



MOX-Report No. 13/2025

**Estimating Non-Stationarity in Spatial Processes: an approach based  
on Random Domain Decomposition**

Scimone, R.; Menafoglio, A.; Secchi, P.

MOX, Dipartimento di Matematica  
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

[mox-dmat@polimi.it](mailto:mox-dmat@polimi.it)

<https://mox.polimi.it>

# Estimating Non-Stationarity in Spatial Processes: an approach based on Random Domain Decomposition

Riccardo Scimone, Alessandra Menafoglio, Piercesare Secchi

MOX - Department of Mathematics, Politecnico di Milano,  
Piazza Leonardo da Vinci 32, 20133 Milano, Italy

## Abstract

The present work addresses the problem of flexible and efficient parameter estimation for non-stationary Gaussian random fields. This problem is crucial to enable modeling and stochastic simulation of complex natural phenomena in the Earth Sciences. Building on the non-stationary Matérn model of [Paciorek and Schervish \(2006\)](#), we propose a novel computational method that leverages random and repeated domain partitions to construct locally stationary estimates. Unlike existing approaches that rely on fixed grids of knots, our method employs a bagging-type strategy to mitigate the influence of domain decompositions in a divide-and-conquer fashion. This results in more robust and adaptive estimations, overcoming key limitations of traditional methods. Through extensive simulations and a real case study, we demonstrate that while fixed grids noticeably impact the final estimated models, our approach produces grid-free estimations, thanks to the additional source of randomness introduced by the aleatory partitions of the domain.

*Keywords:* Non Stationarity, Covariance Modeling, Gaussian Processes, Random Domain Decomposition

# 1 Introduction

Gaussian Processes (GP) have become ubiquitous in many fields of Probability, Statistics and Machine Learning, being used as a powerful and flexible tool for estimation and prediction in regression and classification problems (Rasmussen and Williams, 2006). We focus here on the use of GPs, and in particular of non-stationary GPs, in the field of Spatial Statistics (Cressie, 1993). Many connections between GPs and traditional Spatial Statistics have been established, the most representative example being *kriging* (Cressie, 1993), which can be framed as a particular or limit case of Gaussian Process Regression (see Rasmussen and Williams (2006); Stein (2012); Gelfand and Schliep (2016)).

For spatial modeling, Gaussian Processes are appealing because they are completely specified by their mean and covariance functions. Assume that  $D \subset \mathbb{R}^d$  is a spatial domain of interest, then a real GP is a real stochastic process  $\{Y_s\}_{s \in D}$  whose finite-dimensional distributions are multivariate Gaussian (Rasmussen and Williams, 2006). In this case, we write

$$Y \sim GP(\mu, C + \tau^2 I)$$

where  $\mu : D \rightarrow \mathbb{R}$  is the (spatial) mean function, naturally defined as  $\mu(s) = \mathbb{E}[Y_s]$ , for  $s$  in  $D$ ,  $C : D^2 \rightarrow \mathbb{R}$  is a *valid* (i.e., positive definite, (Cressie, 1993; Stein, 2012)) (spatial) covariance function, defined as  $C(s_1, s_2) = Cov(Y_{s_1}, Y_{s_2})$  for  $s_1, s_2$  in  $D$ ,  $I$  is the identity matrix and  $\tau^2 : D \rightarrow \mathbb{R}^+$  is the *nugget effect*, usually included to account for noisy measurements (Rasmussen and Williams, 2006; Stein, 2012). For a GP to be *stationary*,  $\mu$  and  $\tau$  must be constant over space, and  $C$  must only depend on the increment vector  $s_1 - s_2$ . If in addition  $C$  depends only on the Euclidean distance between  $s_1$  and  $s_2$ , yielding no preferred correlation directions, the process is said to be *stationary isotropic*; when this is not true the process is called *stationary anisotropic*. Among different kinds of anisotropy, *geometric anisotropy* is particularly meaningful and prone to straightforward geometric interpretations. In this case,  $C$  is a function of the Mahalanobis distance between  $s_1$  and  $s_2$  induced by the symmetric positive definite *anisotropy matrix*  $\Sigma$ , whose eigenvalues and eigenvectors define the main correlation directions (Cressie, 1993; Stein, 2012).

If  $C$  cannot be expressed as a function of  $s_1 - s_2$ , for all  $s_1, s_2 \in D$ , we say that  $C$  is a non-stationary covariance function and  $Y$  is a non-stationary GP. In Fig. 1, we show realizations of different GP processes, where  $D = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$ . In Fig. 1a, we present a realization from a stationary isotropic process, while in Fig. 1b a realization of a stationary anisotropic process characterized by an anisotropy angle of 135 degrees with the positive horizontal axis. Fig. 1c displays a realization of a non-stationary process, where both  $\mu$  and  $C$  are non-stationary, with a correlation pattern clearly varying in space and a mean function increasing from the low-left angle to the up-right angle of the domain. In this last case, non-stationarity in  $C$  is induced by letting the anisotropy matrix vary over space, in a sense which will become clear in Section 2. Even this very simple illustration makes intuitively clear how non-stationary GPs can be very flexible modeling tools for the statistical analysis of geo-referenced natural phenomena characterized by a space-varying structure in their mean and in their covariance, like, for instance, those typically studied in climatology. However, while the stationarity assumption allows one to leverage a solid theoretical background for spatial estimation, simulation, interpolation, prediction and posterior inference (Cressie, 1993; Stein, 2012), the non-stationary setting presents relevant challenges, particularly from the computational viewpoint. In this con-

## Stationary and non-stationary processes

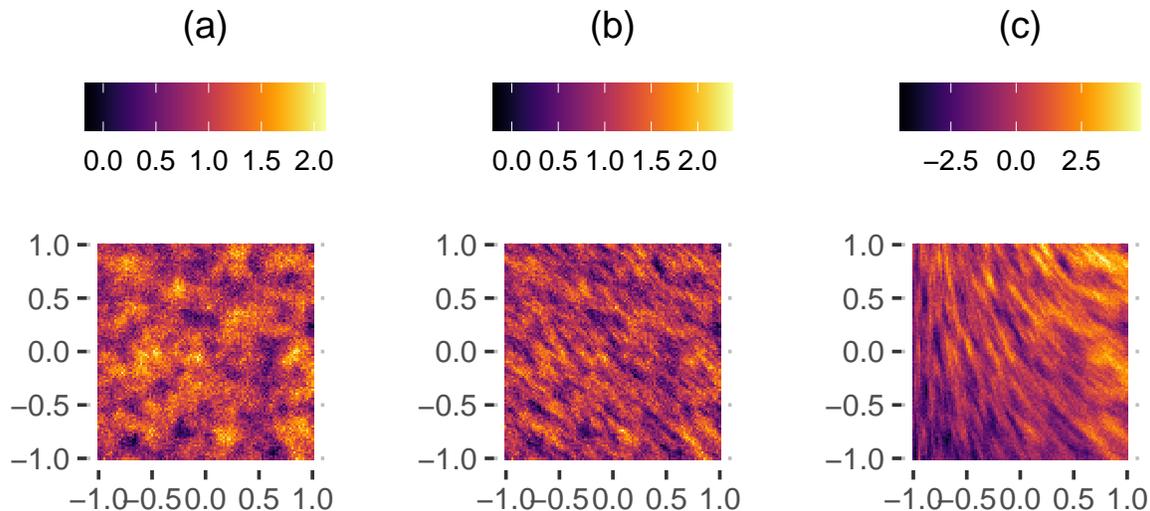


Figure 1: Realizations of (a) a stationary isotropic process, (b) a stationary anisotropic process and (c) a non-stationary process (right). With the notation of Section 2, the process (a) corresponds to setting  $\mu = 1$ ,  $\tau = 0.03$ ,  $\sigma = 0.3$ ,  $\lambda_1 = \lambda_2 = \rho^2 = e^{-7}$ ,  $\nu = 2$ ; the process (b) to  $\lambda_1 = e^{-7}$ ,  $\lambda_2 = e^{-5}$ ,  $\theta = \pi/4$ ; the process (c) to  $\tau = 0.1$  and  $\mu(s) = x_1 + x_2$ ,  $\sigma(s) = e^{\frac{1}{4}x_2}$ ,  $\lambda_1(s) = e^{-8+x_1}$ ,  $\lambda_2(s) = e^{-5}$ ,  $\theta(s) = \frac{\pi}{6}(x_1 + 1)$ , denoting by  $s = (x_1, x_2)^T$ .

text, much work has been recently devoted to the development of sound methodological frameworks for estimation and predictions in a non-stationary setting, keeping, at the same time, the required computational efforts to a reasonable level (see (Fouedjio, 2017) for a recent review).

Within this literature, we want here to contribute to the rich research stream which focuses on *convolution* methods. Starting with the seminal work of (Higdon et al. (1999)), this line of research aims to model non-stationarity by convolving a single stationary process with a spatially-varying kernel, or alternatively, by convolving different stationary processes with a spatially constant kernel (Fuentes, 2001). A major achievement was reached in (Paciorek and Schervish, 2006), where a closed form, analytic, non-stationary extension of the Matérn covariance model was introduced. Stemming from its stationary counterpart, whose appealing properties for spatial modeling are discussed in (Anders and Stein, 2011), this non-stationary Matérn model allows one to include directly a spatially-varying smoothness, variance and anisotropy matrix, thus generating a very flexible, interpretable, and yet analytically tractable, family of processes. The huge challenge which comes with all these advantages, and which motivates our contribution, is the necessity of estimating parameters which are spatial functions, making the dimension of the parameter space, in principle, infinite.

The inferential perspective followed in (Paciorek and Schervish, 2006) is Bayesian, since the authors suggest to elicit stationary GP priors for the spatially varying parameters. This approach is fully general, but is affected by a heavy computational burden, which stems from difficult MCMC computations. This forces several kind of approximations of the GP priors and likelihood to work out the posterior Bayesian inference (Paciorek and

---

(Schervish, 2006; Risser and Calder, 2015; Risser and Turek, 2020). More parsimonious models have also been defined, with (Risser and Calder, 2015) explicitly accounting for exogenous covariates. In (Risser and Turek, 2020) all the models presented in (Paciorek and Schervish, 2006) and (Risser and Calder, 2015) are carefully reviewed and R software for their implementation is provided in the `BayesNSGP` package ((Turek and Risser, 2022)).

Less taxing estimation methods, not relying on Bayesian modeling or MCMC computations, are framed in (Anderes and Stein, 2011), (Risser and Calder, 2017) and (Fouedjio et al., 2016). They are all grounded on the concept of *local stationarity*, i.e, the idea that the behavior of the process in a small neighbourhood of any spatial location should be approximately stationary. This idea, which has been recently extended to local non-stationarity in (Li and Sun, 2019), is the starting point for several *local likelihood* estimation methods ((Anderes and Stein, 2011), (Risser and Calder, 2017)) or methods based on *anchor points* (Fouedjio et al., 2016), i.e., fixed spatial locations where stationarity is assumed in small neighbourhoods.

Fixing some knot locations, typically through a regular spatial grid, is actually a key element of many of the cited estimation methods: knots are needed to approximate stationary GP priors on the spatially-varying parameters ((Paciorek and Schervish, 2006; Risser and Turek, 2020)) or to exploit the local stationarity assumption (Fouedjio et al., 2016; Risser and Calder, 2017). However, the grid choice can deeply affect estimation, as for instance discussed in (Banerjee et al., 2008), which is focused on Gaussian Predictive Processes. We show empirical evidence of this fact in Section 3.

We here propose a novel estimation approach for the spatially varying parameters of the non-stationary Matérn model developed in (Paciorek and Schervish, 2006), with the aim of pairing the cited methods with an alternative that avoids taxing computations as well as potential inconveniences coming from the necessity of fixing a grid of knots. The approach is based on *Random Domain Decomposition* (RDD, Secchi et al. (2013); Menafoglio et al. (2018, 2021)). Initially developed as a tool for predicting the values of the process in unobserved locations (Menafoglio et al., 2018, 2021), we here extend it to provide spatially-varying estimations for all the parameters of the non stationary Matérn model developed in (Paciorek and Schervish, 2006), and empirically investigate its capability to identify and estimate non-stationarity. RDD is an ensemble method, based on *bagging* (Breiman, 1996; Secchi et al., 2013), where each weak estimator is obtained on a *random partition* of the spatial domain. The randomness is then integrated out in the aggregation phase, by taking expectations or medians, achieving smoothness by averaging on many partitions rather than by imposing smooth priors on parameters or by using smoothing kernels, and avoiding the choice of a fixed grid of knots. Since in each sub-domain of each partition stationarity is assumed, we use standard variogram estimation techniques for stationary anisotropic processes (Cressie, 1993; Sherman, 2011; Stein, 2012), with the same approach followed by Fouedjio et al. (2016) with their *anchor points*. This results in an algorithm that is much faster than alternatives requiring demanding likelihood optimization steps or MCMC computations (Paciorek and Schervish, 2006; Risser and Calder, 2015; Risser and Turek, 2020). We also show, on simulated and real case studies, that RDD is able to capture spatial-varying features of the process under study, performing comparably better than other models.

The rest of the paper is organized as follows. In Section 2 we describe the non-stationary Matérn model introduced in (Paciorek and Schervish, 2006) and we set some notation; in

Section 3 we briefly revise available estimation methods and highlight some of their potential setbacks, motivating the introduction of our proposal, which is described in Section 4. Section 5 is devoted to an assessment of the performance of the algorithm on simulated data, making comparisons with the other methods which, to the best of our knowledge, have not been extensively tested on simulated scenarios; we then proceed, in Section 6, with the analysis of the Colorado Rain data, a real dataset which has already been used in (Paciorek and Schervish, 2006). Concluding remarks and lines of future research are discussed in Section 7.

## 2 Non-Stationary Matérn Model

The Matérn Covariance model, in its basic stationary isotropic version, is

$$C_{\nu,\rho}^S(h) = \sigma^2 R_{\nu,\rho}^S(h), \quad \rho > 0, \quad \nu > 0, \quad \sigma > 0,$$

where  $R_{\nu,\rho}^S(h)$  is the Matérn stationary correlation function, giving the correlation between the value of the process in two locations whose Euclidean distance is  $h$  and defined as

$$R_{\nu,\rho}^S(h) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left[ 2\sqrt{\nu} \frac{h}{\rho} \right]^\nu \mathcal{K}_\nu \left( 2\sqrt{\nu} \frac{h}{\rho} \right). \quad (1)$$

In (1),  $\sigma$ ,  $\rho$  and  $\nu$  are the covariance parameters:  $\sigma^2$  is the process variance (also known as *partial sill* (Cressie, 1993; Stein, 2012),  $\rho$  controls the correlation length of the process (the so-called *range*), and  $\nu$  influences the process smoothness. The function  $\mathcal{K}_\nu$  is the modified Bessel function of the second kind. A complete discussion on the model can be found in (Stein, 2012). These parameters need to be estimated from the data, sometimes with the exception of  $\nu$  which may be kept fixed for interpretation purposes and identifiability issues, being controversial how smoothness can be learned from discretely sampled data. The basic model (1) can be made anisotropic, *preserving stationarity*, by substituting the scaled Euclidean distance  $h/\rho$  with the Mahalanobis distance

$$h_{ij} = \sqrt{[s_i - s_j]^T \Sigma^{-1} [s_i - s_j]}, \quad (2)$$

$\Sigma$  being a symmetric positive definite matrix named *anisotropy ellipse*. To be rigorous, for all  $s_i, s_j \in D$ , this yields

$$C_{\nu,\Sigma}^S(s_i, s_j) = \sigma^2 R_{\nu,\Sigma}^S(s_i - s_j),$$

where

$$R_{\nu,\Sigma}^S(s_i - s_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}} [2\sqrt{\nu} h_{ij}]^\nu \mathcal{K}_\nu(2\sqrt{\nu} h_{ij}). \quad (3)$$

Note that stationarity is guaranteed by the fact that the covariance function depends on locations only via  $s_i - s_j$ . The symmetric, positive definite matrix  $\Sigma$  needs to be estimated from the data, so that the estimation of  $\rho$  is replaced by the need of estimating the  $d(d+1)/2$  independent entries of  $\Sigma$ . Of course, the special case  $\Sigma = \rho^2 I$  corresponds to the isotropic case.

The 2D case is particularly easy to interpret. Indeed,  $\Sigma$  can be parametrized in terms (i) of its eigenvalues  $\lambda_1, \lambda_2 > 0$ , named *maximum* and *minimum correlation length* (related with the axes-lengths of the elliptic neighbourhoods induced by  $\Sigma$ ), and (ii) of an

anisotropy angle  $\theta \in [0, \pi/2)$ , characterizing the eigenvector corresponding to  $\lambda_1$ , named *principal correlation direction*. Again,  $\lambda_1 = \lambda_2 = \rho^2$  brings us back to case (1),  $\theta$  being irrelevant in this particular case. The eccentricity of  $\Sigma$ , also known as *anisotropy ratio*  $\phi = \sqrt{\frac{\lambda_1}{\lambda_2}}$ , measures how far the process is from being isotropic. In the stationary setting, all parameters are usually estimated from data via variogram estimation or via maximum likelihood (see (Cressie, 1993; Stein, 2012) for a detailed review). For two sample realization from this model, see Figure 1.

In (Paciorek and Schervish, 2006), the Matérn model is extended, allowing for non-stationarity, by letting standard deviation and anisotropy vary in space. The authors show that smooth functions  $\sigma(s)$  and  $\Sigma(s)$  can be plugged-in in the stationary anisotropic Matérn model, still obtaining a valid covariance model. More precisely, denote by  $\sigma : D \rightarrow \mathbb{R}^+$  and  $\Sigma : D \rightarrow \mathcal{S}_d$  smooth functions (with respect to the Euclidean metric on  $D$  and the 2-norm on the matrix space  $\mathcal{S}_d$ ) of the spatial location  $s \in D \subseteq \mathbb{R}^d$ ,  $\mathcal{S}_d$  being the cone of positive definite symmetric matrices of dimension  $d$ , and let  $q_{ij}$  be a “non-stationary Mahalanobis distance”

$$q_{ij} = q(s_i, s_j) = \sqrt{[s_i - s_j]^T \left( \frac{\Sigma(s_i) + \Sigma(s_j)}{2} \right)^{-1} [s_i - s_j]}.$$

Note that  $q_{ij}$  does not depend only on  $s_i - s_j$ , rather being a more complex function of both locations, although the dependence is dropped to ease the notation. Moreover,  $q_{ij}$  does not define a metric in  $D$ , failing to satisfy the triangular inequality. On this bases, for all  $s_i, s_j \in D$ , (Paciorek and Schervish, 2006) build a non-stationary kernel  $C_{\nu, \Sigma}^{NS}(s_i, s_j)$  defined as

$$C_{\nu, \Sigma}^{NS}(s_i, s_j) = \sigma(s_i)\sigma(s_j)R_{\nu, \Sigma}^{NS}(s_i, s_j), \quad (4)$$

where the associated non-stationary correlation  $R_{\nu, \Sigma}^{NS}$  reads

$$R_{\nu, \Sigma}^{NS}(s_i, s_j) = |\Sigma(s_i)|^{\frac{1}{4}}|\Sigma(s_j)|^{\frac{1}{4}} \left| \frac{\Sigma(s_i) + \Sigma(s_j)}{2} \right|^{-\frac{1}{2}} \frac{1}{\Gamma(\nu)2^{\nu-1}} [2\sqrt{\nu}q_{ij}]^\nu \mathcal{K}_\nu(2\sqrt{\nu}q_{ij}).$$

Model (4) trivially includes the stationary models (1) and (3). The function  $\Sigma$  has the role of *spatially varying anisotropy matrix*, and, in the convolution literature, it corresponds to a *spatially varying convolution kernel* (Higdon et al., 1999; Paciorek and Schervish, 2006). If one couples model (4) with non-constant mean and nugget function, full flexibility is achieved and one obtains a model with spatially-varying mean, nugget, process standard deviation, and geometric anisotropy (controlled by the function  $\Sigma$ ). Note that, also in this case, it is particularly meaningful to parametrize  $\Sigma$  in terms of eigenvalues and rotation angles. In particular, when  $d = 2$ , the symmetric positive definite matrix  $\Sigma$  can be defined via three smooth functions  $\lambda_1 : D \rightarrow \mathbb{R}^+$ ,  $\lambda_2 : D \rightarrow \mathbb{R}^+$ ,  $\theta : D \rightarrow [0, \pi/2)$  assigning, to each point in  $D$ , the minimum and maximum correlation length, and the anisotropy angle, respectively. For the sake of illustration, Fig. 2 reports a representation of the non-stationary realization presented in Fig. 1c, overlaid, for some spatial locations, by the ellipses defined by  $\Sigma$ . One can clearly see that the variation of the anisotropy parameters strongly affects the shape of the realization.

## Non-stationary process realization, detail

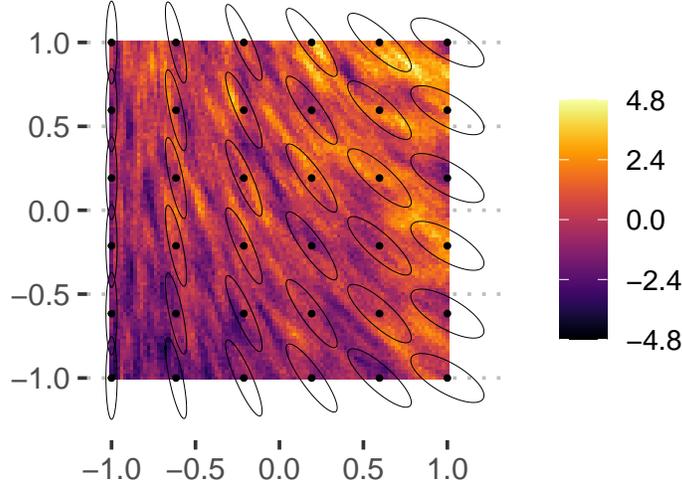


Figure 2: Realization of a non-stationary process, generated according to model (4), where we set  $\nu = 2$ ,  $\mu(s) = x_1 + x_2$ ,  $\sigma(s) = e^{\frac{1}{4}x_2}$ ,  $\tau(s) = 0.03$ ,  $\lambda_1(s) = e^{-8+x_1}$ ,  $\lambda_2(s) = e^{-5}$ ,  $\theta(s) = \frac{\pi}{6}(x_1 + 1)$ .

## 3 Available Estimation Methods and Their Drawbacks

### 3.1 State of the art

The extreme flexibility of model (4), while well-suited for non-stationary modelling and simulations, poses challenges when aiming to estimate the functional parameters  $\mu$ ,  $\tau$ ,  $\sigma$  and  $\Sigma$ . To the best of our knowledge, three different estimation frameworks explicitly based on model (4) have been proposed so far. The first one is the Bayesian framework introduced in (Paciorek and Schervish, 2006), and later extended and implemented in the works (Risser and Calder, 2015) and (Risser and Turek, 2020). The second one consists of the local likelihood estimation proposed and implemented in (Risser and Calder, 2017), while the third is the proposal made in (Fouedjio et al., 2016), based on spatially smoothed variogram estimations, carried out in fixed locations.

In the following, we assume  $D \subset \mathbb{R}^2$ . Let the random vector  $Y = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$  represent the GP realized at the locations  $s_1, \dots, s_n \in D$ . For ease of notation, for  $s \in D$ , we define  $\psi(s) = (\mu(s), \tau(s), \sigma(s), \lambda_1(s), \lambda_2(s), \theta(s))^T$  a vector containing the mean function, the nugget, and all the spatially varying parameters on which the covariance function  $\Sigma$  appearing in (4) is defined upon.

With a slight abuse of notation, we can elicit the following Bayesian model which, conditionally on  $(\psi(s_1), \dots, \psi(s_n))$ , assumes

$$Y \sim \mathcal{N}(\mu, \tau^2 + C_\psi^{NS}) \quad (5)$$

where  $\mu$  and  $\tau$  now represent the random vectors collecting the realizations of the functions  $\mu(s)$  and  $\tau(s)$  at the locations  $s_1, \dots, s_n \in D$ , while the matrix  $C_\psi^{NS}$  collects the relevant covariances, modeled as in (4). For  $s \in D$ , the six components  $\psi_j(s)$  of the vector  $\psi(s)$  define

six random fields on  $D$ . These random fields are assumed to be independent and such that  $g_j(\psi_j) \sim GP(\mu_{\psi_j}, C_{\psi_j})$ , where  $g_j$  is a link function – allowing the distributions of  $\psi_j(s)$ , for  $s \in D$ , to be compatible with the meaning and constraint inherent in the definition of the parameter –  $\mu_{\psi_j}$  is a constant mean and  $C_{\psi_j}$  is a *stationary* covariance function, acting as hyperparameters of the model. For example, [Risser and Turek \(2020\)](#) propose to use as link function the logarithm of  $\tau$ ,  $\sigma$ ,  $\lambda_1$  and  $\lambda_2$  and a scaled inverse logit transformation of  $\theta$ .

Model [\(5\)](#) is the most general model proposed in [\(Paciorek and Schervish, 2006\)](#) and implemented in [\(Risser and Turek, 2020\)](#). Moreover, it is the only Bayesian model not requiring additional knowledge on the dependence of the spatial parameters on exogenous covariates (see [\(Paciorek and Schervish, 2006; Risser and Calder, 2015; Risser and Turek, 2020\)](#)). Its generality comes with costly MCMC computations to carry out posterior inference, and thus with the necessity of approximating the stationary GP priors. This is done via a standard radial basis function approximation requiring a fixed grid of knots  $b_k \in D, k = 1, \dots, K$ , usually in the form of a regular grid. This approximation provides the linear representation

$$\psi_j(s) = \mu_{\psi_j} + \sum_{k=1}^K w_{kj}(s)u_{jk}, \quad (6)$$

where the  $u_{jk}$  are independent standard Gaussian random variables and the  $w_{kj}$  are the appropriate smoothing radial basis functions (see [\(Paciorek and Schervish, 2006\)](#), section 3.2.2, for the details). Clearly, the above representation greatly reduces the computational burden, since at each MCMC step only the  $u_{jk}$ , for all  $j, k$  are sampled, in contrast with the full GP priors required by model [\(5\)](#). Model [\(5\)](#), together with the radial basis function approximation, is implemented in the `BayesNSGP` R package [\(Risser and Turek, 2020\)](#), with a solution which is suitable for moderately large datasets, but which cannot be applied to intensive simulation studies where all parameters are allowed to vary spatially. The need for knots – and, consequently, variants of representation [\(6\)](#) – is also shared by the other previously mentioned methods [\(Fouedjio et al., 2016; Risser and Calder, 2017\)](#), which indeed can be applied in simulation studies, since they avoid MCMC computation altogether.

In [\(Risser and Calder, 2017\)](#), which is partially based on the work [\(Anderes and Stein, 2011\)](#), each knot  $b_k$  of the grid is the center of a neighborhood, i.e. a ball  $N_k$  defined by a *fitting radius*  $r$ , where a stationary covariance model is assumed to hold. In each  $N_k$ , this covariance model has a different vector of parameters  $\psi^{b_k}$ , defined analogously to  $\psi$ . So one needs to estimate the components  $\psi_j^{b_k}$ , in each of the  $k = 1, \dots, K$  neighbourhoods. These estimates are carried out independently in each neighbourhood, as is typical of local likelihood estimation methods [\(Tibshirani and Hastie, 1987\)](#). Once estimates  $\hat{\psi}_j^{b_k}$  are obtained, the final, spatially varying estimates are defined via a representation analogous to [\(6\)](#):  $\hat{\psi}_j(s) = \sum_{k=1}^K w_k(s)\hat{\psi}_j^{b_k}$ . In [\(Fouedjio et al., 2016\)](#), where knots are called *anchor points*, the approach is quite similar, but in each knot the estimation is carried out by weighted variogram fitting [\(Cressie, 1993\)](#), resulting in a dramatically faster algorithm, while the smoothing phase is exactly identical.

Both methodologies require the choice of a stationary covariance model to perform estimation in each knot: [Risser and Calder \(2017\)](#) suggests an anisotropic stationary Matérn model, while [Fouedjio et al. \(2016\)](#) adopts an anisotropic exponential model. Apart from the problem of tuning bandwidth parameters, as the fitting radius  $r$  or the parameters

governing the decay of the smoothing kernel, which can be tackled by means of cross-validation, the main issue comes with the initial choice of the grid where the latent values  $\psi_j^{b_k}$  must be estimated.

## 3.2 A simulated example

To the best of our knowledge, none of these methods have been tested on simulations generated directly by model (4), despite being based on it. In particular, the local likelihood method has only been tested on mixtures of stationary models (Risser and Calder, 2017), while the method presented in (Fouedjio et al., 2016) has been tested exclusively on real world data. An extensive simulation study will be the object of Section 5. Here we focus on an illustrative example of the possible instabilities associated with the above-mentioned estimation methods. For this purpose, we consider the same non-stationary process used to generate Fig 2 (refer to the corresponding captions for its parametric specification ); Fig. 3 displays its spatially-varying parameters  $\mu$ ,  $\sigma$  and  $\Sigma$ , together with three independent realizations of the resulting process. These charts should be taken as reference for comparison with the following Figs. 4 and 5.

Figure 4 reports the estimates of  $\sigma(s)$  and of  $\Sigma(s)$  based on the method of (Fouedjio et al., 2016) (hereafter named FOU16), the estimates of the remaining parameters being available as supplementary material. Each column corresponds to one of the three independent realizations depicted in Fig. 3. Estimates are obtained by applying the method to samples of  $N = 10000$  spatial observations, with  $K = 64$  equispaced knots. Visual inspection of Fig. 4b and 4e suggests that this case is associated with fairly good estimates of the anisotropy ellipses, capturing the behaviour expected from Fig. 3. This is also reflected in good estimates of the other parameters (e.g., of  $\sigma$ , see Fig. 4b). In Fig. 4c and 4f, the estimates of  $\sigma$  appear to be less faithful, even though their scale of variation is correctly captured. In contrast, Fig. 4a and 4d show very irregular variations for the estimated ellipses, as well as a strong influence of the structure of the fixed anchor points in the patterns of the estimated  $\sigma$ .

For the local likelihood methods of (Risser and Calder (2017) (hereafter named RC17), we refer to the results depicted in Fig. 5 – having an analogous structure as Fig. 4. In this case, the method fails to correctly capture the range of variability of  $\sigma(s)$  as well as the variation of eccentricity in the field of  $\Sigma(s)$ , probably due to the greater instability of likelihood optimization of RC17 with respect to the variogram fitting underlying FOU16.

These preliminary observations make us argue that it is of interest to develop a grid-free estimation framework or, at least, a framework where instabilities caused by estimation problems in some knots do not hugely impact on the results of estimation. This will be the object of the next Section.

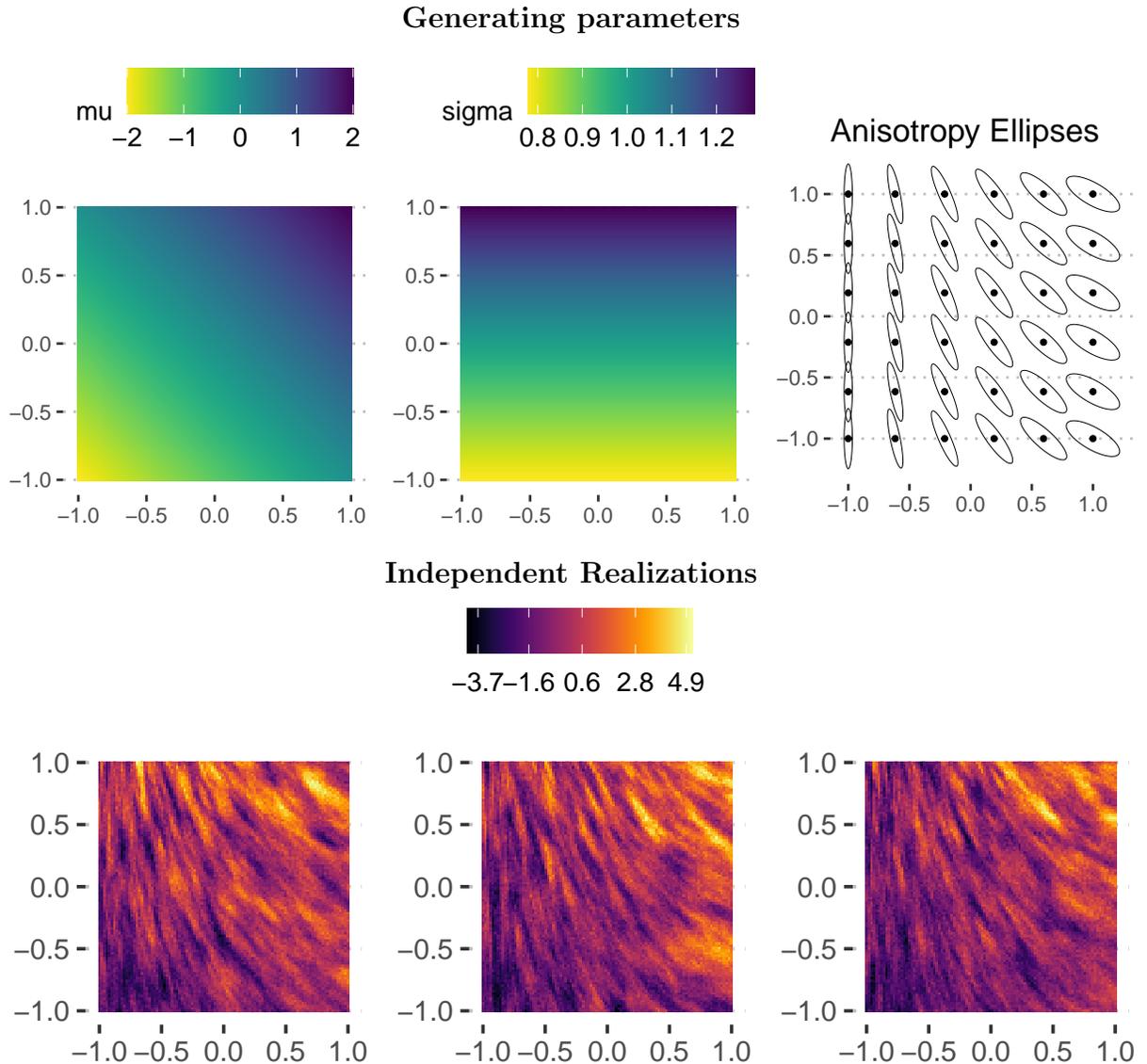


Figure 3: First row: Parameters  $\mu(s) = x_1 + x_2$ , and  $\sigma(s) = e^{\frac{1}{4}x_2}$  and the anisotropy ellipses corresponding to  $\lambda_1(s) = e^{-8+x_1}$ ,  $\lambda_2(s) = e^{-5}$ ,  $\theta(s) = \frac{\pi}{6}(x_1 + 1)$ . Second row: three independent realizations of the corresponding non-stationary Gaussian process.

## 4 Random Domain Decomposition for Non-Stationary Estimation

We here introduce the Random Domain Decomposition (RDD) approach as a generalization of the seminal proposal illustrated in (Menafoglio et al., 2018), adapting it to the non-stationary parametric framework defined by model (4). The cornerstone of the methodology is the introduction of a random partition  $\mathcal{P}$  of  $D$ , generated according to some law  $\mathcal{L}$ . Any realization  $\mathcal{P}^b$  of  $\mathcal{P}$  is then an array  $(\mathcal{P}_1^b, \dots, \mathcal{P}_K^b)$  of mutually disjoint subsets partitioning  $D$ , where the number of elements  $K$  of the partition can, in general, be random as well. A

## Estimates of $\sigma(s)$ and $\Sigma(s)$ using FOU16

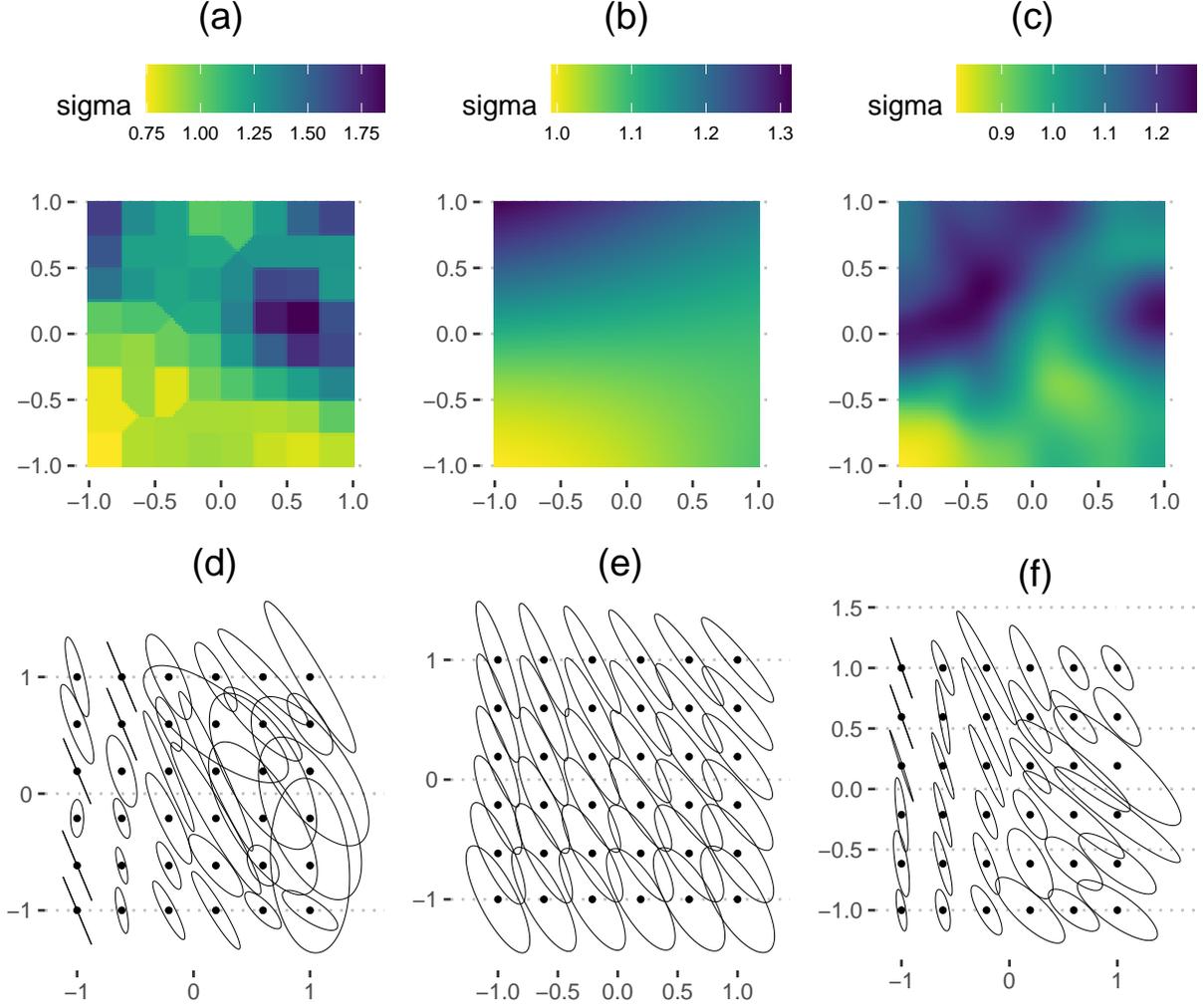


Figure 4: Estimates of the variance  $\sigma(s)$  and of the anisotropy ellipses  $\Sigma(s)$ , for the three independent realizations in Fig. 3, carried out using the method of Fouedjio et al. (2016) (FOU16) from a sample of  $N = 10000$  spatial observations, with  $K = 64$ .

straightforward example of a valid law  $\mathcal{L}$  is the *Random Voronoi Tessellation*, which is the only example so far explored (Secchi et al., 2013; Menafoglio et al., 2018, 2021), and will also be our choice throughout the present work. In this case,  $\mathcal{L}$  is defined via a random vector of spatial locations  $\mathbf{P} \in D^K$ ,  $K$  being fixed, whose components are independent and uniformly distributed on  $D$ . Conditionally on a realization  $\mathbf{P}^b$  of  $\mathbf{P}$ , one defines the corresponding partition as the cells of the Voronoi tessellation of  $D$  defined by  $\mathbf{P}^b$ , that is

$$\mathcal{P}_k^b = \{s \in D \mid d(s, P_k^b) \leq d(s, P_r^b) \forall r \neq k\}, \quad k = 1, \dots, K.$$

Given  $\mathcal{P} \sim \mathcal{L}$ , and the vector of spatially varying parameters  $\psi$ , we define the *weak estimator*  $\psi_*$  as the random variable defined as follows. Conditionally on  $\mathcal{P} = \mathcal{P}^b$ :

1. Assume that, for  $k = 1, \dots, K$ , a stationary Matérn covariance model holds true in  $\mathcal{P}_k^b$

## Estimates of $\sigma(s)$ and $\Sigma(s)$ using RC17

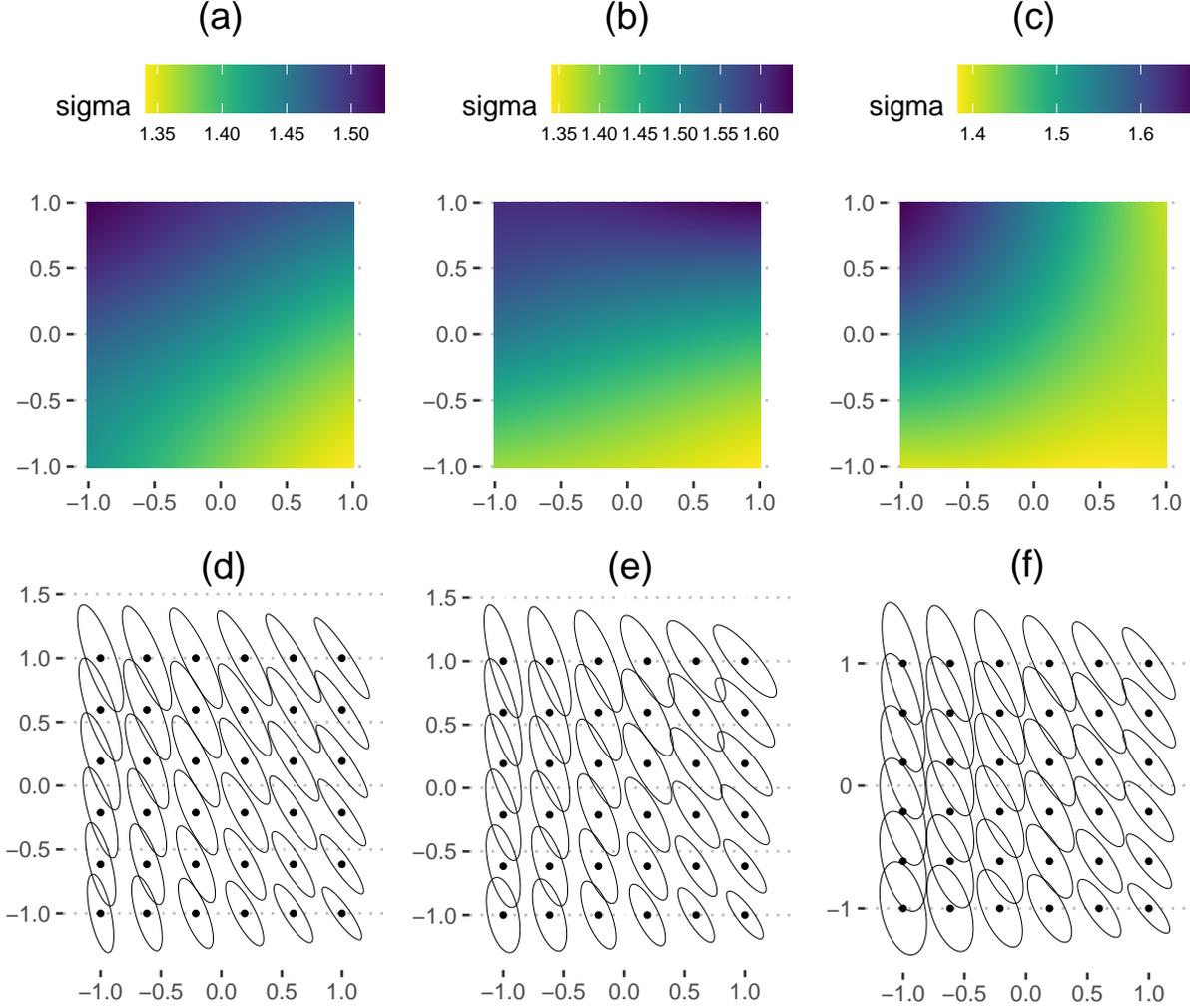


Figure 5: Estimates of the variance  $\sigma(s)$  and of the anisotropy ellipses  $\Sigma(s)$ , for the three independent realizations in Fig. 3, carried out using the method of Risser and Calder (2017) (RC17) from a sample of  $N = 10000$  spatial observations, with  $K = 64$  grid points.

(this is the local stationarity assumption);

2. Assign to each  $\mathcal{P}_k^b$  a (spatially constant) estimate  $\psi_{*k}$  obtained by standard anisotropic variogram fitting and Kriging of the mean in  $\mathcal{P}_k^b$ ;
3. For a given  $s \in D$ , define  $\psi_*(s)$  as  $\psi_{*k(s)}$  where  $k(s)$  is the index of the unique element of the partition containing  $s$ .

Finally, we propose to estimate  $\psi(s)$  with

$$\hat{\psi}(s) = \mathbb{E}_{\mathcal{L}}[\psi_*(s)],$$

where the expectation in the right hand side is computed via crude Monte Carlo: as suggested in (Menafoglio et al., 2018), one can generate, according to  $\mathcal{L}$ ,  $B$  independent

realizations  $\mathcal{P}^b$ ,  $b = 1, \dots, B$ , of the random partition, entailing then the  $B$  realizations  $\psi_*^b$ . This leads to the estimate

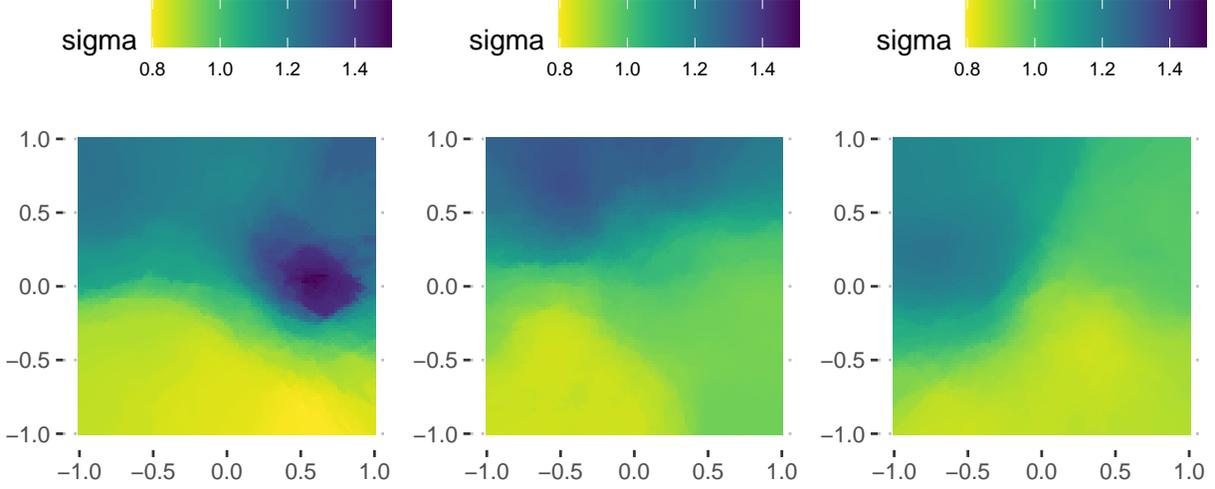
$$\hat{\psi}(s) = \frac{1}{B} \sum_{b=1}^B \psi_*^b(s). \quad (7)$$

Medians could be preferred to simple averaging for the sake of robustness. Using the terminology of ensemble estimation, each  $\psi_*^b$ , which is a piecewise constant random variable, is a *weak estimator* (Breiman, 1996), while  $\hat{\psi}(s)$ , obtained by aggregating all the weak estimators, has the desired property of varying smoothly in space as a result of taking expectations.

From now on, we will assume that  $\mathcal{L}$  is the Random Voronoi Tessellation, so that the above Monte Carlo estimation amounts to (i) generating  $B$  random (and non-regular) grids of knots  $\mathbf{P}^b \stackrel{\text{iid}}{\sim} U(D^K)$ ,  $b = 1, \dots, B$ , where  $U$  is the product measure whose marginal distributions are uniform on  $D$ ; (ii) computing the corresponding random partitions and weak estimators; (iii) output  $\hat{\psi}(s)$  for any given  $s$ .

In our proposal, knots change location for each weak estimator, so that their position have a very weak influence on the final estimation, if any. Moreover, our choice is also robust with respect to poor local estimates: the averaging in equation (7) not only overcomes the need for a final smoothing phase via radial basis functions, but also prevents  $\hat{\psi}(s)$  from being significantly influenced by few anomalous weak estimators. We note that the estimator proposed in (Fouedjio et al., 2016) can be seen, within our framework, as a specific weak estimator corresponding to random knots whose realization happens to be arranged as a regular grid. Another advantageous feature of estimator (7) lays in the fact that, when  $\mathcal{L}$  is the Random Voronoi Tessellation, the only relevant hyperparameter is the integer  $K$  ( $K = 1$  corresponding to standard stationary estimation), which could be easily tuned by means of cross-validation. In Figure 6, we show the estimates obtained by means of RDD for the three realizations drawn in 3 (setting  $K = 8$ ). It is interesting to note the similarities between the heatmaps for  $\sigma$  obtained via RDD and via the method in (Fouedjio et al., 2016): RDD correctly identifies the range of  $\sigma$  and approximately mimic its spatial behavior. However, it avoids unnatural behaviours due to any underlying grid (compare the left panel in Figure 4 against the left panel in Figure 6). Note moreover that the overestimation region of  $\sigma$  (the deep blue part of the heatmaps in the left panels of Figure 4 and 6) is characterized by a much less severe error for RDD, as can be noted by inspecting the colour scale: this is arguably due to the beneficial smoothing effect of averaging many weak predictors. We propose more thorough comparisons in the next Section.

### Estimates by RDD: estimate of $\sigma$



### Estimates by RDD: estimate of anisotropy

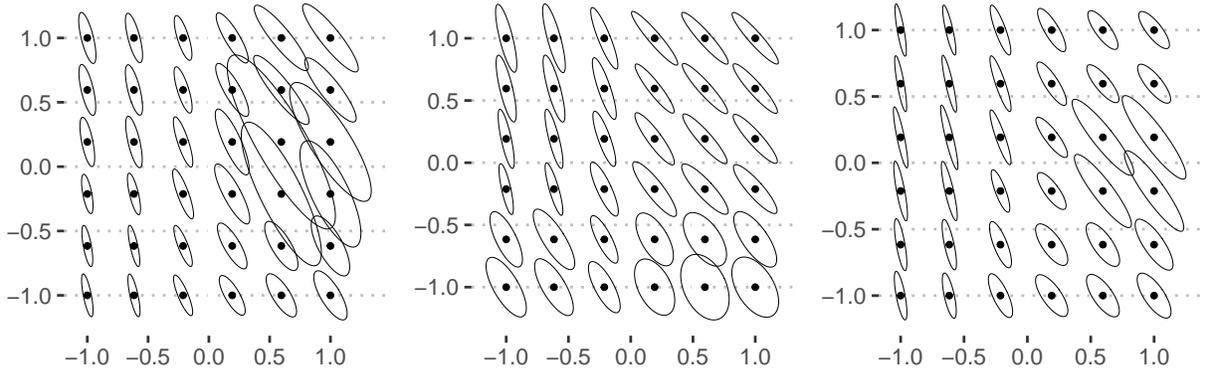


Figure 6: Estimates, obtained by our RDD proposal, carried out on the realizations drawn in Figure 3

## 5 Performance Assessment in Simulations

We show in Fig. 7 the mean and the standard deviation of the spatially averaged error made on each parameter, by performing RDD and the methods (Fouedjio et al., 2016) and (Risser and Calder, 2017) on 10 independent realizations of the non stationary Matérn process described in Figure 2. We show the results for RDD over different sample sizes  $N \in \{1000, 5000, 8000, 10000\}$  and different values of the integer parameter  $K$ , controlling the number of elements of the partitions used to compute the weak estimators;  $B$  is set to 100. To be rigorous, for each independent realization  $p = 1, \dots, 10$ , each parameter  $\psi_j$  and each value of  $N$  and  $K$ , we compute  $RMSE(\psi_j)_{p,N,K}$ . This is the root mean square error for parameter  $\psi_j$  in realization  $p$ , averaged over the spatial domain  $D$ , obtained when applying RDD with those particular values of  $N$  and  $K$ . We then draw in Fig. 7, for each value of  $N$  and  $K$  and each parameter, the average  $RMSE(\psi_j)_{N,K} = \frac{1}{10} \sum_{p=1}^{10} RMSE(\psi_j)_{p,N,K}$

## Spatially averaged errors and standard deviations over 10 realizations

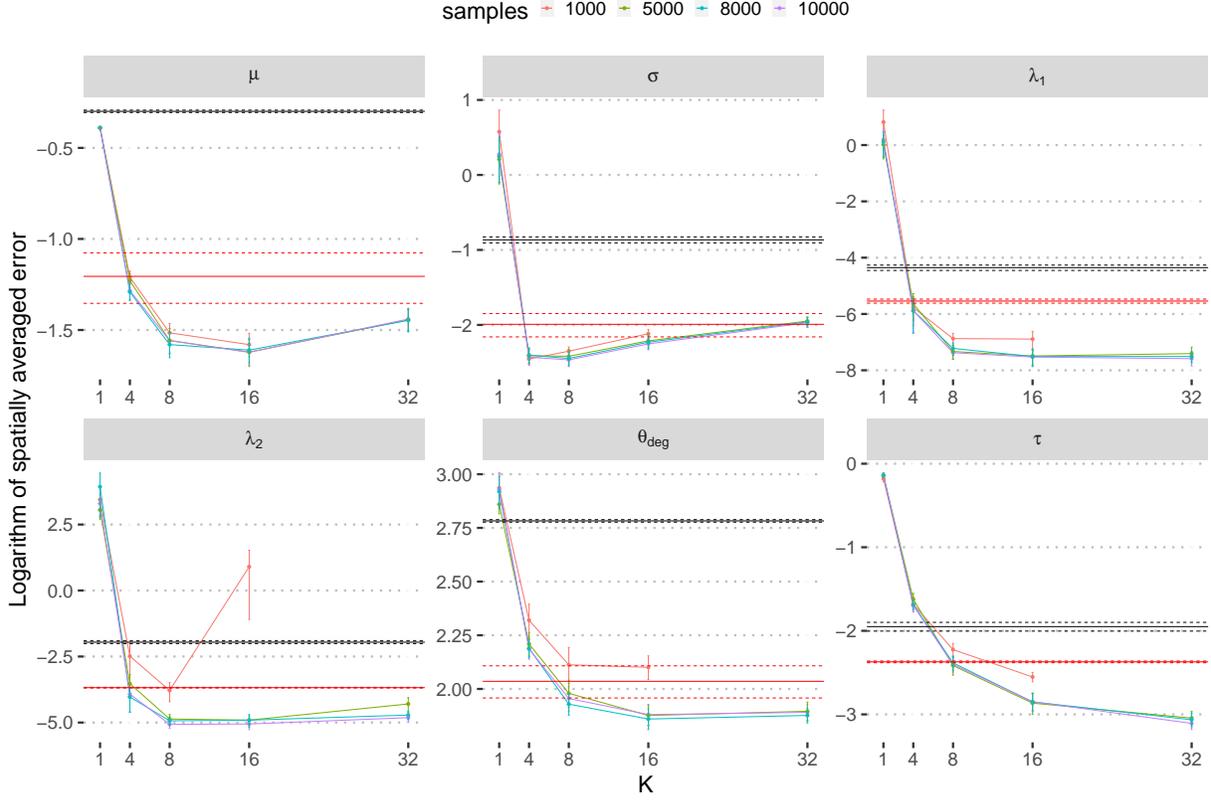


Figure 7: Mean and standard deviations of the spatially averaged errors made by performing RDD on 10 independent realizations of the non-stationary Matérn process described in Figure 2, as a function of  $K$ . Red horizontal lines refer to FOU16, black horizontal lines to RC17, in both cases applied to the most favorable sample size  $N = 10000$ . Remaining colored lines refer to RDD applied for different sample sizes  $N \in \{1000, 5000, 8000, 10000\}$ .

computed over the 10 different realizations (the solid points in the chart) and the corresponding standard deviation (the intervals around the points).

In Fig. 7, the results for FOU16 (in red) and RC17 (black) are drawn in correspondence of  $N = 10000$  (i.e., the most favorable sample size), and refer to a grid of 64 equi-spaced knots.

Notice that the parameter  $K$  controls the *locality* of the weak estimators. Indeed, larger values of  $K$  correspond to finer partitions and hence to more local estimates, while  $K = 1$  means standard stationary estimates. In our case, coherently with the non-stationary nature of the process under study, the quality of estimation benefits from higher values of  $K$ , with  $K = 4$  or  $K = 8$  being the best choices.

The method of (Fouedjio et al., 2016) produces better estimates than those of (Risser and Calder, 2017), probably because of better stability in variogram estimation with respect to maximum likelihood estimation, which is coherent with the empirical evidence in Fig. 4 and Fig. 5. It can be noted that RDD seems to perform better than the other methods, even when  $N$  is only moderately large; interestingly, RDD for a sample of size  $N = 1000$  seems

to perform comparably with the method of FOU16 when the sample size is  $N = 10000$ . We point out that the method in (Risser and Calder, 2017) has been implemented by the authors in the R package `convospat`, which has then been used for our tests, while the proposal by (Fouedjio et al., 2016) did not come with software. This latter method was later independently implemented in R package `LocallyStationaryModels` (De Carlo and Crippa, 2022), which was used for our tests.

## 6 A Case Study: Analysis of Colorado Rain data

We focus on the dataset used in the seminal work of (Paciorek and Schervish, 2006) to test the Bayesian estimation framework (5). We here compare our RDD estimates with the estimates produced by such model, obtained in (Paciorek and Schervish, 2006) and reproduced in (Risser and Turek, 2020), where the `BayesNSGP` R package is introduced. The dataset consists of the average yearly precipitations in Colorado during 1981, measured on  $N = 271$  different locations, a limited sample size which makes model (5) computable. Colorado is an interesting example of a spatial domain where non-stationary modeling may be appropriate: as can be seen from Fig. 8a, this state is characterized by a very heterogeneous orography, with flat plains crossed by large rivers in the East and mountain peaks in the West, as well as a large plateau encompassed by mountains in the South. This results in different precipitation amounts across the country, as we show in Fig. 8b, where the yearly average in each observed location is drawn.

Fig. 8c and 8d show the estimates for  $\mu$  and  $\sigma$  obtained via RDD with  $K = 6$ . Parameter  $\mu$  seems to coherently capture the variation due to orography, with higher values corresponding to concentration of mountains, and lower values in flat plains and plateaus. More interestingly, the estimated standard deviation  $\sigma$  peaks in the area where elevation varies more abruptly, i.e., in the South area in correspondence of the large plateau.

Notably, in (Paciorek and Schervish, 2006) and (Risser and Turek, 2020), when applying model (5) to the same dataset,  $\mu$  and  $\sigma$  were not allowed to vary according to GP priors, which were only elicited for  $\lambda_1$ ,  $\lambda_2$  and  $\theta$ . More precisely,  $\mu$  and  $\sigma$  are constrained to be (unknown) constants in (Paciorek and Schervish, 2006), or simple parametric functions of the terrain slope in (Risser and Turek, 2020). This is probably due to the notoriously difficult MCMC computations mentioned in Section 3. Indeed, we were not able to reach a satisfactory convergence of Markov Chains by imposing general GP priors on all parameters, using the package `BayesNSGP`.

We can however compare how RDD and model (5) estimate the anisotropy ellipses, which are drawn in Fig. 9. While both methods tend to agree in assigning shorter correlation ranges in correspondence of mountain peaks (see the first row of Fig. 9), the inclination of the anisotropy ellipses (i.e.,  $\theta$ ) is quite different. A closer inspection of the estimation of the parameter  $\theta$  shows that, similarly to the method FOU16 in the simulation study described in the previous section, the estimate for  $\theta$  is affected by the grid of knots which was used to approximate the GP priors. Instead, RDD produces for  $\theta$  a smooth, interpretable and fully data-driven map, as in the case of  $\mu$  and  $\sigma$  in Fig. 8.

## Colorado: Orography and Yearly precipitation in 1981

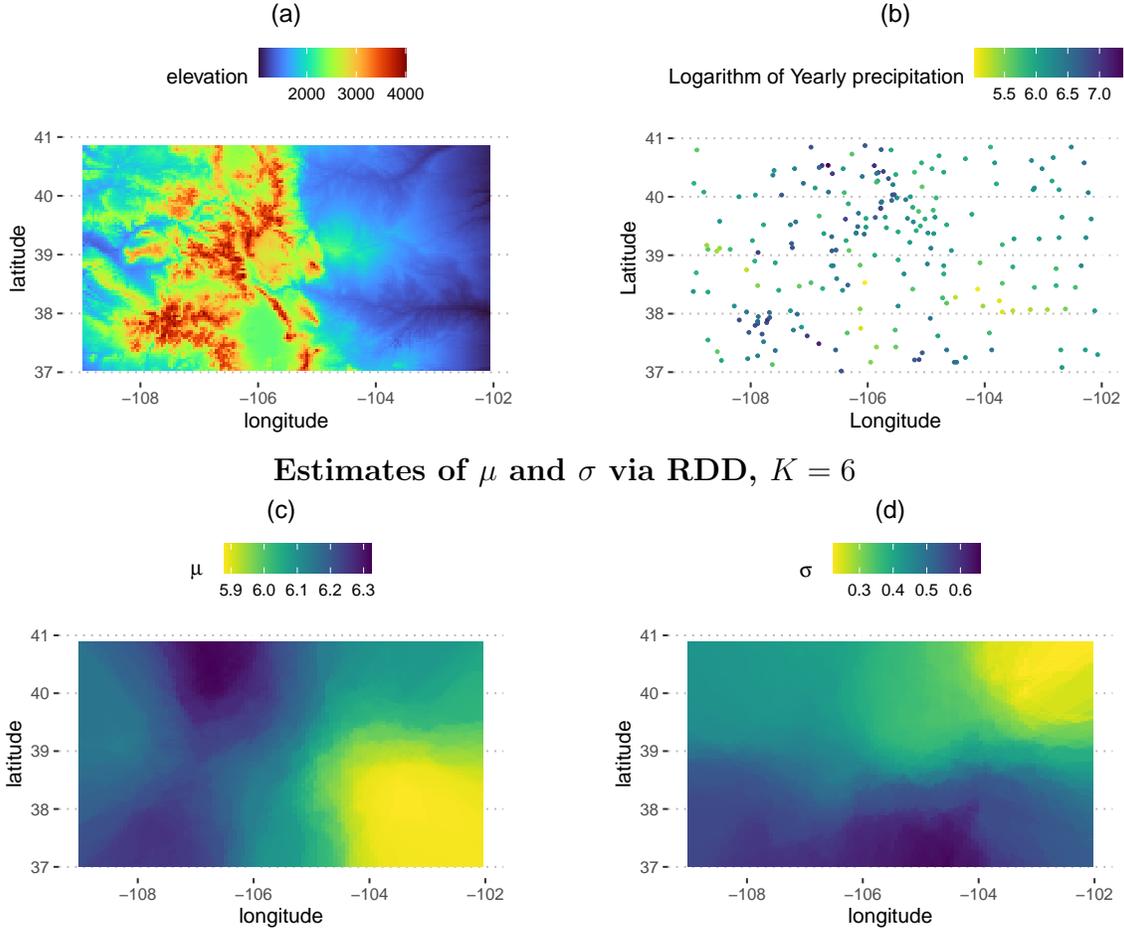


Figure 8: (a) Orography of the state of Colorado; (b) logarithm of the average yearly precipitation in the 217 observed locations; (c) and (d) estimates, obtained via RDD, for the parameter functions  $\mu$  and  $\sigma$ , respectively.

## 7 Conclusion

In this work we have introduced a novel, robust estimation framework for non-stationary spatial processes in which non-stationarity is modeled by the Matérn model introduced in (Paciorek and Schervish, 2006). We have empirically shown the possible drawbacks of the current estimation methods, on a fully non-stationary simulation study which, up to our knowledge, was never attempted before. Some of these drawbacks are arguably due to the presence of a preliminary, arbitrary grid of knots: this is particularly true for the estimation frameworks developed in (Fouedjio et al., 2016) and in (Paciorek and Schervish, 2006; Risser and Turek, 2020). RDD overcomes altogether this issue, providing estimations which are at the same time smooth and fully data-driven, for all the parameters of model (4). On the other hand, the analyses carried out in this work give rise to a number of interesting question and issues. It would be of great interest to compare the performance of RDD against other methods when it comes to *prediction* and *uncertainty quantification* (e.g., through stochastic simulation) of the values of the process of interest in new unobserved

Estimates of anisotropy and  $\theta$  via RDD and via model (5) as fitted in (Risser and Turek, 2020)

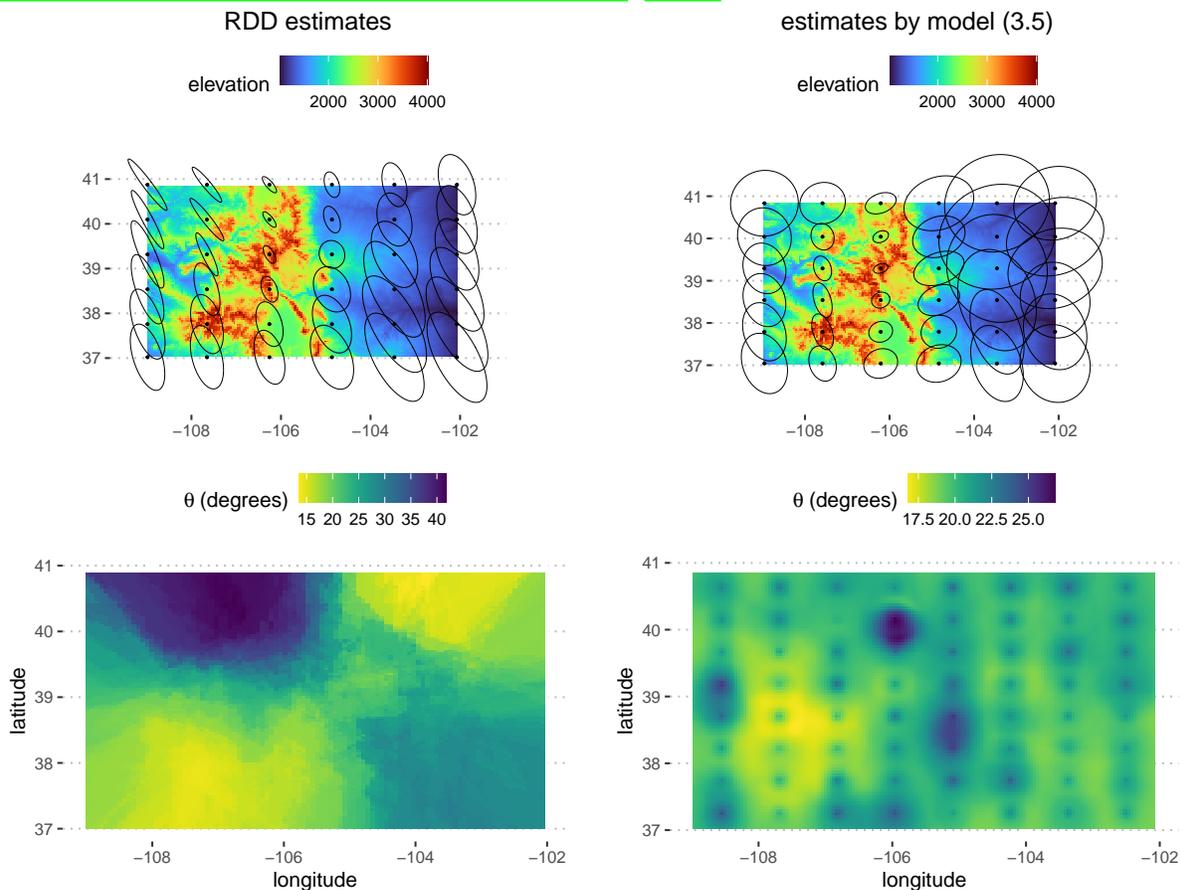


Figure 9: Left panels: RDD estimation of the anisotropy ellipses in representative locations (top) and of  $\theta$  (bottom). Right panels: RC17 estimation of the anisotropy ellipses in representative locations (top) and of  $\theta$  (bottom); these were obtained by fitting model (5) using the `BayesNSGP` package.

locations. Indeed, artifacts in estimated parameters and instabilities affecting methods for their inference (e.g., induced by the grid of estimation) may potentially propagate over new realizations of the field, hindering their use in practice. Another area of further exploration lays in possible refinement of the law governing the RDD partitions, which might be shaped to include information which is available a priori. Moreover, while model (4) has been formulated for scalar and multivariate data, the development of an analogous theoretical framework for functional data would be of great interest and will be the scope of future work. Developing an asymptotic theory for RDD estimators, whether in finite or infinite-dimensional settings, remains a critical yet open challenge. Such a theory could provide additional support for the excellent performance of the proposed method, which has been empirically demonstrated on both simulated and real data in this work.

## 8 Supplementary Material

The R code to reproduce all the analyses described in the present work is available at [https://github.com/RiccardoScimone/RDD\\_code\\_paper](https://github.com/RiccardoScimone/RDD_code_paper).

### Acknowledgements

Alessandra Menafoglio and Piercesare Secchi acknowledge the initiative “Dipartimento di Eccellenza 2023–2027”, MUR, Italy, Dipartimento di Matematica, Politecnico di Milano.

### References

- Anderes, E. B. and Stein, M. L. (2011). Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, 102(3):506–520.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 23:123–140.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics Applied probability and statistics. J. Wiley & Sons, New York, revised edition edition.
- De Carlo, C. and Crippa, L. (2022). *LocallyStationaryModels*. R package.
- Fouedjio, F. (2017). Second-order non-stationary modeling approaches for univariate geostatistical data. *Stochastic Environmental Research and Risk Assessment*, 31:1887–1906.
- Fouedjio, F., Desassis, N., and Rivoirard, J. (2016). A generalized convolution model and estimation for non-stationary random functions. *Spatial Statistics*, 16:35–52.
- Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12(5):469–483.
- Gelfand, A. E. and Schliep, E. M. (2016). Spatial statistics and gaussian processes: A beautiful marriage. *Spatial Statistics*, 18:86–104. Spatial Statistics Avignon: Emerging Patterns.
- Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. *Bayesian statistics*, 6(1):761–768.
- Li, Y. and Sun, Y. (2019). Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation. *Statistica Sinica*, 29(3):1209–1231.

- Menafoglio, A., Gaetani, G., and Secchi, P. (2018). Random domain decompositions for object-oriented kriging over complex domains. *Stochastic Environmental Research and Risk Assessment*, 32:3421–3437.
- Menafoglio, A., Pigoli, D., and Secchi, P. (2021). Kriging riemannian data via random domain decompositions. *Journal of Computational and Graphical Statistics*, 30(3):709–727.
- Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506.
- Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA.
- Risser, M. D. and Calder, C. A. (2015). Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, 26(4):284–297.
- Risser, M. D. and Calder, C. A. (2017). Local likelihood estimation for covariance functions with spatially-varying parameters: The convospat package for r. *Journal of Statistical Software*, 81(14):1–32.
- Risser, M. D. and Turek, D. (2020). Bayesian inference for high-dimensional nonstationary gaussian processes. *Journal of Statistical Computation and Simulation*, 90(16):2902–2928.
- Secchi, P., Vantini, S., and Vitelli, V. (2013). Bagging voronoi classifiers for clustering spatial functional data. *International Journal of Applied Earth Observation and Geoinformation*, 22:53–64. *Spatial Statistics for Mapping the Environment*.
- Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. J. Wiley & Sons, New York.
- Stein, M. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York.
- Tibshirani, R. and Hastie, T. J. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82:559–567.
- Turek, D. and Risser, M. (2022). *BayesNSGP: Bayesian Analysis of Non-Stationary Gaussian Process Models*. R package version 0.1.2.

## MOX Technical Reports, last issues

Dipartimento di Matematica  
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 12/2025** Alessandro Andrea Zecchi, Claudio Sanavio, Simona Perotto e Sauro Succi  
*Improved amplitude amplification strategies for the quantum simulation of classical transport problems*
- 11/2025** Tonini, A.; Dede', L.  
*Enhanced uncertainty quantification variational autoencoders for the solution of Bayesian inverse problems*
- 10/2025** Botti, M.; Mascotto, L.; Vacca, G.; Visinoni, M.  
*Stability and interpolation estimates of Hellinger-Reissner virtual element spaces*
- 09/2025** Quarteroni, A.; Gervasio, P.; Regazzoni, F.  
*Combining physics-based and data-driven models: advancing the frontiers of research with Scientific Machine Learning*
- 08/2025** Botti, M.; Fumagalli, I.; Mazzieri, I.  
*Polytopal discontinuous Galerkin methods for low-frequency poroelasticity coupled to unsteady Stokes flow*
- 07/2025** Patanè, G.; Nicolussi, F.; Krauth, A.; Gauglitz, G.; Colosimo, B. M.; Dede', L.; Menafoglio, A.  
*Functional-Ordinal Canonical Correlation Analysis With Application to Data from Optical Sensors*
- 06/2025** Torzoni, M.; Manzoni, A.; Mariani, A.  
*Enhancing Bayesian model updating in structural health monitoring via learnable mappings*
- 04/2025** Andrini, D.; Magri, M.; Ciarletta, P.  
*Nonlinear morphoelastic theory of biological shallow shells with initial stress*
- 05/2025** Buchwald, S.; Ciaramella, G.; Verani, M.  
*Greedy reconstruction algorithms for function approximation*
- 02/2025** Corda, A.; Pagani, S.; Vergara, C.  
*Influence of patient-specific acute myocardial ischemia maps on arrhythmogenesis: a computational study*