Efficient spatial modelling using the SPDE approach with bivariate splines

Xiaoyu Liu
Department of Statistical Science, University College London
Serge Guillas
Department of Statistical Science, University College London
and
Ming-Jun Lai
Department of Mathematics, University of Georgia

October 13, 2015

Abstract

Gaussian fields (GFs) are frequently used in spatial statistics for their versatility. The associated computational cost can be a bottleneck, especially in realistic applications. It has been shown that computational efficiency can be gained by doing the computations using Gaussian Markov random fields (GMRFs) as the GFs can be seen as weak solutions to corresponding stochastic partial differential equations (SPDEs) using piecewise linear finite elements. We introduce a new class of representations of GFs with bivariate splines instead of finite elements. This allows an easier implementation of piecewise polynomial representations of various degrees. It leads to GMRFs that can be inferred efficiently and can be easily extended to non-stationary fields. The solutions approximated with higher order bivariate splines converge faster, hence the computational cost can be alleviated. Numerical simulations using both real and simulated data also demonstrate that our framework increases the flexibility and efficiency. Supplementary materials are available online.

Keywords: Gaussian Markov random field; Mapping; Multivariate splines; Non-stationary spatial process; Spatial approximation.

1 Introduction

Gaussian fields (GFs) are at the core of spatial statistics, especially in the class of structured additive regression models, named latent Gaussian models, which are flexible and extensively used (Cressie, 1993; Banerjee et al., 2004; Diggle and Ribeiro, 2007). However, when making statistical inference, it is usually needed to evaluate the likelihood function or the latent Gaussian field distribution, for which we need to make computations on dense matrices, e.g. the covariance matrix $\mathbf{\Sigma}(\boldsymbol{\theta})$, typically of order $\mathcal{O}(n^3)$ where n is the dimension of $\mathbf{\Sigma}(\boldsymbol{\theta})$. Rue et al. (2009) overcome this computational hurdle. They approximate Bayesian inference in latent Gaussian models by assuming that the latent field is Gaussian Markov random field (GMRF). With only a few hyperparameters, integrated nested Laplace approximations (INLA) produce faster inference than simulation based approaches such as MCMC. To take advantage of the computational efficiency of GMRF, Lindgren et al. (2011) constructed an explicit link between GFs and GMRFs. They considered the GFs with Matérn covariance function,

$$r(\mathbf{u}, \mathbf{v}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\mathbf{v}-\mathbf{u}\|)^{\nu}K_{\nu}(\kappa\|\mathbf{v}-\mathbf{u}\|), \tag{1}$$

where $\|\mathbf{v}-\mathbf{u}\|$ is the Euclidean distance between two locations $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^D$, K_ν is the modified Bessel function of the second kind and order $\nu > 0$, $\kappa > 0$ controls the nominal correlation range through $\rho = \sqrt{8\nu}/\kappa$ corresponding to correlations near 0.1 at the Euclidean distance ρ, and σ^2 is the marginal variance. The integer value of ν determines the mean-square differentiability of the underlying process. They noticed that a Gaussian field $x(\mathbf{u})$ with Matérn covariance (1) is a solution to the linear fractional SPDE

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\mathbf{u})) = W(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^D, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0, \tag{2}$$

where the innovation process W is spatial Gaussian white noise with unit variance (Whittle, 1954, 1963), $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian operator, and τ controls the marginal variance through the relationship

$$\tau^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\sigma^2}.$$

Denoting the inner product of two functions f and g on $\mathbb{R}^D$ as $\langle f, g\rangle = \int_{\mathbb{R}^D} f(\mathbf{u})g(\mathbf{u})\,\mathrm{d}\,\mathbf{u}$,

we consider the stochastic weak formulation of SPDE (2)

$$\left\{\langle\phi_t, (\kappa^2 - \Delta)^{\alpha/2}\tau x\rangle, t = 1, ..., n_t\right\} \stackrel{\mathrm{d}}{=} \left\{\langle\phi_t, W\rangle, t = 1, ..., n_t\right\}, \tag{3}$$

for every finite set of suitable test functions $\{\phi_t(\mathbf{u}), t = 1, ..., n_t\}$, where '$\stackrel{\mathrm{d}}{=}$' denotes equality in distribution (Walsh, 1986). Lindgren et al. (2011) constructed a finite element representation (Brenner and Scott, 2008) of the Gaussian random field over an unstructured triangulation of the form

$$x_h(\mathbf{u}) = \sum_{k=1}^{n} w_k \psi_k(\mathbf{u}), \tag{4}$$

where $\{\psi_k\}_{k=1}^n$ are piecewise linear basis functions. By requiring (3) to hold for only a specific set of test functions, they showed that the Gaussian weights $\{w_k\}_{k=1}^n$ are GMRFs when $\alpha = 1$, and can be approximated with GMRFs when $\alpha \geq 2$. Therefore the inference for GFs can be carried out using GMRFs and the computational efficiency can be improved dramatically. This work is closely related to the spatial spline regression models by Sangalli et al. (2013) where a spatial surface is approximated with finite elements. Another recent related work is Nychka et al. (2015), where the authors proposed a representation of a random field using multi-resolution radial basis functions on a regular grid. They also assumed that the coefficients associated with the basis functions to be distributed according to a GMRF to speed up the computation.

It is stated in Lindgren et al. (2011) and Simpson et al. (2012) that the convergence rate of a finite element approximation to the full solution to the SPDE (2) is of order $\mathcal{O}(h^2)$ where h is the length of longest edge in the triangulation. Hence the convergence can be achieved by refining the underlying triangulation which is usually called the h-version finite elements. An alternative is to increase the approximation order over any fixed triangulation with higher degree polynomials over each triangle, which is called the p-version finite elements (the degree of polynomials is usually denoted by p). It has been illustrated that the convergence rate of the p-version cannot be worse than the h-version in most cases (Babuska et al., 1981). To do so, multivariate splines over triangulations can be employed instead of conventional finite elements. This provides a flexible and easy construction of splines with piecewise polynomials of various degrees and smoothness. Basic concepts and theories of multivariate splines can be found in the monograph by Lai and

Schumaker (2007). Multivariate splines have been shown to be more efficient and flexible than conventional finite element method in data fitting problems and solving PDEs, see Awanou et al. (2006). It has been applied in spatial statistics. For example, Guillas and Lai (2010) introduced a spatial data analysis model with bivariate splines by penalizing the roughness with a partial differential operator; this has been demonstrated to be more efficient and accurate than thin-place splines in the application of ozone concentration forecasting (Ettinger et al., 2012). In this paper, we introduce bivariate splines to represent the GFs on $\mathbb{R}^2$ and show its advantages over the piecewise linear finite elements in Lindgren et al. (2011). Within our framework of the SPDE approach using bivariate splines, it is allowed to choose piecewise polynomial representations of arbitrary degrees to adapt to the various data structures and features, and make the inference more computationally efficient.

The paper is structured as follows. In Section 2 some basics of bivariate splines in the Bernstein form (B-form) are reviewed first. Then we show how to link the GFs with GMRFs within the framework of bivariate splines, establish the theoretical properties of the bivariate spline approximations and discuss extensions to non-stationary fields. In Section 3, we conduct several numerical simulations to illustrate our method and compare with the approach of Lindgren et al. (2011) on both real and simulated data sets. Section 4 consists of conclusion and discussion. Proofs are in the Appendix in supplementary materials.

2 SPDE approach using bivariate splines

2.1 B-form bivariate splines

Let $\mathbf{\Delta}$ be a triangulation of a bounded domain $\Omega \subset \mathbb{R}^2$. We consider the continuous spline spaces

$$S_d^0(\mathbf{\Delta}) = \{s \in C^0(\Omega), s|_T \in \mathcal{P}_d, \forall T \in \mathbf{\Delta}\},$$

where $\mathcal{P}_d$ is the space of bivariate polynomials of degree $d \geq 1$, $C^0(\Omega)$ is the space of all continuous functions on Ω. For any $d \geq 1$, the spline space $S_d^0(\mathbf{\Delta})$ contains all possible continuous spline functions which are bivariate polynomials of degree d over each triangle $T \in \mathbf{\Delta}$. We apply the B-form representation of splines in $S_d^0(\mathbf{\Delta})$ proposed by Awanou et al.

(2006) in this paper. We only give a brief introduction to the bivariate splines here and some details are relegated to Appendix A.1. For more complete and in-depth explanations, see Lai and Schumaker (2007).

Let $T = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$ be a non-degenerate (i.e. with non-zero area) triangle with vertices $\mathbf{v}_1 = (x_1, y_1)$, $\mathbf{v}_2 = (x_2, y_2)$ and $\mathbf{v}_3 = (x_3, y_3)$. Then every point $\mathbf{v} = (x, y) \in \mathbb{R}^2$ has a unique representation in the form

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3, \tag{5}$$

with $b_1 + b_2 + b_3 = 1$, where b_1, b_2 and b_3 are named the barycentric coordinates of the point $\mathbf{v} = (x, y)$ relative to the triangle T. The polynomials

$$B_{ijk}^{T,d}(\mathbf{v}) = \frac{d}{i!j!k!} b_1^i b_2^j b_3^k, \ i + j + k = d, \tag{6}$$

are called the Bernstein polynomials of degree d relative to triangle T. Then for each spline function $s \in S_d^0(\mathbf{\Delta})$, we can write

$$s|_T = \sum_{i+j+k=d} c_{ijk}^T B_{ijk}^{T,d}, \quad T \in \mathbf{\Delta},$$

where the coefficients $\mathbf{c} = \{c_{ijk}^T, i + j + k = d, T \in \mathbf{\Delta}\}$ are called B-coefficients of s. Note that linear finite elements are typical splines in $S_1^0(\mathbf{\Delta})$.

2.2 SPDE modelling with B-form bivariate splines

Let $\{\psi_1, \psi_2, ..., \psi_m\}$ be a set of locally supported basis functions of $S_d^0(\mathbf{\Delta})$ for any $d \geq 1$, where $m = \dim S_d^0(\mathbf{\Delta})$ (see Appendix A.1 and Lai and Schumaker (2007) for more details on these basis functions) on a triangulation of a bounded domain $\Omega \subset \mathbb{R}^2$. For any $h = 1, ..., m$, the corresponding B-coefficients of ψ_h are denoted by $\mathbf{c}_h$. The results in this paper hold for any triangulations. But the quality of approximations depend on the triangulation properties. In practice, we suggest the Delaunay triangulations that are chosen to maximize the minimum interior triangle angle following Lindgren et al. (2011).

Then we can construct a bivariate spline representation of the solution to SPDE (2) in the spline space $S_d^0(\mathbf{\Delta})$ as

$$x_{\mathbf{\Delta}}(\mathbf{u}) = \sum_{h=1}^{m} w_h \psi_h(\mathbf{u}). \tag{7}$$

Following Lindgren et al. (2011) we approximate a weak solution to the SPDE with respect to the spline space $S_d^0(\mathbf{\Delta})$ by finding the distribution of the weights $\{w_h, h = 1, ..., m\}$ that fulfils the stochastic weak formulation (3) for only a specific set of test functions such that the integrals at both sides of (3) exist. The distribution of the approximate solution $x_{\mathbf{\Delta}}(\mathbf{u})$ can be obtained through the stochastic weights. Specifically, we choose $\phi_h = (\kappa^2 - \Delta)^{1/2}\psi_h$ for $\alpha = 1$ leading to the least squares solution. For $\alpha = 2$, we can choose either $\phi_h = \psi_h$ for any $d \geq 1$ or $\phi_h = (\kappa^2 - \Delta)\psi_h$ for $d \geq 2$, leading to the Galerkin or least squares solution respectively. For $\alpha \geq 3$, if we let $\alpha = 2$ on the left-hand side of the SPDE (2) then the right-hand side is a Gaussian process generated by the operator $(\kappa^2 - \Delta)^{(\alpha-2)/2}$. Then we can choose $\phi_h = \psi_h$ for this innovative SPDE. Hence we get a recursive Galerkin solutions ending with either $\alpha = 1$ or 2. We also assume appropriate boundary conditions to avoid the solutions in the null space of the differential operator. Throughout this paper, the Neumann condition (zero normal derivative at the boundary) is imposed. Then we have the main results as below.

Theorem 1 *The vector of weights* $\mathbf{w} = (w_h, h = 1, ..., m)'$ *of bivariate spline representation* (7) *is Gaussian with mean zero and the precision matrix* $\mathbf{Q}_\alpha$ *that are given as follows:*

(1) for $\alpha = 1$,

$$\mathbf{Q}_1 = \tau^2(\kappa^2\mathbf{M} + \mathbf{K}),$$

(2) for $\alpha = 2$,

$$\mathbf{Q}_2^G = \tau^2(\kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{K}\mathbf{M}^{-1}\mathbf{K}),$$

$$\mathbf{Q}_2^{LS} = \tau^2(\kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{R}),$$

where $\mathbf{Q}_2^G$ *and* $\mathbf{Q}_2^{LS}$ *are the Galerkin and least squares solutions respectively,*

(3) for $\alpha \geq 3$,

$$\mathbf{Q}_\alpha = \kappa^4\mathbf{Q}_{\alpha-2} + \kappa^2(\mathbf{Q}_{\alpha-2}\mathbf{M}^{-1}\mathbf{K} + \mathbf{K}\mathbf{M}^{-1}\mathbf{Q}_{\alpha-2}) + \mathbf{K}\mathbf{M}^{-1}\mathbf{Q}_{\alpha-2}\mathbf{M}^{-1}\mathbf{K},$$

where

$$\mathbf{M} = \mathbf{C}'\mathbf{M_0}\mathbf{C}, \quad \mathbf{K} = \mathbf{C}'\mathbf{K_0}\mathbf{C}, \quad \mathbf{R} = \mathbf{C}'\mathbf{R_0}\mathbf{C},$$

and $\mathbf{M_0} = \text{diag}(\mathbf{M}_T, T \in \boldsymbol{\Delta})$, $\mathbf{K_0} = \text{diag}(\mathbf{K}_T, T \in \boldsymbol{\Delta})$ *and* $\mathbf{R_0} = \text{diag}(\mathbf{R}_T, T \in \boldsymbol{\Delta})$ *are block diagonal square matrix with square blocks*

$$\mathbf{M}_T = \left[\int_T B^T_{ijk}(x,y)B^T_{i'j'k'}(x,y)dxdy\right]^{i'+j'+k'=d}_{i+j+k=d},$$

$$\mathbf{K}_T = \left[\int_T \nabla B^T_{ijk}(x,y)\nabla B^T_{i'j'k'}(x,y)dxdy\right]^{i'+j'+k'=d}_{i+j+k=d},$$

and

$$\mathbf{R}_T = \left[\int_T \Delta B^T_{ijk}(x,y)\Delta B^T_{i'j'k'}(x,y)dxdy\right]^{i'+j'+k'=d}_{i+j+k=d},$$

respectively and $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_m)$ *whose h-th column is the B-coefficient vector of basis function* ψ_h.

Since the basis functions $\{\psi_1, \psi_2, ..., \psi_m\}$ are locally supported in $S^0_d(\boldsymbol{\Delta})$, the matrices $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{R}$ are guaranteed to be sparse. However in the Galerkin solution for $\alpha = 2$ and recursive solutions for $\alpha \geq 3$, the inverse matrix $\mathbf{M}^{-1}$ is not necessarily sparse, making the precision matrix dense. The mass lumping technique (Chen and Thomée, 1985) can be applied by replacing $\mathbf{M}$ with a diagonal matrix $\tilde{\mathbf{M}}$ whose elements are the sum of each row of $\mathbf{M}$, i.e. $\tilde{\mathbf{M}}_{ii} = \sum_j \mathbf{M}_{ij}$. The same technique is also deployed by Lindgren et al. (2011) and they discussed the implications in detail in their Appendix C.5 showing that the convergence rate would not be affected. We will discuss the properties of the mass lumping approximation in our case later. Therefore the precision matrix is sparse and the underlying coefficients $\mathbf{w}$ are approximated with a GMRF.

2.3 Approximation properties

Define the Hilbert space H^1 associated with the differential operator $(\kappa^2 - \Delta)$ to be the space of square integrable functions $f(x, y)$ for which $\|f\|^2_{H^1} = \kappa^2 \int_\Omega f(x,y)^2 dxdy + \int_\Omega \nabla f(x,y) \cdot \nabla f(x,y) dxdy$ is finite following Lindgren et al. (2011). Approximation results for bivariate splines, e.g. Th. 5.19 in Lai and Schumaker (2007), show that the bivariate spline space $S^0_d(\boldsymbol{\Delta})$ for any $d \geq 1$ spanned by a finite set of basis functions $\{\psi_1, ..., \psi_m\}$ is dense in H^1: for every $f \in H^1$, there is a sequence $\{f_m\}$, $f_m \in S^0_d(\boldsymbol{\Delta})$ such that $\lim_{m\to\infty} \|f - f_m\|_{H^1} = 0$ where the limit scenario $m \to \infty$ corresponds to $|\boldsymbol{\Delta}| \to 0$ where $|\boldsymbol{\Delta}|$ is the length of the

longest edge in the triangulation $\mathbf{\Delta}$. Using this fact, it follows directly from the Th.3-4 in Appendix C.2 of Lindgren et al. (2011) that, the bivariate spline approximation $x_{\mathbf{\Delta}}$ converges weakly to the weak solution to the SPDE. Note that the weak convergence of $x_{\mathbf{\Delta}}$ obtained for $\mathbf{Q}_2^{LS}$ cannot be derived directly but can be easily proved in the same fashion with just a few modifications. In addition, we can derive rates of convergence results. We first define the associated Sobolev space on Ω in $\mathbb{R}^2$ for any $1 \leq q \leq \infty$ and $d \geq 1$ as

$$W_q^d(\Omega) = \{f : \|f\|_{d,q,\Omega} < \infty\},$$

where

$$\|f\|_{d,q,\Omega} = \begin{cases} \left(\sum_{k=0}^{d} |f|_{k,q,\Omega}^q \right)^{1/q}, & 1 \leq q < \infty \\ \sum_{k=0}^{d} |f|_{k,\infty,\Omega}, & q = \infty, \end{cases}$$

with

$$|f|_{k,q,\Omega} = \begin{cases} \left(\sum_{\nu+\mu=k} \|D_x^\nu D_y^\mu f\|_{q,\Omega}^q \right)^{1/q}, & 1 \leq q < \infty \\ \max_{\nu+\mu=k} \|D_x^\nu D_y^\mu f\|_{\infty,\Omega}, & q = \infty, \end{cases}$$

and

$$\|f\|_{q,\Omega} = \begin{cases} \left(\int_\Omega |f(u)du|^q \right)^{1/q}, & 1 \leq q < \infty, \\ \operatorname{ess\,sup}_{u \in \Omega} |f(u)|, & q = \infty. \end{cases}$$

Then we have the proposition below regarding to the Galerkin solutions when $\alpha = 2$.

Proposition 1 *Let $L = (\kappa^2 - \Delta)$, $x_{\mathbf{\Delta}}(s)$ is the bivariate spline approximation of the random Gaussian field $x(s)$ in the spline space $S_d^0(\mathbf{\Delta})$, $d \geq 1$. Then for any $f \in H^1 \cap W_2^{m+1}(\Omega)$ with $1 \leq m \leq d$, we have*

$$\mathrm{E}\left(\int_\Omega f(s) L(x(s) - x_{\mathbf{\Delta}}(s)) ds \right)^2 \leq K |\mathbf{\Delta}|^{m+1} |f|_{m+1,2,\Omega}$$

where K is a constant, $|\mathbf{\Delta}|$ is the length of the longest triangle edge in the triangulation $\mathbf{\Delta}$.

It is clear that we are able to achieve a faster convergence rate by using bivariate splines with higher degree d. For example, when $d = 3$ the convergence rate can be as high as $\mathcal{O}(|\mathbf{\Delta}|^4)$, which is two magnitude higher than $\mathcal{O}(|\mathbf{\Delta}|^2)$ in Lindgren et al. (2011).

As we have mentioned, the matrix $\mathbf{M}$ in the Galerkin solutions is lumped by replacing $\mathbf{M}$ with a diagonal matrix $\tilde{\mathbf{M}}$ which yields a Markov approximation $\tilde{x}_{\boldsymbol{\Delta}}$ to the bivariate spline solution $x_{\boldsymbol{\Delta}}$. Let f and g be test functions in H^1 and let $f_{\boldsymbol{\Delta}}$ and $g_{\boldsymbol{\Delta}}$ be their projections onto the bivariate spline space $S_d^0(\boldsymbol{\Delta})$ for any $d \geq 1$, with basis weights $\mathbf{w}_f$ and $\mathbf{w}_g$. Since the recursive algorithm for $\alpha \geq 3$ is based on $\alpha = 2$ at each iteration, here we only investigate the effect of the Markov approximation on the Galerkin solutions for $\alpha = 2$. When $\alpha = 2$, the difference between the covariances for the Markov approximation $\tilde{x}_{\boldsymbol{\Delta}}$ and the bivariate spline solution $x_{\boldsymbol{\Delta}}$ is

$$\epsilon_{\boldsymbol{\Delta}}(f_{\boldsymbol{\Delta}}, g_{\boldsymbol{\Delta}}) = \mathrm{cov}(\langle f, L\tilde{x}_{\boldsymbol{\Delta}}\rangle_\Omega, \langle g, L\tilde{x}_{\boldsymbol{\Delta}}\rangle_\Omega) - \mathrm{cov}(\langle f, Lx_{\boldsymbol{\Delta}}\rangle_\Omega, \langle g, Lx_{\boldsymbol{\Delta}}\rangle_\Omega) = \mathbf{w}_f'\tilde{\mathbf{M}}\mathbf{w}_g - \mathbf{w}_f'\mathbf{M}\mathbf{w}_g.$$

We have the following result showing that such a difference can be bounded.

Proposition 2 *For $f_{\boldsymbol{\Delta}}$, $g_{\boldsymbol{\Delta}} \in S_d^0(\boldsymbol{\Delta})$, we have*

$$|\epsilon_{\boldsymbol{\Delta}}(f_{\boldsymbol{\Delta}}, g_{\boldsymbol{\Delta}})| \leq K|\boldsymbol{\Delta}|^2$$

where K is a positive constant dependent on $\|f_{\boldsymbol{\Delta}}\|_{2,2,\Omega}, \|g_{\boldsymbol{\Delta}}\|_{2,2,\Omega}, \|f_{\boldsymbol{\Delta}}\|_{\infty,\Omega}, \|g_{\boldsymbol{\Delta}}\|_{\infty,\Omega}$ and $|\boldsymbol{\Delta}|$ is the length of the longest triangle edge in the triangulation $\boldsymbol{\Delta}$.

We can see that the mass lumping error is at most of order $\mathcal{O}(|\boldsymbol{\Delta}|^2)$. We are not sure whether this is the lowest bound in theory. Thus the overall approximation error of bivariate splines representations to the SPDE solutions is at most $\mathcal{O}(|\boldsymbol{\Delta}|^2)$, which is the same as Lindgren et al. (2011).

These results broadly reach an agreement with the numerical simulations in Bolin and Lindgren (2013). The authors have shown that the higher order splines are more efficient in covariance approximation but become less efficient using mass lumping approximation. It seems to suggest higher order bivariate splines may be less helpful to the convergence. However for practical applications, we want to make some points clear here. Firstly, the actual approximation errors in both propositions also depend on the unknown constants K. Secondly, as stated in Bolin and Lindgren (2013), parameter inference is also very important to the accuracy and efficiency in application, which could be affected by using different representations. Last but not least, higher order splines may lose efficiency in approximating the true Matérn field with mass lumping, but not necessarily in approximating

the spatial field. Thus efficiency can still be improved in practice by using higher order bivariate splines. The numerical simulations later illustrate that higher order bivariate splines are more efficient in spatial prediction.

2.4 Non-stationary fields

Lindgren et al. (2011) showed that the SPDE (2) can be extended to a non-stationary version

$$(\kappa^2(\mathbf{u}) - \Delta)^{\alpha/2}(\tau(\mathbf{u})x(\mathbf{u})) = W(\mathbf{u}), \tag{8}$$

where the parameters κ^2 and τ are not constants but depend on the location $\mathbf{u}$. The two parameters are assumed to vary slowly and have the general form of low dimensional representations

$$\log(\kappa^2(\mathbf{u})) = \sum_{j=1}^{n_{\kappa^2}} \theta_j^{(\kappa^2)} B_j^{(\kappa^2)}(\mathbf{u}), \quad \log(\tau(\mathbf{u})) = \sum_{j=1}^{n_\tau} \theta_j^{(\tau)} B_j^{(\tau)}(\mathbf{u}),$$

where the number of smooth basis functions n_{κ^2} and n_τ should not be large to guarantee computational efficiency. The inner product can be approximated with $\langle \psi_t, \kappa^2 \psi_s \rangle \approx \kappa^2(\mathbf{u}_s^\star)\langle \psi_t, \psi_s \rangle$, where $\mathbf{u}_s^\star$ is some point in the support of ψ_s which can be chosen to be the domain point associated with the non-zero B-coefficients of ψ_s (see Appendix A.1 for more details about domain points and their relationship to ψ_s). Defining the diagonal matrices

$$\boldsymbol{\kappa}^2 = \text{diag}(\kappa^2(\xi_h), h = 1, ..., m), \quad \boldsymbol{\tau} = \text{diag}(\tau(\xi_h), h = 1, ..., m),$$

where ξ_h is the domain point associated with the non-zero B-coefficients of basis function ψ_h for $h = 1, ..., m$. It can be easily shown with minor modification to the proof of Theorem 1 that the weights $\mathbf{w}$ in the bivariate spline representation (7) can be approximated with GMRF as well. For example when $\alpha = 2$, the precision matrix of $\mathbf{w}$ is $\mathbf{Q}_2^G(\boldsymbol{\kappa}^2, \boldsymbol{\tau}) = \boldsymbol{\tau}(\boldsymbol{\kappa}^2\mathbf{M}\boldsymbol{\kappa}^2 + 2\boldsymbol{\kappa}^2\mathbf{K} + \mathbf{K}\mathbf{M}^{-1}\mathbf{K})\boldsymbol{\tau}$ or $\mathbf{Q}_2^{LS}(\boldsymbol{\kappa}^2\boldsymbol{\tau}) = \boldsymbol{\tau}(\boldsymbol{\kappa}^2\mathbf{M}\boldsymbol{\kappa}^2 + 2\boldsymbol{\kappa}^2\mathbf{K} + \mathbf{R})\boldsymbol{\tau}$ for Galerkin or least squares solutions respectively. As stated in Lindgren et al. (2011), by assuming the parameters κ^2 and τ to be constant locally, the solution to the SPDE (8) can still be interpreted as a Matérn field over a local area and the associated global non-stationary field can be achieved by combining all the local Matérn fields automatically via the SPDE.

3 Numerical simulations

We conduct several numerical simulations to evaluate the performance of the SPDE approach with bivariate splines and compare with the linear finite element approach in Lindgren et al. (2011) in terms of spatial prediction. In all simulations over $\mathbb{R}^2$ we fix $\alpha = 2$ which corresponds to the smoothness parameter $\nu = 1$ in the Matérn covariance function. The Bayesian inference for the model is run in `R-inla (www.r-inla.org)` using the integrated nested Laplace approximation (Rue et al., 2009). For brevity, our proposed bivariate spline approximation in $S_d^0(\mathbf{\Delta})$ is denoted BS-SPDE with $d = 1, 2, ...$ (BS-SPDE-G or BS-SPDE-LS for Galerkin or least squares solution respectively) and the linear finite element approximation is denoted LFE-SPDE.

3.1 Comparison of LFE-SPDE and BS-SPDE

Study 1

In this simulation, we compare the LFE-SPDE method and BS-SPDE of degree $d \geq 2$ in data fitting for some common surfaces. Elevations of different surfaces are collected on a grid over square $[-2, 2] \times [-2, 2]$ that is equally spaced every 0.2. Then we make predictions on another finer grid that is equally spaced every 0.01 over square $[-2, 2] \times [-2, 2]$ using the SPDE approach. The prediction accuracy for the whole surface can be measured with mean squared error MSE $= \sum_{i=1}^{n} (\hat{f}(\mathbf{u}_i) - f(\mathbf{u}_i))^2/n$, where $f(\mathbf{u}_i)$ is the true elevation on location $\mathbf{u}_i$ and $\hat{f}(\mathbf{u}_i)$ is the prediction using corresponding posterior means.

Four different surfaces are considered: $2\sin(x)\cos(y)$ and $2\exp(-\frac{x^2+y^2}{s})$ with three different shape parameters $s = 2$, 1, 0.5. We construct 35 different meshes that have 2, 3, 6, 13, 19, 28, 53, 96, 112, 148, 212, 279, 342, 390, 444, 520, 705, 874, 1065, 1368, 1802, 2416, 2798, 3176, 3708, 4428, 5514, 6696, 8460, 10958, 15009, 21832, 26718, 33776, 43875 triangles respectively to demonstrate the convergence (mesh size $|\mathbf{\Delta}|$ monotonically decreases roughly from 6.6 to 0.026). For each approach, the associated number of basis functions (denoted by N_B) and CPU time for calling `inla` (denoted by T_{cpu} in seconds) are recorded when the corresponding MSEs reach levels of 10^{-l}, $l = 1, 2, ..., 8$. N_B is also the dimension of corresponding precision matrix of the weights $\mathbf{w}$ and directly relates to

the computational complexity. For example the samples and likelihoods can be computed in $\mathcal{O}(N_B^{3/2})$ operations for two dimensional GMRFs. For comparison, the simulation stops when the number of basis functions of BS-SPDE with $d \geq 2$ exceeds the number of basis functions of LFE-SPDE using the densest mesh. The results are presented in Figure 1 where the y-axes for N_B and T_{cpu} are taken on a logarithmic scale.

From Figure 1 we can see that in general BS-SPDE with $d \geq 2$ can be more efficient than LFE-SPDE both in terms of number of basis functions and computing time needed to reach specific levels of MSE, especially those lower than 10^{-4}. In the left side of Figure 1, the dash lines for BS-SPDE-LS are invisible as they coincide with the solid lines for BS-SPDE-G, while BS-SPDE-LS is more computationally efficient in general than BS-SPDE-G as shown in the right side. Specifically, for the surface $2\sin(x)\cos(y)$, to reach high precision levels such as 10^{-6} or 10^{-7}, BS-SPDE-G and BS-SPDE-LS with high degree $d \geq 3$ are more efficient since they require only less than 10% of the basis functions and computing time required by LFE-SPDE. For the Gaussian surface $2\exp(-\frac{x^2+y^2}{2})$, BS-SPDE with $d \geq 2$ are generally much more efficient than LFE-SPDE for the MSE levels up to 10^{-7} with about 50% gains in the computing time. But BS-SPDE-LS does not reach the MSE level 10^{-8} and BS-SPDE-G with $d = 3$ takes more computing time than the others to reach the MSE level 10^{-8}. For the next Gaussian shape surface $2\exp(-\frac{x^2+y^2}{1})$ which is steeper than the previous one, BS-SPDE with high degrees can be better than LFE-SPDE for the MSE levels around 10^{-4} to 10^{-6} but their efficiency is decreased to reach the higher precision levels 10^{-7} and 10^{-8}. However, BS-SPDE-LS with $d = 4$ reaches the low MSE level 10^{-7} within only 20% of the computing time required by LEF-SPDE. For the last surface which is quite steep, BS-SPDE-G with $d = 2$ is comparable with LFE-SPDE and reaches the high precision levels 10^{-6}, 10^{-7} by requiring slightly less number of basis functions and similar time, and BS-SPDE-LS with $d = 4$ is more efficient to reach the MSE level 10^{-6}.

From these results, we can conclude that BS-SPDE can be much more efficient in many cases especially when the high precision levels are desired and the target functions are smooth. For functions that are not that smooth, lower degree representations such as LFE-SPDE or BS-SPDE with $d = 2$ might be more appropriate, which is consistent with the general comments by Babuska et al. (1981). Note that even for the last Gaussian

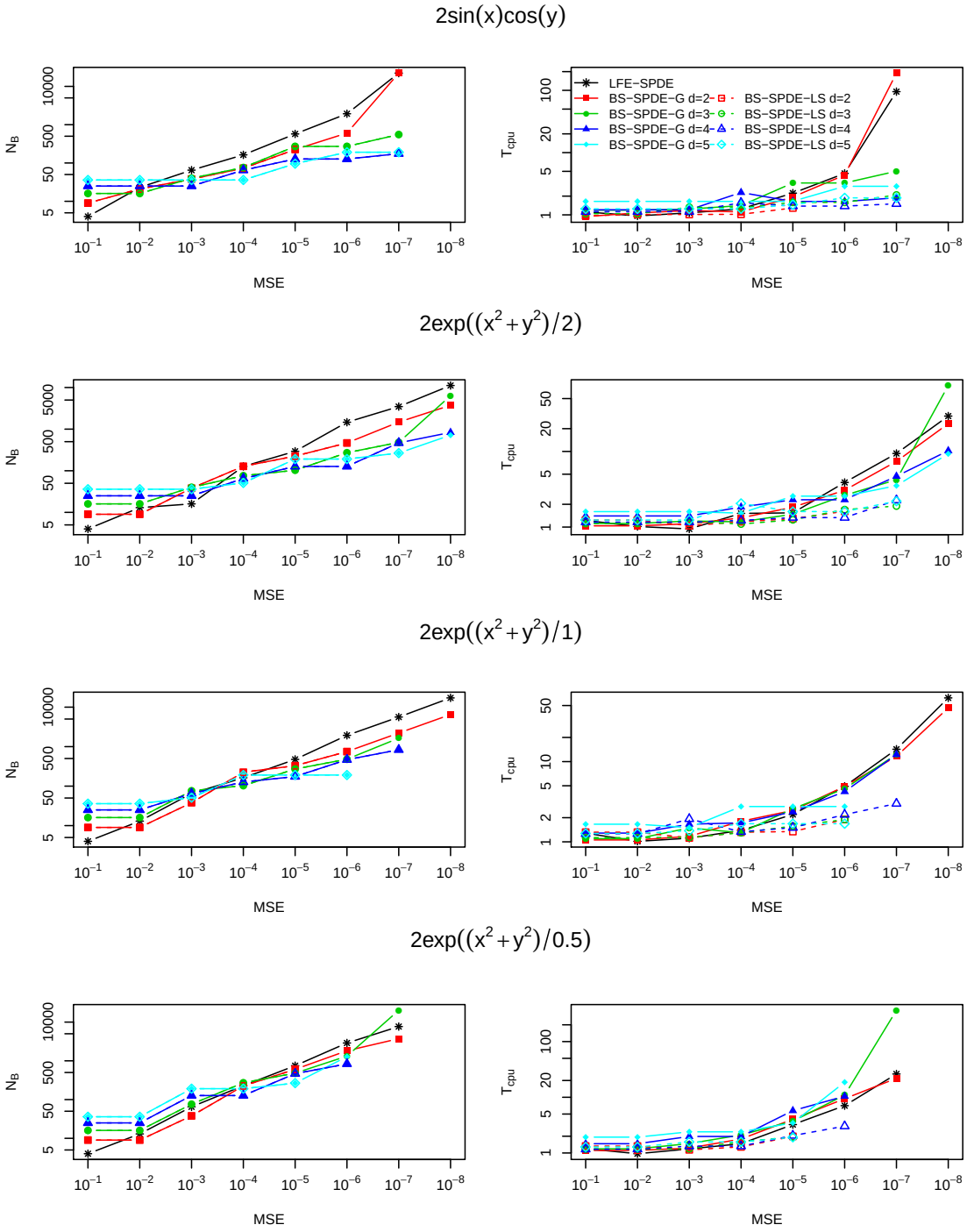

Figure 1: Number of basis functions (N_B) and CPU time for `inla` (T_{cpu} in seconds) required by LFE-SPDE, BS-SPDE-G and BS-SPDE-LS with $d = 2, 3, 4, 5$ respectively to reach specific MSE levels for different surfaces

shape surface which is much less smooth than the others, BS-SPDE-G with $d = 2$ still can be comparable with LFE-SPDE; and we obtain 50% gains in the computing time using BS-SPDE-LS with $d = 4$ if the MSE level 10^{-6} is desired.

Study 2

In this study, we compare LFE-SPDE and BS-SPDE in spatial estimation and prediction with real data sets that are extracted from the ETOPO1 Global Relief Model (Amante and Eakins, 2009), which is a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry. The data is available from National Geophysical Data Center (NGDC), USA. Four different regions around the the Strait of Juan de Fuca area are chosen for this study as shown in Figure 2. In general, region 1 covers near shore seabed with relatively simple and gradual variations while the seabed in the other three regions is quite complicated.

Figure 2: Four regions extracted from ETOPO 1 around the Strait of Juan de Fuca

For comparison, both in-sample and out-of-sample predictive fit performance are explored using LFE-SPDE, BS-SPDE-G and BS-SPDE-LS with $d = 2, 3, 4$ based on various meshes. We denote the observations by y_1, y_2, ..., y_n. As for the in-sample fit measurement, root mean square error (RMSE) between the observations and the predictions at the

observed locations

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

are calculated where the predictions $\hat{y}_i$ are taken to be the associated posterior mean. Since the SPDE approach aims to estimate the whole surface, smaller RMSE suggests the estimated surface is closer to the measurements at the observed locations. To measure the predictive performance, leave-one-out cross validation is employed using the embedded function within `R-inla`. The logarithmic score (Log Score) of prediction is defined as

$$\text{Log Score} = -\frac{1}{n}\sum_{i=1}^{n}\log\left[\pi(y_i|y_{-i})\right],$$

where $\pi(y_i|y_{-i})$ is the posterior predictive density of y_i given all the other observations y_{-i}. Therefore the smaller Log Score is, the more certain we are with the predictions. Furthermore six meshes are built as shown in Figure 3. The meshes are extended with coarse triangles to avoid boundary effect (Lindgren et al., 2011). The number of basis functions for each combination of mesh and SPDE method is shown in Table 1.

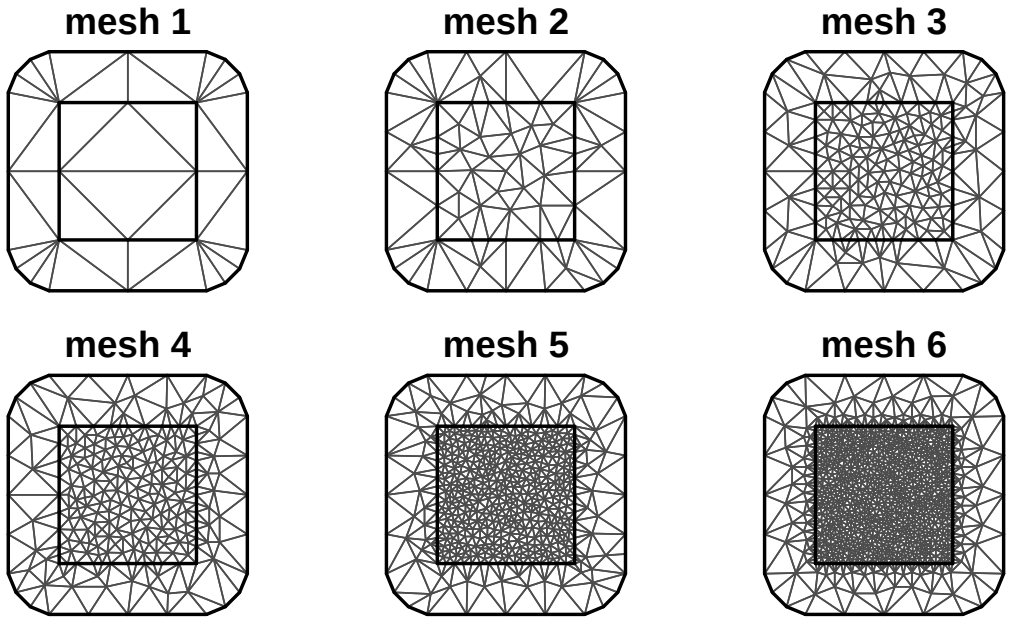

Figure 3: Six meshes for Study 2

Table 2 presents the RMSEs and Log Scores using LFE-SPDE, BS-SPDE-G and BS-SPDE-LS with $d = 2, 3, 4$ based on the six meshes respectively. In general, as the tri-

Table 1: Number of basis functions using LFE-SPDE and BS-SPDE-G/LS with $d = 2, 3, 4$ based on six meshes

mesh	LFE-SPDE	BS-SPDE $d = 2$	BS-SPDE $d = 3$	BS-SPDE $d = 4$
1	28	89	184	313
2	70	252	547	955
3	195	749	1663	2937
4	244	945	2104	3721
5	536	2111	4726	8381
6	978	3881	8710	15465

angulation becomes denser, the estimations and predictions are more accurate using both LFE-SPDE and BS-SPDE-G/LS in most cases. For a particular mesh, the RMSEs and Log Scores of BS-SPDE-G/LS with $d \geq 2$ are generally smaller than those of LFE-SPDE. In terms of the number of basis functions, BS-SPDE-G/LS with $d \geq 2$ also demonstrate better performance than LFE-SPDE in most cases. For example, for region 4, BS-SPDE-G with $d = 2$ based on mesh 4 and BS-SPDE-G with $d = 4$ based on mesh 2 yield much smaller RMSEs and Log Scores than LFE-SPDE based on mesh 6 while they have similar numbers of basis functions. In general, BS-SPDE-LS yields smaller RMSEs than BS-SPDE-G in most cases. However, in terms of Log Score, BS-SPDE-G performs better than BS-SPDE-LS in most cases for the other three regions except region 1. We notice the sudden change in the model performance. For example, the RMSE obtained using LFE-SPDE for region 2 is about 15.87 based on mesh 4; it is decreased suddenly to only 0.037 based on mesh 5 or 0.012 using BS-SPDE-G with $d = 2$. This may because the finite elements or splines reach some level of degree of freedom that is enough to model the surface well.

Based on Table 2, we can select the models with good performance for continuous map reconstruction among the different combinations of meshes and SPDE approaches. In most cases it is difficult to have a model with the smallest RMSE and Log Score at the same time, so we only choose the one with relatively small RMSE and Log Score. In this way, the reconstructed map can be close to the elevations at the observed locations; meanwhile we are more confident with the predictions at the other locations. As marked with asterisks in Table 2, we choose BS-SPDE-LS with $d = 4$ based on mesh 3 for region 1, BS-SPDE-LS

Table 2: Study 2: RMSE and Log Score (RMSE|Log Score) using LFE-SPDE, BS-SPDE-G and BS-SPDE-LS with $d = 2, 3, 4$. The values in red colors are obtained based on similar number of basis functions between 945 to 978. The values marked with asterisks ($*$) are the selected model fit for map reconstruction

mesh	LFE-SPDE	BS-SPDE-G			BS-SPDE-LS		
		$d = 2$	$d = 3$	$d = 4$	$d = 2$	$d = 3$	$d = 4$
region 1							
1	1.85|2.07	1.41|1.84	1.18|1.73	1.11|1.73	1.41|1.84	1.20|1.72	1.16|1.71
2	1.17|1.71	0.76|1.53	0.55|1.47	0.29|1.21	0.88|1.65	0.81|1.70	0.74|1.73
3	0.83|2.01	0.46|1.48	0.33|1.42	0.16|1.09	0.43|1.64	0.048|1.18	0.027|0.98$*$
4	0.79|1.61	0.46|1.48	0.34|1.43	0.17|1.16	0.23|1.43	0.039|1.08	0.028|0.99
5	0.62|1.51	0.49|1.50	0.42|1.45	0.36|1.38	0.024|1.51	0.022|1.30	0.021|1.08
6	0.63|1.52	0.50|1.51	0.49|1.48	0.44|1.53	0.021|1.74	0.020|1.19	0.019|0.92
region 2							
1	79.81|5.82	66.01|5.66	52.26|5.41	47.01|5.30	66.41|5.65	52.21|5.42	45.81|5.27
2	36.20|5.02	21.27|4.53	9.09|3.86	0.02|0.50	22.47|4.58	14.52|4.23	12.48|4.10
3	17.24|4.34	0.013|0.51	0.0091|0.43	0.0088|0.45	0.015|0.43	0.011|0.41$*$	0.0094|0.50
4	15.87|4.28	0.012|0.52	0.010|0.46	0.010|0.50	0.013|0.47	0.0098|0.46	0.0091|0.54
5	0.037|1.85	0.013|0.85	0.010|0.55	0.010|0.48	0.0086|0.61	0.0069|0.55	0.0065|0.57
6	0.032|1.98	0.011|0.72	0.010|0.55	0.012|0.53	0.0075|0.71	0.0062|0.64	0.0073|0.81
region 3							
1	41.88|5.18	37.52|5.09	30.83|4.89	25.44|4.69	37.52|5.10	30.51|4.89	25.21|4.68
2	26.64|4.72	11.62|3.93	6.06|3.43	0.021|0.45	11.70|3.94	7.83|3.59	0.029|0.52
3	11.45|3.95	0.028|0.74	0.016|0.50	0.015|0.48$*$	0.022|0.58	0.021|0.56	0.017|0.61
4	10.18|3.86	0.025|0.72	0.017|0.53	0.015|0.48	0.021|0.55	0.019|0.60	0.016|0.59
5	7.21|3.66	0.021|1.01	0.018|0.64	0.016|0.50	0.014|0.69	0.012|0.68	0.012|0.72
6	0.053|2.08	0.019|0.99	0.019|0.80	0.019|0.64	0.012|0.82	0.011|0.79	0.012|0.96
region 4							
1	47.12|5.30	35.98|5.07	28.46|4.83	22.95|4.61	36.02|5.07	28.46|4.83	22.73|4.59
2	26.36|4.71	12.56|4.02	8.08|3.71	0.020|0.48	12.47|4.01	9.24|3.78	0.033|0.55
3	13.85|4.13	0.023|0.62	0.016|0.52	0.013|0.43$*$	0.019|0.55	0.016|0.54	0.014|0.56
4	12.42|4.04	0.021|0.76	0.017|0.54	0.014|0.46	0.019|0.54	0.015|0.56	0.014|0.59
5	0.061|2.19	0.019|0.93	0.017|0.64	0.015|0.49	0.012|0.63	0.011|0.63	0.010|0.68
6	0.049|2.14	0.017|0.93	0.018|0.78	0.017|0.58	0.011|0.77	0.010|0.77	0.010|0.88

with $d = 3$ based on mesh 3 for region 2, BS-SPDE-G with $d = 4$ based on mesh 3 for region 3, and BS-SPDE-G with $d = 4$ based on mesh 3 for region 4. Note that for region 1, BS-SPDE-LS with $d = 4$ based on mesh 6 yields both smallest RMSE and Log Score among all the combinations. However the associated computational cost is much heavier than the others. There is some trade off between model performance and computational cost. Hence we select the one with relatively good performance and is also computationally efficient. Then the posterior means and standard deviations of the four regions predicted using the respective selected models are displayed in Figure 4. The posterior means in general capture the main features of the corresponding regions and the posterior standard deviations provide uncertainty estimates of the predictions. Note that the selection rule of predictive model here is quite simple and subjective. More appropriate model selection techniques can be employed in application.

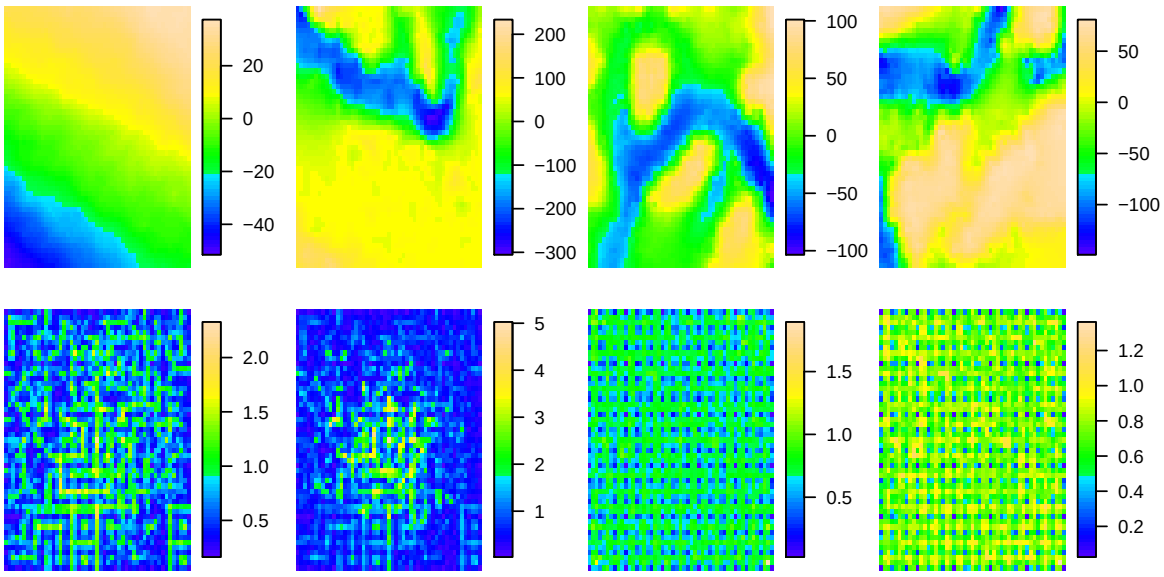

Figure 4: Posterior mean (top) and standard deviation (bottom) for the regions 1-4 (left to right), after selection of the appropriate approximation models

3.2 Spatial analysis of ozone levels data over Eastern USA

In this section, we analyse a data set of ozone levels at a certain hour in one of days in September, 2005 around the Eastern United States, which is available from the Air Explorer

Database of Environmental Protection Agency (EPA), using the non-stationary BS-SPDE-G method. The data set has 546 locations where ozone levels are recorded. As shown in Figure 5, the observations of ozone concentration are distributed unevenly and the domain is irregular. Denote the ozone levels by z_i and the associated locations by $\mathbf{s}_i = (x_i, y_i)$, $i = 1, ..., 546$. We consider a simple spatial model

$$z_i \sim b_0 + f(\mathbf{s}_i), \quad i = 1, ..., 546,$$

where b_0 is the intercept and the spatial effect $f(\mathbf{s}_i)$ is assumed to be a non-stationary GF generated by the non-stationary version SPDE (8), represented with bivariate splines in $S_d^0(\mathbf{\Delta})$ with $d = 1, 2, 3, 4, 5$. The triangulation $\mathbf{\Delta}$ is shown in Figure 5.

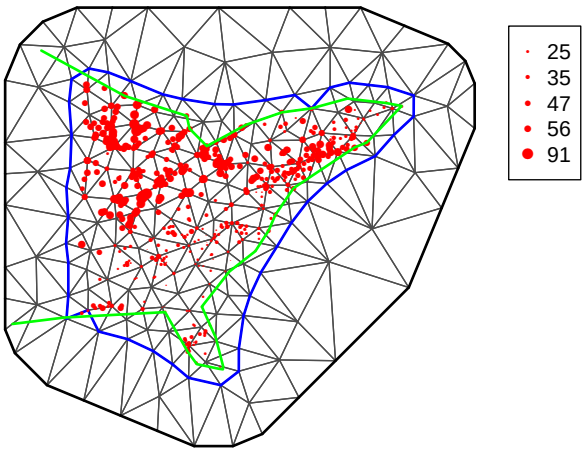

Figure 5: Triangulation over Eastern United States; green line: U.S. boundary; red dots: locations of ozone monitoring stations; size proportional to the ozone levels in ppb (parts per billion)

The non-stationary parameters $\tau(\mathbf{u})$ and $\kappa^2(\mathbf{u})$ are represented with two-dimensional B-splines that have n_x and n_y basis functions in the x-direction and y-direction respectively. Therefore at any location $\mathbf{s} = (x, y)$, the basis functions of the associated B-spline can be calculated as $B_{lk}(\mathbf{s}) = B_l^x(x)B_k^y(y)$, where $B_l^x(\cdot)$ and $B_k^y(\cdot)$ are the basis functions in x and y directions respectively, for $l = 1, ..., n_x$ and $k = 1, ..., n_y$. Hence there are $n_x n_y$ basis functions in total for each of the parameters $\kappa^2(\cdot)$ and $\tau(\cdot)$. We consider 12 models A-L with different combinations of the number of basis functions in Table 3. Note that with one basis function, the B-spline is constant so model A corresponds to the stationary SPDE

model (2). The number of basis functions represents the number of basis functions in both x-direction and y-direction; for example in model C, there are 3 basis functions for $\tau(\cdot)$ in each direction which means there are actually $3 \times 3 = 9$ basis functions for $\tau(\cdot)$.

Table 3: Number of basis functions for the B-spline in each direction for the parameters

	A	B	C	D	E	F	G	H	I	J	K	L
$\kappa^2(\cdot)$	1	1	1	1	1	2	3	4	5	2	3	4
$\tau(\cdot)$	1	2	3	4	5	1	1	1	1	2	3	4

To measure the fit and predictive performance and select the appropriate representations for $\kappa^2(\cdot)$ and $\tau(\cdot)$, we employ the leave-one-out cross validation and aim to find the model with the smallest Log Score. Figure 6 presents the Log Scores of the 12 models for the two parameters $\kappa^2(\cdot)$ and $\tau(\cdot)$ as shown in Table 3 using BS-SPDE-G approach with $d = 1, 2, 3, 4, 5$ respectively.

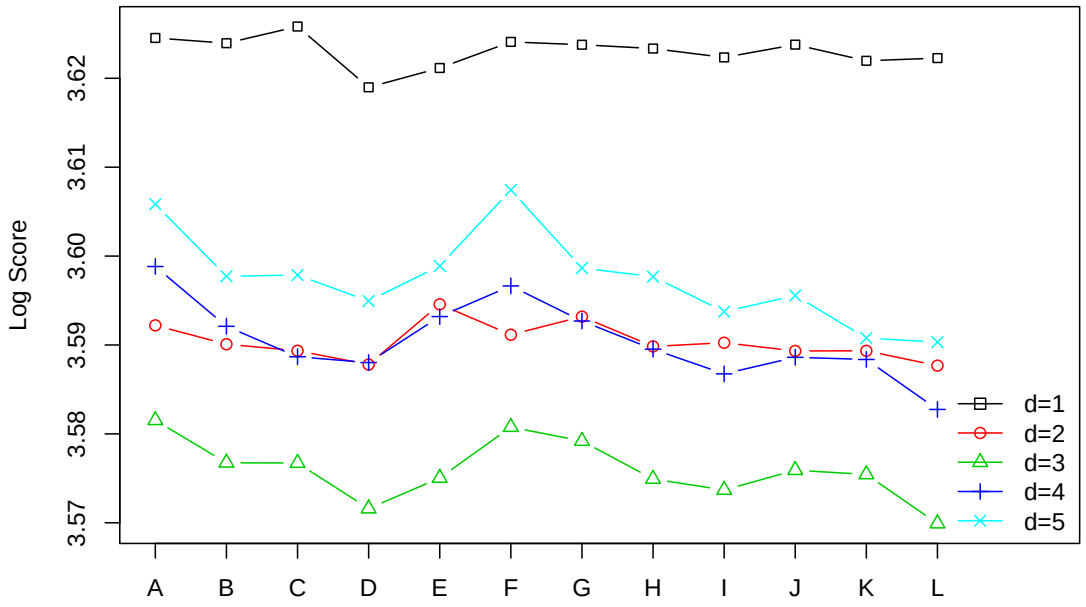

Figure 6: Log Scores for the models A-L using BS-SPDE-G with $d = 1, ..., 5$

It is easy to see that the Log Scores obtained from BS-SPDE-G with higher d are generally smaller than those obtained from BS-SPDE-G with lower d. Using BS-SPDE-G with a specific d, the Log Scores for different representations of $\kappa^2(\cdot)$ and $\tau(\cdot)$ are different.

In general the non-stationary models B-L yield smaller Log Scores than the stationary model A. The overall smallest Log Score is obtained with model L and BS-SPDE-G $d = 3$. As shown in Table 3, model L corresponds to 4 basis functions for $\kappa^2(\cdot)$ and 4 basis functions for $\tau(\cdot)$ in both x and y direction. This suggests that both $\kappa^2(\cdot)$ and $\tau(\cdot)$ display spatial variation over the domain. The number of parameters may be considered for model selection. In fact, we notice that the Log Score obtained using BS-SPDE-G $d = 3$ with model D is only slightly higher than model L while the number of parameters used to represent the non-stationarity is only half of model L. A proper model selection technique would account for it, e.g. AIC and BIC, but this is beyond the scope of this paper.

Then we apply the non-stationary model L and predict ozone levels using the BS-SPDE-G approach with $d = 3$. Figure 7 displays the posterior mean and standard deviation of the predictions given the observations presented in Figure 5. As we can see, the predicted ozone level is low in the south-east corner and at the top of the north-east corner and high in the north and middle area, which is consistent with the observations. Furthermore, the posterior predictive standard deviation shows some spatial variation over the entire domain because of the irregular distribution of the observations and the non-stationarity of the SPDE model.

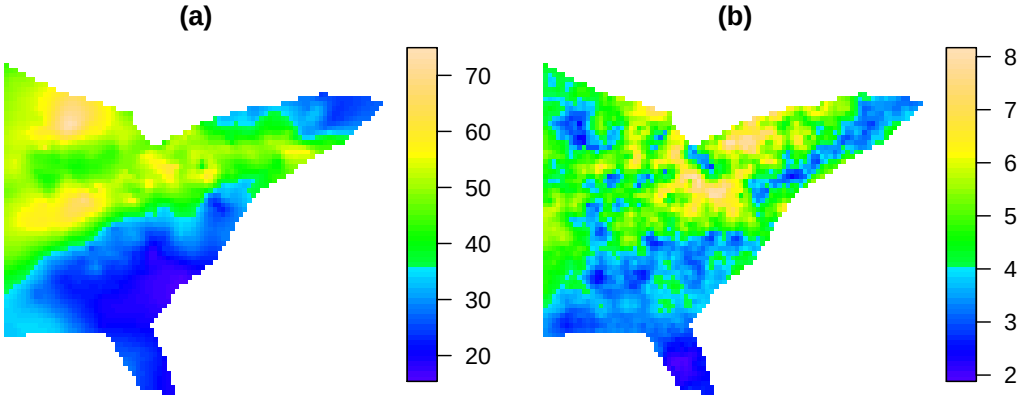

Figure 7: (a) Posterior mean; (b) posterior standard deviation: ozone levels over Eastern United States predicted using BS-SPDE-G with $d = 3$ and non-stationary model L

4 Discussion

We have shown that higher order polynomial basis can be easily implemented in the SPDE framework for GFs using bivariate splines. Both the theoretical results and numerical simulations have demonstrated the advantages of this new approach over the linear finite element approach in terms of approximation accuracy and computational efficiency. By using higher degree representations, we can also implement the least squares solutions to the SPDE (2) for $\alpha = 2$. This is more computationally efficient than the corresponding Galerkin solutions due to the sparser structures. We have shown that the SPDE approach can be applied to the spatial modelling of bathymetry. The current commonly used mapping tools, e.g. Generic Mapping Tools (GMT) used by NOAA (Eakins and Taylor, 2010), do not include uncertainty estimates of the maps. GMT also requires high smoothness conditions, see Smith and Wessel (1990) and Wessel and Bercovici (1998), that may not be appropriate for the bathymetry/topography. Hence the computationally efficient SPDE approach has promising potential in spatial mapping.

There is still room for further investigation in the SPDE approach with bivariate splines. It has been suggested by the numerical simulations that the degree of polynomial basis has an impact on the performance of the SPDE approach. Thus it is essential to choose appropriate degrees. In fact, the degree of bivariate splines can be adaptive. Hu et al. (2007) proposed a new spline method which allows automatic degree raising over triangles of interest. This new method is able to solve linear PDEs very effectively and efficiently. Another extension to manifolds could be considered. Lai et al. (2009) discussed the application of spherical splines in geopotential approximation where the techniques of triangulated spherical splines can be applied to represent the Matérn fields on manifolds. Furthermore, when α is larger than 2, which means the smoothness parameter ν in (1) increases as well, sample paths of the Matérn fields are smoother (Paciorek and Schervish, 2004). In this case, smoother representations of the GFs are desired. However, it is quite difficult to implement higher orders of smoothness in conventional finite elements. But within the bivariate splines framework, higher orders of smoothness conditions can be implemented easily by imposing linear constraints on the B-coefficients (Lai and Schumaker, 2007). However the implementation within the SPDE framework is non-trivial and needs to be investigated as

the large number of linear constraints are computationally expensive.

5 Supplementary materials

Appendix: More details about bivariate splines relevant to this paper and the proofs for Theorem 1 and Propositions 1 - 2. (pdf file)

Code and data for simulations: R code and data sets for the numerical simulations. (zip file)

6 Acknowledgements

We are grateful to the Editor, the Associate Editor and the two referees for their helpful comments and suggestions. Xiaoyu Liu acknowledges the financial support by the China Scholarship Council. Ming-Jun Lai would like to thank the Simons Foundation for supporting his research in a Simons collaboration grant 2013-2018.

A Appendix

We include some details about bivariate splines relevant to this paper. Then the sketch of proofs for Theorem 1 and Propositions 1 - 2 are provided.

A.1 Preliminaries on bivariate splines

To evaluate a polynomial of degree d in B-form over any triangle, say $p = \sum_{i+j+k=d} c_{ijk} B^d_{ijk}$, at the point $\mathbf{v} = (x, y)$ whose barycentric coordinates are $b = (b_1, b_2, b_3)$ with $b_1+b_2+b_3 = 1$, let $c^{(0)}_{ijk} = c_{ijk}$ and for all $l = 1, ..., d$,

$$c^{(l)}_{ijk} = b_1 c^{(l-1)}_{i+1,j,k} + b_2 c^{(l-1)}_{i,j+1,k} + b_3 c^{(l-1)}_{i,j,k+1}.$$

For $i + j + k = d - l$, we have

$$p(\mathbf{v}) = \sum_{i+j+k=d-l} c^{(l)}_{ijk} B^{d-l}_{ijk}(\mathbf{v}),$$

for all $0 \leq l \leq d$. In particular, $p(\mathbf{v}) = c_{000}^{(d)}$. This is called the de Casteljau algorithm (Lai and Schumaker, 2007).

Each vector $\mathbf{u}$ can be uniquely described by a triple (a_1, a_2, a_3) called directional coordinates of $\mathbf{u}$, that is $a_i = \alpha_i - \beta_i$, $i = 1, 2, 3$, where $(\alpha_1, \alpha_2, \alpha_3)$ and $(\beta_1, \beta_2, \beta_3)$ are the barycentric coordinates of two points $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ such that $\mathbf{u} = \boldsymbol{\omega} - \tilde{\boldsymbol{\omega}}$. It is easy to see that the barycentric coordinates of a point sum to 1, while the directional coordinates of a vector sum to 0. Suppose $\mathbf{u}$ is a vector in $\mathbb{R}^2$ whose directional coordinates are $a = (a_1, a_2, a_3)$, then for $i+j+k = d$, we define the directional derivative of B_{ijk}^d at location $\mathbf{v}$ with respect to directional vector $\mathbf{u}$ to be

$$D_u B_{ijk}^d(\mathbf{v}) = d\left[a_1 B_{i-1,j,k}^{d-1}(\mathbf{v}) + a_2 B_{i,j-1,k}^{d-1}(\mathbf{v}) + a_3 B_{i,j,k-1}^{d-1}(\mathbf{v})\right]. \tag{9}$$

The integrals and inner products of the Bernstein polynomials can be calculated precisely as presented in the following lemma.

Lemma 1 *Let $p = \sum_{i+j+k=d} c_{ijk} B_{ijk}^d$ be a polynomial of degree d on triangle T (with area A_T), then*

$$\int_T p(x,y)dxdy = \frac{A_T}{\binom{d+2}{2}} \sum_{i+j+k=d} c_{ijk}. \tag{10}$$

Let $q = \sum_{\nu+\mu+\kappa=d} \tilde{c}_{\nu\mu\kappa} B_{\nu\mu\kappa}^d$ be another polynomial of degree d on triangle T, then the inner product of p and q is

$$\int_T p(x,y)q(x,y)dxdy = \frac{A_T}{\binom{2d}{d}\binom{2d+2}{2}} \sum_{\substack{i+j+k=d \\ \nu+\mu+\kappa=d}} \binom{i+\nu}{i}\binom{j+\mu}{j}\binom{k+\kappa}{k} c_{ijk}\tilde{c}_{\nu\mu\kappa}. \tag{11}$$

For the spline space $S_d^0(\boldsymbol{\Delta})$, the domain points are defined to be the set

$$\mathcal{D}_{d,\boldsymbol{\Delta}} = \{\xi_{ijk} = (i\mathbf{v}_1 + j\mathbf{v}_2 + k\mathbf{v}_3)/d, i+j+k = d, T = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\rangle \in \boldsymbol{\Delta}\}.$$

Therefore the spline function can also be denoted by

$$s|_T = \sum_{\xi \in \mathcal{D}_{d,T}} c_\xi B_\xi^{T,d},$$

where $B_\xi^{T,d}$ stands for $B_{ijk}^{T,d}$ for $\xi = \xi_{ijk} \in \mathcal{D}_{d,T}$ and c_ξ is the corresponding B-coefficient c_{ijk}. Note that since s is continuous, if ξ lies on an edge shared by two different triangles

T and $\tilde{T}$, then the corresponding coefficients c_ξ for $s|_T$ and $s|_{\tilde{T}}$ should be the same. Then we show that the basis for $S_d^0(\mathbf{\Delta})$ can be constructed easily with spline functions in $S_d^0(\mathbf{\Delta})$ with specific B-coefficients. For each $\xi \in \mathcal{D}_{d,\mathbf{\Delta}}$, let ψ_ξ be the spline in $S_d^0(\mathbf{\Delta})$ having all zero B-coefficients except for $c_\xi = 1$, then we have the following result

Lemma 2 *The set of splines $\mathcal{B} = \{\psi_\xi, \xi \in \mathcal{D}_{d,\mathbf{\Delta}}\}$ forms a basis for the spline space $S_d^0(\mathbf{\Delta})$ which satisfies $\psi_\xi(\mathbf{v}) \geq 0$ and $\sum_{\xi \in \mathcal{D}_{d,\mathbf{\Delta}}} \psi_\xi(\mathbf{v}) = 1$ for all $\mathbf{v} \in \Omega$.*

It is obvious that ψ_ξ is identically zero on all triangles that do not contain ξ since the corresponding B-coefficients are all zeros so that ψ_ξ is locally supported.

A.2 Proof of Theorem 1

By plugging the bivariate spline representation of $x(\mathbf{u})$ in to the equality (3), we have

$$\left\{ \langle \phi_t, \sum_{h=1}^{m} (\kappa^2 - \Delta)^{\alpha/2} \tau \psi_h w_h \rangle, t = 1, ..., n_t \right\} \stackrel{d}{=} \left\{ \langle \phi_t, W \rangle, t = 1, ..., n_t \right\}, \tag{12}$$

for any appropriate set of test functions $\{\phi_t, t = 1, ..., n_t\}$.

When $\alpha = 1$:

By choosing a set of test functions to be $\phi_h = (\kappa^2 - \Delta)^{1/2} \psi_h$, we have

$$\left\{ \langle (\kappa^2 - \Delta)^{1/2} \psi_t, \sum_{s=1}^{m} (\kappa^2 - \Delta)^{1/2} \tau \psi_s w_s \rangle, t = 1, ..., m \right\} \stackrel{d}{=} \left\{ \langle (\kappa^2 - \Delta) \psi_t, W \rangle, t = 1, ..., m \right\}. \tag{13}$$

Following Lemma 2 of Lindgren et al. (2011), the left hand side of (13) is

$$\left\{ \sum_{s=1}^{m} \tau (\kappa^2 \langle \psi_t, \psi_s \rangle + \langle \nabla \psi_t, \nabla \psi_s \rangle) w_s, t = 1, ..., m \right\},$$

when the Neumann boundary condition holds. The integral on the right hand side is in fact Gaussian with mean zero and covariance matrix whose (t, s)-th element is

$$\begin{aligned} \mathrm{Cov}(\langle (\kappa^2 - \Delta)^{1/2} \psi_t, W \rangle, \langle (\kappa^2 - \Delta)^{1/2} \psi_s, W \rangle) &= \langle (\kappa^2 - \Delta)^{1/2} \psi_t, (\kappa^2 - \Delta)^{1/2} \psi_s \rangle \\ &= \kappa^2 \langle \psi_t, \psi_s \rangle + \langle \nabla \psi_t, \nabla \psi_s \rangle. \end{aligned}$$

Then we can write (13) in the matrix form as

$$\tau(\kappa^2 \mathbf{M} + \mathbf{K}) \mathbf{w} \sim N(\mathbf{0}, \kappa^2 \mathbf{M} + \mathbf{K}),$$

where the (t,s)-th entry of the matrices $\mathbf{M}$, $\mathbf{K}$ are respectively $\mathbf{M}_{ts} = \langle \psi_t, \psi_s \rangle$, $\mathbf{K}_{ts} = \langle \nabla\psi_t, \nabla\psi_s \rangle$. $\mathbf{M}$ and $\mathbf{K}$ are usually named mass matrix and stiffness matrix respectively in bivariate spline literature. Therefore it is easy to show that the precision matrix of $\mathbf{w}$ is $\mathbf{Q} = \tau^2(\kappa^2\mathbf{M} + \mathbf{K})$.

When $\alpha = 2$:

We can choose the specific set of test functions to be $\phi_h = (\kappa^2 - \Delta)\psi_h$ or $\phi_h = \psi_h$, leading to the least squares or Galerkin solutions respectively.

(1) When $\phi_h = (\kappa^2 - \Delta)\psi_h$, we have

$$\left\{ \langle (\kappa^2 - \Delta)\psi_t, \sum_{s=1}^{m} (\kappa^2 - \Delta)\tau\psi_s w_s \rangle, t = 1, ..., m \right\} \stackrel{d}{=} \left\{ \langle (\kappa^2 - \Delta)\psi_t, W \rangle, t = 1, ..., m \right\}. \quad (14)$$

The left hand side of (14) is

$$\left\{ \sum_{s=1}^{m} \tau(\kappa^4 \langle \psi_t, \psi_s \rangle + 2\kappa^2 \langle \nabla\psi_t, \nabla\psi_s \rangle + \langle \Delta\psi_s, \Delta\psi_t \rangle) w_s, t = 1, ..., m \right\},$$

by applying the stochastic Green's first identity along with the Neumann boundary condition. The integral on the right hand side is Gaussian with mean zero and covariance matrix whose (t,s)-th element is

$$\mathrm{Cov}(\langle (\kappa^2 - \Delta)\psi_t, W \rangle, \langle (\kappa^2 - \Delta)\psi_s, W \rangle) = \kappa^4 \langle \psi_t, \psi_s \rangle + 2\kappa^2 \langle \nabla\psi_t, \nabla\psi_s \rangle + \langle \Delta\psi_s, \Delta\psi_t \rangle.$$

Then we can write (14) in the matrix form as

$$\tau(\kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{R})\mathbf{w} \sim N(\mathbf{0}, \kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{R}),$$

where the (t,s)-th entry of the matrix $\mathbf{R}$ is $\mathbf{R}_{ts} = \langle \Delta\psi_t, \Delta\psi_s \rangle$. The matrices $\mathbf{M}$ and $\mathbf{K}$ are defined above, and $\mathbf{R}$ is usually called roughness matrix. Then the precision matrix of $\mathbf{w}$ for the least squares solution can be easily shown to be $\mathbf{Q}^{LS} = \tau^2(\kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{R})$.

(2) When $\phi_h = \psi_h$, we have

$$\left\{ \langle \psi_t, \sum_{s=1}^{m} (\kappa^2 - \Delta)\tau\psi_s w_s \rangle, t = 1, ..., m \right\} \stackrel{d}{=} \left\{ \langle (\kappa^2 - \Delta)\psi_t, W \rangle, t = 1, ..., m \right\}. \quad (15)$$

Following the same procedure as for the lease squares solution, we have the left hand side of (15) is in fact

$$\left\{ \sum_{s=1}^{m} \tau(\kappa^2 \langle \psi_t, \psi_s \rangle + \langle \nabla\psi_t, \nabla\psi_s \rangle) w_s, t = 1, ..., m \right\},$$

and the integral on the right hand side is Gaussian with mean zero and covariance matrix whose (t, s)-th element is

$$\text{Cov}(\langle\psi_t, W\rangle, \langle\psi_s, W\rangle) = \langle\psi_t, \psi_s\rangle.$$

Then (15) can be re-written as

$$\tau(\kappa^2\mathbf{M} + \mathbf{K})\mathbf{w} \sim N(\mathbf{0}, \mathbf{M}).$$

Then the precision matrix of $\mathbf{w}$ for the Galerkin solution is $\mathbf{Q}^G = \tau^2(\kappa^2\mathbf{M}+\mathbf{K})\mathbf{M}^{-1}(\kappa^2\mathbf{M}+\mathbf{K}) = \tau^2(\kappa^4\mathbf{M} + 2\kappa^2\mathbf{K} + \mathbf{K}\mathbf{M}^{-1}\mathbf{K})$.

When $\alpha \geq 3$:

Following the recursive algorithm, we can find the solution to the SPDE $\tau(\kappa^2-\Delta)^{\alpha/2}x(\mathbf{u}) = W(\mathbf{u})$ by solving the innovative SPDE $(\kappa^2 - \Delta)x(\mathbf{u}) = \tilde{x}(\mathbf{u})$, where $\tilde{x}(\mathbf{u})$ is the solution to the SPDE $\tau(\kappa^2 - \Delta)^{(\alpha-2)/2}\tilde{x}(\mathbf{u}) = W(\mathbf{u})$. Then by choosing the test functions $\phi_h = \psi_h$, $h = 1, ..., m$ and following the same procedure for the Galerkin solution when $\alpha = 2$, we have

$$\mathbf{Q}_\alpha = (\kappa^2\mathbf{M} + \mathbf{K})\mathbf{M}^{-1}\mathbf{Q}_{\alpha-2}\mathbf{M}^{-1}(\kappa^2\mathbf{M} + \mathbf{K}),$$

which can be expanded to the expression in the theorem.

Then we show the calculations of the matrix components $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{R}$. Following Lemma 1 and $\nabla p = \sum_{i+j+k=d} c_{ijk}\nabla B^d_{ijk}$, $\Delta p = \sum_{i+j+k=d} c_{ijk}\Delta B^d_{ijk}$ for any $p = \sum_{i+j+k=d} c_{ijk}B^d_{ijk}$, we have the contribution of triangle T to the (t, s)-th entry of $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{R}$ for $t, s = 1, ..., m$ are

$$\mathbf{M}_{ts}|_T = \langle\psi_t, \psi_s\rangle_T = \mathbf{c}_t'|_T M_T \mathbf{c}_s|_T,$$

$$\mathbf{K}_{ts}|_T = \langle\nabla\psi_t, \nabla\psi_s\rangle_T = \mathbf{c}_t'|_T K_T \mathbf{c}_s|_T,$$

$$\mathbf{R}_{ts}|_T = \langle\nabla\psi_t, \nabla\psi_s\rangle_T = \mathbf{c}_t'|_T R_T \mathbf{c}_s|_T,$$

where M_T, K_T and R_T are defined in Theorem 1, and $\mathbf{c}_h|_T$ is the column vector of B-coefficients of ψ_h associated with triangle T, $h = 1, ..., m$. Then it is followed that

$$\mathbf{M}_{ts} = \sum_T \mathbf{M}_{ts}|_T = \mathbf{c}_t'\mathbf{M_0}\mathbf{c}_s, \quad \mathbf{K}_{ts} = \sum_T \mathbf{K}_{ts}|_T = \mathbf{c}_t'\mathbf{K_0}\mathbf{c}_s, \quad \mathbf{R}_{ts} = \sum_T \mathbf{M}_{ts}|_T = \mathbf{c}_t'\mathbf{R_0}\mathbf{c}_s,$$

where $\mathbf{M_0} = \text{diag}(\mathbf{M}_T, T \in \mathbf{\Delta})$, $\mathbf{K_0} = \text{diag}(\mathbf{K}_T, T \in \mathbf{\Delta})$ and $\mathbf{R_0} = \text{diag}(\mathbf{R}_T, T \in \mathbf{\Delta})$. Therefore we have the following simple matrix representation that

$$\mathbf{M} = \mathbf{C}'\mathbf{M_0}\mathbf{C}, \quad \mathbf{K} = \mathbf{C}'\mathbf{K_0}\mathbf{C}, \quad \mathbf{R} = \mathbf{C}'\mathbf{R_0}\mathbf{C}.$$

A.3 Proof of Proposition 1

Let $f_{\mathbf{\Delta}}(\mathbf{s})$ be the H^1-orthogonal projection of $f \in H^1 \cap W_2^{m+1}(\Omega)$ onto the bivariate spline space $S_d^0(\mathbf{\Delta})$, it follows that

$$\begin{aligned}\int_\Omega f(\mathbf{s}) L x_{\mathbf{\Delta}}(\mathbf{s}) d\mathbf{s} &= \int_\Omega (f(\mathbf{s}) - f_{\mathbf{\Delta}}(\mathbf{s})) L x_{\mathbf{\Delta}}(\mathbf{s}) d\mathbf{s} + \int_\Omega f_{\mathbf{\Delta}}(\mathbf{s}) L x_{\mathbf{\Delta}}(\mathbf{s}) d\mathbf{s} \\ &= \int_\Omega f_{\mathbf{\Delta}}(\mathbf{s}) L x_{\mathbf{\Delta}}(\mathbf{s}) d\mathbf{s} \\ &= \int_\Omega f_{\mathbf{\Delta}}(\mathbf{s}) dW(\mathbf{s}),\end{aligned}$$

where the second equality follows from the orthogonality of $f(\mathbf{s}) - f_{\mathbf{\Delta}}(\mathbf{s})$ to $S_d^0(\mathbf{\Delta})$ with respect to H^1 inner product. Then we have

$$\int_\Omega f(\mathbf{s}) L(x(\mathbf{s}) - x_{\mathbf{\Delta}}(\mathbf{s})) d\mathbf{s} = \int_\Omega (f(\mathbf{s}) - f_{\mathbf{\Delta}}(\mathbf{s})) dW(\mathbf{s}).$$

Hence it follows from the white noise integrals that

$$\begin{aligned}\mathrm{E}\left(\int_\Omega f(\mathbf{s}) L(x(\mathbf{s}) - x_{\mathbf{\Delta}}(\mathbf{s})) d\mathbf{s}\right)^2 &= \mathrm{E}\left(\int_\Omega (f(\mathbf{s}) - f_{\mathbf{\Delta}}(\mathbf{s})) dW(\mathbf{s})\right)^2 \\ &= \int_\Omega (f(\mathbf{s}) - f_{\mathbf{\Delta}}(\mathbf{s}))^2 d\mathbf{s}.\end{aligned}$$

Then it follows from standard results in bivariate splines literatures, for example Th. 5.19 in Lai and Schumaker (2007) that under some suitable assumptions on the triangulation, we have for $1 \le m \le d$,

$$\|f - f_{\mathbf{\Delta}}\|_{2,\Omega} \le K |\mathbf{\Delta}|^{m+1} |f|_{m+1,2,\Omega}.$$

A.4 Proof of Proposition 2

First of all, it is easy to see that

$$w_f' \mathbf{M} w_g = \langle f_{\mathbf{\Delta}}, g_{\mathbf{\Delta}} \rangle_{\mathbf{\Delta}} = \sum_{T \in \mathbf{\Delta}} \int_T f_{\mathbf{\Delta}} g_{\mathbf{\Delta}} dx dy \tag{16}$$

since $f_{\mathbf{\Delta}}, g_{\mathbf{\Delta}} \in S_d^0(\mathbf{\Delta})$. Next we can see

$$\begin{aligned}\mathbf{w}_f' \tilde{\mathbf{M}} \mathbf{w}_g &= \sum_{T \in \mathbf{\Delta}} \sum_{\xi \in \mathcal{D}_{d,T}} c_\xi(f_{\mathbf{\Delta}}) \sum_\eta \int_T \phi_\xi \phi_\eta dx dy c_\xi(g_{\mathbf{\Delta}}) \\ &= \sum_{T \in \mathbf{\Delta}} \sum_{\xi \in \mathcal{D}_{d,T}} \frac{A_T}{\binom{d+2}{2}} c_\xi(f_{\mathbf{\Delta}}) c_\xi(g_{\mathbf{\Delta}}),\end{aligned}$$

where $\mathcal{D}_{d,T} = \{(i\mathbf{v}_1 + j\mathbf{v}_2 + k\mathbf{v}_3)/d, i+j+k=d\}$ is the set of associated domain points of triangle $T = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$, A_T is the area of triangle T and $c_\xi(s)$ is the B-coefficient of s. When $f_{\boldsymbol{\Delta}} = C$ is a constant C, it is easy to see that

$$\sum_{\xi \in \mathcal{D}_{d,T}} \frac{A_T}{\binom{d+2}{2}} c_\xi(f_{\boldsymbol{\Delta}}) c_\xi(g_{\boldsymbol{\Delta}}) = \int_T f_{\boldsymbol{\Delta}} g_{\boldsymbol{\Delta}} dxdy$$

and hence, we have

$$\mathbf{w}_f' \tilde{\mathbf{M}} \mathbf{w}_g = \sum_{T \in \boldsymbol{\Delta}} \int_T f_{\boldsymbol{\Delta}} g_{\boldsymbol{\Delta}} dxdy$$

which is $w_f' \mathbf{M} w_g$ by (16). Similar when $g_{\boldsymbol{\Delta}}$ is a piecewise constant. Also, when $d = 1$, this result follows from Lemma 1 in Chen and Thomée (1985). We now prove it for general $d \geq 1$.

We first note that

$$\begin{aligned} c_\xi(f_{\boldsymbol{\Delta}}) c_\xi(g_{\boldsymbol{\Delta}}) =& (c_\xi(f_{\boldsymbol{\Delta}}) - f_{\boldsymbol{\Delta}}(\xi)) c_\xi(g_{\boldsymbol{\Delta}}) + f_{\boldsymbol{\Delta}}(\xi)(c_\xi(g_{\boldsymbol{\Delta}}) - g_{\boldsymbol{\Delta}}(\xi)) \\ &+ f_{\boldsymbol{\Delta}}(\xi) g_{\boldsymbol{\Delta}}(\xi). \end{aligned}$$

Then we claim that

$$\sum_{\xi \in \mathcal{D}_{d,T}} \frac{A_T}{\binom{d+2}{2}} f_{\boldsymbol{\Delta}}(\xi) g_{\boldsymbol{\Delta}}(\xi) \text{ approximates } \int_T f_{\boldsymbol{\Delta}}(x,y) g_{\boldsymbol{\Delta}}(x,y) dxdy.$$

Indeed, let us recall the Bernstein-Bézier approximation of arbitrary continuous function F on T. That is, using Th. 2.45 in Lai and Schumaker (2007), we have

$$\|F - B_d(F)\|_{T,\infty} \leq \frac{|T|^2}{d} |F|_{2,T} \tag{17}$$

where $B_d(F) = \sum_{\xi \in \mathcal{D}_{d,T}} F(\xi) B_\xi$ and B_ξ are the Bernestein-Bézier polynomials of degree d. Letting $F(x,y) = f_{\boldsymbol{\Delta}}(x,y) g_{\boldsymbol{\Delta}}(x,y)$, we have

$$\begin{aligned} & |\int_T F(x,y) dxdy - \int_T B_d(F) dxdy| \\ =& |\int_T f_{\boldsymbol{\Delta}}(x,y) g_{\boldsymbol{\Delta}}(x,y) dxdy - \sum_{\xi \in \mathcal{D}_{d,T}} f_{\boldsymbol{\Delta}}(\xi) g_{\boldsymbol{\Delta}}(\xi) \frac{A_T}{\binom{d+2}{2}}| \\ \leq& \frac{|T|^2}{d} \int_T |f_{\boldsymbol{\Delta}} g_{\boldsymbol{\Delta}}|_{2,T} dxdy \\ \leq& K \frac{|T|^2}{d} |f_{\boldsymbol{\Delta}} g_{\boldsymbol{\Delta}}|_{2,1,T} \\ \leq& K \frac{|T|^2}{d} |f_{\boldsymbol{\Delta}}|_{2,2,T} |g_{\boldsymbol{\Delta}}|_{2,2,T}, \end{aligned}$$

where we have used the fact that $f_{\mathbf{\Delta}} g_{\mathbf{\Delta}}$ is a polynomial of degree $2d$ in the second inequality and the Cauchy-Schwarz inequality in the last inequality. This finishes the proof of the claim.

Next we consider

$$I_1(T) := \sum_{\xi \in \mathcal{D}_{d,T}} \frac{A_T}{\binom{d+2}{2}} (f_{\mathbf{\Delta}}(\xi) - c_\xi(f_{\mathbf{\Delta}})) c_\xi(g_{\mathbf{\Delta}}).$$

We have

$$|I_1(T)| = A_T \|\{f_{\mathbf{\Delta}}(\xi) - c_\xi(f_{\mathbf{\Delta}})\}_{\xi \in \mathcal{D}_{d,T}}\|_\infty \|\{c_\xi(g_{\mathbf{\Delta}})\}_{\xi \in \mathcal{D}_{d,T}}\|_\infty$$

and hence, by Th. 2.6 in Lai and Schumaker (2007),

$$|I_1(T)| \le A_T K^2 \|B_d(f_{\mathbf{\Delta}}) - f_{\mathbf{\Delta}}\|_T |g_{\mathbf{\Delta}}|_T,$$

where K is a positive constant. We use the property of Bernstein-Bézier approximation again, i.e. the estimate in (17) to have

$$\begin{aligned} |I_1(T)| \le & K^2 A_T \frac{|T|^2}{d} |f_{\mathbf{\Delta}}|_{2,T} \|g_{\mathbf{\Delta}}\|_{\infty,\Omega} \\ \le & K^2 \|f_{\mathbf{\Delta}}\|_{2,1,T} \|g_{\mathbf{\Delta}}\|_{\infty,\Omega}. \end{aligned}$$

Therefore we have

$$\sum_{T \in \mathbf{\Delta}} |I_1(T)| \le K^2 \frac{|T|^2}{d} \|g_{\mathbf{\Delta}}\|_{\infty,\Omega} \|f_{\mathbf{\Delta}}\|_{2,1,\Omega}.$$

Similarly, we can discuss

$$I_2(T) := \sum_{\xi \in \mathcal{D}_{d,T}} \frac{A_T}{\binom{d+2}{2}} f_{\mathbf{\Delta}}(\xi)(c_\xi(g_{\mathbf{\Delta}}) - g_{\mathbf{\Delta}}(\xi))$$

to have a similar estimate as $I_1(T)$. Putting these three estimates above we have obtained

$$|\epsilon_{\mathbf{\Delta}}(f_{\mathbf{\Delta}}, g_{\mathbf{\Delta}})| \le K |\mathbf{\Delta}|^2 (\|f_{\mathbf{\Delta}}\|_{2,2,\Omega} \|g_{\mathbf{\Delta}}\|_{2,2,\Omega} + \|f_{\mathbf{\Delta}}\|_{2,1,\Omega} \|g_{\mathbf{\Delta}}\|_{\infty,\Omega} + \|f_{\mathbf{\Delta}}\|_{\infty,\Omega} \|g_{\mathbf{\Delta}}\|_{2,1,\Omega}),$$

where K is a positive constant, $|\mathbf{\Delta}|$ is the length of the longest edge in the triangulation $\mathbf{\Delta}$. These complete the proof.

References

Amante, C. and B. W. Eakins (2009). *ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division.

Awanou, G., M.-J. Lai, and P. Wenston (2006). The multivariate spline method for scattered data fitting and numerical solutions of partial differential equations. *Wavelets and splines: Athens*, 24–74.

Babuska, I., B. A. Szabo, and I. N. Katz (1981). The p-version of the finite element method. *SIAM Journal on Numerical Analysis 18*(3), 515–545.

Banerjee, S., A. E. Gelfand, and B. P. Carlin (2004). *Hierarchical modeling and analysis for spatial data*. CRC Press.

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

Brenner, S. C. and L. R. Scott (2008). *The mathematical theory of finite element methods*, Volume 15. Springer.

Chen, C. M. and V. Thomée (1985). The lumped mass finite element method for a parabolic problem. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics 26*(03), 329–354.

Cressie, N. (1993). *Statistics for spatial data*, Volume 900. Wiley New York.

Diggle, P. and P. J. Ribeiro (2007). *Model-based geostatistics*. Springer.

Eakins, B. W. and L. A. Taylor (2010). *Seamlessly integrating bathymetric and topographic data to support tsunami modeling and forecasting efforts*, Chapter 2. ESRI Press, Redlands.

Ettinger, B., S. Guillas, and M.-J. Lai (2012). Bivariate splines for ozone concentration forecasting. *Environmetrics 23*(4), 317–328.

Guillas, S. and M.-J. Lai (2010). Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics 22*(4), 477–497.

Hu, X.-L., D.-F. Han, and M.-J. Lai (2007). Bivariate splines of various degrees for numerical solution of partial differential equations. *SIAM Journal on Scientific Computing 29*(3), 1338–1354.

Lai, M.-J. and L. L. Schumaker (2007). *Spline functions on triangulations*, Volume 110. Cambridge University Press.

Lai, M.-J., C. K. Shum, V. Baramidze, and P. Wenston (2009). Triangulated spherical splines for geopotential reconstruction. *Journal of Geodesy 83*(8), 695–708.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(4), 423–498.

Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain (2015). A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics 24*(2), 579–599.

Paciorek, C. J. and M. J. Schervish (2004). Nonstationary covariance functions for Gaussian process regression. *Advances in Neural Information Processing Systems 16*, 273–280.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Sangalli, L. M., J. O. Ramsay, and T. O. Ramsay (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 681–703.

Simpson, D., F. Lindgren, and H. Rue (2012). Think continuous: Markovian gaussian models in spatial statistics. *Spatial Statistics 1*, 16–29.

Smith, W. H. F. and P. Wessel (1990). Gridding with continuous curvature splines in tension. *Geophysics 55*(3), 293–305.

Walsh, J. B. (1986). *An introduction to stochastic partial differential equations*. Springer.

Wessel, P. and D. Bercovici (1998). Interpolation with splines in tension: A Green's function approach. *Mathematical Geology 30*(1), 77–93.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 434–449.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute 40*(2), 975–994.