Supplementary Material

1 DESCRIBING GMRFS THROUGH SPDES

Whittle (1954) showed that the solution, $u(\boldsymbol{s})$, of the stochastic partial differential equation (SPDE)

$$(\kappa(\boldsymbol{s})^2 - \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^3, \tag{S1}$$

is a Gaussian Markov random field (GMRF) with a Matérn covariance function. The approach was popularized by Lindgren et al. (2011); then, extended to non-stationary and anisotropic 2D fields in Fuglstad et al. (2015a,b), and recently to 3D fields by Berild and Fuglstad (2023).

In Equation (S1), $\mathcal{W}(\boldsymbol{s})$ is Gaussian white noise, while κ is a parameter controlling both variance and range of the GMRF. The component $\mathbf{H}$ is also used to regulate the variance and the range, but more importantly, it is controlling the anisotropy of the Laplacian, $\nabla \cdot \mathbf{H}\nabla$, and thus the anisotropy of the resulting field. With this anisotropy, the model can account for varying properties depending on directions. Also, note that $\kappa(\boldsymbol{s})$ and $\mathbf{H}(\boldsymbol{s})$ in Equation (S1) are allowed to vary through space.

In the following, we will describe the parametrization of the anisotropy (Section 1.1), the non-stationarity (Section 1.2), properties of the GMRF described through the SPDE (Section 1.3), and how we infer parameters from data (Section 1.4).

1.1 Parametrizing Anisotropy

The spatially and directionally varying covariances are described by controlling the eigenvalues and eigenvectors of the matrix $\mathbf{H}$ in the anisotropic Laplacian. Berild and Fuglstad (2023) proposed the following parsimonious and interpretable parameterization

$$\mathbf{H}(\boldsymbol{s}) = \gamma(\boldsymbol{s})\mathbf{I}_3 + \boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}} + \boldsymbol{\omega}(\boldsymbol{s})\boldsymbol{\omega}(\boldsymbol{s})^{\mathrm{T}}, \tag{S2}$$

where $\gamma(\boldsymbol{s}) > 0$, $\boldsymbol{v} = (v_x, v_y, v_z)^{\mathrm{T}} \in \mathbb{R}^3$ and $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^{\mathrm{T}} \in \mathbb{R}^3$ whereby $\boldsymbol{v} \perp \boldsymbol{\omega}$. Thus, the eigenvalues are $\lambda_1 = \gamma$, $\lambda_2 = \gamma + \|\boldsymbol{v}\|^2$, and $\lambda_3 = \gamma + \|\boldsymbol{\omega}\|^2$ with eigenvectors $\boldsymbol{v}_1 = \boldsymbol{v} \times \boldsymbol{\omega}$, $\boldsymbol{v}_2 = \boldsymbol{v}$, and $\boldsymbol{v}_3 = \boldsymbol{\omega}$.

The vector $\boldsymbol{v}$ is simply parametrized with its Cartesian components, v_x, v_y, and v_z. Further, $\boldsymbol{\omega}$ is parametrized by two scalars, ρ_1 and ρ_2, controlling the linear combination of two orthogonal vector, $\boldsymbol{\omega}_1 = (-v_y, v_x, 0)^{\mathrm{T}}$ and $\boldsymbol{\omega}_2 = \boldsymbol{v} \times \boldsymbol{\omega}_1$, in the plane with $\boldsymbol{v}$ as normal vector such that

$$\boldsymbol{\omega} = \rho_1 \frac{\boldsymbol{\omega}_1}{\|\boldsymbol{\omega}_1\|} + \rho_2 \frac{\boldsymbol{\omega}_2}{\|\boldsymbol{\omega}_2\|}. \tag{S3}$$

1.2 Parametrizing non-stationarity

The non-stationarity is obtained by allowing the parameters κ, γ, v_x, v_y, v_z, ρ_1 and ρ_2 to vary throughout space. This is achieved by describing these parameters as spline functions:

$$g(\boldsymbol{s}) = (\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\alpha}_{\mathbf{g}}. \tag{S4}$$

Here, $\boldsymbol{\alpha}_g \in \mathbb{R}^p$ is a vector of weights or the *new* parameters for the spline function $g(\cdot)$ or the *old* parameter, and $(\boldsymbol{s}) = (\mathbf{f_1}(\boldsymbol{s}), \ldots, \mathbf{f_p}(\boldsymbol{s}))^{\mathrm{T}}$ is a p-dimensional vector of B-spline basis functions evaluated at location $\boldsymbol{s}$. These basis splines are constructed as a tensor product of three clamped 1D second-order B-splines in each dimension as

$$f_{ijk}(\boldsymbol{s}) = B_{x,i}(x) \cdot B_{y,j}(y) \cdot B_{z,k}(z), \quad \boldsymbol{s} = (x, y, z)^{\mathrm{T}} \in \mathcal{D}, \tag{S5}$$

where $B_{x,i}$ is the $i-th$ component in the x-direction and similarly for the other directions. Thereby, the number of parameters for each spline function $g(\cdot)$ is $p = 27$ which in total is 189 parameters or B-spline weights for all spline functions in our model. We collect them as

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

1.3 General properties

As mentioned in Section 1.4 the parameters controlling the covariance structure in the spatial effect are learned from the residuals of an autoregressive model of order one fit to the SINMOD dataset $\boldsymbol{y}$. This gives the underlying process

$$\boldsymbol{x}_{\mathrm{R}}|\boldsymbol{\theta} \sim \mathcal{N}_n\left(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta})\right), \tag{S6}$$

with zero mean and the inverse covariance matrix $\mathbf{Q}(\boldsymbol{\theta})$. In this case $\boldsymbol{x}_R = \boldsymbol{u}$, where $\boldsymbol{u}$ is the random variable specified through the SPDE.

The marginal variance for the solution $\boldsymbol{u}$ of the SPDE, can be derived through a series of calculations. The transfer function for Equation (S1) is $g(w) = (\kappa^2 + \boldsymbol{w}^{\mathrm{T}}\mathbf{H}\boldsymbol{w})^{-1}$ and given the spectral density of Gaussian white noise in $\mathbb{R}^3$ is $(2\pi)^{-3}$ the spectral density of the solution of the SPDE is

$$f_{\mathrm{S}}(\boldsymbol{w}) = (2\pi)^{-3}(\kappa^2 + \boldsymbol{w}^{\mathrm{T}}\mathbf{H}\boldsymbol{w})^{-1}. \tag{S7}$$

Lastly, the variance of the solution is found by integrating its spectral density over the whole domain $\mathbb{R}^3$, and thus, the variance of the process we are trying to describe is

$$\mathrm{Var}\left(x(\boldsymbol{s})\right) = \frac{1}{8\pi\kappa(\boldsymbol{s})\sqrt{\det(\mathbf{H}(\boldsymbol{s}))}}. \tag{S8}$$

Moreover, in a stationary case where $\kappa(\boldsymbol{s}) = \kappa$ and $\mathbf{H}(\boldsymbol{s}) = \mathbf{H}$ for all $\boldsymbol{s}$ the covariance of the process between two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ in $\mathbb{R}^3$ can be written as

$$\mathrm{Cov}\left(x(\boldsymbol{s}_1), x(\boldsymbol{s}_2)\right) = \frac{1}{8\pi\kappa\sqrt{\det(\mathbf{H})}} \exp\left(-\kappa||\mathbf{H}^{-1/2}(\boldsymbol{s}_1 - \boldsymbol{s}_2)||\right), \tag{S9}$$

or namely an exponential covariance function.

1.4 Parameter inference

Following the notation from the process defined in Equation (S6) the data model for the innovations are

$$\boldsymbol{y}_R|\boldsymbol{x}_R, \sigma^2_{\mathrm{S}} \sim \mathcal{N}_m\left(\mathbf{A}\boldsymbol{x}_R, \sigma^2_{\mathrm{S}}\mathbf{I}_n\right). \tag{S10}$$

Here, matrix $\mathbf{A}$ is a m by n matrix linking the locations of the observation $\boldsymbol{y}_R$ to our discretization in the process $\boldsymbol{x}_R$, and σ_{S}^2 is the independent noise in the innovations.

To find the optimal parameters such that our models best describe the innovations we will maximize the likelihood function of the parameters. Specifically, the parameters have a multivariate Gaussian distribution, and following common practice we will optimize the logarithmic transformation of this likelihood, thereby the log-likelihood:

$$\begin{aligned}\ell(\boldsymbol{\theta}, \sigma_{\mathrm{S}}^2|\boldsymbol{y}) = \text{Const} + \log \pi(\boldsymbol{\theta}, \sigma_{\mathrm{S}}^2) + \frac{1}{2}\log\det(\mathbf{Q}) - \frac{m}{2}\log(\sigma_{\mathrm{S}}^2) \\ - \frac{1}{2}\log\det(\mathbf{Q}_{\mathrm{C}}) - \frac{1}{2}\boldsymbol{\mu}_{\mathrm{C}}^{\mathrm{T}}\mathbf{Q}_{\mathrm{C}}\boldsymbol{\mu}_{\mathrm{C}} - \frac{1}{2\sigma_{\mathrm{S}}^2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu}_{\mathrm{C}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu}_{\mathrm{C}}).\end{aligned} \tag{S11}$$

The reader is referred to the supplementary material of Berild and Fuglstad (2023) for a full derivation of the log-likelihood. In Equation (S11), $\mathbf{Q}_{\mathrm{C}}$ is the conditional precision matrix, i.e. the precision matrix given that the model has seen $\boldsymbol{y}$,

$$\mathbf{Q}_{\mathrm{C}} = \mathbf{Q} + \mathbf{A}^T\mathbf{A}/\sigma_{\mathrm{S}}^2. \tag{S12}$$

Similarly, the conditional mean is

$$\boldsymbol{\mu}_{\mathrm{C}} = \mathbf{Q}_{\mathrm{C}}^{-1}\mathbf{A}^T\mathbf{A}\boldsymbol{y}/\sigma_{\mathrm{S}}^2. \tag{S13}$$

Note that this is similar to the equations used in the model updating, but with $\boldsymbol{\mu} = 0$ as assumed by our process of these innovations.

The parameter space is quite challenging to explore so we use an analytical expression for the gradient to determine the search directions. In order to speed up the gradient calculations a stochastic version is calculated, and therefore, ultimately the optimization strategy is a stochastic gradient descent algorithm. Furthermore, we have employed a root mean square propagation (RMSprop) in the optimization to improve the stability and convergence.

2 EIBV DESIGN CRITERIA

For any design $\boldsymbol{d}$ at stage t giving data $\boldsymbol{y} = \boldsymbol{y}_{\boldsymbol{d}}$, the EIBV reduction in can be rephrased as

$$\begin{aligned}\mathrm{EIBV}(\boldsymbol{\mu}_{\mathrm{C},t-1}, \mathbf{Q}_{\mathrm{C},t-1}) &= \int E_{\boldsymbol{y}|\mathcal{Y}_{t-1}}\left\{p_{\boldsymbol{s}}(\boldsymbol{y}, \mathcal{Y}_{t-1})[1 - p_{\boldsymbol{s}}(\boldsymbol{y}, \mathcal{Y}_{t-1})]\right\} d\boldsymbol{s}, \\ p_{\boldsymbol{s}}(\boldsymbol{y}, \mathcal{Y}_{t-1}) &= P(\boldsymbol{x}(\boldsymbol{s}) \leq \ell|\boldsymbol{y}, \mathcal{Y}_{t-1}),\end{aligned} \tag{S14}$$

where $\mathcal{Y}_{t-1}$ denotes all the data gathered at stages before t. The probability $p_{\boldsymbol{s}}(\boldsymbol{y}, \mathcal{Y}_{t-1})$ is a Gaussian cumulative distribution function (CDF) with linear conditioning to $\boldsymbol{y}$ in the mean and with a variance that does not depend on the outcome of the data.

The conditional mean at stage $t-1$ is $\boldsymbol{\mu}_{\mathrm{C},t-1}$ with entries $\mu_{\mathrm{C},t-1}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$. The conditional covariance matrix is $\boldsymbol{\Sigma}_{\mathrm{C},t-1} = \mathbf{Q}_{\mathrm{C},t-1}^{-1}$ with diagonal entries $\sigma_{\mathrm{C},t-1}^2(\boldsymbol{s}_i)$, and after the updating we have $\boldsymbol{\Sigma}_{\mathrm{C},t} = \mathbf{Q}_{\mathrm{C},t}^{-1}$ with diagonal entries $\sigma_{\mathrm{C},t}^2(\boldsymbol{s}_i)$. Based on results of Chevalier et al. (2014) and Fossum et al. (2021), the EIBV in Equation (S14) can be evaluated in closed form as a bivariate Gaussian CDF depending

on these parameters.

$$\begin{aligned} \mathrm{EIBV}(\boldsymbol{\mu}_{\mathrm{C},t-1}, \mathbf{Q}_{\mathrm{C},t-1}) &= \sum_{i=1}^{n} \mathrm{EBV}(\mu_{\mathrm{C},t-1}(\boldsymbol{s}_i), \sigma^2_{\mathrm{C},t-1}(\boldsymbol{s}_i), \sigma^2_{\mathrm{C},t}(\boldsymbol{s}_i)) \\ \mathrm{EBV}(\eta, v^2, w^2) &= \Phi_2\left(\begin{bmatrix} \ell \\ -\ell \end{bmatrix}; \begin{bmatrix} \eta \\ -\eta \end{bmatrix}, \begin{bmatrix} v^2 & (w^2 - v^2) \\ (w^2 - v^2) & v^2 \end{bmatrix}\right), \end{aligned} \tag{S15}$$

where Φ_2 denotes the bivariate Gaussian cumulative distribution function.

We note that this closed form EIBV calculation in equation (S15) relies on the variance terms and their reduction in the updating step ($w^2 = \sigma^2_{\mathrm{C},t}(\boldsymbol{s}_i)$ compared with $v^2 = \sigma^2_{\mathrm{C},t-1}(\boldsymbol{s}_i)$). In the formulation with the precision matrix, these terms are not immediately available. Matrix recursions exist for computing the marginal variance terms from the precision matrix, see e.g. Zammit-Mangion and Rougier (2018), but the solution can be computationally challenging in 3D because of rather large fill-in of non-zeros in the sparse matrix structure during the recursion. We instead approximate the required variance terms by Monte Carlo sampling from the GMRF model.

Conditional samples are here generated by a trick known as conditioning by Kriging equation (see e.g. Wackernagel (2003)). This relies on the following steps; first an unconditional sample $\boldsymbol{x}^b_{\mathrm{C},t-1}$ of the field is generated. In our case this comes from the Gaussian distribution with mean $\boldsymbol{\mu}_{\mathrm{C},t-1}$ and precision matrix $\mathbf{Q}_{\mathrm{C},t-1}$. Next, a synthetic data sample $\boldsymbol{y}^b = \mathbf{A}\boldsymbol{x}^b_{\mathrm{C},t-1} + \boldsymbol{\epsilon}^b$, $\boldsymbol{\epsilon}^b \sim N(0, \sigma^2_{\mathrm{auv}}\boldsymbol{I}_m)$ is generated according to the specified design. Finally, a conditional sample is formed by solving the equation for the conditional mean (also known as the Kriging equation), given the synthetic data:

$$\boldsymbol{x}^b_{\mathrm{C},t} = \boldsymbol{x}^b_{\mathrm{C},t-1} + \mathbf{Q}^{-1}_{\mathrm{C},t-1}\mathbf{A}^{\mathrm{T}}\left(\boldsymbol{y}^b - \mathbf{A}\boldsymbol{x}^b_{\mathrm{C},t-1}\right)/\sigma^2_{\mathrm{auv}}. \tag{S16}$$

This procedure is repeated for $b = 1, \ldots, B$ independent Monte Carlo samples. We used $B = 100$ in our implementation. Equation (S16) requires matrix-vector solves with the sparse precision matrix which can be done very fast.

When the optimal design is selected, the AUV acts to move in the direction of the selected design. It senses salinity data $\boldsymbol{y}$, and then Equation (S16) is used with these in-situ observations in place of $\boldsymbol{y}^b$ to get the conditional samples for time stage t. The resulting Monte Carlo sample $\boldsymbol{x}^b_{\mathrm{C},t}$, $b = 1 \ldots, B$ forms the basis for the EIBV evaluation at the next time step, when $t \to t + 1$.

REFERENCES

Berild, M. O. and Fuglstad, G.-A. (2023). Spatially varying anisotropy for gaussian random fields in three-dimensional space. *Spatial Statistics* 55, 100750

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics* 56, 455–465

Fossum, T. O., Travelletti, C., Eidsvik, J., Ginsbourger, D., and Rajan, K. (2021). Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling. *The Annals of Applied Statistics* 15, 597 – 618. doi:10.1214/21-AOAS1451

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015a). Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica* 25, 115–133

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015b). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics* 14, 505–531

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

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications* (Springer Science & Business Media)

Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika* 41, 434–449. doi:10.2307/2332724

Zammit-Mangion, A. and Rougier, J. (2018). A sparse linear algebra algorithm for fast computation of prediction variances with gaussian markov random fields. *Computational Statistics & Data Analysis* 123, 116–130