

Elling Svee

# Spatial Modelling Using Fractional SPDE Models with Spatially Varying Local Anisotropy

Project in Applied Physics and Mathematics  
Supervisor: Geir-Arne Fuglstad  
2024

Norwegian University of Science and Technology  
Faculty of Information Technology and Electrical Engineering  
Department of Mathematical Sciences





---

## Abstract

---

Over the last decade, the SPDE approach has become widely adopted in spatial statistics due to its computational benefits. The approach also naturally extends to approximating non-stationary Gaussian random fields (GRFs) with a Matérn-like covariance function, which can be useful in scenarios where properties are spatially varying. Originally, the smoothness of the GRFs had to be fixed to specific values, but recent advancements have extended the approach to any smoothness, giving so-called fractional SPDE models. This thesis introduces a fractional SPDE model incorporating non-stationarity in range, anisotropy and marginal variance. The model can approximate a wide range of GRFs, and enables computationally efficient inference. Spatial anisotropy is included through an identifiable parameterization, improving parameter estimation. Extending the model to support every smoothness makes it harder to determine a gradient for the likelihood function, which we address by leveraging PyTorch for automatic differentiation.

The parameter estimation of the proposed model is evaluated through a simulation study. Comparing against simpler model using data generated by a selected fractional and non-stationary model, we find the proposed model to have a lower estimated bias under repeated simulations. This lower bias is evident with as few as 250 observations, and when the signal-to-noise ratio is as high as 25%. The proposed model also avoids overfitting when estimating parameters on data from a non-fractional or stationary model. This makes it well-suited for applications where the true nature of the underlying field is uncertain.

Predictions are assessed using the root-mean-square error (RMSE) and continuously ranked probability scores (CRPS). In the simulation study, using data from a fractional and non-stationary model, the proposed model achieves a 8.0% better average CRPS score compared to the non-fractional and stationary model. However, for this particular dataset, it is only 1.3% better than a non-fractional and non-stationary model, suggesting that estimating the smoothness does not always enhance predictions. We also evaluate the predictive capabilities using a dataset of precipitation measurements. The proposed model captures some non-stationary trends, but its predictions are not noticeably better than those of the simpler models.



---

## Contents

---

<b>Abstract</b>	i
<b>Contents</b>	iii
<b>1 Introduction</b>	1
<b>2 Background</b>	5
2.1 Gaussian random fields . . . . .	5
2.2 Generalized Gaussian random fields . . . . .	6
2.3 Spatial regression . . . . .	8
2.4 Gaussian Markov random fields . . . . .	9
2.5 Scoring rules . . . . .	11
<b>3 SPDE model</b>	13
3.1 Link between SPDEs and Matérn fields . . . . .	13
3.2 Proposed model . . . . .	14
3.3 Parameterizing the diffusion matrix . . . . .	15
3.4 Parameterizing the non-stationarity . . . . .	17
3.4.1 Specifying the basis representation . . . . .	17
3.4.2 Computing the penalties . . . . .	19
<b>4 Discretization</b>	21
4.1 Outline of discrete approach . . . . .	21
4.2 Finite element method . . . . .	21
4.2.1 Discretizing the SPDE . . . . .	21
4.2.2 Computing the non-stationary finite element matrices . . . . .	24
4.2.3 Computing non-fractional precision matrix . . . . .	24
4.3 Approximating the fractional SPDE . . . . .	25
4.3.1 Fundamentals . . . . .	25
4.3.2 Computing the fractional precision matrix . . . . .	26
4.4 Summary of discrete approach . . . . .	27
<b>5 Inference</b>	29

5.1	Hierarchical model . . . . .	29
5.2	Parameter estimation . . . . .	30
5.3	Prediction . . . . .	31
5.4	Proposed optimization strategy . . . . .	31
<b>6</b>	<b>Simulation study</b>	<b>35</b>
6.1	Motivation and goals . . . . .	35
6.2	Evaluation . . . . .	36
6.3	Candidate models . . . . .	37
6.4	Stationary and non-fractional data . . . . .	38
6.5	Fractional and non-stationary data . . . . .	40
6.6	Varying signal-to-noise ratio . . . . .	41
<b>7</b>	<b>Analysis of precipitation data</b>	<b>45</b>
7.1	Dataset and goals . . . . .	45
7.2	Modelling the precipitation . . . . .	46
7.3	Investigating non-stationarity in the data . . . . .	47
7.4	Model comparison . . . . .	47
7.4.1	10-fold cross-validation . . . . .	47
7.4.2	Varying number of observations . . . . .	49
7.5	Interpretation of results . . . . .	50
<b>8</b>	<b>Discussion</b>	<b>53</b>
<b>A</b>	<b>Scaling factor for the marginal variance</b>	<b>55</b>

# CHAPTER 1

---

## Introduction

---

Spatial modeling and data analysis play a critical role in understanding complex systems across many scientific fields. For example, in meteorology, it is used to analyze atmospheric phenomena such as temperature or precipitation. These are often influenced by factors like local terrain and proximity to bodies of water, leading to strong spatial dependence (Hu et al., 2015). A convenient way of capturing the spatial characteristics is using a *spatial regression model*. Here, the observed value at each location is assumed to be the sum of fixed effects due to covariates, a spatial effect and an unstructured random effect (Gelfand et al., 2010). In a statistical model for precipitation, the fixed effects might include altitude, whereas the spatial effect accounts for spatially structured variation not explained by the fixed effects. Capturing these spatial variations allows the model to reflect gradual changes in precipitation across a region, such as the decrease in rainfall moving inland from a coastal area.

A common approach for representing the spatial effect is through a *Gaussian random field* (GRF). However, increased model size and complexity bring higher computational demands, and the use of GRFs pose challenges for large-scale spatial modelling (Banerjee et al., 2003). The SPDE approach, proposed by Lindgren et al. (2011), offers a way of constructing spatial regression models that enables efficient inference. Here, a GRF within the class of Matérn fields is defined via a *stochastic partial differential equation* (SPDE). Alternatively, in a *non-stationary* SPDE model, a non-stationary GRF can be constructed by making some coefficients in the SPDE spatially varying. Then, a *Gaussian Markov random field* (GMRF) approximation of its solution is computed by a *finite element method* (FEM). This link enables doing the spatial modelling using GRFs, while the computations are performed with the corresponding GMRFs (Lindgren et al., 2011). Since the precision matrix of the GMRF will typically be sparse, many mathemati-

cal operations can be done efficiently. The original approach was limited to specific smoothnesses, but Bolin and Kirchner (2020) generalized this by approximating the differential operator of a fractional SPDE. We refer to this new class of SPDE models as *fractional* SPDE models, and describe smoothness parameters requiring such models as *fractional smoothnesses*. Notably, since the fractional models can handle spatially varying parameters, it is of interest to investigate models which are both fractional and non-stationary.

The purpose of this thesis is to extend existing work on spatial anisotropy and fractional models to construct a more flexible SPDE-based model than previously considered. Further, we want to develop a computationally efficient approach for parameter estimation and prediction, and assess in which situations parameter estimates and predictions are improved by the more complex model. The proposed model incorporates non-stationarity is range, anisotropy and marginal variance, and allow any smoothness parameter. This type of model is not entirely new, as Bolin and Kirchner (2020) provided an example of a fractional model with a non-stationary range and marginal variance. However, their example lacked many details regarding implementation and inference. Also, as non-stationary anisotropy was omitted, their model was unable to capture this important feature. In fact, even though non-stationary anisotropy has been extensively studied (Fuglstad et al., 2015a,b; Llamazares-Elias et al., 2024), all previous models have been restricted to a specific non-fractional smoothness.

We first consider the derivation and implementation of the proposed model. It combines the fractional model of Bolin and Kirchner (2020) with the work on non-stationary anisotropy developed by Fuglstad et al. (2015a,b). Formulating the model is relatively simple, but solving the fractional SPDE complicates many computational aspects. In particular, performing parameter estimation involves optimizing a likelihood function which is hard to differentiate analytically or numerically, making a gradient term difficult to obtain. This requires a new approach for fitting the SPDE models, and we explore utilizing PyTorch. Its capabilities of automatic differentiation allows derivatives of the likelihood to be computed exactly, drastically increasing the convergence rate during optimization.

Introducing non-stationarity requires careful assessment of its parameterization, as there is a real danger of overfitting the model (Fuglstad et al., 2015b). We show how to address this by controlling the non-stationarity through a penalty term. For incorporating non-stationary anisotropy, Fuglstad et al. (2015a) proposed using a vector field specifying the direction and magnitude of the local anisotropic effect at every location. However, two different parameter-combinations could yield the same field, making the parameterization non-identifiable. Llamazares-Elias et al. (2024) proposes an alternative identifiable parameterization, and we incorporate this into our model. Moreover, we suggest an extra non-stationary term in the SPDE, which weakens the relation between the non-stationary parameters of the model.

Lastly, we investigate whether the proposed model yields any benefits when performing inference. We are interested in determining if it leads to more correct parameter estimates, evaluate its robustness to overfitting, and see if it improves

predictions. Evaluating the model against simpler alternatives is done in two parts. First, a simulation study is conducted, where we use simulated data from models with known parameters. Here, we assess the estimated parameters by their estimated bias and empirical standard deviation from repeated simulations. Predictions are evaluated with the empirical mean and standard deviation of the root-mean-square error (RMSE) and continuously ranked probability scores (CRPS). Secondly, we analyze a real-world dataset over the conterminous U.S. containing precipitation observations. The predictions for the different model are compared using the average RMSE and CRPS scores from a 10-fold cross-validation.

The thesis begins with Chapter 2 establishing the theoretical foundation for spatial regression and SPDE models. Chapter 3 introduces the proposed SPDE model, and details the parameterization of the non-stationarity and anisotropy. Next, Chapter 4 describes the process of discretizing the model using a FEM, as well incorporating the fractional smoothness. Chapter 5 shows how to use the spatial regression model for parameter estimation and prediction. Chapter 6 presents a simulation study, followed by the analysis of precipitation data in Chapter 7. Finally, Chapter 8 concludes the thesis by summarizing our findings and suggesting directions for future work.



# CHAPTER 2

---

## Background

---

### 2.1 Gaussian random fields

Observed data is often connected spatially in ways that traditional regression models cannot explain, and spatial regression models handle this by including a spatial effect. Due to their mathematical simplicity, GRFs are commonly used in spatial regression models, as they provide a flexible way of modelling spatial dependencies and uncertainty. Before defining the GRF, recall the definition of the multivariate normal distribution.

**Definition 2.1.1** (Multivariate Gaussian distribution). *The probability density function (PDF) of a multivariate Gaussian distribution is given by*

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where  $\boldsymbol{\mu} \in \mathbb{R}^n$  and  $\boldsymbol{\Sigma}$  is an  $n \times n$  matrix with  $\boldsymbol{\Sigma} > 0$ .  $|\cdot|$  denotes the determinant. We write  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  to denote that  $\mathbf{x}$  is normally distributed with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ .

A GRF is a generalization of the multivariate normal distribution to a continuous spatial domain. While a multivariate normal distribution characterizes a finite collection of jointly Gaussian random variables with a specified covariance matrix, a GRF defines a collection of continuously-indexed random variables.

**Definition 2.1.2** (Gaussian Random Field).  *$\{u(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$  on a domain  $\mathcal{D} \subset \mathbb{R}^d$  is a Gaussian random field if  $\forall m \in \mathbb{N}$ ,  $\forall a_1, \dots, a_m \in \mathbb{R}$  and  $\forall \mathbf{s}_1, \dots, \mathbf{s}_m \in \mathcal{D}$ ,  $\sum_{i=1}^m a_i u(\mathbf{s}_i)$  is a Gaussian random variable.*

The spatial dependence in a GRF is fully described through its mean function

$\mu : \mathcal{D} \rightarrow \mathbb{R}$  and covariance function  $c : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ . Based on the choices of  $\mu$  and  $c$ , the field can exhibit different spatial properties. In many situations, spatial dependence is approximately uniform in different directions, and the GRFs are therefore often assumed to be *isotropic*. This means having a constant mean, and that the covariance between every pair of points  $\mathbf{s}, \mathbf{s}' \in \mathcal{D}$  is a function of the Euclidean distance  $\|\mathbf{s} - \mathbf{s}'\|$ . A common isotropic covariance function is the Matérn covariance function, which has separate parameters to control the smoothness, range and marginal variance. These are often the most important properties, and the ability to specify each one independently makes the Matérn covariance function a popular choice (Stein, 1999).

**Definition 2.1.3** (Matérn Covariance Function). *The Matérn covariance function is on the form*

$$c(\mathbf{h}; \nu, \kappa^2, \sigma^2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|\mathbf{h}\|)^\nu K_\nu(\kappa\|\mathbf{h}\|), \quad \mathbf{h} \in \mathcal{D}, \quad (2.1)$$

for  $\nu, \kappa^2, \sigma^2 > 0$ . Here,  $\|\cdot\|$  denotes the Euclidean distance,  $\Gamma(\cdot)$  is the Gamma function and  $K_\nu(\cdot)$  is the modified Bessel function of the second kind with order  $\nu$ .

The parameter  $\nu$  determines the smoothness,  $\kappa^2$  controls the range, and  $\sigma^2$  the marginal variance. We often use *correlation range* to describe at which distance the correlation between two points becomes negligible. In this thesis it is defined as  $\rho = \sqrt{8\nu}/\kappa$ , corresponding to correlations near 0.1 at the distance  $\rho$ . Denoting the ceiling function as  $\lceil \cdot \rceil$ , a smoothness parameter  $\nu$  will yield realizations which are  $\lceil \nu \rceil - 1$  times differentiable.

Another way of utilizing GRFs is generating new fields through a SPDE. This is the foundation for the SPDE approach detailed in the coming chapters.

**Example 2.1.1.** *A simple example of a SPDE generating a GRF  $u$  on  $\mathbb{R}^2$  is*

$$(1 - \Delta) u(x, y) = v(x, y), \quad (x, y) \in \mathbb{R}^2, \quad (2.2)$$

where  $v$  is a known GRF.  $\Delta = \nabla \cdot \nabla$  is the Laplace operator, where  $\nabla$  represents the gradient.

From Equation (2.2), it is clear that the solution  $u$  should make the left- and right-hand side equal, but it is not clear in what sense they should be equal. For this thesis, we specify a *solution* of an SPDE to be the GRF which makes the two sides of the equation equal in distribution. In Example 2.1.1 a solution  $u$  should therefore be the GRF giving the two sides the same mean and covariance.

## 2.2 Generalized Gaussian random fields

In Example 2.1.1, the right-hand side of Equation (2.2) was assumed to be a GRF. However, for most SPDEs in this thesis, the right-hand side is chosen as Gaussian white noise  $\mathcal{W}$ , which is a generalized GRF. Informally, this can be seen as a GRF that is discontinuous everywhere and has infinite variance. Since  $\mathcal{W}$  lacks a well-defined pointwise meaning, we must relax the concept of a random field to

allow for more flexibility.

**Example 2.2.1** (Dirac delta function). Consider the function

$$f_\sigma(x) = (\sqrt{2\pi}\sigma)^{-1} \exp\left\{-x^2/2\sigma^2\right\}, \quad \sigma > 0, x \in \mathbb{R}.$$

Let  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  be a function which is sufficiently smooth and non-zero on a bounded domain, referred to as a test function. The limit of the integral  $\int_{\mathbb{R}} \varphi(x) f_\sigma(x) dx$  as  $\sigma$  approaches zero becomes

$$\lim_{\sigma \rightarrow 0^+} \int_{\mathbb{R}} \varphi(x) f_\sigma(x) dx = \varphi(0) \quad (2.3)$$

We can then define a generalized function  $\delta$  through the action  $\delta[\varphi] = \varphi(0)$ , and see that this can be interpreted as the limit of the integral considered in Equation (2.3). Note that the generalized function  $\delta$  is not a function itself, but gets its meaning through its action on test functions.

In an analogous way as the  $\delta$  is Example 2.2.1, a generalized random field  $u$  on  $\mathcal{D} \subset \mathbb{R}^d$  is determined by the random properties of its action  $u[\varphi]$  on smooth and compactly supported test functions  $\varphi : \mathcal{D} \rightarrow \mathbb{R}$ . We often use the abuse of notation

$$\langle \varphi, u \rangle_{\mathcal{D}} \stackrel{\text{def}}{=} u[\varphi],$$

where we can heuristically think of the action  $u[\varphi]$  as an inner product between  $u$  and test function  $\varphi$ . The derivative of a generalized field is understood in a distributional sense. Rather than computing derivatives at individual points, the *weak* derivative  $u'$  is interpreted as satisfying integration by parts. Specifically, the action of  $u'$  on  $\varphi$  is given by

$$\langle \varphi, u' \rangle_{\mathcal{D}} = -\langle \varphi', u \rangle_{\mathcal{D}},$$

where  $\varphi'$  denotes the classical derivative of the test function.

The notion of Gaussianity can be extended to generalized fields through generalized GRFs, which gives Gaussian properties to the inner products.

**Definition 2.2.1** (Generalized Gaussian random field). A generalized GRF  $u$  on  $\mathcal{D}$  is a generalized field such that for every  $m \in \mathbb{N}$ , the inner products

$$\langle \varphi_i, u \rangle_{\mathcal{D}}, \quad i = 1, \dots, m,$$

are jointly Gaussian for every set of smooth and compactly supported test functions  $\{\varphi_i\}_{i=1}^m$ . If there exists a constant  $b \geq 0$  such that  $E[\langle \varphi, u \rangle_{\mathcal{D}}] \leq b \langle \varphi, \varphi \rangle_{\mathcal{D}}$  for every  $\varphi$ , we say the generalized GRF has bounded second moments.

A particularly important example of a generalized GRF is Gaussian white noise  $\mathcal{W}$ , which will serve as a key component in the SPDE model.

**Definition 2.2.2** (Gaussian white noise). Gaussian white noise  $\mathcal{W}$  on  $\mathcal{D}$  is a generalized GRF with bounded second moments such that, for every  $m \in \mathbb{N}$  and test functions  $\{\varphi_i\}_{i=1}^m$ , the inner products  $\langle \varphi_i, \mathcal{W} \rangle_{\mathcal{D}}$  for  $i = 1, \dots, m$  are jointly Gaussian with expectation  $E[\langle \varphi_i, \mathcal{W} \rangle_{\mathcal{D}}] = 0$  and covariance

$$\text{Cov}[\langle \varphi_i, \mathcal{W} \rangle_{\mathcal{D}}, \langle \varphi_j, \mathcal{W} \rangle_{\mathcal{D}}] = \langle \varphi_i, \varphi_j \rangle_{\mathcal{D}}. \quad (2.4)$$

Similarly as for the GRFs, concept of equality between two generalized GRFs is in this thesis interpreted as almost sure equality in distribution. This means that the action of the two sides on every choice of test functions should have the same distribution.

**Example 2.2.2** (Equality in distribution). *For a generalized GRF  $u$  on  $\mathcal{D}$ , Gaussian white noise  $\mathcal{W}$  and differential operator  $L$ , the equality  $Lu = \mathcal{W}$  holds in distribution if*

$$\begin{aligned} \mathbb{E} [\langle \varphi, Lu \rangle_{\mathcal{D}}] &= \mathbb{E} [\langle \varphi, \mathcal{W} \rangle_{\mathcal{D}}] \\ \text{Cov} [\langle \varphi, Lu \rangle_{\mathcal{D}}, \langle \psi, Lu \rangle_{\mathcal{D}}] &= \text{Cov} [\langle \varphi, \mathcal{W} \rangle_{\mathcal{D}}, \langle \psi, \mathcal{W} \rangle_{\mathcal{D}}] \end{aligned}$$

for all test functions  $\varphi$  and  $\psi$ .

## 2.3 Spatial regression

In this thesis, for the GRFs and generalized GRFs to be useful for inference, they must be understood within the broader framework of spatial regression models. There are many ways a spatial structure shows up in data, and introducing space into regression models can in some cases yield better estimates and predictions. Spatial structures usually improve regression models in two main ways. They can capture a known underlying spatial process that directly affects the data, or they can account for unknown factors by reflecting spatial patterns in the prediction errors.

The spatial regression model consists of a latent GRF  $\eta$  on  $\mathcal{D}$  and unstructured noise  $\epsilon$ . We commonly think of  $\eta$  as the true underlying field we want to predict, whereas the noise representing observation error in our data. For every location  $\mathbf{s} \in \mathcal{D}$ , we model the latent GRF as the sum of fixed effects due to covariates and a spatial field. This gives

$$\eta(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta} + u(\mathbf{s}),$$

where  $\mathbf{x}(\mathbf{s})^T$  is a  $p$ -dimensional row vector describing the covariates at location  $\mathbf{s} \in \mathcal{D}$ ,  $\boldsymbol{\beta}$  is an unknown  $p$ -dimensional vector defining the influence of each covariate on the response, and  $u$  is a GRF with zero mean and unknown covariance. Incorporating data, consider a set of  $n$  observations made at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$ . Then, for each  $i = 1, \dots, n$ , the observation  $y_i$  is modelled as

$$y_i = \eta(\mathbf{s}_i) + \epsilon_i, \quad \epsilon_i | \sigma_e^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2),$$

for some unknown variance  $\sigma_e^2$  determining the impact of the unstructured noise. This structure forms the basis for the hierarchical model constructed in Chapter 5.

Utilizing the framework of the spatial regression model, we can now compute the maximum likelihood estimator of the model parameters by conditioning on observed data. Let  $u$  be a zero mean GRF on  $\mathcal{D} \subset \mathbb{R}^2$  described by unknown parameters  $\boldsymbol{\theta}$ . Assume no fixed effects, and that we have made exact observations  $\mathbf{u} = (u(\mathbf{s}_1), \dots, u(\mathbf{s}_n))^T$  of the GRF at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$ . Denoting the distribution  $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$ , the parameters can be estimated using a maximum

likelihood estimator

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \left\{ -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma(\boldsymbol{\theta})| - \frac{1}{2} \mathbf{u}^T \Sigma(\boldsymbol{\theta})^{-1} \mathbf{u} \right\}. \quad (2.5)$$

Notice that evaluating the likelihood involves computing the log-determinant and inverting the covariance matrix, and that this must be done for each optimization step. Here computing the Cholesky factorization of  $\Sigma$  required, which has a time complexity of  $\mathcal{O}(n^3)$  for dense  $n \times n$  matrices.

## 2.4 Gaussian Markov random fields

As illustrated in the previous section, modelling using GRFs requires factorizing a dense covariance matrix. This bottleneck makes it challenging to scale the spatial models effectively, and restricts their capabilities for handling large models. We are therefore interested in methods for doing spatial modelling in a computationally less costly way, and Lindgren et al. (2011) found one alternative by constructing an explicit link between GRFs and GMRFs.

A GMRF can be seen as a way of applying the conditional independence structure of a GRF on a finite set of points, and is characterized by a sparse matrix  $\mathbf{Q}$  called the *precision matrix*. Compared to the continuously-indexed GRF, the discretely-indexed GMRF offers many computational benefits. In particular, the sparseness of  $\mathbf{Q}$  means it can be factorized efficiently. Additionally, sparse matrices require less memory to store, something which allows for the consideration of much larger systems without encountering memory issues.

To rigorously define the dependence between points in the GMRF, the concept of a graph is required. For the purpose of this thesis, only finite and undirected graphs is considered.

**Definition 2.4.1** (Graph). *A graph is a tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is a finite set of nodes and  $\mathcal{E}$  is the set of edges  $\{i, j\}$  for  $i, j \in \mathcal{V}$  and  $i \neq j$ . If  $\{i, j\} \in \mathcal{E}$  there is an edge from node  $i$  to node  $j$ , and otherwise there is no edge between the nodes. If  $\mathcal{V} = \{1, 2, \dots, n\}$ ,  $\mathcal{G}$  is called a labeled graph.*

Figure 2.1a provides a simple example of how a graph might look. Often, it has some intrinsic interpretation, like the vertices representing cities and the edges being highways connecting them. However, for the SPDE model, the graph correspond to a triangulation of a spatial domain, something which will be discussed in Chapter 4. The following definition of the GMRF is given in Rue and Held (2005).

**Definition 2.4.2** (Gaussian Markov random field). *A random vector  $\mathbf{u} \in \mathbb{R}^n$  is called a GMRF with respect to a labeled graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with mean  $\boldsymbol{\mu}$  and precision  $\mathbf{Q} > 0$ , if and only if its density has the form*

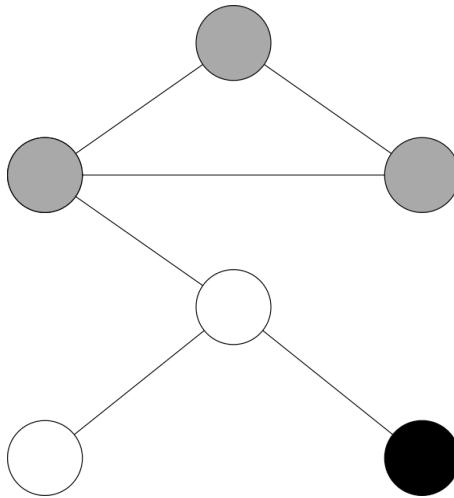
$$f(\mathbf{u}; \boldsymbol{\mu}, \mathbf{Q}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{u} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{u} - \boldsymbol{\mu}) \right\}$$

and

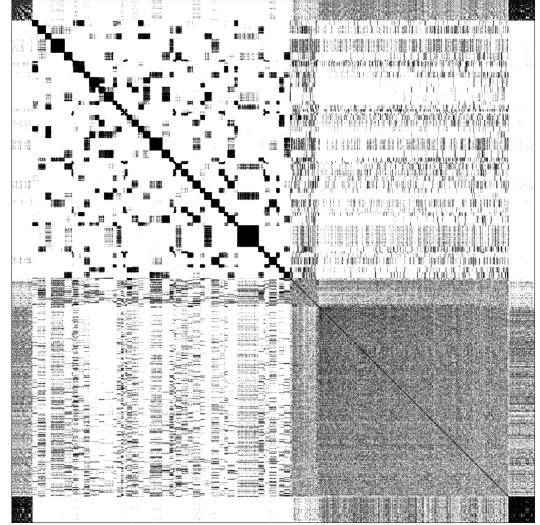
$$\mathbf{Q}_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \quad \forall i \neq j.$$

Note that since the graph can be directly inferred from the non-zero elements of the precision matrix, it is not included in the notation unless explicitly required. Comparing the density of the GMRF to that of a Gaussian distribution, see that the precision matrix  $\mathbf{Q}$  is the inverse of the covariance matrix  $\Sigma$ . However, this is not just a re-parametrization, but also a change of perspective. The off-diagonal elements of the precision matrix characterizes the conditional dependence properties of the Gaussian distribution.

The main reason for using GMRFs in this thesis is that when solving the considered SPDEs using a finite element method, it results in GMRFs where each variable is only conditionally dependent on the variables closest to itself. As we will discuss in Chapter 4, this in turn means that the precision matrices becomes very sparse. Figure 2.1b visualizes an example of how such a matrix can look like after permuting its rows and columns.



(a) Graph with three categories of nodes. The black node has no direct edge to any of the gray nodes, and is only connected to one of the white nodes.



(b) The sparsity structure of a  $11505 \times 11505$  permuted precision matrix  $\mathbf{Q}$ . Non-zero elements are colored black.

**Figure 2.1:** Examples of how graph representations relates to the sparsity structure of the precision matrix. Nodes not connected with edges will have corresponding zero value in  $\mathbf{Q}$ .

For spatial modelling using GRFs, the Cholesky factorization of the covariance matrix  $\Sigma$  is typically the most computationally expensive part. Utilizing the sparseness of the precision matrix, Rue and Held (2005, Algorithm 2.8) provides an efficient method for factorizing it. In particular, given an  $n \times n$  precision matrix  $\mathbf{Q}$  with a two-dimensional spatial sparsity structure, the time-complexity is typically  $\mathcal{O}(n^{3/2})$ . To compute the likelihood we utilize the relation  $|\mathbf{Q}| = |\mathbf{L}| |\mathbf{L}^T| = |\mathbf{L}|^2$ , where  $\mathbf{L}$  denotes the lower Cholesky factor of  $\mathbf{Q}$ . By the density in Equation

(2.4.2), the log-likelihood becomes

$$\log f(\mathbf{u}; \boldsymbol{\mu}, \mathbf{Q}) = -\frac{n}{2} \log(2\pi) + \sum_{i=1}^n \log \mathbf{L}_{ii} - \frac{1}{2} (\mathbf{u} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{u} - \boldsymbol{\mu}), \quad \mathbf{u} \in \mathbb{R}^n.$$

The factorization can also be used for generating simulations. The following theorem shows how this can be done by solving a lower triangular system of equations.

**Theorem 2.4.1** (Simulating from a GMRF). *Let  $\mathbf{u}$  be a GMRF with mean  $\boldsymbol{\mu}$  and precision matrix  $\mathbf{Q} > 0$ , and let  $\mathbf{L}$  be the lower triangular Cholesky factor of  $\mathbf{Q}$ . Then we have the distribution*

$$\mathbf{L}^T (\mathbf{u} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

*Proof.* Can be found in Rue and Held (2005, Section 2.3.2) □

One disadvantage of working with precision matrices instead of covariance matrices, is that the marginal properties of the distribution are no longer directly available. For example, determining the marginal variances requires computing the diagonal elements of the inverse precision matrix. However, Gelfand et al. (2010, Section 12.1.7.11) provides an algorithm for utilizing the sparseness for computing the *partial inverse*. Here, by applying Takahashi recursions (Takahashi, 1973) to the Cholesky factor of the precision matrix, marginal variances and neighbor covariances is obtained in  $\mathcal{O}(n(\log n)^2)$  time (Lindgren et al., 2022). In Python, the C++ code provided by Simpson (2024) can be wrapped to a function performing this computation.

## 2.5 Scoring rules

When comparing models in the simulation study and precipitation analysis, it is important to have clear ways of evaluating their parameter estimates and predictions. For this purpose, numerical scores are used for summarizing how well the models perform. We evaluate the models using two scoring rules, the RMSE and the CRPS. Each score provides a different perspective on the performance of the model.

Assume there are  $n$  observations  $\mathbf{y} = (y_1, \dots, y_n)^T$  at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$ . For the predictive Gaussian distribution of the spatial model, let

$$\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^T \quad \text{and} \quad \hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1, \dots, \hat{\sigma}_n)^T,$$

denote the predicted mean and standard deviation at the corresponding locations. The RMSE is defined as

$$\text{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

and measures the square difference between predicted and observed values. When comparing different predictors, a lower RMSE indicates a better score.

The CRPS compares the cumulative distribution of predictions to the true observations (Gneiting and Raftery, 2007). In general, for an observed value  $y^*$  and a predictive distribution with cumulative distribution function  $F$ , it is defined as

$$\text{CRPS}(y^*, F) = \int_{-\infty}^{\infty} (F(y^*) - I(x \geq y^*))^2 dx,$$

where  $I$  denotes the indicator function. In our case, as the predictive distribution is Gaussian, we can for every  $i = 1, \dots, n$  express the CRPS on the form

$$\text{CRPS}(y_i, \mathcal{N}(\hat{y}_i, \hat{\sigma}_i^2)) = \hat{\sigma}_i \left[ 2\phi\left(\frac{y_i - \hat{y}_i}{\hat{\sigma}_i}\right) + \frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \left( 2\Phi\left(\frac{y_i - \hat{y}_i}{\hat{\sigma}_i}\right) - 1 \right) - \frac{1}{\sqrt{\pi}} \right].$$

Here,  $\phi$  and  $\Phi$  denote the probability density function and the cumulative density function of the standard normal distribution, respectively. As for the RMSE, a lower CRPS is preferred when comparing models. For convenience, it is useful to have a single CRPS score for a set of observations. The average CRPS, hereby simply referred to as the CRPS, is defined as

$$\overline{\text{CRPS}}(\mathbf{y}, \hat{\mathbf{y}}, \hat{\boldsymbol{\sigma}}) = \frac{1}{n} \sum_{i=1}^n \text{CRPS}(y_i, \mathcal{N}(\hat{y}_i, \hat{\sigma}_i^2)).$$

# CHAPTER 3

---

## SPDE model

---

### 3.1 Link between SPDEs and Matérn fields

In this chapter, we present the proposed SPDE model, and explain how to parameterize it. The foundation of the model is a result by Whittle (1954, 1963), which found a connection between a zero-mean and isotropic GRF  $u$  and a stochastic partial differential equation. In particular, a GRF  $u$  on  $\mathbb{R}^2$  with the Matérn covariance function from Equation (2.1) can be represented as the solution to the SPDE

$$(\kappa^2 - \Delta)^\beta u(\mathbf{s}) = \sigma \sqrt{4\pi \frac{\Gamma(2\beta)}{\Gamma(2\beta - 1)}} \kappa^{2\beta-1} \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2, \quad (3.1)$$

where  $\kappa^2, \sigma^2 > 0$ ,  $\Delta = \nabla \cdot \nabla$  is the Laplace operator and  $\mathcal{W}$  is Gaussian white noise. The fractional power  $\beta > 1/2$  is related to the smoothness parameter  $\nu > 0$  via the relation  $2\beta = \nu + 1$ . The interpretation of the parameters  $\nu$ ,  $\kappa^2$  and  $\sigma^2$  remains the same as for the covariance function.

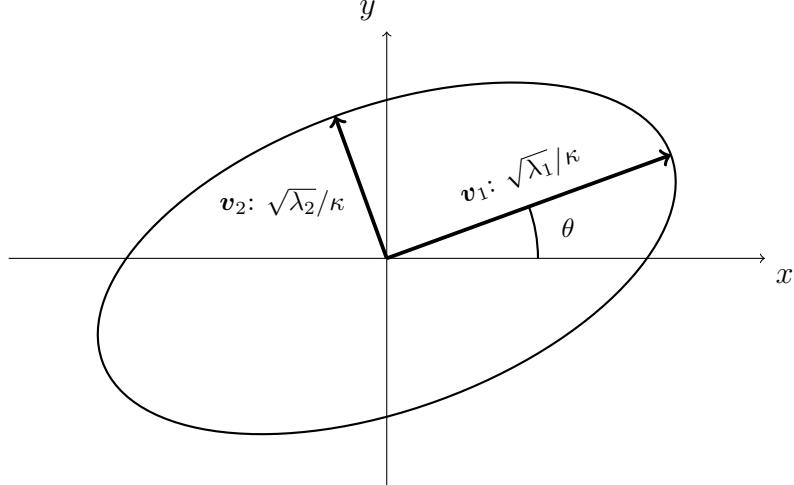
To generate a GRF with an anisotropic Matérn-like covariance function, the Laplacian  $\Delta$  is replaced by the operator  $\nabla \cdot \mathbf{H} \nabla$ .  $\mathbf{H}$  is a  $2 \times 2$  symmetric positive definite matrix. Now, the SPDE

$$(\kappa^2 - \nabla \cdot \mathbf{H} \nabla)^\beta u(\mathbf{s}) = \sigma \sqrt{4\pi \frac{\Gamma(2\beta)}{\Gamma(2\beta - 1)}} \kappa^{2\beta-1} (\det \mathbf{H})^{1/4} \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2, \quad (3.2)$$

gives a GRF with covariance function

$$c(\mathbf{h}; \nu, \kappa^2, \sigma^2, \mathbf{H}) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} (\kappa \|\mathbf{H}^{-1/2} \mathbf{h}\|)^\nu K_\nu (\kappa \|\mathbf{H}^{-1/2} \mathbf{h}\|), \quad \mathbf{h} \in \mathbb{R}^2.$$

Here, the right-hand side of Equation (3.2) is derived analytically in Appendix A. Compared to the isotropic Matérn covariance function in Equation (2.1), a change in the directionality is introduced. Figure 3.1 shows how the eigenpairs of  $\mathbf{H}$  and the value of  $\kappa^2$  act together to control the direction dependent range. The construction leads to elliptic iso-covariance curves.



**Figure 3.1:** Iso-correlation curve for the anisotropic covariance function in  $\mathbb{R}^2$ . Visualizes all points with correlation 0.6 to the center, where this level is a result of choosing  $\nu = 1$ . Observe the directionality and strength introduced by  $\kappa^2$  and the diffusion matrix  $\mathbf{H}$ .  $(\lambda_1, \mathbf{v}_1)$  and  $(\lambda_2, \mathbf{v}_2)$  are the eigenpairs of  $\mathbf{H}$  where  $\lambda_1 \geq \lambda_2$ .  $\theta$  denotes the angle between the  $x$ -axis and  $\mathbf{v}_1$ . The semi-major and semi-minor axes of the ellipse point in the directions of  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , and their lengths are  $\sqrt{\lambda_1}/\kappa$  and  $\sqrt{\lambda_2}/\kappa$ , respectively.

## 3.2 Proposed model

The model proposed in this thesis extends the SPDE in Equation (3.2) by making the parameters controlling range, anisotropy and marginal variance location dependent. Loosely speaking, the non-stationarity can be thought of as different Matérn-like fields, locally with their own set of fixed parameters, being combined into a full non-stationary process (Fuglstad et al., 2015a). As will be discussed further in Chapter 4, when discretizing the problem, we have to restrict the SPDE to a bounded and closed domain  $\mathcal{D} \subset \mathbb{R}^2$ . For this, Lindgren et al. (2011) enforces zero Neumann boundary conditions, so we will do the same. The proposed SPDE takes the form

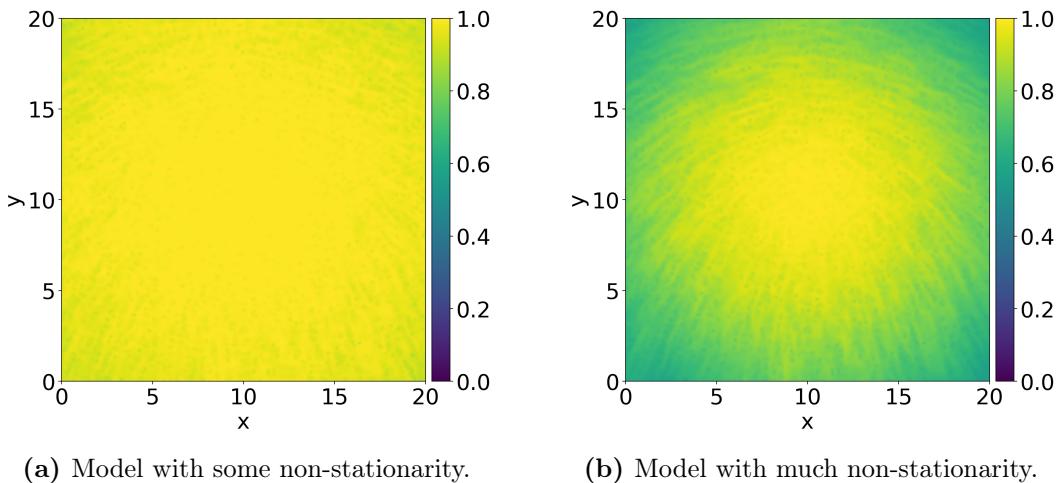
$$\begin{cases} (\kappa^2(\mathbf{s}) - \nabla \cdot \mathbf{H}(\mathbf{s}) \nabla)^\beta u(\mathbf{s}) = \tau(\mathbf{s}) \mathcal{W}(\mathbf{s}) & \mathbf{s} \in \mathcal{D}, \\ (\mathbf{H}(\mathbf{s}) \nabla u(\mathbf{s})) \cdot \mathbf{n} = 0 & \mathbf{s} \in \partial \mathcal{D}. \end{cases} \quad (3.3)$$

where  $\mathbf{n}$  denotes the outwards pointing unit normal vector and  $\partial \mathcal{D}$  denotes the boundary of the domain. The scaling term on the right-hand side of the equation is

$$\tau(\mathbf{s}) = \sigma(\mathbf{s}) \sqrt{4\pi \frac{\Gamma(2\beta)}{\Gamma(2\beta-1)}} \kappa(\mathbf{s})^{2\beta-1} (\det \mathbf{H}(\mathbf{s}))^{1/4}. \quad (3.4)$$

See that  $\tau$  closely resembles the right-hand side of the stationary SPDE from Equation (3.2), but that the parameters has been made non-stationary. This is intended to act as a stabilizing term, weakening the connection between the non-stationary parameters controlling the range, anisotropy and marginal variance.

Another reason for the choice of  $\tau$  is that, when the parameters are constant over the spatial domain, the scaling ensures that they retain the same interpretation as in the stationary model. However, note that although the term makes the equation invariant to the scaling of  $\kappa^2$  and  $\mathbf{H}$ , it is still not invariant to non-stationary variation in the parameters. Figure 3.2 plots the marginal standard deviation for two different models, one which is slightly non-stationary and another which is very non-stationary. Both models have a fixed parameter  $\sigma = 1$ , but we see that the marginal standard deviation deviates from 1 in areas with high non-stationarity.



**Figure 3.2:** Marginal standard deviations for two different models with a fixed parameter  $\sigma = 1$ . Observe how the amount of non-stationary anisotropy impacts the marginal standard deviation. Non-stationarity is introduced near the edges of the domain, while the center remains almost stationary. A realization of the field in Figure 3.2b is shown in Figure 3.3b.

As noted in the introduction, what makes the proposed model different from previous models is its inclusion of both a fractional smoothness parameter  $\nu = 2\beta - 1$  and a non-stationary spatial anisotropy. Additionally, although the parameterization of the right-hand side is inspired by that of Llamazares-Elias et al. (2024), the  $\tau$  from Equation (3.4) handles any smoothness.

### 3.3 Parameterizing the diffusion matrix

To use the model for inference, the terms of the SPDE have to be parameterized in some meaningful manner. Parameterizing the diffusion matrix requires some restrictions to assure positive definiteness, and Fuglstad et al. (2015b) proposed decomposing it into scalar functions. The spatially varying diffusion matrix was defined as

$$\mathbf{H}(\mathbf{s}) = \gamma(\mathbf{s})\mathbf{I} + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^T, \quad \mathbf{s} \in \mathcal{D}, \quad (3.5)$$

where  $\gamma(\mathbf{s})$  determined the isotropic effect and the vector field  $\mathbf{v}(\mathbf{s}) = (v_x(\mathbf{s}), v_y(\mathbf{s}))^\top$  specified the direction and magnitude of the local anisotropic effect. An issue here is that decomposition of  $\mathbf{H}$  is invariant to changing the sign of  $\mathbf{v}$  to  $-\mathbf{v}$ , meaning that two different choices of parameters yields the same matrix. As demonstrated in Fuglstad et al. (2015a), this creates undesirable local maxima making the parameter estimation more challenging.

An alternative parameterization was presented in Llamazares-Elias et al. (2024). Here, instead of using the vector field  $\mathbf{v} = (v_x, v_y)^\top$ , we consider the half angle version  $\tilde{\mathbf{v}}$  defined as

$$\tilde{\mathbf{v}} = \|\mathbf{v}\| \exp\{i\alpha/2\} \quad \text{with } \alpha = \arg(\mathbf{v}) \in [0, 2\pi).$$

The dependence on  $\mathbf{s} \in \mathcal{D}$  is here omitted for notational simplicity. After the transformation,  $\tilde{\mathbf{v}}$  retains the same magnitude  $\|\mathbf{v}\|$  as the original vector, but takes angles  $[0, \pi)$  instead of  $[0, 2\pi)$ . Thus, the two vectors  $\mathbf{v}$  and  $-\mathbf{v}$  now map to distinct  $\tilde{\mathbf{v}} \in \mathbb{C}^2$ , and no longer belong to the same parameter space.

Defining the transformed vector  $\tilde{\mathbf{v}} = (\tilde{v}_x, \tilde{v}_y)^\top$  and the orthogonal counterpart  $\tilde{\mathbf{v}}_\perp = (-\tilde{v}_y, \tilde{v}_x)^\top$ , we express the diffusion matrix as

$$\mathbf{H} = \frac{\exp\{\|\mathbf{v}\|\}}{\|\mathbf{v}\|^2} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^\top + \frac{\exp\{-\|\mathbf{v}\|\}}{\|\mathbf{v}\|^2} \tilde{\mathbf{v}}_\perp \tilde{\mathbf{v}}_\perp^\top.$$

Note that both  $\tilde{\mathbf{v}} \tilde{\mathbf{v}}^\top$  and  $\tilde{\mathbf{v}}_\perp \tilde{\mathbf{v}}_\perp^\top$  are matrices of rank 1, and the orthogonality of the vectors gives determinant

$$\det \mathbf{H} = \left( \frac{\exp\{\|\mathbf{v}\|\}}{\|\mathbf{v}\|^2} \right) \left( \frac{\exp\{-\|\mathbf{v}\|\}}{\|\mathbf{v}\|^2} \right) \|\mathbf{v}\|^4 = 1.$$

The motivation behind this choice of parameterization can be formulated in the following theorem.

**Theorem 3.3.1.**  *$\mathbf{H}$  is a smooth and bijective parameterization on the space of symmetric positive definite matrices with determinant 1.*

*Proof.* Can be found in Llamazares-Elias et al. (2024, Supplementary Material S1).  $\square$

The smoothness of the parameterization ensures that small changes in  $\mathbf{v}$  result in correspondingly small changes in  $\mathbf{H}$ , and vice versa. Due to it being injective, we can recover the vector  $\mathbf{v}$  uniquely when given a matrix  $\mathbf{H}$ . Additionally, the surjectivity means any symmetric positive definite matrix with determinant 1 can be generated with the parameterization.

For computing the diffusion matrix, we can avoid explicitly using the complex vectors  $\tilde{\mathbf{v}}$  and  $\tilde{\mathbf{v}}_\perp$ . By considering the half angle formulas for  $\sin(\alpha/2)$  and  $\cos(\alpha/2)$  we find that

$$\tilde{\mathbf{v}} \tilde{\mathbf{v}}^\top = \frac{\|\mathbf{v}\|^2}{2} \mathbf{I} + \frac{\|\mathbf{v}\|}{2} \begin{bmatrix} v_x & v_y \\ v_y & -v_x \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{v}}_\perp \tilde{\mathbf{v}}_\perp^\top = \frac{\|\mathbf{v}\|^2}{2} \mathbf{I} - \frac{\|\mathbf{v}\|}{2} \begin{bmatrix} v_x & v_y \\ v_y & -v_x \end{bmatrix}.$$

Rewriting using the hyperbolic functions yields the final parameterization

$$\mathbf{H}(\mathbf{s}) = \cosh(\|\mathbf{v}(\mathbf{s})\|) \mathbf{I} + \frac{\sinh(\|\mathbf{v}(\mathbf{s})\|)}{\|\mathbf{v}(\mathbf{s})\|} \begin{bmatrix} v_x(\mathbf{s}) & v_y(\mathbf{s}) \\ v_y(\mathbf{s}) & -v_x(\mathbf{s}) \end{bmatrix}, \quad \mathbf{s} \in \mathcal{D}. \quad (3.6)$$

One drawback of the new parameterization is that the connection between the vector field  $\mathbf{v}$  and the directionality of the anisotropy is less clear. As an example, Figure 3.3 shows realizations for the parameterizations in Equations (3.5) and (3.6) together with the underlying vector field. Using the parameterization by Fuglstad et al. (2015b), the direction of anisotropy at each location aligns with the vector field. However, this alignment is absent in the parameterization by Llamazares-Elias et al. (2024). Despite this minor limitation, we choose to implement the non-stationary models using the identifiable parameterization.

## 3.4 Parameterizing the non-stationarity

### 3.4.1 Specifying the basis representation

For implementing the non-stationarity of the model, we need a way to control how the parameters  $\kappa^2$ ,  $\sigma^2$ ,  $v_x$  and  $v_y$  vary over the spatial domain. This involves decomposing the non-stationarity into basis functions which influence can be controlled by coefficients. Fuglstad et al. (2015b) also proposed using a penalty for making sure the parameters does not vary too much, and we want to incorporate this. Note that to avoid having to directly handle the restriction of  $\kappa^2$  and  $\sigma^2$  only being positive, the logarithm of the parameters is parameterized instead.

All the non-stationary parameters will employ the same parameterization, so to keep the notation general we parameterize some  $f : \mathcal{D} \rightarrow \mathbb{R}$ . The function can be decomposed into a stationary contribution  $\alpha_0$  and a non-stationary  $f_{\text{NS}}$ , giving

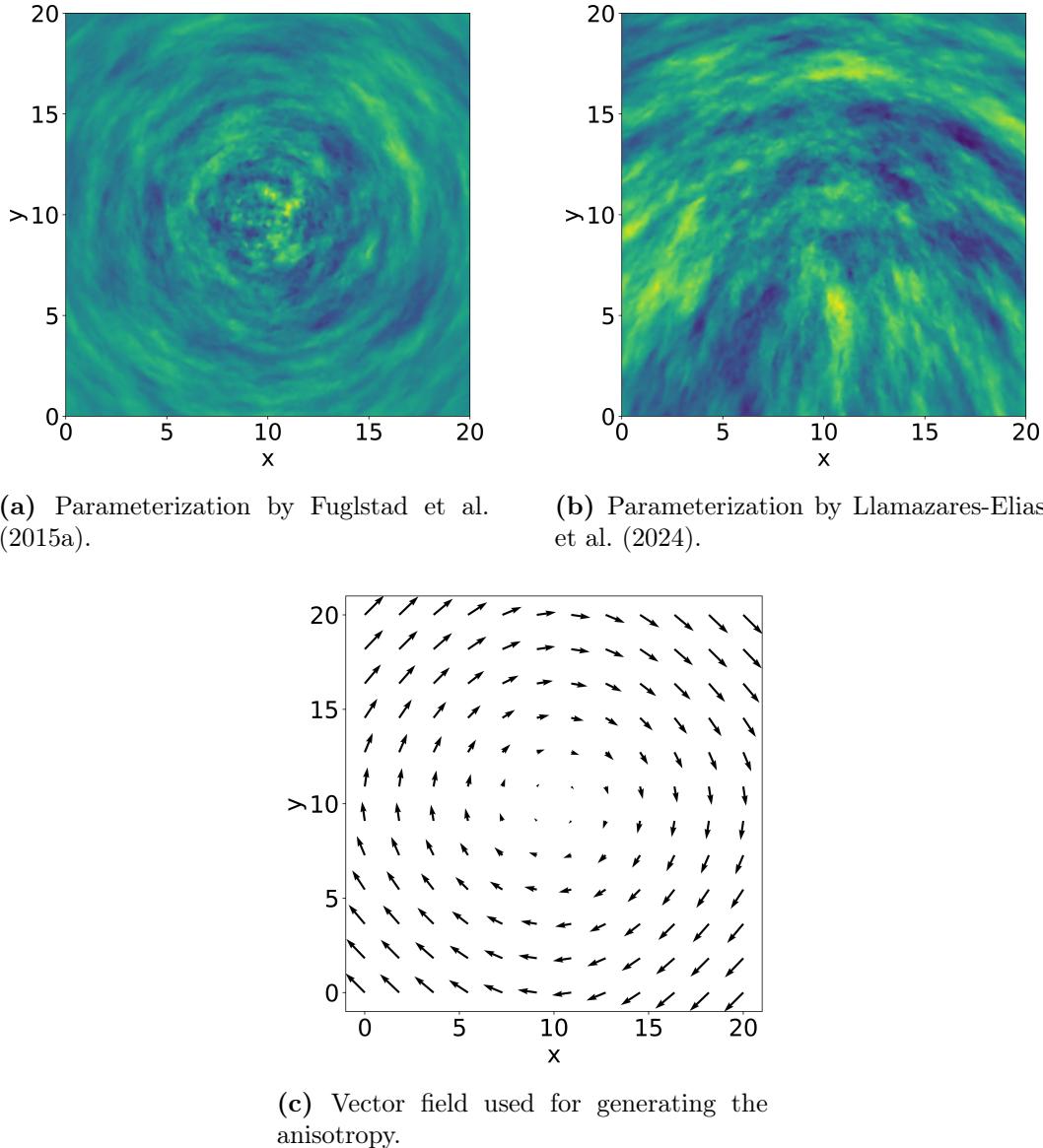
$$f(\mathbf{s}) = \alpha_0 + f_{\text{NS}}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D}.$$

Following the procedure proposed in Fuglstad et al. (2015b) we choose to give the  $f_{\text{NS}}$  a spline-like penalty according to the distribution generated from the SPDE

$$\begin{cases} -\Delta f_{\text{NS}}(\mathbf{s}) = \mathcal{W}(\mathbf{s})/\sqrt{\tau_f}, & \mathbf{s} \in \mathcal{D}, \\ \nabla f_{\text{NS}}(\mathbf{s}) \cdot \mathbf{n} = 0, & \mathbf{s} \in \partial\mathcal{D}. \end{cases} \quad (3.7)$$

Here,  $\tau_f > 0$  is a selected value controlling the penalty, determining how much  $f_{\text{NS}}$  is allowed to vary from a constant function. A larger  $\tau_f$  gives a stricter penalty on the non-stationarity, resulting in the  $f_{\text{NS}}$  term having less influence. The zero Neumann boundary condition is applied to restrict the resulting distribution, making it only invariant to the addition of a constant function (Fuglstad et al., 2015b).

To decompose the non-stationarity into a controllable set of basis functions, begin by choosing a domain  $[A_1, A_2] \times [B_1, B_2] \supseteq \mathcal{D}$ , and denote  $A = A_2 - A_1$  and  $B = B_2 - B_1$ . Let  $E = \{(0, 1), (1, 0), (1, 1), \dots\}$  be a set specifying the number basis elements, which also determines the number of coefficients which needs to be estimated. The  $f_{\text{NS}}$  is expanded as a linear combination of basis functions



**Figure 3.3:** Example of how the different parameterization of the diffusion matrix corresponds to the underlying vector field. Note that since the two parameterizations controls this strength of the anisotropic effect in different ways, the strength is different between the two realizations.

$\{f_{kl}\}_{(k,l) \in E}$ . For convenience, split

$$f_{kl}(x, y) = g_k(x)h_l(y), \quad (k, l) \in E, \quad (x, y) \in [A_1, A_2] \times [B_1, B_2],$$

into separate bases for the  $x$ - and  $y$ -coordinate. Due to the selected penalty in Equation (3.7), we choose our basis functions

$$g_l(x) = \frac{1}{M_k} \cos\left(\frac{k\pi(x - A_1)}{A}\right) \quad \text{and} \quad h_l(y) = \frac{1}{N_l} \cos\left(\frac{l\pi(y - B_1)}{B}\right), \quad (3.8)$$

which are the eigenfunctions of  $\Delta$ . Here,  $M_k$  and  $N_l$  are normalizing constants determined by integrating the corresponding cos-terms over their respective domains. See that these basis functions always have zero derivatives on the boundaries, naturally satisfying the boundary conditions. The function  $f$  can therefore be decomposed as

$$f(\mathbf{s}) \approx \alpha_0 + \sum_{(k,l) \in E} \alpha_{kl} g_k(x)h_l(y), \quad \mathbf{s} = (x, y) \in [A_1, A_2] \times [B_1, B_2], \quad (3.9)$$

where  $\alpha_{kl} \in \mathbb{R}$  are coefficients determining the contribution of each basis element  $f_{kl}$ .

### 3.4.2 Computing the penalties

For determining how to penalize the non-stationarity, we can analytically solve the SPDE in Equation (3.7) analytically using the basis representation. This will yield the distribution of the coefficients controlling the non-stationarity, whose likelihood can be used as a penalizing term. Let  $\boldsymbol{\alpha}$  denote the vector containing all the  $\alpha_{kl}$ -coefficients for  $(k, l) \in E$ . Inserting into the SPDE yields

$$-\Delta \left( \alpha_0 + \sum_{(k,l) \in E} \alpha_{kl} g_k(x)h_l(y) \right) = \frac{1}{\sqrt{\tau_f}} \sum_{(k,l) \in E} z_{kl} g_k(x)h_l(y),$$

with  $z_{kl} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . The right-hand side follows from expanding the Gaussian white noise in terms of the basis functions. Using the linearity and that the basis elements are orthogonal, solving the SPDE is equivalent finding the  $\boldsymbol{\alpha}$  satisfying

$$-\Delta (\alpha_0 + \alpha_{kl} g_k(x)h_l(y)) = \alpha_{k,l} \left[ \left( \frac{\pi k}{A} \right)^2 + \left( \frac{\pi l}{B} \right)^2 \right] g_k(x)h_l(y) = \frac{z_{kl}}{\sqrt{\tau_f}} g_k(x)h_l(y),$$

for all  $(k, l) \in E$ . Cancelling the cosine terms on both sides and solving for  $\alpha_{k,l}$  we find

$$\alpha_{kl} = \frac{1}{\sqrt{\tau_f}} \left[ \left( \frac{\pi k}{A} \right)^2 + \left( \frac{\pi l}{B} \right)^2 \right]^{-1} z_{kl}.$$

This yields the distribution  $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\text{NS}}^{-1}/\tau_f)$  where  $\mathbf{Q}_{\text{NS}}$  is diagonal matrix with elements  $\left[ \left( \frac{\pi k}{A} \right)^2 + \left( \frac{\pi l}{B} \right)^2 \right]^2$  corresponding to the ordering of elements in  $\boldsymbol{\alpha}$ .



# CHAPTER 4

---

## Discretization

---

### 4.1 Outline of discrete approach

This chapter will discuss the discretization of the proposed model from Equation (3.3). In particular, we want to solve the SPDE

$$L^\beta(\mathbf{s})u(\mathbf{s}) \stackrel{\text{def}}{=} (\kappa^2(\mathbf{s}) - \nabla \cdot \mathbf{H}(\mathbf{s})\nabla)^\beta u(\mathbf{s}) = \tau(\mathbf{s})\mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D},$$

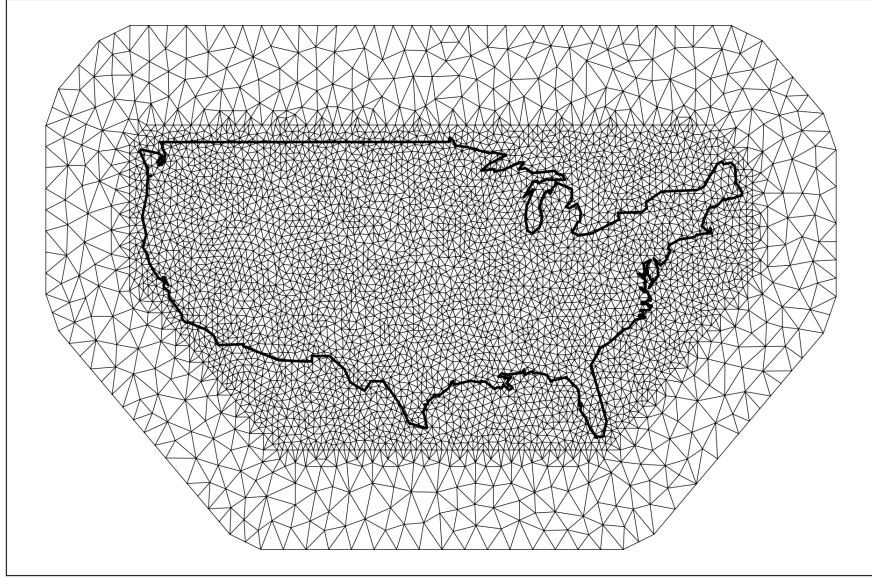
in a way that yields a GMRF approximation with a sparse precision matrix  $\mathbf{Q}$ . Following the method proposed by Lindgren et al. (2011), we do this using FEM. The FEM also provides a natural framework for incorporating the non-stationary parameters, as these can be included directly in the finite element matrices.

The complicating part of solving the SPDE is the fractional power  $\beta$ . For a non-fractional smoothness  $\beta \in \mathbb{N}$ , the  $L^\beta$  is a local operator. This results in the FEM yielding a sparse precision matrix  $\mathbf{Q}$ . On the other hand, for a fractional  $\beta \notin \mathbb{N}$ , the  $L^\beta$  is no longer local, and we cannot expect the resulting precision matrix to be sparse. Thus, incorporating the fractional smoothness requires an alternative approach, and here we utilize an approximation proposed by Bolin and Kirchner (2020).

### 4.2 Finite element method

#### 4.2.1 Discretizing the SPDE

The FEM is a numerical technique for solving partial differential equations (PDE). It works by discretizing the domain into smaller, simpler parts called *elements*. Within each element, the solution is approximated using a set of simple basis func-



**Figure 4.1:** Example of triangulation generated by fmesher (Lindgren, 2024) using observations within the conterminous U.S. The triangulation has 4388 vertices and 8664 triangles. Border is drawn on top to illustrate the area of interest.

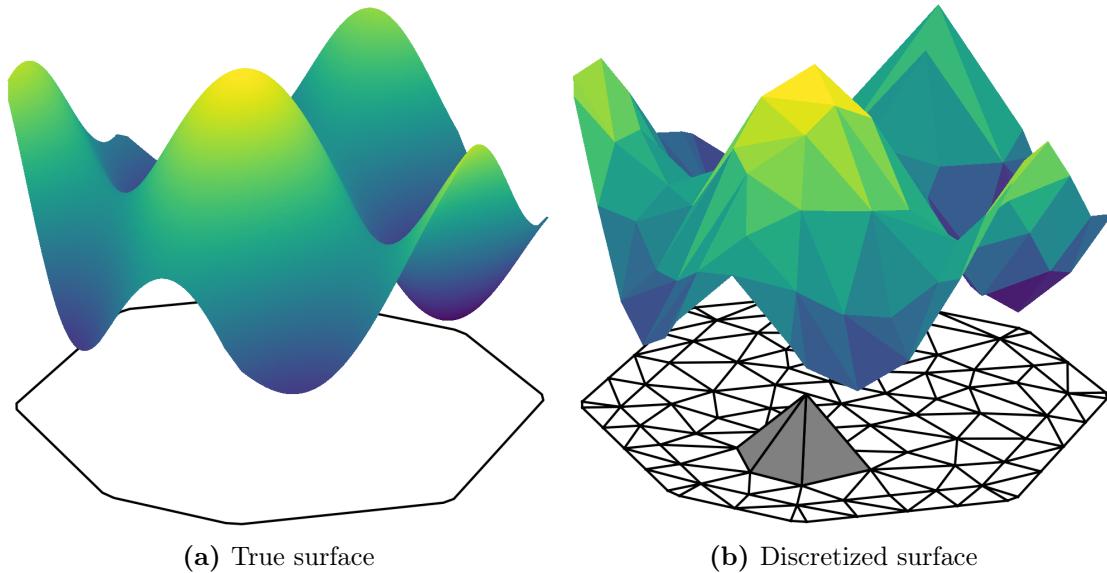
tions, which in this thesis are linear polynomials. By assembling the contributions from all elements, a system of algebraic equations is formed, which can be solved to obtain an approximate solution to the PDE.

Figure 4.1 displays one example of a domain discretized into triangles, called a *triangulation*. Denote the triangulation as  $\mathcal{T}_h$ , where  $h = \max_{T \in \mathcal{T}_h} \text{diam}(T)$  indicates the size of the triangles. The number  $n_h$  is specified by the number of vertices in the triangulation. Observe that the triangulated domain has two regions, with the inner having smaller triangles than the outer. In practice, only the inner part of the domain which is of interest. However, Lindgren et al. (2011) suggested expanding the domain by approximately two times the correlation range. This is done to reduce boundary effects when applying Neumann boundary conditions. To lower the computational cost, the triangulation in the outer region is made coarser than in the inner.

For every triangle in the triangulation, a linear basis function  $\varphi$  with support on the triangle is defined. The set of all such basis functions,  $\{\varphi_i\}_{i=1}^{n_h}$ , forms a basis for the piecewise linear function space, which is compactly supported and finite-dimensional. Figure 4.2b illustrates how the linear functions related to a chosen vertex grows out of the mesh, and how they together form a piecewise linear approximation of the surface in Figure 4.2a.

The main idea of the FEM is to determine some approximation  $u_h$  of the true solution  $u$  using the basis functions. For multivariate Gaussian weights  $\mathbf{u} = (u_1, \dots, u_{n_h})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$  with unknown precision  $\mathbf{Q}$ , this approximation can be expressed as

$$u_h(\mathbf{s}) = \sum_{i=1}^{n_h} u_i \varphi_i(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D}. \quad (4.1)$$



**Figure 4.2:** Discretization of a smooth surface using linear basis functions. Mesh plotted beneath the discretization. The gray pyramid consists of all non-zero basis functions for a selected vertex.

Solving the SPDE therefore reduces to determining the precision matrix  $\mathbf{Q}$ . Although how to do this has already been explained in Lindgren et al. (2011), we feel it is necessary so repeat some key ideas before the later discussion on how to handle the fractional power.

From the SPDE in Equation (3.3), let  $\beta = 1$  and consider the action of the left- and right-hand sides on test functions chosen as the basis elements  $\{\varphi\}_{i=1}^{n_h}$ . Then, by the independence of the basis elements, the SPDE can be unraveled into the system

$$\left[ \langle \varphi_i, (\kappa^2 - \nabla \cdot \mathbf{H} \nabla) u \rangle_{\mathcal{D}} \right]_{i=1,\dots,n_h} \stackrel{d}{=} \left[ \langle \varphi_i, \tau \mathcal{W} \rangle_{\mathcal{D}} \right]_{i=1,\dots,n_h}.$$

The spatial dependence is again omitted for notational simplicity. As discussed in Section 2.2, the equality in distribution must be understood in the context of the generalized fields, meaning the two sides must have the same expectation and covariance. Using integration by parts and applying our Neumann boundary conditions, the term  $-\langle \varphi_i, \nabla \cdot \mathbf{H} \nabla u \rangle_{\mathcal{D}}$  simplifies to  $\langle \nabla \varphi_i, \mathbf{H} \nabla u \rangle_{\mathcal{D}}$ . Further, by replacing  $u$  by the discretization  $u_h$ , we obtain

$$\left[ \sum_{j=1}^{n_h} u_j (\langle \varphi_i, \kappa^2 \varphi_j \rangle_{\mathcal{D}} + \langle \nabla \varphi_i, \mathbf{H} \nabla \varphi_j \rangle_{\mathcal{D}}) \right]_{i=1,\dots,n_h} \stackrel{d}{=} [\langle \varphi_i, \tau \mathcal{W} \rangle_{\mathcal{D}}]_{i=1,\dots,n_h} \quad (4.2)$$

Notice that since the basis functions  $\{\varphi_i\}_{i=1}^{n_h}$  are only non-zero within their corresponding element, most of the inner products in Equation (4.2) become zero. This is what will later give a GMRF approximation with a sparse precision matrix.

### 4.2.2 Computing the non-stationary finite element matrices

To solve Equation (4.2), we can construct a matrix system by computing the inner products. However, observe that the inner products  $\langle \varphi_i, \kappa^2(\mathbf{s})\varphi_j \rangle_{\mathcal{D}}$ ,  $\langle \varphi_i, \tau(\mathbf{s})\mathcal{W} \rangle_{\mathcal{D}}$  and  $\langle \nabla \varphi_i, \mathbf{H}(\mathbf{s})\nabla \varphi_j \rangle_{\mathcal{D}}$  all contain spatially varying parameters. To simplify the inner products we assume the parameters to be constant within the support of the individual basis functions (Lindgren et al., 2011). For the term with the non-stationary  $\kappa^2$ , this gives

$$\langle \varphi_i, \kappa^2(\mathbf{s})\varphi_j \rangle_{\mathcal{D}} \approx \kappa^2(\mathbf{s}_j^*) \langle \varphi_i, \varphi_j \rangle_{\mathcal{D}}, \quad i, j = 1, \dots, n_h,$$

where  $\mathbf{s}_j^*$  is defined as the centroid on the support of  $\varphi_j$ . To assure the sparsity of the precision matrix, we also follow Lindgren et al. (2011) by replacing  $\langle \varphi_i, \varphi_j \rangle_{\mathcal{D}}$  with  $\langle 1, \varphi_j \rangle_{\mathcal{D}}$ . This is a common technique in FEM called mass lumping, and was in Bolin and Lindgren (2013) shown to have little effect on the resulting approximation. Define the  $n_h \times n_h$  matrix  $\mathbf{C}$  with elements

$$\mathbf{C}_{ii} = \langle 1, \varphi_i \rangle_{\mathcal{D}}, \quad i = 1, \dots, n_h,$$

which is referred to as the *mass matrix*.

Handling the non-stationary diffusion matrix can also be done by assuming it to be constant over the support of the basis functions, but we are no longer able to move the matrix out of the inner product. Instead, define the  $n_h \times n_h$  *stiffness matrix*  $\mathbf{G}$  as

$$\mathbf{G}_{ij} = \langle \nabla \varphi_i, \mathbf{H}(\mathbf{s}_j^*) \nabla \varphi_j \rangle_{\mathcal{D}}, \quad i, j = 1, \dots, n_h.$$

Explicit formulas for computing the matrices  $\mathbf{C}$  and  $\mathbf{G}$  can be found in Lindgren et al. (2011, Appendix A). Finally, corresponding to the operator  $L = \kappa^2 - \nabla \cdot \mathbf{H} \nabla$ , define  $\mathbf{L} = \mathbf{D}_{\kappa^2} \mathbf{C} + \mathbf{G}$ . Here,  $\mathbf{D}_{\kappa^2}$  is a diagonal matrix with the values of  $\kappa^2(\mathbf{s}_i^*)$  on the diagonal.

To compute the matrix corresponding to the right-hand side of Equation (4.2), we refer back to Equation (2.4). This equation shows that Gaussian white noise produces jointly Gaussian random variables, with their covariance determined by the test functions they act upon. In particular, assuming  $\tau$  to be constant over the support of the basis functions we see

$$\begin{aligned} \text{Cov} [\langle \varphi_i, \tau \mathcal{W} \rangle_{\mathcal{D}}, \langle \varphi_j, \tau \mathcal{W} \rangle_{\mathcal{D}}] &\approx \tau(\mathbf{s}_i^*) \tau(\mathbf{s}_j^*) \langle \varphi_i, \varphi_j \rangle_{\mathcal{D}} \\ &\approx \tau^2(\mathbf{s}_j^*) \mathbf{C}_{ij}, \quad i, j = 1, \dots, n_h \end{aligned}$$

Hence, the matrix representing the right-hand side will be  $\mathbf{D}_{\tau^2} \mathbf{C}$ , where  $\mathbf{D}_{\tau^2}$  is a diagonal matrix defined similarly as for  $\mathbf{D}_{\kappa^2}$ .

### 4.2.3 Computing non-fractional precision matrix

Utilizing the non-stationary finite element matrices, the solution of the system in Equation (4.2) can now be computed. Define the discrete functions

$$f_h(\mathbf{s}) = \sum_{i=1}^{n_h} w_{f,i} \varphi_i(\mathbf{s}) \quad \text{and} \quad g_h(\mathbf{s}) = \sum_{i=1}^{n_h} w_{g,i} \varphi_i(\mathbf{s}),$$

with weights  $\mathbf{w}_f = (w_{f,1}, \dots, w_{f,n_h})^T$  and  $\mathbf{w}_g = (w_{g,1}, \dots, w_{g,n_h})^T$ . Inserting into the system we see

$$\langle f_h, Lu_h \rangle_{\mathcal{D}} = \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} w_{f,i} u_j (\langle \varphi_i, \kappa^2 \varphi_j \rangle_{\mathcal{D}} + \langle \nabla \varphi_i, \mathbf{H} \nabla \varphi_j \rangle_{\mathcal{D}}) \approx \mathbf{w}_f^T \mathbf{L} \mathbf{u},$$

and

$$\text{Cov} [\langle f_h, Lu_h \rangle_{\mathcal{D}}, \langle g_h, Lu_h \rangle_{\mathcal{D}}] \approx \mathbf{w}_f^T \mathbf{L} \text{Cov} [\mathbf{u}, \mathbf{u}] \mathbf{L}^T \mathbf{w}_g = \mathbf{w}_f^T \mathbf{L} \mathbf{Q}^{-1} \mathbf{L}^T \mathbf{w}_g.$$

For the covariance of the right-hand side, we obtain

$$\begin{aligned} \text{Cov} [\langle f_h, \tau \mathcal{W} \rangle_{\mathcal{D}}, \langle g_h, \tau \mathcal{W} \rangle_{\mathcal{D}}] &\approx \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} w_{f,i} w_{g,j} \tau(\mathbf{s}_i^*) \tau(\mathbf{s}_j^*) \langle \varphi_i, \varphi_j \rangle_{\mathcal{D}} \\ &\approx \mathbf{w}_f^T \mathbf{D}_{\tau^2} \mathbf{C} \mathbf{w}_g. \end{aligned}$$

Thus, equation (4.2) has equality in distribution when  $\mathbf{Q} = \mathbf{L}^T (\mathbf{D}_{\tau^2} \mathbf{C})^{-1} \mathbf{L}$ . As we apply mass lumping to  $\mathbf{C}$ , the precision matrix will be sparse (Lindgren et al., 2011).

Following the procedure described in Lindgren et al. (2011, Appendix D), similar sparse matrices can be found for SPDEs with  $\beta \in \{2, 3, \dots\}$ . The details behind this are a bit technical, but is done by iteratively by utilizing precision matrices with lower smoothnesses. For completeness, the matrices become

$$\begin{cases} \mathbf{Q}_1 = \mathbf{L}^T (\mathbf{D}_{\tau^2} \mathbf{C})^{-1} \mathbf{L}, & \beta = 1, \\ \mathbf{Q}_{\beta} = \mathbf{L}^T \mathbf{C}^{-1} \mathbf{Q}_{\beta-1} \mathbf{C}^{-1} \mathbf{L}, & \beta = 2, 3, \dots \end{cases} \quad (4.3)$$

Although the expressions allow us to cover a range of different smoothnesses, the fixed values set clear limitations which are not necessarily coherent with the true nature of the GRFs we want to model. For example, a field with an exponential covariance-structure, corresponding to  $\beta = 0.75$ , cannot be approximated using Equation (4.3). Therefore, to allow all  $\beta \geq 1/2$ , another approach is required. Be aware that in the original approach, Lindgren et al. (2011) did also present methods for approximating the fixed smoothnesses  $\beta \in \{0.5, 1.5, \dots\}$ . However, we choose to handle these using the alternative approach.

## 4.3 Approximating the fractional SPDE

### 4.3.1 Fundamentals

To account for fractional smoothness parameters  $\beta \notin \mathbb{N}$ , Bolin and Kirchner (2020) proposed the *rational SPDE approach*. This section will highlight the main ideas of the method, but will primarily focus on how to apply the method to our SPDE. The reader is referred to Bolin et al. (2020) and Bolin and Kirchner (2020) for a more in-depth explanation of the theoretical details behind the approach.

Consider the fractional SPDE

$$L^{\beta} u = \tau \mathcal{W} \quad \text{with } 1/2 \leq \beta \notin \mathbb{N},$$

and where the spatial dependence is omitted for notational simplicity. Following the same discretization as in Section 4.2, we can heuristically think of  $\mathbf{C}^{-1}\mathbf{L}$  as the discrete version of the differential operator  $L$ . Hence, for Gaussian distributed weights  $\mathbf{u}$ , the discretized equation can be written as the system

$$(\mathbf{C}^{-1}\mathbf{L})^\beta \mathbf{u} = (\mathbf{C}^{-1}\mathbf{D}_{\tau^2})^{1/2} \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (4.4)$$

Further, inverting the term on the right-hand side gives

$$\mathbf{u} = (\mathbf{C}^{-1}\mathbf{L})^{-\beta} (\mathbf{C}^{-1}\mathbf{D}_{\tau^2})^{1/2} \mathbf{z}. \quad (4.5)$$

Notice that if we can approximate the inverse fractional term  $(\mathbf{C}^{-1}\mathbf{L})^{-\beta}$ , then Equation (4.5) can be used to determine  $\mathbf{u}$ . Bolin and Kirchner (2020) proposed doing this using a *rational approximation*.

To construct the rational approximation, begin by splitting the smoothness  $\beta = (\beta - m_\beta) + m_\beta$  into a fractional part and a non-fractional part. Here  $m_\beta = \max\{1, \lfloor \beta \rfloor\}$  where  $\lfloor \cdot \rfloor$  denotes the floor function. Consider a rational approximation  $r(x)$  of the function  $x^{\beta-m_\beta}$ . Then, we can express  $x^{-\beta}$  as

$$\begin{aligned} x^{-\beta} &\approx r(x^{-1}) x^{-m_\beta} = \frac{c_m \prod_{i=0}^m (1 - r_{1,i}x)}{b_{m+1} x^{-1} \prod_{i=0}^{m+1} (1 - r_{2,i}x)} x^{-m_\beta} \\ &= \frac{c_m \prod_{i=0}^m (1 - r_{1,i}x)}{b_{m+1} x^{m_\beta-1} \prod_{i=0}^{m+1} (1 - r_{2,i}x)}. \end{aligned} \quad (4.6)$$

In this expression, the coefficients  $c_m$ ,  $b_{m+1}$ ,  $\{r_{1,i}\}_{i=0}^m$  and  $\{r_{2,i}\}_{i=0}^{m+1}$  are determined by the rational approximation  $r(x)$ . The integer  $m$  is selected manually, and controls the degree of the approximation. The motivation behind using polynomial of degrees  $m$  in the numerator and  $m+1$  in the denominator is explained in Bolin and Kirchner (2020). This ensures that the rational approximation is universally applicable for all values of  $\beta > 1/2$ . The products show up as a result of factorizing polynomials of the rational approximation in terms of their roots, which avoids  $x$  being raised to any power.

### 4.3.2 Computing the fractional precision matrix

Now that we have found a way of approximating the function  $x^{-\beta}$ , we must find a way of computing it in a way that solves the matrix system. For calculating the rational coefficients  $c_m$ ,  $b_{m+1}$ ,  $\{r_{1,i}\}_{i=0}^m$  and  $\{r_{2,i}\}_{i=0}^{m+1}$ , Bolin and Kirchner (2020) suggested using the Chenshaw-Lord Chebyshev-Padé algorithm (Baker Jr and Graves-Morris, 1996). This requires bounding the spectrum of  $(\mathbf{C}^{-1}\mathbf{L})^{-1}$  to a known interval  $J_h$ . For this purpose we rescale  $\mathbf{L} = \mathbf{D}_{\kappa^2} \mathbf{C} + \mathbf{G}$  to

$$\mathbf{L} = \mathbf{C} + \mathbf{D}_{\kappa^2}^{-1} \mathbf{G},$$

where  $\mathbf{I}$  is the identity matrix. By considering the eigenvalues of the rescaled  $(\mathbf{C}^{-1}\mathbf{L})^{-1}$  we now have  $J_h \subset [0, 1]$  for all mesh sizes  $h$ . This rescaling also improves the efficiency of the approximation, as for any given  $\beta$ , the rational coefficients be pre-computed and looked up from a table when needed.

Due to technical details provided in Bolin and Kirchner (2020), we can insert the

matrix  $\mathbf{C}^{-1}\mathbf{L}$  from Equation (4.4) into the rational approximation. Define the matrices

$$\mathbf{P}_l = b_{m+1} \mathbf{D}_{\kappa^2}^\beta \mathbf{C} (\mathbf{C}^{-1}\mathbf{L})^{m\beta-1} \mathbf{L}_{m+1} \quad \text{and} \quad \mathbf{P}_r = c_m \prod_{i=1}^m (\mathbf{I} - r_{1,i} \mathbf{C}^{-1}\mathbf{L}), \quad (4.7)$$

where  $\mathbf{L}_{m+1} = \prod_{i=1}^{m+1} (\mathbf{I} - r_{2,i} \mathbf{C}^{-1}\mathbf{L})$ . Observe that the matrices  $\mathbf{P}_l$  and  $\mathbf{P}_r$  correspond to the denominator and numerator in Equation (4.6), except that the scaling  $\mathbf{D}_{\kappa^2}^\beta$  is applied to  $\mathbf{P}_l$  to revert the effect of the rescaled  $\mathbf{L}$ . Additionally, as they are constructed using products of sparse matrices, both  $\mathbf{P}_l$  and  $\mathbf{P}_r$  will be sparse as long as  $m$  is reasonably small.

Going back to Equation (4.5), the discrete solution  $\mathbf{u}$  can now be approximated as

$$\mathbf{u} \approx \mathbf{P}_l^{-1} \mathbf{P}_r (\mathbf{C}^{-1}\mathbf{D}_{\tau^2})^{1/2} \mathbf{z}.$$

To avoid having to invert  $\mathbf{P}_l$ , we can split the equality into a system of equations. Introduce Gaussian weights  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$ , and consider

$$\mathbf{P}_l \mathbf{w} = (\mathbf{C}^{-1}\mathbf{D}_{\tau^2})^{1/2} \mathbf{z}, \quad (4.8)$$

$$\mathbf{u}^R = \mathbf{P}_r \mathbf{w}, \quad (4.9)$$

where  $\mathbf{u}^R$  denotes the rational approximation of the solution. Following a similar procedure as in Section 4.2.3, we take the covariances of both sides of Equation (4.8) and solve for the precision. This gives the matrix

$$\mathbf{Q} = \mathbf{P}_l^T (\mathbf{D}_{\tau^2}\mathbf{C})^{-1} \mathbf{P}_l. \quad (4.10)$$

The fractional solution is therefore approximated with  $\mathbf{u}^R = \mathbf{P}_r \mathbf{w}$ , where  $\mathbf{w}$  has precision given in Equation (4.10). For simplicity, we will in the remaining text denote the rational approximation  $\mathbf{u}^R$  as  $\mathbf{u}$ .

## 4.4 Summary of discrete approach

Before moving on, we briefly want to recap the full approach for discretizing the model. Begin by computing the matrices  $\mathbf{C}$ ,  $\mathbf{G}$ ,  $\mathbf{D}_{\kappa^2}$  and  $\mathbf{D}_{\tau^2}$ . Then, depending on the smoothness parameter  $\beta$ , proceed with one of the two approaches:

- If  $\beta \in \mathbb{N}$ , choose  $\mathbf{P}_r = \mathbf{I}$  where  $\mathbf{I}$  is the  $n_h \times n_h$  identity matrix. Define  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$  with  $\mathbf{Q}$  computed according to formulas in Equation (4.3).
- If  $\beta \notin \mathbb{N}$ , choose degree  $m \in \mathbb{N}$  and compute matrices  $\mathbf{P}_l$  and  $\mathbf{P}_r$  according to formulas in Equation (4.7). Define  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$  with  $\mathbf{Q} = \mathbf{P}_l^T (\mathbf{D}_{\tau^2}\mathbf{C})^{-1} \mathbf{P}_l$ .

For both the fractional and non-fractional cases define  $\mathbf{u} = \mathbf{P}_r \mathbf{w}$ , and use this when doing inference.



# CHAPTER 5

---

## Inference

---

### 5.1 Hierarchical model

In this thesis, the main goal of constructing the spatial regression model is to use it for inference. We focus on two key tasks, where the first is to estimate the model parameters from observed data, and the second is to predict the field at unobserved locations. Constructing a hierarchical model is a natural way for handling these tasks. It provides a structure for representing the relationships between different components of the data, and allows uncertainty to be incorporated at each level. From the hierarchical model, relevant posterior distributions for the parameters and latent variables can be computed using Bayes theorem.

Following the construction of the regression model as presented in Section 2.3, assume a set of  $n$  observations  $\mathbf{y} = (y_1, \dots, y_n)^T$  made at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{D}$ . Describe the covariates at location  $\mathbf{s}_i$  by a  $p$ -dimensional row vector  $\mathbf{x}(\mathbf{s}_i)^T$ , and collect them in an  $n \times p$  dimensional matrix  $\mathbf{X} = [\mathbf{x}(\mathbf{s}_1) \ \dots \ \mathbf{x}(\mathbf{s}_n)]^T$ . Let the spatial effect  $u$  be modelled and parameterized as described in Chapter 3.

Discretizing the model, construct a triangulation of the domain with linear basis functions  $\{\varphi_j\}_{j=1}^{n_h}$ . To use the GMRF approximation  $\mathbf{u} = \mathbf{P}_r \mathbf{w}$  for computations, introduce the  $n \times n_h$  matrix  $\mathbf{A}$  with elements  $A_{ij} = \varphi_j(\mathbf{s}_i)$ , giving a mapping between the observed locations and the triangulation. As a result, the discretized regression model becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbf{P}_r \mathbf{w} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} | \sigma_e^2 \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}),$$

where  $\boldsymbol{\beta}$  is a  $p$ -variate vector for the coefficients of the covariates and  $\boldsymbol{\epsilon}$  is unstructured noise.

The underlying model for the  $p$ -dimensional random variable  $\beta$  is made proper by introducing a Gaussian penalty  $\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}/\tau_\beta)$ . Here,  $\tau_\beta$  is a manually selected value. Further, denote the coefficients that control the non-stationary  $\log(\kappa^2)$ ,  $\log(\sigma^2)$ ,  $v_x$  and  $v_y$  as  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ . Let the corresponding penalty parameters be  $\tau_1, \tau_2, \tau_3$  and  $\tau_4$ , and note that these must also be manually selected. The full hierarchical model becomes

$$\begin{aligned} \text{Level 1: } & \mathbf{y}|\beta, \mathbf{w}, \sigma_e^2 \sim \mathcal{N}(\mathbf{X}\beta + \mathbf{A}\mathbf{P}_r \mathbf{w}, \sigma_e^2 \mathbf{I}), \\ \text{Level 2: } & \mathbf{w}|\alpha_1, \alpha_2, \alpha_3, \alpha_4 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}), \\ \text{Level 3: } & \beta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}/\tau_\beta), \\ & \alpha_i | \tau_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\text{NS}}^{-1}/\tau_i) \quad \text{for } i = 1, 2, 3, 4. \end{aligned}$$

Notice that we for simplicity have parameterized all the non-stationary parameters using the same basis functions. Thus the same  $\mathbf{Q}_{\text{NS}}$  matrix is used for all the penalties. However, it is easy to generalize the model to different basis representations for the different non-stationary parameters. This which would only slightly alter a few of the following expressions.

## 5.2 Parameter estimation

We use a maximum likelihood estimator to estimate the parameters

$$\boldsymbol{\theta} = (\nu, \sigma_e^2, \alpha_1, \alpha_2, \alpha_3, \alpha_4).$$

For this the integrated likelihood is needed, where the latent field consisting of the fixed and spatial effects is integrated out. As both the parameters for the fixed effects and the spatial field by construction are Gaussian, this integration can be done explicitly. Begin by collecting the fixed effect and the spatial effect in  $\mathbf{z} = (\mathbf{w}^T, \beta^T)$ . Defining the matrices

$$\mathbf{S} = [\mathbf{A}\mathbf{P}_r \quad \mathbf{X}] \quad \text{and} \quad \mathbf{Q}_z = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \tau_\beta \mathbf{I} \end{bmatrix},$$

we from the hierarchical model obtain the distributions

$$\begin{aligned} \mathbf{y}|\mathbf{z}, \boldsymbol{\theta} & \sim \mathcal{N}(\mathbf{S}\mathbf{z}, \sigma_e^2 \mathbf{I}) \\ \mathbf{z}|\boldsymbol{\theta} & \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_z^{-1}). \end{aligned}$$

Denoting the posterior distribution as  $\pi(\boldsymbol{\theta}|\mathbf{y})$  and following the procedure described in Fuglstad et al. (2015b, Appendix C), we obtain the penalized log-likelihood

$$\begin{aligned} \log \pi(\boldsymbol{\theta}|\mathbf{y}) \propto & -\frac{1}{2} \sum_{i=1}^4 \tau_i \alpha_i^T \mathbf{Q}_{\text{NS}} \alpha_i + \frac{1}{2} \log |\mathbf{Q}_z| - \frac{n}{2} \log \sigma_e^2 \\ & - \frac{1}{2} \log |\mathbf{Q}_C| - \frac{1}{2} \boldsymbol{\mu}_C^T \mathbf{Q}_z \boldsymbol{\mu}_C \\ & - \frac{1}{2} \sigma_e^{-2} (\mathbf{y} - \mathbf{S}\boldsymbol{\mu}_C)^T (\mathbf{y} - \mathbf{S}\boldsymbol{\mu}_C), \end{aligned} \tag{5.1}$$

where  $\mathbf{Q}_C = \mathbf{Q}_z + \mathbf{S}^T \mathbf{S} / \sigma_e^2$  and  $\boldsymbol{\mu}_C = \mathbf{Q}_C^{-1} \mathbf{S}^T \mathbf{y} / \sigma_e^2$ . The MLE  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  will be the parameters maximizing this log-likelihood.

### 5.3 Prediction

To determine the predictive distributions for unmeasured locations, continue using the same construction as for the parameter estimation. From Bayes' rule we obtain the distribution

$$\begin{aligned}\pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}) &= \frac{\pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}) \pi(\mathbf{z}|\boldsymbol{\theta})}{\pi(\mathbf{y}|\boldsymbol{\theta})} \\ &\propto \pi(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}) \pi(\mathbf{z}|\boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{1}{2} \mathbf{z}^T \mathbf{Q}_C \mathbf{z} + \mathbf{z}^T \mathbf{S}^T \mathbf{y} / \sigma_e^2\right\}.\end{aligned}$$

Recognize this as a Gaussian distribution with mean  $\boldsymbol{\mu}_C$  and precision  $\mathbf{Q}_C$ , implying that  $\mathbf{z}|\mathbf{y}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_C, \mathbf{Q}_C^{-1})$ .

For a new set of  $k$  locations  $\mathbf{s}_1, \dots, \mathbf{s}_k \in \mathcal{D}$ , the mean and variance at the locations can be predicted using the density of  $\mathbf{z}|\mathbf{y}, \hat{\boldsymbol{\theta}}$ . Here,  $\hat{\boldsymbol{\theta}}$  is the MLE estimated using the observed data. Define the projection matrix for the predictions as

$$\mathbf{S}_p = [\mathbf{A}_p \mathbf{P}_r \quad \mathbf{X}_p],$$

where  $\mathbf{X}_p$  is the  $k \times p$  dimensional design matrix for the prediction locations, and  $\mathbf{A}_p$  is the  $k \times n_h$  dimensional matrix mapping the prediction locations to the triangulation. The predictive density for the new set of prediction locations therefore becomes

$$\mathbf{S}_p \mathbf{z} | \mathbf{y}, \hat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\mathbf{S}_p \boldsymbol{\mu}_C, (\mathbf{S}_p \mathbf{Q}_C \mathbf{S}_p^T)^{-1}\right).$$

The marginal variance related to each prediction is found on the diagonal of  $(\mathbf{S}_p \mathbf{Q}_C \mathbf{S}_p^T)^{-1}$ . However, as this will be a dense matrix, computing the full inverse is undesirable. We therefore use the algorithm mentioned in Section 2.4 to obtain the diagonal elements through computing the partial inverse.

### 5.4 Proposed optimization strategy

When computing the maximum likelihood estimator for the model parameters, the likelihood function in Equation (5.1) must be optimized numerically. By estimating the non-stationary parameters, the likelihood function will be more sensitive to local optimum (Fuglstad et al., 2015a), and we must therefore find a suitable optimization strategy that can handle both the fractional and non-stationary model. For the fractional model, Bolin and Kirchner (2020) estimated their parameters using a gradient-free algorithm. Although such an optimization algorithm can work as long as the initial values are close enough to the true values, we found it to not be very robust for estimating the non-stationary parameters. Utilizing a gradient-based optimization algorithm will likely improve the convergence, as it yields more accurate steps in the parameter space.

For the non-fractional case with  $\beta = 1$ , Fuglstad et al. (2015a, Appendix D) found analytical expressions for the gradient of the non-stationary model. However, due to the  $\mathbf{P}_l$  and  $\mathbf{P}_r$  matrices in the rational approximation, a gradient is no longer as easily derived. While attempting to analytically compute the derivatives through the chain-rule, we realized that a very appealing approach for handling this issue could be to use automatic differentiation. This is a technique for evaluating the partial derivative of a function specified by a computer program. It is often well-suited for functions that involve complex arithmetic operations, and is currently the best alternative we have found for computing the gradient of the likelihood.

One modern way of performing automatic differentiation is through the PyTorch library (Paszke et al., 2019). Although this library is mainly known for machine learning, the `autograd` feature can be applied for differentiating functions. Additionally, `pytorch.optim` contains many modern optimization algorithms. One major disadvantage with PyTorch is its limited support for sparse matrix operations. Even though many sparse matrix operations exist, it lacks the possibility of computing a sparse Cholesky factorization. Therefore, since the sparse matrices are factorized densely, performing the optimization in PyTorch results in higher runtimes. This restricts how fine we can make our meshes, which limits the amount of short range details the model can capture.

Another complicating issue is that the likelihood does not have well-defined partial derivative with respect to the smoothness  $\nu$ . The cause of this is that the precision  $\mathbf{Q}$  is approximated differently based on the size of the smoothness. Additionally, as mentioned in Section 4.3.2, the coefficients of the rational approximation are not computed for every new  $\nu$  value, but instead looked up from a pre-computed table. Therefore, since the likelihood cannot be differentiated with respect to  $\nu$ , its value must be fixed during the gradient-based optimization.

For getting the advantages of the gradient terms, while still capturing short-range details and estimating the smoothness, we propose a hybrid optimization consisting of two steps.

1. **Gradient-based optimization:** This step is designed to capture broader non-stationary trends in the data. Choose initial values based on the problem at hand. Use PyTorch for optimizing parameters with gradients computed via automatic differentiation. To accommodate the requirement of a fixed smoothness, we use multiple-starts procedure, where the optimization is repeated several times for different fixed smoothnesses. For Chapters 6 and 7, we use the set of values  $\{0.5, 0.75, 1, 1.25, 1.5\}$ , but these must be chosen based on the problem and th desired runtime. After completing optimization for all five smoothness values, select the parameters associated with the highest log-likelihood.
2. **Gradient-free optimization:** Use the optimized parameters from the previous step as initial values in a gradient-free optimization algorithm. Since we now can utilize a sparse matrix factorization, choosing a fine mesh allows us to capture short range details. Moreover, as the smoothness no longer needs to be fixed, this parameter  $\nu$  can also be optimized. For the optimization we use the COBYLA algorithm (Powell, 1998), as implemented in the NLOpt package

(Johnson, 2007).

Both Adam and COBYLA require termination criteria, and for simplicity we set the algorithms to stop after a pre-selected number of iterations. Since more complex models often need more optimization steps to converge, and convergence depends on the amount and quality of data, the exact number of iterations is determined experimentally on a case-to-case basis.

We want to stress that the hybrid approach currently serves as a temporary solution to the unresolved challenge of optimizing the fractional and non-stationary model. Improving the optimization can be the subject of future work, but is outside the scope of this thesis. Currently, the runtime of the approach is very high compared to when fitting a non-fractional model, as the gradient can in the non-fractional case be computed analytically without relying on PyTorch. Therefore, we will not analyze runtimes between different types of models, as it would not be a fair comparison.



# CHAPTER 6

---

## Simulation study

---

### 6.1 Motivation and goals

Justifying the added complexity of the proposed fractional and non-stationary model requires empirical evidence to demonstrate its benefits over simpler alternatives. This validation will be carried out through two complementary methods, a simulation study and an application to real-world data. The simulation study, through simulated data with known parameters, is primarily meant to assess the accuracy of the parameter estimation. On the other hand, the analysis of precipitation data will focus on predictions.

The proposed model offers greater flexibility than simpler models, and we expect that this will improve its parameter estimates and predictions in certain situations. However, the increased flexibility can also introduce restrictions, such as the amount or quality of data required to reliably estimate model parameters. Furthermore, the inclusion of additional parameters increases the risk of overfitting, which may lead to poor performance when applied to data from stationary or non-fractional model. The main questions we consider in the simulation study are therefore

- Q1: How robust is the proposed model to data from a non-fractional or stationary model, and how does this depend on the number of observations?
- Q2: For data from a fractional and non-stationary model, does the proposed model yield significantly better parameter estimates compared to simpler models? Are the predictions any better? How does this depend on the number of observations?
- Q3: How sensitive is the model to the signal-to-noise ratio in the data?

## 6.2 Evaluation

To generate data for the answering the research questions, we choose four true models describing the situations we want to investigate. Each model is visualized in the same sections as where their data is analyzed. In general, for the fractional models, a smoothness of  $\nu = 0.5$  is used. For the non-stationarity, only the diffusion matrix  $\mathbf{H}$  is spatially varying, and is parameterized using 8 basis functions. Additionally, in order to get a realistic comparison between the stationary and non-stationary models, the non-stationarity is deliberately chosen to not be too extreme. Consequently, it provides an opportunity to examine the nuances of stationary versus non-stationary models when the optimal choice is unclear.

The procedure for parameter estimation and prediction the same when answering Q1 and Q2, and is summarized by the following steps.

1. Using the true model, generate a realization on the fine mesh. From the inner part of the domain, randomly pick 2000 vertex-locations  $\mathbf{s}_1, \dots, \mathbf{s}_{2000}$  along with their corresponding values  $u_1, \dots, u_{2000}$ . Apply iid. noise with distribution  $\mathcal{N}(0, \sigma_e^2)$  to the data, simulating the observation error, giving  $y_1, \dots, y_{2000}$ .
2. Randomly split the 2000 observations into two equally sized sets  $D_{\text{est}}$  and  $D_{\text{pred}}$  used for parameter estimation and prediction evaluation, respectively. Further, divide  $D_{\text{est}}$  into the nested subsets

$$D_i = \{(\mathbf{s}_1, y_1), \dots, (\mathbf{s}_i, y_i)\} \quad \text{for } i \in \{125, 250, 500, 1000\}.$$

This nesting property of the subsets ensures that the observations included in smaller datasets are also part of larger ones. Let the data for evaluating the predictions be the remaining

$$\mathbf{s}_{\text{pred}} = (\mathbf{s}_{1001}, \dots, \mathbf{s}_{2000}) \quad \text{and} \quad \mathbf{y}_{\text{pred}} = (y_{1001}, \dots, y_{2000}).$$

3. For each considered model:
  - 3.1. Using the data in each subset  $D_i$  with  $i \in \{125, 250, 500, 1000\}$ , estimate the corresponding parameters  $\hat{\boldsymbol{\theta}}$ .
  - 3.2. With the estimated parameters  $\hat{\boldsymbol{\theta}}$ , predict the latent field  $u(\mathbf{s})$  at locations  $\mathbf{s} \in \mathbf{s}_{\text{pred}}$ . This gives the predicted means  $\hat{\mathbf{u}}$  and standard deviations  $\hat{\boldsymbol{\sigma}}$ , which can be used for evaluating the predictions by computing the RMSE and CRPS scores.
4. Repeat steps 1-4 a total of 25 times. Use this to compute the empirical means and standard deviations of the parameter estimates and predictive scores. Evaluate the bias of the parameter estimates by considering the difference between the true parameters and the empirical means.

For answering Q3, the steps are slightly altered. Instead of varying the number of observations, fix  $D_{\text{est}} = D_{500}$  and vary the  $\sigma_e^2$  of the noise applied to the data. Then estimate the model parameters for the different amounts of noise.

When presenting the estimated parameters, the estimates based on 500 observations are shown. We empirically found this to be a suitable amount of observations for describing how well the models estimate their parameters. Only the stationary parts of the non-stationary parameters are provided. Although this will not give insight into the estimated non-stationarity, it will be an indication of the model fit. For more interpretable results, the stationary parts of the estimated parameters are used to compute the iso-covariance curve describing the correlation ranges and anisotropy. As illustrated in Figure 3.1, we will provide estimates for the correlation ranges  $\rho_1$  and  $\rho_2$  along the main and secondary axes  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , in addition to the direction  $\theta$ .

### 6.3 Candidate models

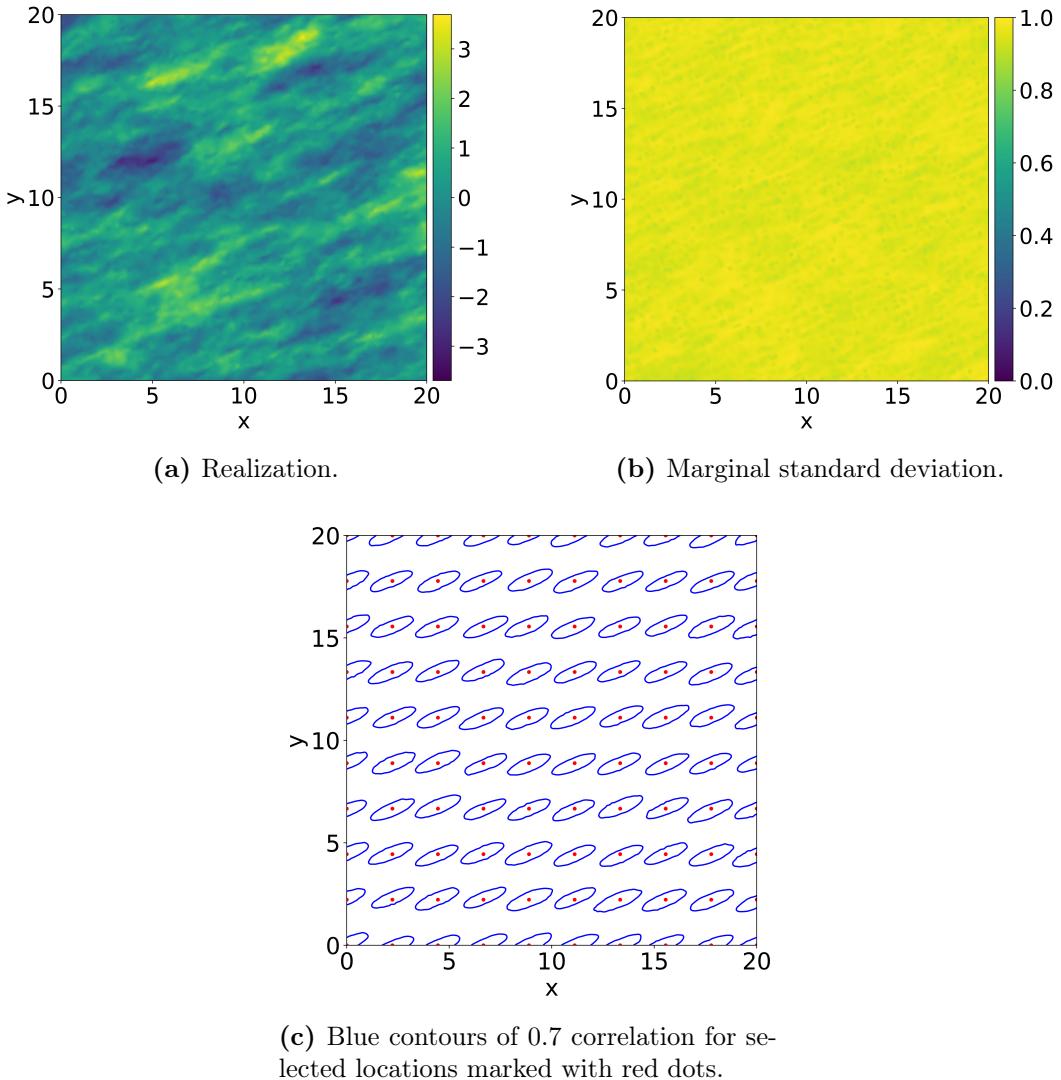
Four models are compared in the simulation study. All are based on the spatial regression model presented discussed in Section 2.3. The fixed effects  $\mathbf{x}^T \boldsymbol{\beta}$  are omitted, and only the GRF  $u$  is varied between models.

The simplest of the four models is a non-fractional and stationary model (NF-S), whose GRF has a fixed smoothness  $\nu = 1$  and stationary parameters. Making the parameters spatially varying, the non-fractional and non-stationary model (NF-NS) has a GRF with fixed  $\nu = 1$  and non-stationary range, anisotropy and marginal variance. The non-stationarity is parameterized using 8 basis functions, with penalty parameters  $\tau_i = 1$  for  $i = 1, 2, 3, 4$ . This only imposes a small penalty on the non-stationarity, as we are interested in seeing how prone the non-stationary models are to overfitting. Note that the non-stationarity is parameterized with the same amount of basis functions as in the true models. This means that the NF-NS is, in theory, able to perfectly replicate the non-stationary behavior present in the data. Moving on to fractional models, the GRF of the fractional and stationary model (F-S) has a fractional smoothness parameter and stationary parameters. The rational approximation has degree  $m = 2$ , as this is shown in Bolin and Kirchner (2020) to be sufficient for most cases. Lastly, the proposed fractional and non-stationary model (F-NS) is the most complicated, and has a GRF with both a fractional smoothness and spatially varying parameters. The non-stationarity is parameterized the same way as for NF-NS.

The parameters of all models are estimated using the hybrid approach described in Section 5.4. The coarse mesh for the first step consists of approximately 4000 vertices and 8100 triangles, while the fine mesh has 12000 vertices and 26000 triangles. From empirical testing, we found that 300 iterations for the Adam optimization and 600 for the NLOpt optimization were sufficient for determining the parameter estimates, so this is what is used when optimizing the likelihood. For the initial values of the optimization, the stationary parameters are assigned random values within  $\pm 50\%$  from the truth, while non-stationary parameters are set to zero.

## 6.4 Stationary and non-fractional data

Begin by considering a dataset generated by a fractional and stationary model. Figure 6.1 visualizes the model from which the data was sampled, where we see it exhibits a clear anisotropic behavior. Table 6.1 presents the estimated parameters. Both the F-S and F-NS estimates the smoothness relatively accurately, but F-NS has a larger empirical standard deviation. For the stationary components of NF-NS and F-NS, the estimated  $\rho_1$  and  $\rho_2$  parameters are biased. Yet, note that the true value is contained within the empirical standard deviation of F-NS. One possibility is that the bias in estimates of the non-stationary models can be attributed to the non-stationary parameters having an impact, meaning the models slightly overfit to the data. Additionally, for the F-NS, it estimates  $\sigma^2$  accurately, but the empirical standard deviation is larger than for the other models.



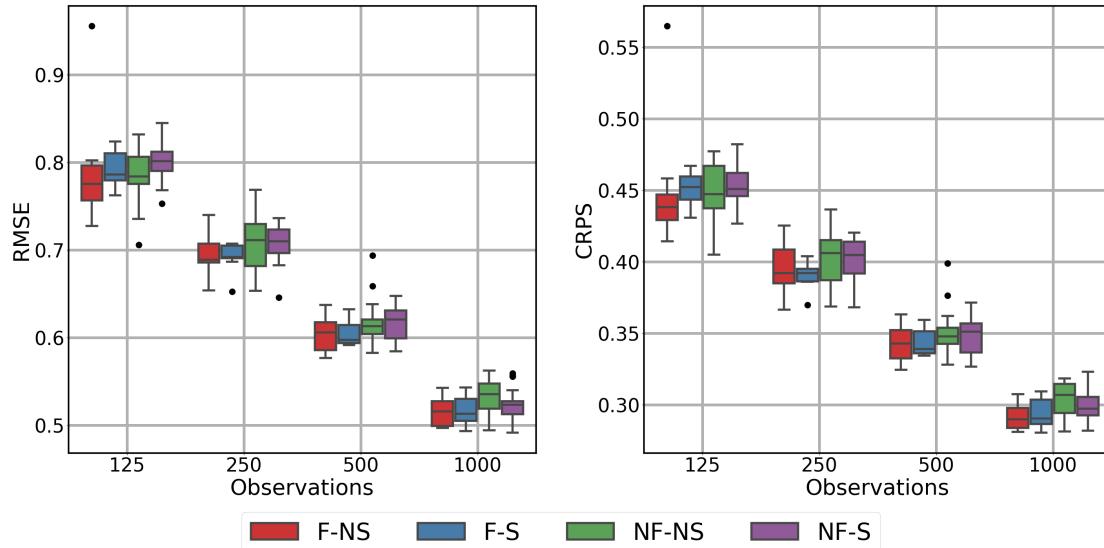
**Figure 6.1:** Visualizations of the fractional and stationary model.

Comparing the predictive scores, see from Figure 6.2 that the NF-NS and F-NS have approximately the same empirical mean and standard deviation as the NF-S and F-S. The only exception is one outlier in F-NS for 125 observations.

**Table 6.1:** Mean parameter estimates with standard deviation in parentheses. Data from a fractional and stationary model, and estimates based on 500 observations of a single realization.

Model	Parameters					
	$\nu$	$\rho_1$	$\rho_2$	$\theta$	$\sigma^2$	$\sigma_e^2$
True values	0.50	5.74	1.40	22.50	1.00	0.10
F-NS	0.55 (0.37)	6.69 (2.68)	1.96 (0.86)	22.94 (3.55)	1.07 (0.36)	0.10 (0.03)
F-S	0.51 (0.10)	5.99 (1.29)	1.53 (0.50)	22.52 (2.59)	0.97 (0.14)	0.08 (0.04)
NF-NS	1.00	4.26 (0.61)	1.43 (0.30)	22.82 (4.06)	0.89 (0.12)	0.13 (0.04)
NF-S	1.00	5.28 (0.81)	1.34 (0.19)	21.48 (2.54)	0.84 (0.11)	0.17 (0.03)

Interestingly, for only 125 observations, it appears that NF-NS and F-NS has slightly lower means than their stationary counterparts. Comparing the fractional and non-fractional models, the fractional models overall achieves slightly better mean scores than the non-fractional models. However, the difference is very small.



**Figure 6.2:** Scores for predictions using data from a stationary model with smoothness  $\nu = 0.5$ .

Considering data generated by a stationary and non-fractional model, we let all parameters remain the same, except setting the smoothness to  $\nu = 1$ . Table 6.2 displays the estimated parameters for the different models. Both the F-S and F-NS estimate the non-fractional smoothness parameter accurately. Further, similar as for the fractional case, the estimated correlation ranges of the NF-NS and F-NS are still biased. Yet, the parameter estimates for F-NS no longer has a much higher empirical standard deviation compared to the other models. Unexpectedly, see that the F-NS overall has a smaller estimated bias than the NF-NS, even though the model had to estimate the smoothness.

Overall, the estimated parameters of F-NS are close to the true values when using data from the selected stationary models. The proposed model is able to accurately estimate both fractional and non-fractional smoothnesses. Moreover, its predictions gives scores with approximately the same empirical mean and stan-

**Table 6.2:** Mean parameter estimates with standard deviation in parentheses. Data from a non-fractional and stationary model, and estimates based on 500 observations of a single realization.

Model	Parameters					
	$\nu$	$\rho_1$	$\rho_2$	$\theta$	$\sigma^2$	$\sigma_e^2$
True values	1.00	8.11	1.97	22.50	1.00	0.10
F-NS	1.06 (0.31)	7.83 (0.97)	1.90 (0.17)	23.15 (2.03)	0.99 (0.08)	0.10 (0.02)
F-S	0.98 (0.12)	8.08 (1.26)	1.95 (0.36)	23.08 (1.55)	1.05 (0.11)	0.09 (0.02)
NF-NS	1.00	7.05 (1.37)	2.27 (0.50)	30.56 (6.41)	0.99 (0.13)	0.10 (0.01)
NF-S	1.00	7.97 (1.05)	1.92 (0.26)	22.86 (1.59)	1.02 (0.13)	0.10 (0.02)

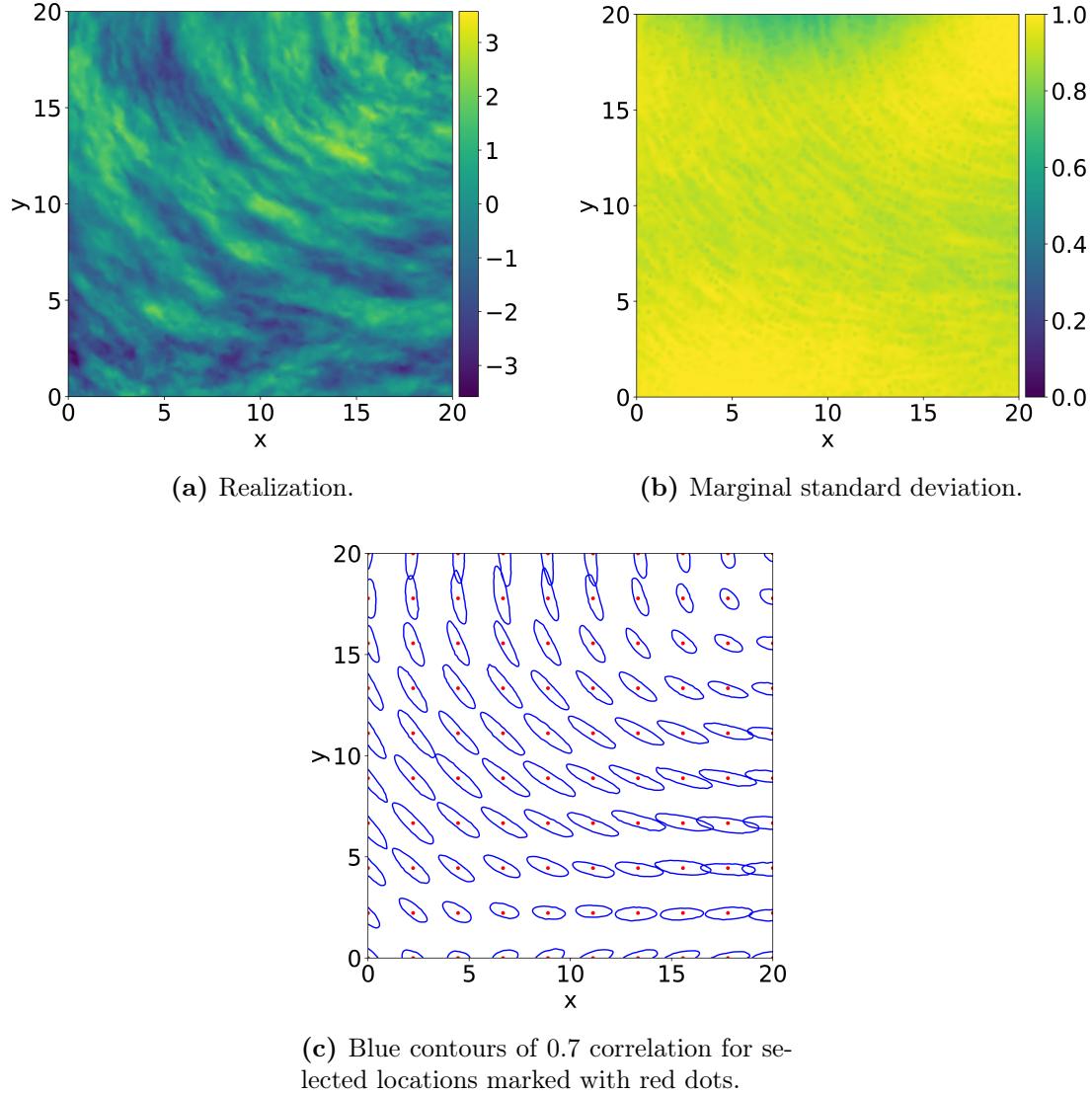
dard deviation as the stationary models. The only limitation is that the F-NS requires at least a moderate amount of observations, and that the estimated parameters in some cases have a higher empirical standard deviation. These results are therefore an indication that using the proposed model can be beneficial when the true nature of the underlying field is unknown.

## 6.5 Fractional and non-stationary data

Figure 6.3 visualizes the fractional and non-stationary model used for the subsequent estimation and prediction. Observe that the anisotropy exhibits a clear spatial variation. However, a stationary model with anisotropy oriented downward to the right could approximate the field reasonably well. Although a stationary marginal variance is used, Figure 6.3b indicates slight variations due to the non-stationarity.

The estimated parameters are displayed in Table 6.3. Be aware that the true model has no stationary contribution to the anisotropy, so large values for the estimated  $\theta$  are somewhat misleading for the non-stationary models. Overall, the F-NS achieves estimates very close to the true parameters. The NF-NS also obtains adequate results, but there is again a bias in the estimated correlation ranges. As expected, the stationary models estimate an anisotropic behavior pointing downwards to the right, indicated by a negative  $\theta$ -value. Both fractional models estimate a bit higher smoothness than the true value, but F-S is more biased than F-NS. For F-S, a possible explanation for the higher smoothness is that it has to compensate for only having stationary parameters.

The prediction scores from the fitted models are shown in Figure 6.4. Unlike in the previous section, the non-stationary models have better scores than the stationary models, with the F-NS achieving the smallest mean and standard deviation. However, the non-stationary models are also more sensitive to the number of observations. With only 125 observations, their mean scores are higher than those of the stationary models, but this is no longer the case when the number of observations increases to 250. Comparing fractional and non-fractional models, the scores are similar. Yet, when using at least 250 observations, the fractional models have a slightly lower mean and a noticeably smaller standard deviation than their non-fractional counterparts.



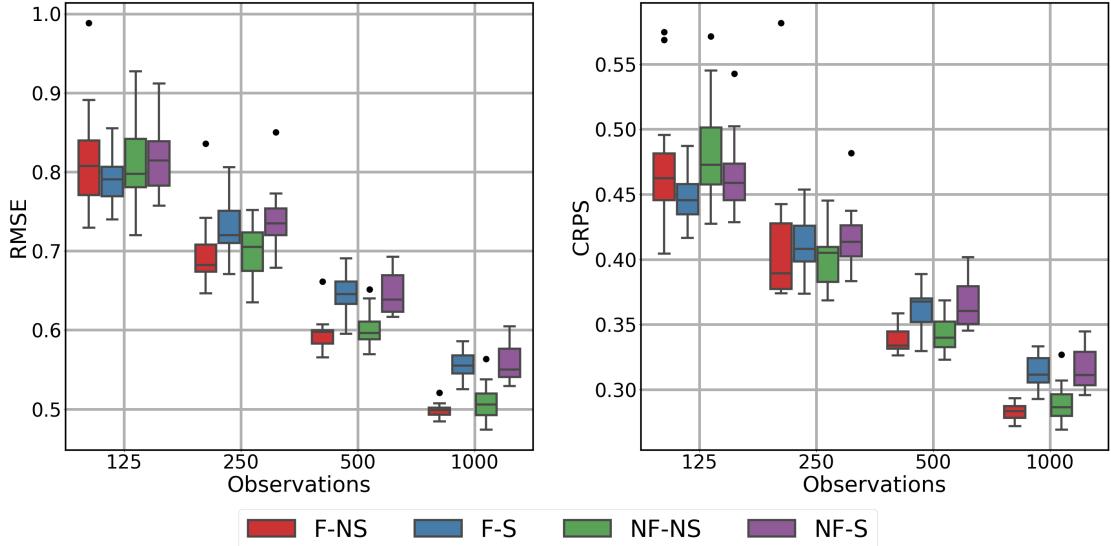
**Figure 6.3:** Visualizations of the fractional and non-stationary model.

**Table 6.3:** Mean parameter estimates with standard deviation in parentheses. Data from a fractional and non-stationary model, and estimates based on 500 observations of a single realization.

Model	Parameters					
	$\nu$	$\rho_1$	$\rho_2$	$\theta$	$\sigma^2$	$\sigma_e^2$
True values	0.50	2.83	2.83	-	1.00	0.10
F-NS	0.57 (0.15)	2.97 (0.73)	2.64 (0.67)	-19.50 (50.25)	1.03 (0.14)	0.10 (0.01)
F-S	0.61 (0.40)	3.57 (0.89)	1.92 (0.47)	-35.25 (11.59)	0.88 (0.09)	0.08 (0.05)
NF-NS	1.00	2.58 (0.28)	2.18 (0.30)	-6.26 (54.83)	0.96 (0.12)	0.13 (0.02)
NF-S	1.00	3.38 (0.73)	1.84 (0.43)	-30.66 (10.36)	0.76 (0.10)	0.16 (0.06)

## 6.6 Varying signal-to-noise ratio

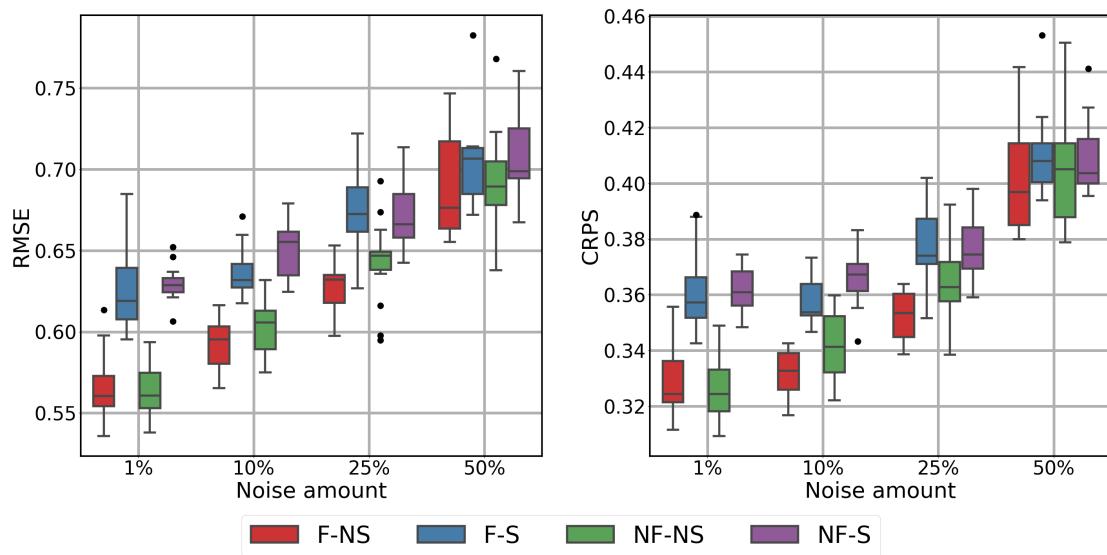
Lastly, consider the same fractional and non-stationary model as in the previous section, except that we now use different  $\sigma_e^2$  values when applying noise to the



**Figure 6.4:** Scores for predictions using data from a non-stationary model with smoothness  $\nu = 0.5$ .

data. Figure 6.5 visualizes the prediction scores as functions of the ratio  $\sigma_e^2/\sigma^2$ . As expected, the predictive accuracy decreases across all models when increasing the noise level. Comparing the stationary and non-stationary models, the non-stationary models have a smaller empirical mean and standard deviation for noise levels of 1%, 10% and 25%. Notably, the mean scores of the non-stationary models at 25% noise is approximately the same as for the stationary models at only 1% noise. At 50% noise, both types of models achieve similar scores, but the non-stationary models have a significantly larger empirical standard deviation.

Another interesting observation is the difference between NF-NS and F-NS. At 1% noise, both have approximately equal scores. However, when increasing the noise level, the fractional model consistently has lower scores than its non-fractional counterpart. These results indicate that the use of fractional models is not necessarily restricted by the signal-to-noise ratio, and can in fact be beneficial for some amounts of observation noise.



**Figure 6.5:** Scores for predictions using data from a non-stationary model with smoothness  $\nu = 0.5$  and varying observation noise  $\sigma_e^2$ . Noise amount denotes the ratio  $\sigma_e^2/\sigma^2$ . Parameter estimates based on 500 observations, and predictions on 1000 unobserved locations.



# CHAPTER 7

---

## Analysis of precipitation data

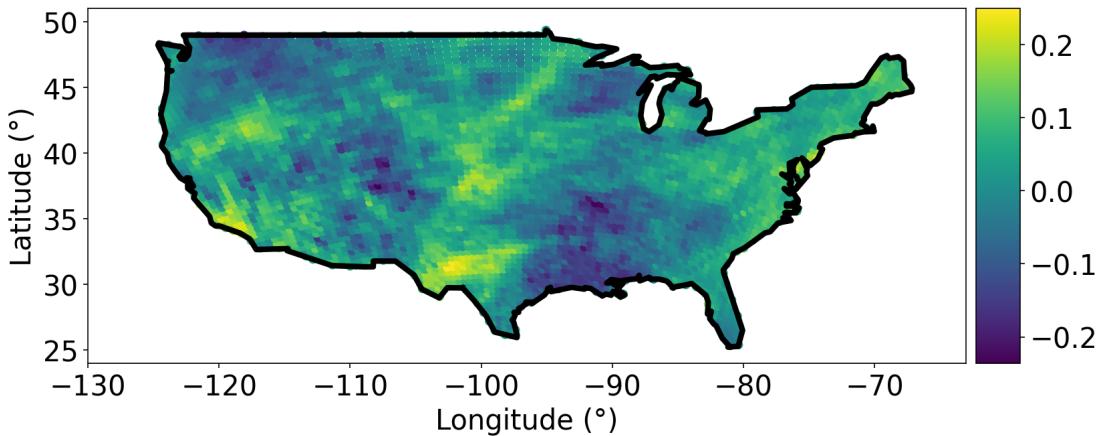
---

### 7.1 Dataset and goals

As the second part of evaluating the proposed model, we analyze a dataset of climate reanalysis data describing precipitation. Climate reanalysis combines a climate model with observations to describe climate patterns. The dataset is generated by the Experimental Climate Prediction Center Regional Spectral Model (ECPCRSM) for the North American Regional Climate Change Assessment Program (NARCCAP). It contains the average summer precipitation over the conterminous U.S. for a 26-year period from 1979 to 2004.

The data was selected because previous studies by Genton and Kleiber (2015) and Bolin and Kirchner (2020) identified it as having a fractional smoothness. In particular, Genton and Kleiber (2015) suggested that an exponential covariance model could be used for a subset of the data. Further, since the data consists of observations made densely across a large spatial region and where the data has been gathered over many years, it is reasonable to expect that a non-stationary model can yield better predictions. Fuglstad et al. (2015b) explored a similar dataset using a non-stationary model, and obtained smaller RMSE and CRPS scores when introducing non-stationarity.

We want to investigate whether using the proposed model can be advantageous compared to simpler models. The main focus is if it improves predictions, and how this depends on the amount of observed data. Additionally, we want to check if the non-stationary model is able to detect any spatial varying behavior, and if this behavior can be explained by real properties of the domain.



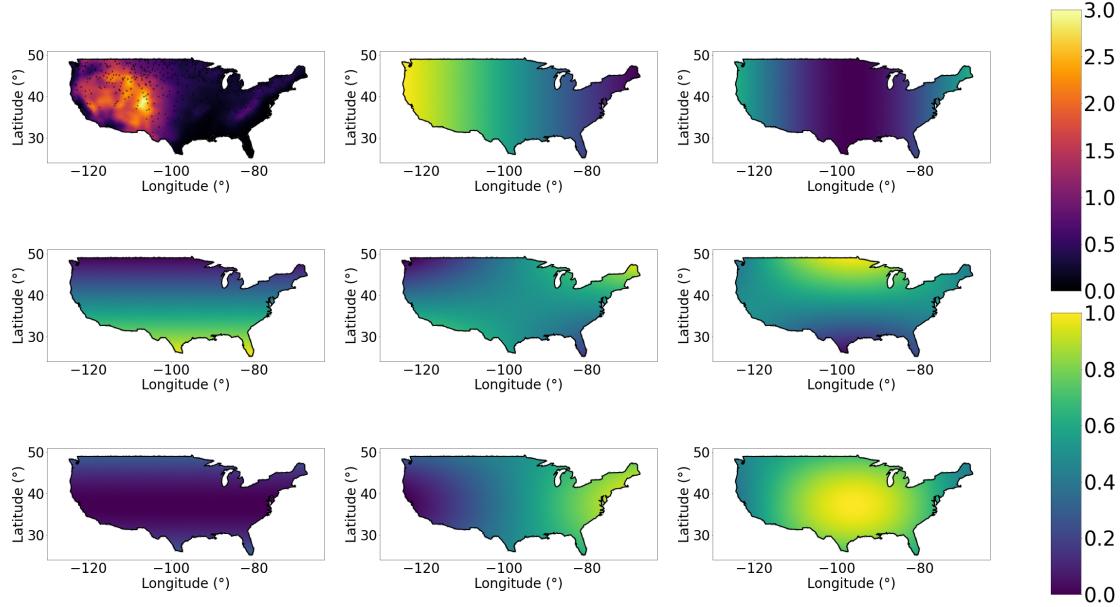
**Figure 7.1:** Processed observations for the year 1979.

## 7.2 Modelling the precipitation

To make the data suitable for spatial modelling, we begin by preprocessing it according to the procedure described in Bolin and Kirchner (2020). For obtaining data where a Gaussian distribution can be assumed, it is transformed by taking the cube root (Genton and Kleiber, 2015). To avoid having to estimate the mean covariates, the mean over the 26 years is subtracted from each location observed, giving data for which a zero mean can be assumed. Figure 7.1 visualizes the resulting data for the year 1979. Observe that there appears to be a shorter correlation range in the south-west, and that the range increases when moving east. This trend is not unique to this specific year, and shows up in most of the yearly data.

For each year, the  $n = 4112$  observations are modelled as described in Chapter 5. As the field is assumed to zero mean, the fixed effects are omitted. Similarly as in Chapter 6, we compare the NF-S, NF-NS, F-S and F-NS. The models are mostly the same as the ones described in Section 6.3, except that the non-stationarity is parameterized using 16 basis functions. The first 15 are constructed according to the procedure described in Chapter 3. After experimentation, we found that using penalty parameters of  $\tau_i = 20$  for  $i = 1, 2, 3, 4$  yielded a good results. The 16-th basis function describes the altitude in kilometers. This is added to accommodate the expected influence of the Rocky Mountains in the western region. For penalizing this additional basis, the likelihood from the normal distribution  $\mathcal{N}(0, 1)$  is used. The variance was chosen empirically after experimenting with different values. In Figure 7.2, the altitude basis and the first eight of the other bases are visualized.

To perform computations, the domain is expanded by two times the correlation range estimated in Bolin and Kirchner (2020), and triangulated into a coarse and a fine mesh with approximately 4000 and 12000 vertices, respectively. Figure 4.1 from Chapter 4 displays the coarse mesh. For parameter estimation using the hybrid approach, we found 400 iterations for the Adam optimization and 800 for the NLOpt optimization to be a reasonable balance between model fit and runtime. Choosing initial values for the models, the stationary parameters are randomly set



**Figure 7.2:** The altitude basis and first eight other basis functions used for parameterizing the non-stationary parameters.

within  $\pm 50\%$  of the parameters estimated by Bolin and Kirchner (2020), while the non-stationary parameters are set to zero.

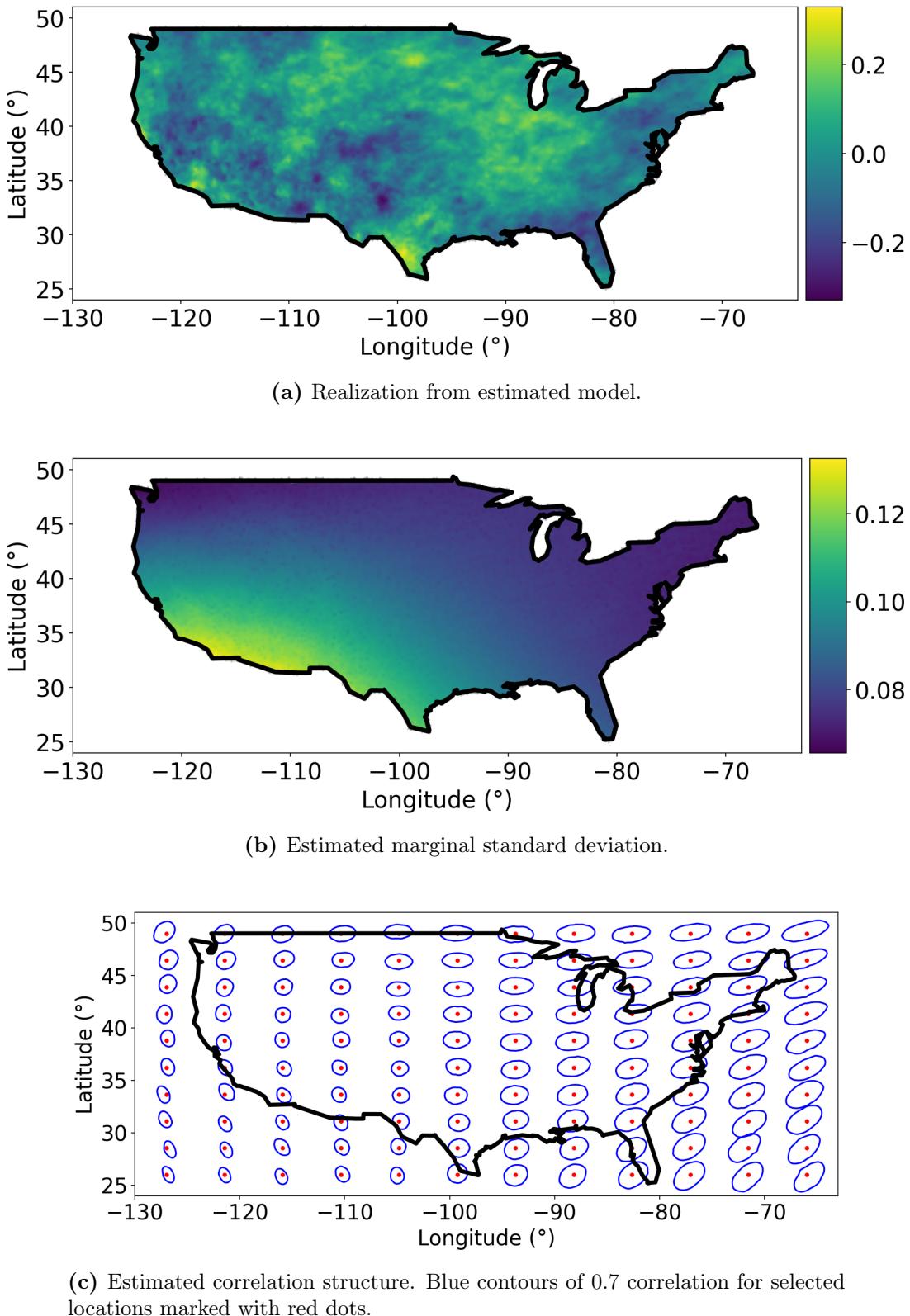
## 7.3 Investigating non-stationarity in the data

We begin by estimating the parameters of the F-NS using all the available data. Figure 7.3 visualizes a realization from the fitted model, in addition to the marginal standard deviation and correlation ranges. Figure 7.3c clearly indicates a trend of a lower correlation range in the western part of the domain. Additionally, the eastern part exhibits a slight non-stationary anisotropic behavior, where the correlation is stronger along the longitudinal axis. One possible explanation of this is the Gulf steam impacting the eastern climate. Observe from Figure 7.3b that the model estimates a higher marginal standard deviation in the south-west. This could be due to the complex terrain of the Rocky Mountains or Sierra Nevada introducing localized weather phenomena. The proximity to the Pacific Ocean might also have an impact.

## 7.4 Model comparison

### 7.4.1 10-fold cross-validation

To compare the models, we perform a 10-fold cross-validation. The dataset is split randomly into ten roughly equal-sized subsets, called *folds*. Each fold is sequentially used as the test set, while the remaining nine folds serve as the training set. The models estimate the parameters using the training set, and then predict the precipitation for the locations in the test set. The predictive performance is measured using the RMSE and CRPS scores, and are computed using the predic-



**Figure 7.3:** Estimated fractional and non-stationary model fitted on all available data. In total approximately  $10^5$  observations across 26 realizations.

tions and true values from the test fold. This process is repeated for all ten folds, and the average of the scores across the folds is reported as an estimate of the

predictive performance.

Inspecting the estimated parameters, Table 7.1 displays the mean estimates and empirical standard deviations over the ten folds. The F-NS estimates a lower smoothness compared to the F-S. Additionally, see that compared to the non-fractional models, the fractional have a larger empirical standard deviation for the estimated correlation ranges. Comparing the two non-stationary models, F-NS and NF-NS estimate slightly different correlation ranges. Another noteworthy observation is how the NF-S estimates a very different correlation range and marginal variance compared to the other models.

**Table 7.1:** Mean parameter estimates from 10-fold cross-validation, with standard deviation in parentheses.

Model	Parameters					
	$\nu$	$\rho_1$	$\rho_2$	$\theta$	$\sigma^2 [10^{-2}]$	$\sigma_e^2 [10^{-2}]$
F-NS	0.47 (0.04)	6.07 (0.18)	4.03 (0.14)	-1.41 (0.41)	0.44 (0.07)	0.0401 (0.0004)
F-S	0.75 (0.07)	4.56 (0.27)	3.61 (0.19)	-1.13 (0.68)	0.48 (0.03)	0.0357 (0.0010)
NF-NS	1.00	4.35 (0.02)	2.34 (0.01)	-1.31 (0.27)	0.44 (0.07)	0.0367 (0.0011)
NF-S	1.00	14.10 (0.05)	11.31 (0.04)	5.57 (0.24)	5.31 (0.10)	0.0407 (0.0014)

The average RMSE and CRPS scores from the predictions are displayed in Table 7.2. Observe that the differences between the models are very small, but that the F-NS has the smallest average in both measures. Interestingly, even when using very different parameter compared to the other models, the NF-S obtains an approximately equal score. One likely explanation for the small differences between the models is that every unobserved location in the held-out fold is located very close to an observed location. Since the field changes little between nearby locations, all models make accurate predictions. As a result, the estimate of the observation noise dominates the results.

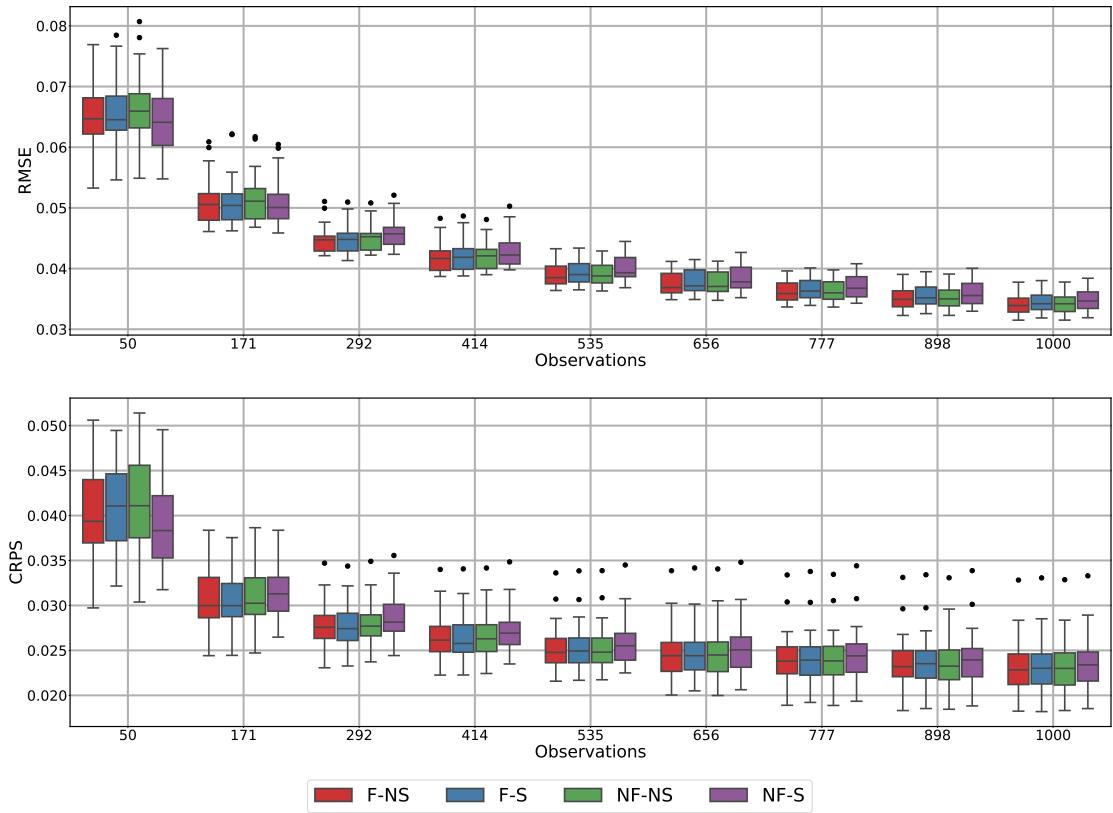
**Table 7.2:** Average RMSE and CRPS scores from the 10-fold cross-validation.

Scores	Models			
	F-NS	F-S	NF-NS	NF-S
RMSE	0.025692	0.025713	0.025762	0.025732
CRPS	0.018973	0.019026	0.019019	0.019028

### 7.4.2 Varying number of observations

To address the issue of observation noise dominating the results of the 10-fold cross-validation, an alternative approach is proposed. This method assumes a scenario where precipitation data is available at every location over multiple years, but only collected at a subset of locations for a new year. The goal is to evaluate how well the method can predict precipitation at unobserved locations for the new year, based on the number of observed locations distributed across the spatial domain.

The procedure begins by selecting a range of observation counts. Following a



**Figure 7.4:** Predictive scores as a function of the number of observations.

similar strategy as in Chapter 6, nested observation locations are randomly selected for each count in the range. For a given year, the model parameters are estimated based on all available data from the other 25 years. Using these parameter estimates and the observations in the selected year, predictions are made for all unobserved locations. Note that the number of unobserved locations varies depending on the number of observations, but this variation remains consistent across models. The predictions are evaluated by using the predictions and true values to compute RMSE and CRPS scores. This procedure is repeated by holding out each of the 26 years in turn, providing the empirical mean and standard deviation of the prediction scores.

Figure 7.4 presents the resulting scores an increasing number observations. As expected, for every model, the empirical mean of the scores increases as more data is observed. However, there is little difference between the different models. As the number of observations increases beyond 50, all deviation between the models become negligible.

## 7.5 Interpretation of results

Based on the results from Table 7.2 and Figure 7.4, there is no indication of the F-NS achieving better results than the simpler alternatives. In fact, there is neither any indication that incorporating non-stationary has any significant impact on the predictions. These results are surprising, as it was expected that at least the non-stationarity would influence the scores.

Contrary to our findings, the analysis by Bolin and Kirchner (2020) concluded that a non-stationary  $\kappa^2$  and  $\sigma^2$  improved the fractional model. However, we do not agree with their approach for doing the cross-validation. They train the models on all available data, and only hold out training and test sets when doing the predictions. The results presented here are therefore not necessarily contradictory to their findings, as their results might suffer from a poorly constructed evaluation.

Although it is possible that under these conditions, neither fractional nor non-stationary properties improve the predictions, other factors could also have influenced our results. The following list summarizes some aspects that may have had an impact.

- The resolution of the meshes was too low. This could have resulted in the model not being able to capture the short-range details necessary to estimate the correct smoothness.
- The optimization algorithm for the parameter estimation did not converge properly. This could have been caused by too few iterations or poor choices of initial values. The parameters not being estimated correctly can also have been caused by wrong penalty parameters, or parameterizing the non-stationarity with too few basis functions.
- The flexibility of the model could not represent the correct covariance-structure. Fuglstad et al. (2015b) argues for a similar dataset that the  $\sigma_e^2$  has to be different in the eastern and western regions. According to their findings, only adding non-stationarity in the range, anisotropy, and marginal variance leads to overfitting. The model incorrectly adapts to the higher observation error in the western region.
- The precipitation varies significantly from year to year. If this is the case, determining a common covariance-structure for all 26 years is suboptimal.
- It is wrong to assume the data is Gaussian. Genton and Kleiber (2015) only considered a smaller region, and expanding the assumption of Gaussianity to the entire conterminous U.S. could be incorrect. For example, areas like the Mojave Desert are known to have little to no yearly precipitation, which means assuming Gaussianity could potentially result in a heavy-tailed distribution that has a higher likelihood of producing extreme values. If this was the case, even the optimal covariance structure for the model could yield poor results.
- Even though we have subtracted the mean precipitation over the 26 year from the data, we cannot be certain that the mean structure is correctly removed from the data. For example, it is likely that the climate was subject to some change over the time-span. Phenomena like the El Niño-Southern Oscillation could also have influenced the precipitation over some periods.

Addressing these issues can be the subject of future work, but is outside the scope of this thesis.



# CHAPTER 8

---

## Discussion

---

This thesis extends the class of non-stationary SPDE models to simultaneously handle non-stationary anisotropy and fractional smoothness parameters. We do this by combining independent work with previously explored approaches. In particular, the non-stationary spatial regression model by Fuglstad et al. (2015b), the parameterization by Llamazares-Elias et al. (2024), and the rational approximation by Bolin and Kirchner (2020) have all influenced the resulting model.

Incorporating non-stationarity into the model is simple, as it follows as a natural extension of making the parameters in the SPDE spatially varying. On the other hand, introducing a fractional smoothness is more challenging, as it involves estimating the fractional differential operator in a way which gives a sparse precision matrix. As discussed in Section 5.4, the approximation of the fractional SPDE also significantly complicates the inference. Although PyTorch can be used to compute the gradient of the likelihood, it currently only supports a dense Cholesky factorization. This limitation sets restrictions on the size of the systems which we can consider, and is handled through the hybrid optimization approach proposed in Section 5.4.

When parameterizing the non-stationarity, we follow a similar approach as proposed in Fuglstad et al. (2015b). This yields a parameterization where the non-stationarity can be controlled with a penalty parameter. However, an issue is that the penalty has to be chosen manually. As we do not always have knowledge of what value is suitable, it must be selected through trial and error. For parameterizing the diffusion matrix, the half-angle approach by Llamazares-Elias et al. (2024) was used. Compared to the parameterization by Fuglstad et al. (2015a), it has the significant benefit of being identifiable. Although not provided in the thesis, we found from empirical evaluation that the new parameterization made the

model more resilient to getting stuck in local optimum. One drawback is that the direction of the underlying vector field no longer aligns with the direction of the anisotropy. Yet, this does not pose any practical issues, as we are rarely interested in the vector field itself.

The simulation study finds that the proposed model is beneficial for parameter estimation. It estimates parameters accurately based on a single realization of data from a non-fractional or stationary model, and for most signal-to-noise ratios. For data from a fractional and non-stationary model, its estimated bias is lower than for simpler alternatives. However, with very few observations, the bias of the non-stationary models are higher than for the stationary.

The simulation study shows that the proposed model generally achieves better prediction scores than simpler models. Specifically, its empirical mean and standard deviation of the RMSE and CRPS scores are often smaller than those of the other models. When evaluating data from the selected fractional and non-stationary model, the non-stationary models produce significantly better scores. Additionally, fractional models typically outperform non-fractional models across different signal-to-noise ratios. However, for most of the datasets used in the evaluation, the performance of the fractional and non-fractional models is nearly identical. The main distinction is that the scores of the fractional models tend to exhibit a smaller empirical standard deviation when applied to fractional data.

Applying the models to real-world data, the proposed model estimated a fractional smoothness consistent with the results of Bolin and Kirchner (2020) and a non-stationary behavior similar to the one found by Fuglstad et al. (2015b). Although we expected the non-stationary models to perform better than the stationary, our results indicated little difference between the models. Yet, we believe more work is needed to provide a clear answer of whether the proposed model is beneficial for this particular dataset.

Overall, the proposed fractional and non-stationary model is a step towards a fully flexible SPDE-based model, and shows potential of being useful in some situations. However, from our findings, it is hard to argue for why estimating the fractional smoothness is necessary for most applications. It complicates many aspects of the implementation and inference, and the predictions does not yield significantly better RMSE and CRPS scores than when using a non-fractional alternative. For future work, we are interested in improving the optimization. One way to doing this is by implementing a sparse Cholesky factorization in PyTorch, removing the need for the hybrid approach. There are also better ways of choosing the penalty parameters, which would make it easier to fit non-stationary models. Moreover, we want to investigate other scenarios where a fractional and non-stationary model can be applicable. Even though we were unable to show its usability for the precipitation data, there might be different situations where it is beneficial. Lastly, Bolin et al. (2024) recently proposed a modified version of the rational SPDE model, which they claim improves some aspects of the approximation. Incorporating this version can be relevant, but will also require some modifications to the model.

## APPENDIX A

---

### Scaling factor for the marginal variance

---

An analytical formula for the scaling factor can be computed by finding an expression for the marginal variance of

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

We can then use this to determine how we must scale the white noise for the solution  $u$  to have the desired marginal variance  $\sigma^2$ . To analytically compute the marginal variance, spectral theory is applied. For a continuous stationary process defined of a one-dimensional domain (Lindgren, 2013, Theorem 3.7), the spectral theorem states the following.

**Theorem A.0.1** (The spectral theorem). *If  $\{x(t), t \in \mathbb{R}\}$  is a zero mean, continuous stationary process with spectral distribution  $F(w)$ , there exists a complex-valued spectral process  $\{Z(w), w \in \mathbb{R}\}$ , with orthogonal increments, such that*

$$E[|Z(w_2) - Z(w_1)|^2] = F(w_2) - F(w_1)$$

for all  $w_1 < w_2$  and

$$x(t) = \int_{-\infty}^{\infty} e^{iwt} dZ(w).$$

The spectral theorem naturally expands to  $d$ -dimensional processes, and can therefore be utilized in  $\mathbb{R}^2$ . The Fourier transform of the previously specified model will be

$$(\kappa^2 + \mathbf{w}^T \mathbf{H} \mathbf{w})^\beta \hat{u}(\mathbf{w}) = \widehat{\mathcal{W}}(\mathbf{w}), \quad \mathbf{w} \in \mathbb{R}^2,$$

where  $\hat{u}$  and  $\widehat{\mathcal{W}}$  are the Fourier transformed processes of  $u$  and  $\mathcal{W}$ , respectively. The spectral density function can be found by evaluating  $E[\hat{u} \hat{u}]$  where  $\hat{u}$  denotes

the complex conjugate. By the fact that the spectral density of a standard Gaussian white noise process on  $\mathbb{R}^2$  is identically equal to  $1/(2\pi)^2$  we obtain

$$f_S(\mathbf{w}) = \left(\frac{1}{2\pi}\right)^2 \frac{1}{(\kappa^2 + \mathbf{w}^T \mathbf{H} \mathbf{w})^{2\beta}}. \quad (\text{A.1})$$

The martial variance can be determined by integrating the spectral density  $f_S$  over  $\mathbb{R}^2$ , i.e.

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

As  $\mathbf{H}$  is symmetric positive definite it has a symmetric positive definite square root  $\mathbf{H}^{1/2}$ . Therefore, we can use the change of variables  $\mathbf{w} = \kappa \mathbf{H}^{1/2} \mathbf{z}$ . Hence,

$$\begin{aligned} \sigma_m^2 &= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \frac{1}{(\kappa^2 + \kappa^2 \mathbf{z}^T \mathbf{z})^{2\beta}} \det(\kappa \mathbf{H}^{1/2}) d\mathbf{z} \\ &= \frac{1}{4\pi^2 (\kappa^2)^{2\beta-1} \sqrt{\det(\mathbf{H})}} \int_{\mathbb{R}^2} \frac{1}{(1 + \mathbf{z}^T \mathbf{z})^{2\beta}} d\mathbf{z} \\ &= \frac{1}{4\pi (\kappa^2)^{2\beta-1} \sqrt{\det(\mathbf{H})}} \cdot \frac{\Gamma(2\beta - 1)}{\Gamma(2\beta)}. \end{aligned} \quad (\text{A.2})$$

Here, the final integral was computed by comparing the solution to the analytical expressions given by Whittle (1963). Using the result from Equation (A.2), scaling the Gaussian white noise with the factor

$$\tau = \sigma \sqrt{4\pi \frac{\Gamma(2\beta)}{\Gamma(2\beta - 1)} \kappa^{2\beta-1} (\det \mathbf{H})^{1/4}}$$

gives the GRF a marginal variance  $\sigma^2$ . Note that the determinant disappears when parameterizing  $\mathbf{H}$  as in Equation (3.6).

---

## Bibliography

---

George A Baker Jr and Peter Graves-Morris. Padé approximants, volume 59 of. *Encyclopedia of Mathematics and its Applications*, 22, 1996.

Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.

David Bolin and Kristin Kirchner. The rational spde approach for gaussian random fields with general smoothness. *Journal of Computational and Graphical Statistics*, 29(2):274–285, 2020.

David Bolin and Finn Lindgren. A comparison between markov approximations and other methods for large spatial data sets. *Computational Statistics & Data Analysis*, 61:7–21, 2013.

David Bolin, Kristin Kirchner, and Mihály Kovács. Numerical solution of fractional elliptic stochastic pdes with spatial white noise. *IMA Journal of Numerical Analysis*, 40(2):1051–1073, 2020.

David Bolin, Alexandre B Simas, and Zhen Xiong. Covariance–based rational approximations of fractional spdes for computationally efficient bayesian inference. *Journal of Computational and Graphical Statistics*, 33(1):64–74, 2024.

Geir-Arne Fuglstad, Finn Lindgren, Daniel Simpson, and Håvard Rue. Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, pages 115–133, 2015a.

Geir-Arne Fuglstad, Daniel Simpson, Finn Lindgren, and Håvard Rue. Does non-stationary spatial data always require non-stationary random fields?, 2015b.  
URL <https://arxiv.org/abs/1409.0743>.

Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

- Marc G Genton and William Kleiber. Cross-covariance functions for multivariate geostatistics. 2015.
- Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.
- Xiangping Hu, Ingelin Steinsland, Daniel Simpson, Sara Martino, and Håvard Rue. Spatial modelling of temperature and humidity using systems of stochastic partial differential equations, 2015. URL <https://arxiv.org/abs/1307.1402>.
- Steven G. Johnson. The NLOpt nonlinear-optimization package. <https://github.com/stevengj/nlopt>, 2007.
- Finn Lindgren. *fmesher: Triangle Meshes and Related Geometry Tools*, 2024. URL <https://inlabru-org.github.io/fmesher/>. R package version 0.1.7.9008, <https://github.com/inlabru-org/fmesher>.
- Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498, 2011.
- Finn Lindgren, David Bolin, and Håvard Rue. The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599, August 2022. ISSN 2211-6753. doi: 10.1016/j.spasta.2022.100599. URL <http://dx.doi.org/10.1016/j.spasta.2022.100599>.
- Georg Lindgren. *Stationary stochastic processes*. Texts in Statistical Science Series. CRC Press, 2013. Theory and applications.
- Liam Llamazares-Elias, Jonas Latz, and Finn Lindgren. A parameterization of anisotropic gaussian fields with penalized complexity priors, 2024.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019. URL <https://arxiv.org/abs/1912.01703>.
- Michael JD Powell. Direct search algorithms for optimization calculations. *Acta numerica*, 7:287–336, 1998.
- Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.
- Dan Simpson. Random C++ Part 2: Sparse partial inverses in Eigen, 2024. URL <https://dansblog.netlify.app/posts/2024-09-05-partial-inverse/partial-inverse>.
- ML Stein. Interpolation of spatial data. Springer series in statistics, 1999.

Kazuhiro Takahashi. Formation of sparse bus impedance matrix and its application to short circuit study. In *Proc. PICA Conference, June, 1973*, 1973.

Peter Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

Peter Whittle. Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994, 1963.