with a two-dimensional Gaussian RF with non-informative prior $\mathcal{NG}$ distribution (cf. Eq. (49)).

Typically, massive concrete structures made of tamped concrete from that time have been built in layers [55]. Hence, we employ a transverse anisotropic exponential correlation function, where the correlation length $l_{c,1}$ in direction z_1 differs from the correlation length $l_{c,2}$ in direction z_2 [56]:

$$\rho\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right) = \exp\left(-\sqrt{\frac{\left(\Delta_1\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)\right)^2}{l_{c,1}^2} + \frac{\left(\Delta_2\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)\right)^2}{l_{c,2}^2}}\right), \tag{63}$$

where $\Delta_1\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)$ and $\Delta_2\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)$ denote the canonical distances of $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ in directions z_1 and z_2 respectively. Assuming a uniform prior on $l_{c,1}$ and $l_{c,2}$ results in the following MAP estimate for the two correlation lengths:

$$\boldsymbol{l}_c^* = \left[l_{c,1}^*,\ l_{c,2}^*\right] = [1.52\,\mathrm{m}, 0.58\,\mathrm{m}] \tag{64}$$

These values are used in the Bayesian updating to obtain the posterior predictive RF for f_c. As the marginal posterior predictive PDF is a log-Student's t-distribution, the moments cannot be evaluated and thus, Fig. 7 illustrates the median (panel a)) and the corresponding 5% quantile value (panel b)) of $f_c(z)$ given $\mathbf{M}$ across the ship lock wall. The measured values and the information about their location are clearly reflected, as regions close to low measurement values show low median and 5% quantile values, and regions close to high measurement values show higher median and 5% quantile values. This is illustrated by the example of three points (A, B and C) at different locations of the ship lock wall, where the median and 5% quantile values have been extracted and listed in Table 2. Point A, located close to a low measurement value, features a posterior median of 10.7 MPa and a 5% quantile value of 6.6 MPa, both of which are significantly lower than those at point C with a median of 23.1 MPa and a 5% quantile value of 14.1 MPa. Contrary to point A, point C is located close to a high measurement value (cf. Table 1). The median of 15.9 MPa and 5% quantile of 8.2 MPa at point B are representative values for all locations far away from the measurements, i.e., all points with negligible spatial correlation to any measurement location.

Section 2.10 demonstrates the connection of the proposed RF approach and the established Bayesian approach in EN 1990. Next, we compare this approach to the results of the proposed hierarchical RF model using the data of Table 1. The mean and standard deviation of the log-transformed measurement values are $\bar{\mu}_X = 2.75$ and $\bar{\sigma}_X = 0.37$ with a corresponding k_n value of $k_n(n_m = 24) = 1.75$. Applying Eq. (60)

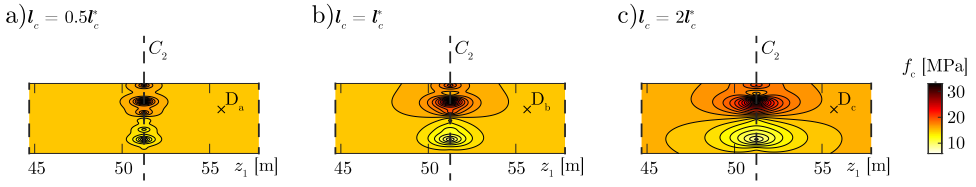

Fig. 8. Posterior predictive median of the concrete compressive strength of a ship lock wall in the area around core sample C_2 obtained with varying correlation lengths.

Table 2
Median (50% quantile value, $f_{c,0.5}$) and 5% quantile value ($f_{c,0.05}$) of the marginal posterior predictive concrete compressive strength at three different locations (A, B and C) of the ship lock wall.

	z_1 [m]	z_2 [m]	$f_{c,0.5}$ [MPa]	$f_{c,0.05}$ [MPa]
Point A	25.95	2.5	10.7	6.6
Point B	37.25	5.3	15.9	8.2
Point C	78.3	6.85	23.1	14.1

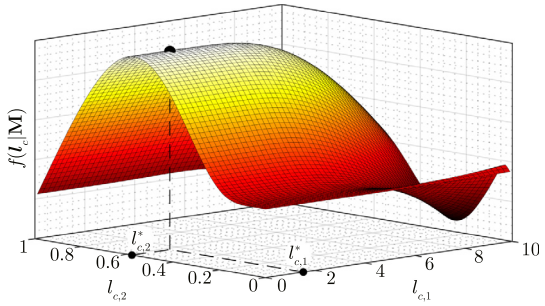

Fig. 9. Two-dimensional posterior distribution of the correlation lengths in z_1 (horizontal) and z_2 (vertical) direction, $f(\boldsymbol{l}_c|\mathbf{M})$. The maximum of $f(\boldsymbol{l}_c|\mathbf{M})$ is located at $l_{c,1} = 1.53$ m and $l_{c,2} = 0.58$ m, which is equivalent to the MAP estimate $\boldsymbol{l}_c^*$.

gives a characteristic value (5% quantile value) of $f_{c,k} = 8.2$ MPa. This value matches the 5% quantile value at locations without spatial correlation to the measurement points (cf. point C in Table 2). We note that this congruence depends on the chosen prior parameters of the RF and, thus, is the exception, not the rule.

The correlation length is an important parameter in any RF model. To illustrate this, the Bayesian analysis has been carried out for $\boldsymbol{l}_c = 0.5\boldsymbol{l}_c^*$ and $\boldsymbol{l}_c = 2\boldsymbol{l}_c^*$. The resulting marginal median of $f_c(\boldsymbol{z})$ in the area around the core sample C_2 is illustrated in Fig. 8. Obviously, the larger the correlation length, the bigger the area that is influenced by the spatial correlation to the measurements. For $\boldsymbol{l}_c = 0.5\boldsymbol{l}_c^*$ the spatial effect of the measurements on the median is restricted to a domain of length ≈ 2 m, wheres for $\boldsymbol{l}_c = 2\boldsymbol{l}_c^*$ this effect spans over a length of ≈ 10 m. It is mentioned that this is the effect of the final step of the Bayesian approach, where the posterior predictive distribution is obtained. The whole RF is influenced by the data and the chosen correlation length by the global posterior parameters of the $\mathcal{NG}$ distribution, as can be seen by the different median values of $f_c(\boldsymbol{z})$ at locations D_a, D_b and D_c indicated in Fig. 8. These locations are chosen exemplarily for all points with negligible spatial correlation to any measurement location. For $\boldsymbol{l}_c = 0.5\boldsymbol{l}_c^*$ the median is 15.7 MPa and for $\boldsymbol{l}_c = 2\boldsymbol{l}_c^*$ it is 16.3 MPa, compared to 15.9 MPa when $\boldsymbol{l}_c = \boldsymbol{l}_c^*$.

When employing the MAP procedure to approximate $f(\boldsymbol{\tau}|\mathbf{M})$, it is important to be aware of the sensitivity of the estimate and the amount of information provided by the data. In this example, the vertical distance of the measurement locations is relatively small while the horizontal distance is either 0 or very large. Hence, the MAP estimate for $l_{c,1}$ is subject to larger uncertainty than the MAP estimate for $l_{c,2}$, which is illustrated in Fig. 9 . While $f(\boldsymbol{\tau}|\mathbf{M})$ has a distinct maximum in direction $l_{c,2}$ at $l_{c,2} = 0.58$ m, it is relatively flat in direction $l_{c,1}$. In fact, any $l_{c,1}$ smaller than 5 m is approximately equally likely given the data at hand. Only for $l_{c,1} > 5$ m the measurements of different core samples are noticeably correlated. This behavior is of special interest when no prior information on the correlation length is assumed, since in such case the MAP estimate is only controlled by the data. In general, learning the correlation length from limited amount of data is not a trivial task, especially if no prior information on the RF parameters is available. In a study on the correlation length of soil parameters, a minimum of 5 measurement values within one correlation length are recommended for learning the correlation length of the exponential correlation model [57].

4. Conclusion

This paper presents a comprehensive hierarchical Bayesian approach to model random material properties with spatially distributed data. It is based on modeling a Gaussian random field assuming a normal-gamma prior distribution on its parameters. Closed-form expressions for the posterior normal-gamma distribution of the parameters of the random field are derived by making use of the conjugacy of the normal-gamma distribution and a multivariate Gaussian likelihood function. Subsequently, closed-form expressions for the spatial parameter function of the posterior predictive random field are derived, resulting in a non-homogeneous Student's t-random field. That is, the marginal distribution of the posterior predictive random field is a Student's t-distribution with location-specific parameters.

Sampling from such a random field can be achieved by expressing the Student's t-random field in terms of a Gaussian random field and one additional chi-squared random variable. For estimating the correlation parameters, a maximum a-posteriori estimation approach is proposed that accounts for the available data and potential prior information. In addition, an extension of the approach to non-Gaussian translation prior random fields is discussed and closed-form expressions for the case of a lognormal marginal prior distribution are derived.

The applicability of the presented approach to different engineering fields is illustrated by two examples, one from the field of geotechnical engineering and one from structural engineering. The derived posterior random field models reflect the location-specific information from the measurements, whereas their uncertainty increases with increasing distance from the measurement locations. Furthermore, it is demonstrated that the uncertainty can be reduced by increasing the amount of data. The spatial fluctuation of the posterior random field is sensitive to the choice of the correlation length parameter. When no information is available on the prior autocorrelation function, the maximum a-posteriori estimate for the correlation length is sensitive to the measurement data and should be handled with care, especially in the case where limited data is available.

The presented modeling approach can be extended to account for a trend function in the prior random field parameters. A trend in the prior mean can be included by employing a linear basis function model, similar to the work of [23]. A parametric dependence can also be included in the prior precision parameter, which leads to a model known as weighted Bayesian linear regression [7]. Investigation of these models in the context of material modeling is left to future studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been financially supported by the *Bundesanstalt für Wasserbau* (Federal Waterways Engineering and Research Institute, Germany). The corresponding author would like to thank Claus Kunz from the *Bundesanstalt für Wasserbau* who motivated the idea to develop the presented approach and Prof. Lori Graham-Brady from Johns Hopkins University for many discussions that have considerably enhanced the quality of the paper.

Appendix A. The posterior normal-gamma distribution

The posterior parameters of the normal-gamma distribution for the parameter vector $\boldsymbol{\theta} = \left[\mu_X, \lambda_X\right]^\mathsf{T}$, as specified in Eqs. (10) to (13) are derived in the following.

According to Bayes' theorem, the posterior distribution $f(\boldsymbol{\theta}|\mathbf{M})$ is proportional to the product of prior distribution $f(\boldsymbol{\theta})$ and likelihood $\mathrm{L}(\boldsymbol{\theta}|\mathbf{M})$, which are defined in Eqs. (3) and (5). Using the normal-gamma prior distribution and the multivariate Gaussian likelihood gives the following expression:

$$f(\boldsymbol{\theta}|\mathbf{M}) \propto \lambda_X^{\alpha_0+\frac{n}{2}-\frac{1}{2}} \cdot \exp\left(-\frac{\kappa_0\lambda_X}{2}\left(\mu_X-\mu_0\right)^2\right) \cdot \exp\left(-\lambda_X\beta_0\right) \cdot \cdot \exp\left(-\frac{\lambda_X}{2}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)^\mathsf{T}\right). \tag{A.1}$$

With the definition of $\mathbf{A} = \kappa_0\left(\mu_X-\mu_0\right)^2 - 2\mu_X\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T} + \mu_X^2\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}$, Eq. (A.1) can be rearranged as

$$f(\boldsymbol{\theta}|\mathbf{M}) \propto \lambda_X^{\alpha_0+\frac{n}{2}-\frac{1}{2}} \cdot \exp\left(-\lambda_X\left(\beta_0+\frac{1}{2}\boldsymbol{x}_m\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}+\frac{1}{2}\mathbf{A}\right)\right). \tag{A.2}$$

Initially, the focus lies on $\mathbf{A}$ which is expanded and modified as follows:

$$\mathbf{A} = \left(\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}\right)\left(\mu_X^2 - 2\mu_X\frac{\kappa_0\mu_0+\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}}{\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}}\right) + \kappa_0\mu_0^2. \tag{A.3}$$

Next, the square of the expression inside the second parenthesis is completed:

$$\mathbf{A} = \underbrace{\left(\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}\right)\left(\mu_X-\frac{\kappa_0\mu_0+\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}}{\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}}\right)^2}_{\mathbf{B}} + \underbrace{\kappa_0\mu_0^2 - \frac{\left(\kappa_0\mu_0+\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}\right)^2}{\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}}}_{\mathbf{C}}. \tag{A.4}$$

The terms of $\mathbf{C}$ in Eq. (A.4) are expanded and converted to a common denominator:

$$\mathbf{C} = \left(\kappa_0\mu_0^2\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T} - 2\kappa_0\mu_0\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T} - \left(\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}\right)^2\right)\left(\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}\right)^{-1}. \tag{A.5}$$

Inserting the expression for $\mathbf{B}$ and $\mathbf{C}$ into Eq. (A.2) gives:

$$f(\boldsymbol{\theta}|\mathbf{M}) \propto \lambda_X^{\alpha_0+\frac{n}{2}-\frac{1}{2}} \cdot \exp\left(-\lambda_X\left(\beta_0+\frac{1}{2}\boldsymbol{x}_m\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}+\frac{1}{2}\mathbf{C}\right)\right) \cdot \exp\left(-\frac{\lambda_X}{2}\mathbf{B}\right). \tag{A.6}$$

The parametric form of the posterior normal-gamma distribution as defined in Section 2.4 is as follows:

$$\mathcal{NG}\left(\mu_X,\lambda_X|\mu_n,\kappa_n,\alpha_n,\beta_n\right) = C_n\lambda_X^{\alpha_n-\frac{1}{2}} \cdot \exp\left(-\frac{\kappa_n\lambda_X}{2}\left(\mu_X-\mu_n\right)^2\right) \cdot \exp\left(-\lambda_X\beta_n\right). \tag{A.7}$$

Writing out all the terms in Eq. (A.6) and comparing it to (A.7) one can see that up to the normalizing constant C_n, the resulting expression of Eq. (A.6) is a normal-gamma distribution with parameters as follows:

$$\mu_n = \frac{\kappa_0\mu_0+\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}}{\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}}, \tag{A.8}$$

$$\kappa_n = \kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}, \tag{A.9}$$

$$\alpha_n = \alpha_0+\frac{n}{2}, \tag{A.10}$$

$$\beta_n = \beta_0+\frac{1}{2}\left(\boldsymbol{x}_m\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T} + \frac{\kappa_0\mu_0^2\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T} - 2\kappa_0\mu_0\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T} - \left(\mathbf{1}_n\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}\right)^2}{\kappa_0+\mathbf{1}_n\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}}\right). \tag{A.11}$$

The normalizing constant is

$$C_n = \frac{\beta_n^{\alpha_n}\kappa_n^{\frac{1}{2}}}{\Gamma\left(\alpha_n\right)\left(2\pi\right)^{\frac{1}{2}}}. \tag{A.12}$$

Appendix B. The marginal posterior predictive Student's t-distribution

In Section 2.5, the Student's t-distribution is introduced as the marginal posterior predictive distribution of the RF $X(\boldsymbol{z})$ for the normal-gamma conjugate prior distribution of the RF parameters. This appendix derives the analytical expressions for the parameters of the marginal posterior predictive distribution as given in Eqs. (20) to (22).

The marginal posterior predictive PDF at any point $\boldsymbol{z} \in \Omega$ is defined by

$$f(x;\boldsymbol{z}|\mathbf{M}) = \int_{\boldsymbol{\Theta}} f(x;\boldsymbol{z}|\boldsymbol{\theta},\mathbf{M})\, f(\boldsymbol{\theta}|\mathbf{M})\,\mathrm{d}\boldsymbol{\theta}. \tag{B.1}$$

$f(\boldsymbol{\theta}|\mathbf{M})$ is the posterior normal-gamma distribution as defined in Eq. (8) and $f(x;\boldsymbol{z}|\boldsymbol{\theta},\mathbf{M})$ is a location-specific normal distribution with parameters $\mu_{\boldsymbol{z}}''$ and $\lambda_{\boldsymbol{z}}''$ given by Eqs. (16) and (17). Hence, Eq. (B.1) can be expanded as follows:

$$f(x;\boldsymbol{z}|\mathbf{M}) = \int_{\lambda_X=0}^{+\infty}\int_{\mu_X=-\infty}^{+\infty} \mathcal{N}\left(x|\mu_{\boldsymbol{z}}'',\lambda_{\boldsymbol{z}}''\right)\mathcal{N}\left(\mu_X|\mu_n,\kappa_n\lambda_X\right)\mathrm{d}\mu_X\,\mathcal{G}\left(\lambda_X|\alpha_n,\beta_n\right)\mathrm{d}\lambda_X. \tag{B.2}$$

The inner integral involves the convolution of two normal densities:

$$\int_{\mu_X=-\infty}^{+\infty} \mathcal{N}\left(x|\mu_{\boldsymbol{z}}'',\lambda_{\boldsymbol{z}}''\right)\mathcal{N}\left(\mu_X|\mu_n,\kappa_n\lambda_X\right)\mathrm{d}\mu_X = f\left(x;\boldsymbol{z}|\lambda_X,\mathbf{M}\right). \tag{B.3}$$

For the solution of the integral, the expression for $\mu_{\boldsymbol{z}}''$, given in Eq. (16) is rewritten as follows:

$$\mu_{\boldsymbol{z}}'' = \mu_X + \boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)^\mathsf{T} = \underbrace{\mu_X\left(1-\boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}\right)}_{\psi} + \underbrace{\boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\boldsymbol{x}_m^\mathsf{T}}_{\xi}. \tag{B.4}$$

For this special case and noting that $\lambda_{\boldsymbol{z}}''$ does not depend on μ_X, the marginalization in Eq. (B.3) can be solved analytically and results in a normal density $f\left(x;\boldsymbol{z}|\lambda_X,\mathbf{M}\right) = \mathcal{N}\left(x|\tilde{\mu}_{\boldsymbol{z}},\tilde{\lambda}_{\boldsymbol{z}}\right)$, where $\tilde{\mu}_{\boldsymbol{z}}$ and $\tilde{\lambda}_{\boldsymbol{z}}$ are given by the following equations [36]:

$$\tilde{\mu}_{\boldsymbol{z}} = \psi\mu_n + \xi, \tag{B.5}$$

$$\tilde{\lambda}_{\boldsymbol{z}} = \left(\left(\lambda_{\boldsymbol{z}}''\right)^{-1} + \psi^2\lambda_X^{-1}\kappa_n^{-1}\right)^{-1} = \underbrace{\lambda_X\left(1-\boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\boldsymbol{R}_{\boldsymbol{z},m}^\mathsf{T} + \left(1-\boldsymbol{R}_{\boldsymbol{z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^\mathsf{T}\right)^2\kappa_n^{-1}\right)^{-1}}_{\tilde{\kappa}_{\boldsymbol{z}}}. \tag{B.6}$$

Inserting in Eq. (B.2) results in

$$f(x;\boldsymbol{z}|\mathbf{M}) = \frac{\beta_n^{\alpha_n}\left(\tilde{\kappa}_{\boldsymbol{z}}\right)^{\frac{1}{2}}}{\Gamma\left(\alpha_n\right)(2\pi)^{\frac{1}{2}}}\int_{\lambda_X=0}^{+\infty}\lambda_X^{\alpha_n-\frac{1}{2}}\exp\left(-\lambda_X\left(\beta_n+\frac{\tilde{\kappa}_{\boldsymbol{z}}}{2}\left(x_{\boldsymbol{z}}-\tilde{\mu}_{\boldsymbol{z}}\right)^2\right)\right)\mathrm{d}\lambda_X. \tag{B.7}$$

A solution of the integral in Eq. (B.7) is readily available and the resulting expression is as follows [36]:

$$f(x;\boldsymbol{z}|\mathbf{M}) = \frac{\beta_n^{\alpha_n}\left(\tilde{\kappa}_{\boldsymbol{z}}\right)^{\frac{1}{2}}}{\Gamma\left(\alpha_n\right)(2\pi)^{\frac{1}{2}}}\left(\beta_n+\frac{\tilde{\kappa}_{\boldsymbol{z}}}{2}\left(x_{\boldsymbol{z}}-\tilde{\mu}_{\boldsymbol{z}}\right)^2\right)^{-\frac{1}{2}-\alpha_n}\Gamma\left(\alpha_n+\frac{1}{2}\right). \tag{B.8}$$

To bring $f(x;\boldsymbol{z}|\mathbf{M})$ into a standardized format, we define $\mu_{z,t} = \tilde{\mu}_{\boldsymbol{z}}$, $\lambda_{z,t} = \frac{\tilde{\kappa}_{\boldsymbol{z}}\alpha_n}{\beta_n}$ and $\nu_t = 2\alpha_n$ [36]. This gives

$$f(x;\boldsymbol{z}|\mathbf{M}) = \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\left(\frac{\lambda_{z,t}}{\pi\nu_t}\right)^{\frac{1}{2}}\left(1+\frac{\lambda_{z,t}}{\nu_t}\left(x_{\boldsymbol{z}}-\mu_{z,t}\right)^2\right)^{-\frac{\nu_t}{2}-\frac{1}{2}}. \tag{B.9}$$

Eq. (B.9) describes the marginal posterior predictive distribution of the RF $X(\boldsymbol{z})$ given measurement data $\mathbf{M}$, which is a Student's t-distribution with location parameter $\mu_{z,t}$, scale parameter $\lambda_{z,t}$ and degrees of freedom ν_t defined as follows:

$$\mu_{z,t} = \mu_n + \boldsymbol{R}_{z,m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m-\mu_n\mathbf{1}_n\right)^{\mathsf{T}}, \tag{B.10}$$

$$\lambda_{z,t} = \frac{\alpha_n}{\beta_n\left(1-\boldsymbol{R}_{z,m}\mathbf{R}_m^{-1}\boldsymbol{R}_{z,m}^{\mathsf{T}}+\left(1-\boldsymbol{R}_{z,m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)^2\kappa_n^{-1}\right)}, \tag{B.11}$$

$$\nu_t = 2\alpha_n, \tag{B.12}$$

where μ_n, κ_n, α_n and β_n are the posterior parameters of the normal-gamma distribution given by Eqs. (10) to (13).

Appendix C. The multivariate posterior predictive student's t-distribution

This section extends the derivation of Appendix B to the multivariate case to derive the parameters for the kth order posterior predictive Student's t-distribution as given by Eqs. (22), (28) and (29) in Section 2.6.

The PDF of the posterior predictive distribution of the RF $X(\boldsymbol{z})$ is

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta},\mathbf{M})\, f(\boldsymbol{\theta}|\mathbf{M})\,\mathrm{d}\boldsymbol{\theta}, \tag{C.1}$$

with $\boldsymbol{x}\in\mathbb{R}^k$ and $\mathbf{Z} = [\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k]\in\mathbb{R}^{k\times d}$ denoting any set of spatial points in Ω. $f(\boldsymbol{x};\mathbf{Z}|\boldsymbol{\theta},\mathbf{M})$ is a k-variate normal density with mean vector $\boldsymbol{\mu}_{\mathbf{Z}}''$ and precision matrix $\boldsymbol{\Lambda}_{\mathbf{Z}}''$ given by Eqs. (24) and (25). $f(\boldsymbol{\theta}|\mathbf{M})$ is a normal-gamma distribution as defined in Eq. (8) and is independent of the locations $\mathbf{Z}$. Eq. (C.1) is expanded as follows:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\lambda_X=0}^{+\infty}\int_{\mu_X=-\infty}^{+\infty}\mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_{\mathbf{Z}}'',\boldsymbol{\Lambda}_{\mathbf{Z}}''\right)\cdot\mathcal{N}\left(\mu_X|\mu_n,\lambda_X\kappa_n\right)\mathrm{d}\mu_X\cdot\mathcal{G}\left(\lambda_X|\alpha_n,\beta_n\right)\mathrm{d}\lambda_X. \tag{C.2}$$

The inner integral can be solved by rewriting Eq. (24) as follows:

$$\boldsymbol{\mu}_{\mathbf{Z}}'' = \mu_X\mathbf{1}_k + \mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m-\mu_X\mathbf{1}_n\right)^{\mathsf{T}} = \mu_X\underbrace{\left(\mathbf{1}_k-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)}_{\boldsymbol{\psi}}+\underbrace{\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\boldsymbol{x}_m^{\mathsf{T}}}_{\boldsymbol{\xi}}. \tag{C.3}$$

Using this expression, the integration over μ_X can be performed analytically and results in the density of a multivariate normal distribution $\mathcal{N}\left(\boldsymbol{x}|\tilde{\boldsymbol{\mu}}_{\mathbf{Z}},\tilde{\boldsymbol{\Lambda}}_{\mathbf{Z}}\right)$ with parameters given as [36]

$$\tilde{\boldsymbol{\mu}}_{\mathbf{Z}} = \mu_n\boldsymbol{\psi}+\boldsymbol{\xi}, \tag{C.4}$$

$$\tilde{\boldsymbol{\Lambda}}_{\mathbf{Z}} = \left(\left(\boldsymbol{\Lambda}_{\mathbf{Z}}''\right)^{-1}+\boldsymbol{\psi}^{\mathsf{T}}\left(\lambda_X\kappa_n\right)^{-1}\boldsymbol{\psi}\right)^{-1}. \tag{C.5}$$

Substituting Eqs. (25) and (C.3) into Eq. (C.5), $\tilde{\boldsymbol{\Lambda}}_{\mathbf{Z}}$ can be expressed as the following linear function of λ_X:

$$\tilde{\boldsymbol{\Lambda}}_{\mathbf{Z}} = \lambda_X\underbrace{\left(\mathbf{R}_{\mathbf{Z}}-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{R}_{\mathbf{Z},m}^{\mathsf{T}}+\left(\mathbf{1}_k-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)^{\mathsf{T}}\kappa_n^{-1}\left(\mathbf{1}_k-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)\right)^{-1}}_{\tilde{\mathbf{K}}_{\mathbf{Z}}}. \tag{C.6}$$

Inserting $\mathcal{N}\left(\boldsymbol{x}|\tilde{\boldsymbol{\mu}}_{\mathbf{Z}},\tilde{\boldsymbol{\Lambda}}_{\mathbf{Z}}\right)$ into Eq. (C.2) gives

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\lambda_X=0}^{+\infty}\mathcal{N}\left(\boldsymbol{x}|\tilde{\boldsymbol{\mu}}_{\mathbf{Z}},\tilde{\mathbf{K}}_{\mathbf{Z}}\lambda_X\right)\mathcal{G}\left(\lambda_X|\alpha_n,\beta_n\right)\mathrm{d}\lambda_X. \tag{C.7}$$

Next, an alternative parametrization is introduced, defining $\nu_t = 2\alpha_n$ and $\eta = \frac{\lambda_X\beta_n}{\alpha_n}$. Inserted into Eq. (C.7), this gives the following [36]:

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \int_{\eta=0}^{+\infty}\mathcal{N}\left(\boldsymbol{x}\middle|\tilde{\boldsymbol{\mu}}_{\mathbf{Z}},\tilde{\mathbf{K}}_{\mathbf{Z}}\frac{\eta\alpha_n}{\beta_n}\right)\mathcal{G}\left(\eta\middle|\frac{\nu_t}{2},\frac{\nu_t}{2}\right)\mathrm{d}\eta, \tag{C.8}$$

for which a solution is available [36]. The resulting expression is

$$f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = \frac{\Gamma\left(\frac{k+\nu_t}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\left|\tilde{\mathbf{K}}_{\mathbf{Z}}\right|^{\frac{1}{2}}\left(\frac{\alpha_n}{\beta_n\pi\nu_t}\right)^{\frac{k}{2}}\left(1+\frac{\alpha_n\left(\boldsymbol{x}-\tilde{\boldsymbol{\mu}}_{\mathbf{Z}}\right)\tilde{\mathbf{K}}_{\mathbf{Z}}\left(\boldsymbol{x}-\tilde{\boldsymbol{\mu}}_{\mathbf{Z}}\right)^{\mathsf{T}}}{\beta_n\nu_t}\right)^{-\frac{k+\nu_t}{2}}, \tag{C.9}$$

which is a k-variate Student's t-distribution with parameters $\tilde{\boldsymbol{\mu}}_{\mathbf{Z}}$, $\frac{\alpha_n}{\beta_n}\tilde{\mathbf{K}}_{\mathbf{Z}}$ and ν_t. Defining $\boldsymbol{\mu}_{\mathbf{Z},t} = \tilde{\boldsymbol{\mu}}_{\mathbf{Z}}$ and $\boldsymbol{\Lambda}_{\mathbf{Z},t} = \frac{\alpha_n}{\beta_n}\tilde{\mathbf{K}}_{\mathbf{Z}}$ yields the expression of Eq. (27) for the kth order posterior predictive distribution of $X(\boldsymbol{z})$ given measurement data $\mathbf{M}$. That is, $f(\boldsymbol{x};\mathbf{Z}|\mathbf{M}) = f_t\left(\boldsymbol{x}|\boldsymbol{\mu}_{\mathbf{Z},t},\boldsymbol{\Lambda}_{\mathbf{Z},t},\nu_t\right)$ with parameters given as

$$\boldsymbol{\mu}_{\mathbf{Z},t} = \mu_n\mathbf{1}_k+\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\left(\boldsymbol{x}_m-\mu_n\mathbf{1}_n\right)^{\mathsf{T}}, \tag{C.10}$$

$$\boldsymbol{\Lambda}_{\mathbf{Z},t} = \frac{\alpha_n}{\beta_n}\left(\mathbf{R}_{\mathbf{Z}}-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{R}_{\mathbf{Z},m}^{\mathsf{T}}+\left(\mathbf{1}_k-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)\kappa_n^{-1}\left(\mathbf{1}_k-\mathbf{R}_{\mathbf{Z},m}\mathbf{R}_m^{-1}\mathbf{1}_n^{\mathsf{T}}\right)^{\mathsf{T}}\right)^{-1}, \tag{C.11}$$

$$\nu_t = 2\alpha_n, \tag{C.12}$$

where μ_n, κ_n, α_n and β_n are the posterior parameters of the normal-gamma distribution given by Eqs. (10) to (13).

Appendix D. The log-student's t-distribution

In Section 2.8, the log-Student's t-distribution is introduced as resulting marginal distribution of the posterior predictive RF when the prior RF has lognormal marginal distribution. In this appendix, the log-Student's t-distribution and some of its properties are described.

When X follows a Student's t-distribution, $Y = \exp(X)$ follows the log-Student's t-distribution [50,52]. The PDF can be derived as follows:

$$f_{t,\ln}(y) = \left|\frac{\mathrm{d}\ln(y)}{\mathrm{d}y}\right| f_t(\ln(y)) = \frac{1}{y}f_t(\ln(y)), \tag{D.1}$$

where $f_t(\cdot)$ is the PDF of the Student's t-distribution, which gives

$$f_{t,\ln}\left(y|\mu_t,\lambda_t,\nu_t\right) = y^{-1}\frac{\Gamma\left(\frac{\nu_t}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)}\left(\frac{\lambda_t}{\pi\nu_t}\right)^{\frac{1}{2}}\left(1+\frac{\lambda_t\left(\ln(y)-\mu_t\right)^2}{\nu_t}\right)^{-\frac{\nu_t}{2}-\frac{1}{2}}, \tag{D.2}$$

where μ_t, λ_t and ν_t are the parameters of the underlying Student's t-distribution. The CDF of Y is given by the CDF of the underlying Student's t-distribution with argument $\ln(y)$:

$$F_{t,\ln}\left(y|\mu_t,\lambda_t,\nu_t\right) = F_t\left(\ln(y)|\mu_t,\lambda_t,\nu_t\right). \tag{D.3}$$

The log-Student's t-distribution does not have finite moments of any order. A simple proof is given in the following. The expected value of Y is defined as:

$$\mathrm{E}[Y] = \mathrm{E}\left[\exp(X)\right], \tag{D.4}$$

where X follows the Student's t-distribution. The exponential function can be written in terms of the following power series [e.g. [58]]:

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \tag{D.5}$$

which can be substituted into Eq. (D.4) to give:

$$\mathrm{E}[Y] = \mathrm{E}\left[\sum_{k=0}^{\infty} \frac{X^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{\mathrm{E}\left[X^k\right]}{k!}. \tag{D.6}$$

$\mathrm{E}\left[X^k\right]$ is the kth raw moment of the Student's t-distributed random variable X. However, the moments of the Student's t-distribution are only finite for orders $k < \nu_t$ [59] and thus, the following holds for $\mathrm{E}[Y]$ due to the sum in Eq. (D.6):

$$\mathrm{E}[Y] \to \infty \text{ for } \nu_t < \infty. \tag{D.7}$$

Since the first-order moment of Y is infinite, all higher-order integer moments of Y, as well as joint moments for the multivariate case, will also be infinite. In the limiting case, when $\nu_t \to \infty$, the log-Student's t-distribution converges to the lognormal distribution, which has finite moments of any order.

References

[1] K.-K. Phoon, F.H. Kulhawy, Characterization of geotechnical variability, Can. Geotech. J. 36 (4) (1999) 612–624.
[2] S. Geyer, I. Papaioannou, C. Kunz, D. Straub, Reliability assessment of large hydraulic structures with spatially distributed measurements, Struct. Infrastruct. Eng. 16 (2020) 599–612.
[3] E. Vanmarcke, Random fields: analysis and synthesis, revised and expanded new ed., World Scientific, 2010.
[4] M. Grigoriu, Crossings of non-Gaussian translation processes, J. Eng. Mech. 110 (4) (1984) 610–620.
[5] P. Abrahamsen, A review of Gaussian random fields and correlation functions, second ed., Norwegian Computing Center, 1997.
[6] Y. Liu, J. Li, S. Sun, B. Yu, Advances in Gaussian random field generation: a review, Comput. Geosci. (2019) 1–37.
[7] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin, Bayesian data analysis, third ed., in: Chapman & Hall/CRC texts in statistical science, CRC, Boca Raton, FL, 2013.
[8] S. Brooks, A. Gelman, G. Jones, X.-L. Meng, Handbook of Markov chain Monte Carlo, in: Chapman & Hall/crc Handbooks of Modern Statistical Methods, CRC, Boca Raton, FL, 2011.
[9] P. Del Moral, A. Doucet, A. Jasra, Sequential Monte Carlo samplers, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (3) (2006) 411–436.
[10] J. Ching, Y.-C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, J. Eng. Mech. 133 (7) (2007) 816–832.
[11] D. Straub, I. Papaioannou, Bayesian updating with structural reliability methods, J. Eng. Mech. 141 (3) (2015) 04014134.
[12] W. Betz, I. Papaioannou, J.L. Beck, D. Straub, Bayesian inference with subset simulation: strategies and improvements, Comput. Methods Appl. Mech. Engrg. 331 (2018) 72–93.
[13] F. Uribe, I. Papaioannou, W. Betz, D. Straub, Bayesian inference of random fields represented with the Karhunen–Loève expansion, Comput. Methods Appl. Mech. Engrg. 358 (2020) 112632.
[14] I. Papaioannou, D. Straub, Learning soil parameters and updating geotechnical reliability estimates under spatial variability – theory and application to shallow foundations, Georisk Assess. Manag. Risk Eng. Syst. Geohazards 11 (1) (2017) 116–128.
[15] M. Stein, Interpolation of spatial data: some theory for kriging, in: Springer Series in Statistics, Springer, New York City, NY, 1999.
[16] M.S. Handcock, M.L. Stein, A Bayesian analysis of kriging, Technometrics 35 (4) (1993) 403–410.
[17] S. Banerjee, B.P. Carlin, A.E. Gelfand, Hierarchical modeling and analysis for spatial data, second ed., in: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC, Boca Raton, FL, 2014.
[18] C. Rasmussen, C. Williams, Gaussian processes for machine learning, in: Adaptive Computation and Machine Learning, MIT, Cambridge, MA, 2006, p. 248.
[19] O. Ditlevsen, N.J. Tarp-Johansen, H. Denver, Bayesian soil assessments combining prior with posterior censored samples, Comput. Geotech. 26 (3–4) (2000) 187–198.
[20] J. Ching, J.-S. Wang, Application of the transitional Markov chain Monte Carlo algorithm to probabilistic site characterization, Eng. Geol. 203 (2016) 151–167.
[21] Y. Wang, Z. Cao, D. Li, Bayesian perspective on geotechnical variability and site characterization, Eng. Geol. 203 (2016) 117–125.
[22] S.-H. Jiang, I. Papaioannou, D. Straub, Bayesian updating of slope reliability in spatially variable soils with in-situ measurements, Eng. Geol. 239 (2018) 310–320.
[23] J. Ching, K.-K. Phoon, Characterizing uncertain site-specific trend function by sparse Bayesian learning, J. Eng. Mech. 143 (7) (2017) 04017028.
[24] J. Ching, W.-H. Huang, K.-K. Phoon, 3D probabilistic site characterization by sparse Bayesian learning, J. Eng. Mech. 146 (12) (2020) 04020134.
[25] J. Ching, Z. Yang, K.-K. Phoon, Dealing with nonlattice data in three-dimensional probabilistic site characterization, J. Eng. Mech. 147 (5) (2021) 06021003.
[26] Y. Wang, T. Zhao, Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling, Géotechnique 67 (6) (2017) 523–536.
[27] Y. Wang, T. Zhao, K.-K. Phoon, Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation, Can. Geotech. J. 55 (6) (2018) 862–880.
[28] Y. Wang, T. Zhao, Y. Hu, K.-K. Phoon, Simulation of random fields with trend from sparse measurements without detrending, J. Eng. Mech. 145 (2) (2019) 04018130.
[29] S. Montoya-Noguera, T. Zhao, Y. Hu, Y. Wang, K.-K. Phoon, Simulation of non-stationary non-Gaussian random fields from sparse measurements using Bayesian compressive sampling and Karhunen–Loève expansion, Struct. Saf. 79 (2019) 66–79.
[30] Y. Hu, T. Zhao, Y. Wang, C. Choi, C.W. Ng, Direct simulation of two-dimensional isotropic or anisotropic random field from sparse measurement using Bayesian compressive sampling, Stoch. Environ. Res. Risk Assess. 33 (8) (2019) 1477–1496.
[31] T. Zhao, Y. Wang, Simulation of cross-correlated random field samples from sparse measurements using Bayesian compressive sensing, Mech. Syst. Signal Process. 112 (2018) 384–400.
[32] J. Benjamin, C. Cornell, Probability, statistics, and decision for Civil Engineers, in: Dover Books on Engineering, McGraw-Hill, New York City, NY, 1970.
[33] R. Rackwitz, Predictive distribution of strength under control, Matér. Construct. 16 (1983) 259–267.
[34] European Committee for Standardization, EN 1990:2002-04: Eurocode 0: Basis of structural design, European Standard, European Committee for Standardization, 2002.
[35] G. Box, G. Tiao, Bayesian inference in statistical analysis, revised new ed., in: Wiley Classics Library, John Wiley & Sons, 2011.
[36] C.M. Bishop, Pattern recognition and machine learning, in: Information Science and Statistics, Springer, New York City, NY, 2006.
[37] E.H. Vanmarcke, Probabilistic modeling of soil profiles, J. Geotech. Eng. Div. 103 (11) (1977) 1227–1246.
[38] K.-K. Phoon, S.-T. Quek, P. An, Identification of statistically homogeneous soil layers using modified Bartlett statistics, J. Geotech. Geoenviron. Eng. 129 (7) (2003) 649–659.
[39] F.V. Jensen, T.D. Nielsen, Bayesian networks and decision graphs, second ed., in: Information Science and Statistics, Springer, New York City, NY, 2007, p. XVI, 448.
[40] H. Raiffa, R. Schlaifer, Applied statistical decision theory, in: Studies in Managerial Economics, Division of Research, Graduate School of Business Administration, Harvard University, Cambridge, MA, 1961.
[41] B. Matérn, Spatial variation, second ed., in: Lecture Notes in Statistics, Springer, New York City, NY, 1986.
[42] M. DeGroot, Optimal statistical decisions, in: McGraw-Hill Series in Probability and Statistics, McGraw-Hill, New York City, NY, 1969.
[43] S. Kotz, S. Nadarajah, Multivariate t-distributions and their applications, Cambridge University, 2004.
[44] S. Banerjee, Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework, Spatial Stat. 37 (2020) 100417.
[45] R.G. Ghanem, P.D. Spanos, Stochastic finite elements. A spectral approach, revised ed., in: Dover Civil and Mechanical Engineering, Dover, 2012.
[46] K.P. Murphy, Machine learning: a probabilistic perspective, in: Adaptive Computation and Machine Learning, MIT, Cambridge, MA, 2013.
[47] D. Savvas, I. Papaioannou, G. Stefanou, Bayesian identification and model comparison for random property fields derived from material microstructure, Comput. Methods Appl. Mech. Engrg. 365 (2020) 113026.
[48] C.-C. Li, A.D. Kiureghian, Optimal discretization of random fields, J. Eng. Mech. 119 (6) (1993) 1136–1154.
[49] P.-L. Liu, A.D. Kiureghian, Multivariate distribution models with prescribed marginals and covariances, Probab. Eng. Mech. 1 (2) (1986) 105–112.
[50] D.T. Cassidy, M.J. Hamp, R. Ouyed, Pricing European options with a log Student's t-distribution: a Gosset formula, Physica A 389 (24) (2010) 5736–5748.
[51] D.T. Cassidy, M.J. Hamp, R. Ouyed, Log Student's t-distribution-based option sensitivities: Greeks for the Gosset formulae, Quant. Finance 13 (8) (2013) 1289–1302.

[52] L.H. Vanegas, G.A. Paula, Log-symmetric distributions: statistical properties and parameter estimation, Braz. J. Probab. Stat. 30 (2) (2016) 196–220.
[53] P.W. Mayne, Seismic CPT soundings at ANSS stations. Oct-Nov 2002, June-September 2003, 2003, geosystems.ce.gatech.edu/Faculty/Mayne/Research/sound2003/anssall/index.htm. (Accessed 09 Mar 2020).
[54] C.P. Robert, The Bayesian choice: from decision-theoretic foundations to computational implementation, second ed., in: Springer Texts in Statistics, Springer, 2007.
[55] A. Westendarp, H. Becker, J. Bödefeld, H. Fleischer, C. Kunz, M. Maisner, H. Müller, A. Rahimi, T. Reschke, F. Spörel, Erhaltung und Instandsetzung von massiven Verkehrswasserbauwerken [Maintenance and repair of massive hydraulic structures], Beton-Kalender 2015, in: Beton-Kalender 2015, John Wiley & Sons, 2015, pp. 185–246.
[56] H. Zhu, L. Zhang, Characterizing geotechnical anisotropic spatial variations using random field theory, Can. Geotech. J. 50 (7) (2013) 723–734.
[57] X. Nie, J. Zhang, H. Huang, Z. Liu, S. Lacasse, Scale of fluctuation for geotechnical probabilistic analysis, in: Geotechnical Safety and Risk V, IOS, 2015, pp. 834–840.
[58] W. Rudin, Real and complex analysis, third ed., in: Higher Mathematics Series, McGraw-Hill, New York City, NY, 1987.
[59] G. Casella, R. Berger, Statistical inference, second ed., in: Duxbury Advanced Series in Statistics and Decision Sciences, Thomson Learning, Pacific Grove, CA, 2002.