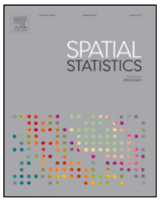
Spatial Statistics 55 (2023) 100750

Contents lists available at ScienceDirect

Spatial Statistics

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



Spatially varying anisotropy for Gaussian random fields in three-dimensional space

Martin Outzen Berild *, Geir-Arne Fuglstad

Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway

ARTICLE INFO

Article history:
Received 12 January 2023
Received in revised form 12 April 2023
Accepted 12 April 2023
Available online 25 April 2023

Keywords:
Spatial non-stationarity
Spatially-varying anisotropy
Stochastic partial differential equations
Gaussian Markov random fields

ABSTRACT

Isotropic covariance structures can be unreasonable for phenomena in three-dimensional spaces. In the ocean, the variability of a response may vary with depth, and ocean currents may lead to spatially varying anisotropy. We construct a class of non-stationary anisotropic Gaussian random fields (GRFs) in three dimensions through stochastic partial differential equations (SPDEs), where computations are done efficiently using Gaussian Markov random field approximations. A key novelty is the parametrization of the spatially varying anisotropy through vector fields.

In a simulation study, we find that simple stationary models obtain reasonable parameter estimates with a moderate number of observations and a single realization, whereas the most complex non-stationary anisotropic model requires dense observations and multiple realizations. Further, we construct a stationary and a non-stationary GRF prior for salinity in an ocean mass outside Trondheim, Norway, based on simulations from the complex numerical ocean model SINMOD. These GRF priors are then evaluated using in-situ measurements collected with an autonomous underwater vehicle. We find that the new model outperforms the stationary anisotropic GRF prior for real-time prediction of unobserved locations both in terms of root mean square error and continuous rank probability score.

* Corresponding author.
E-mail address: martin.o.berild@ntnu.no (M.O. Berild).

https://doi.org/10.1016/j.spasta.2023.100750

1. Introduction

Gaussian random fields (GRFs) are a powerful tool for spatial and spatio-temporal geostatistical modeling (Diggle et al., 1998; Cressie and Wikle, 2015). When the key goal is predictions at unobserved locations, i.e., Kriging, isotropic covariance functions often perform well, and more flexible covariance structures should be used with care (Fuglstad et al., 2015b). In data-rich settings, this is often attributed to the screening effect (Stein, 2002), which causes the closest observations to become the most important for prediction. In other settings, the primary goal is not prediction, but to estimate the covariance structure, for example, to describe internal variability in a climate model ensemble (Castruccio et al., 2019), or to produce a spatial prior based on numerical simulations that will later be used to guide autonomous sampling (Fossum et al., 2021; Foss et al., 2022). For the former, Fuglstad and Castruccio (2020) and Hu et al. (2022) demonstrated that flexible covariance structures can perform better than stationary covariance structures.

There are many approaches to constructing flexible covariance structures (Sampson, 2010; Salvaña and Genton, 2021; Schmidt et al., 2011). Some early approaches are the deformation method (Sampson and Guttorp, 1992) and kernel convolutions (Paciorek and Schervish, 2006), but they both involve the covariances between any pair of locations. This means standard implementations are infeasible for large datasets. There are many ways to overcome such computational issues in spatial statistics and some are applicable for flexible covariance structures (Heaton et al., 2019). The stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011) is interesting because it directly gives rise to computationally efficient models and easily extends to non-stationary covariance models.

A challenge with increasing the degree of flexibility in the covariance structure is that general parametrizations of such flexibility require many parameters. The common family of isotropic Matérn covariance functions (Stein, 2012) are parametrized through 3 parameters: marginal variance, range, and smoothness. Flexible models can have 100s or more parameters (Fuglstad et al., 2015b). An appealing way to reduce dimensionality is to describe the covariance structure through covariates (Schmidt et al., 2011; Neto et al., 2014; Ingebrigtsen et al., 2014, 2015; Risser and Calder, 2015), but this requires knowledge about what type of non-stationarity is expected and suitable covariates to explain such non-stationarity.

There has been much recent development on the theoretical and numerical side for generalized Whittle-Matérn GRFs using the SPDE approach (Lindgren et al., 2022). However, the literature is sparse with respect to parametrization and practical application of models with spatially varying anisotropy in three-dimensional space. For example, stationary models have been used in the context of seismic inversion (Zhang et al., 2016), simple anisotropic covariance structures have been used in the context of fMRI data from the brain (Sidén et al., 2021), and more complex covariance structures in the context of astronomy (Lee and Gammie, 2021), though this was two-dimensional space and time treated as three-dimensional space. The spatially varying anisotropy parametrized by a vector field (Fuglstad et al., 2015a) has not been extended to three-dimensional space.

The aim of this paper is to develop a new approach for parametrization and estimation with spatially varying anisotropy in three-dimensional space. A key advantage of using the SPDE approach is that the SPDE guarantees a valid covariance structure. The main challenge is how to describe and parametrize non-stationary covariance structures. Fuglstad et al. (2015a) used one vector field to describe spatially varying anisotropy, but in three dimensions, two spatially varying orthogonal vector fields are necessary for full generality.

In a simulation study, we investigate how much data is necessary to recover parameters for three different model complexities: stationary isotropic, stationary anisotropic, and non-stationary anisotropic. We then estimate stationary and non-stationary GRF priors that encode knowledge about the ocean from a numerical forecast generated by the numerical model SINMOD by SINTEF. These priors are updated using sparse in-situ measurements by an autonomous underwater vehicle (AUV), and the predictive ability under the priors is compared on a mission conducted in Trondheimsfjorden, Norway, on May 27, 2021. Improved predictions are key, for example, in autonomous sampling of the oceans (Fossum et al., 2019, 2021), but current approaches in autonomous ocean sampling are limited to stationary GRFs.

In Section 2, we describe how to model anisotropy and non-stationarity in three dimensions using SPDEs. Then in Section 3, we describe how to perform inference for the new model in a computationally efficient way. In Section 4, we describe the simulation study and discuss the results, and then continue, in Section 5, with the application to ocean sampling. We end with a discussion in Section 6.

2. Constructing SPDEs with spatially varying anisotropy

2.1. Existing models

The Matérn covariance function on $\mathbb{R}^3$ is given by

$$r(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)^{\nu} K_{\nu}(\kappa\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|), \quad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^3, \tag{1}$$

where $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^3$, $\sigma > 0$ the marginal standard deviation, K_{ν} is the modified Bessel function of the second kind and order $\nu > 0$, and $\kappa > 0$ is an inverse spatial scale parameter. As discussed in Lindgren et al. (2011), GRFs with this covariance function are the stationary solutions of the SPDE

$$(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2}(\tau u(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{R}^3, \tag{2}$$

where $\alpha = \nu + 3/2$, $\tau = \sqrt{8\pi\kappa}/\sigma$, $\nabla \cdot \nabla$ is the Laplacian, and $\mathcal{W}$ is a standard Gaussian white noise process.

Lindgren et al. (2011) proposed to introduce non-stationarity by allowing κ and τ to vary in space (Ingebrigtsen et al., 2014, 2015) or by deformations of space (Hildeman et al., 2021). Fuglstad et al. (2015a,b) consider a version of the SPDE, where the Laplacian is replaced by an anisotropic Laplacian where the direction and degree of anisotropy vary spatially. This was further extended to spherical geometry in Fuglstad and Castruccio (2020), Hu et al. (2022) and Zhang et al. (2023). However, all of these works were in two-dimensional base spaces, and only simpler models have been applied for three-dimensional base spaces (Zhang et al., 2016; Sidén et al., 2021).

The key idea in Fuglstad et al. (2015a) was to replace $\nabla \cdot \nabla$ by $\nabla \cdot \mathbf{H}(\boldsymbol{s}) \nabla$, where $\mathbf{H}(\boldsymbol{s})$ is everywhere a symmetric positive definite 2×2 matrix that controls the strength and direction of anisotropy. The matrix-valued function was specified as $\mathbf{H}(\boldsymbol{s}) = \gamma(\boldsymbol{s})\mathbf{I}_2 + \boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}}$, $\boldsymbol{s} \in \mathbb{R}^2$, where $\gamma(\cdot)$ is a positive function and $\boldsymbol{v}(\cdot)$ is a vector field. This allows $\gamma(\cdot)$ to control the baseline strength of dependence in all directions, and $\boldsymbol{v}(\cdot)$ to control the strength and direction of additional spatial dependence. However, the same parametrization in $\mathbb{R}^3$ is not sufficiently general to control anisotropy fully.

2.2. Stationary anisotropy in $\mathbb{R}^3$

We follow the idea in Fuglstad et al. (2015a) for $\mathbb{R}^2$, and change the SPDE in Eq. (2) to

$$(\kappa^2 - \nabla \cdot \mathbf{H} \nabla) u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^3, \tag{3}$$

where $\nabla \cdot \mathbf{H} \nabla$ is an anisotropic Laplacian and the symmetric positive definite 3×3 matrix $\mathbf{H}$ controls the anisotropy. The parameter τ has been dropped since κ and $\mathbf{H}$ together control both marginal variance and correlation.

As shown in Appendix A.1, the resulting marginal variance is

$$\sigma_m^2 = \frac{1}{8\pi\kappa\sqrt{\det(\mathbf{H})}} \tag{4}$$

and the covariance function is explicitly known as

$$r(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{1}{8\pi\kappa\sqrt{\det(\mathbf{H})}} \exp\left(-\kappa\|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|\right) \tag{5}$$

for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^3$. The latter is derived in Appendix A.2, and is a generalization of results in Lindgren et al. (2011) and Sidén et al. (2021). This corresponds to geometric anisotropy in the Matérn

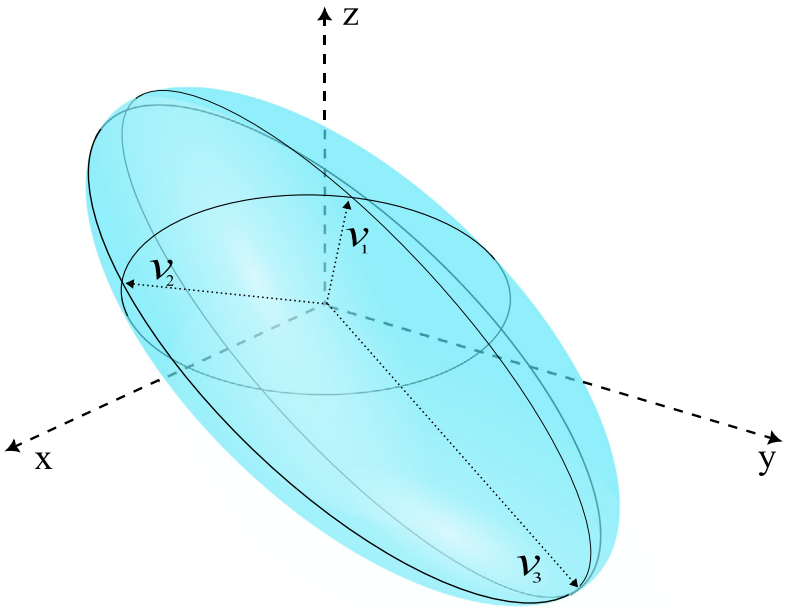


Fig. 1. The three-dimensional iso-correlation surface at the ∼0.37 level of Eq. (5), where $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ are the eigenvectors of some $\mathbf{H}$ with lengths $\sqrt{\lambda_1}/\kappa$, $\sqrt{\lambda_2}/\kappa$ and $\sqrt{\lambda_3}/\kappa$.

covariance function with smoothness $\nu = 1/2$. To understand the behavior of the covariance function, it is useful to think about $\mathbf{H}$ in terms of its eigenvalue decomposition. Let $\tilde{\boldsymbol{v}}_1$, $\tilde{\boldsymbol{v}}_2$, and $\tilde{\boldsymbol{v}}_3$ be orthonormal eigenvectors corresponding to eigenvalues λ_1, λ_2 and λ_3, respectively. Then Fig. 1 shows an example of the 0.37 level iso-correlation surface that will arise from the covariance function in Eq. (5). The semi-axes of the ellipsoid in the figure are $\boldsymbol{v}_1 = (\sqrt{\lambda_1}/\kappa)\tilde{\boldsymbol{v}}_1$, $\boldsymbol{v}_2 = (\sqrt{\lambda_2}/\kappa)\tilde{\boldsymbol{v}}_2$, and $\boldsymbol{v}_3 = (\sqrt{\lambda_3}/\kappa)\tilde{\boldsymbol{v}}_3$, which by evaluating the covariance function with either of these semi-axes will yield the relationship and the iso-correlation level $r(\boldsymbol{v})/\sigma_m^2 = e^{-1} \approx 0.37$.

We generalize the parametrization described in Section 2.2 and $\mathbf{H}$ is decomposed as

$$\mathbf{H} = \gamma \mathbf{I}_3 + \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}. \tag{6}$$

where $\boldsymbol{v} = (v_x, v_y, v_z)^{\mathrm{T}} \in \mathbb{R}^3$ and $\boldsymbol{w} = (\omega_x, \omega_y, \omega_z)^{\mathrm{T}} \in \mathbb{R}^3$, $\boldsymbol{v} \perp \boldsymbol{\omega}$, and $\gamma > 0$. The eigenvalue decomposition of $\mathbf{H}$ has eigenvalues $\lambda_1 = \gamma$, $\lambda_2 = \gamma + \|\boldsymbol{v}\|^2$ and $\lambda_3 = \gamma + \|\boldsymbol{w}\|^2$ with the corresponding eigenvectors $\boldsymbol{v}_1 = \boldsymbol{v} \times \boldsymbol{\omega}$, $\boldsymbol{v}_2 = \boldsymbol{v}$ and $\boldsymbol{v}_3 = \boldsymbol{\omega}$, respectively. We construct $\boldsymbol{\omega}$ by a linear combination of two orthogonal vectors in the plane with $\boldsymbol{v}$ as a normal vector. First, let $\boldsymbol{\omega}_1 = (-v_y, v_x, 0)^{\mathrm{T}}$, which satisfies $\boldsymbol{v} \perp \boldsymbol{\omega}_1$. Second, let $\boldsymbol{\omega}_2 = \boldsymbol{v} \times \boldsymbol{\omega}_1 = (-v_z v_x, -v_z v_y, v_x^2 + v_y^2)^{\mathrm{T}}$, which also satisfies $\boldsymbol{v} \perp \boldsymbol{\omega}_2$. We parametrize $\boldsymbol{\omega}$ through

$$\boldsymbol{\omega} = \rho_1 \frac{\boldsymbol{\omega}_1}{\|\boldsymbol{\omega}_1\|} + \rho_2 \frac{\boldsymbol{\omega}_2}{\|\boldsymbol{\omega}_2\|}, \tag{7}$$

where $\rho_1, \rho_2 \in \mathbb{R}$ which works whenever $v_x = v_y \neq 0$. An alternative solution is to use Euler–Rodrigues parametrization (Euler, 1771; Rodrigues, 1840) to obtain both $\boldsymbol{v}$ and $\boldsymbol{\omega}$; however, in this case, the parameters are less interpretable.

The above parametrization for $\mathbf{H}$ uses six parameters, γ, v_x, v_y, v_z, ρ_1, and ρ_2, to describe all forms of geometric anisotropy. The parametrization is interpretable: (1) γ controls the isotropic effect, (2) v_x, v_y, and v_z controls extra anisotropy in one direction, and (3) ρ_1 and ρ_2 controls extra anisotropy in a second direction orthogonal to the first. Lastly, κ simultaneously controls scaling of spatial dependence equally in all directions, and the variance of the GRF together with the six other parameters as seen in Eq. (4).

2.3. Spatially varying anisotropy on bounded domain $\mathcal{D} \subset \mathbb{R}^3$

Non-stationarity and spatially varying anisotropy is achieved by making the coefficients in Eq. (3) spatially varying,

$$(\kappa(\boldsymbol{s})^2 - \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^3, \tag{8}$$

where $\kappa(\cdot)$ is a positive function, and $\mathbf{H}(\cdot)$ is a differentiable spatially varying symmetric positive definite 3×3 matrix. There are further requirements to make this well-defined such as ensuring that $\kappa(\cdot)$ is bounded from below by a non-negative number and that $\mathbf{H}(\cdot)$ is uniformly positive definite; see Bolin and Kirchner (2020) for more details on the technical requirements on $\kappa(\cdot)$ and $\mathbf{H}(\cdot)$. Heuristically, one can imagine that the SPDE is gluing together different local behavior described by ellipsoids, as discussed in Section 2.2, to a valid non-stationary covariance structure.

In practice, we need to limit Eq. (8) to a bounded domain to parametrize the non-stationarity. The SPDE we propose is

$$(\kappa(\boldsymbol{s})^2 - \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^3, \tag{9}$$

where $\mathcal{D}$ is bounded, and we enforce the boundary condition

$$(\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^{\mathrm{T}}\boldsymbol{n}(\boldsymbol{s}) = 0, \quad \boldsymbol{s} \in \partial\mathcal{D},$$

where $\boldsymbol{n}(\boldsymbol{s})$ is the outward normal vector of $\mathcal{D}$. This corresponds to no flux through the boundary. The effect of the boundary condition is increased marginal variance on the boundary and increased spatial dependency due to the "reflective" boundary condition. As discussed in Lindgren et al. (2011) and Fuglstad et al. (2015b), one can extend the domain $\mathcal{D}$ outside the area with observations to reduce boundary effects, or one can consider the boundary effects a feature that the non-stationary model can adjust for if necessary.

3. Estimating SPDEs with spatially varying anisotropy

3.1. Parameterizing the non-stationarity

Before using the SPDE in Eq. (9) in inference, we need to parametrize the non-stationarity through a finite number of parameters. This involves expanding $\log(\kappa(\cdot))$, $\log(\gamma(\cdot))$, $v_x(\cdot)$, $v_y(\cdot)$, $v_z(\cdot)$, $\rho_1(\cdot)$, and $\rho_2(\cdot)$ in basis functions. The log-transform is used for $\kappa(\cdot)$ and $\gamma(\cdot)$ since they must be positive functions.

Let $g : \mathbb{R}^3 \to \mathbb{R}$ denote a generic function that we want to expand in a basis, and let $p > 0$ the number of basis functions. We use basis splines similar to Fuglstad et al. (2015b), and set

$$g(\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\alpha}_g, \tag{10}$$

where $\boldsymbol{\alpha}_g \in \mathbb{R}^p$, and $\boldsymbol{f}(\boldsymbol{s}) = (f_1(\boldsymbol{s}), \ldots, f_p(\boldsymbol{s}))^{\mathrm{T}}$ is a p-dimensional vector with the basis functions evaluated at location $\boldsymbol{s}$.

In this paper, we use rectangular domains such as $\mathcal{D} = [A_1, B_1] \times [A_2, B_2] \times [A_3, B_3]$, and a basis constructed as a tensor product of three one-dimensional second-order B-splines. We use clamped splines as discussed in Appendix A.3, and Fig. 2 shows an example of the resulting basis functions in 1-dimension. The use of a rectangular grid is typically not a serious restriction, but tetrahedral grids can be natural for special geometries such as the globe which has a spherical surface (Zhang et al., 2016). For the globe, one could consider a basis based on spherical harmonic functions for the coefficients in the SPDE.

Let $B_{x,i}$ denote the ith basis function of the second-order basis in the x-dimension, and similarly $B_{y,j}$ and $B_{z,k}$ for the y- and z-dimension. The resulting tree-dimensional basis is then

$$f_{ijk}(\boldsymbol{s}) = B_{x,i}(s_1) \cdot B_{y,j}(s_2) \cdot B_{z,k}(s_3), \quad \boldsymbol{s} = (s_1, s_2, s_3)^{\mathrm{T}} \in \mathcal{D}, \tag{11}$$

for all combinations $i, j, k \in \{1, \ldots, m\}$ (see Fig. 3 for visualization in 3D). This means that $\boldsymbol{\alpha}_g \in \mathbb{R}^{m^3}$, and m^3 parameters must be estimated for each of the seven functions described at the start of the section.

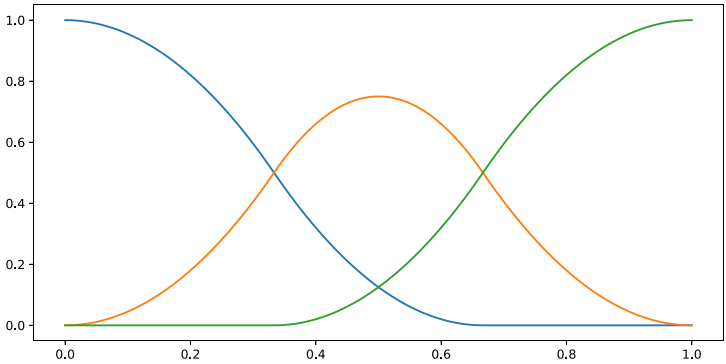


Fig. 2. Clamped B-spline basis with three basis functions in 1D.

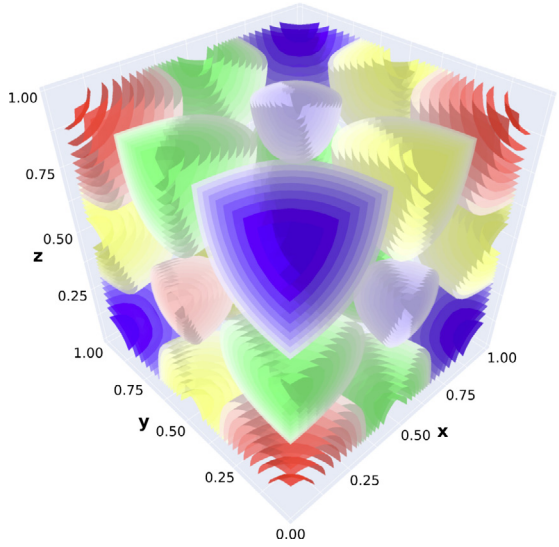


Fig. 3. Parametrized function representation with B-splines in 3D.

In Sections 4 and 5, we use 3 basis functions for each spatial dimension such that $p = m^3 = 3^3 = 27$, resulting in a total of 189 parameters controlling the functions of $\log(\kappa(\cdot))$, $\log(\gamma(\cdot))$, $v_x(\cdot)$, $v_y(\cdot)$, $v_z(\cdot)$, $\rho_1(\cdot)$, and $\rho_2(\cdot)$. When data is sparse, such a model can easily result in overfitting (Fuglstad et al., 2015b), and it is necessary to introduce penalties on the seven functions. In Fuglstad et al. (2015b), this was achieved by a hierarchical model where

$$\tau_g \Delta g(\boldsymbol{s}) = \mathcal{W}_g(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D},$$

together with Neumann boundary conditions of zero derivatives on the boundary of the domain. However, this requires selecting a reasonable value for $\tau_g > 0$ for each of the seven functions and is computationally expensive if it is done using cross-validation. In the context of this paper, we are constructing a stochastic model that mimics the behavior of a densely "observed" numerical simulation model and we do not include penalties beyond the restriction of using 27 basis functions. In the simulation study in Section 4, we demonstrate the ability of this model to be estimated and investigate the amount of data needed to estimate the model.

3.2. Hierarchical model and discretization

Consider a bounded domain $\mathcal{D} \subset \mathbb{R}^3$, and observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ made at locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in \mathcal{D}$. We assume a Gaussian observation model

$$y_i|\eta(\boldsymbol{s}_i), \sigma_{\mathrm{N}}^2 \sim \mathcal{N}(\eta(\boldsymbol{s}_i), \sigma_{\mathrm{N}}^2), \quad i = 1, \ldots, n,$$

where $\sigma_{\mathrm{N}}^2 > 0$ is the nugget variance and

$$\eta(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + u(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D},$$

describes true spatial variation as a combination of covariates and a GRF. Here $\boldsymbol{x}(\cdot)$ is a spatially varying vector of k covariates, $\boldsymbol{\beta} \in \mathbb{R}^k$ are the coefficients of the covariates, and $u(\cdot)$ is a GRF with spatially varying anisotropy as presented in Section 2.

As described in Appendix B, the GRF $u(\cdot)$ is discretized using a regular grid with l cells, and we get a Gaussian Markov random field $\boldsymbol{w} = (w_1, \ldots, w_l)^{\mathrm{T}}$. Let $\boldsymbol{\theta}$ be the vector of all parameters controlling $u(\cdot)$, then

$$\boldsymbol{w}|\boldsymbol{\theta} \sim \mathcal{N}_l(\mathbf{0}, \mathbf{Q}^{-1}),$$

where dependence on $\boldsymbol{\theta}$ is suppressed for $\mathbf{Q}$, and $\mathbf{Q}$ is a $l \times l$ precision matrix with a three-dimensional spatial sparsity structure. The vector $\boldsymbol{w}$ is linked to $u(\cdot)$ through a linear transformation $u(\boldsymbol{s}) = \boldsymbol{a}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{w}$, where $\boldsymbol{a}$ has only one non-zero entry corresponding to which grid cell location $\boldsymbol{s}$ belongs. This gives $\boldsymbol{u} = (u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n))^{\mathrm{T}} = \mathbf{A}\boldsymbol{w}$, where the $n \times l$ matrix $\mathbf{A}$ only has one non-zero entry on each row.

The coefficients of the fixed effect, $\boldsymbol{\beta}$, is assigned the weak penalty $\boldsymbol{\beta} \sim \mathcal{N}_K(\mathbf{0}, V\mathbf{I}_K)$ for a fixed $V > 0$. Thus we can write $\boldsymbol{y}$ as

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{w} + \boldsymbol{\epsilon}, \tag{12}$$

where $\mathbf{X}$ is the design matrix of covariates, and $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n\sigma_{\mathrm{N}}^2)$ is an n-dimensional vector of random noise. This gives rise to the hierarchical formulation

$$\begin{aligned}
&\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{w}, \sigma_{\mathrm{N}}^2 \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{w}, \sigma_{\mathrm{N}}^2\mathbf{I}_n),\\
&\boldsymbol{\beta} \sim \mathcal{N}_k(\mathbf{0}, V\mathbf{I}_k), \quad \boldsymbol{w}|\boldsymbol{\theta} \sim \mathcal{N}_l(\mathbf{0}, \mathbf{Q}^{-1}).
\end{aligned}$$

Let $\boldsymbol{s}^* \in \mathcal{D}$ be an unobserved location. After parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma_{\mathrm{N}}^2}$ are estimated, one can predict the underlying value $\eta(\boldsymbol{s}^*) = \boldsymbol{x}(\boldsymbol{s}^*)^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{a}(\boldsymbol{s}^*)^{\mathrm{T}}\boldsymbol{w}$ or a new observation $y^* = \boldsymbol{x}(\boldsymbol{s}^*)^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{a}(\boldsymbol{s}^*)^{\mathrm{T}}\boldsymbol{w} + \epsilon^*$, where $\epsilon^* \sim \mathcal{N}(0, \hat{\sigma_{\mathrm{N}}^2})$ is a new nugget. The predictions are made using the conditional distributions $\eta(\boldsymbol{s}^*)|\boldsymbol{y}, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \sigma_{\mathrm{N}}^2 = \hat{\sigma_{\mathrm{N}}^2}$ and $y^*|\boldsymbol{y}, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \sigma_{\mathrm{N}}^2 = \hat{\sigma_{\mathrm{N}}^2}$. The estimation of parameters is detailed in the next section.

3.3. Parameter inference

Simplify notation by letting $\boldsymbol{z} = (\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$. Then

$$\boldsymbol{z}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_z^{-1}), \text{ where } \mathbf{Q}_z = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & V\mathbf{I}_k \end{bmatrix}.$$

Let $\mathbf{S} = \begin{bmatrix} \mathbf{A} & \mathbf{X} \end{bmatrix}$, then the observation model can be rewritten as

$$\boldsymbol{y}|\boldsymbol{z}, \sigma_{\mathrm{N}}^2 \sim \mathcal{N}_n(\mathbf{S}\boldsymbol{z}, \mathbf{I}_n\sigma_{\mathrm{N}}^2). \tag{13}$$

Using this notation the log-likelihood can be expressed as

$$\begin{aligned}
\log \pi(\boldsymbol{\theta}, \sigma_{\mathrm{N}}^2|\boldsymbol{y}) = \text{Const} + \log \pi(\boldsymbol{\theta}, \sigma_{\mathrm{N}}^2) + \frac{1}{2}\log\det(\mathbf{Q}_z) - \frac{n}{2}\log(\sigma_{\mathrm{N}}^2)\\
- \frac{1}{2}\log\det(\mathbf{Q}_C) - \frac{1}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\mathbf{Q}_C\boldsymbol{\mu}_C - \frac{1}{2\sigma_{\mathrm{N}}^2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)^{\mathrm{T}}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C).
\end{aligned} \tag{14}$$

Here dependence on $\boldsymbol{\theta}$ is suppressed for $\boldsymbol{\mu}_C$, $\mathbf{Q}_z$ and $\mathbf{Q}_C$, and $\pi(\boldsymbol{\theta}, \sigma_N^2)$ can be used to assign a penalty on $\boldsymbol{\theta}$, e.g., like the random-walk penalty used in Fuglstad et al. (2015b). The conditional precision matrix $\mathbf{Q}_C$ is

$$\mathbf{Q}_C = \mathbf{Q}_z + \mathbf{S}^T\mathbf{S}/\sigma_N^2 \tag{15}$$

and $\boldsymbol{\mu}_C$ is the conditional mean,

$$\boldsymbol{\mu}_C = \mathbf{Q}_C^{-1}\mathbf{S}^T\boldsymbol{y}/\sigma_N^2. \tag{16}$$

Parameter inference is done by maximizing Eq. (14) with respect to $\boldsymbol{\theta}$ and σ_N^2. The parameter vector $\boldsymbol{\theta}$ includes all coefficients for the basis functions, and when using 27 basis functions for each function,

$$\boldsymbol{\theta} = \left(\boldsymbol{\alpha}_{\log(\kappa^2)}, \boldsymbol{\alpha}_{\log\gamma}, \boldsymbol{\alpha}_{v_x}, \boldsymbol{\alpha}_{v_y}, \boldsymbol{\alpha}_{v_z}, \boldsymbol{\alpha}_{\rho_1}, \boldsymbol{\alpha}_{\rho_2}\right),$$

has 189 parameters. The parameter space is challenging to search and we use an analytical expression for the gradient in the optimization algorithm. The derivation of the analytical gradient involves many nested chain rules and a technique to calculate a partial inverse of sparse matrices (Rue and Held, 2010), see Appendix A.5 for a complete description.

4. Simulation study

In this section, we perform a simulation study to investigate the amount of data required to acquire reasonable parameter estimates for the SPDE models. We consider three different levels of complexity for the spatial dependence structure and use the observation model

$$\boldsymbol{y}_{\text{Mod}} = \mathbf{A}\boldsymbol{w}_{\text{Mod}} + \boldsymbol{\epsilon}, \tag{17}$$

where $\boldsymbol{w}_{\text{Mod}}$ is the GMRF controlled by the parameters $\boldsymbol{\theta}_{\text{Mod}}$ for a given model Mod, and $\boldsymbol{\epsilon}$ is i.i.d. Gaussian noise with mean zero and standard deviation $\boldsymbol{\sigma}_N = 0.1$. All models use the same domain, $\boldsymbol{s} \in \mathcal{D} = [0, 40] \times [0, 40] \times [0, 40]$, with a grid of size $(M, N, P) = (30, 30, 30)$, which results in a total of 27000 grid cells.

The simplest model is a Stationary Isotropic (SI) model which has a covariance structure controlled by the three parameters $\boldsymbol{\theta}_{\text{SI}} = (\log\kappa^2, \log\gamma, \log\sigma_N^2)$. We assign the values $\kappa^2 = 0.2$, $\gamma = 2.5$, and $\sigma_N = 0.1$, which corresponds to a spatial range 10.6, around 25% the extent in each dimension, and a marginal variance 0.023.

The second model is a Stationary Anisotropic (SA) model described by 8 parameters $\boldsymbol{\theta}_{\text{SA}} = (\log\kappa^2, \log\gamma, v_x, v_y, v_z, \rho_1, \rho_2, \log\sigma_N^2)$. We assign the values $\kappa^2 = 0.35$, $\gamma = 0.5$, $v_x = 1.9$, $v_y = 1.4$, $v_z = 0.4$, $\rho_1 = 1.4$, $\rho_2 = 0.6$ and $\sigma_N = 0.1$. These values correspond to spatial ranges 10.1 along the x-dimension, 6.8 along y, and 3.9 along z, and marginal variance 0.023. This gives around the same range as SI in x-direction, but smaller ranges in y- and z-directions.

The third model is the Non-stationary Anisotropic (NA) model described in Section 3, which has three B-spline basis functions in each spatial dimension for each coefficient in the SPDE. NA has 190 parameters controlling the non-stationary anisotropic covariance structure. We select the parameters

$$\boldsymbol{\theta}_{\text{NA}} = \left(\boldsymbol{\alpha}_{\log(\kappa^2)}, \boldsymbol{\alpha}_{\log\gamma}, \boldsymbol{\alpha}_{v_x}, \boldsymbol{\alpha}_{v_y}, \boldsymbol{\alpha}_{v_z}, \boldsymbol{\alpha}_{\rho_1}, \boldsymbol{\alpha}_{\rho_2}, \log\sigma_N\right)$$

so that the covariance structure and variance resemble a vortex as illustrated in Fig. 4.

The aim of the simulation study is to investigate how much data is needed to estimate the chosen covariance structures in SI, SA, and NA. It is known that for SI observed on a bounded domain, the parameters are not consistently estimable (Zhang, 2004). In the isotropic SPDE in Eq. (2), κ cannot be consistently estimated, while τ can be consistently estimated (Fuglstad et al., 2019). Bolin and Kirchner (2023) provide more general results also applicable to the SPDE with spatially varying coefficients in Eq. (9), which suggests that $\kappa(\cdot)$ is also not consistently estimable in this case. We will not pursue exact theoretical results in this paper, but this suggests that even with a densely observed realization of NA on a bounded domain, there is a limit to how well the parameters

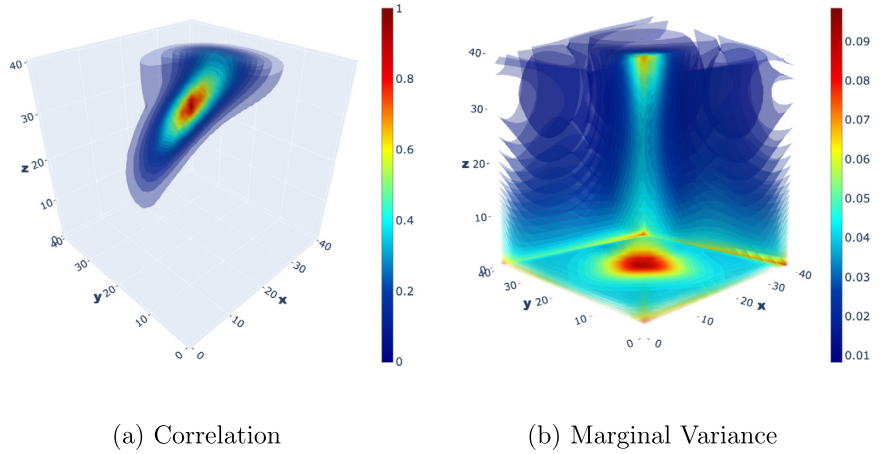


Fig. 4. Spatial correlation at location [20,20,20] (a) and variance of the spatial effect (b) in the non-stationary anisotropic model.

can be estimated. On the other hand, under both in-fill asymptotics and an increasing number of realizations of the GRF, we expect the parameters to be consistently estimable.

These issues motivate an exploration of two aspects of the data: (1) the density of observations and (2) the number of realizations (replicated observation of these spatial locations). For (1), we use low density with 100 locations, high density with 10000 locations, and full observation of the grid with 27000 locations. This corresponds to 1%, 37%, and 100% of the total spatial locations, respectively. For (2), we consider 1 realization, 10 realizations, and 100 realizations. This gives a total of nine different settings for the amount of data. We use 100 repetitions for each setting, where the observed locations are randomly drawn in each trial.

First, for each data setting and repetition, we simulate from SI, SA, and NA and generate data according to Eq. (17). Then for each dataset, we estimate the parameters of the corresponding model using maximum likelihood. Finally, the estimated parameters are compared with their true value. For SI and SA, we consider root mean square error (RMSE) for each parameter separately since there are few parameters to show. For NA, we summarize the error in the spatially varying coefficients by the average RMSE across space since the individual parameters are less interesting than whether we capture the correct spatially varying functions.

Table 1 shows RMSE between the parameter values in each model and their values inferred in the nine different data settings. The estimates are obtained using the inference method described in Section 3.3 with the observation model in Eq. (17). For a further analysis of the results, Table S2 in Appendix C presents the relative RMSE which also considers the scale of the parameters. Note that some parts of the table are omitted to simplify the presentation of the results for the reader as the removed values are not of interest to the conclusion of the study. From Table 1 we observe that SI and SA perform well with dense data or multiple realizations, but for sparse data, SI requires 10 realizations while SA requires 100 realizations. For NA, much more data is required and we need to observe the whole grid with 10 or more realizations. There is a large discrepancy between 10000 observed points (37%) and 27000 (100%), so it would be interesting to investigate where in this range reasonable estimates are obtained. However, we leave this for future work. We also note that these results are specific to the chosen covariance structures, and further investigation would be interesting to evaluate behavior over a large range of scenarios.

Table 1
The Root Mean Square Error (RMSE) of parameter estimates for SI, SA, and NA from 100 repetitions for each data setting, which varies the number of observed locations (No. loc.) and the number of replicated observations of these locations (No. real.).

	No. loc.	100			10000			27000		
	No. real.	1	10	100	1	10	100	1	10	100
Stat. Iso.	$\log\kappa$	0.761	**0.168**	**0.055**	**0.147**					
	$\log\gamma$	0.796	**0.160**	**0.065**	**0.044**					
	$\log\tau$	2.515	**0.649**	**0.187**	**0.067**					
Stationary Anisotropic	$\log\kappa$	0.993	0.185	**0.093**	**0.097**	**0.038**				
	$\log\gamma$	7.583	43.234	**0.454**	**0.218**	**0.080**				
	$\lvert v_x\rvert$	1.317	0.972	**0.469**	**0.189**	**0.071**				
	$\lvert v_y\rvert$	1.252	0.730	**0.461**	**0.144**	**0.036**				
	$\lvert v_z\rvert$	1.183	0.571	**0.256**	**0.077**	**0.028**				
	$\lvert\rho_1\rvert$	1.110	0.862	**0.263**	**0.119**	**0.038**				
	$\lvert\rho_2\rvert$	1.413	0.610	**0.297**	**0.074**	**0.028**				
	$\log\tau$	1.816	1.253	**0.194**	**0.169**	**0.031**				
Non-Stationary Anisotropic	$\log\kappa$						2.620	0.864	**0.329**	**0.245**
	$\log\gamma$						2.664	1.248	**0.566**	**0.497**
	$\lvert v_x\rvert$						3.253	0.675	**0.459**	**0.447**
	$\lvert v_y\rvert$						3.962	0.615	**0.363**	**0.356**
	$\lvert v_z\rvert$						1.655	0.594	**0.290**	**0.250**
	$\lvert\rho_1\rvert$						0.221	0.797	**0.336**	**0.274**
	$\lvert\rho_2\rvert$						0.642	1.130	**0.716**	**0.622**
	$\log\tau$						1.146	0.016	**0.006**	**0.005**

5. GRF prior for statistical sampling of the ocean

5.1. Motivation

We aim to determine the three-dimensional extent of a plume in the ocean by combining forecasts from a numerical ocean model with in-situ measurements from an autonomous underwater vehicle (AUV). In this particular application, we consider a freshwater plume in the ocean. Forecasts produced by numerical ocean models provide physically realistic and detailed behavior for the ocean. They can accurately capture many large-scale features, but local behavior such as plumes created by freshwater discharge will not be accurate. On the other, an AUV can sample accurately locally, but the observations will be sparse for the region of interest.

We aim to construct a prior based on the numerical ocean model that informs prior beliefs about the ocean. Then we will update this prior with in-situ measurements by the AUV, to provide a prediction of the current state of the ocean. In this paper, we assume a spatial prior and do not include a dynamical model. This means that we assume that operation time is short enough that the ocean mass is reasonably stable. It is common practice in ocean sampling with AUVs not to use dynamical models (Fossum et al., 2019, 2021), and the simulation from the numerical ocean models suggests that behavior is stable around 2–3 h, but we will investigate the effect this has on predictions when performing cross-validation.

There are two steps in our approach. Step 1 is to estimate a stationary GRF prior and a non-stationary GRF prior based on a simulation from the numerical ocean model as described in Section 5.2. Step 2 is to combine each of the estimated priors with an observation model, and evaluate the predictive ability on in-situ observations from AUV as described in Section 5.3. The GRFs that we estimate based on the numerical ocean model can be viewed as statistical emulators of the ocean.

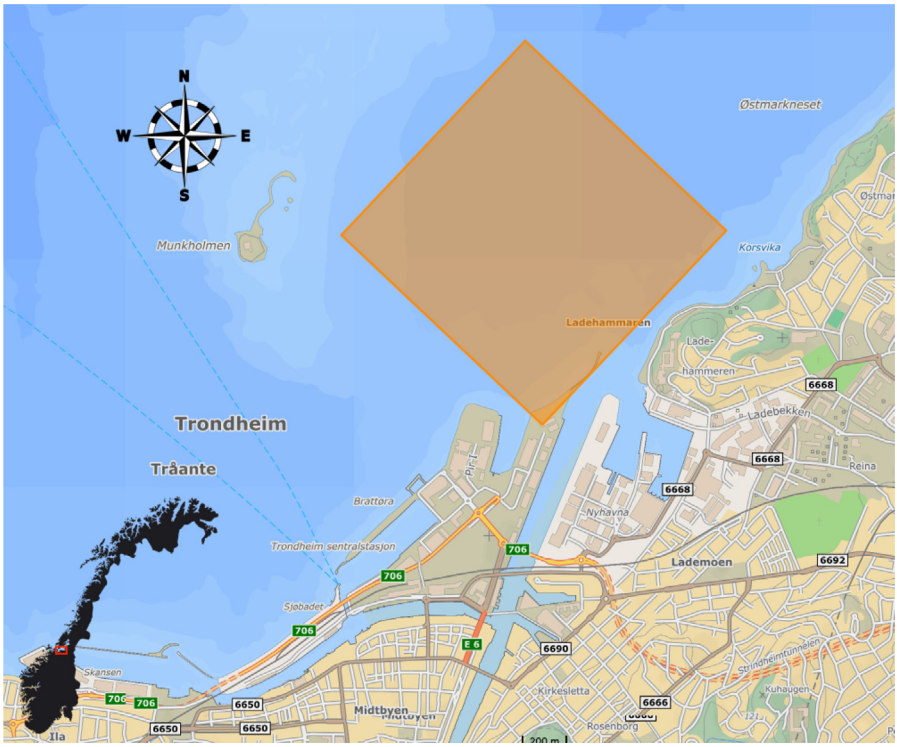


Fig. 5. The area of operation in Trondheimsfjorden at Ladehammaren just outside of Trondheim, Norway. The compass shows the cardinal directions relative to the map.

5.2. The numerical ocean model and the GRF prior

The model training data used in this application is from a forecast produced by the ocean model SINMOD. Data is provided by SINTEF Ocean which developed and ran the simulation. SINMOD is a three-dimensional numerical ocean model based on primitive equations that are solved using finite difference methods on a regular grid with horizontal cell sizes of 20 km×20 km and is nested in several steps down to 32 m × 32 m. Moreover, it uses z* vertical layers which allow for varying grid resolutions depending on the depth and help capture the higher variability of the surface. SINMOD is driven by atmospheric forces, freshwater outflows, and tides, and it provides numerical simulations of multiple variables such as salinity, temperature, and currents. The reader is referred to Slagstad and McClimans (2005) for a more detailed description of the method.

The area of operation is located in Trondheimsfjorden at Ladehammaren just outside of Trondheim, Norway, and the operation time was May 27, 2021, between 10:30 and 14:30. The outlined area in Fig. 5 indicates the operational area which covers 1408 m × 1408 m in the horizontal plane. At the southeast side of this field, the Nidelva river flows into the fjord. This causes a very complex salinity field that is unfeasible to describe with a stationary covariance model. Therefore, we will use the numerical simulations from SINMOD to estimate a non-stationary GRF. As demonstrated in the simulation study, complex covariance structures can reliably be estimated based on such dense data.

In this application, we will focus on univariate modeling of the salinity and we choose the fine-scale horizontal grid sizes $h_x = 32\,\text{m}$ $h_y = 32\,\text{m}$, which in total gives $N = 45$ and $M = 45$ grid nodes for both the numerical and the statistical model. Moreover, in the vertical plane, we use 1-meter increments between the depth layers, i.e., $h_z = 1\,\text{m}$. To avoid any major effects of the

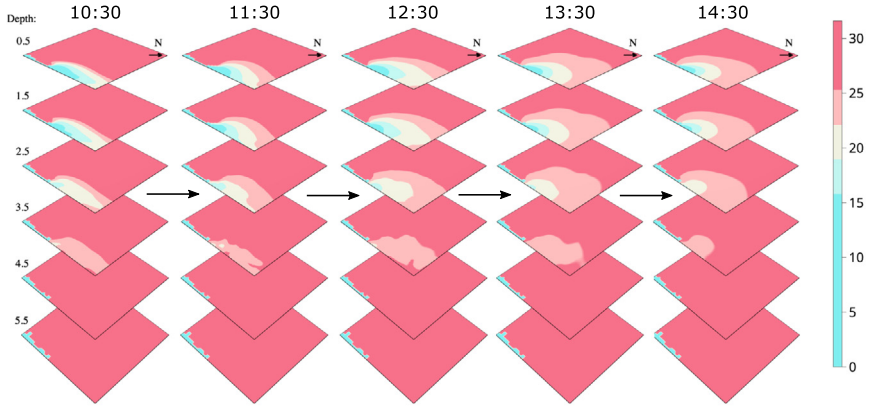


Fig. 6. Five timesteps of the dataset simulated with the numerical ocean model SINMOD for May 27, 2021. The timestamps are displayed over their respective timesteps. The N-arrow is the cardinal north.

boundaries in this direction $P = 11$ depth layers are used resulting in a depth range of 0.5 m to 10.5 m. SINMOD outputs $\boldsymbol{z}_t$, $t = 0, 1, 2, \ldots, 143$, which are vectors of salinity values in all cells in the three-dimensional grid at different time points throughout the whole May 27, 2021. The timesteps are 10 minutes, and Fig. 6 shows five timesteps from SINMOD for the top six depth layers during the operation. Note that the varying vertical layers in the numerical model are either with 0.5 m or 1 m increments, so the SINMOD simulations do not require any additional modification to fit within our statistical model.

We first estimate the model

$$\boldsymbol{z}_t = \Phi \boldsymbol{z}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, 143,$$

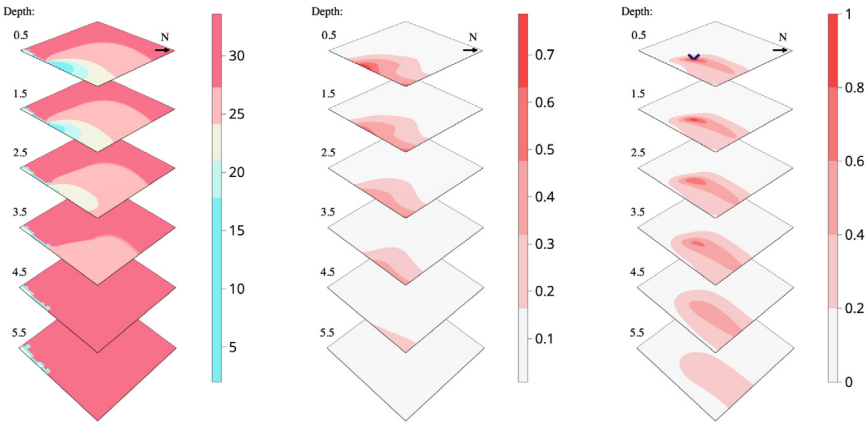
where Φ is a diagonal matrix of AR(1) coefficients. The diagonal entries of Φ are estimated with maximum likelihood separately for each spatial location such that $\hat{\Phi}_{ii} = \sum_{t=1}^{143} z_{t,i} z_{t-1,i} / \sum_{t=1}^{143} z_{t-1,i}^2$ for $i = 1, \ldots, NMP$, where $z_{t,i}$ is the value in cell i at time t. We then compute empirical innovations $\hat{\boldsymbol{\epsilon}}_t = \boldsymbol{z}_t - \hat{\Phi} \boldsymbol{z}_{t-1}$, $t = 1, \ldots, 143$. These empirical innovations describe the spatial covariance structure for short-term changes in salinity.

We fit the flexible non-stationary anisotropic model with 190 parameters, $\hat{\boldsymbol{\theta}}_{\text{NA}} = (\boldsymbol{\alpha}_{\log \kappa}, \boldsymbol{\alpha}_{\log \gamma}, \boldsymbol{\alpha}_{v_x}, \boldsymbol{\alpha}_{v_y}, \boldsymbol{\alpha}_{v_z}, \boldsymbol{\alpha}_{\rho_1}, \boldsymbol{\alpha}_{\rho_2}, \log \sigma_N^2)$, and the stationary anisotropic model with 8 parameters, $\hat{\boldsymbol{\theta}}_{\text{SA}} = (\log \kappa^2, \log \gamma, v_x, v_y, v_z, \rho_1, \rho_2, \log \sigma_N^2)$, to the assumed independent realization from a GRF $\hat{\boldsymbol{\epsilon}}_1, \ldots . \hat{\boldsymbol{\epsilon}}_{143}$. Note that there are $NMP = 22275$ spatial locations and the 144 empirical innovations cover the whole day of May 27, 2021. Figs. 7(b) show the resulting variance of the spatial effect and Fig. 7(c) the spatial correlation with location $(x, y, z) = (22, 10, 0)$ of the non-stationary anisotropic model. The same figures of the stationary anisotropic model can be found in Appendix C, Fig. S3.

In the next step, we construct the expected value of the GRF using the time average of the whole day, $\boldsymbol{\mu} = \sum_{t=0}^{143} \boldsymbol{z}_t / 144$. The mean is shown in Fig. 7(a) and shows the overall tendency for freshwater near the river outlet and saltwater further out in the ocean. We choose the prior

$$\boldsymbol{\eta} = \boldsymbol{\mu} + \boldsymbol{e}, \tag{18}$$

where we combine the fixed mean vector, $\boldsymbol{\mu}$, with a new realization, $\boldsymbol{e}$, of the estimated stationary anisotropic model or the non-stationary anisotropic model. This is a spatial prior on a 32 m $\times$ 32 m $\times$ 1 m resolution.



(a) SINMOD prior (b) Marginal Variance (c) Correlation

Fig. 7. Prior field **(a)** found from SINMOD simulations, the variance of the spatial effect **(b)** and spatial correlation of point [22,10,0] (marked) **(c)** in the non-stationary anisotropic model. The N-arrow is the cardinal north.

5.3. In-situ data collection and emulator evaluation

In-situ measurements were made with the AUV on May 27, 2021, between 10:30 and 14:30. The AUV followed 9 pre-planned paths within the area of operation: two intersects at 0.5 m depth one northbound and one north-westbound starting from the river, two zig-zags in each depth layer (0.5 m,2 m,5 m), and one up-and-down pattern in depth ranging from 0.5 m to 10.5 m moving north-westbound starting from the river. The design was chosen to explore three depth levels in detail with zig-zag patterns combined with an up-and-down pattern to collect measurements also at other depths. This is not constructed based on any optimality criterion, and we chose a fixed design so that model comparison is not biased by decisions made by the AUV during sampling. Fig. 8 displays the locations of the measurements in the top 5 layers of the field.

The AUV is moving at 1.5 m/s and continuously samples the salinity. This means that multiple measurements are made within each 32 m $\times$ 32 m $\times$ 1 m grid cell. Observations are represented as y_i, $i = 1, \ldots, n_{\text{obs}}$, whereby y_i is the average value of the measurements in grid cell i. Therefore, Fig. 8 shows that even though the AUVs zig-zag path was fixed at 2 m and 5 m depth, the majority of these measurements were allocated to the cell with centers at 1.5 m and 4.5 m depth rather than at 2.5 m or 5.5 m. The observations are then combined with the prior in Eq. (18) using

$$\begin{aligned} y_i|\boldsymbol{\eta}, \sigma_{\text{N}}^2 &\overset{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{a}_i^{\text{T}}\boldsymbol{\eta}, \sigma_{\text{meas}}^2), \quad i = 1, \ldots, n_{\text{obs}}, \\ \boldsymbol{\eta} &\sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}_{\text{Prior}}^{-1}), \end{aligned}$$

where $\boldsymbol{a}_i$ selects the correct grid cell, $\mathbf{Q}_{\text{Prior}}^{-1}$ is the estimated precision matrix for the GMRF, and the Gaussian likelihood with nugget variance σ_{meas}^2 describes measurement noise and sub-grid variation. In general, we would estimate σ_{meas}^2 using a trial run, but here, we estimated σ_{meas}^2 using the average empirical variance over all observed grid cells in the total dataset. Note that we have not accounted for the uncertainty in the AUVs positions in these models. As the AUV dive, it loses its GPS signal and only relies on estimated location. When the GPS signal is returned a linear interpolation is made to account for drift but no uncertainty is included.

We evaluated the two priors, or emulators, by randomly ordering the 9 segments and then sequentially including more observations to predict the remaining hold-out data. The random

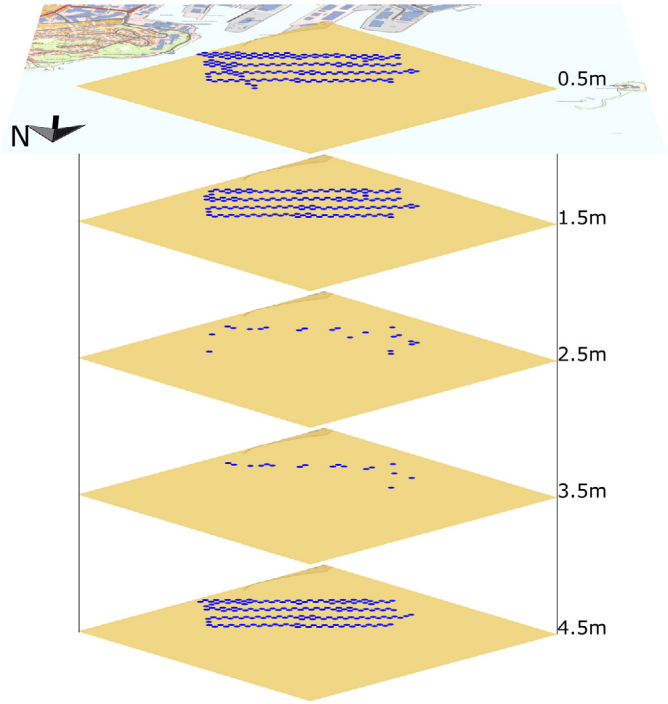


Fig. 8. Measurement locations of the AUV in the top 6 depth layers of the spatial field on May 27th, 2021, in Trondheimsfjorden at Ladehammaren just outside of Trondheim, Norway. The N-arrow is the cardinal north.

permutation of the segments was done repeatedly to determine the variation in scores over different paths. This scheme evaluates the AUVs' ability to predict future observations while maintaining the sequential structure of measurements. Fig. 9 shows that the non-stationary model provides a better prior for the salinity in the ocean than the stationary model. The differences are largest when little data is available, which is consistent with the idea that the prior is most important in this case. The non-stationary model can leverage knowledge about which areas are most uncertain using the spatially varying marginal variance and update the prior based on expected similarities from the spatially varying anisotropy. The improvements are seen both in point predictions through RMSE and in predictive distributions as measured by CPRS (Gneiting and Raftery, 2007).

Note that the standard deviations in both RMSE and CRPS are increasing after about 70% of observed data to 95% of observed data. This is partly because there are dynamic changes in salinity during the operation time. The pre-planned path which consisted of an up-and-down pattern in depth ranging from 0.5 m to 10.5 m was captured later than the others, and the measurements in this path are not well predicted by earlier measurements. This causes an increased variation between folds for 70% observed data and higher. It would be interesting to investigate if space–time models can account for the dynamic changes. Fig. S4 in the Appendix shows the cross-validation repeated without this segment. There is still a slight increase in standard deviation for a high proportion of observed data. This is because between-fold variation is increasingly emphasized as more and more data is collected since there are only eight possible choices for the segment placed last in the shuffled paths.

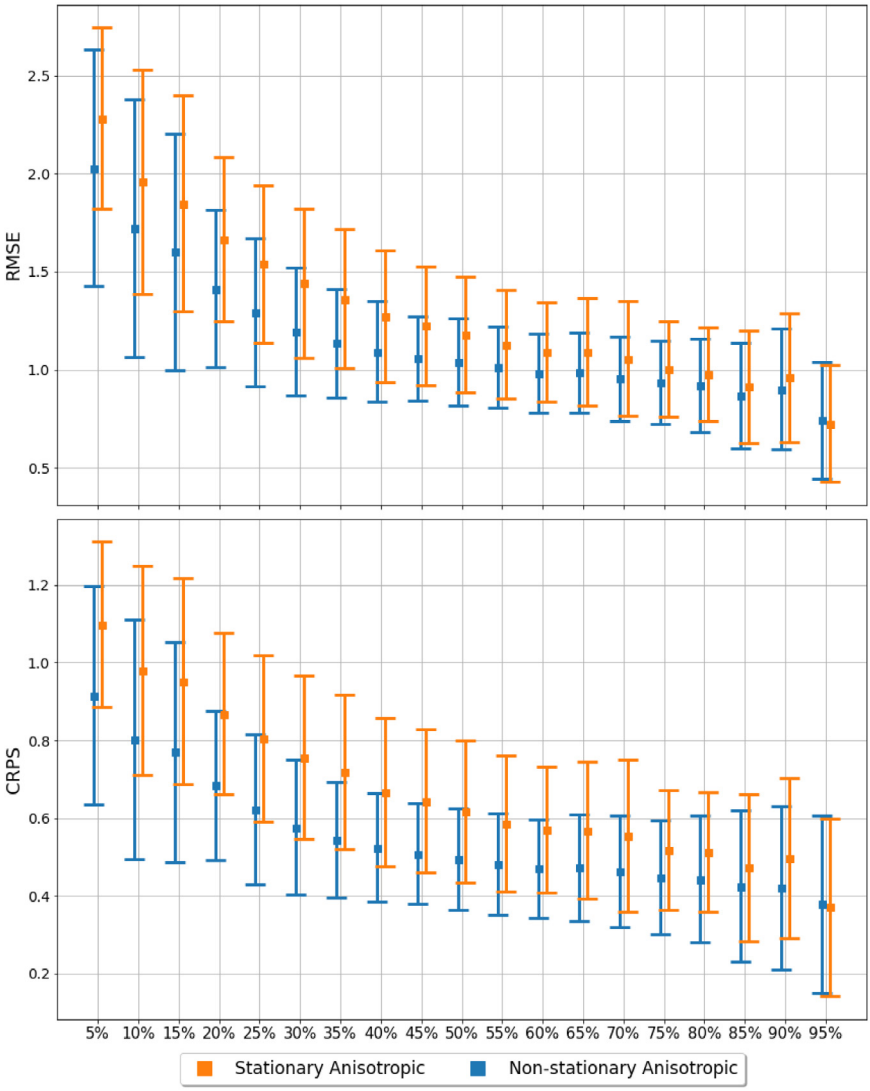


Fig. 9. The root mean square error (RMSE, top) and the continuous ranked probability score (CRPS, bottom) of predictions from the stationary anisotropic (orange) and non-stationary anisotropic models (blue) given different proportions of observed data (5%, 95%). The error bars are the standard deviations of the different measures under random permutations of the 9 segments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Discussion

We extend the class of SPDE-based GRFs introduced in Fuglstad et al. (2015a) to three-dimensional space by overcoming two key issues: parametrization and computation. For the former, we developed a specification of spatially varying anisotropy through a spatially varying

baseline isotropic dependence, and two orthogonal spatially varying vector fields that describe extra dependence. This allows for an interpretable description of the 3×3 positive definite matrix describing anisotropy. For the latter, we use a finite volume method to construct a GMRF that approximates the solution of the SPDE.

Specifying spatially varying marginal variance and spatially varying anisotropy requires specifying 7 spatially varying real functions. In this paper, we expand each function with a clamped B-spline basis. If each function uses P^3 basis functions, this gives in total $7P^3$ coefficients. As demonstrated in the simulation study, an unpenalized estimation of these parameters requires a densely observed area and multiple realizations. Application of the new models in data-sparse situations will require penalties that restrict the regularity of the 7 spatially varying functions. However, more research is needed to come up with a practical way to determine the appropriate strength of penalization for each of the functions.

While we did not experience any practical issues with the chosen way to describe the two orthogonal vector fields, the construction has a "gimbal lock" type issue. If one vector field points exactly along the z-axis, there is no unique choice for the second vector field. A potential way to avoid this issue is by describing the orientation of the two orthogonal vector fields through quaternions or Euler–Rodrigues parameters.

Moving from two-dimensional space to three-dimensional space introduces an asymptotically higher computation cost as a function of grid size. For a regular three-dimensional grid with N nodes, the computational cost is $\mathcal{O}(N^2)$ compared to $\mathcal{O}(N^{3/2})$ in two-dimensional space. This increased computational cost arises from increased fill-in in the Cholesky factor. However, the application demonstrates that the use of a grid size of $N = 22275$ is unproblematic even for real-time updates on an AUV.

In this paper, we have considered regular three-dimensional grids with cubical grid cells. Previous work has considered tetrahedral mesh using finite element methods (Zhang et al., 2016), and software is available for creating tetrahedral meshes (Lindgren et al., 2022, Sec. 6.4). Tetrahedral meshes may have computational benefits if the domain of interest is far from cubical since one can use fewer grid cells in the discretization. Further, they have the potential to better describe bathymetric boundaries such as the sea bed, or boundaries such as islands. A major motivation for deriving the finite volume method matrices in this paper is the desire to later extend them to space–time with advection–diffusion equations. Finite volume methods have advantages in local conservation and stabilization of fluxes with approaches such as upwind schemes. Further, finite volume methods avoid the mass lumping step required for finite element methods. An interesting future direction of study would be to develop the theoretical properties, which are not currently known, for the finite volume approach for this class of SPDEs.

For the predictions of salinity in the Trondheim's fjord, we see the highest improvement of the complex GRF prior compared to an isotropic GRF, for sparse in-situ measurements. As more data is collected, the difference between the models decreases. This suggests that the key advantage of training the more complex GRF is encoding prior physical knowledge to more effectively update knowledge about unobserved locations. Salinity was used as an example, but in general, the same approach could be used to map other biologically interesting quantities such as phytoplankton (Fossum et al., 2019). The GRFs developed in this paper are a step forward in quantifying beliefs about unobserved regions in the ocean, which is essential for optimal decisions and more effective autonomous sampling (Fossum et al., 2021).

In future work, it would be interesting to add a dynamic component to the model to capture physical processes such as diffusion and advection. However, this substantially increases computational cost, and it is not clear to which degree an advection field from a numerical model should be trusted and which boundary conditions are best in an advection-dominated problem. Future work may also consider whether non-Gaussian random fields (Bolin and Wallin, 2019) would lead to any improvements. The new class of GRFs shows great promise for encoding prior knowledge about a phenomenon in a computationally efficient way. However, overfitting is an important issue, and we must consider ways to penalize the complexity. In particular, we need to consider ways to allow flexibility in an area where it is needed such as a river outlet and restrict flexibility in areas where we expect stationarity.

Acknowledgments

Berild and Fuglstad are supported by the Research Council of Norway, project number 305445. The authors are grateful to Ingrid Ellingsen and SINTEF for providing the simulations from the numerical ocean model SINMOD.

Appendix A. General properties

A.1. Marginal variance

Here, we will derive the expression for the marginal variance in a general sense and then specify it for three-dimensional spaces with exponential covariance functions. The SPDE considered in this work is

$$(\kappa^2 - \nabla \cdot \mathbf{H}\nabla)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \tag{S1}$$

where $\boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^d$ a spatial location in the domain of dimension d and $\alpha = \nu + d/2$ where $\nu > 0$ is the smoothness. Any solution of this SPDE is a Matérn field and let $\sigma_m > 0$ be its marginal standard deviation; then, its covariance function is

$$r(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\sigma_m^2}{2^{\nu-1}\Gamma(\nu)} (\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|)^{\nu} K_{\nu}(\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|). \tag{S2}$$

The transfer function of the SPDE is

$$g(\boldsymbol{w}) = (\kappa^2 + \boldsymbol{w}^T \mathbf{H} \boldsymbol{w})^{-\alpha/2}.$$

Using this and by including the spectral density of standard Gaussian white noise in $\mathbb{R}^d$ is $(2\pi)^{-d}$, the spectral density of the solution of the SPDE is

$$f_S(\boldsymbol{w}) = (2\pi)^{-d} (\kappa^2 + \boldsymbol{w}^T \mathbf{H} \boldsymbol{w})^{-\alpha}.$$

Lastly, to find the marginal variance of the field the integral of the spectral density is made over $\mathbb{R}^d$ as

$$\sigma_m^2 = \int_{\mathbb{R}^d} f_S(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}.$$

Including the change of variables $\boldsymbol{w} = \kappa \mathbf{H}^{-1/2} \boldsymbol{z}$ the expression becomes

$$\begin{aligned} \sigma_m^2 &= (2\pi)^{-d} \int_{\mathbb{R}^d} (\kappa^2 + \kappa^2 \boldsymbol{z}^T \boldsymbol{z})^{-\alpha} \det(\kappa \mathbf{H}^{-1/2}) \mathrm{d}\boldsymbol{z} \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \kappa^{d-2\alpha} (1 + \boldsymbol{z}^T \boldsymbol{z})^{-\alpha} \det(\mathbf{H})^{-1/2} \mathrm{d}\boldsymbol{z} \\ &\overset{\alpha=\nu+d/2}{=} (2\pi)^{-d} \kappa^{-2\nu} \det(\mathbf{H})^{-1/2} \int_{\mathbb{R}^d} (1 + \boldsymbol{z}^T \boldsymbol{z})^{-\alpha} \mathrm{d}\boldsymbol{z}, \end{aligned} \tag{S3}$$

which by specifying a exponential covariance in $\mathbb{R}^3$ with $\alpha = 2$, $\nu = 1/2$ and $d = 3$ is

$$\sigma_m^2 = \frac{1}{8\pi\kappa\sqrt{\det(\mathbf{H})}}.$$

Note that the integral in is solved by converting to polar coordinates as

$$\int_{\mathbb{R}^3} \frac{1}{(1 + \boldsymbol{z}^T \boldsymbol{z})^2} \mathrm{d}\boldsymbol{z} = \int_0^{\pi} \sin(\phi) \mathrm{d}\phi \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \frac{\rho^2}{(1 + \rho^2)^2} \mathrm{d}\rho = \pi^2.$$

A.2. Covariance function

Evaluating Eq. (S2) at $\nu = 1/2$ and including the expression for the marginal variance the covariance function can be formalized as

$$r(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sqrt{\frac{2}{\pi}} \frac{1}{8\pi\kappa\sqrt{\det(\mathbf{H})}} \sqrt{\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|} K_{\frac{1}{2}}(\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|).$$

Then, consider the modified Bessel function of the second kind

$$K_n(z) = \sqrt{\frac{\pi}{2z}} \frac{e^{-z}}{(n-\frac{1}{2})!} \int_0^\infty e^{-t} t^{n-1/2} \left(1 - \frac{t}{2z}\right)^{n-1/2} \mathrm{d}t,$$

and evaluate this at order 1/2 gives

$$K_{\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}} e^{-z}.$$

The covariance function can then be formalized as

$$\begin{aligned} r(\boldsymbol{s}_1, \boldsymbol{s}_2) =& \sqrt{\frac{2}{\pi}} \sigma_m^2 \sqrt{\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|} \\ & \times \sqrt{\frac{\pi}{2 \cdot \kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|}} \exp\left(-\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|\right) \\ =& \sigma_m^2 \exp\left(-\kappa \|\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)\|\right). \end{aligned} \tag{S4}$$

A.3. One-dimensional clamped B-splines

We illustrate the construction of 1-dimensional B-splines using the interval $[A, B] \in \mathbb{R}$. Let $A = t_0 < t_1 < \cdots < t_m = B$ be the knot points. Then the zero-order B-splines are constructed recursively as

$$B_{i,0}(t) = \begin{cases} 1, & t_i \le t \le t_{i+1}, \\ 0, & \text{otherwise}, \end{cases} \qquad t \in [A, B],$$

for $i = 0, \ldots, p-1$. Let r denote the order of the B-splines. The first- and second-order basis splines are constructed as

$$B_{i,r}(t) = \frac{t - t_i}{t_{i+r} - t_i} B_{i,r-1}(t) + \frac{t_{i+r+1} - t}{t_{i+r+1} - t_{i+1}} B_{i+1,r-1}(t), \quad t \in [A, B],$$

for $i = 0, \ldots, p - r - 1$.

Using the r-order B-spline basis, we construct a function $g : [A, B] \to \mathbb{R}$ by

$$g(t) = \sum_{i=0}^{p-r-1} \alpha_i B_{i,r}(t).$$

where $\alpha_0, \ldots, \alpha_{p-r-1} \in \mathbb{R}$ are coefficients. We use a clamped spline where $g'(A) = g'(B) = 0$ and need the additional requirement that $\alpha_0 = \alpha_1$ and $\alpha_{p-r-2} = \alpha_{p-r-1}$.

A.4. Integrated likelihood

The distribution of $\boldsymbol{z} = (\boldsymbol{u}, \boldsymbol{\beta})$ is given by

$$\boldsymbol{z}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_z^{-1}),$$

and the observation model is

$$\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}, \sigma_N^2 \sim \mathcal{N}_n(\mathbf{S}\boldsymbol{z}, \mathbf{I}_n \sigma_N^2).$$

From this the distribution of $\boldsymbol{z}$ given some observations $\boldsymbol{y}$ is

$$\begin{aligned}
\pi(\boldsymbol{z}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y}) &\propto \pi(\boldsymbol{z}, \boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y}) \\
&= \pi(\boldsymbol{\theta}, \sigma_N^2)\pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{z}) \\
&\propto \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{z} - \frac{1}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{z})^{\mathrm{T}}\mathbf{I}_n\sigma_N^{-2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{z})\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{z}^{\mathrm{T}}\left(\mathbf{Q}_z + \sigma_N^{-2}\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)\boldsymbol{z} - 2\boldsymbol{z}^{\mathrm{T}}\mathbf{S}^{T}\boldsymbol{y}\cdot\sigma_N^{-2}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_C)^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{z} - \boldsymbol{\mu}_C)\right) \\
&\Downarrow \\
\boldsymbol{z}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y} &\sim \mathcal{N}_n\left(\boldsymbol{\mu}_C, \mathbf{Q}_C^{-1}\right)
\end{aligned}$$

Here, $\mathbf{Q}_C = \mathbf{Q}_z + \mathbf{S}^{\mathrm{T}}\mathbf{S}\cdot\sigma_N^{-2}$ is the conditional precision matrix and $\boldsymbol{\mu}_C = \mathbf{Q}_C^{-1}\mathbf{S}^{\mathrm{T}}\boldsymbol{y}\cdot\sigma_N^{-2}$ is the conditional mean.

Then, integrating out $\boldsymbol{z}$ from the joint distribution gives

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{z}, \sigma_N^2, \boldsymbol{y})}{\pi(\boldsymbol{z}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y})} \\
&= \frac{\pi(\boldsymbol{\theta}, \sigma_N^2)\pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{z})}{\pi(\boldsymbol{z}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y})},
\end{aligned}$$

where the left-hand side does not depend on $\boldsymbol{z}$ such that it may be evaluated for any given value. Let us evaluate it for $\boldsymbol{z} = \boldsymbol{\mu}_C$ such that

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y}) \propto& \frac{\pi(\boldsymbol{\theta}, \sigma_N^2)\pi(\boldsymbol{z} = \boldsymbol{\mu}_C|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{z} = \boldsymbol{\mu}_C)}{\pi(\boldsymbol{z} = \boldsymbol{\mu}_C|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y})} \\
\propto& \pi(\boldsymbol{\theta})\frac{|\mathbf{Q}_z|^{1/2}|\mathbf{I}_n\cdot\sigma_N^{-2}|^{1/2}}{|\mathbf{Q}_C|^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{\mu}_C\right) \\
&\times \exp\left(-\frac{1}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)^{\mathrm{T}}\mathbf{I}_n\cdot\sigma_N^{-2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)\right).
\end{aligned}$$

The last term $\pi(\boldsymbol{z}|\boldsymbol{\theta}, \sigma_N^2, \boldsymbol{y})$ is removed since it is equal to 1. Thereby, conditioning on $\boldsymbol{y}$ and taking the log we have the log-likelihood

$$\begin{aligned}
\log(\pi(\boldsymbol{\theta}, \sigma_N^2|\boldsymbol{y})) =& \text{Constant} + \log(\pi(\boldsymbol{\theta}, \sigma_N^2)) + \frac{1}{2}\log(\det(\mathbf{Q}_z)) + \frac{n}{2}\log(\sigma_N^{-2}) \\
&- \frac{1}{2}\log(\det(\mathbf{Q}_C)) - \frac{1}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{\mu}_C - \frac{1}{2\cdot\sigma_N^2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)^{\mathrm{T}}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C).
\end{aligned} \tag{S5}$$

A.5. Gradient of the log-likelihood

This section is similar to the derivation of the gradient presented in the supplementary material of Fuglstad et al. (2015b).

$$\begin{aligned}
\log(\pi(\boldsymbol{\theta}, \tau_N|\boldsymbol{y})) =& \text{Constant} + \log(\pi(\boldsymbol{\theta}, \tau_N)) + \frac{1}{2}\log(\det(\mathbf{Q}_z)) + \frac{n}{2}\log(\sigma_N^{-2}) \\
&- \frac{1}{2}\log(\det(\mathbf{Q}_C)) + \frac{1}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\mathbf{Q}_C\boldsymbol{\mu}_C - \frac{\tau_N}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}.
\end{aligned}$$

Note that the last two terms are rewritten for simplicity in the gradient calculation and that the variance of the Gaussian noise term, σ_N^2 is re-parametrized with its inverse $\tau_N = 1/\sigma_N^2$ (precision).

Derivatives of the log-likelihood are taken with respect to θ_i, the elements of $\boldsymbol{\theta}$, and the precision on log scale as $\log(\tau_N)$.

The first term is a constant and therefore its derivative is zero with respect to any of the parameters. The next term, the penalty or the prior of the parameters, is not used in this paper and otherwise depends on the choice of penalty so gradient calculation is not specified for this term.

To continue note the derivatives of the precision matrix

$$\frac{\partial \mathbf{Q}_C}{\partial \theta_i} = \frac{\partial \mathbf{Q}_z}{\partial \theta_i} \quad \text{and} \quad \frac{\partial \mathbf{Q}_C}{\partial \log(\tau_N)} = \mathbf{S}^T \mathbf{S} \tau_N,$$

which is used in the following derivations. First, the derivatives with respect to θ_i are considered. The derivative of the log determinant terms are

$$\begin{aligned}\frac{\partial}{\partial \theta_i}\left(\log(\det(\mathbf{Q})) - \log(\det(\mathbf{Q}_C))\right) =& \operatorname{Tr}\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) - \operatorname{Tr}\left(\mathbf{Q}_C^{-1}\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) \\ =& \operatorname{Tr}\left((\mathbf{Q}^{-1} - \mathbf{Q}_C^{-1})\frac{\partial \mathbf{Q}}{\partial \theta_i}\right),\end{aligned}$$

and the derivative of the quadratic terms are

$$\begin{aligned}\frac{\partial}{\partial \theta_i}\left(\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\tau_N + \frac{1}{2}\boldsymbol{\mu}_C^T\mathbf{Q}_C\boldsymbol{\mu}_C\right) =& \frac{\partial}{\partial \theta_i}\left(\frac{1}{2}\boldsymbol{\mu}_C^T\mathbf{Q}_C\boldsymbol{\mu}_C\right) \\ =& -\frac{1}{2}\boldsymbol{y}^T\tau_N\mathbf{S}\mathbf{Q}_C^{-1}\left(\frac{\partial \mathbf{Q}_C}{\partial \theta_i}\right)\mathbf{Q}_C^{-1}\mathbf{S}^T\tau_N\boldsymbol{y} \\ =& -\frac{1}{2}\boldsymbol{\mu}_C^T\left(\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)\boldsymbol{\mu}_C.\end{aligned}$$

Then, combining these the derivative of the log-likelihood with respect to θ_i is

$$\frac{\partial}{\partial \theta_i}\log(\pi(\boldsymbol{\theta}, \tau_N|\boldsymbol{y})) = \frac{\partial}{\partial \theta_i}\log(\pi(\boldsymbol{\theta}, \tau_N)) + \frac{1}{2}\operatorname{Tr}\left((\mathbf{Q}^{-1} - \mathbf{Q}_C^{-1})\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) - \frac{1}{2}\boldsymbol{\mu}_C^T\left(\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)\boldsymbol{\mu}_C$$

Next, the derivative with respect to the log precision, $\log \tau_N$, is considered. The derivative of the log determinant terms are

$$\begin{aligned}\frac{\partial}{\partial \log(\tau_N)}\left(\frac{n}{2}\log(\tau_N) - \frac{1}{2}\log(\det(\mathbf{Q}_C))\right) =& \frac{n}{2} - \frac{1}{2}\operatorname{Tr}\left(\mathbf{Q}_C^{-1}\frac{\partial}{\partial \log(\tau_N)}\mathbf{Q}_C\right) \\ =& \frac{n}{2} - \frac{1}{2}\operatorname{Tr}\left(\mathbf{Q}_C^{-1}\mathbf{S}^T\mathbf{S}\cdot\tau_N\right)\end{aligned}$$

Further, the derivative of $1/2\boldsymbol{y}^T\boldsymbol{y}\cdot\tau_N$ with respect to $\log(\tau_N)$ is just the same expression so the remaining quadratic term becomes

$$\begin{aligned}\frac{\partial \frac{1}{2}\boldsymbol{\mu}_C^T\mathbf{Q}_C\boldsymbol{\mu}_C}{\partial \log(\tau_N)} =& \frac{\partial \frac{1}{2}\boldsymbol{y}^T\tau_N\mathbf{S}\mathbf{Q}_C^{-1}\mathbf{S}^T\tau_N\boldsymbol{y}}{\partial \log(\tau_N)} \\ =& \boldsymbol{y}^T\tau_N\mathbf{S}\mathbf{Q}_C^{-1}\mathbf{S}^T\frac{\partial \tau_N}{\partial \log(\tau_N)}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}^T\tau_N\mathbf{S}\mathbf{Q}_C^{-1}\frac{\partial \mathbf{Q}_C}{\partial \log(\tau_N)}\mathbf{Q}_C^{-1}\mathbf{S}^T\tau_N\boldsymbol{y} \\ =& \boldsymbol{\mu}_C^T\mathbf{S}^T\tau_N\boldsymbol{y} - \frac{1}{2}\boldsymbol{\mu}_C^T\mathbf{S}^T\mathbf{S}\boldsymbol{\mu}_C\tau_N,\end{aligned}$$

and then, by adding the last quadratic term, the expression simplifies to

$$-1/2\boldsymbol{y}^T\boldsymbol{y}\cdot\tau_N + \boldsymbol{\mu}_C^T\mathbf{S}^T\boldsymbol{y}\cdot\tau_N - \frac{1}{2}\boldsymbol{\mu}_C^T\mathbf{S}^T\mathbf{S}\boldsymbol{\mu}_C\cdot\tau_N = -\frac{1}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)^T(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)\cdot\tau_N.$$

Finally, combining all these terms we have the derivative of the log-likelihood with respect to $\log(\tau_N)$:

$$\frac{\partial \log(\pi(\boldsymbol{\theta}, \tau_N|\boldsymbol{y}))}{\partial \log(\tau_N)} = \frac{\partial \log(\pi(\boldsymbol{\theta}, \tau_N))}{\partial \log(\tau_N)} + \frac{n}{2} - \frac{1}{2}\text{Tr}\left(\mathbf{Q}_C^{-1}\mathbf{S}^{\text{T}}\mathbf{S} \cdot \tau_N\right) - \frac{1}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C)^{\text{T}}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_C) \cdot \tau_N$$

Note that the derivative of $\mathbf{Q}_C$ can be calculated quickly and it is derived from a series of chain rules; first on $\mathbf{Q}_C$, then on $\mathbf{A}$ and $\mathbf{A}_\mathbf{H}$, and finally within $\mathbf{H}$. The most computationally heavy calculation in the gradient of the log-likelihood is to calculate the inverses in the difference $\mathbf{Q}^{-1} - \mathbf{Q}_C^{-1}$. However, since this term is multiplied with the derivative of $\mathbf{Q}$ with respect to θ_i, which carries the non-zero structure of $\mathbf{Q}$, only elements of $\mathbf{Q}^{-1}$ and $\mathbf{Q}_C^{-1}$ which correspond to the non-zero structure of $\mathbf{Q}$ need to be calculated. This is done by calculating a partial inverse of two matrices as described in Rue and Held (2010).

Appendix B. Derivation

B.1. Discretization

To find the local solution of the SPDE the domain $\mathcal{D} = [A_1, B_1] \times [A_2, B_2] \times [A_3, B_3]$ is divided into equally sized rectangular cubes or cells. We use M cells to divide $[A_1, B_1]$ in the x-direction, N cells on $[A_2, B_2]$ in y-direction and P cells on $[A_3, B_3]$ in z-direction. The cells have sides parallel to each axis of size $h_x = (B_1 - A_1)/M$, $h_y = (B_2 - A_2)/N$, and $h_z = (B_3 - A_3)/P$. The cells are assigned an index with regards to their cell number along each axes starting from number 0; $i \in [0, M]$ along x, $j \in [0, N]$ along y, and $k \in [0, P]$ along z. For a specific cell, its domain can be denoted as

$$E_{i,j,k} = [ih_x, (i+1)h_x] \times [jh_y, (j+1)h_y] \times [kh_z, (k+1)h_z],$$

and Fig. S1 shows this cell and its closest neighbors. Furthermore, as a regular grid is employed the volume of a cell is $V = h_x h_y h_z$.

To further define the local solution of the SPDE we denote the faces of a grid cell as $\sigma^F_{i,j,k}$ (front), $\sigma^B_{i,j,k}$ (back), $\sigma^L_{i,j,k}$ (left), $\sigma^R_{i,j,k}$ (right), $\sigma^U_{i,j,k}$ (up) and $\sigma^D_{i,j,k}$ (down) with their respective face centers $\boldsymbol{s}_{i,j-1/2,k}$, $\boldsymbol{s}_{i,j+1/2,k}$, $\boldsymbol{s}_{i-1/2,j,k}$, $\boldsymbol{s}_{i+1/2,j,k}$, $\boldsymbol{s}_{i,j,k+1/2}$ and $\boldsymbol{s}_{i,j,k-1/2}$. Fig. S2 describes the different faces of a cell.

B.2. Local solution of the SPDE

Note that this description is an extension to three dimensions of the derivation described in Fuglstad et al. (2015a), and the reader is referred to there for further details. To locally solve the SPDE a finite volume scheme is derived. First, Eq. (S1) is integrated over a cell $E_{i,j,k}$ as

$$\int_{E_{ijk}} \kappa^2(\boldsymbol{s})u(\boldsymbol{s})\text{d}\boldsymbol{s} - \int_{E_{ijk}} \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s})\text{d}\boldsymbol{s} = \int_{E_{ijk}} \mathcal{W}(\boldsymbol{s})\text{d}\boldsymbol{s}, \tag{S6}$$

where $\text{d}\boldsymbol{s}$ is a volume element. The integral of the Gaussian white noise on the right-hand side is a Gaussian variable with mean zero and variance equal to the volume of a cell which is independent of neighboring cells. Let z_{ijk} be an standard Gaussian variable; then, Eq. (S6) becomes

$$\int_{E_{ijk}} \kappa^2(\boldsymbol{s})u(\boldsymbol{s})\text{d}\boldsymbol{s} - \int_{E_{ijk}} \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s})\text{d}\boldsymbol{s} = \sqrt{V}z_{ijk}.$$

Then, applying the divergence theorem to the second integral with the divergence operator gives

$$\int_{E_{ijk}} \kappa^2(\boldsymbol{s})u(\boldsymbol{s})\text{d}\boldsymbol{s} - \oint_{\partial E_{ijk}} (\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^T \boldsymbol{n}(\boldsymbol{s})\text{d}\sigma = \sqrt{V}z_{ijk}.$$

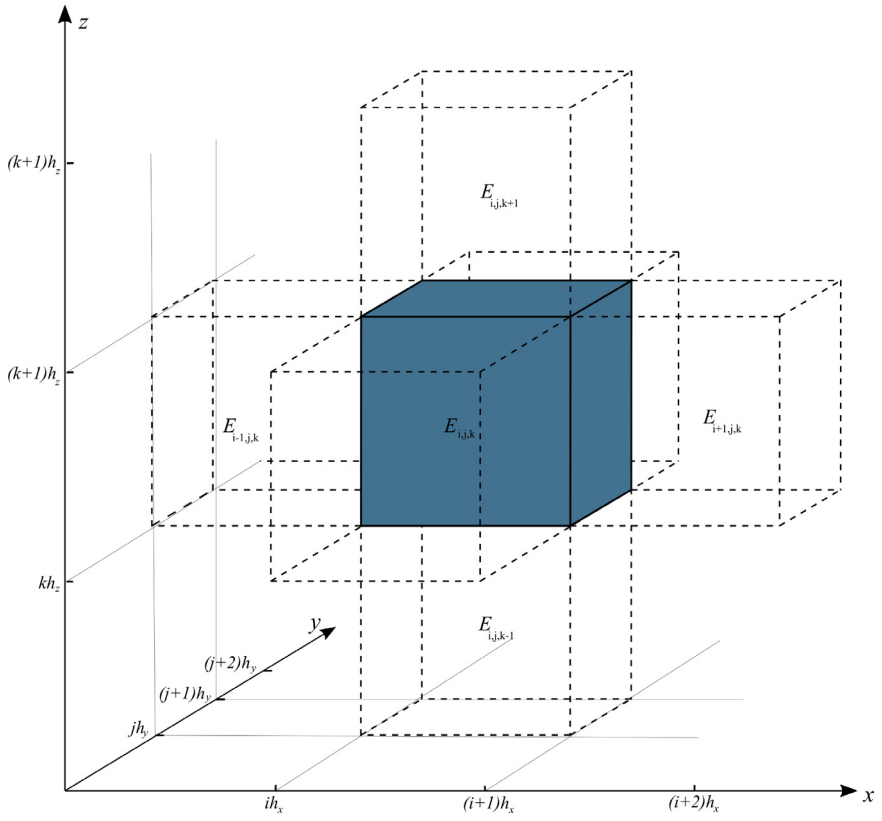


Fig. S1. One cell $E_{i,j,k}$ in the discretization with its closest neighbors; $E_{i+1,j,k}$, $E_{i-1,j,k}$, $E_{i,j+1,k}$, $E_{i,j-1,k}$, $E_{i,j,k+1}$, and $E_{i,j,k-1}$.

The first integral is approximated by letting k_{ijk}^2 be the average value of the continuous function $\kappa^2(\boldsymbol{s})$ within a cell, i.e. $\kappa_{ijk}^2 = 1/V \int_{E_{ijk}} \kappa^2(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$, resulting in

$$V\kappa_{ijk}^2 u_{ijk} - \oint_{\partial E_{ijk}} (\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^T \boldsymbol{n}(\boldsymbol{s}) \mathrm{d}\sigma = \sqrt{V} z_{ijk}. \tag{S7}$$

To describe the solution of the second integral it is divided into integrals over each surface as

$$\oint_{\partial E_{ijk}} (\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^T \boldsymbol{n}(\boldsymbol{s}) \mathrm{d}\sigma = W_{ijk}^L + W_{ijk}^R + W_{ijk}^B + W_{ijk}^F + W_{ijk}^U + W_{ijk}^D, \tag{S8}$$

or $W_{ijk}^{\text{dir}} = \int_{\sigma_{ijk}^{\text{dir}}} (\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^T \boldsymbol{n}(\boldsymbol{s}) \mathrm{d}\sigma$, where dir denotes the surface; R (positive x-direction), L (negative x-direction), B (positive y-direction), F (negative y-direction), U (positive z-direction), and D (negative z-direction). Now, an approximation of this surface integral over each face is required. It is assumed that the gradient of $u(\boldsymbol{s})$ is constant over each face and equal to the value at the center of each face. The resulting scheme for the gradient on each face is described in Table S1. Furthermore, let $\mathbf{H}$ be approximated by its value at the center of the face, and then, we have the

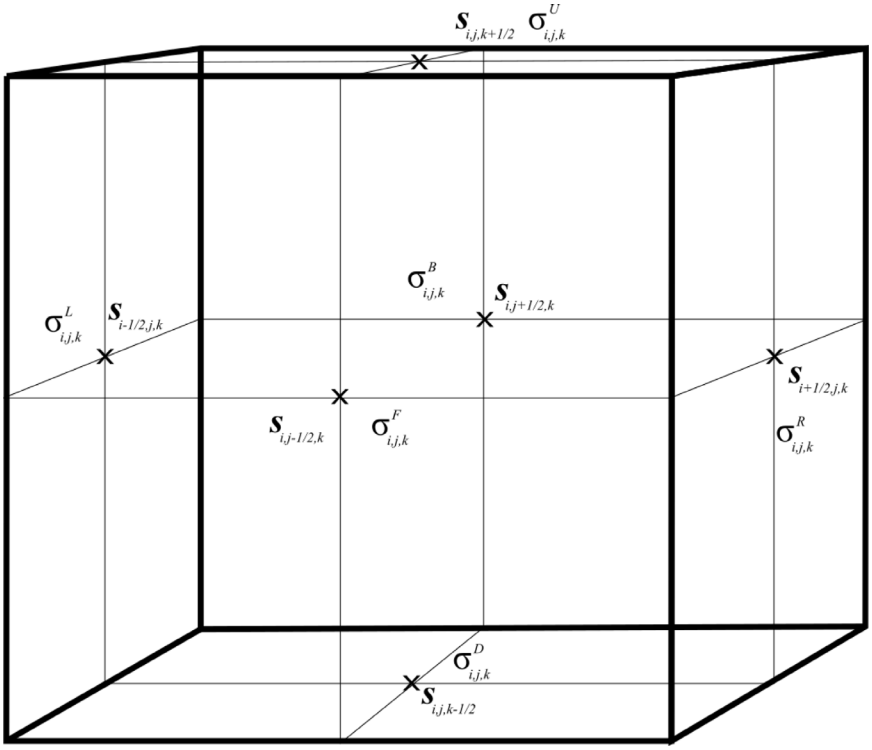


Fig. S2. One cell $E_{i,j,k}$ of the discretization with all its faces; $\sigma^{F}_{i,j,k}$ (front), $\sigma^{B}_{i,j,k}$ (back), $\sigma^{L}_{i,j,k}$ (left), $\sigma^{R}_{i,j,k}$ (right), $\sigma^{U}_{i,j,k}$ (up), and $\sigma^{D}_{i,j,k}$ (down) each with its respective face centers.

approximation

$$\begin{aligned} W^{\text{dir}}_{ijk} &= \int_{\sigma^{\text{dir}}_{ijk}} \nabla u(\boldsymbol{s})^T \mathbf{H}(\boldsymbol{s}) \boldsymbol{n}(\boldsymbol{s}) \mathrm{d}\sigma \\ &\approx \nabla u(\boldsymbol{c}^{\text{dir}}_{ijk})^T \mathbf{H}(\boldsymbol{c}^{\text{dir}}_{ijk}) \boldsymbol{n}(\boldsymbol{c}^{\text{dir}}_{ijk}) \int_{\sigma^{\text{dir}}_{ijk}} \mathrm{d}\sigma \\ &= \nabla u(\boldsymbol{c}^{\text{dir}}_{ijk})^T \mathbf{H}(\boldsymbol{c}^{\text{dir}}_{ijk}) \boldsymbol{n}(\boldsymbol{c}^{\text{dir}}_{ijk}) \mathrm{A}(\sigma^{\text{dir}}_{ijk}), \end{aligned} \tag{S9}$$

where $\boldsymbol{c}^{\text{dir}}_{ijk}$ is the center of face dir in the cell E_{ijk}, and $A(\sigma^{\text{dir}}_{ijk})$ is the area of the face. Combining Eq. (S9) with the scheme of $\nabla u(\boldsymbol{c}^{\text{dir}}_{ijk})$ from Table S1, and denoting the components of $\mathbf{H}$ as

$$\mathbf{H}(\boldsymbol{s}) = \begin{bmatrix} H^{11}(\boldsymbol{s}) & H^{12}(\boldsymbol{s}) & H^{13}(\boldsymbol{s}) \\ H^{21}(\boldsymbol{s}) & H^{22}(\boldsymbol{s}) & H^{23}(\boldsymbol{s}) \\ H^{31}(\boldsymbol{s}) & H^{32}(\boldsymbol{s}) & H^{33}(\boldsymbol{s}) \end{bmatrix}$$

the approximations for each face become

$$\hat{W}^{R}_{i,j,k} = h_y h_z \left[\mathbf{H}^{11}(\boldsymbol{s}_{i+1/2,j,k}) \frac{u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i,j,k})}{h_x} \right] +$$

Table S1
Numerical scheme of the partial derivative with respect to x, y and z of u_{ijk} on the different faces of cell E_{ijk}.

Face	Scheme
$\sigma^R_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i+1/2,j,k}) \simeq \frac{1}{h_x}\left(u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i,j,k})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i+1/2,j,k}) \simeq \frac{1}{4h_y}\left(u(\boldsymbol{s}_{i+1,j+1,k}) + u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i+1,j-1,k}) - u(\boldsymbol{s}_{i,j-1,k})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i+1/2,j,k}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i+1,j,k+1}) + u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i+1,j,k-1}) - u(\boldsymbol{s}_{i,j,k-1})\right)$
$\sigma^L_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i-1/2,j,k}) \simeq \frac{1}{h_x}\left(u(\boldsymbol{s}_{i,j,k}) - u(\boldsymbol{s}_{i-1,j,k})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i-1/2,j,k}) \simeq \frac{1}{4h_y}\left(u(\boldsymbol{s}_{i,j+1,k}) + u(\boldsymbol{s}_{i-1,j+1,k}) - u(\boldsymbol{s}_{i,j-1,k}) - u(\boldsymbol{s}_{i-1,j-1,k})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i-1/2,j,k}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i,j,k+1}) + u(\boldsymbol{s}_{i-1,j,k+1}) - u(\boldsymbol{s}_{i,j,k-1}) - u(\boldsymbol{s}_{i-1,j,k-1})\right)$
$\sigma^B_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i,j+1/2,k}) \simeq \frac{1}{4h_x}\left(u(\boldsymbol{s}_{i+1,j+1,k}) + u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i-1,j+1,k}) - u(\boldsymbol{s}_{i-1,j,k})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i,j+1/2,k}) \simeq \frac{1}{h_y}\left(u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j,k})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i,j+1/2,k}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i,j+1,k+1}) + u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j+1,k-1}) - u(\boldsymbol{s}_{i,j,k-1})\right)$
$\sigma^F_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i,j-1/2,k}) \simeq \frac{1}{4h_x}\left(u(\boldsymbol{s}_{i+1,j,k}) + u(\boldsymbol{s}_{i+1,j-1,k}) - u(\boldsymbol{s}_{i-1,j,k}) - u(\boldsymbol{s}_{i-1,j-1,k})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i,j-1/2,k}) \simeq \frac{1}{h_y}\left(u(\boldsymbol{s}_{i,j,k}) - u(\boldsymbol{s}_{i,j-1,k})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i,j-1/2,k}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i,j,k+1}) + u(\boldsymbol{s}_{i,j-1,k+1}) - u(\boldsymbol{s}_{i,j,k-1}) - u(\boldsymbol{s}_{i,j-1,k-1})\right)$
$\sigma^U_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i,j,k+1/2}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i+1,j,k+1}) + u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i-1,j,k+1}) - u(\boldsymbol{s}_{i-1,j,k})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i,j,k+1/2}) \simeq \frac{1}{4h_y}\left(u(\boldsymbol{s}_{i,j+1,k+1}) + u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j-1,k+1}) - u(\boldsymbol{s}_{i,j-1,k})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i,j,k+1/2}) \simeq \frac{1}{h_x}\left(u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j,k})\right)$
$\sigma^D_{i,j,k}$	$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i,j,k-1/2}) \simeq \frac{1}{4h_z}\left(u(\boldsymbol{s}_{i+1,j,k}) + u(\boldsymbol{s}_{i+1,j,k-1}) - u(\boldsymbol{s}_{i-1,j,k}) - u(\boldsymbol{s}_{i-1,j,k-1})\right)$ $\frac{\partial}{\partial y}u(\boldsymbol{s}_{i,j,k-1/2}) \simeq \frac{1}{4h_y}\left(u(\boldsymbol{s}_{i,j+1,k}) + u(\boldsymbol{s}_{i,j+1,k-1}) - u(\boldsymbol{s}_{i,j-1,k}) - u(\boldsymbol{s}_{i,j-1,k-1})\right)$ $\frac{\partial}{\partial z}u(\boldsymbol{s}_{i,j,k-1/2}) \simeq \frac{1}{h_x}\left(u(\boldsymbol{s}_{i,j,k}) - u(\boldsymbol{s}_{i,j,k-1})\right)$

$$h_y h_z\left[\mathbf{H}^{21}(\boldsymbol{s}_{i+1/2,j,k})\frac{u(\boldsymbol{s}_{i+1,j+1,k}) + u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i+1,j-1,k}) - u(\boldsymbol{s}_{i,j-1,k})}{4h_y}\right] +$$
$$h_y h_z\left[\mathbf{H}^{31}(\boldsymbol{s}_{i+1/2,j,k})\frac{u(\boldsymbol{s}_{i+1,j,k+1}) + u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i+1,j,k-1}) - u(\boldsymbol{s}_{i,j,k-1})}{4h_z}\right],$$

$$\hat{W}^L_{i,j,k} =$$
$$h_y h_z\left[\mathbf{H}^{11}(\boldsymbol{s}_{i-1/2,j,k})\frac{u(\boldsymbol{s}_{i-1,j,k}) - u(\boldsymbol{s}_{i,j,k})}{h_x}\right] +$$
$$h_y h_z\left[\mathbf{H}^{21}(\boldsymbol{s}_{i-1/2,j,k})\frac{u(\boldsymbol{s}_{i,j-1,k}) + u(\boldsymbol{s}_{i-1,j-1,k}) - u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i-1,j+1,k})}{4h_y}\right] +$$
$$h_y h_z\left[\mathbf{H}^{31}(\boldsymbol{s}_{i-1/2,j,k})\frac{u(\boldsymbol{s}_{i,j,k-1}) + u(\boldsymbol{s}_{i-1,j,k-1}) - u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i-1,j,k+1})}{4h_z}\right],$$

$$\hat{W}^B_{i,j,k} =$$
$$h_x h_z\left[\mathbf{H}^{12}(\boldsymbol{s}_{i,j+1/2,k})\frac{u(\boldsymbol{s}_{i+1,j+1,k}) + u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i-1,j+1,k}) - u(\boldsymbol{s}_{i-1,j,k})}{4h_x}\right] +$$
$$h_x h_z\left[\mathbf{H}^{22}(\boldsymbol{s}_{i,j+1/2,k})\frac{u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j,k})}{h_y}\right] +$$
$$h_x h_z\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j+1/2,k})\frac{u(\boldsymbol{s}_{i,j+1,k+1}) + u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j+1,k-1}) - u(\boldsymbol{s}_{i,j,k-1})}{4h_z}\right],$$

$$
\begin{aligned}
&\hat{W}^F_{i,j,k} = \\
&h_x h_z \left[\mathbf{H}^{12}(\boldsymbol{s}_{i,j-1/2,k}) \frac{u(\boldsymbol{s}_{i-1,j,k}) + u(\boldsymbol{s}_{i-1,j-1,k}) - u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i+1,j-1,k})}{4h_x} \right] + \\
&h_x h_z \left[\mathbf{H}^{22}(\boldsymbol{s}_{i,j-1/2,k}) \frac{u(\boldsymbol{s}_{i,j-1,k}) - u(\boldsymbol{s}_{i,j,k})}{h_y} \right] + \\
&h_x h_z \left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j-1/2,k}) \frac{u(\boldsymbol{s}_{i,j,k-1}) + u(\boldsymbol{s}_{i,j-1,k-1}) - u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j-1,k+1})}{4h_z} \right], \\
&\hat{W}^U_{i,j,k} = \\
&h_x h_y \left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i+1,j,k+1}) + u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i-1,j,k+1}) - u(\boldsymbol{s}_{i-1,j,k})}{4h_x} \right] + \\
&h_x h_y \left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i,j+1,k+1}) + u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j-1,k+1}) - u(\boldsymbol{s}_{i,j-1,k})}{4h_y} \right] + \\
&h_x h_y \left[\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j,k})}{h_z} \right], \\
&\hat{W}^D_{i,j,k} = \\
&h_x h_y \left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i-1,j,k}) + u(\boldsymbol{s}_{i-1,j,k-1}) - u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i+1,j,k-1})}{4h_x} \right] + \\
&h_x h_y \left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i,j-1,k}) + u(\boldsymbol{s}_{i,j-1,k-1}) - u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j+1,k-1})}{4h_y} \right] + \\
&h_x h_y \left[\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i,j,k-1}) - u(\boldsymbol{s}_{i,j,k})}{h_z} \right], \\
&\hat{W}^T_{i,j,k} = \\
&h_x h_y \left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i+1,j,k+1}) + u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i-1,j,k+1}) - u(\boldsymbol{s}_{i-1,j,k})}{4h_x} \right] + \\
&h_x h_y \left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i,j+1,k+1}) + u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j-1,k+1}) - u(\boldsymbol{s}_{i,j-1,k})}{4h_y} \right] + \\
&h_x h_y \left[\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k+1/2}) \frac{u(\boldsymbol{s}_{i,j,k+1}) - u(\boldsymbol{s}_{i,j,k})}{h_z} \right], \\
&\hat{W}^B_{i,j,k} = \\
&h_x h_y \left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i-1,j,k}) + u(\boldsymbol{s}_{i-1,j,k-1}) - u(\boldsymbol{s}_{i+1,j,k}) - u(\boldsymbol{s}_{i+1,j,k-1})}{4h_x} \right] + \\
&h_x h_y \left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i,j-1,k}) + u(\boldsymbol{s}_{i,j-1,k-1}) - u(\boldsymbol{s}_{i,j+1,k}) - u(\boldsymbol{s}_{i,j+1,k-1})}{4h_y} \right] + \\
&h_x h_y \left[\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k-1/2}) \frac{u(\boldsymbol{s}_{i,j,k-1}) - u(\boldsymbol{s}_{i,j,k})}{h_z} \right].
\end{aligned}
$$

Next, a vectorization of the discretization is made; first moving along the z-direction, then along x-direction, and lastly along the y-direction. Let us denote this with the common index $l = j \cdot M \cdot P + i \cdot P + k$ so $\boldsymbol{s}_{ijk} = \boldsymbol{s}_{j \cdot M \cdot P + i \cdot P + k} = \boldsymbol{s}_l$ which gives $u(\boldsymbol{s}_{ijk}) = u_l$ and $\kappa^2(\boldsymbol{s}_{ijk}) = \kappa^2_l$, and let the last index be $L = (N-1)MP + (M-1)P + P - 1$. Further, the vectorization results in the linear system of equations

$$
(\mathbf{D}_V \mathbf{D}_{\kappa^2} - \mathbf{A}_H)\boldsymbol{u} = \mathbf{D}_V^{1/2} \boldsymbol{z}, \tag{S10}
$$

where $\mathbf{D}_V = V \cdot \mathbf{I}_{MNP}$, $\mathbf{D}_{\kappa^2} = \left[\kappa_0^2, \ldots, \kappa_l^2, \ldots, \kappa_L^2\right] \mathbf{I}_{MNP}$, and $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{MNP})$. For simplicity the indices of the neighbors are denoted $k_p = k+1$, $k_n = k-1$, $j_p = j+1$, $j_n = j-1$, $i_p = i+1$, and

$i_n = i - 1$. The development of $\mathbf{A_H}$ is done by the sum $\hat{W}^L_{ijk} + \hat{W}^R_{ijk} + \hat{W}^B_{ijk} + \hat{W}^F_{ijk} + \hat{W}^U_{ijk} + \hat{W}^D_{ijk}$ and accounting for the index in u_{ijk} to form the linear relationship. In the following, non-zero elements of the $(jMN + iP + k)$-th row of $\mathbf{A_H}$ are formalized, and the index in $(\mathbf{A_H})\_$ denotes the column being assigned. The resulting coefficient with the point itself is

$$\begin{aligned}(\mathbf{A_H})_{j\cdot M\cdot P+i\cdot P+k} = &- \frac{h_y h_z}{h_x}\left[\mathbf{H}^{11}(\boldsymbol{s}_{i+1/2,j,k}) + \mathbf{H}^{11}(\boldsymbol{s}_{i-1/2,j,k})\right]\\ &- \frac{h_x h_z}{h_y}\left[\mathbf{H}^{22}(\boldsymbol{s}_{i,j+1/2,k}) + \mathbf{H}^{22}(\boldsymbol{s}_{i,j-1/2,k})\right]\\ &- \frac{h_x h_y}{h_z}\left[\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k+1/2}) + \mathbf{H}^{22}(\boldsymbol{s}_{i,j,k-1/2})\right],\end{aligned}$$

with the six closest neighbors are

$$\begin{aligned}(\mathbf{A}_H)_{j\cdot M\cdot P+i\cdot P+k_p} = &\frac{h_x h_y}{h_z}\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k+1/2})\\ &+ \frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i+1/2,j,k}) - \mathbf{H}^{31}(\boldsymbol{s}_{i-1/2,j,k})\right]\\ &+ \frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j+1/2,k}) - \mathbf{H}^{32}(\boldsymbol{s}_{i,j-1/2,k})\right]\end{aligned}$$

$$\begin{aligned}(\mathbf{A}_H)_{j\cdot M\cdot P+i\cdot P+k_n} = &\frac{h_x h_y}{h_z}\mathbf{H}^{33}(\boldsymbol{s}_{i,j,k-1/2})\\ &- \frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i+1/2,j,k}) - \mathbf{H}^{31}(\boldsymbol{s}_{i-1/2,j,k})\right]\\ &- \frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j+1/2,k}) - \mathbf{H}^{32}(\boldsymbol{s}_{i,j-1/2,k})\right]\end{aligned}$$

$$\begin{aligned}(\mathbf{A}_H)_{j\cdot M\cdot P+i_p\cdot P+k} = &\frac{h_z h_y}{h_x}\mathbf{H}^{11}(\boldsymbol{s}_{i+1/2,j,k})\\ &+ \frac{h_y}{4}\left[\mathbf{H}^{12}(\boldsymbol{s}_{i,j,k+1/2}) - \mathbf{H}^{12}(\boldsymbol{s}_{i,j,k-1/2})\right]\\ &+ \frac{h_z}{4}\left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j+1/2,k}) - \mathbf{H}^{13}(\boldsymbol{s}_{i,j-1/2,k})\right]\end{aligned}$$

$$\begin{aligned}(\mathbf{A}_H)_{j\cdot M\cdot P+i_n\cdot P+k} = &\frac{h_z h_y}{h_x}\mathbf{H}^{11}(\boldsymbol{s}_{i-1/2,j,k})\\ &- \frac{h_y}{4}\left[\mathbf{H}^{12}(\boldsymbol{s}_{i,j,k+1/2}) - \mathbf{H}^{12}(\boldsymbol{s}_{i,j,k-1/2})\right]\\ &- \frac{h_z}{4}\left[\mathbf{H}^{13}(\boldsymbol{s}_{i,j+1/2,k}) - \mathbf{H}^{13}(\boldsymbol{s}_{i,j-1/2,k})\right]\end{aligned}$$

$$\begin{aligned}(\mathbf{A}_H)_{j_p\cdot M\cdot P+i\cdot P+k} = &\frac{h_x h_z}{h_y}\mathbf{H}^{22}(\boldsymbol{s}_{i,j+1/2,k})\\ &+ \frac{h_x}{4}\left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2}) - \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2})\right]\\ &+ \frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i+1/2,j,k}) - \mathbf{H}^{21}(\boldsymbol{s}_{i-1/2,j,k})\right]\end{aligned}$$

$$
\begin{aligned}
(\mathbf{A}_H)_{j_n \cdot M \cdot P + i \cdot P + k} = & \frac{h_x h_z}{h_y} \mathbf{H}^{22}(\boldsymbol{s}_{i,j-1/2,k}) \\
& - \frac{h_x}{4}\left[\mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2}) - \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2})\right] \\
& - \frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i+1/2,j,k}) - \mathbf{H}^{21}(\boldsymbol{s}_{i-1/2,j,k})\right],
\end{aligned}
$$

and with the twelve closest diagonals are

$$
\begin{aligned}
(\mathbf{A}_H)_{j \cdot M \cdot P + i_p \cdot P + k_p} &= \frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i+1/2,j,k}) + \mathbf{H}^{13}(\boldsymbol{s}_{i,j,k+1/2})\right], \\
(\mathbf{A}_H)_{j \cdot M \cdot P + i_n \cdot P + k_n} &= \frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i-1/2,j,k}) + \mathbf{H}^{13}(\boldsymbol{s}_{i,j,k-1/2})\right], \\
(\mathbf{A}_H)_{j \cdot M \cdot P + i_n \cdot P + k_p} &= -\frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i-1/2,j,k}) + \mathbf{H}^{13}(\boldsymbol{s}_{i,j,k+1/2})\right] \\
(\mathbf{A}_H)_{j \cdot M \cdot P + i_p \cdot P + k_n} &= -\frac{h_y}{4}\left[\mathbf{H}^{31}(\boldsymbol{s}_{i+1/2,j,k}) + \mathbf{H}^{13}(\boldsymbol{s}_{i,j,k-1/2})\right], \\
(\mathbf{A}_H)_{j_p \cdot M \cdot P + i \cdot P + k_p} &= \frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j+1/2,k}) + \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2})\right], \\
(\mathbf{A}_H)_{j_n \cdot M \cdot P + i \cdot P + k_n} &= \frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j-1/2,k}) + \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2})\right], \\
(\mathbf{A}_H)_{j_n \cdot M \cdot P + i \cdot P + k_p} &= -\frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j-1/2,k}) + \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k+1/2})\right], \\
(\mathbf{A}_H)_{j_p \cdot M \cdot P + i \cdot P + k_n} &= -\frac{h_x}{4}\left[\mathbf{H}^{32}(\boldsymbol{s}_{i,j+1/2,k}) + \mathbf{H}^{23}(\boldsymbol{s}_{i,j,k-1/2})\right], \\
(\mathbf{A}_H)_{j_p \cdot M \cdot P + i_p \cdot P + k} &= \frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i+1/2,j,k}) + \mathbf{H}^{12}(\boldsymbol{s}_{i,j+1/2,k})\right], \\
(\mathbf{A}_H)_{j_n \cdot M \cdot P + i_n \cdot P + k} &= \frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i-1/2,j,k}) + \mathbf{H}^{12}(\boldsymbol{s}_{i,j-1/2,k})\right], \\
(\mathbf{A}_H)_{j_n \cdot M \cdot P + i_p \cdot P + k} &= -\frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i+1/2,j,k}) + \mathbf{H}^{12}(\boldsymbol{s}_{i,j-1/2,k})\right], \\
(\mathbf{A}_H)_{j_p \cdot M \cdot P + i_n \cdot P + k} &= -\frac{h_z}{4}\left[\mathbf{H}^{21}(\boldsymbol{s}_{i-1/2,j,k}) + \mathbf{H}^{12}(\boldsymbol{s}_{i,j+1/2,k})\right].
\end{aligned}
$$

Note that the corner points are not included in this scheme. Denoting $\mathbf{A} = \mathbf{D}_V \mathbf{D}_{k^2} - \mathbf{A}_{\mathbf{H}}$, Eq. (S10) can be written as

$$
\boldsymbol{z} = \mathbf{D}_V^{-1/2} \mathbf{A} \boldsymbol{u},
$$

and thus, the joint distribution of $\boldsymbol{u}$ is

$$
\begin{aligned}
\pi(\boldsymbol{u}) &\propto \pi(\boldsymbol{z}) \propto \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\right) \\
\pi(\boldsymbol{u}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{D}_V^{-1}\mathbf{A}\boldsymbol{u}\right) \\
\pi(\boldsymbol{u}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}\boldsymbol{u}\right).
\end{aligned}
$$

Here, $\mathbf{Q} = \mathbf{A}^{\mathrm{T}}\mathbf{D}_V^{-1}\mathbf{A}$ which is a sparse matrix of 93 non-zero elements per row. This corresponds to the point, the 18 closest neighbors, and their 18 closest neighbors. Then removing duplicates results in 93 points.

Table S2
The Relative Root Mean Square Error (RRMSE) of parameter estimates for SI, SA, and NA from 100 repetitions for each data setting, which varies the number of observed locations (No. loc.) and the number of replicated observations of these locations (No. real.).

No. loc.		100			10000			27000		
No. real.		1	10	100	1	10	100	1	10	100
Stat. Iso.	$\log\kappa$	0.473	**0.105**	**0.034**	**0.091**					
	$\log\gamma$	0.869	**0.175**	**0.071**	**0.048**					
	$\log\tau$	0.546	**0.141**	**0.041**	**0.015**					
Stationary Anisotropic	$\log\kappa$	0.946	0.176	**0.089**	**0.092**	**0.036**				
	$\log\gamma$	10.940	62.373	**0.655**	**0.315**	**0.115**				
	$\lvert v_x\rvert$	0.693	0.512	**0.247**	**0.100**	**0.037**				
	$\lvert v_y\rvert$	0.894	0.522	**0.330**	**0.103**	**0.026**				
	$\lvert v_z\rvert$	2.957	1.428	**0.641**	**0.192**	**0.070**				
	$\lvert\rho_1\rvert$	0.793	0.616	**0.188**	**0.085**	**0.027**				
	$\lvert\rho_2\rvert$	2.354	1.016	**0.495**	**0.123**	**0.047**				
	$\log\tau$	0.394	0.272	**0.042**	**0.037**	**0.007**				
Non-Stationary Anisotropic	$\log\kappa$						2.119	0.698	**0.266**	**0.198**
	$\log\gamma$						2.901	1.359	**0.617**	**0.542**
	$\lvert v_x\rvert$						1.292	0.268	**0.182**	**0.178**
	$\lvert v_y\rvert$						1.573	0.244	**0.144**	**0.141**
	$\lvert v_z\rvert$						0.585	0.210	**0.103**	**0.088**
	$\lvert\rho_1\rvert$						2.211	7.970	**3.361**	**2.741**
	$\lvert\rho_2\rvert$						1.071	1.884	**1.194**	**1.038**
	$\log\tau$						0.249	0.003	**0.001**	**0.001**

Appendix C. Additional figures and tables

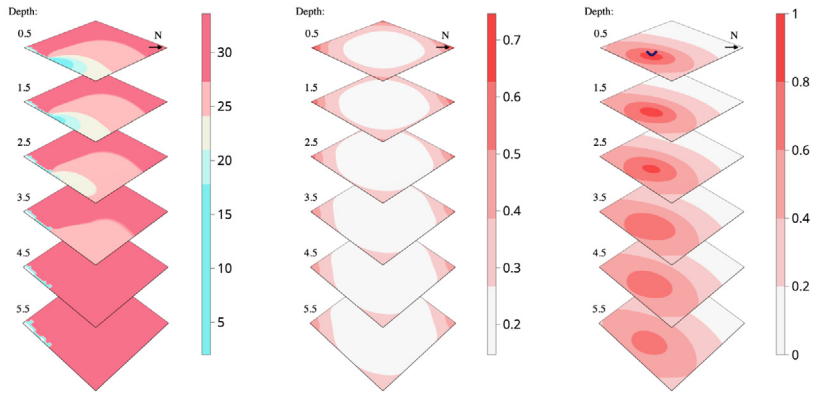
In Section 4, we ran a simulation study investigating the amount of data required to estimate models of different complexity. The results of this simulation study were presented in Table 1 which showed the root mean square errors (RMSEs) of the parameter estimates for different data settings. The parameters can be at different scales so the RMSE might not be sufficient for the interpretation of the results. Therefore, Table S2 is included and shows the relative root mean square error (RRMSE) of the parameters.

For the non-stationary model, the relative errors are not so easily available as some of its parameters are zero. To account for this, the NA model Table S2 shows the average RMSE of the parameters for each spline function relative to the average parameter value of each spline function and not the average RRMSE of the parameters for each spline function. To clarify the RRMSE for the stationary models are

$$\text{RRMSE}(\hat{\theta}_i) = \text{RMSE}(\hat{\theta}_i) \Big/ |\theta_i|,$$

where $\hat{\theta}_i$ is the estimated ith element of $\boldsymbol{\theta} = (\log\kappa^2, \log\gamma, \ldots, \log\sigma_N)$. Then for the non-stationary case, we have

$$\begin{aligned}\overline{\text{RRMSE}(\hat{\boldsymbol{\alpha}}_g)} &= \overline{\text{RMSE}(\hat{\boldsymbol{\alpha}}_g)} \Big/ \overline{|\boldsymbol{\alpha}_g|} \\ &= \sum_{j=1}^{p} \text{RMSE}(\alpha_{g_j}) \Big/ \sum_{j=1}^{p} |\alpha_{g_j}|.\end{aligned}$$



(a) SINMOD prior (b) Marginal Variance (c) Correlation

Fig. S3. Prior field **(a)** found from SINMOD simulations, the variance of the spatial effect **(b)** and spatial correlation of point [22,10,0] **(c)** in the stationary anisotropic model. The N-arrow shows the cardinal north.

Here, $\boldsymbol{\alpha}_{\boldsymbol{g}}$ denotes the p-dimensional vector of parameters for an arbitrary spline function $g(\cdot)$ in the set of all parameters for all the spline functions

$$\boldsymbol{\theta} = (\boldsymbol{\alpha}_{\kappa}, \boldsymbol{\alpha}_{\gamma}, \boldsymbol{\alpha}_{v_x}, \boldsymbol{\alpha}_{v_y}, \boldsymbol{\alpha}_{v_z}, \boldsymbol{\alpha}_{\rho_1}, \boldsymbol{\alpha}_{\rho_2}),$$

and j specifies the index of the parameter from that arbitrary spline function.

In the application, Section 5, we estimate the parameters of a non-stationary anisotropic and stationary anisotropic model on a simulated dataset from the numerical ocean model SINMOD. The resulting properties of the non-stationary model are presented in Fig. 7 in Section 5.2 since this is the main focus of the applications. The properties of the stationary anisotropic model fit on the same dataset are presented in Fig. S3. The marginal variance in Fig. S3(b), which should be constant for this stationary model, shows some variability caused by the boundary conditions. Notice that this boundary effect is also bigger in the direction of the strongest dependency directions seen in the south and north corners. Notice also the large discrepancies between the correlations in these two models, Fig. S3(c) and Fig. 7c, as the stationary anisotropic model kind of captures an average correlation within the field.

In Fig. 9 in Section 5, we observed an increase given more data in the standard deviation of the root mean square error of predictions under random permutation after about 70% of observed data. We noted that this was due to our assumption that the model is purely spatial given a small operation time, and that the last segment in the field test (the up and down pattern) did not fit well within this assumption. Fig. S4 shows the same results as Fig. 9 but with the last segment removed. Here, we observe a slight reduction of this effect but admit that there is still an increase. This is because all or most of the held-out data comes from the same segment.

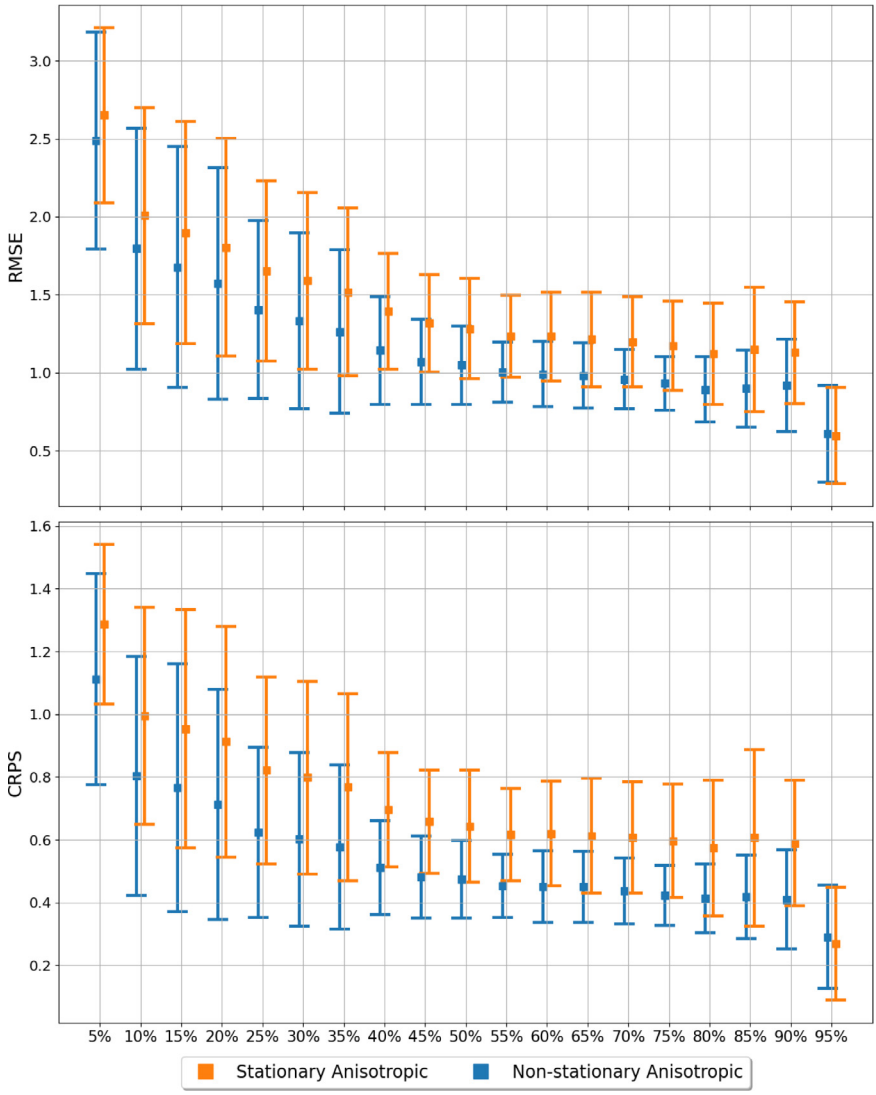


Fig. S4. The root mean square error (RMSE, top) and the continuous ranked probability score (CRPS, bottom) of predictions from the stationary anisotropic (orange) and non-stationary anisotropic models (blue) given different proportions of observed data (5%, 95%). The error bars are the standard deviations of the different measures under random permutations of 8 segments. Here, one segment is removed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Bolin, D., Kirchner, K., 2020. The rational SPDE approach for Gaussian random fields with general smoothness. J. Comput. Graph. Statist. 29 (2), 274–285.

Bolin, D., Kirchner, K., 2023. Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. Bernoulli 29 (2), 1476–1504.

Bolin, D., Wallin, J., 2019. Multivariate type G Matérn stochastic partial differential equation random fields. J. R. Stat. Soc. Ser. B Stat. Methodol. 82 (1), 215–239, URL https://doi.org/10.1111/rssb.12351.

Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A., Hammerling, D., 2019. Reproducing internal variability with few ensemble runs. J. Clim. 32 (24), 8511–8522.
Cressie, N., Wikle, C.K., 2015. Statistics for Spatio-Temporal Data. John Wiley & Sons.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C. Appl. Stat. 47 (3), 299–350.
Euler, L., 1771. Problema algebraicum ob affectiones prorsus singulares memorabile. Novi Commentarii Academiae Scientiarum Petropolitanae 75–106, URL https://scholarlycommons.pacific.edu/euler-works/407.
Foss, K.H., Berget, G.E., Eidsvik, J., 2022. Using an autonomous underwater vehicle with onboard stochastic advection-diffusion models to map excursion sets of environmental variables. Environmetrics 33 (1), e2702.
Fossum, T.O., Fragoso, G.M., Davies, E.J., Ullgren, J.E., Mendes, R., Johnsen, G., Ellingsen, I., Eidsvik, J., Ludvigsen, M., Rajan, K., 2019. Toward adaptive robotic sampling of phytoplankton in the coastal ocean. Science Robotics 4 (27), eaav3041.
Fossum, T.O., Travelletti, C., Eidsvik, J., Ginsbourger, D., Rajan, K., 2021. Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling. Ann. Appl. Stat. 15 (2), 597–618, URL https://doi.org/10.1214/21-AOAS1451.
Fuglstad, G.-A., Castruccio, S., 2020. Compression of climate simulations with a nonstationary global SpatioTemporal SPDE model. Ann. Appl. Stat. 14 (2), 542–559.
Fuglstad, G.-A., Lindgren, F., Simpson, D., Rue, H., 2015a. Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Statist. Sinica 115–133.
Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2015b. Does non-stationary spatial data always require non-stationary random fields? Spatial Stat. 14, 505–531.
Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), URL https://doi.org/10.1198/016214506000001437.
Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W., Sun, F., Zammit-Mangion, A., 2019. A case study competition among methods for analyzing large spatial data. J. Agric. Biol. Environ. Stat. 24, 398–425.
Hildeman, A., Bolin, D., Rychlik, I., 2021. Deformed SPDE models with an application to spatial modeling of significant wave height. Spatial Stat. 42, 100449.
Hu, W., Fuglstad, G.-A., Castruccio, S., 2022. A stochastic locally diffusive model with neural network-based deformations for global sea surface temperature. Stat 11.
Ingebrigtsen, R., Lindgren, F., Steinsland, I., 2014. Spatial models with explanatory variables in the dependence structure. Spatial Stat. 8, 20–38.
Ingebrigtsen, R., Lindgren, F., Steinsland, I., Martino, S., 2015. Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. Spatial Stat. 14, 338–364.
Lee, D., Gammie, C.F., 2021. Disks as inhomogeneous, anisotropic Gaussian random fields. Astrophys. J. 906 (1), 39.
Lindgren, F., Bolin, D., Rue, H., 2022. The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. Spatial Stat. 50, 100599, Special Issue: The Impact of Spatial Statistics URL https://www.sciencedirect.com/science/article/pii/S2211675322000057.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.
Neto, J.H.V., Schmidt, A.M., Guttorp, P., 2014. Accounting for spatially varying directional effects in spatial covariance structures. J. R. Stat. Soc. Ser. C. Appl. Stat. 103–122.
Paciorek, C.J., Schervish, M.J., 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17 (5), 483–506.
Risser, M.D., Calder, C.A., 2015. Regression-based covariance functions for nonstationary spatial modeling. Environmetrics 26 (4), 284–297.
Rodrigues, 1840. Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. J. Math. Pures Appl. 380–440, URL http://eudml.org/doc/234443.
Rue, H., Held, L., 2010. Markov Random Fields. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (Eds.), Chapman & Hall/CRC Handbooks of Modern Statistical Methods. pp. 171–200, URL http://www.crcpress.com/product/isbn/9781420072877.
Salvaña, M.L.O., Genton, M.G., 2021. Lagrangian spatio-temporal nonstationary covariance functions. In: Advances in Contemporary Statistics and Econometrics. Springer, pp. 427–447.
Sampson, P.D., 2010. Constructions for nonstationary spatial processes. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (Eds.), Handbook of Spatial Statistics. CRC Press, Boca Rotan, FL, pp. 119–130.
Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure. J. Amer. Statist. Assoc. 87 (417), 108–119.
Schmidt, A.M., Guttorp, P., O'Hagan, A., 2011. Considering covariates in the covariance structure of spatial processes. Environmetrics 22 (4), 487–500.
Sidén, P., Lindgren, F., Bolin, D., Eklund, A., Villani, M., 2021. Spatial 3D Matérn priors for fast whole-brain fMRI analysis. Bayesian Anal. 16 (4), 1251–1278.
Slagstad, D., McClimans, T.A., 2005. Modeling the ecosystem dynamics of the Barents sea including the marginal ice zone: I. Physical and chemical oceanography. J. Mar. Syst. 58 (1), 1–18, URL https://www.sciencedirect.com/science/article/pii/S0924796305001296.
Stein, M.L., 2002. The screening effect in Kriging. Ann. Statist. 30 (1), 298–323.

Stein, M.L., 2012. Interpolation of Spatial Data: Some Theory for Kriging. Springer Science & Business Media.
Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99 (465), 250–261.
Zhang, J., Bonas, M., Bolster, D., Fuglstad, G.-A., Castruccio, S., 2023. High resolution global precipitation downscaling with latent Gaussian models and nonstationary SPDE structure. arXiv preprint arXiv:2302.03148.
Zhang, R., Czado, C., Sigloch, K., 2016. Bayesian spatial modelling for high dimensional seismic inverse problems. J. R. Stat. Soc. Ser. C. Appl. Stat. 65 (2), 187–213, URL http://www.jstor.org/stable/24772417.