Statistics and Computing (2026) 36:261
https://doi.org/10.1007/s11222-026-10990-x

ORIGINAL PAPER

regTPS: a novel approach to approximate a Gaussian random field for Bayesian spatial modeling

Joaquin Cavieres[1,2] · **Sebastian Krumscheid**[3]

Received: 22 January 2026 / Accepted: 23 September 2026
© The Author(s) 2026

Abstract

Gaussian random field is a ubiquitous model for spatial phenomena in diverse scientific disciplines. Its approximation is often crucial for computational feasibility in simulation, inference, and uncertainty quantification. The Karhunen–Loéve Expansion provides an optimal low-rank representation of a (spatial) stochastic process, but its practical implementation relies heavily on how the underlying covariance structure is specified. In this work, we propose a novel approach called regTPS, which constructs a spatial covariance by formulating the inverse of a regularized thin plate spline operator. By computing the eigendecomposition of the regularized biharmonic operator, regTPS induces a smoothness prior that hierarchically allocates variance across spatial frequencies. We evaluate the statistical performance and computational efficiency of regTPS relative to the established stochastic partial differential equations approach within an MCMC algorithm in two different simulation designs. Across various simulation scenarios, the regTPS method achieved predictive accuracy comparable to the SPDE approach. Moreover, regTPS substantially reduced the effective parameter dimension, with reductions of approximately 43–52% for large sample sizes. Also, within a mesh independent simulation setting, regTPS consistently maintains its predictive accuracy under both Matérn and exponential covariance functions, keeping an optimal balance between basis-dimension reduction and sampling efficiency. Lastly, application to real NO_2 data concentrations using 10-fold and leave-one-out cross-validation establishes that regTPS obtain a competitive predictive accuracy against SPDE and nearest-neighbor Gaussian process models. Thus, regTPS presents a rigorous, low-dimensional, and computational efficient alternative for Bayesian spatial modeling.

Keywords Gaussian random field · thin plates splines · Bayesian inference · Spatial modeling

1 Introduction

Spatial models have attracted growing interest within the scientific community in recent years due to their ability to model spatial dependencies between observations (measurements) in a two-dimensional space, an aspect often inadequately addressed by conventional statistical approaches. In particular, referenced spatial data (also called geostatistical data) involve the analysis of data collected by sampling a continuous spatial process, denoted $u(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2$, at a finite number of spatial locations $\mathbf{s}_i \in \mathcal{D} \subset \mathbb{R}^2$, where $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)^\top$ typically represents coordinates such as latitude and longitude. Because direct observations of $u(\mathbf{s})$ are often subject to measurement error, a common assumption is that these errors are additive. Denoting by y_i the observed values at spatial location $\mathbf{s}_i$, a general spatial model (geostatistical model) can be expressed as:

$$y_i = \mu + u(\mathbf{s}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where the error terms ε_i are commonly assumed to be independent and identically distributed with zero mean and standard deviation σ_ε (Cressie 1989; Møller 2013). This model is easily extended to a regression framework by defining the mean structure as $\mu = \mathbf{x}_i^\top \boldsymbol{\beta}$, where $\mathbf{x}_i$ includes

✉ Joaquin Cavieres
joaquin.cavieres@imng.uni-stuttgart.de

Sebastian Krumscheid
sebastian.krumscheid@kit.edu

[1] Optimization and Inverse Problems, Universität Stuttgart, Stuttgart, Germany

[2] Spatial Data Science and Statistical Learning, Universität Göttingen, Göttingen, Germany

[3] Scientific Computing Center - Uncertainty Quantification, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Published online: 06 November 2026

explanatory variables related to y_i. The primary goal of spatial modelling is to predict the value of the underlying spatial process $u(\mathbf{s})$ at unsampled locations or to estimate the average of $u(\mathbf{s})$ over a region of interest $\mathcal{D}$. A common strategy for handling uncertainty in the continuous spatial stochastic process $u(\mathbf{s})$ is to model it as a Gaussian random field (GRF). Despite its flexibility, assuming a GRF as the spatial random field becomes computationally intensive when dealing with large datasets, motivating the development of scalable approximations. A comprehensive review of scalable spatial models is provided by Heaton et al. (2019). Notable approaches include low-rank approximations, such as Gaussian predictive process models (Banerjee et al. 2008), as well as methods based on sparse matrix algebra, including the Vecchia approximation (Vecchia 1988; Katzfuss and Guinness 2021). One of the most popular approaches to approximating a GRF was proposed by Lindgren et al. (2011), who developed an efficient framework based on stochastic partial differential equations (SPDEs). This approach establishes an explicit link between GRFs and a Gaussian Markov random field (GMRF) (Rue and Held 2005), yielding sparse precision matrices and enabling fast computation without sacrificing accuracy in latent Gaussian models. The SPDE method has been widely adopted across a range of disciplines (Lindgren et al. 2022). An alternative to GRF-based models is the use of radial basis functions (RBFs; Wendland, 2004, 2006), and in particular thin plate splines (TPS). Originally introduced by Duchon (1977), TPS are spline-based techniques for spatial interpolation and smoothing (Wahba 1990; Green 1987). However, TPS suffer from poor scalability, with computational demands increasing rapidly with the number of observations. To overcome this limitation, Wood (2003) proposed thin plate regression splines (TPRS), which approximate the spline basis using truncated spectral decomposition, yielding improved computational efficiency (Wood 2017; Cavieres et al. 2023). Within a Bayesian framework, early work by Wahba (1983) and Nychka (1988) introduced methods to construct credible intervals for smoothing splines. Building on these ideas and the correspondence between spline smoothing and stochastic processes established by Kimeldorf and Wahba (1970), White (2006) derived a spatial prior based on TPS, enabling Bayesian inference without resorting to Markov chain Monte Carlo methods. For broader perspectives on Bayesian spatial modeling, see Handcock and Stein (1993), Gelfand (2012), and Banerjee et al. (2014). In spatial statistics and uncertainty quantification, the Karhunen–Loève expansion (KLE) is a widely used spectral method for approximating GRFs (Lord et al. 2014; Alexanderian 2015). The KLE provides a truncated series representation of a GRF in terms of eigenfunctions and eigenvalues derived from the covariance kernel. Early theoretical foundations were laid by Loève (1978) and Adler (1981), with later applications in stochastic finite element methods (Ghanem and Spanos 1991). In geostatistics, KLE has been used to model spatial variability through orthogonal decompositions of covariance structures (Zhu and Ghosh 1997; Brenner et al. 2017). Further applications include environmental modeling, Bayesian inverse problems (Marzouk et al. 2009; Biegler et al. 2011), Gaussian process emulation (Higdon et al. 2008), and hierarchical Bayesian frameworks (Cotter et al. 2010). Despite its advantages, the KLE requires the solution of an often expensive eigenvalue problem, motivating ongoing research into efficient numerical approximations, including adaptive and sparse variants (Le Maître and Knio 2010; Litvinenko et al. 2022).

In this work, we propose a novel methodology to approximate a GRF based on the KLE, where the covariance function (covariance kernel) is defined through the inverse of a regularized elliptic operator. We refer to this approach as regTPS. The proposed framework establishes a direct and explicit connection between classical spline smoothing theory and spatial stochastic process modeling, yielding a covariance-driven KLE whose modes inherit the analytical and smoothing properties of thin plate splines. The main contributions of this work can be summarized as follows:

1. We develop a theoretical approximation of a GRF covariance function defined as the inverse of a regularized elliptic operator,

   $$L_\alpha = \mathbf{I} + \alpha\Delta^2.$$

   While the Bayesian interpretation of penalized regression is well established, our approach explicitly derives the associated spectral representation using Hilbert-Schmidt theory and Mercer's theorem. This provides a rigorous foundation for the KLE and clarifies its relationship with the TPS geometry. The resulting probabilistic formulation unifies penalized regression and spatial process modeling, enabling principled uncertainty quantification through the full Bayesian posterior over latent coefficients and hyperparameters.
2. Through regularization, we convert the conditionally positive definite TPS kernel into a strictly positive definite Hilbert-Schmidt kernel, thereby ensuring the validity of the Mercer decomposition. In contrast to standard constructions, we provide an explicit spectral treatment of the polynomial null space: eigenfunctions with zero bending energy are retained and assigned unit prior variance, rather than being implicitly constrained or projected out.
3. The KLE basis functions are obtained from the eigensystem of the regularized elliptic operator and inherit the smoothness hierarchy induced by the bending-energy functional. The resulting eigenvalue ordering naturally separates unpenalized polynomial trends from increasingly penalized high-frequency modes, providing both

theoretical justification and practical guidance for adaptive truncation of the expansion.
4. The proposed regTPS framework produces spatially smooth realizations by construction, as smoothness is directly encoded in the operator-based prior via the bending-energy penalty. This contrasts with many parametric covariance models, such as the Matérn class, where smoothness must be carefully controlled through additional parameters, and provides a natural prior for applications requiring guaranteed regularity of spatial processes.
5. By explicitly assigning unit prior variance to the polynomial null space, the regTPS approach accommodates large-scale spatial trends without imposing smoothing penalties. This treatment constitutes a theoretical and practical advantage over standard KLE formulations, in which the null space is often discarded or handled inadequately.

For comparative purposes, we conduct spatial modeling using the widely adopted SPDE approach under Markov chain Monte Carlo (MCMC) sampling, considering both simulated data and a real data application. The remainder of the paper is organized as follows. Section 2 reviews GRFs and KLE theories, followed by a detailed discussion of TPS and their connection to Hilbert-Schmidt kernels, where the proposed regTPS methodology is introduced. In Sect. 3, these theoretical developments are extended to a Bayesian framework for spatial modeling. Section 4 presents numerical analyzes on simulated data, including comparisons with the SPDE approach in terms of root mean squared error (RMSE), mean absolute error (MAE), and R^2, as well as computational efficiency, under various scenarios and covariance specifications. In Sect. 5, the SPDE and regTPS approaches are applied to NO_2 concentration data in Germany. Predictive performance is evaluated using leave-one-out cross-validation and K-fold cross-validation, along with a comparison of computational runtime. Additionally, we include the spNNGP model (Finley et al. 2022), with the purpose of providing a comprehensive and complete benchmark, as it represents another well-established framework for Bayesian spatial modeling. In Sect. 6, we discuss the results of our analyzes, and Sect. 7 concludes the paper.

2 Methodology

2.1 Gaussian random field

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded index set representing the spatial domain, commonly $d = 2$ for the spatial domain. A spatial random field can be defined as a function $\mathbf{u}(\mathbf{s}, \omega) : \mathcal{D} \times \Omega \to \mathbb{R}$, where $\mathbf{s} \in \mathcal{D}$ is a vector of spatial coordinates $\mathbf{s} = (s_1, s_2)$, and $\omega \in \Omega$ is a generic outcome of the sample space. Furthermore, a spatial random field is a collection of random variables at each spatial coordinate $\mathbf{s}$ in $\mathcal{D}$. Now, let's consider a set of spatial locations $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_N)^\top$ associated with a set of random variables $\mathbf{u}(\mathbf{s}, \omega) = \{u(\mathbf{s}_1, \omega), \ldots, u(\mathbf{s}_N, \omega)\}$, where N is the number of observed random variables. For simplicity in notation, we will only consider to $\mathbf{u}(\mathbf{s})$ as the spatial random field instead of $\mathbf{u}(\mathbf{s}, \omega)$. A spatial random field is called a *Gaussian* random field (GRF) when any finite set of its realizations $\{u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_N)\}$ follows a multivariate normal distribution, i.e., $\{u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_N)\} \sim \mathcal{N}_N(\boldsymbol{\mu}, \Sigma)$ (Adler 2010; Lord et al. 2014). A GRF is entirely defined by its mean function $\mu(\boldsymbol{s}) = \mathbb{E}[u(\boldsymbol{s})]$ and its covariance function $C(\mathbf{s}, \mathbf{s}') = \mathrm{Cov}(u(\mathbf{s}), u(\mathbf{s}'))$. Moreover, a GRF is stationary if the associated finite-dimensional distributions of the field are invariant under shifts in space, and isotropic, if the covariance between any two points depends solely on the distance separating them, meaning $C(\mathbf{s}, \mathbf{s}') = C(\mathbf{d})$, where $\mathbf{d} = \|\mathbf{s} - \mathbf{s}'\|$ (Euclidean distance) (Abrahamsen 1997).

2.2 The Karhunen–Loéve expansion

Consider a second-order (finite variance) spatial random field $u(\mathbf{s})$ with mean $\mu(\mathbf{s})$ and covariance $C(\mathbf{s}, \mathbf{s}')$. Covariance functions are symmetric and positive semidefinite functions that belong to the class of Hilbert-Schmidt kernels (in the Euclidean space, these are functions with finite Frobenius norm; Abrahamsen (1997)). These properties ensure that the associated covariance operator admits a countable orthonormal basis of eigenfunctions, with the corresponding eigenvalues being real and non-negative (Uribe et al. 2020). Considering Mercer's theorem (Mercer 1909), the covariance kernel can be represented by a series expansion based on the spectral representation of the covariance operator (Kolmogorov and Fomin 1975) as:

$$C(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{s}) \phi_k(\mathbf{s}'), \tag{2}$$

where $\lambda_k \in [0, \infty)$ and $\phi_k(\mathbf{s}) : \mathcal{D} \to \mathbb{R}$ (where $\phi_k(\mathbf{s}) \in L^2(\mathcal{D})$, with $L^2(\mathcal{D})$ as the Hilbert space of square-integrable functions (Uribe et al. 2020)). In this way, a spatial random field $u(\mathbf{s})$ can be approximated by $\tilde{u}(\mathbf{s})$ using the KLE as an infinite series of orthogonal functions (eigenfunctions) and uncorrelated random variables as:

$$u(\mathbf{s}) \approx \tilde{u}(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{k=1}^{M} \sqrt{\lambda_k}\, \phi_k(\mathbf{s})\, \epsilon_k, \tag{3}$$

where $u(\mathbf{s})$ is the GRF at spatial location $\mathbf{s}$, $\mu(\mathbf{s})$ is the mean function of the GRF, M is the truncation parameter (denot-

ing the number of retained Karhunen–Loéve basis functions), λ_k are the eigenvalues, $\phi_k(\mathbf{s})$ are the eigenfunctions forming an orthonormal basis, and ϵ_k are uncorrelated i.i.d. random errors with $\epsilon_k \sim \mathcal{N}(0,1)$ (Fasshauer and McCourt 2015; Lord et al. 2014). The eigenfunctions $\phi_k(\mathbf{s})$ and the eigenvalues λ_k are determined by the covariance function $C(\mathbf{s},\mathbf{s}')$ of the GRF, specifically by solving the following Fredholm integral eigenvalue problem over the spatial domain $\mathcal{D}$:

$$\int_{\mathcal{D}} C(\mathbf{s},\mathbf{s}')\phi_k(\mathbf{s}')\,d\mathbf{s}' = \lambda_k\phi_k(\mathbf{s}), \tag{4}$$

whose analytical solution exists only for specific cases of covariance functions (Ghanem and Spanos 2003). Covariance kernels play a central role in random field modeling by characterizing the spatial correlation structure of the field. Among the widely used and flexible families of isotropic covariance functions are the Whittle-Matérn kernels (Matérn 1960), which form a class of Hilbert-Schmidt kernels. These kernels provide a tunable framework to model varying degrees of smoothness and spatial dependence, and are mathematically defined as

$$C(\mathbf{d}) = \sigma_u^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\mathbf{d}}{\rho}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\mathbf{d}}{\rho}\right), \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function, $K_\nu(\cdot)$ is the modified Bessel function of the second kind, ρ is the range parameter (correlation length), and $\nu > 0$ is the smoothness parameter. Particular cases of Whittle-Matérn kernels include $C_{1/2}(\mathbf{d}) = \sigma_u^2 \exp\left(-\frac{\mathbf{d}}{\rho}\right)$, called the *exponential* kernel, and $C_\infty(\mathbf{d}) = \sigma_u^2 \exp\left(-\frac{\mathbf{d}^2}{2\rho^2}\right)$, called the *squared exponential* (also known as Gaussian) covariance kernel, respectively (Banerjee et al. 2014; van Lieshout 2019).

2.3 Thin plate splines as a Hilbert–Schmidt kernel

Consider a set of spatial data points $\{(\mathbf{s}_i, y_i)\}_{i=1}^N$, where $\mathbf{s}_i = (s_{i1}, s_{i2}) \in \mathcal{D} \subset \mathbb{R}^2$ represent the spatial locations and $y_i \in \mathbb{R}$ are the observed values. Our main objective is to construct an interpolant $f : \mathcal{D} \to \mathbb{R}$ such that $f(\mathbf{s}_i) = y_i$ for $i = 1,\ldots,N$. This interpolant is built as a linear combination of *radial basis functions* (RBFs) centered at the spatial data points such that:

$$f(\mathbf{s}) = \sum_{i=1}^{N} c_i \phi(\|\mathbf{s} - \mathbf{s}_i\|), \tag{6}$$

where $\phi : [0,\infty) \to \mathbb{R}$ is the RBF, $\|\cdot\|$ denotes the Euclidean norm, c_i are the coefficients to be determined, and $\mathbf{s}_i \in \mathbb{R}^d$ are the known spatial locations. Originally introduced by Duchon (1977), thin plate splines (TPS) is a special case of RBF interpolation derived from variational principles and widely used for smooth surface fitting. For TPS, the RBF (aka kernel function) is $\phi(\mathbf{d}) = \mathbf{d}^2\log(\mathbf{d})$, where the natural setting for a smoothing TPS is the Sobolev space $H^m(\mathcal{D})$. For a smoothing TPS defined in a two-dimensional space, the bending energy is defined for functions $f \in H^2(\mathcal{D})$. In this way, for $L^2(\mathcal{D})$ (the space of square-integrable functions with inner product $\langle f, g\rangle_{L^2} = \int_{\mathcal{D}} f(\mathbf{s})g(\mathbf{s})\,d\mathbf{s}$) and $H^2(\mathcal{D})$ (the space of functions in $L^2(\mathcal{D})$), the bending energy seminorm $J(f)$ for a two-dimensional TPS is given by Wahba (1990); Green and Silverman (1993); Cavieres et al. (2023):

$$J(f) = \int_{\mathcal{D}} \left[\left(\frac{\partial^2 f}{\partial s_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial s_1 \partial s_2}\right)^2 + \left(\frac{\partial^2 f}{\partial s_2^2}\right)^2\right] d\mathbf{s} = \int_{\mathcal{D}} |\Delta f(\mathbf{s})|^2\,d\mathbf{s}, \tag{7}$$

where Δ is the Laplacian operator and $J(f)$ is a seminorm on $H^2(\mathcal{D})$. Specifically, $J(f) = 0$ if and only if $f(\mathbf{s})$ is a polynomial of total degree less than 2 (i.e., $f(\mathbf{s}) = c_0 + c_1 s_1 + c_2 s_2$) (Berlinet and Thomas-Agnan 2011; Fasshauer and McCourt 2015). Denoting this polynomial space as $\mathcal{P}_1$, then $\mathcal{P}_1$ is the null space of the bending energy operator. Finally, the TPS aims to find $f \in H^2(\mathcal{D})$ minimizing the following functional:

$$\mathcal{L}(f) = \sum_{i=1}^{N} (y_i - f(\mathbf{s}_i))^2 + \alpha J(f), \tag{8}$$

where $\alpha > 0$ is a smoothing parameter and $J(f)$ was defined in Equation (7).

2.3.1 Thin plate splines as a Green's function solution

A Green's function is a fundamental solution to a linear differential operator. Essentially, it allows us to solve linear partial differential equations (PDEs) by converting them into integral equations (Hilbert 1985; Wahba 1990; Fasshauer and McCourt 2015). The formal definition is the following: given a linear differential operator $\mathcal{L}$ on the domain $\mathcal{D} \subset \mathbb{R}^d$, the Green's kernel G of $\mathcal{L}$ is defined as the solution of:

$$\mathcal{L}G(\mathbf{s},\mathbf{s}') = \delta(\mathbf{s}-\mathbf{s}'), \quad \mathbf{s}' \in \mathcal{D} \text{ fixed.} \tag{9}$$

Here $\delta(\mathbf{s}-\mathbf{s}')$ is the Dirac delta functional evaluated at $\mathbf{s}-\mathbf{s}'$. In this case, δ acts as a point evaluator for any $f \in L^2(\mathcal{D})$, i.e.,

$$\int_{\mathcal{D}} f(\mathbf{s}')\delta(\mathbf{s}-\mathbf{s}')d\mathbf{s}' = f(\mathbf{s}) \tag{10}$$

Duffy (2015); Evans (2022). Note that, although this property of the Dirac delta function is analogous to the reproducing

property of a reproducing kernel K (Fasshauer and McCourt 2015), δ is not the reproducing kernel of $L^2(\mathcal{D})$ since $\delta \notin L^2(\mathcal{D})$. For the biharmonic operator Δ^2, the Green's function $G(\mathbf{s}, \mathbf{s}')$ satisfies:

$$\Delta^2 G(\mathbf{s}, \mathbf{s}') = \delta(\mathbf{s} - \mathbf{s}'), \quad \mathbf{s}, \mathbf{s}' \in \mathbb{R}^d. \tag{11}$$

For $\mathbb{R}^2$, the biharmonic Green's function is given by Aronszajn and Smith (1957); Duffy (2015); Fasshauer and McCourt (2015):

$$G(\mathbf{s}, \mathbf{s}') = \frac{1}{8\pi} \left\|\mathbf{s} - \mathbf{s}'\right\|^2 \log \left\|\mathbf{s} - \mathbf{s}'\right\|. \tag{12}$$

Since the scale of the kernel is absorbed into α and the coefficients c_i, the TPS kernel $K(\mathbf{s}, \mathbf{s}') = \left\|\mathbf{s} - \mathbf{s}'\right\|^2 \log \left\|\mathbf{s} - \mathbf{s}'\right\|$ is the fundamental solution (Green's function) for the biharmonic operator on $\mathbb{R}^2$.

2.3.2 Hilbert–Schmidt operators and kernels

A linear operator $T : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ is a Hilbert-Schmidt operator if its Hilbert-Schmidt norm is finite:

$$\|T\|_{\text{HS}}^2 = \sum_{k=1}^{\infty} \|T e_k\|_{L^2}^2 < \infty, \tag{13}$$

for any orthonormal basis $\{e_k\}$ of $L^2(\mathcal{D})$ (Reed and Simon 1980; Engl and Ramlau 2015). If T is an integral operator with kernel $K(\mathbf{s}, \mathbf{s}')$:

$$(Tf)(\mathbf{s}) = \int_{\mathcal{D}} K(\mathbf{s}, \mathbf{s}') f(\mathbf{s}') d\mathbf{s}', \tag{14}$$

then T is a Hilbert-Schmidt operator if and only if the kernel $K(\mathbf{s}, \mathbf{s}')$ is square-integrable, i.e., $K \in L^2(\mathcal{D} \times \mathcal{D})$.

To recap, the main requirement for the KLE is a covariance function $C(\mathbf{s}, \mathbf{s}')$ that is symmetric and positive semi-definite. However, the TPS kernel $K(\mathbf{s}, \mathbf{s}') = \mathbf{d}^2 \log(\mathbf{d})$ is only conditionally positive definite (Fasshauer 2007; Wendland 2004). To utilize the spectral properties required for Bayesian inference, we must reframe the problem. Instead of working with the singular biharmonic operator directly, we define a regularized operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$ that yields a proper Hilbert-Schmidt kernel.

2.3.3 Connecting thin plate splines to Hilbert–Schmidt kernels

To obtain a valid Hilbert-Schmidt kernel associated with TPS, we introduce a regularized elliptic operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$, where $\alpha > 0$ is a regularization parameter and $\mathbf{I}$ denotes the identity operator on the Sobolev space $H^2(\mathcal{D})$. The operator L_α is elliptic, self-adjoint, and invertible. Its inverse, $L_\alpha^{-1} = (\mathbf{I} + \alpha\Delta^2)^{-1}$, defines a compact, self-adjoint, and strictly positive definite integral operator. In the context of continuous variational problems, Mercer's theorem (Mercer 1909; Steinwart and Scovel 2012) and Hilbert-Schmidt theory (Renardy and Rogers 2004) apply to compact, self-adjoint operators with positive definite kernels. As we already know, TPS is conditionally positive definite and therefore falls outside the direct scope of these results. To address this issue, we reformulate the problem by replacing the TPS kernel $K(\mathbf{s}, \mathbf{s}')$ with the kernel induced by the inverse of the regularized operator L_α. Let's consider the following continuous variational problem:

$$\min_{f \in H^2(\mathcal{D})} \left(\int_{\mathcal{D}} (f(\mathbf{s}) - g(\mathbf{s}))^2 \, d\mathbf{s} + \alpha J(f) \right), \tag{15}$$

where $g(\mathbf{s})$ is a target function, $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain, and $J(f) = \int_{\mathcal{D}} (\Delta^2 f)^2 \, d\mathbf{s}$ is the bending-energy functional. Applying the Euler-Lagrange equation to this continuous problem yields the necessary condition for a minimizer:

$$f(\mathbf{s}) - g(\mathbf{s}) + \alpha\Delta^2 f(\mathbf{s}) = 0. \tag{16}$$

Rearranging, we obtain the elliptic partial differential equation (PDE):

$$\begin{aligned} f(\mathbf{s}) + \alpha\Delta^2 f(\mathbf{s}) &= g(\mathbf{s}), \\ L_\alpha f &= g, \end{aligned} \tag{17}$$

which is a fourth-order elliptic PDE (Trudinger 1983; Courant and Hilbert 2024). Under appropriate boundary conditions—such as clamped plate conditions (fixed values and normal derivatives on $\partial\mathcal{D}$)—the biharmonic operator Δ^2 is an unbounded, self-adjoint operator on $L^2(\mathcal{D})$ with domain $H^4(\mathcal{D}) \cap H_0^2(\mathcal{D})$. The regularized operator L_α is then boundedly invertible, and its inverse L_α^{-1} is a compact operator. Most importantly, the integral kernel associated with L_α^{-1}, denoted $K_\alpha(\mathbf{s}, \mathbf{s}')$, is a well-defined Hilbert-Schmidt kernel satisfying:

$$(L_\alpha^{-1} g)(\mathbf{s}) = \int_{\mathcal{D}} K_\alpha(\mathbf{s}, \mathbf{s}') g(\mathbf{s}') \, d\mathbf{s}'. \tag{18}$$

Furthermore, by Mercer's theorem, the kernel K_α admits the spectral expansion:

$$K_\alpha(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^{\infty} \lambda_{k,\alpha} \, \phi_k(\mathbf{s}) \, \phi_k(\mathbf{s}'), \tag{19}$$

where $\{\lambda_{k,\alpha}\}_{k=1}^{\infty}$ are the eigenvalues and $\{\phi_k(\mathbf{s})\}_{k=1}^{\infty}$ are the corresponding orthonormal eigenfunctions of the compact

operator L_α^{-1}, satisfying:

$$\int_{\mathcal{D}} \phi_k(\mathbf{s})\phi_\ell(\mathbf{s})\, d\mathbf{s} = \delta_{k\ell}, \tag{20}$$

where $\delta_{k\ell}$ is the Kronecker delta, and with $\sum_{k=1}^{\infty} \lambda_{k,\alpha}^2 < \infty$ (Hilbert-Schmidt property). These eigenfunctions are solutions to:

$$L_\alpha^{-1}\phi_k(\mathbf{s}) = \lambda_{k,\alpha}\, \phi_k(\mathbf{s}), \tag{21}$$

or equivalently, applying L_α to both sides, we obtain the eigenvalue problem for the operator L_α:

$$(\mathbf{I} + \alpha\Delta^2)\phi_k(\mathbf{s}) = \mu_k(\alpha)\, \phi_k(\mathbf{s}), \tag{22}$$

where $\mu_k(\alpha) = 1/\lambda_{k,\alpha}$ denotes the eigenvalues of the operator L_α itself (not its inverse). Let $\{v_k\}_{k=1}^{\infty}$ denote the eigenvalues of the biharmonic operator Δ^2, ordered in ascending order, so that $\Delta^2\phi_k(\mathbf{s}) = v_k\phi_k(\mathbf{s})$. Substituting this into Equation (22), we obtain:

$$\phi_k(\mathbf{s}) + \alpha v_k\phi_k(\mathbf{s}) = \mu_k(\alpha)\, \phi_k(\mathbf{s}), \tag{23}$$

which simplifies to:

$$(1 + \alpha v_k)\phi_k(\mathbf{s}) = \mu_k(\alpha)\, \phi_k(\mathbf{s}). \tag{24}$$

This establishes the relationship between the eigenvalues:

$$\mu_k(\alpha) = 1+\alpha v_k, \qquad \text{and consequently} \qquad \lambda_{k,\alpha} = \frac{1}{1+\alpha v_k}. \tag{25}$$

The eigenvalue spectrum of Δ^2 has a special structure. The biharmonic operator Δ^2 has a null space $\mathcal{N}(\Delta^2)$ consisting of polynomial functions of degree at most $m-1$ (where $m = 2$ for standard TPS, yielding polynomials of degree up to 1). In two dimensions, this null space has dimension $d_{\text{null}} = \binom{d+m-1}{d}$ (e.g. $d_{\text{null}} = 3$ in two dimensions, corresponding to the constant, x and y terms, and $d_{\text{null}} = 4$ in three dimensions, adding the third linear coordinate). For these polynomial functions, we have $\Delta^2 p(\mathbf{s}) = 0$, which corresponds to $v_k = 0$ for $k \le d_{\text{null}}$. Thus, the eigenvalue structure can be characterized as follows:

- **Null space modes** ($k \le d_{\text{null}}$, $v_k = 0$): These eigenvalues correspond to the polynomial null space $\mathcal{N}(\Delta^2)$. For these modes,

$$\mu_k(\alpha) = 1 + \alpha \cdot 0 = 1, \qquad \lambda_{k,\alpha} = \frac{1}{1} = 1.$$

  This implies that L_α^{-1} maps these polynomials to themselves with unit eigenvalue, and they are *unpenalized* by the bending-energy functional. The eigenfunctions ϕ_k for $k \le d_{\text{null}}$ span the polynomial space $\mathcal{P}_{m-1}(\mathcal{D})$ and capture large-scale spatial trends with unit prior variance.
- **Penalized modes** ($k > d_{\text{null}}$, $v_k > 0$): These eigenvalues correspond to the non-polynomial bending modes of the plate, which are orthogonal to the null space in the L^2 inner product. For these modes,

$$\mu_k(\alpha) = 1 + \alpha v_k > 1, \qquad \lambda_{k,\alpha} = \frac{1}{1+\alpha v_k} < 1.$$

  The eigenvalues v_k increase with k, corresponding to increasingly oscillatory eigenfunctions with higher bending energy. The regularization parameter α controls the decay rate of $\lambda_{k,\alpha}$, where larger α causes faster decay, imposing stronger smoothing on high-frequency components.

In the classical TPS solution framework, the polynomial part is handled separately from the radial basis function part. The coefficients are typically constrained such that the radial basis contribution is orthogonal to the null space of polynomials, ensuring well-posedness and uniqueness of the "bending" part of the solution (Wahba 1975; Green 1987; Cavieres et al. 2023). The regularization parameter $\alpha > 0$ makes the entire problem well-posed without requiring explicit orthogonality constraints, as the operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$ is boundedly invertible. Therefore, the kernel of the regularized TPS, denoted $K_\alpha(\mathbf{s}, \mathbf{s}')$ and associated with the inverse operator $L_\alpha^{-1} = (\mathbf{I} + \alpha\Delta^2)^{-1}$, is a Hilbert-Schmidt, self-adjoint, positive definite kernel with spectral expansion:

$$K_\alpha(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^{\infty} \lambda_{k,\alpha}\, \phi_k(\mathbf{s})\, \phi_k(\mathbf{s}'), \tag{26}$$

where $\lambda_{k,\alpha} = (1 + \alpha v_k)^{-1}$ and $\{\phi_k(\mathbf{s})\}_{k=1}^{\infty}$ are the complete orthonormal system of eigenfunctions of Δ^2 (equivalently, of L_α and L_α^{-1}), including both the polynomial null space modes and the orthogonal complement. This spectral representation forms the theoretical foundation of the regTPS approach for practical Bayesian inference on spatial data described in Sect. 3.

3 Approximating a GRF by a regTPS: a Bayesian approach

In the previous section, we established that the regularized TPS covariance kernel admits the spectral expansion:

$$K_\alpha(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^{\infty} \lambda_{k,\alpha}\, \phi_k(\mathbf{s})\, \phi_k(\mathbf{s}'), \tag{27}$$

where $\phi_k(\mathbf{s})$ are the eigenfunctions of the regularized biharmonic operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$, satisfying the eigenvalue problem:

$$(\mathbf{I} + \alpha\Delta^2)\phi_k(\mathbf{s}) = \mu_k(\alpha)\, \phi_k(\mathbf{s}), \tag{28}$$

where $\mu_k(\alpha) = 1 + \alpha v_k$ are the eigenvalues of the operator L_α, and v_k are the eigenvalues of the biharmonic operator Δ^2. The corresponding covariance eigenvalues (KLE prior variances) are:

$$\lambda_{k,\alpha} = \frac{1}{\mu_k(\alpha)} = \frac{1}{1 + \alpha v_k}. \tag{29}$$

While this continuous formulation provides a theoretical foundation, the eigenfunctions ϕ_k are generally not available in closed form for arbitrary domains $\mathcal{D} \subset \mathbb{R}^d$ and boundary conditions. To obtain a computationally tractable framework, we employ the Ritz-Galerkin approximation by projecting the infinite-dimensional problem onto a finite-dimensional subspace spanned by TPS basis functions.

3.1 Finite-dimensional approximation via TPS basis functions

Let N denote the number of spatial locations $\{\mathbf{s}_i\}_{i=1}^N \subset \mathcal{D}$, and let K denote the number of TPS basis functions used in the approximation. We introduce a finite-dimensional approximation space $\mathcal{V}_K$ spanned by K TPS basis functions:

$$\mathcal{V}_K = \text{span}\{\tilde{\phi}_1, \tilde{\phi}_2, \ldots, \tilde{\phi}_K\}, \quad \text{where} \quad \tilde{\phi}_j(\mathbf{s}) = \phi_{\text{TPS}}(\|\mathbf{s} - \boldsymbol{\xi}_j\|), \tag{30}$$

and $\phi_{\text{TPS}}(d)$ is the TPS basis function, with knot locations $\{\boldsymbol{\xi}_j\}_{j=1}^K$ typically chosen as a subset of the spatial locations or placed on a regular grid. At this point, it is important to distinguish clearly between two types of functions:

- $\phi_k(\mathbf{s})$: the k-th eigenfunction of the continuous regularized operator L_α, belonging to the infinite-dimensional Sobolev space $H^2(\mathcal{D})$.
- $\tilde{\phi}_j(\mathbf{s})$: the j-th TPS basis function, which is *not* an eigenfunction but serves as a computational basis element for constructing the approximation space $\mathcal{V}_K$.

The connection between these functions is established through the Galerkin projection. Since TPS basis functions form a dense subset of $H^2(\mathcal{D})$ (Duchon 1977), any eigenfunction $\phi_k \in H^2(\mathcal{D})$ can be approximated within $\mathcal{V}_K$ by a linear combination:

$$\phi_k(\mathbf{s}) \approx \hat{\phi}_k(\mathbf{s}) = \sum_{j=1}^{K} a_{kj}\, \tilde{\phi}_j(\mathbf{s}), \tag{31}$$

where $\hat{\phi}_k \in \mathcal{V}_K$ denotes the discrete approximation to ϕ_k. The approximation error $\|\phi_k - \hat{\phi}_k\|_{H^2(\mathcal{D})} \to 0$ as $K \to \infty$, provided the knot points $\{\boldsymbol{\xi}_j\}$ become dense in $\mathcal{D}$ and the domain satisfies standard smoothness conditions (Brenner and Scott 2008).

3.2 Discrete Eigenvalue problem

The coefficients $\mathbf{a}_k = (a_{k1}, \ldots, a_{kK})^\top$ are obtained by diagonalizing the discrete bending-energy operator associated with Δ^2. This operator is represented in $\mathcal{V}_K$ by the penalty matrix $\mathbf{S} \in \mathbb{R}^{K\times K}$, whose entries are defined by the bending-energy bilinear form:

$$\mathbf{S}_{jl} = \int_{\mathcal{D}} (\Delta\tilde{\phi}_j(\mathbf{s}))(\Delta\tilde{\phi}_l(\mathbf{s}))\, d\mathbf{s}. \tag{32}$$

This matrix provides a finite-dimensional representation of the biharmonic operator Δ^2 restricted to $\mathcal{V}_K$ and defines the natural geometry for TPS smoothing, the reproducing kernel Hilbert space (RKHS) associated with the bending-energy seminorm (Wahba 1990; Brenner and Scott 2008). In practice, the penalty matrix $\mathbf{S}$ is constructed using the method described in Wood (2017), which provides numerically stable computation from the TPS basis representation without requiring explicit evaluation of the integral in Equation (32). Here, we perform the standard eigendecomposition of $\mathbf{S}$:

$$\mathbf{S} = \Psi_S \mathbf{V}_S \Psi_S^\top, \tag{33}$$

where $\mathbf{V}_S = \text{diag}(v_1, v_2, \ldots, v_K)$ contains the eigenvalues of $\mathbf{S}$ ordered in ascending order, and the columns of $\Psi_S \in \mathbb{R}^{K\times K}$ contain the corresponding eigenvectors, which are orthonormal: $\Psi_S^\top \Psi_S = \mathbf{I}_K$. The eigenvalues v_k of the discrete penalty matrix $\mathbf{S}$ provide finite-dimensional approximations to the eigenvalues of the continuous biharmonic operator Δ^2 restricted to $\mathcal{V}_K$, with the approximation improving as K increases and the basis becomes dense. This standard eigenvalue problem corresponds to an implicit diagonal approximation of the L^2 mass matrix, which is computationally efficient and widely used in TPS software implementations (Wood 2017). A fully consistent Galerkin discretization would require solving a generalized eigenvalue problem incorporating an explicit mass matrix to preserve

L^2 orthogonality of the discrete eigenfunctions. We provide a detailed theoretical justification for the standard eigenvalue approximation in Appendix A.

3.2.1 Eigenvalue ordering and the polynomial null space

A critical feature of the TPS penalty matrix is its null space, which corresponds to the space of polynomial functions of degree $m-1$ in dimension d (where $m=2$ for standard TPS in two-dimensional space, yielding polynomials up to degree 1). This null space has dimension $d_{\text{null}} = \binom{d+m-1}{d}$ (e.g., $d_{\text{null}} = 3$ for two-dimensional space, corresponding to the constant, x, and y polynomial terms). The penalty matrix $\mathbf{S}$ assigns zero bending energy to these polynomial functions, resulting in d_{null} eigenvalues that are numerically zero (within machine precision $\epsilon \approx 10^{-12}$). We order the eigenvalues in *ascending* order:

$$v_1 \leq v_2 \leq \cdots \leq v_K, \tag{34}$$

such that

$$v_k < \epsilon \quad \text{for } k \leq d_{\text{null}}, \quad \text{and} \quad v_k > 0 \quad \text{for } k > d_{\text{null}}. \tag{35}$$

The first d_{null} eigenvectors span the polynomial null space and represent unpenalized large-scale trends in the spatial random field. The remaining eigenvectors correspond to increasingly rough, high-frequency spatial features that are penalized by the bending-energy norm. Specifically, larger eigenvalues v_k (for $k > d_{\text{null}}$) correspond to eigenfunctions with greater curvature and higher oscillation frequency. The columns of Ψ_S define coefficient vectors for basis functions in $\mathcal{V}_K$ that diagonalize the discrete bending-energy operator. Specifically, the k-th column $\boldsymbol{\psi}_{S,k} \in \mathbb{R}^K$ yields the k-th discrete Ritz eigenfunction:

$$\hat{\phi}_k(\mathbf{s}) = \sum_{j=1}^{K} \psi_{S,jk}\, \tilde{\phi}_j(\mathbf{s}) = \tilde{\boldsymbol{\phi}}(\mathbf{s})^\top \boldsymbol{\psi}_{S,k}, \tag{36}$$

where $\tilde{\boldsymbol{\phi}}(\mathbf{s}) = (\tilde{\phi}_1(\mathbf{s}), \ldots, \tilde{\phi}_K(\mathbf{s}))^\top$ is the vector of TPS basis functions evaluated at the spatial location $\mathbf{s}$.

3.3 Discrete KLE with adaptive truncation

Using the discrete eigenfunctions $\hat{\phi}_k$ and applying the regularization parameter $\alpha > 0$, we obtain the discrete analogue of Equation (27):

$$K_\alpha(\mathbf{s}, \mathbf{s}') \approx \sum_{k=1}^{M} \lambda_{k,\alpha}\, \hat{\phi}_k(\mathbf{s})\, \hat{\phi}_k(\mathbf{s}'), \tag{37}$$

where the discrete KLE eigenvalues (prior variances) are given by:

$$\lambda_{k,\alpha} = \frac{1}{1+\alpha v_k}, \tag{38}$$

which approximates the continuous relationship in Equation (29), where v_k are the eigenvalues of the discrete penalty matrix $\mathbf{S}$ from Equation (33). For components in the null space ($k \leq d_{\text{null}}$), we have $v_k \approx 0$, which yields $\lambda_{k,\alpha} \approx 1$. These components are therefore unpenalized and capture the low-frequency polynomial trends in the spatial field with unit prior variance. For $k > d_{\text{null}}$, we have $v_k > 0$, and $\lambda_{k,\alpha}$ decreases monotonically with increasing k. Larger eigenvalues v_k correspond to rougher (higher-frequency) basis functions, which are increasingly shrunk by the regularization parameter α. The decay rate is controlled by α: larger values of α impose stronger regularization, causing $\lambda_{k,\alpha}$ to decay more rapidly. In practice, we truncate the expansion at $M \leq K$ components rather than using all available discretized basis functions K. The truncation level M (the number of retained basis functions) is selected adaptively to retain a specified proportion (typically 95% – 99%) of the total spectral variance:

$$M = \min\left\{ m : \frac{\sum_{k=1}^{m} \lambda_{k,\alpha}}{\sum_{k=1}^{K} \lambda_{k,\alpha}} \geq \gamma \right\}, \tag{39}$$

where $\gamma \in (0,1)$ is the variance retention threshold (e.g., $\gamma = 0.95$). This adaptive truncation exploits the rapid decay of $\lambda_{k,\alpha}$ with increasing k, retaining only those components that contribute meaningfully to the covariance structure while discarding negligible high-frequency modes.

3.4 The regTPS construction on *d*-dimensional domains

The construction above is stated for $\mathcal{D} \subset \mathbb{R}^2$, the setting of the numerical experiments in Sect. 4. Because it is formulated at the level of the operator rather than of a particular covariance family, it extends to $\mathcal{D} \subset \mathbb{R}^d$ without structural modification. We record here what changes, and what does not, when the Laplace operator acts on a three-dimensional domain.

Admissibility. Replacing $\mathcal{L}_\alpha = I + \alpha\Delta^2$ by the regularized polyharmonic operator $\mathcal{L}_\alpha^{(m)} = I + \alpha(-\Delta)^m$, the covariance eigenvalues retain the form $\lambda_{k,\alpha} = (1+\alpha v_k)^{-1}$, where the v_k are now the eigenvalues of $(-\Delta)^m$ on $\mathcal{D}$. By Weyl's law on a bounded domain, $v_k \asymp C_d\, k^{2m/d}$, and hence

$$\lambda_{k,\alpha} \asymp (\alpha C_d)^{-1}\, k^{-2m/d}, \qquad k \to \infty. \tag{40}$$

The operator $\left(\mathcal{L}_\alpha^{(m)}\right)^{-1}$ is trace class, and therefore admissible as a covariance operator within the KLE, precisely when $\sum_k \lambda_{k,\alpha} < \infty$, that is, when

$$2m > d. \tag{41}$$

Condition presented in Equation (41) is the same requirement $m > d/2$ under which the Duchon semi-norm defines a reproducing kernel Hilbert space of continuous functions (Duchon 1977), and it coincides with the condition $\alpha_{\mathrm{SPDE}} > d/2$ required by the SPDE representation (Appendix B) under the identification $\alpha_{\mathrm{SPDE}} = m$ established below. The bending-energy operator used throughout this work ($m = 2$) is thus admissible for $d \leq 3$ and requires no modification in three dimensions; for $d \geq 4$ the penalty order must be raised to $m > d/2$.

If $d = 3$ in regTPS then three elements take different values:

(i) The null-space dimension $d_{\mathrm{null}} = \binom{d+m-1}{d}$ increases from 3 to 4, the unpenalized modes being the constant and the three linear coordinates.
(ii) The Duchon radial basis changes from $\phi(r) = r^2 \log r$ to $\phi(r) = r$, since the exponent $2m - d = 1$ is odd; the general form is r^{2m-d} for $2m - d$ odd and $r^{2m-d} \log r$ for $2m - d$ even.
(iii) The KLE spectrum decays more slowly, from k^{-2} in the plane to $k^{-4/3}$ in a volume, by Equation (40).

The eigendecomposition presented in Equation (33), the truncation rule (Equation (39)) and the Bayesian formulation of Sect. 3.5 are otherwise unchanged.

Cost of the slower decay. The truncation level required to retain a variance fraction $\gamma = 1 - \varepsilon$ follows from Equation (40). The neglected tail satisfies $\sum_{k>M} \lambda_{k,\alpha} \asymp M^{-(2m-d)/d}$, so that

$$M(\varepsilon) \asymp \varepsilon^{-d/(2m-d)}. \tag{42}$$

For $m = 2$ this gives $M \asymp \varepsilon^{-1}$ in $d = 2$ but $M \asymp \varepsilon^{-3}$ in $d = 3$. Tightening the retained variance from 95% to 99% therefore costs a factor of roughly five in the plane and roughly 125 in a volume. The dimension reduction reported in Sect. 4 should accordingly not be expected to transfer unchanged to volumetric domains.

Implied smoothness. Relation Equation (40) also identifies the regularity of the induced prior. A covariance spectrum decaying as $k^{-2m/d}$ corresponds to sample paths in $H^s(\mathcal{D})$ for every $s < m - d/2$, that is, to a Matérn field of smoothness

$$\nu = m - \tfrac{d}{2}. \tag{43}$$

The SPDE representation of Appendix B yields $\nu = \alpha_{\mathrm{SPDE}} - d/2$, so the two constructions describe the same regularity class under the identification $\alpha_{\mathrm{SPDE}} = m$: with $m = 2$ we obtain $\nu = 1$ in $d = 2$, matching the specification $\alpha_{\mathrm{SPDE}} = 2$ used throughout, and $\nu = 1/2$ in $d = 3$, again matching $\alpha_{\mathrm{SPDE}} = 2$. The two approximations, therefore, remain directly comparable in three dimensions, but both describe a rougher field than in the plane. A practitioner requiring in $d = 3$ the regularity that $m = 2$ supplies in $d = 2$ must raise the penalty order to $m = 3$.

3.5 Bayesian interpretation

Having established the discrete KLE representation, we now formulate the Bayesian inference framework. Consider the penalized spatial least-squares problem:

$$\mathcal{L}(f) = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{N} (y_i - f(\mathbf{s}_i))^2 + \alpha J(f), \tag{44}$$

where $y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(f(\mathbf{s}_i), \sigma_\varepsilon^2)$, $\sigma_\varepsilon > 0$ denotes the standard deviation of the observation noise, N is the number of observations at spatial locations $\mathbf{s}_i \in \mathbb{R}^d$, and $J(f) = \int_{\mathcal{D}} (\Delta^2 f)^2 \, d\mathbf{s}$ is the bending-energy functional. Let $f(\mathbf{s}) = \sum_{j=1}^{K} c_j \tilde{\phi}_j(\mathbf{s})$ with coefficient vector $\mathbf{c} = (c_1, \ldots, c_K)^\top$ and design matrix $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{N \times K}$ with entries $\tilde{\Phi}_{ij} = \tilde{\phi}_j(\mathbf{s}_i)$. Then $\mathbf{y} = \tilde{\mathbf{\Phi}}\mathbf{c} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$, and the Gaussian likelihood is:

$$p(\mathbf{y} \mid \mathbf{c}, \sigma_\varepsilon) = (2\pi\sigma_\varepsilon^2)^{-N/2} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \|\mathbf{y} - \tilde{\mathbf{\Phi}}\mathbf{c}\|^2 \right). \tag{45}$$

In the Bayesian interpretation of TPS, the roughness penalty $J(f) = \mathbf{c}^\top \mathbf{S} \mathbf{c}$ corresponds to a Gaussian prior on the coefficients $\mathbf{c}$ defined through the precision matrix:

$$\mathbf{P}_\alpha = \mathbf{I} + \alpha \mathbf{S}, \tag{46}$$

where $\mathbf{I}$ is the $K \times K$ identity matrix. Thus, the prior on $\mathbf{c}$ is:

$$\pi(\mathbf{c} \mid \alpha) \propto \exp\left(-\frac{1}{2} \mathbf{c}^\top \mathbf{P}_\alpha \mathbf{c} \right), \tag{47}$$

which implies $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_\alpha)$, with covariance matrix $\mathbf{K}_\alpha = \mathbf{P}_\alpha^{-1} = (\mathbf{I} + \alpha \mathbf{S})^{-1}$. Using the eigendecomposition $\mathbf{S} = \Psi_S \mathbf{V}_S \Psi_S^\top$ from Equation (33) and defining the transformed coefficients:

$$\mathbf{z} = \Psi_S^\top \mathbf{c}, \tag{48}$$

the prior transforms to:

$$\pi(\mathbf{z} \mid \alpha) \propto \exp\left(-\frac{1}{2}\mathbf{z}^\top(\mathbf{I} + \alpha\mathbf{V}_S)\mathbf{z}\right), \tag{49}$$

revealing that the components are independent with diagonal covariance:

$$z_k \sim \mathcal{N}\left(0,\ \lambda_{k,\alpha}\right), \qquad \lambda_{k,\alpha} = \frac{1}{1 + \alpha v_k}, \tag{50}$$

where v_k are the eigenvalues of the penalty matrix $\mathbf{S}$ from Equation (33). This demonstrates that the eigendecomposition of $\mathbf{S}$ yields a diagonal prior covariance in the transformed coordinate system $\mathbf{z}$, with the KLE eigenvalues $\lambda_{k,\alpha}$ directly governing the prior variance of each component. For null space components ($k \leq d_{\text{null}}$), we have $v_k \approx 0$, yielding $\lambda_{k,\alpha} \approx 1$—these components have unit prior variance and are unpenalized. For penalized components ($k > d_{\text{null}}$), increasing v_k leads to decreasing $\lambda_{k,\alpha}$, imposing stronger shrinkage on rougher basis functions. Moreover, the spatial field at any location $\mathbf{s}$ can be expressed in terms of the transformed coefficients as:

$$f(\mathbf{s}) \approx \sum_{k=1}^{M} z_k \hat{\phi}_k(\mathbf{s}) = \mathbf{\Phi}(\mathbf{s})^\top \mathbf{z}, \tag{51}$$

where $\mathbf{\Phi}(\mathbf{s}) = (\hat{\phi}_1(\mathbf{s}), \ldots, \hat{\phi}_M(\mathbf{s}))^\top$ is the vector of discrete eigenfunctions, and we truncate at M components as specified in Equation (39). At the observation locations, this gives the linear predictor $\boldsymbol{\mu} = \mathbf{\Phi}_{\text{obs}}\mathbf{z}$, where $\mathbf{\Phi}_{\text{obs}} \in \mathbb{R}^{N \times M}$ has entries $(\mathbf{\Phi}_{\text{obs}})_{ik} = \hat{\phi}_k(\mathbf{s}_i)$. This Bayesian formulation provides the foundation for posterior inference, where we estimate the transformed coefficients $\mathbf{z}$, the regularization parameter α, and the noise level σ_ε from the observed data $\mathbf{y}$. The prior structure in Equation (46) allows efficient computation while maintaining the interpretability of the TPS framework, with automatic regularization through the KLE eigenvalue decay.

3.5.1 Non-centered parameterization

For computational efficiency and improved posterior geometry in MCMC sampling, we employ a non-centered parameterization (Papaspiliopoulos et al. 2007; Betancourt and Girolami 2015). We introduce non-centered parameters $\tilde{z}_k \sim \mathcal{N}(0, 1)$ and transform them according to:

$$z_k = \sqrt{\lambda_{k,\alpha}}\,\tilde{z}_k = \begin{cases} \tilde{z}_k, & k \leq d_{\text{null}} \\ \frac{\tilde{z}_k}{\sqrt{1+\alpha v_k}}, & k > d_{\text{null}} \end{cases} \tag{52}$$

For null space components, the scaling factor is unity, ensuring these components remain unpenalized. For penalized components, the scaling factor $1/\sqrt{1 + \alpha v_k}$ implements the shrinkage induced by regularization. This parameterization decorrelates the latent coefficients $\{\tilde{z}_k\}$ from the hyperparameter α, substantially improving MCMC mixing and convergence, particularly when α is also being estimated from the data.

3.5.2 Computational implementation

In practice, we work with truncated representations to improve computational efficiency. We truncate to M components as specified in Equation (39), so that $\mathbf{z} = (z_1, \ldots, z_M)^\top$ and $\Psi_S \in \mathbb{R}^{K \times M}$ contains only the first M columns of the full eigenvector matrix. We pre-compute the KLE design matrix at observation locations:

$$\mathbf{\Phi}_{\text{obs}} = \tilde{\mathbf{\Phi}}\Psi_S \in \mathbb{R}^{N \times M}, \tag{53}$$

where $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{N \times K}$ contains the TPS basis functions $\tilde{\phi}_j(\mathbf{s}_i)$ evaluated at observation locations, and the entries of $\mathbf{\Phi}_{\text{obs}}$ are $(\mathbf{\Phi}_{\text{obs}})_{ik} = \hat{\phi}_k(\mathbf{s}_i)$, representing the evaluation of the k-th discrete eigenfunction at location $\mathbf{s}_i$. The likelihood in terms of the non-centered parameters $\tilde{\mathbf{z}}$ is then:

$$p(\mathbf{y} \mid \tilde{\mathbf{z}}, \alpha, \sigma_\varepsilon) = (2\pi\sigma_\varepsilon^2)^{-N/2} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\|\mathbf{y} - \mathbf{\Phi}_{\text{obs}}\mathbf{z}(\tilde{\mathbf{z}}, \alpha)\|^2\right), \tag{54}$$

where $\mathbf{z}(\tilde{\mathbf{z}}, \alpha)$ denotes the transformation in Equation (52).

3.5.3 Prior specification and posterior inference

We complete the Bayesian model by assigning hyperpriors. For the observation noise standard deviation σ_ε, we use a penalized complexity (PC) prior (Simpson et al. 2017):

$$\pi(\sigma_\varepsilon) = \lambda_\sigma \exp(-\lambda_\sigma \sigma_\varepsilon), \quad \lambda_\sigma = -\frac{\log(\alpha_0)}{\sigma_0}, \tag{55}$$

where σ_0 and α_0 are chosen such that $P(\sigma_\varepsilon > \sigma_0) = \alpha_0$. We set $\sigma_0 = 0.5$ and $\alpha_0 = 0.05$, encoding weak prior information that the noise standard deviation is unlikely to exceed 0.5 with high probability. For the regularization parameter α, we use a lognormal prior:

$$\log \alpha \sim \mathcal{N}(0, \sigma_\alpha^2), \tag{56}$$

with $\sigma_\alpha = 1.0$ to allow flexibility while maintaining weak informativeness. This prior places most mass on moderate smoothing levels while allowing the data to determine the appropriate degree of regularization. The non-centered parameters have standard normal priors: $\tilde{z}_k \sim \mathcal{N}(0, 1)$ for $k = 1, \ldots, M$. Thus, combining the likelihood in Equation (50) with the priors, the joint posterior distribution is:

$$p(\tilde{\mathbf{z}}, \alpha, \sigma_\varepsilon \mid \mathbf{y}) \propto p(\mathbf{y} \mid \tilde{\mathbf{z}}, \alpha, \sigma_\varepsilon) \prod_{k=1}^{M} p(\tilde{z}_k)\, p(\alpha)\, p(\sigma_\varepsilon), \tag{57}$$

where the likelihood depends on $\tilde{\mathbf{z}}$ through the transformation in Equation (52).

To characterize the uncertainty in all model parameters, including α and σ_ε, we employ the Hamiltonian Monte Carlo (HMC) algorithm implemented in Stan (Carpenter et al. 2017; Team 2018), via Template Model Builder (TMB) (Kristensen et al. 2015) and `tmbstan` (Monnahan and Kristensen 2018) packages of the statistical software R. The pre-computation of the fixed eigendecomposition in Equation (33) allows for fast, linear-time calculation of the prior density and likelihood gradient in each step of the sampler, making it computationally feasible to obtain a robust representation of the full posterior distribution.

3.5.4 The SPDE method

For comparison purposes, we will mainly to consider the popular SPDE method for Bayesian spatial modeling (Lindgren et al. 2011). Briefly, instead of working directly with dense covariance matrices, the GRF is represented as solution of a stochastic partial differential equation (SPDE). By the Finite Element Method (FEM), the spatial domain is discretized, and one obtains a sparse precision matrix that approximates the covariance structure of the GRF. In simple terms, for large-scale spatial modelling purposes, this method allows for efficient computation while retaining theoretical guaranties from the Matérn class of covariance families (Whittle 1954, 1963; Lindgren et al. 2011). For a summarized explanation of this method and the prior distributions used for the hyperparameters participating in the MCMC sampling, see Appendix B.

4 Numerical analysis

4.1 Simulation 1

As previously mentioned, we will refer to our proposed approach (the regularized TPS kernel) to approximate a GRF using the KLE method as regTPS. We will assess its statistical and computational performance through a series of analyzes and compare the results against the widely used SPDE method to approximate a GRF (Lindgren et al. 2010). For this purpose, we adopt the spatial statistical model specified in Equation (1) without the mean component (μ); that is, no predictors were considered. The models are evaluated under four scenarios with increasing numbers of spatial locations: 50, 100, 150, and 200 points, referred to as Sce. 1, Sce. 2, Sce. 3, and Sce. 4, respectively. For each scenario, a single realization of the true GRF is simulated on the nodes of a common triangular mesh. Specifically, a covariance matrix is constructed using a Matérn covariance function evaluated at the mesh nodes, and a multivariate normal sample is drawn to obtain the spatial field. The true GRF is then projected from the mesh nodes to the spatial locations and to a regular grid using projection matrices derived from the SPDE discretization. Observations are generated by adding independent Gaussian noise to the projected field at the observation locations, that is; $y_i = u(\mathbf{s}_i) + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon)$. Because the true field is projected onto the finite element basis, the target of Simulation 1 lies by construction within the function space of the SPDE predictor, while regTPS must approximate it from a distinct spline eigenbasis. Simulation 1 is therefore a conservative point of reference for regTPS, and Simulation 2 removes this asymmetry (Sect. 4.2). Thus, for each scenario, we have the following (Table 1):

Finally, the mathematical formulation of the models can be found in Appendix C and the R code to simulate the data in Appendix D.

4.1.1 Simulation 1 results

4.1.2 regTPS evaluations

Figure 1 shows the prior and posterior distribution of α (penalty parameter). In this case, for the simulated data, the posterior distribution in all scenarios (Sce. 1 to Sce. 4) are narrower than before, meaning that the data provided much more information than previously, allowing us to compute the value of α accordingly, since it could normally be a problematic parameter in the sampling estimation process.

To assess whether the posterior distribution is coherent with the prior structure and to diagnose potential identifiability issues, we examine the relationship between prior standard deviations and posterior means of the basis coefficients. For each retained component $k = 1, \ldots, M$, we compute:

- **Prior standard deviation**: $\mathrm{SD}_{\text{prior}}(z_k) = \sqrt{\lambda_{k,\alpha}}$, where $\lambda_{k,\alpha} = (1 + \alpha v_k)^{-1}$.
- **Posterior mean (absolute value)**: $|\bar{z}_k| = |\mathbb{E}[z_k \mid \mathbf{y}]|$, estimated from MCMC samples.

Figure 2 plots $|\bar{z}_k|$ against $\mathrm{SD}_{\text{prior}}(z_k)$ for all scenarios in a $\log_{10}$ scale, where the dashed line represents the theoretical prior relationship $\mathbb{E}[|z_k|] = \mathrm{SD}_{\text{prior}}(z_k)/\sqrt{\pi/2}$, and the shaded region indicates a ± 2SD credible band. Points falling along the theoretical line suggest that the posterior is dominated by the prior (weak data information), while

Table 1 Degrees of freedom (DoF) comparison between SPDE and regTPS approaches

Scenario	Spatial locations	SPDE			regTPS			Ratio
		n_{mesh}	Hyper.	Total	M	Hyper.	Total	
Sce. 1	50	101	3	104	45	2	47	≈0.45
Sce. 2	100	151	3	154	90	2	92	≈0.60
Sce. 3	150	204	3	207	131	2	133	≈0.64
Sce. 4	200	244	3	247	177	2	179	≈0.72

Hyper. = Hyperparameters; Ratio = M/n_{mesh} (spatial DoF ratio)
SPDE hyperparameters: $\sigma_\varepsilon, \sigma_u, \rho$; regTPS: $\sigma_\varepsilon, \alpha$

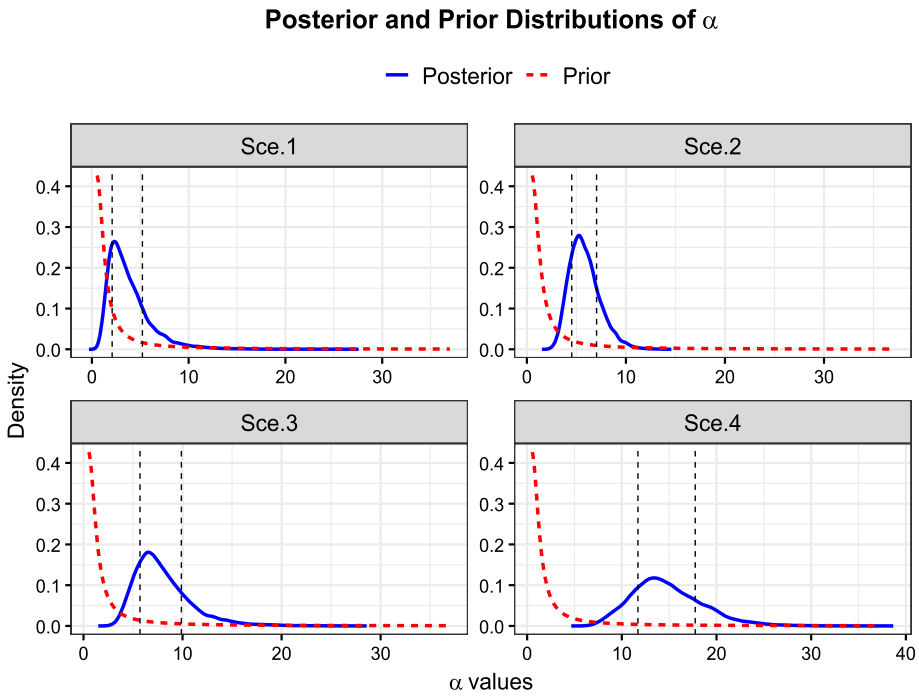


Fig. 1 Posterior (blue color) and prior distributions (red color) for the penalty parameter α. The vertical dashed lines indicates the quantiles (0.2 and 0.8) associated with the posterior distributions

deviations indicate data-driven learning. The clear separation between null space modes (unit prior SD) and penalized modes (decaying prior SD) confirms that the eigenvalue spectrum correctly encodes the smoothness hierarchy of the spatial field. It verifies that the prior standard deviations computed from the discrete eigenvalue problem correctly approximate the continuous theory and that the MCMC sampler successfully explores the posterior without numerical artifacts.

Since the truncation point should depend on the required level of accuracy in reconstructing the covariance function, Fig. 3 shows the cumulative variance explained by the components of the regTPS model. As shown, 99% of the total variance is explained using fewer components than the number of basis functions, confirming that the model effectively reduces the dimensionality of the spatial data (with the curves rapidly approaching 99% of explained variance). Selecting a lower threshold of explained variance would further improve computational efficiency; however, in this case, the number of estimated TPS basis functions does not approximate to the number of mesh nodes derived from the discretized spatial domain in the SPDE approach for a fair comparison. Figure 4 shows the eigenvalue spectrum of the **S** penalty matrix. The eigenvalues increase rapidly after the null-space components, reflecting the growing roughness associated with higher-order basis functions. This behavior is expected, as the penalty operator assigns small eigenvalues to smooth, low-frequency modes and increasingly large eigenvalues to highly oscillatory components. While the penalty eigenvalues grow with increasing roughness, the corresponding KLE eigenvalues decrease monotonically, leading to fast saturation of the cumulative variance explained (Fig. 3). These results collectively demonstrate that the regTPS model concentrates most of the spatial variability in a small number of smooth modes.

To assess the sensitivity of the posterior distribution of the penalty parameter α to the choice of prior scale, and to determine whether this choice affects the reconstruction quality of the spatial field, we examine how both the posterior distribution of α and the resulting reconstruction error vary with σ_α. Building on the baseline analysis, which uses

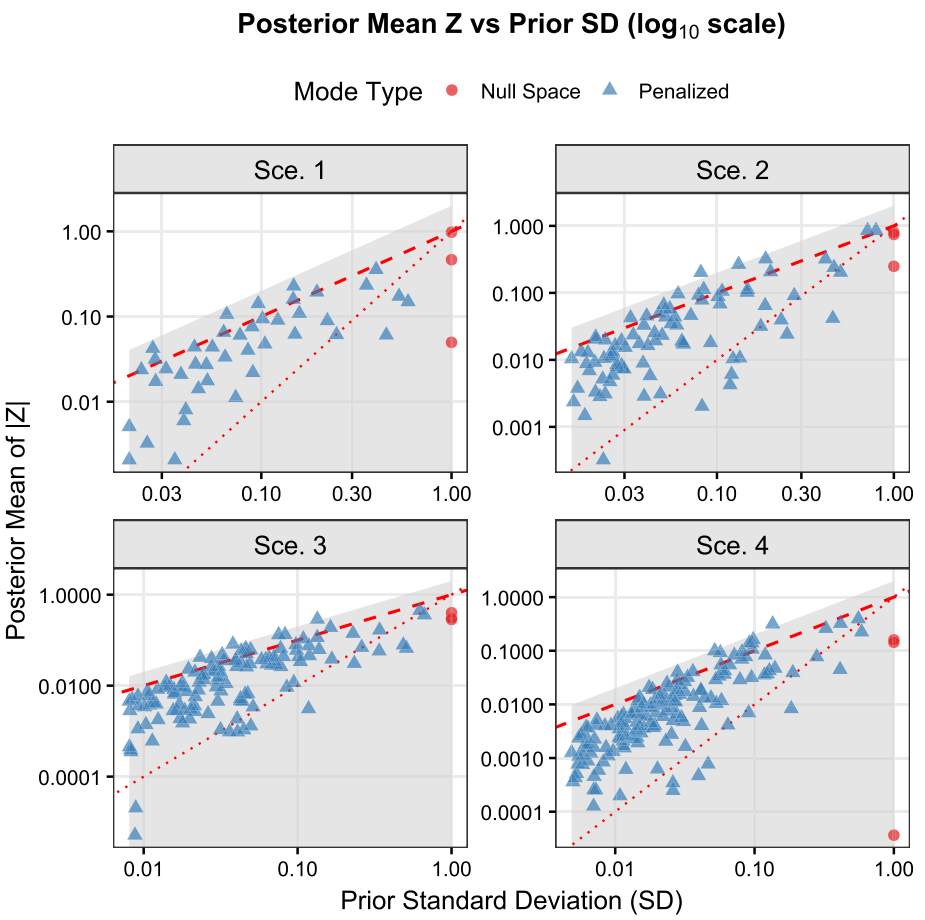


Fig. 2 Posterior means $|\mathbb{E}[z_k \mid \mathbf{y}]|$ versus prior standard deviations $\sqrt{\lambda_{k,\alpha}}$ in a $\log_{10}$ scale for all scenarios. The red circles indicates the null space modes ($k \leq 3$) with unit prior variance and blue triangles indicates the penalized modes with $\lambda_{k,\alpha} = (1 + \alpha v_k)^{-1}$. The dashed line is the theoretical prior expectation $\mathbb{E}[|z_k|] = \sqrt{2\lambda_{k,\alpha}/\pi}$ while the gray region is ± 2SD band

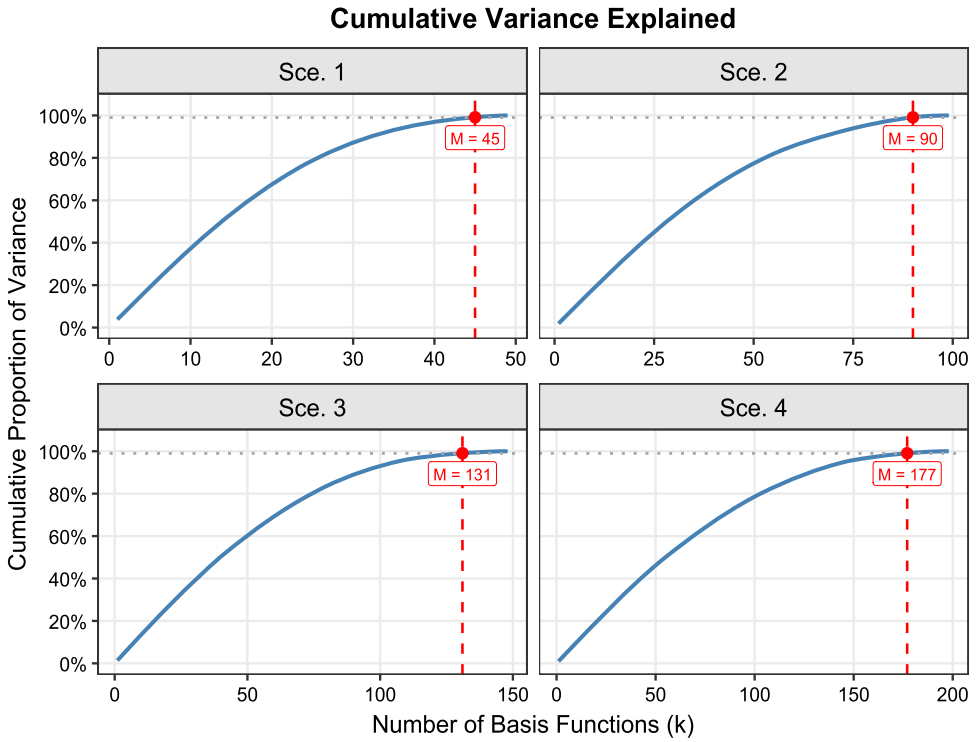


Fig. 3 Cumulative variance explained by the truncated KLE basis functions across scenarios. The dashed horizontal line indicates the 99% variance retention threshold used to select the truncation level M

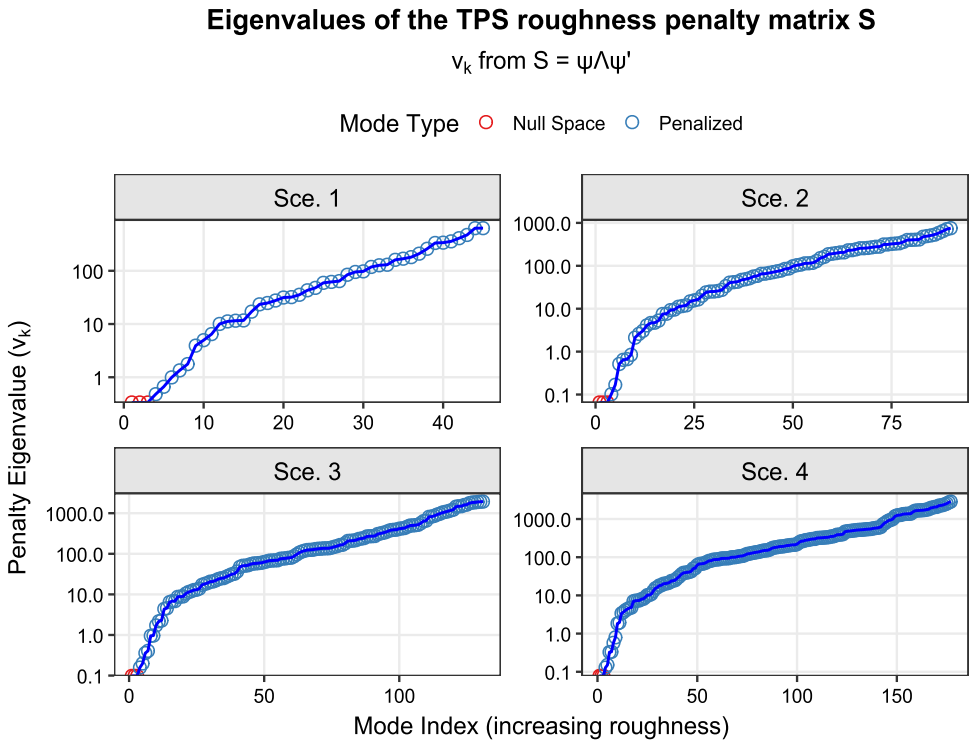


Fig. 4 The penalty eigenvalues v_k increase rapidly after the null-space components, reflecting increasing roughness of higher-order TPS basis functions. Small values of v_k correspond to smooth, low-frequency modes, whereas large values penalize highly oscillatory components

Table 2 Sensitivity analysis of the posterior distribution of the penalty parameter α and reconstruction error for different prior scales σ_α. The base case ($\sigma_\alpha = 1.0$) is highlighted in bold

σ_α	α_{mean}	α_{sd}	RMSE
1	**7.975**	**2.858**	**0.286**
2	9.598	4.009	0.284
4	10.974	5.223	0.283
6	10.939	4.934	0.284
8	10.984	4.987	0.283

$\sigma_\alpha = 1.0$ (Appendix C), we perform this sensitivity analysis for the intermediate simulation setting, Scenario 3 (Sce. 3), described in Table 1, while varying the prior scale σ_α. Table 2 shows that the posterior mean of α remains relatively stable across the different prior scales considered. Specifically, the posterior mean varies only slightly, from 7.975 (base case) to 10.984, suggesting that posterior inference for α is not highly sensitive to the choice of σ_α in Sce. 3. The posterior standard deviation also remains relatively stable, ranging from 2.858 (base case) to 5.223. The reconstruction error exhibits an even higher degree of stability, with RMSE values (calculated as ($\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$)) remains between 0.283 and 0.286 in all prior scales. These results suggest that varying the prior scale σ_α over the range considered has little practical effect on either posterior inference for α or the reconstruction quality of the spatial field.

Another potential source of approximation error arises from truncating the KLE basis. To assess the effect of the truncation level, we varied the variance-explained threshold from 0.85 to 0.99, while keeping the maximum number of available basis functions fixed. For each threshold, we computed a truncation-only benchmark using a least-squares interpolation (LSI) of the known true spatial field onto the retained KLE basis. This provides the best reconstruction achievable using only the retained M basis functions and therefore isolates the approximation error due solely to basis truncation, without contributions from observation noise or statistical estimation uncertainty. As in the previous analysis, Sce. 3 was considered and a fixed $\sigma_\alpha = 1.0$ (base case). Therefore, we compare this benchmark with the posterior RMSE obtained from the regTPS model (Table 3).

As presented in Table 3, increasing the variance-explained threshold from 0.85 to 0.99 augment the number of retained basis functions (M) from 83 to 131 as expected. While the (RMSE_{LSI}) decreases from 0.093 to 0.070, the corresponding error for the regTPS model ($\text{RMSE}_{\text{regTPS}}$) exhibited only a marginal decrease, shifting from 0.288 to 0.287 (reaching a minimum of 0.287 at $\gamma = 0.95$ and $\gamma = 0.90$). These findings indicate that incorporating additional basis functions yields only a minimal reduction in truncation error, exercising negligible influence on the overall posterior reconstruction accuracy of the regTPS model.

Figure 5 shows the sensitivity of the prediction error to different dimensions of M, controlled by varying the cumulative variance-explained threshold from 0.80 to 0.99 (in increments of 0.02 to 0.03). As expected, higher thresholds

Table 3 Reconstruction error (RMSE) as a function of the KLE variance-explained threshold for the least-squares interpolation (LSI) and the regTPS model

Variance explained (threshold)	M	RMSE_{LSI}	$\text{RMSE}_{\text{regTPS}}$
0.85	83	0.093	0.288
0.90	93	0.078	0.287
0.95	107	0.073	0.287
0.99	131	0.070	0.287

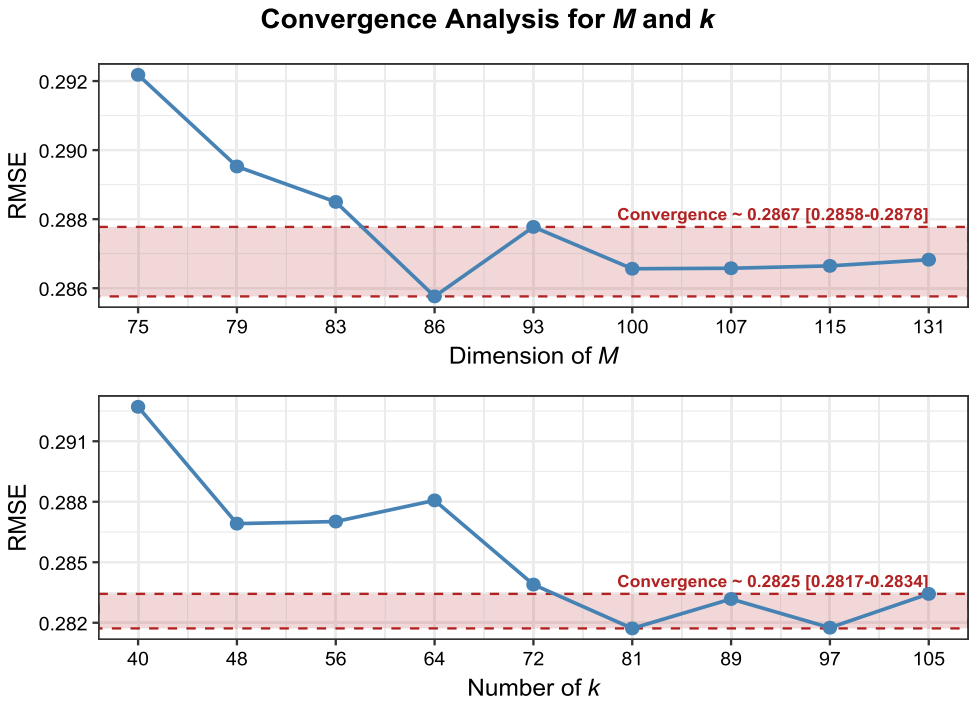


Fig. 5 Convergence analysis for RMSE values with respect to model dimension M (upper) and basis dimension k (lower) for the regTPS method in Simulation 1

require a larger number of modes, expanding M from 75 to 131. The values of RMSE shows rapid initial stabilization, decreasing from 0.2922 at $M = 75$ to around 0.286 – 0.287 for $M \geq 100$. Additionally, Fig. 5 shows the prediction performance across increasing spline basis dimensions, k from 40 to 105, alongside their corresponding truncated dimensions M (ranging from 39 to 99). The mean RMSE decreases from 0.293 at $k = 40$ and reaches a minimum of 0.2817 at $k = 81$ (with $M = 78$). In the other side, for $k = 81$, the error is stabilized between 0.282 and 0.284, demonstrating that increasing the basis resolution beyond this points leads to diminishing improvements in accuracy, while incurring substantially higher computational costs.

According to the Table 3 and Fig. 5, expanding the basis set provides negligible improvements in posterior reconstruction accuracy, confirming that truncation accounts for well under a tenth of the total squared error, while the majority is attributable to estimation uncertainty and observation noise.

4.1.3 SPDE and regTPS comparisons

Since covariance matrices obtained from the SPDE and regTPS methods are defined in different spaces, they cannot be compared directly. Thus, to enable a fair comparison, we project them into a common space. Specifically, the covariance matrix $\mathbf{K}_\alpha$ from regTPS is mapped from its low-dimensional basis function space to the high-dimensional mesh node space (the mesh built for the SPDE method). For example, according to the variance explained for Sce. 1, only 45 basis functions are needs to build the approximated GRF; hence we evaluate these basis functions at the number of mesh node locations coming from the discretized domain using the SPDE method (101 mesh nodes). Essentially, we are building the spatial field at the mesh nodes from the basis functions coefficients. After this, the difference between the true covariance matrix and the obtained from the SPDE and regTPS methods were plotted (Fig. 6). Although there is a small positive difference between the values of the true and regTPS covariance matrices, as we expect, the patterns of the points are very similar for all the scenarios. On the contrary, the covariance matrix estimated for the

Table 4 Comparison of SPDE and regTPS models with the true GRF using a Matérn covariance function across scenarios in Simulation 1

Scenario	Method	RMSE	R^2	MAE
Sce. 1	SPDE	0.421	0.828	0.330
	regTPS	0.399	0.860	0.307
Sce. 2	SPDE	0.315	0.875	0.235
	regTPS	0.351	0.866	0.268
Sce. 3	SPDE	0.269	0.940	0.207
	regTPS	0.287	0.932	0.226
Sce. 4	SPDE	0.230	0.914	0.177
	regTPS	0.256	0.900	0.199

Table 5 Comparison of SPDE and regTPS models with the true GRF using an exponential covariance function across scenarios in Simulation 1

Scenario	Method	RMSE	R^2	MAE
Sce. 1	SPDE	0.453	0.671	0.348
	regTPS	0.461	0.700	0.354
Sce. 2	SPDE	0.408	0.723	0.312
	regTPS	0.399	0.776	0.307
Sce. 3	SPDE	0.343	0.900	0.273
	regTPS	0.357	0.893	0.283
Sce. 4	SPDE	0.312	0.854	0.243
	regTPS	0.398	0.788	0.304

SPDE method gives us negative difference values compared with the true covariance matrix. For comparison purposes, we evaluated the predictive performance of the models using Root Mean Square Error (RMSE, defined as before), the Pearson correlation $R^2 = \left(\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \right)^2$ and Mean Absolute Error (MAE $= \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$).

Table 4 summarizes the common metrics used to evaluate the predictive performance of a spatial statistical model as presented before. As expected, the SPDE approach outperforms the regTPS models. This is because the data were generated under a Matérn covariance function, for which the SPDE method has a direct representation. Nevertheless, the statistical performance difference between the SPDE and regTPS models is relatively small. In fact, for the RMSE, MAE and R^2 metrics, the discrepancy between the true GRF and the predictions from regTPS is close to zero. Moreover, in Sce. 1, the regTPS model exhibits better predictive performance than the SPDE model.

Figure 7 presents the interpolation of the approximated GRF obtained with the SPDE and regTPS models and their comparison with the true GRF. Across all scenarios, both approaches are able to correctly reproduce the underlying spatial structure of the true spatial field. This holds even when the number of spatial observations is relatively small (for example, Sce. 1 only considers 50 spatial observations associated with their respective 50 spatial locations).

Figure 8 shows the posterior predictive distributions for both approaches. This allows us to evaluate the complete probabilistic description of model predictions, rather than limiting inference to point estimates. By incorporating both process and measurement uncertainty, the predictive posterior allows for rigorous quantification of uncertainty around predictions. From the figure, we can see that both methods can correctly reproduce new simulated data from the posterior distribution obtained for the parameters in the models for all scenarios.

As an additional comparison based on the RMSE, R^2, and MAE metrics, we simulated a GRF using the exponential covariance function (Sect. 2.1) and applied the same procedure described in Sect. 4.1 to generate the observed data. The results, presented in Table 5, indicate that SPDE achieves consistently lower RMSE and MAE values than regTPS in three of the four scenarios (Sce. 1, Sce. 3, and Sce. 4), although regTPS remains highly competitive in all of them. However, regTPS outperforms SPDE for all the metrics in Sce. 2. These findings suggest that the regTPS approach can be effectively extended beyond the Matérn covariance function to exponential covariance functions while retaining robust predictive performance, even when the underlying data-generating mechanism is based on the SPDE framework.

Figure 9 presents the computational efficiency of the SPDE and regTPS models across all scenarios. Computational efficiency is measured as the logarithm of the median Bulk Effective Sample Size (Bulk ESS) generated per unit time, $\log\left(\frac{\text{median(Bulk ESS)}}{t}\right)$, where (t) denotes the total runtime in minutes. The Bulk ESS is computed across all monitored parameters and provides a robust summary of the sampling efficiency of the HMC algorithm. Higher values indicate that more effectively independent posterior samples are obtained per unit time. As shown in Fig. 9, the regTPS model is more computationally efficient than the SPDE model in all scenarios. Thus, regTPS consistently demonstrates better computational performance than SPDE models.

Finally, to evaluate scalability larger sample sizes, we extended the comparison to four additional scenarios with $N = 400, 800, 1200, 1600$ spatial locations. The SPDE mesh resolution was refined with sample size according to cutoff $= \rho/(2\sqrt{N/50})$, thereby linking mesh density to the spatial range parameter ρ. For the regTPS approach, the number of spline basis functions was set to $k = \min(0.7\, n_{\text{mesh}},\ 450)$, preventing k from approaching N and

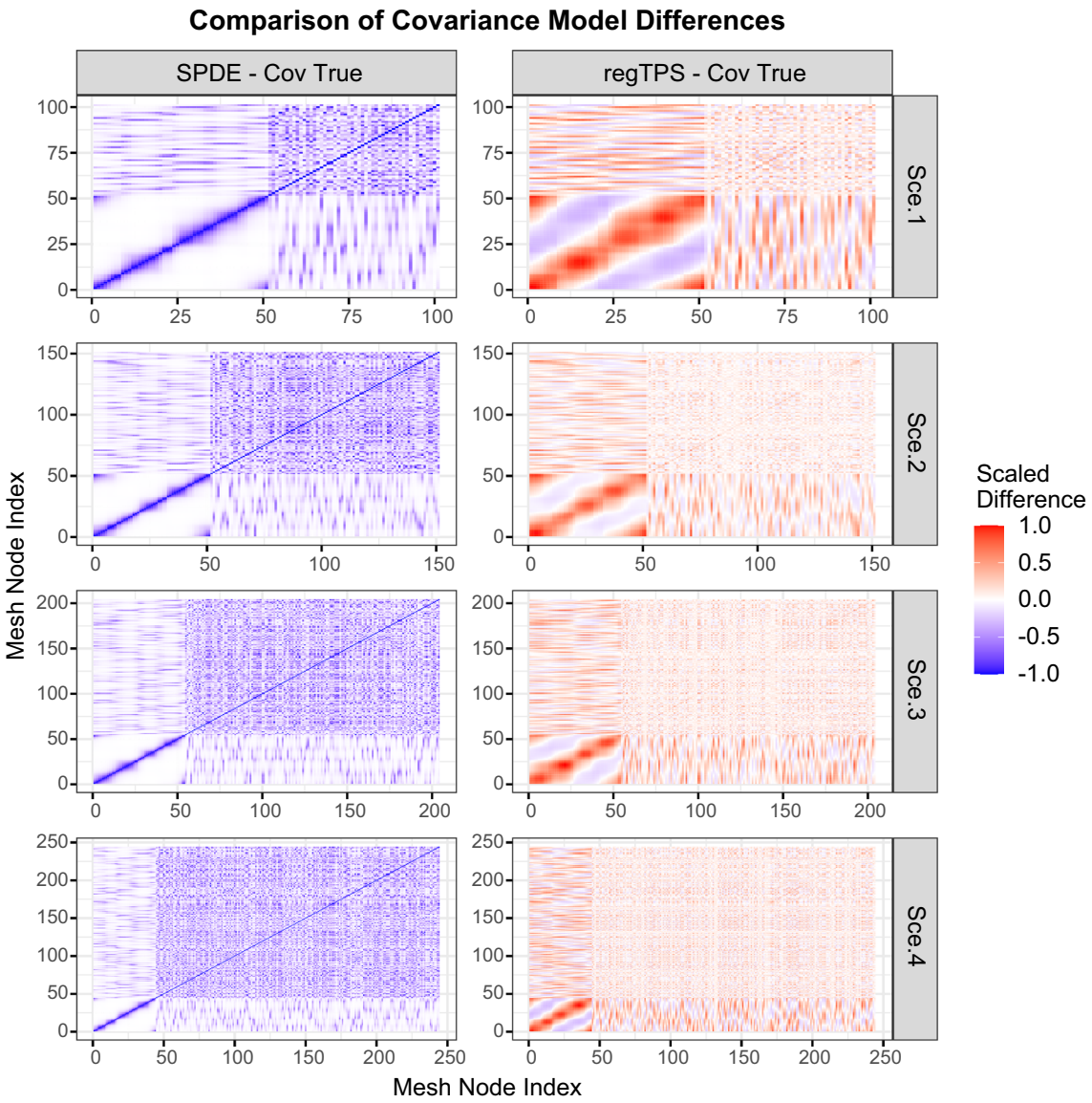


Fig. 6 Normalized estimation errors of the SPDE (left side) and regTPS model (right side) relative to the true Matérn covariance in Simulation 1. Spatial patterns of the heatmap reveal underestimation (blue) and overestimation (red) for each model

avoiding instability in spline extrapolation. The number of retained basis functions (M) was selected adaptively using a 95% variance-explained criterion (Table 6). In this case, both models were fitted only to simulated spatial field using the Matérn covariance function under identical MCMC settings, enabling a direct comparison of effective dimensionality, sampling efficiency, and reconstruction accuracy as a function of N under a fixed and reproducible scheme.

In most evaluated scenarios, the regTPS approach demonstrated better computational advantage over the SPDE method, delivering superior dimension-adjusted computational efficiency due to its substantially lower effective dimensionality. For both methods, RSME values decreased monotonically as the number of spatial locations increased. SPDE models yielded RMSE values ranging from 0.279 down to 0.187, whereas regTPS achieved slightly lower values ranging from 0.270 down to 0.181. Consequently, while overall differences in predictive accuracy between the two methods remained marginal across nearly all scenarios, regTPS consistently maintained a slight advantage (Fig. 10).

4.2 Simulation 2

In addition to the Simulation 1 study, we consider a design in which the true spatial field is generated independently of the discretization method. Specifically, the spatial field is simulated jointly at the observation and prediction-grid locations directly from both Matérn and exponential covariance func-

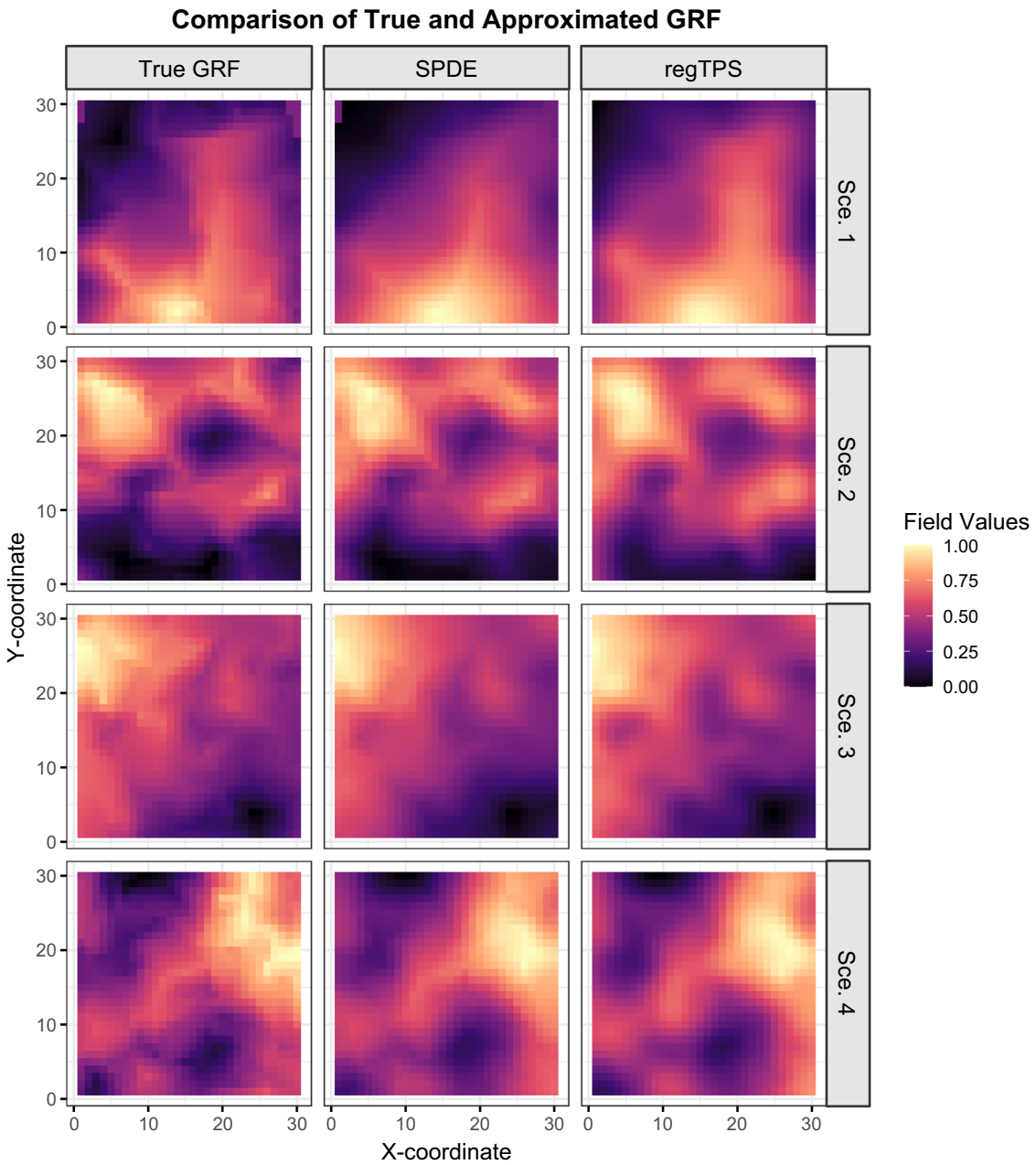


Fig. 7 Posterior medians of the true GRF and the approximated GRF by the SPDE and regTPS methods across Sce. 1 to Sce. 4. in Simulation 1

tions, without reference to any triangulation mesh or spline basis; a single realization is drawn over the union of these locations to preserve the correct spatial dependence between observed and predicted values, which is then partitioned into the observed and grid components. Both the SPDE and regTPS models are subsequently fitted to this same simulated dataset independently, each approximating the underlying GRF using its own discretization without either representation being privileged by the simulation mechanism itself. In summary, Simulation 1 provides a conservative benchmark that structurally favors SPDE over regTPS, a design asymmetry resolved by Simulation 2 (Sect. 4.1).

4.2.1 Simulation 2 results

For a spatial field simulated using a Matérn covariance function, the regTPS approach demonstrated strong predictive performance relative to the SPDE method for all the metrics (Table 7). In Simulation 2, regTPS outperformed SPDE in three out of the four scenarios (Sce. 1, Se. 2, and Sce. 4). The main difference occurred in Sce. 2, where regTPS

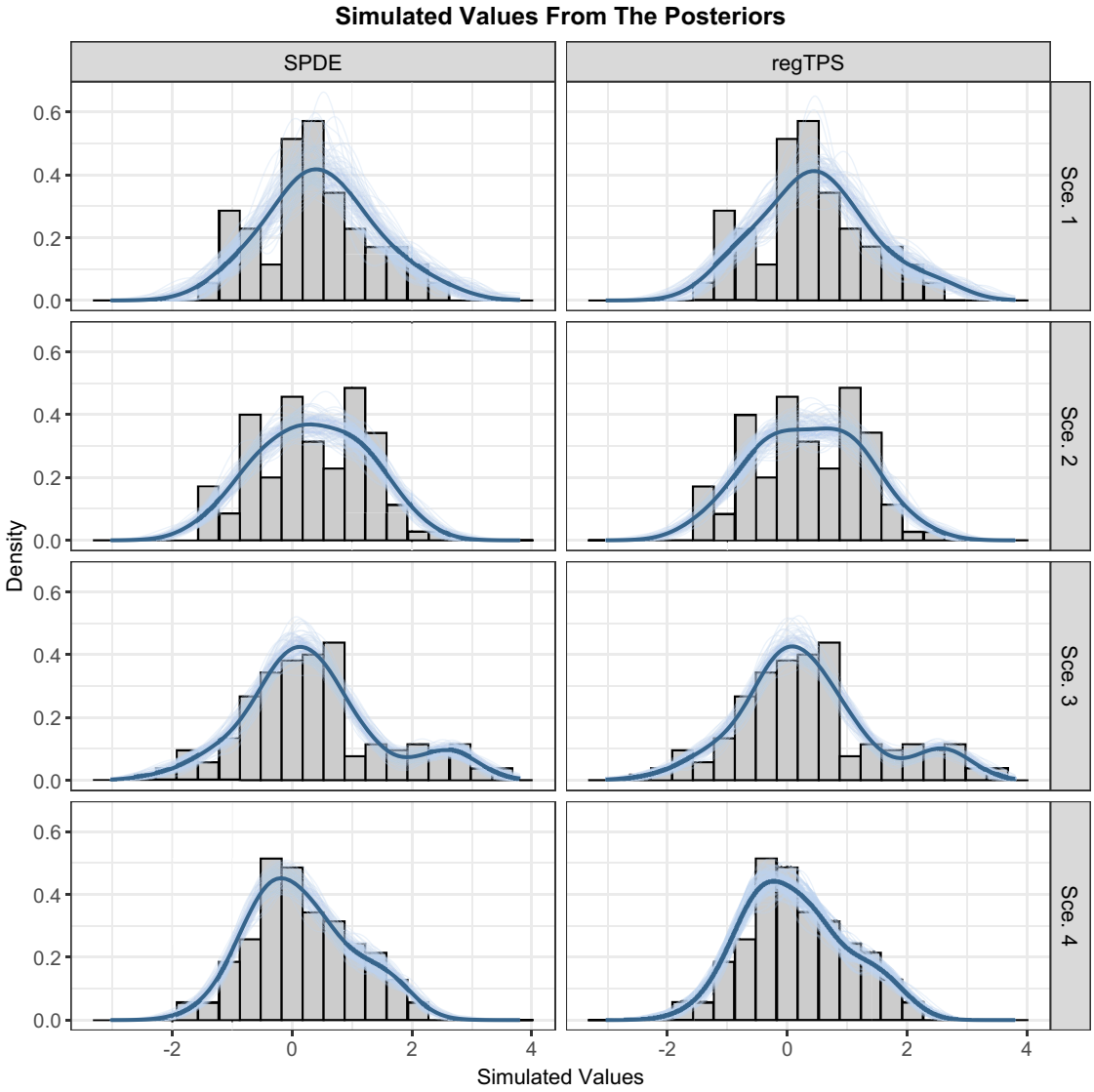


Fig. 8 Observed values (histogram) and 100 posterior predictive distributions (light blue lines) from the SPDE and regTPS methods for Sce.1 to Sce. 4 in Simulation 1. The "dark" blue line is the posterior mean of all the posterior predictive distributions

reduced RMSE from 0.471 to 0.455 and MAE from 0.353 to 0.350, while improving R^2 from 0.682 to 0.717. In Sce. 4, regTPS maintained a small difference across all three evaluation metrics. In contrast, both methods yielded comparable predictive accuracy in Sce. 3, where performance differences were minimal (RMSE of 0.323 for SPDE versus 0.324 for regTPS).

For an exponential covariance function, the results in Table 8 show that the regTPS method maintains competitive accuracy relative to the SPDE model in Simulation 2. In particular, regTPS outperforms SPDE for all the metrics in Sce. 3, whereas SPDE obtain better results in Sce. 1, Sce. 2, and Sce. 4. Overall, the predictive performance of regTPS remains stable and closely aligned with that of the SPDE model when extended to an exponential covariance function.

As expected, the predictive performance of the SPDE and regTPS models in Simulation 2 is lower than those observed in Simulation 1. This occurs because the simulated spatial field is no longer smoothed via FEM projection, causing it to preserve sub-element-scale variability that falls outside the representation capacity of both bases.

Regarding computational efficiency, the regTPS model consistently outperformed the SPDE method for all scenarios with different covariance functions. Under the Matérn covariance function, regTPS achieved higher efficiency values, ranging from 8.57 to 7.31, compared to SPDE ranging from 7.26 to 5.11. A similar advantage was maintained under the exponential covariance function, where regTPS recorded efficiency values between 8.53 to 6.34, whereas SPDE ranged from 6.88 to 4.61. Overall, regTPS demonstrated a systematic

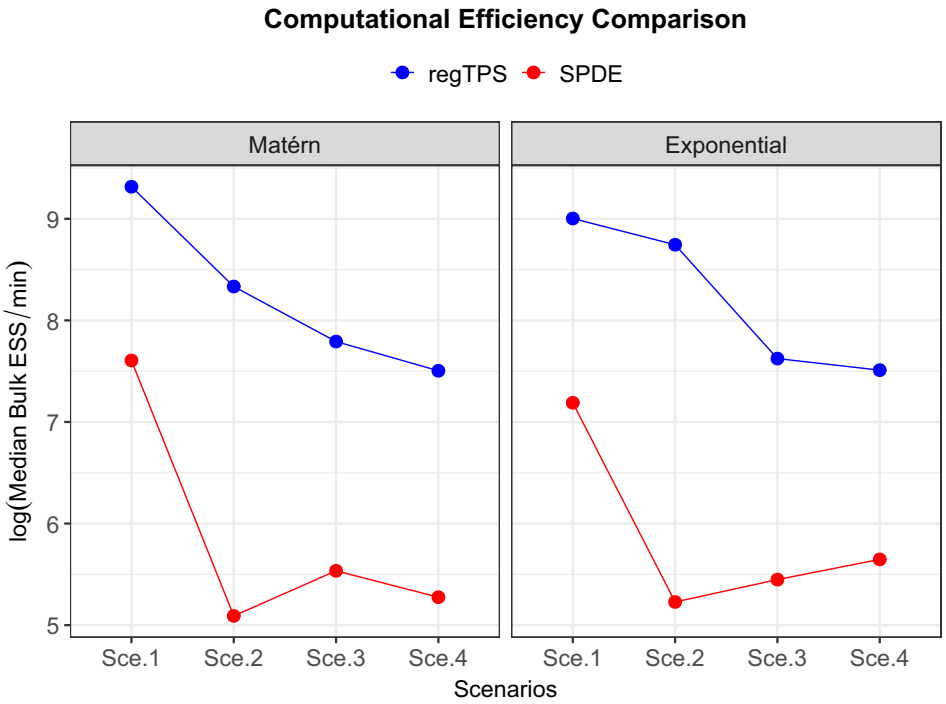


Fig. 9 Computational efficiency $\log\left(\frac{\text{median(Bulk ESS)}}{\text{minutes}}\right)$ calculated for the different scenarios in the SPDE and regTPS models for Simulation 1

Table 6 Model dimensions across scalability scenarios. n_{mesh}: number of SPDE mesh nodes; k: number of TPS basis functions; M: number of retained basis functions

Scenario	Spatial locations (N)	n_{mesh} (SPDE)	k (regTPS)	M (regTPS)
Sce. 5	400	305	214	171
Sce. 6	800	450	315	252
Sce. 7	1200	601	421	341
Sce. 8	1600	754	450	364

Table 7 Comparison of SPDE and regTPS models with the true GRF using a Matérn covariance function in Simulation 2

Scenario	Method	RMSE	R^2	MAE
Sce. 1	SPDE	0.482	0.784	0.380
	regTPS	0.480	0.799	0.372
Sce. 2	SPDE	0.471	0.682	0.353
	regTPS	0.455	0.717	0.350
Sce. 3	SPDE	0.323	0.923	0.253
	regTPS	0.324	0.925	0.256
Sce. 4	SPDE	0.335	0.854	0.262
	regTPS	0.329	0.860	0.257

Table 8 Comparison of SPDE and regTPS models with the true GRF using a exponential covariance function in Simulation 2

Scenario	Method	RMSE	R^2	MAE
Sce. 1	SPDE	0.685	0.552	0.539
	regTPS	0.727	0.525	0.574
Sce. 2	SPDE	0.593	0.549	0.455
	regTPS	0.610	0.561	0.462
Sce. 3	SPDE	0.517	0.788	0.411
	regTPS	0.514	0.793	0.408
Sce. 4	SPDE	0.493	0.702	0.393
	regTPS	0.500	0.698	0.397

computational advantage across the scenarios, maintaining high efficiency regardless of the underlying covariance specification (Fig. 11).

5 Real application: NO_2 data

For our analysis, we use the concentration measurements from 2017 collected national ground monitoring stations in Germany and the Netherlands (474 in total). The original hourly data were obtained from the European Environment Agency (EEA). Negative values were treated as missing; an annual average concentrations were calculated by taking the mean of available hourly measurements while omitting the missing values. As shown in Fig. 12, the observed data are not Gaussian distributed; therefore, a square root transformation was applied to have an approximate Gaussian distribution.

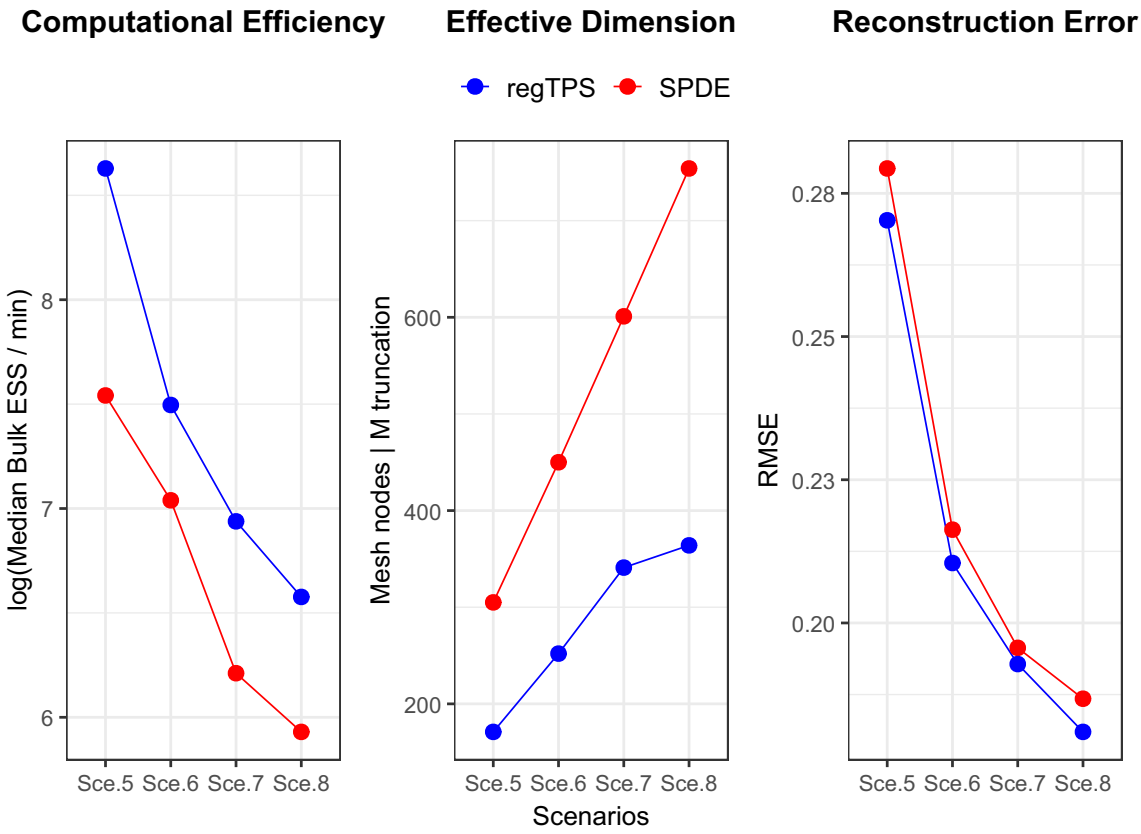


Fig. 10 Comparison of SPDE and regTPS across scenarios in terms of computational efficiency, effective dimension, and reconstruction error (RMSE) in Simulation 1

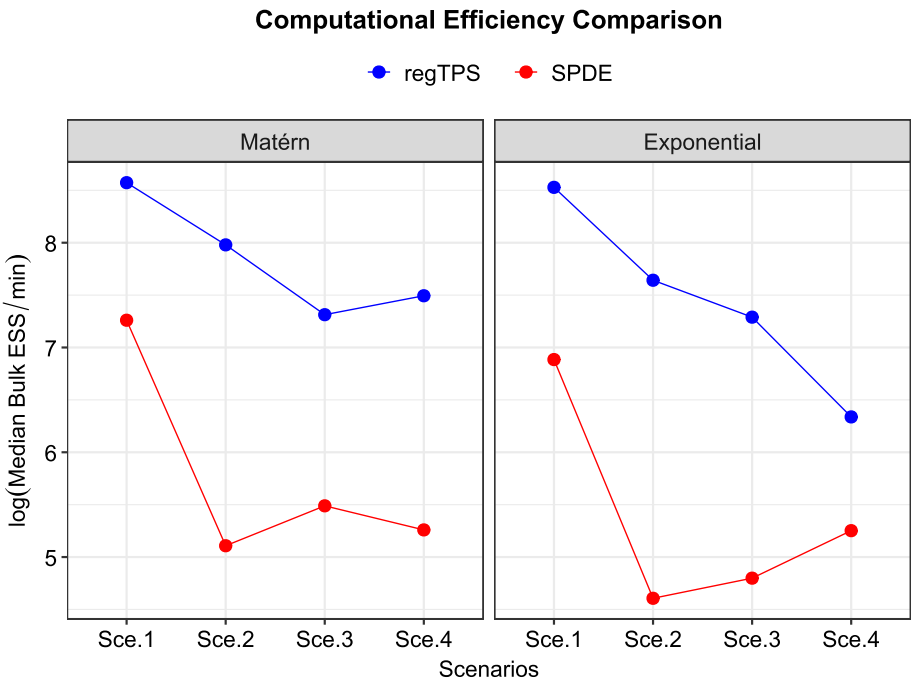


Fig. 11 Computational efficiency $\log\left(\frac{\text{median(Bulk ESS)}}{\text{minutes}}\right)$ calculated for the different scenarios in the SPDE and regTPS models in Simulation 2

Furthermore, Lu et al. (2023) showed that a Gaussian likelihood provides the best fit to the observed data. Therefore, we focus only on the German dataset, given the higher number of monitoring stations compared with the Netherlands (386 against 88, respectively). A detailed description of the data can be found in (Lu et al. 2020).

Thus, to predict/interpolate nitrogen dioxide (NO_2) concentrations across Germany using the 386 monitoring stations, we use the SPDE and regTPS methods for comparative purposes. Moreover, to rigorously evaluate the predictive accuracy of regTPS., we additionally use the package spNNGP available in the software R, which fits univariate Bayesian spatial regression models for large datasets using nearest neighbor Gaussian processes (Finley et al. 2022). spNNGP package represents the gold standard for scalable, Bayesian spatial process modeling using large datasets, since its primary advantage lies in using directed acyclic graphs to restrict spatial conditioning to a small set of nearest neighbors, drastically reducing computational complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N)$ while maintaining full posterior predictive inference over large spatial grids. In this way, by leveraging monitoring station observations mapped to a common high-resolution spatial grid, all three models reconstruct the underlying spatial field to predict/interpolate the NO_2 concentrations to ensure direct comparability across predictions and posterior uncertainty estimates. The details of the spNNGP model employed for comparative purposes, can be find it in the Appendix E

5.1 Results for the SPDE, spNNGP and regTPS models applied to NO_2 data

Three models quickly converged across all MCMC chains. While PC priors for SPDE models enable a simple hyperprior specification, MCMC sampling remains challenging for the SPDE method. In this case, the parameters ρ and σ_u used a Fréchet and exponential prior distributions, respectively, following the suggestion of Cavieres et al. (2024). However, the chains convergence also is influenced by the mesh configuration used to discretize the spatial domain. In contrast, the regTPS approach is simpler to implement. The only parameter that could cause convergence problems is α. For this analysis, assigning a lognormal prior distribution to α is sufficient to ensure reliable convergence. The spNNGP model was specified using a Matérn covariance function with an intercept and a spatially correlated spatial field. We assigned uniform and inverse-Gamma prior distributions to the spatial range and variance parameters, respectively. For the nugget variance, we also assigned an inverse-Gamma prior, whereas the Matérn smoothness parameter was assigned a Uniform$(0.1, 2)$ prior to have a nearly fair comparison with the SPDE model where $\nu = 1$. We used 15 nearest neighbors to balance the accuracy of the NNGP approximation against its computational cost. Three independent MCMC chains were run, each yielding 3000 post-warm-up samples. Regarding predictive performance, the SPDE model successfully predicted new observations drawn from the posterior, confirming its suitability for modeling the observed NO_2 data. The regTPS model can also correctly represent the NO_2 data, but with higher posterior predictive values compared to the SPDE and spNNGP models (Fig. 13).

The SPDE, spNNGP and regTPS models were applied to predict/interpolate the NO_2 concentrations across Germany (Fig. 14). The three models exhibit comparable large-scale spatial patterns, particularly in the lower quantile and posterior median, where most of the average pollution in the regions is consistently identified. However, the SPDE model shows a tendency to overestimate the NO_2 concentrations in the posterior median, which may reflect its spatial smoothing properties. However, there is a high discrepancy in the upper quantiles since the SPDE model predicts elevated concentrations in the north and southeast of Germany, whereas the regTPS model attributes the highest concentrations values to the western regions. The predictions from the spNNGP model were mostly consistent with those from the regTPS model. Both models indicated moderate NO_2 concentrations in western and southern of Germany, with high concentrations (0.975 quantile) concentrated near the southwestern border. However, although the three models exhibit similar spatial prediction patterns, the SPDE model produces higher predicted values than the spNNGP and regTPS models. This difference is particularly evident in the posterior means and the 0.975 posterior quantiles, especially in the north of Germany.

Given the similar spatial patterns of the predicted and interpolated NO_2 concentrations, we evaluated the predictive performance of the SPDE, spNNGP and regTPS models, using the leave-one-out cross validation criterion (loo-cv, Vehtari et al. (2017)). It is a technique that evaluates predictive performance by systematically omitting each observation from the NO_2 data and assessing how well the model predicts the held-out value. Essentially, its use Pareto-smoothed importance sampling (PSIS-LOO), which re-weights posterior draws to approximate the leave-one-out distribution providing an efficient estimate of out-of-sample predictive accuracy. In this case, the regTPS model achieved the highest ELPD and was therefore used as the reference model. However, its lead over spNNGP was modest, with an ELPD difference of 6.8 (SE 5.1), indicating that the performance advantage is minimal. Relative to regTPS, the spNNGP model had a ΔELPD of -6.8 with a standard error of 5.1, whereas the SPDE model had a ΔELPD of -23.9 with a standard error of 7.6. The latter result indicates a substantial and statistically significant reduction in predictive accuracy for the SPDE model (Table 9).

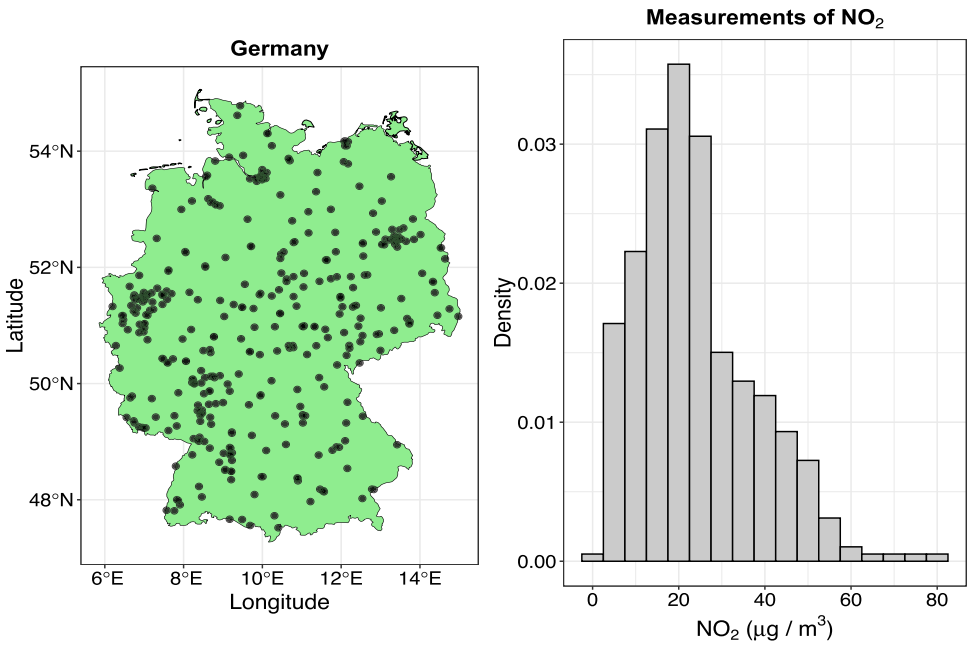


Fig. 12 Geographical distribution of ground stations in Germany and histogram for NO_2 measurements used in this study

Table 9 LOO-CV model comparison between the SPDE, spNNGP, and regTPS models. ΔELPD denotes the difference in expected log predictive density relative to the best-performing model, and SE denotes the standard error of this difference. In the third column: M is the number of retained basis functions for regTPS, n_{mesh} is the number of SPDE mesh nodes, and Nei denotes the number of nearest neighbors used in the spNNGP

Model	ΔELPD	SE	M / n_{mesh} / Nei
regTPS	**0.0**	–	$M = 207$
spNNGP	−6.8	5.1	Nei = 15
SPDE	−23.9	7.6	$n_{\text{mesh}} = 300$

Table 10 Comparison of SPDE, spNNGP, regTPS models in the prediction/interpolation of NO_2 across Germany using 10-fold cross-validation

Model	RMSE	R^2	MAE	ELPD (10-fold)
SPDE	1.301	0.209	1.06	−649.6
regTPS	1.270	0.184	1.03	−638.7
spNNGP	1.220	0.220	1.01	−628.5

According the Pareto-k diagnostics (see Vehtari et al. (2017)), most observations were well supported by the PSIS

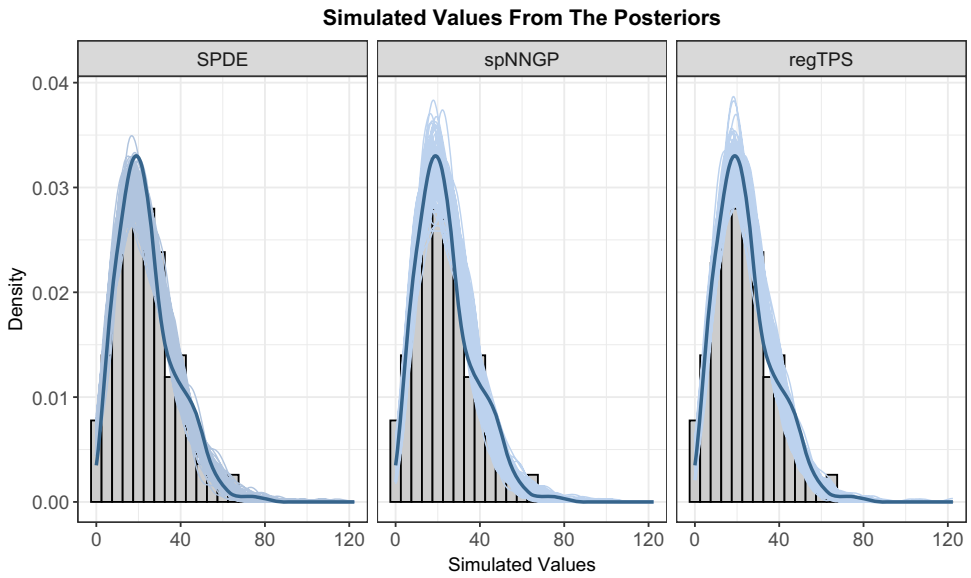


Fig. 13 Observed data (histogram) and 100 posterior predictive distributions (light blue lines) from the SPDE, spNNGP and regTPS models applied to the NO_2 pollution in Germany. The "dark" blue line is the posterior mean of all the posterior predictive distributions

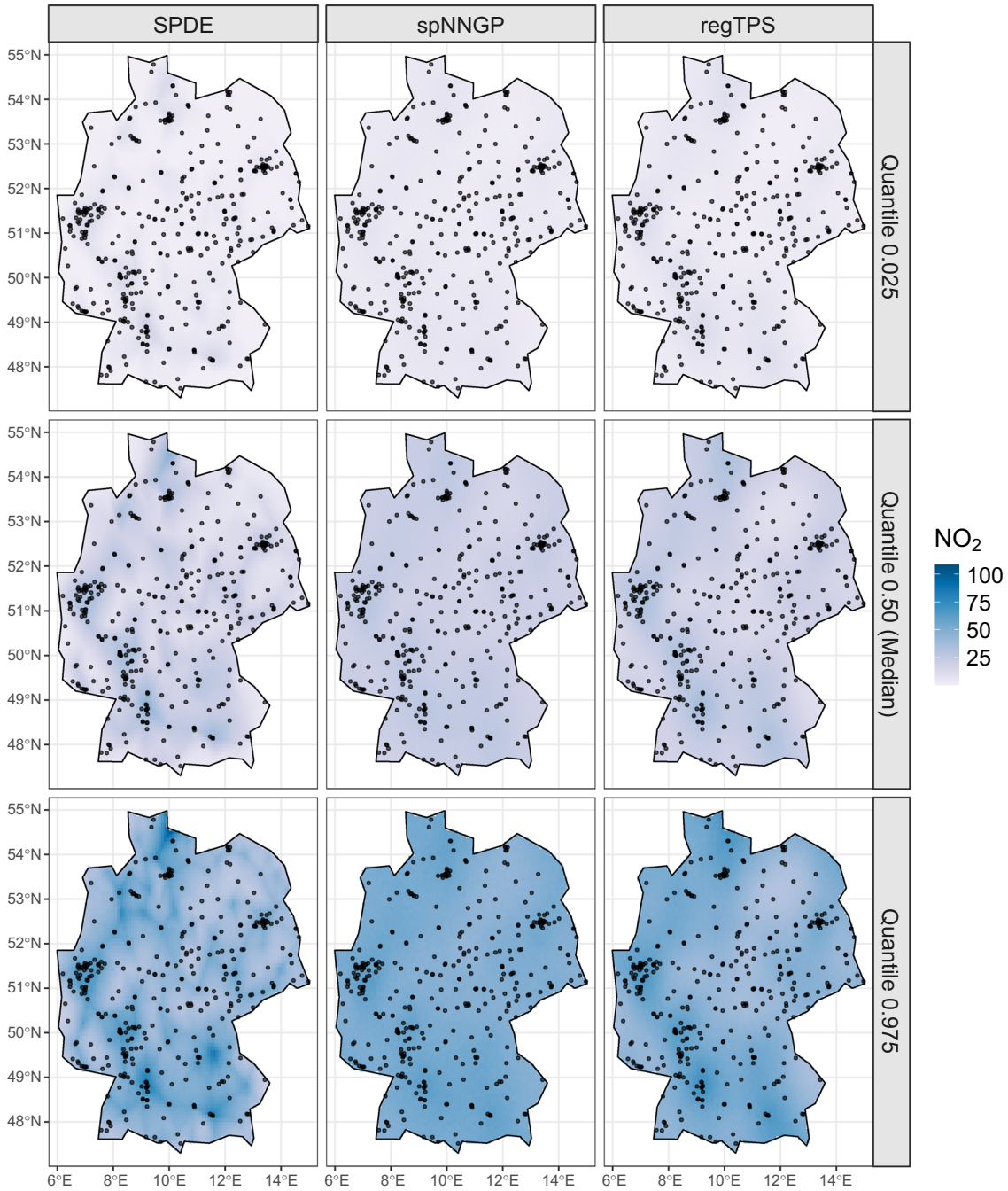


Fig. 14 Posterior median of NO_2 predictions/interpolations, and lower and upper quantiles (2.5% and 97.5% respectively) in Germany for the SPDE, spNNGP and regTPS models

approximation; however, the SPDE model exhibited several observations with relatively high Pareto-k values. From the SPDE model, 43 observations $0.7 < k \leq 1$ and any one observation had $k > 1$ (diagnostic metric k indicates that when $k < \min(1 - 1/\log 10(S), 0.7)$, where S is the sample size, then the model is well specified, otherwise the model could indicate problems (e.g. influence of extreme values in the data), suggesting that the PSIS-LOO approximation is less reliable for these observations. These higher values do not necessarily indicate problems with model convergence, but rather that a small number of observations may have a strong influence on the leave-one-out predictive assessment (Fig. 15).

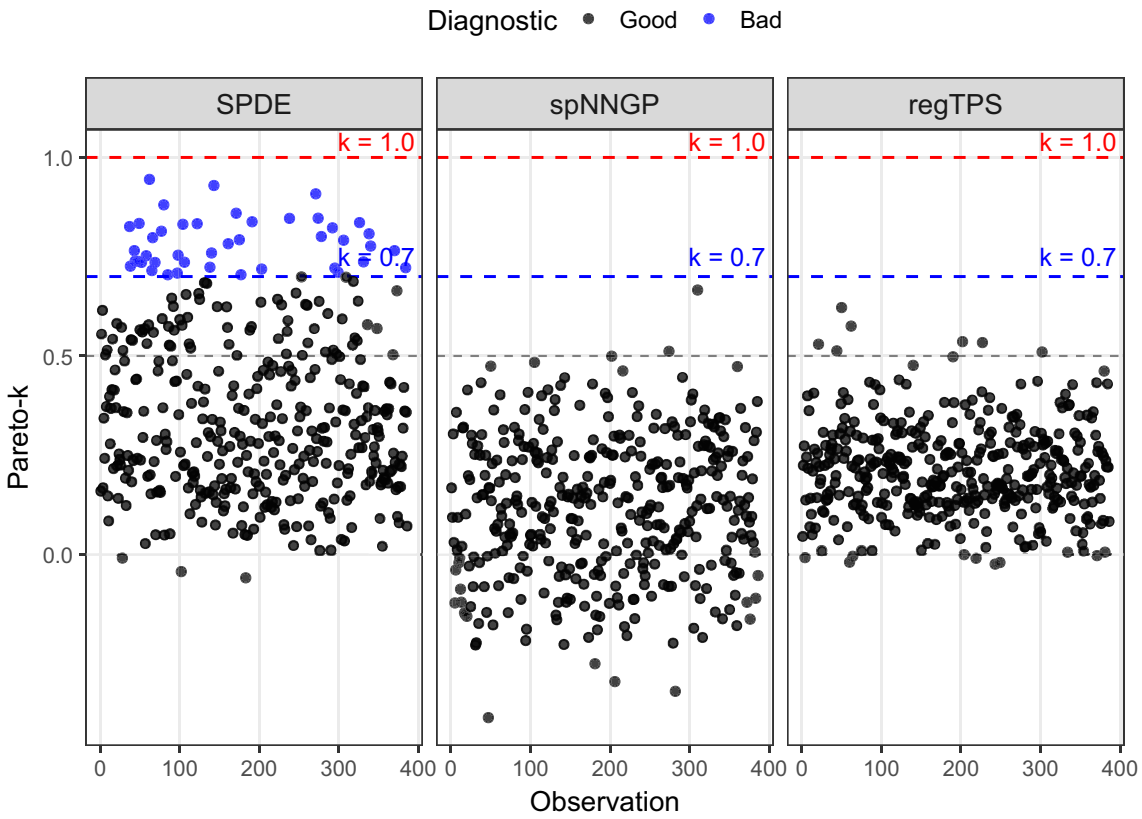


Fig. 15 Diagnostic Pareto shape values (k) for SPDE, spNNGP, and regTPS models. Values $k > 1$ indicate an unreliable PSIS-LOO approximation (red line), while values $0.7 < k \leq 1$ denote moderate observation-level instability (blue line)

To complement these results, we conducted a 10-fold cross-validation analysis, in which the observations were randomly partitioned into ten approximately equal-sized folds. Each fold was sequentially used as a held-out test set, while the remaining observations were used for model fitting. The same folds and spatial locations were used for the SPDE, spNNGP, and regTPS models to ensure a direct comparison of their predictive performance. As in the simulation studies, performance was assessed using the RMSE, R^2 and MAE (Table 10). The spNNGP model exhibited the best predictive performance, with an RMSE of 1.220, an MAE of 1.01, and an R^2 value of 0.220. The regTPS model showed the second best performance in terms of RMSE and MAE, with values of 1.270 and 1.03, respectively. However, the R^2 was lower than that of the SPDE model (0.184 versus 0.209, respectively). Furthermore, the ELPD value for 10-fold in the SPDE model aligns with the PSIS-LOO results, reinforcing its limited predictive capability in the real application.

Computational performance was evaluated by separately recording posterior sampling and prediction times for each model. Because SPDE, spNNGP and regTPS exhibit different model structures, sampling and prediction times are reported separately, together with their sum as the total computational cost. This distinction is particularly relevant for spNNGP, which provides comparatively fast posterior sampling but requires additional computational effort for spatial prediction, whereas SPDE and regTPS medium sampling costs but allow substantially faster prediction through their explicit spatial representations. Table 11 shows that, the computational times differed substantially across the three models. SPDE and regTPS models took similar posterior sampling times (9.46 and 9.40 min, respectively), whereas spNNGP was considerably faster during sampling (1.21 min). However, for the spatial predictions, spNNGP required 29.74 min compared with only 0.59 min for SPDE and 0.52 min for regTPS models. Consequently, the total computational time was lowest for regTPS (9.92 min), closely followed by SPDE (10.05 min), while spNNGP required more time for the complete computation (30.95 min).

6 Discussion

Considering all the previous analyses and comparisons between the SPDE and regTPS approaches, it is important to recognize their fundamental differences in modeling

Table 11 Time of computation for the SPDE, spNNGP, and regTPS models. Sampling time denotes the computational time required for posterior sampling, whereas prediction time denotes the time required to obtain posterior predictions at the prediction locations. Total computational time is simply the sum of the previous two times

Model	Sampling time (min)	Prediction time (min)	Total time (min)
SPDE	9.46	0.59	10.05
spNNGP	1.21	29.74	30.95
regTPS	9.40	0.52	9.92

assumptions. The regTPS approach, while providing strong theoretical development and offering computational advantages through eigendecomposition, represents a restrictive subclass of GRFs compared to the SPDE framework. A key distinction lies in the treatment of spatial smoothness. The SPDE method provides flexibility in controlling the smoothness parameter ν of the Matérn covariance family (Whittle 1963; Stein 1999), which directly governs the mean-square differentiability of the spatial random field. Specifically, a GRF with Matérn covariance is $\lfloor \nu \rfloor$ times mean-square differentiable (Williams and Rasmussen 2006; Gneiting et al. 2010). By varying ν, practitioners can accommodate spatial processes ranging from rough, non-differentiable fields ($\nu = 0.5$, corresponding to the exponential covariance) to infinitely smooth fields ($\nu \to \infty$, approaching the squared exponential covariance). More generally, the connection between SPDE operators and Matérn smoothness is given by the relationship $\nu = \ell - d/2$, where ℓ denotes the order of the differential operator and d is the spatial dimension (Lindgren et al. 2011). For the biharmonic operator in two dimensions, we have $\ell = 2$ (since Δ^2 is a fourth-order operator involving second derivatives), yielding $\nu = 2 - 2/2 = 1$ for the standard TPS formulation (Wahba 1990). While the regularization parameter α controls the effective spatial range (correlation length) of the field, with larger α inducing stronger smoothing and longer-range spatial dependence, it cannot alter the fundamental smoothness class determined by the operator order. To achieve different levels of differentiability within the TPS framework, one would need to change the order of the differential operator itself (e.g., using a higher-order operator such as Δ^3 in two dimensions to obtain $\nu \approx 2$), though this would require reconfiguring the entire basis and penalty structure (Duchon 1977; Meinguet 1979). In this way, the smoothness constraint has both advantages and limitations. On the positive side, fixing ν reduces the number of hyperparameters to estimate and eliminates potential identifiability issues that can arise when jointly estimating smoothness and range parameters in the Matérn model (Zhang 2004; Stein et al. 2004). The mean-square differentiable smoothness class $\nu = 1$ is often appropriate for many environmental and physical processes, providing a reasonable default that balances flexibility with computational tractability (Cressie 1989; Stein 1999). However, the fixed smoothness assumption may be restrictive for applications where the true spatial process exhibits markedly different regularity properties, such as processes with sharp discontinuities or, conversely, phenomena requiring substantially smoother representations. Another important consideration when comparing the SPDE and regTPS approaches is the relationship between model complexity (degrees of freedom) and computational efficiency. In the SPDE models the number of mesh nodes substantially exceeded the number of retained basis functions M in the regTPS approach across all scenarios. In the Simulation 1, the regTPS method achieved dimension reductions ranging from 55% to 28% relative to the SPDE method (Sce. 1 to Sce.4, respectively), while retaining 99% of the cumulative variance explained. This difference directly impacts computational performance: fewer degrees of freedom translate to more efficient MCMC sampling. The SPDE approach faces a fundamental trade-off between mesh resolution and computational tractability. Reducing the number of mesh nodes to improve computational speed often leads to convergence difficulties in MCMC sampling and potential loss of spatial resolution. Conversely, constructing high-quality, fine-resolution triangulations, necessary for capturing detailed spatial structure, produces accurate results but substantially increases computational burden. This trade-off between mesh resolution, computational efficiency, and MCMC convergence represents a key limitation of the SPDE approach when implemented via MCMC sampling (Cavieres et al. 2024). In contrast, the regTPS framework provides a more automated dimension reduction mechanism: the truncation parameter M is selected objectively via variance retention, and the eigenvalue decay structure ensures that discarded high-frequency modes contribute negligible information. This adaptive truncation typically yields a parsimonious representation without requiring manual mesh tuning. Additionally, the two approaches differ in the number of hyperparameters requiring estimation. The SPDE model estimates three hyperparameters (σ_u, ρ, σ_ε), whereas the regTPS method estimates two (α, σ_ε). While this difference appears minor, the reduction from three to two hyperparam-

eters marginally simplifies the posterior geometry and can improve MCMC mixing, particularly when using the non-centered parameterization described in Sect. 3. Moreover, the regularization parameter α in the regTPS method has a direct interpretation as controlling the trade-off between data fidelity and smoothness, whereas the joint interpretation of σ_u and ρ in the Matérn covariance function is less intuitive for practitioners unfamiliar with spatial statistics. For large sample sizes, regTPS retained a clear practical advantage over the SPDE models in terms of dimensionality reduction and computational efficiency. In all scenarios, the number of retained basis functions (M) was approximately 43–52% lower than the corresponding SPDE mesh dimension. This reduction translated into higher computational efficiency, particularly in Sce. 5, Sce. 7 and Sce. 8, where regTPS achieved substantially larger effective sample sizes per minute. The advantage decreased in Sce. 6 but increased again in Sce. 7, coinciding with an increase in the computational cost of the SPDE models as the mesh became denser. Despite using less basis functions, the accuracy results demonstrate that regTPS consistently matches or marginally outperforms the SPDE approach in Sce. 5 to Sce. 8. For both methods, the RMSE values decreased monotonically as sample size (N_{sp}) increased from 400 to 1600. Specifically, SPDE models yielded RMSE values dropping from 0.279 to 0.187, whereas regTPS achieved slightly lower errors ranging from 0.270 down to 0.181. These results indicate that the dimension reduction inherent to regTPS does not introduce limitations for achieve high accuracy in predictions. While low-rank representations risk discarding local spatial variability by omitting higher-frequency eigencomponents (Stein 2014; Wikle 2010), the truncation levels chosen in the large sample sizes, retained sufficient spectral information to preserve predictive accuracy. A related question is how the construction behaves when the Laplace operator acts on a three-dimensional domain. As shown in Sect. 3.4, the framework transfers without structural modification: the regularized biharmonic operator remains trace class for $d \leq 3$, the truncation rule is unchanged, and the induced prior corresponds to a Matérn field of smoothness $\nu = m - d/2$, so the correspondence with the SPDE approach established in the plane carries over. What changes is the efficiency of the representation. Because the spectrum decays as $k^{-2m/d}$, the number of modes required to retain a fixed fraction of the spectral variance grows as $\varepsilon^{-d/(2m-d)}$, from ε^{-1} in two dimensions to ε^{-3} in three. Two practical consequences follow. First, the reduction in effective dimension reported here should be expected to be smaller in volumetric applications. Second, the comparison with the SPDE method shifts: the TPS penalty eigenproblem is dense and scales cubically in the number of knots, whereas the SPDE precision matrix remains sparse, so the sparsity advantage of the finite-element representation becomes more valuable as d increases. Quantifying this trade-off requires a three-dimensional benchmark with volumetric observations, which we leave to future work. We note that the common case of a planar domain observed repeatedly over time is better treated by a separable space-time extension of the present construction than by an isotropic three-dimensional penalty. Simulation 2 evaluated both approaches under a data-generating process independent of either discretization framework. When the true spatial process was drawn directly from a Matérn covariance function, regTPS demonstrated superior predictive accuracy compared to the SPDE approach, achieving lower RMSE and MAE alongside higher R^2 values in three out of four scenarios. For data generated using an exponential covariance function, regTPS maintained highly competitive predictive performance, closely matching SPDE for Sce.1 to Sce. 4 while outperforming it in Sce. 3. These results show that regTPS achieved similar or slightly better predictive accuracy while requiring less computation for both covariance functions. This suggests that its computational advantage is not specific to the simulation design but is instead related to its ability to model complex spatial data and few computational resources. In real data application, the SPDE model exhibited the weakest predictive performance, showing a significant reduction in expected log predictive density ($\Delta\text{ELPD} = -23.9, \text{SE} = 7.6$), highest error rates (RMSE/MAE), and Pareto-k diagnostic problems. According to Cavieres et al. (2024), these issues arises from the difficulty of balancing mesh construction with the computational efficiency of MCMC methods. Moreover, selecting appropriate prior distributions for the hyperparameters is another important consideration. Despite these challenges, SPDE models generally perform well; however, these topics remain limitations when developing accurate and computationally efficient Bayesian models. While Polynomial Chaos Expansion (PCE, Wiener (1938); Ghanem et al. (2017)) and sparse methods (Smolyak 1963; Bungartz and Griebel 2004) provide forward propagation in low-dimensional parametric settings. Sparse Gaussian process approximations based on inducing points (Snelson and Ghahramani 2007; Titsias 2009) address a related low-rank objective. Importantly, regTPS differs in deriving its basis from the spectrum of a regularized differential operator rather than from optimized inducing locations, which yields a deterministic and interpretable ordering of modes by bending energy. Spatial Nearest Neighbor Gaussian Processes (spNNGP, Finley et al. (2022)) offer a more methodologically aligned benchmark for fitting and interpolating continuous spatial data. Incorporating spNNGP for comparison purposes enables a clear, rigorous assessment of the capacity of regTPS for data-driven field estimation in our real applications. Ultimately, this comparison highlights the predictive accuracy and computational efficiency of regTPS for direct field reconstruction from noisy observational data.

7 Conclusion

In the present work, we propose a novel approach to approximate a Gaussian random field (GRF) using a thin plate spline (TPS) as a covariance kernel. Since the TPS is a conditionally positive definite kernel, it cannot be used directly to approximate a GRF within the Karhunen–Loéve expansion (KLE) method. We therefore establish a direct connection between TPS and Hilbert-Schmidt kernels by considering the inverse of an elliptic operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$, which yields L_α^{-1}, a compact, self-adjoint, and positive definite operator. Under appropriate boundary conditions, the integral kernel associated with L_α^{-1} is a well-defined Hilbert-Schmidt kernel and can therefore be used directly within the KLE method. The construction is also dimension-agnostic: it remains admissible on three-dimensional domains without modification, at the price of a slower spectral decay and hence a weaker dimension reduction, as quantified in Sect. 3.4. This construction offers a practical advantage over standard covariance families such as the Matérn or exponential covariance functions: it does not require specifying informative prior distributions for additional correlation hyperparameters, which reduces the need for iterative prior tuning and simplifies MCMC sampling by avoiding the strong parameter correlations that such hyperparameters can otherwise induce. In the regTPS approach, the polynomial null space is additionally given a unit prior variance, allowing the model to represent large-scale spatial trends without incurring a smoothing penalty; explicitly handling the null space in this way offers a theoretical advantage over standard KLE constructions, which frequently omit the mean or handle it less consistently. In Simulation 1, the regTPS approach accurately recovers spatial fields simulated under the (assumed) Matérn covariance function and performs well even when the fields are generated from an exponential covariance function across Sce. 1 to Sce. 4. Although the simulated data were produced using the SPDE approach, regTPS exhibits better predictive performance than the SPDE model in Sce. 1 (Matérn) and Sce. 2 (exponential). We further extended this comparison to larger sample sizes, with up to 1600 spatial observations (Sce. 5 to Sce. 8), confirming that regTPS consistently required substantially fewer effective components M than the corresponding SPDE mesh dimension and remained more computationally efficient. While differences in predictive accuracy between the two methods are marginal across Sce. 5 to Sce. 8, regTPS maintains a slight numerical advantage over the SPDE models. Besides, the convergence analysis in Simulation 1 demonstrates that the prediction error minimizes at a substantially reduced basis dimension ($k = 81$) and stabilizes thereafter. This indicates that a compact basis cuts computational overhead while marginally improving accuracy, providing direct empirical support for basis dimension reduction. In Simulation 2, using a Matérn covariance function, regTPS achieved superior predictive accuracy in Sce. 1, Sce. 2 and Sce. 4 compared to SPDE models, reducing both RMSE and MAE while improving R^2. For an exponential covariance function used to generate the data, regTPS remained highly competitive and outperforming SPDE in Sce. 3. In all the scenarios in Simulation 2, regTPS systematically preserved its computational advantage, recording consistently higher overall and dimension-adjusted efficiency values. In the modeling of NO_2 concentrations, both approaches achieved accurate computation of posterior predictive distributions and spatial interpolation across the spatial domain (Germany). The results indicate similar large-scale spatial patterns in the posterior median of the predicted values, but with some differences in the upper tail of the posterior distribution (97.5% quantile). The SPDE model predicts higher NO_2 concentrations in the northern and southeastern regions of Germany, whereas the regTPS model assigns higher values primarily to the western region. These differences arise from the distinct ways in which the two approaches approximate the GRF. In the SPDE framework, the marginal variance of the spatial field is governed by the hyperparameters ρ and σ_u, which directly parameterize the Matérn covariance function and enter the construction of the sparse precision matrix $\mathbf{Q}$. In contrast, in the regTPS approach, the spatial field variance is determined by the regularization parameter α through the eigenvalue spectrum $\{\lambda_{k,\alpha}\}_{k=1}^{M}$, where each coefficient z_k has prior variance $\lambda_{k,\alpha} = (1 + \alpha v_k)^{-1}$. This induces a smoothness-adaptive prior that allocates variance hierarchically: unit variance to polynomial null space modes ($k \leq d_{\text{null}}$) and progressively shrinking variance to higher-frequency components ($k > d_{\text{null}}$), with the decay rate controlled by α and the bending-energy eigenvalues v_k. Thus, while the SPDE approach uses a parametric Matérn covariance family, the regTPS approach builds the covariance structure non-parametrically through the eigendecomposition of the regularized biharmonic operator. Overall, these results position regTPS as a theoretically well-structured and computationally efficient alternative to the SPDE method, offering consistent gains in effective dimensionality reduction and sampling efficiency alongside good predictive performance. Although the regTPS and SPDE methods are based on fundamentally different methodological principles, their empirical performance was broadly comparable. In several simulation settings, regTPS yielded superior predictive accuracy. Furthermore, the real-data applications showed that spNNGP achieved the lowest point-prediction errors in the 10-fold cross-validation, whereas regTPS obtained the highest expected log predictive density under PSIS-LOO cross-validation. In terms of computational time, regTPS and SPDE were generally more efficient than spNNGP. Thus, regTPS is a flexible, well-structured, and computationally efficient model that performs competitively with widely used methodologies for Bayesian spatial modeling.

A Mass-matrix formulation and generalized Ritz eigenproblem

In the main text, we employ the standard eigenvalue problem (Eq. 33) to obtain the discrete eigenfunctions of the penalty matrix $\mathbf{S}$. Here we provide the theoretical foundation for this choice by contrasting it with the fully consistent Galerkin formulation based on a generalized eigenvalue problem. The Galerkin discretization of the biharmonic operator presented in the main text relies on the bending-energy bilinear form

$$a(u, v) = \int_{\mathcal{D}} (\Delta u)(\Delta v)\, d\mathbf{s},$$

which induces the stiffness matrix $\mathbf{S}$ defined in Equation (32). This formulation corresponds to measuring function roughness in the reproducing kernel Hilbert space (RKHS) associated with TPS. To obtain a fully consistent discrete analogue of the continuous eigenproblem $\Delta^2\phi = v\phi$ in the $L^2(\mathcal{D})$ sense, the Galerkin procedure must also incorporate the $L^2(\mathcal{D})$ inner product

$$m(u, v) = \int_{\mathcal{D}} u(\mathbf{s})\, v(\mathbf{s})\, d\mathbf{s},$$

whose discrete counterpart is the mass matrix

$$\mathbf{M}_{jk} = \int_{\mathcal{D}} \tilde{\phi}_j(\mathbf{s})\, \tilde{\phi}_k(\mathbf{s})\, d\mathbf{s}. \tag{A1}$$

Incorporating both bilinear forms leads to the generalized eigenvalue problem

$$\mathbf{S}\boldsymbol{\psi}_k = v_k^{(N)}\, \mathbf{M}\boldsymbol{\psi}_k, \tag{A2}$$

where $(v_k^{(N)}, \boldsymbol{\psi}_k)$ are discrete Ritz pairs approximating the eigenpairs (v_k, ϕ_k) of the continuous biharmonic operator Δ^2. The associated discrete eigenfunctions are

$$\hat{\phi}_k(\mathbf{s}) = \sum_{i=1}^{K} \psi_{ik}\, \tilde{\phi}_i(\mathbf{s}), \tag{A3}$$

and satisfy orthonormality with respect to the $L^2(\mathcal{D})$ inner product,

$$\int_{\mathcal{D}} \hat{\phi}_k(\mathbf{s})\, \hat{\phi}_\ell(\mathbf{s})\, d\mathbf{s} = \boldsymbol{\psi}_k^\top \mathbf{M}\boldsymbol{\psi}_\ell = \delta_{k\ell}.$$

Under standard regularity conditions, the eigenvalues $v_k^{(N)}$ and eigenfunctions $\hat{\phi}_k$ converge to their continuous counterparts as the approximation space $\mathcal{V}_N$ becomes dense in $H^2(\mathcal{D})$. The Galerkin discretization of the regularized operator $L_\alpha = \mathbf{I} + \alpha\Delta^2$ in this setting yields the matrix representation

$$\mathbf{P}_\alpha^{(M)} = \mathbf{M} + \alpha\mathbf{S}, \tag{A4}$$

where the superscript (M) emphasizes the inclusion of the mass matrix. The corresponding discrete covariance operator is

$$\mathbf{K}_\alpha^{(M)} = (\mathbf{M} + \alpha\mathbf{S})^{-1}. \tag{A5}$$

Diagonalizing the generalized eigenproblem in Equation (A2) yields

$$\mathbf{K}_\alpha^{(M)} = \Psi_M \boldsymbol{\Lambda}_\alpha \Psi_M^\top, \qquad \lambda_{k,\alpha} = \frac{1}{1 + \alpha v_k^{(N)}}, \tag{A6}$$

where $\Psi_M \in \mathbb{R}^{N\times N}$ has the eigenvectors $\boldsymbol{\psi}_k$ as its columns. Thus, $\mathbf{K}_\alpha^{(M)}$ constitutes a discrete KLE consistent with the $L^2(\mathcal{D})$ geometry of the continuous problem. While the formal mass-matrix formulation in Equation(A2) requires solving the generalized eigenvalue problem $\mathbf{S}\boldsymbol{\psi}_k = v_k^{(N)}\mathbf{M}\boldsymbol{\psi}_k$, our implementation adopts a computationally efficient approximation standard in TPS smoothing practice. Specifically, we compute eigenvalues v_k from the standard eigenvalue problem

$$\mathbf{S}\boldsymbol{\psi}_k = v_k\boldsymbol{\psi}_k, \tag{A7}$$

and employ these eigenvalues directly in the regularized covariance formula from Equation (A6):

$$\lambda_{k,\alpha} = \frac{1}{1 + \alpha v_k}. \tag{A8}$$

This approach is justified as a computational approximation commonly adopted in practical TPS implementations and low-rank Gaussian process models (Wood 2017). For irregular domains or highly clustered observation points, the mass matrix $\mathbf{M}$ is not diagonal, and the generalized eigenvalues reflect both roughness and local sampling density. By using the standard eigenvalue problem (A7), we rank modes purely by bending energy, which provides a stable, interpretable low-rank representation and allows the regularization param-

eter α to absorb any constant scaling. The null space of $\mathbf{S}$ (corresponding to polynomial trends with $v_k \approx 0$) is handled by assigning unit variance in the prior, $\lambda_k = 1$ for $k \leq M_P$, ensuring these components remain unpenalized as required by TPS theory (Duchon 1977; Green 1987; Wood 2017).

B The SPDE method

Consider a GRF $u(\mathbf{s})$ with a Matérn covariance function

$$\mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sigma_u^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa \|\mathbf{s}_i - \mathbf{s}_j\|\right)^{\nu} K_{\nu}\left(\kappa \|\mathbf{s}_i - \mathbf{s}_j\|\right), \tag{A9}$$

where $\Gamma(\cdot)$ is the Gamma function, $\kappa > 0$ is the spatial scale parameter, σ_u is the marginal standard deviation, $\nu > 0$ controls smoothness, and K_ν is the modified Bessel function of the second kind. Following the results of Whittle (1954, 1963), the stationary solution of the stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha_{\mathrm{SPDE}}/2}\left(\tau\, u(\mathbf{s})\right) = \mathcal{W}(\mathbf{s}), \qquad \mathbf{s} \in \mathbb{R}^d, \tag{A10}$$

has a Matérn covariance function, where $\tau > 0$, $\alpha_{\mathrm{SPDE}} > d/2$, Δ is the Laplacian, and $\mathcal{W}$ is Gaussian spatial white noise. The condition $\alpha_{\mathrm{SPDE}} > d/2$ required here for the SPDE representation is the exact counterpart of the admissibility condition $2m > d$ derived in Sect. 3.4, under the identification $\alpha_{\mathrm{SPDE}} = m$ implied by Equation (43). If the spatial domain $D \subset \mathbb{R}^d$ is bounded, then $u(\mathbf{s})$ can be approximated by a Gaussian Markov random field (GMRF) using the finite element method (FEM) (Lindgren et al. 2011). Specifically, the GRF is represented as

$$u(\mathbf{s}) = \sum_{k=1}^{n} \psi_k(\mathbf{s}) u_k, \tag{A11}$$

where $\{\psi_k\}$ are piecewise linear FEM basis functions defined on a spatial triangulation of D, and $\mathbf{u} = (u_1, \ldots, u_n)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$. Here, $\mathbf{Q}$ is a sparse precision matrix constructed from FEM mass and stiffness matrices, denoted $\mathbf{C}$ and $\mathbf{G}$, respectively. The FEM matrices are defined as

$$\begin{aligned} C_{ii} &= \langle \psi_i, 1 \rangle, \\ G_{ij} &= \langle \nabla\psi_i, \nabla\psi_j \rangle, \\ \mathbf{K} &= \kappa^2\mathbf{C} + \mathbf{G}. \end{aligned}$$

Using the SPDE framework, the precision matrix $\mathbf{Q}$ can be constructed recursively as a function of κ and α:

$$\begin{aligned} \mathbf{Q}_{\alpha=1} &= \mathbf{K}, \\ \mathbf{Q}_{\alpha=2} &= \mathbf{K}\mathbf{C}^{-1}\mathbf{K}, \\ \mathbf{Q}_{\alpha\geq 3} &= \mathbf{K}\mathbf{C}^{-1}\mathbf{Q}_{\alpha-2}\mathbf{C}^{-1}\mathbf{K}. \end{aligned}$$

In particular, for $\alpha = 2$, one obtains the explicit expression

$$\mathbf{Q} = \tau^2\Big(\kappa^4\mathbf{C} + 2\kappa^2 \underbrace{\mathbf{G}}_{\mathbf{G}_1} + \underbrace{\mathbf{G}\mathbf{C}^{-1}\mathbf{G}}_{\mathbf{G}_2}\Big), \tag{A12}$$

which is the standard formulation used in practice (Lindgren and Rue 2015; Bakka et al. 2018).

B.1 Parameterization

The parameterization of the GRF commonly assumes a joint normal distribution for τ and κ in log scale. Since κ and τ have a joint influence on the variance marginal of the spatial random field, it is often more convenient to parameterize the model through the range ρ and marginal standard deviation σ_u. The general relationships are:

$$\rho = \frac{\sqrt{8\nu}}{\kappa} \qquad \text{and} \qquad \sigma_u = \sqrt{\frac{\Gamma(\nu)}{\tau^2\kappa^{2\nu}(4\pi)^{d/2}\Gamma(\nu + d/2)}}. \tag{A13}$$

For the commonly used case with $d = 2$ (two-dimensional spatial domain) and $\alpha = 2$ (which corresponds to $\nu = \alpha - d/2 = 1$), these relationships simplify to:

$$\kappa = \frac{\sqrt{8}}{\rho} \qquad \text{and} \qquad \tau = \frac{1}{\kappa\sigma_u} = \frac{\rho}{\sqrt{8}\sigma_u}. \tag{A14}$$

This simplified parameterization is used throughout our implementation.

C Model formulations

We consider the following general model for spatial data:

$$y_i = u(\mathbf{s}_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon), \tag{A15}$$

where y_i are the observed values at spatial locations $\mathbf{s}_i \in \mathbb{R}^2$, $u(\mathbf{s})$ is a GRF, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon)$ (with $\sigma_\varepsilon > 0$). From this, we consider two approximations of $u(\mathbf{s})$: the SPDE method and the regTPS approach with the purpose to fit the model.

C.1 SPDE formulation

C.1.1 Non-centered parameterization

To improve sampling efficiency and posterior geometry in MCMC sampling we employ a non-centered parameterization (Papaspiliopoulos et al. 2007; Betancourt and Girolami 2015). Let $\tilde{\mathbf{u}} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\kappa)^{-1})$ be a spatial field. The centered field $\mathbf{u}$ is obtained through:

$$\mathbf{u} = \frac{1}{\tau}\tilde{\mathbf{u}}, \tag{A16}$$

which ensures that $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \tau^{-2}\mathbf{Q}(\kappa)^{-1})$ with the desired marginal standard deviation $\sigma_u = 1/(\tau\kappa)$. This parameterization decorrelates the spatial field from the hyperparameters κ and τ, substantially improving MCMC mixing and convergence. The spatial field at spatial locations is obtained via the projection matrix $\mathbf{A}_{\text{obs}} \in \mathbb{R}^{n\times m}$, which maps the m mesh nodes to the n spatial locations:

$$u(\mathbf{s}_i) \approx (\mathbf{A}_{\text{obs}}\mathbf{u})_i, \quad i = 1, \ldots, n. \tag{A17}$$

which is the approximated GRF $u(\mathbf{s})$ at the spatial locations $\mathbf{s}_i$.

C.1.2 Priors

We specify priors for the hyperparameters:

$$\begin{aligned} \sigma_\varepsilon &\sim \text{Cauchy}^+(0, s), \quad s = 5, \\ \rho^{-1} &\sim \text{Weibull}(\lambda_\rho^{-1}, 1), \\ \sigma_u &\sim \text{Exponential}(\lambda_{\sigma_u}), \end{aligned}$$

where $\text{Cauchy}^+(0, s)$ denotes a half-Cauchy distribution (the positive half of a Cauchy distribution with location 0 and scale s), and the Weibull distribution on ρ^{-1} with shape parameter 1 corresponds to an exponential-like prior that can be parameterized via the probability statement $\Pr(\rho < \rho_0) = p_\rho$ through $\lambda_\rho = -\rho_0 \log(p_\rho)$. The exponential prior on σ_u is specified via the probability statement $\Pr(\sigma_u > \sigma_{0,u}) = p_{\sigma_u}$ through $\lambda_{\sigma_u} = -\log(p_{\sigma_u})/\sigma_{0,u}$.

C.1.3. Joint posterior distribution

Combining the likelihood in Equation (A15) with the GMRF/SPDE prior and hyperpriors, the joint posterior distribution is:

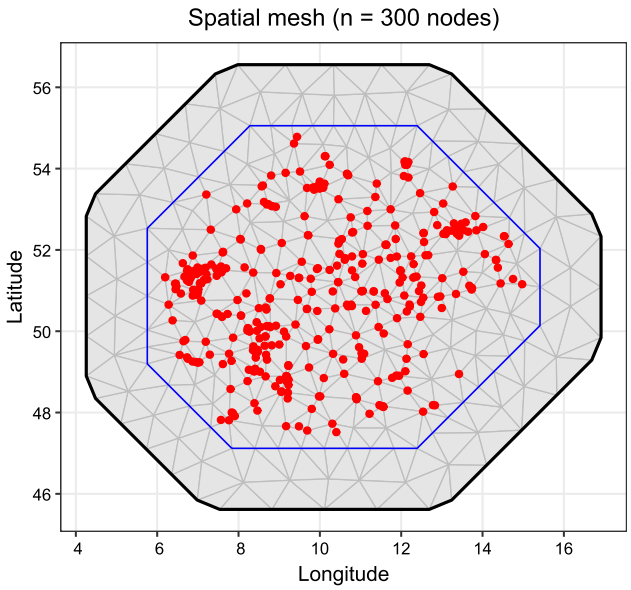


Fig. 16 Mesh constructed for the SPDE model in the real application of NO_2 modeling across Germany

$$p(\tilde{\mathbf{u}}, \sigma_\varepsilon, \rho, \sigma_u \mid \mathbf{y}) \propto \left[\prod_{i=1}^{n} \mathcal{N}\big(y_i \mid (\mathbf{A}_{\text{obs}}\mathbf{u})_i, \sigma_\varepsilon^2\big)\right] \mathcal{N}\big(\tilde{\mathbf{u}} \mid \mathbf{0}, \mathbf{Q}(\kappa)^{-1}\big)\pi(\sigma_\varepsilon)\pi(\rho)\pi(\sigma_u), \tag{A18}$$

where $\mathbf{u} = \tilde{\mathbf{u}}/\tau$ as in Equation (A16), and κ and τ are functions of ρ and σ_u through Equation (A13) (Fig. 16).

C.2. regTPS formulation

In the regTPS approach, the spatial field $u(\mathbf{s})$ is represented using the truncated KLE expansion:

$$u(\mathbf{s}) = \sum_{k=1}^{M} z_k \hat{\phi}_k(\mathbf{s}), \tag{A19}$$

where $\{\hat{\phi}_k\}_{k=1}^{M}$ are the discrete eigenfunctions arising from the eigendecomposition of the TPS penalty matrix $\mathbf{S}$ (Eq. 33), and z_k are the corresponding expansion coefficients. These eigenfunctions form an orthogonal basis in coefficient space, with the null space (first $M_{\text{null}} = d_{\text{null}}$ components) corresponding to low-order polynomials that are unpenalized by the TPS roughness penalty. We truncate the expansion at $M \le K$ components to retain a specified proportion of total variance (typically 95–99%, see Eq. 39).

C.2.1 Non-centered parameterization

To improve posterior geometry and computational efficiency, we adopt a non-centered parameterization. The transforma-

tion is:

$$z_k = \begin{cases} \tilde{z}_k, & k \le M_{\text{null}}, \\ \dfrac{\tilde{z}_k}{\sqrt{1+\alpha v_k}}, & k > M_{\text{null}}, \end{cases} \tag{A20}$$

where $\tilde{z}_k \sim \mathcal{N}(0,1)$ are standard normal random variables, v_k are the eigenvalues of the TPS penalty matrix $\mathbf{S}$, and $\alpha > 0$ is the regularization (smoothing) parameter. This parameterization corresponds to a Gaussian prior with precision matrix $\mathbf{P}_\alpha = \mathbf{I} + \alpha\mathbf{S}$ (Eq. 46), ensuring that null space components remain unpenalized ($\lambda_{k,\alpha} = 1$) while higher-frequency (rougher) modes are increasingly shrunk as α or v_k increases ($\lambda_{k,\alpha} = (1+\alpha v_k)^{-1}$).

Let $\mathbf{\Phi}_{\text{obs}} \in \mathbb{R}^{N\times M}$ denote the pre-computed KLE design matrix whose entries are $(\mathbf{\Phi}_{\text{obs}})_{ik} = \hat{\phi}_k(\mathbf{s}_i)$, i.e., the discrete TPS eigenfunctions evaluated at the observation locations (Eq. 49). The discretized spatial field at observation locations is:

$$\mathbf{u} = \mathbf{\Phi}_{\text{obs}}\mathbf{z}, \tag{A21}$$

where $\mathbf{z} = (z_1, \ldots, z_M)^\top$ with each z_k defined through the transformation in Equation (A20).

C.2.2 Priors

The full Bayesian specification is completed by assigning priors to the model parameters:

$$\begin{aligned} &\sigma_\varepsilon \sim \text{Exponential}(\lambda_\sigma), \quad \lambda_\sigma = -\frac{\log(p_\sigma)}{\sigma_0}, \\ &\log\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha), \\ &\tilde{z}_k \sim \mathcal{N}(0,1), \quad k = 1, \ldots, M. \end{aligned}$$

The exponential prior on σ_ε is a penalized complexity (PC) prior (Simpson et al. 2017) with the probability statement $\Pr(\sigma_\varepsilon > \sigma_0) = p_\sigma$. We set $\sigma_0 = 0.5$ and $p_\sigma = 0.05$. The lognormal prior on α is specified through the normal prior on $\log\alpha$ with mean $\mu_\alpha = 0$ and standard deviation $\sigma_\alpha = 1.0$, allowing flexibility while maintaining weak informativeness and enabling the data to determine the appropriate degree of regularization.

C.2.3 Hyperparameter initialization

To improve MCMC convergence, we initialize the regularization parameter α using a scaling based on the non-zero eigenvalues of the TPS penalty matrix. Let v_k denote the non-zero penalty eigenvalues, excluding the null-space components. We first define a reference value $\alpha_{\text{ref}} = 1$ and rescale it according to the median of the non-zero penalty eigenvalues:

$$\alpha_{\text{init}} = \frac{\alpha_{\text{ref}}}{\text{median}\left(v_k : v_k > 10^{-10}\right)}. \tag{A22}$$

This scaling accounts for the numerical scale of the penalty eigenvalues and provides a standardized initial value for α across different spatial configurations and basis dimensions. The resulting value is used to initialize the log-transformed parameter,

$$\log(\alpha)_{\text{init}} = \log(\alpha_{\text{init}}), \tag{A23}$$

with the initialization restricted to the interval $[-3, 3]$ to avoid excessively large or small starting values. Thus, a reference value $\alpha_{\text{ref}} = 1$ is used throughout the simulations. This initialization is used only to provide a suitable starting point for the MCMC algorithm; the posterior distribution of α is subsequently determined by the likelihood and the specified prior.

C.2.4 Joint posterior distribution

Given the observed values $y_i = u(\mathbf{s}_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, the joint posterior distribution is:

$$p(\tilde{\mathbf{z}}, \sigma_\varepsilon, \alpha \mid \mathbf{y}) \propto \left[\prod_{i=1}^{N} \mathcal{N}\left(y_i \mid (\mathbf{\Phi}_{\text{obs}}\mathbf{z})_i, \sigma_\varepsilon^2\right)\right] \prod_{k=1}^{M} \mathcal{N}\left(\tilde{z}_k \mid 0, 1\right) \pi(\sigma_\varepsilon)\, \pi(\alpha), \tag{A24}$$

where the likelihood depends on $\tilde{\mathbf{z}}$ through the transformation in Equation (A20), and $\mathbf{z}$ is constructed by applying the non-centered transformation to each component.

D R code for data simulation

To evaluate the SPDE and regTPS approaches under controlled conditions, we generate simulated spatial data considering a GRF with known Matérn covariance. We consider four scenarios with sample sizes $N \in \{50, 100, 150, 200\}$, where observation locations $\mathbf{s}_i$ are uniformly distributed on the unit square $[0, 1]^2$.

Listing 1 Simulation code for comparative study of SPDE and regTPS approaches

```
#================================
# Simulating the Data
#================================
base_N_sp <- 50
n_scenarios <- 4
dim_grid <- 30

sigma_u0 <- 1.0
sigma_e0 <- 0.3
nu <- 1.0
rho0 <- 0.3

# Matern covariance function
matern_cov <- function(coords, nu, rho, sigma2) {
  D <- as.matrix(dist(coords))
  D[D == 0] <- 1e-10
  scaling <- (sqrt(2 * nu) * D) / rho
  matern_part <- (2^(1 - nu)) / gamma(nu) * scaling^nu * besselK(scaling, nu)
  diag(matern_part) <- 1
  return(sigma2 * matern_part)
}

# Exponential covariance function
# exponential_cov <- function(coords, rho, sigma2) {
#  D <- as.matrix(dist(coords))
#  cov_mat <- sigma2 * exp(-D / rho)
#  return(cov_mat)
#}


#=====================================================
# Creating the lists for the SPDE/regTPS models
#=====================================================
fits_TMB_spde <- list()
fits_TMB_tps <- list()

for (i in 1:n_scenarios) {
  N_sp <- base_N_sp * i
  set.seed(1234 + i)

  # Create a common mesh and points for this scenario
  sp_points <- data.frame(s1 = runif(N_sp), s2 = runif(N_sp))
  sp_matrix <- as.matrix(sp_points)
  bound1 <- fmesher::fm_nonconvex_hull(sp_matrix)
  mesh <- fmesher::fm_rcdt_2d_inla(loc = sp_matrix,
          boundary = bound1, refine = FALSE, plot.delay = NULL)

  # Create a common grid for this scenario
  grid_total <- expand.grid(s1 = seq(0, 1, length.out = dim_grid),
                            s2 = seq(0, 1, length.out = dim_grid))

  # Simulate the TRUE latent field on the mesh nodes ONCE
  Cov_true <- matern_cov(mesh$loc, nu = nu, rho = rho0, sigma2 = sigma_u0^2)
  # Cov_true <- exponential_cov(mesh$loc, rho = rho, sigma2 = sigma_u^2)
  u_true <- as.numeric(MASS::mvrnorm(1, mu = rep(0, mesh$n), Sigma = Cov_true))

```

```r
  # Create projection matrices to get the true field at other locations
  A_obs_proj <- inla.spde.make.A(mesh = mesh, loc = sp_matrix)
  A_grid_proj <- inla.spde.make.A(mesh = mesh, loc = as.matrix(grid_total))

  # Project the true field to observation points and grid points
  u_true_sp <- as.numeric(A_obs_proj %*% u_true)
  u_true_grid <- as.numeric(A_grid_proj %*% u_true)

  # Add noise to the observations
  y_obs <- u_true_sp + rnorm(N_sp, 0, sigma_e0)

  #=====================================================================
  # Run SPDE model first
  obj_spde <- run_tmb_spde(N_sp, dim_grid, sp_points, mesh, y_obs, u_true_sp,
                                    u_true_grid, Cov_true)

  # Get number of mesh nodes from SPDE
  n_mesh_nodes <- mesh$n

  #====================================
  # COMPARISON STRATEGY FOR regTPS
  #====================================
  k_basis_max <- floor(0.99 * N_sp)     # Conservative: 99% of data points
  k_basis_max <- max(k_basis_max, 10)   # At least 10 basis functions

  # Run TPS model
  obj_tps <- run_tmb_tps(N_sp, dim_grid, sp_points, mesh, y_obs, u_true_sp,
                                    u_true_grid,
                         k_basis = k_basis_max,
                         Cov_true,
                         variance_threshold = 0.99,
                         alpha_ref = 1)

  fits_TMB_spde[[i]] <- obj_spde
  fits_TMB_tps[[i]] <- obj_tps
}


#====================================
# Saving all the TMB models
#====================================
fits_TMB_spde <- list(fits_TMB_spde[[1]], fits_TMB_spde[[2]],
                      fits_TMB_spde[[3]], fits_TMB_spde[[4]])
saveRDS(fits_TMB_spde, file = 'outputs/fits_TMB_spde.RDS')

fits_TMB_tps <- list(fits_TMB_tps[[1]], fits_TMB_tps[[2]],
                     fits_TMB_tps[[3]], fits_TMB_tps[[4]])
saveRDS(fits_TMB_tps, file = 'outputs/fits_TMB_tps.RDS')



#============================
# SPDE - MCMC models
#============================
M = list()
M[[1]] = list()
M[[1]]$model = "spatial_n50"
M[[2]] = list()
M[[2]]$model = "spatial_n100"
M[[3]] = list()
M[[3]]$model = "spatial_n150"
M[[4]] = list()
M[[4]]$model = "spatial_n200"

M[[1]]$formula = fits_TMB_spde[[1]]$obj
M[[2]]$formula = fits_TMB_spde[[2]]$obj
```

```
M[[3]]$formula = fits_TMB_spde[[3]]$obj
M[[4]]$formula = fits_TMB_spde[[4]]$obj




#=================
# Run the models
lwr <- c(1e-4, 1e-4, 1e-4)
upr <- c(10, 10, 10)

for (i in 1:length(M)){
  startTime <- Sys.time()
  print(paste("Running:  ", M[[i]]$model))
  fit <- tmbstan(M[[i]]$formula,
                 chains= 3, open_progress = FALSE,
                 control = list(max_treedepth= 12,  adapt_delta = 0.9),
                 iter = 3000, warmup= 700, cores=no_cores,
                 lower = lwr, upper = upr, seed = 1234)
                 # init = 'last.par.best', seed = 12345)
  endTime <- Sys.time()
  timeUsed = difftime(endTime, startTime, units='mins')
  print(timeUsed)
  saveRDS(fit, file=paste0('outputs/mcmc_spde_', i,'.RDS'))
}

#============================
# regTPS - MCMC models
#============================
M = list()
M[[1]] = list()
M[[1]]$model = "spatial_n50"
M[[2]] = list()
M[[2]]$model = "spatial_n100"
M[[3]] = list()
M[[3]]$model = "spatial_n150"
M[[4]] = list()
M[[4]]$model = "spatial_n200"

M[[1]]$formula = fits_TMB_tps[[1]]$obj
M[[2]]$formula = fits_TMB_tps[[2]]$obj
M[[3]]$formula = fits_TMB_tps[[3]]$obj
M[[4]]$formula = fits_TMB_tps[[4]]$obj

#=================
# Run the models
for (i in 1:length(M)){
  startTime <- Sys.time()
  print(paste("Running:  ", M[[i]]$model))
  fit <- tmbstan(M[[i]]$formula,
                 chains= 3, open_progress = FALSE,
                 control = list(max_treedepth= 12,  adapt_delta = 0.9),
                 iter = 3000, warmup= 700, cores=no_cores,
                 init = 'last.par.best', seed = 1234)
  endTime <- Sys.time()
  timeUsed = difftime(endTime, startTime, units='mins')
  print(timeUsed)
  saveRDS(fit, file=paste0('outputs/mcmc_tps_', i,'.RDS'))
}
```

E spNNGP for univariate Bayesian spatial regression models using nearest neighbor Gaussian processes

The following R code describes the setting used to create the spNNGP model presented in Sect. 5 (real data application).

Listing 2 spNNG model for real NO_2 application

```
# =====================
# Reading the NO2 data
# =====================
y_sqrt <- sqrt(sp_data$y_obs)
hist(y_sqrt)
coords <- as.matrix(sp_data[, c("s1", "s2")])

# Same prediction grid as SPDE and regTPS
grid_total  <- regTPS_KLE_tmb$grid_total
grid_matrix <- as.matrix(grid_total)

# ======================================================
# Starting values and priors for spNNGP
# ======================================================
sigma_sq_start <- var(y_sqrt) * 0.8
tau_sq_start   <- var(y_sqrt) * 0.2

phi_start <- 3 / (0.3 * diff(range(coords[, 1])))

starting <- list(phi = phi_start, sigma.sq = sigma_sq_start,
                 tau.sq = tau_sq_start, nu = 1.0)
tuning <- list(phi = 0.2, sigma.sq = 0.5, tau.sq = 0.5, nu = 0.01)
priors <- list(phi.Unif = c(3 / (2 * diff(range(coords[, 1]))), 3 / (0.01 * diff
    (range(coords[, 1])))),
               sigma.sq.IG = c(2, sigma_sq_start),
               tau.sq.IG   = c(2, tau_sq_start),
               nu.Unif     = c(0.1, 2.0))


# ======================================================
# Fit 3 independent chains
# ======================================================
n_chains  <- 3
n_samples <- 3000
spnngp_fits <- vector("list", n_chains)
spnngp_times <- numeric(n_chains)

for (chain in seq_len(n_chains)) {
  set.seed(1000 + chain)
  t0 <- Sys.time()
  spnngp_fits[[chain]] <- spNNGP(
    y_sqrt ~ 1,
    coords         = coords,
    method         = "latent",
    family         = "gaussian",
    n.neighbors    = 15,
    starting       = starting,
    tuning         = tuning,
    priors         = priors,
    cov.model      = "matern",
    n.samples      = n_samples,
    n.omp.threads  = no_cores,
    n.report       = 500)
  t1 <- Sys.time()
  spnngp_times[chain] <- as.numeric(difftime(t1, t0, units = "min"))
}

# ======================================================
```

```r
# Computing pointwise log-lik for loo-cv comparison
# ================================================
compute_spnngp_loglik <- function(spnngp_fit, X, y, burn_in = 0) {
  w_samples     <- spnngp_fit$p.w.samples
  beta_samples  <- spnngp_fit$p.beta.samples
  theta_samples <- spnngp_fit$p.theta.samples

  n_draws_total <- ncol(w_samples)
  keep_idx <- seq.int(from = min(burn_in + 1, n_draws_total),
                      to = n_draws_total)

  w_samples     <- w_samples[, keep_idx, drop = FALSE]
  beta_samples  <- beta_samples[keep_idx, , drop = FALSE]
  theta_samples <- theta_samples[keep_idx, , drop = FALSE]

  n_draws <- ncol(w_samples)
  n_obs   <- length(y)
   tau_sq_col <- grep("^tau\\.sq$|tau\\.sq", colnames(theta_samples))
  tau_sq_samples <- theta_samples[, tau_sq_col]
   log_lik <- matrix(NA_real_, nrow = n_draws, ncol = n_obs)
  for (d in seq_len(n_draws)) {
    mu_d <- as.numeric(X %*% beta_samples[d, ]) + w_samples[, d]
    sigma_d <- sqrt(tau_sq_samples[d])
    log_lik[d, ] <- dnorm(y, mean = mu_d, sd = sigma_d, log = TRUE)
  }
  log_lik
}

# ====================================
# Recompute with burn-in removed
# ====================================
burn_in <- 700
n_obs <- length(y_sqrt)
X_full <- matrix(1, nrow = n_obs, ncol = 1)
log_lik_chains <- vector("list", n_chains)

for (chain in seq_len(n_chains)) {
  log_lik_chains[[chain]] <- compute_spnngp_loglik(spnngp_fits[[chain]],
                                                   X = X_full, y = y_sqrt,
                                                   burn_in = burn_in)
}

log_lik_3_matrix <- rbind(log_lik_chains[[1]],
                          log_lik_chains[[2]],
                          log_lik_chains[[3]])

loo_3 <- loo(log_lik_3_matrix, cores = no_cores, r_eff = 1)
print(loo_3)
```

Acknowledgements Joaquin Cavieres was partially funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2075 – 390740016.

Author contributions A: Proposed and implemented the statistical methods through programming and conducted the major analyses for the simulation study and the real data application. B: Proposed and critically revised the theory of the manuscript and offered guidance for the simulation and real data application studies. Both authors participated in the conceptualization, writing, review, editing of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability All simulations and real data applications were developed in R software using Template Model Builder (TMB). The codes can be found at: https://github.com/jcavieresg/regTPS/tree/main.

Code availability All simulations and real data applications were developed in R software using Template Model Builder (`TMB`) and `tmbstan` packages. The codes can be found at: https://github.com/jcavieresg/regTPS/tree/main.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abrahamsen, P.: A review of Gaussian random fields and correlation functions (1997)

Adler, R.J.: The Geometry of Random Fields. John Wiley & Sons, Hoboken (1981)

Adler, R.J.: The Geometry of Random Fields. SIAM (2010)

Alexanderian, A.: A brief note on the Karhunen–Lo\eve expansion. arXiv:1509.07526 (2015)

Aronszajn, N., Smith, K.T.: Characterization of positive reproducing kernels. applications to Green's functions. Am. J. Math. **79**(3), 611–622 (1957)

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F.: Spatial modeling with R-INLA: a review. Wiley Interdiscip. Rev. Comput. Stat. **10**(6), e1443 (2018)

Banerjee, S., Carlin, B.P., Gelfand, A.E.: Hierarchical Modeling and Analysis for Spatial Data. CRC Press, Boca Raton (2014)

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H.: Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc. Ser. B Stat Methodol. **70**(4), 825–848 (2008)

Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer Science & Business Media (2011)

Betancourt, M., Girolami, M.: Hamiltonian Monte Carlo for hierarchical models. Curr. Trends Bayesian Methodol. Appl. **79**(30), 2–4 (2015)

Biegler, L.T., Ghattas, O., Heinkenschloss, M., van Bloemen Waanders, B.: Large-scale PDE-constrained optimization for uncertainty quantification. Numerical Analysis of Multiscale Problems, pp. 3–49 (2011)

Brenner, S., Scott, L.R., Zhang, W.: Karhunen–Loève expansion for Gaussian measures on Banach spaces. Math. Comput. **86**(307), 2179–2200 (2017)

Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Springer (2008)

Bungartz, H.-J., Griebel, M.: Sparse grids. Acta Numer **13**, 147–269 (2004)

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: a probabilistic programming language. J. Stat. Softw. **76**(1), 1–32 (2017)

Cavieres, J., Ibacache-Pulgar, G., Contreras-Reyes, J.E.: Thin plate spline model under skew-normal random errors: estimation and diagnostic analysis for spatial data. J. Stat. Comput. Simul. **93**(1), 25–45 (2023)

Cavieres, J., Monnahan, C.C., Bolin, D., Bergherr, E.: Approximated gaussian random field under different parameterizations for MCMC. In International Workshop on Statistical Modelling, pp. 204–210. Springer (2024)

Cotter, S., Dashti, M., Stuart, A.M.: An approximation theory for function? Valued Bayesian inverse problems. SIAM J. Numer. Anal. **48**(1), 322–345 (2010)

Courant, R., Hilbert, D.: Methods of Mathematical Physics, vol. 2. John Wiley & Sons, Hoboken (2024)

Cressie, N.: Geostatistics. Am. Stat. **43**(4), 197–202 (1989)

Duchon, J.: Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Constructive Theory of Functions of Several Variables, pp. 85–100. Springer (1977)

Duffy, D.G.: Green's Functions with Applications. Chapman and Hall/CRC (2015)

Engl, H. W., Ramlau, R.: Regularization of inverse problems. In *Encyclopedia of applied and computational mathematics*, pages 1233–1241. Springer (2015)

Evans, L. C.: *Partial differential equations*, volume 19. American mathematical society (2022)

Fasshauer, G.E.: Meshfree approximation methods with MATLAB, vol. 6. World Scientific (2007)

Fasshauer, G.E., McCourt, M.J.: Kernel-Based Approximation Methods Using Matlab, volume 19. World Scientific Publishing Company (2015)

Finley, A.O., Datta, A., Banerjee, S.: spNNGP r package for nearest neighbor Gaussian process models. J. Stat. Softw. **103**, 1–40 (2022)

Gelfand, A.E.: Hierarchical modeling for spatial data problems. Spatial Stat. **1**, 30–39 (2012)

Ghanem, R., Higdon, D., Owhadi, H.: Handbook of Uncertainty Quantification (2017)

Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Springer (1991)

Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Courier Corporation (2003)

Gneiting, T., Kleiber, W., Schlather, M.: Matérn cross-covariance functions for multivariate random fields. J. Am. Stat. Assoc. **105**(491), 1167–1177 (2010)

Green, P.J.: Penalized likelihood for general semi-parametric regression models. International Statistical Review/Revue Internationale de Statistique 245–259 (1987)

Green, P.J., Silverman, B.W.: Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. CRC Press (1993)
Handcock, M.S., Stein, M.L.: A Bayesian analysis of kriging. Technometrics **35**(4), 403–410 (1993)
Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., et al.: A case study competition among methods for analyzing large spatial data. J. Agric. Biol. Environ. Stat. **24**, 398–425 (2019)
Higdon, D., Gattiker, J., Williams, B., Rightley, M.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (2008)
Hilbert, D.: Methods of mathematical physics. In: CUP Archive (1985)
Katzfuss, M., Guinness, J.: A general framework for Vecchia approximations of Gaussian processes (2021)
Kimeldorf, G.S., Wahba, G.: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Stat. **41**(2), 495–502 (1970)
Kolmogorov, A.N., Fomin, S.V.: Introductory Real Analysis. Courier Corporation (1975)
Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.: TMB: automatic differentiation and laplace approximation (2015). arXiv:1509.00660 arXiv preprint
Le Maître, O.P., Knio, O.M.: Spectral stochastic finite element methods for stochastic partial differential equations. Comput. Methods Appl. Mech. Eng. **194**(12–16), 1645–1660 (2010)
Lindgren, F., Bolin, D., Rue, H.: The SPDE approach for gaussian and non-Gaussian fields: 10 years and still running. Spatial Stat. **50**, 100599 (2022)
Lindgren, F., Lindström, J., Rue, H.: An explicit link between Gaussian fields and Gaussian Markov random fields (2010)
Lindgren, F., Rue, H.: Bayesian spatial modelling with r-inla. J. Stat. Softw. **63**, 1–25 (2015)
Lindgren, F., Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **73**(4), 423–498 (2011)
Litvinenko, A., Matthies, H.G., Schwab, C.: Karhunen–Loève expansions of random fields beyond the classical assumptions. SIAM/ASA J. Uncertain. Quant. **10**(1), 237–263 (2022)
Lord, G.J., Powell, C.E., Shardlow, T.: An Introduction to Computational Stochastic PDEs, vol. 50. Cambridge University Press, Cambridge (2014)
Loève, M.: Probability Theory II. Springer, Berlin (1978)
Lu, M., Cavieres, J., Moraga, P.: A comparison of spatial and nonspatial methods in statistical modeling of no 2: prediction accuracy, uncertainty quantification, and model interpretation. Geogr. Anal. **55**(4), 703–727 (2023)
Lu, M., Schmitz, O., de Hoogh, K., Kai, Q., Karssenberg, D.: Evaluation of different methods and data sources to optimise modelling of no2 at a global scale. Environ. Int. **142**, 105856 (2020)
Marzouk, Y.M., Najm, H.N., Rahn, L.A.: Stochastic spectral methods for efficient Bayesian solution of inverse problems. J. Comput. Phys. **224**(2), 560–586 (2009)
Matérn, B.: Spatial variation. stochastic models and their application to some problems in forest surveys and other sampling investigations (1960)
Meinguet, J.: Multivariate interpolation at arbitrary points made simple. Zeitschrift für angewandte Mathematik und Physik ZAMP **30**(2), 292–304 (1979)
Mercer, J.: Xvi. Functions of positive and negative type, and their connection the theory of integral equations. Philos. Trans. R. Soc. Lond. Ser. A **209**(441-458):415–446 (1909)
Møller, J.: Spatial Statistics and Computational Methods, volume 173. Springer Science & Business Media (2013)
Monnahan, C.C., Kristensen, K.: No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. PLoS ONE **13**(5), e0197954 (2018)
Nychka, D.: Bayesian confidence intervals for smoothing splines. J. Am. Stat. Assoc. **83**(404), 1134–1143 (1988)
Papaspiliopoulos, O., Roberts, G.O., Sköld, M.: A general framework for the parametrization of hierarchical models. Stat. Sci. 59–73 (2007)
Reed, M., Simon, B.: Methods of Modern Mathematical Physics: Functional Analysis, volume 1. Gulf Professional Publishing (1980)
Renardy, M., Rogers, R.C.: An Introduction to Partial Differential Equations. Springer, Berlin (2004)
Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. CRC Press, Boca Raton (2005)
Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H.: Penalising model component complexity: a principled (2017)
Smolyak, S.: Quadrature and interpolation formulas for tensor products of certain classes of functions. Sov. Math. Dokl. **4**, 240–243 (1963)
Snelson, E., Ghahramani, Z.: Local and global sparse Gaussian process approximations. In: Artificial Intelligence and Statistics, pp. 524–531. PMLR (2007)
Stein, M.L.: Interpolation of Spatial Data: Some Theory for Kriging. Springer Science & Business Media
Stein, M.L.: Limitations on low rank approximations for covariance matrices of spatial data. Spatial Stat. **8**, 1–19 (2014)
Stein, M.L., Chi, Z., Welty, L.J.: Approximating likelihoods for large spatial data sets. J. R. Stat. Soc. Ser. B Stat Methodol. **66**(2), 275–296 (2004)
Steinwart, I., Scovel, C.: Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSS. Constr. Approx. **35**(3), 363–417 (2012)
Team, S.D.: Stan modeling language users guide and reference manual, version 2 (2018)
Titsias, M.: Variational learning of inducing variables in sparse Gaussian processes. In: Artificial Intelligence and Statistics, pp. 567–574. PMLR (2009)
Trudinger, N.S.: Elliptic partial differential equations of second order (1983)
Uribe, F., Papaioannou, I., Betz, W., Straub, D.: Bayesian inference of random fields represented with the Karhunen–Loève expansion. Comput. Methods Appl. Mech. Eng. **358**, 112632 (2020)
van Lieshout, M.N.M.: Theory of Spatial Statistics: A Concise Introduction. Chapman and Hall/CRC (2019)
Vecchia, A.V.: Estimation and model identification for continuous spatial processes. J. R. Stat. Soc. Ser. B Stat Methodol. **50**(2), 297–312 (1988)
Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. **27**, 1413–1432 (2017)
Wahba, G.: Smoothing noisy data with spline functions. Numer. Math. **24**(5), 383–393 (1975)
Wahba, G.: Bayesian "confidence intervals" for the cross-validated smoothing spline. J. Roy. Stat. Soc.: Ser. B (Methodol.) **45**(1), 133–150 (1983)
Wahba, G.: Spline Models for Observational Data, volume 59. SIAM (1990)
Wendland, H.: Scattered Data Approximation, vol. 17. Cambridge University Press, Cambridge (2004)
Wendland, H.: Computational aspects of radial basis function approximation. In Studies in Computational Mathematics, volume 12, pp. 231–256. Elsevier (2006)
White, G.: Bayesian semiparametric spatial and joint spatio-temporal modeling. University of Missouri-Columbia (2006)
Whittle, P:. On stationary processes in the plane. Biometrika 434–449 (1954)

Whittle, P.: Stochastic-processes in several dimensions. Bull. Int. Stat. Inst. **40**(2), 974–994 (1963)
Wiener, N.: The homogeneous chaos. Am. J. Math. **60**(4), 897–936 (1938)
Wikle, C.K.: Low-rank representations for spatial processes. Handb. Spatial Stat. **107**, 118 (2010)
Williams, C.K., Rasmussen, C.E.: Gaussian Processes for Machine Learning, vol. 2. MIT Press, Cambridge (2006)
Wood, S.N.: Thin plate regression splines. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **65**(1), 95–114 (2003)
Wood, S.N.: Generalized Additive Models: An Introduction with R. CRC Press, Boca Raton (2017)
Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Am. Stat. Assoc. **99**(465), 250–261 (2004)
Zhu, H., Ghosh, D.: Karhunen–Loève expansion for random fields. Am. Stat. **51**(3), 209–213 (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.