

Multivariate isotropic random fields on spheres: Nonparametric Bayesian modeling and L^p fast approximations

Alfredo Alegría¹, Pier Giovanni Bissiri², Galatia Cleanthous³,
Emilio Porcu⁴, Philip White^{5,*}

¹*Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile*

e-mail: alfredo.alegria@usm.cl

²*Department of Statistics, University of Bologna, Italy*

e-mail: piergiovanni.bissiri@unibo.it

³*Department of Mathematics and Statistics, National University of Ireland, Maynooth*

e-mail: galatia.cleanthous@gmail.com

⁴*Department of Mathematics, Khalifa University, Abu Dhabi, The United Arab Emirates*

e-mail: georgepolya01@gmail.com

⁵*Department of Statistics, Brigham Young University*

e-mail: pwhite@stat.byu.edu

Abstract: We study multivariate Gaussian random fields defined over d -dimensional spheres. First, we provide a nonparametric Bayesian framework for modeling and inference on matrix-valued covariance functions. We determine the support (under the topology of uniform convergence) of the proposed random matrices, which cover the whole class of matrix-valued geodesically isotropic covariance functions on spheres. We provide a thorough inspection of the properties of the proposed model in terms of (a) first moments, (b) posterior distributions, and (c) Lipschitz continuities. We then provide an approximation method for multivariate fields on the sphere for which measures of L^p accuracy are established. Our findings are supported through simulation studies that show the rate of convergence when truncating a spectral expansion of a multivariate random field at a finite order. To illustrate the modeling framework developed in this paper, we consider a bivariate spatial data set of two 2019 NCEP/NCAR Flux Reanalyses.

Keywords and phrases: Bivariate climate data, matrix-valued covariance function, multivariate random field, nonparametric Bayes, L^p approximations, sphere.

Received April 2020.

1. Introduction

The paper deals with multivariate (or vector-valued) Gaussian random fields defined over the d -dimensional unit sphere $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1}, \|\mathbf{x}\| = 1\}$ embedded in \mathbb{R}^{d+1} , having a specified matrix-valued covariance function, see [Ma](#)

*Corresponding author.

(2012), Du, Ma and Li (2013), Ma (2015), Porcu, Bevilacqua and Genton (2016), Alegría et al. (2019a), Bevilacqua, Diggle and Porcu (2019) and Alegría (2020) for recent literature on this topic. Choosing the d -dimensional sphere as an index set is motivated by the increasing availability of global data, covering a big portion of our planet, and we remind the reader of recent applications in Reinsel et al. (1981), Di Lorenzo et al. (2014), Nychka et al. (2015), Combes, Hormázabal and Di Lorenzo (2017), Edwards, Castruccio and Hammerling (2019). In addition to the Earth sciences, Cosmic Microwave Background (CMB) radiation provides a natural application for Gaussian random field models over \mathbb{S}^2 (see Cabella and Marinucci, 2009; Marinucci and Peccati, 2011). Recently, random fields using the circle as an index set have been used to capture temporal seasonality (Shirota and Gelfand, 2017; White and Porcu, 2019; Mastrantonio et al., 2019) or direction (Jona-Lasinio, Gelfand and Jona-Lasinio, 2012; Wang and Gelfand, 2014). Additionally, spheres of dimension $d > 2$ are relevant in various disciplines, such as physics, chemistry and material sciences, as discussed in Dette et al. (2019). For other manifolds we refer the reader to Kerkycharian et al. (2018), Lu, Leonenko and Ma (2020) and Lu and Ma (2020).

There is a rich literature on nonparametric Bayesian approaches for scalar-valued random fields on planar surfaces. See, for instance, Schmidt and O'Hagan (2003); Gelfand, Kottas and MacEachern (2005); Duan, Guindani and Gelfand (2007); Zheng, Zhu and Roy (2010); Reich and Fuentes (2012); Chopin, Rousseau and Liseo (2013); Holbrook et al. (2018) and Müller, Quintana and Page (2018). In addition, there has also been some recent work on nonparametric Bayesian inference for scalar-valued random fields and their covariance functions on spheres (Porcu et al., 2019) and more general manifolds (Castillo, Kerkycharian and Picard, 2014). Despite the activity in this area, multivariate random fields on spheres have not been considered so far under the Bayesian nonparametric framework. The benefit of adopting a nonparametric Bayesian framework are that (a) we focus on Gaussian random fields, so that the matrix-valued covariance function becomes crucial for modeling, inference, and prediction, and (b) we can exploit the simple spectral representation for multivariate covariance functions on spheres (Hannan, 2009; Yaglom, 1987). Second, fast simulation of multivariate Gaussian random fields over spheres has only been considered to a limited extent, and the reader is referred to the HEALPix software (Gorski et al., 2005), and recent contributions by Emery and Porcu (2019) and Alegría, Emery and Lantuéjoul (2020).

We focus our work on developing approaches for nonparametric Bayesian modeling and fast simulation for multivariate random fields over spheres. Specifically, our paper contributes as follows:

1. We provide a strategy to nonparametric Bayesian modeling of matrix-valued covariance functions over d -dimensional spheres. The basic idea is that the matrix-valued covariance function is considered as a random matrix with a given prior probability distribution.
2. We show that our problem is well-posed in the sense that the support of these random matrix-valued covariance functions is, under uniform topol-

- ogy, equal to the whole set of geodesically isotropic covariance functions that are defined over d -dimensional spheres.
3. We then provide a detailed study of the posterior distributions, the first-order moments of the so-defined random matrices, and the geometric properties of these random matrices in terms of Lipschitz continuities.
 4. We propose an approximation method for k -variate random fields with given classes of geodesically isotropic matrix-valued covariance functions. Our method is based on truncation at a final order of a spectral expansion associated with a k -variate random field defined over \mathbb{S}^d . We assess the accuracy of the approximation in the L^p sense, detailed subsequently. Our findings are then supported through a simulation study on bivariate random fields.
 5. We illustrate our findings on a bivariate climate data set where we discuss modeling details.

The plan of the paper is the following: Section 2 contains the necessary mathematical background that is necessary to illustrate our findings. Contributions 1.–3. above are provided in Section 3, and contribution 4. in Section 4. Section 5 presents an analysis of a bivariate global climate dataset (contribution 5.). The paper concludes with a discussion in Section 6. Technical proofs are deferred to the Appendix.

2. Background

2.1. Notation

We use \mathbf{A}' to denote the transpose of a matrix \mathbf{A} . For a positive semi-definite square matrix, \mathbf{A} , a square root, $\mathbf{A}^{1/2}$, is a matrix of the same order of \mathbf{A} such that $\mathbf{A} = \mathbf{A}^{1/2}(\mathbf{A}^{1/2})'$. This decomposition is unique when \mathbf{A} is (strictly) positive definite. For the remainder of the paper, $\mathbb{I}(k)$ denotes the $k \times k$ identity matrix.

Let $\mathbb{R}^{k_1 \times k_2}$ denote the space of matrices of order $k_1 \times k_2$, where $k_1, k_2 \in \mathbb{N}$. The Frobenius inner product in $\mathbb{R}^{k_1 \times k_2}$ is defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle_F := \text{trace}(\mathbf{A}\mathbf{B}'), \quad (1)$$

where the trace of a square matrix is defined as the sum of the elements on its main diagonal. The corresponding norm of a matrix takes the form $\|\mathbf{A}\|_F = \langle \mathbf{A}, \mathbf{A} \rangle_F^{1/2}$. The results provided in Section 4 make use of the following Lebesgue norms. Let (\mathcal{M}, μ) be a measure space and $p \geq 1$. The $L^p(\mathcal{M}; \mathbb{R}^k)$ -norm of a vector valued function $\mathbf{f} : \mathcal{M} \rightarrow \mathbb{R}^k$ is defined as

$$\|\mathbf{f}\|_{L^p(\mathcal{M}; \mathbb{R}^k)} := \left(\int_{\mathcal{M}} \|\mathbf{f}(x)\|_F^p d\mu(x) \right)^{1/p}. \quad (2)$$

Special emphasis is put on the case $\mathcal{M} = \mathbb{S}^d$, equipped with the spherical measure μ . The corresponding inner product of two vector valued functions

$\mathbf{f}, \mathbf{g} : \mathbb{S}^d \rightarrow \mathbb{R}^k$, takes the form:

$$\langle \mathbf{f}, \mathbf{g} \rangle_{L^2(\mathbb{S}^d; \mathbb{R}^k)} := \int_{\mathbb{S}^d} \langle \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}) \rangle_F d\mu(\mathbf{x}). \tag{3}$$

Now, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The vector valued function $\mathbf{f} : \mathbb{S}^d \times \Omega \rightarrow \mathbb{R}^k$ belongs to the mixed Lebesgue space $L^p(\Omega, L^q(\mathbb{S}^d; \mathbb{R}^k))$, $p, q \geq 1$, if

$$\begin{aligned} \|\mathbf{f}\|_{L^p(\Omega, L^q(\mathbb{S}^d; \mathbb{R}^k))} &:= \left(\mathbb{E}(\|\mathbf{f}\|_{L^q(\mathbb{S}^d; \mathbb{R}^k)}^p) \right)^{1/p} \\ &= \left(\int_{\Omega} \|\mathbf{f}(\cdot, \omega)\|_{L^q(\mathbb{S}^d; \mathbb{R}^k)}^p d\mathbb{P}(\omega) \right)^{1/p} \\ &= \left(\int_{\Omega} \left(\int_{\mathbb{S}^d} \|\mathbf{f}(\mathbf{x}, \omega)\|_F^q d\mathbf{x} \right)^{p/q} d\mathbb{P}(\omega) \right)^{1/p} < \infty. \end{aligned} \tag{4}$$

For the special case $p = q = 2$ we have

$$\|\mathbf{f}\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}^2 = \mathbb{E}(\|\mathbf{f}\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)}^2) = \mathbb{E}\left(\int_{\mathbb{S}^d} \|\mathbf{f}(\mathbf{x})\|_F^2 d\mu(\mathbf{x}) \right). \tag{5}$$

2.2. Multivariate random fields on spheres

We define a k -variate zero-mean Gaussian random field

$$\mathbb{Z}(\mathbf{x}) = \{(Z_1(\mathbf{x}), \dots, Z_k(\mathbf{x}))', \mathbf{x} \in \mathbb{S}^d\}$$

as a vector-valued stochastic process that is continuously indexed over \mathbb{S}^d . The components Z_i , $i = 1, \dots, k$, are called scalar random fields. The covariance function $\mathbb{C} : [0, \pi] \rightarrow \mathbb{R}^{k \times k}$, is a matrix valued mapping, whose elements are defined as $C_{i,j}(\theta(\mathbf{x}_1, \mathbf{x}_2)) = \text{cov}\{Z_i(\mathbf{x}_1), Z_j(\mathbf{x}_2)\}$, where $\theta(\cdot, \cdot) : \mathbb{S}^d \times \mathbb{S}^d \rightarrow [0, \pi]$ is the geodesic distance, defined as

$$\theta(\mathbf{x}_1, \mathbf{x}_2) = \arccos(\mathbf{x}'_1 \mathbf{x}_2), \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d,$$

and through the paper, we make use of the shortcut θ whenever there is no confusion. Following [Porcu, Bevilacqua and Genton \(2016\)](#) and [Alegría et al. \(2019a\)](#), we say that \mathbb{C} is geodesically isotropic, as it just depends on \mathbf{x}_1 and \mathbf{x}_2 through their geodesic distance. This paper does not adopt any strategy of covariance modeling through the chordal distance, which has received substantial criticisms in the recent papers by [Gneiting \(2013\)](#) and [Porcu, Bevilacqua and Genton \(2016\)](#), as well as an earlier paper by [Banerjee \(2005\)](#). Throughout, the diagonal elements $C_{i,i}$ are called marginal covariance functions, while the off-diagonal elements $C_{i,j}$ are called cross covariance functions.

For the remainder of the paper, we always assume continuity for the elements $C_{i,j}$ of the matrix-valued mapping \mathbb{C} . This assumption is not too restrictive because discontinuous models are rarely used in practice, with the exception

of the *nugget effect model* which is regularly used as a summand in additive covariance structures for capturing abrupt short scale variations in spatial data applications (Chiles and Delfiner, 2009).

The mapping \mathbb{C} is positive semi-definite; that is,

$$\sum_{\ell=1}^n \sum_{r=1}^n \mathbf{a}'_{\ell} \mathbb{C}(\theta(\mathbf{x}_{\ell}, \mathbf{x}_r)) \mathbf{a}_r \geq 0, \quad (6)$$

for all positive integer n , $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{S}^d$ and $\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \subset \mathbb{R}^k$. A necessary and sufficient condition (Hannan, 2009) for the continuous mapping \mathbb{C} to be the covariance matrix function of a k -variate stationary and isotropic Gaussian random field on \mathbb{S}^d (i.e., to be positive semi-definite) is that there exists a summable sequence of positive semidefinite $k \times k$ matrices $\{\mathbf{B}_n, n = 0, 1, 2, \dots\}$ such that, for $\theta \in [0, \pi]$,

$$\mathbb{C}(\theta) = \sum_{n=0}^{\infty} \mathbf{B}_n c_n(d, \cos(\theta)), \quad (7)$$

where $c_n(d, x)$ is defined as the normalized Gegenbauer polynomial, that is,

$$c_n(d, x) = \frac{C_n^{(d-1)/2}(x)}{C_n^{(d-1)/2}(1)} = \frac{n}{(d-1) \cdots (d+n-2)} C_n^{(d-1)/2}(x), \quad x \in [-1, 1]. \quad (8)$$

Since the Chebyshev polynomials $T_n = C_n^{(0)}$ are already normalized, the last expression is not valid for $d = 1$ and $c_n(1, x) = T_n(x)$. Such a result is the extension to the matrix-valued case of the celebrated Schoenberg's theorem (Schoenberg, 1942). We note that we have been sloppy when writing \mathbf{B}_n instead of $\mathbf{B}_{n,d}$ in (7): clearly, for any dimension d , there exists a different sequence $\mathbf{B}_{n,d}$ satisfying (7). Yet, this is not relevant for the present work, and for other technical aspects, the reader is referred to Porcu, Alegría and Furrer (2018) and the references therein.

Representation (7) has a precise interpretation in terms of the associated zero-mean Gaussian k -variate process, \mathbb{Z} , which admits a uniquely determined expansion of the type

$$\mathbb{Z}(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m \in N_{n,d}} \boldsymbol{\xi}_{n,m} \mathcal{Y}_{n,m}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^d, \quad (9)$$

where $\mathcal{Y}_{n,m}$ denotes spherical harmonics, an orthonormal basis for the space of squared integrable functions (against the Lebesgue measure) on the sphere; for more details see Dai and Xu (2013). The index set $N_{n,d}$ is finite and for details the reader is referred to Porcu, Alegría and Furrer (2018). Here, $\{\boldsymbol{\xi}_{n,m}\}$ is a sequence of random vectors in \mathbb{R}^k that have zero mean and such that $\mathbb{E} \boldsymbol{\xi}_{n,m} \boldsymbol{\xi}'_{k,h} = \delta_{n=k} \delta_{m=h} \mathbf{B}_n$, where δ is the Kronecker delta function and \mathbf{B}_n are the non negative and summable matrices defined at (7). An alternative expansion, with the same covariance structure, will be discussed in Section 4.

3. A Bayesian spectral model

3.1. Random matrix covariances, and their support

For a neater illustration of the results contained in this section, a slight change of notation is needed. We consider a stochastic process, W , being continuously indexed over $\mathbb{S}^d \times \{1, \dots, k\}$, and such that $W(\mathbf{x}, j) = Z_j(\mathbf{x})$ for $\mathbf{x} \in \mathbb{S}^d$ and $j = 1, \dots, k$, with Z_j being the j -th component of the k -variate random field \mathbb{Z} defined at (9). The covariance function K associated with the process W is defined as

$$K\left((\mathbf{x}_1, i), (\mathbf{x}_2, j)\right) = \text{Cov}\left(W(\mathbf{x}_1, i), W(\mathbf{x}_2, j)\right).$$

Clearly, $K\left((\mathbf{x}_1, i), (\mathbf{x}_2, j)\right)$ is the (i, j) -entry of the covariance matrix function $\mathbb{C}(\theta(\mathbf{x}_1, \mathbf{x}_2))$ defined at (7). Thus,

$$K\left((\mathbf{x}_1, i), (\mathbf{x}_2, j)\right) = \sum_{n=0}^{\infty} (\mathbf{B}_n)_{i,j} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)), \tag{10}$$

where $(\mathbf{B}_n)_{i,j}$ denotes the (i, j) -entry of \mathbf{B}_n . We denote by \mathcal{A} the class of continuous covariance functions K on $\mathbb{S}^d \times \{1, \dots, k\}$, for some positive integer k , which admit the representation (10). Apparently, any continuous function K belonging to the class \mathcal{A} implies that the matrix valued function \mathbb{C} having entries $C_{i,j}(\theta(\mathbf{x}_1, \mathbf{x}_2)) = K\left((\mathbf{x}_1, i), (\mathbf{x}_2, j)\right)$ will have a uniquely determined expansion as in (7). This fundamental connection is made explicit several times in the exposition of our results.

Some comments are in order. Since each matrix \mathbf{B}_n in (7) is positive semi-definite, by the Cholesky decomposition, there exists a lower triangular matrix \mathbb{D}_n such that $\mathbf{B}_n = \mathbb{D}_n \mathbb{D}'_n$, for $n = 0, 1, \dots$, where the diagonal entries of \mathbb{D}_n are nonnegative. Cholesky decomposition is unique if \mathbf{B}_n is positive definite, and it is equivalent to:

$$\mathbf{B}_{i,j;n} = \sum_{r=1}^k (\mathbb{D}_n)_{i,r} (\mathbb{D}_n)_{j,r}, \quad n = 0, 1, \dots \tag{11}$$

As a consequence of Schoenberg's representations (7) and (10), the matrices \mathbb{D}_n , $n = 0, 1, 2, \dots$, satisfy

$$\sum_{n=0}^{\infty} (\mathbb{D}_n)_{i,r} (\mathbb{D}_n)_{j,r} < \infty, \tag{12}$$

for every $i, j = 1, \dots, k$.

The primary goal here is to propose a nonparametric Bayesian model by assigning prior distributions to the (random) matrices \mathbf{B}_n in (7). This is equivalent to assign a prior distribution to the matrices \mathbb{D}_n , $n = 0, 1, 2, \dots$, which are therefore replaced by random matrices $\tilde{\mathbb{D}}_n$ that satisfy the following conditions:

- A.1** each matrix $\tilde{\mathbb{D}}_n$ is a lower triangular matrix whose diagonal elements are non-negative, namely $(\tilde{\mathbb{D}}_n)_{i,j} = 0$ for $1 \leq i < j \leq k$, and $(\tilde{\mathbb{D}}_n)_{j,j} \geq 0$, for $j = 1, \dots, k$;
- A.2** the random matrices $\tilde{\mathbb{D}}_0, \tilde{\mathbb{D}}_1, \tilde{\mathbb{D}}_2, \dots$ are independent;
- A.3** for each $n = 0, 1, \dots$, the \mathbb{R}_+ -valued random variables $(\tilde{\mathbb{D}}_n)_{j,j}$ ($j = 1, \dots, k$) are i.i.d. and the real-valued random variables $(\tilde{\mathbb{D}}_n)_{i,j}$ ($j = 1, \dots, i-1, j = 1, \dots, k$) are i.i.d.;
- A.4** $\mathbb{E}((\tilde{\mathbb{D}}_n)_{1,1})^2 = \mathbb{E}((\tilde{\mathbb{D}}_n)_{2,1})^2 = d_n$ where $\sum_{n=0}^\infty d_n < \infty$.

We now define

$$\tilde{\mathbf{B}}_{i,j;n} := \sum_{r=1}^k (\tilde{\mathbb{D}}_n)_{i,r} (\tilde{\mathbb{D}}_n)_{j,r}, \tag{13}$$

for $i, j = 1, \dots, k$. The following result shows that these conditions are indeed sufficient to ensure almost sure convergence of the sequence of random matrices $\{\tilde{\mathbf{B}}_n\}_{n=0}^\infty$, a fundamental ingredient for the uniquely determined expansion (7) to be well-defined.

Proposition 1. *Under Conditions A.1–A.4, we have*

$$\sum_{n=0}^\infty (\tilde{\mathbb{D}}_n)_{i,r} (\tilde{\mathbb{D}}_n)_{j,r} < \infty,$$

and therefore

$$\sum_{n=0}^\infty (\tilde{\mathbf{B}}_n)_{i,j} < \infty,$$

almost surely, for every $i, j = 1, \dots, k$.

Conditions A.1–A.4, in concert with Proposition 1, are necessary to construct a covariance function, \tilde{K} , random version of the covariance function defined at (10). We define it through

$$\tilde{K}(\mathbf{x}_1, \mathbf{x}_2; j_1, j_2) = \sum_{n=0}^\infty (\tilde{\mathbf{B}}_n)_{j_1, j_2} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)), \tag{14}$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$ and every $j_1, j_2 \in \{1, \dots, k\}$. This implicitly defines a random matrix-valued covariance function, $\tilde{\mathbb{C}}$, having expansion

$$\tilde{\mathbb{C}}(\theta) = \sum_{n=0}^\infty \tilde{\mathbf{B}}_n c_n(d, \cos(\theta)), \quad \theta \in [0, \pi], \tag{15}$$

for a converging sequence $\{\tilde{\mathbf{B}}_n\}_{n=0}^\infty$ with entries $(\tilde{\mathbf{B}}_n)_{i,j}$ determined through (14).

Apparently, the random nature of \tilde{K} and $\tilde{\mathbb{C}}$ are ensured provided those are, actually, a random variable, and a random matrix respectively. This is in turn ensured by the following result.

Proposition 2. *Let \mathcal{A} be equipped with the topology induced by the uniform (or sup) norm and consider the sigma-algebra generated by the open sets in such topology. Then the map $\tilde{K} : \Omega \rightarrow \mathcal{A}$ defined in Equation (14) is measurable. Therefore, \tilde{K} is a random variable.*

The final step in the Bayesian nonparametric construction specified through Conditions **A.1–A.4** is to establish the support of the random variable \tilde{K} under a specified topology. We choose to work under the topology of uniform convergence. To establish our next result, an additional condition is needed:

A.5 For each $n \geq 0$, let the support of $(\tilde{\mathbb{D}}_n)_{1,1}$ be $[0, \infty)$ with $\mathbb{P}((\tilde{\mathbb{D}}_n)_{1,1} = 0) = 0$ and let the support of $(\tilde{\mathbb{D}}_n)_{2,1}$ be \mathbb{R} .

With this additional assumption, we are ready to provide the core result of this section.

Theorem 3. *Let conditions **A.1–A.5** hold true. Then, the support of \tilde{K} in the uniform topology is \mathcal{A} .*

Theorem 3 shows that the support of the random matrix-valued covariance function, $\tilde{\mathbb{C}}$, is the whole class of matrix-valued positive definite functions defined by the series representation (7).

3.2. Moments, posteriors, and Lipschitz continuities

The previous section developed an approach for Bayesian nonparametric covariance matrix functions that have a well-defined support under the uniform topology. This section provides further inspection of the properties for this proposed framework. We start by exploring the first-order properties of the random covariance functions, \tilde{K} , defined in (14).

Proposition 4. *Let \tilde{K} be the random covariance function defined through (14). If conditions **A1–A4** hold true, then*

$$\mathbb{E}(\tilde{K}(\mathbf{x}_1, \mathbf{x}_2; j_1, j_2)) = \sum_{n=0}^{\infty} \mathbb{E}((\tilde{\mathbf{B}}_n)_{j_1, j_2}) c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)),$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$ and $j_1, j_2 \in \{1, \dots, k\}$.

We now provide a technical result needed to establish the properties of our proposed model’s posterior distribution.

Proposition 5. *Let conditions **A.1–A.5** hold true. Then, for every positive integer n , and for every finite system of pairwise distinct points $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{S}^d$, for any $(j_1, \dots, j_n) \in \{1, \dots, k\}^n$ and constants $c_1, \dots, c_n \in \mathbb{R}$, where $c_l \neq 0$, for some l , the following holds almost surely:*

$$\sum_{r=1}^n \sum_{l=1}^n c_r c_l \tilde{K}(\mathbf{x}_r, \mathbf{x}_l; j_r, j_l) > 0.$$

Rephrased, Proposition 5 shows that \tilde{K} is almost surely strictly positive definite. This will be useful later on.

We are now able to study the posterior properties of the proposed model. Suppose that we have observations $\mathbf{z} := (\mathbf{z}(\mathbf{x}_1), \dots, \mathbf{z}(\mathbf{x}_n))$ at points $\mathbf{x}_1, \dots, \mathbf{x}_n$ in \mathbb{S}^d sampled from a k -variate Gaussian process on \mathbb{S}^d . Therefore, \mathbf{z} is a $(k \times n)$ -dimensional matrix, and its (j, i) -entry, which we denote by $z_j(\mathbf{x}_i)$, is a realization of the random variable $Z_j(\mathbf{x}_i)$, for $j = 1, \dots, k$, $i = 1, \dots, n$. The unknown covariance function \tilde{K} in (14) is uniquely determined through the sequence of random matrices $\tilde{\mathbb{D}}_0, \tilde{\mathbb{D}}_1, \dots$ defined through (11), which we denote by $\tilde{\mathbb{D}}$.

Thus, conditional on $\tilde{\mathbb{D}}$, \mathbb{Z} is a k -variate stationary and isotropic Gaussian process with covariance matrix function $\tilde{\mathbb{C}}$. Specifying the probability distribution of $\tilde{\mathbb{D}}$ and the conditional distribution of \mathbb{Z} given $\tilde{\mathbb{D}}$ specifies the joint prior distribution of $\tilde{\mathbb{D}}$ and \mathbb{Z} . Because conditioning on observations $(\mathbb{Z}(\mathbf{x}_1), \dots, \mathbb{Z}(\mathbf{x}_n))$ updates the conditional distribution of the random pair $(\mathbb{Z}, \tilde{\mathbb{D}})$, we explore the conditional distribution of the parameter $\tilde{\mathbb{D}}$ given $(\mathbb{Z}(\mathbf{x}_1), \dots, \mathbb{Z}(\mathbf{x}_n))$, namely the posterior distribution of $\tilde{\mathbb{D}}$. To do this, we first define $\text{vec}(\mathbf{z})$ as the vectorization of the matrix \mathbf{z} , *i.e.*,

$$\text{vec}(\mathbf{z}) = (z_1(\mathbf{x}_1), z_2(\mathbf{x}_1), \dots, z_k(\mathbf{x}_1), z_1(\mathbf{x}_2), \dots, z_k(\mathbf{x}_n))',$$

and let mod denote the modulo operation.

Proposition 6. *The posterior $\mathbb{P}^{\mathbf{z}}$ of $\tilde{\mathbb{D}}$ exists, is unique, and is absolutely continuous with respect to the prior, with uniquely determined density evaluated at $\mathbb{D} = \{\mathbb{D}_0, \mathbb{D}_1, \dots\}$ equal to:*

$$\frac{f(\text{vec}(\mathbf{z}); \mathbb{D})}{M(\mathbf{z})}, \quad (16)$$

where $f(\mathbf{z}; \mathbb{D})$ is the kn -dimensional centred multivariate Gaussian density evaluated at $\text{vec}(\mathbf{z})$ with covariance matrix $\Sigma_n(\mathbb{D})$, where

$$\Sigma(\mathbb{D})_{r_1, r_2; n} := \sum_{n=0}^{\infty} \sum_{l=1}^k \mathbb{D}_{j(r_1), l; n} \mathbb{D}_{j(r_2), l; n} c_n(d, \theta(\mathbf{x}_{i(r_1)}, \mathbf{x}_{i(r_2)})), \quad (17)$$

for

$$i(r) = \begin{cases} \lfloor r/k \rfloor & \text{if } r/k \notin \mathbb{N} \\ \lfloor r/k \rfloor - 1 & \text{if } r/k \in \mathbb{N} \end{cases} \quad j(r) = \begin{cases} r \bmod k & \text{if } r \bmod k \neq 0 \\ k & \text{if } r \bmod k = 0 \end{cases}$$

and

$$M(\mathbf{z}) = \mathbb{E}(f(\text{vec}(\mathbf{z}); \mathbb{D})).$$

The matrix (17) is not singular thanks to Proposition 5. This ensures that the density (16) is well defined.

Importantly, we verify that the posterior $\mathbb{P}^{\mathbf{z}}$ distribution of the parameters varies smoothly as a function of \mathbf{z} . This relates to robustness to changes in the observed data. Indeed, this property ensures that small changes in the data

lead to small changes in the posterior distribution, and this is usually established through continuity in the Hellinger metric. If this holds, we say that the estimation task is *well-posed* (see [Stuart, 2010](#); [Dashti and Stuart, 2016](#)).

To address well-posedness of the posterior, recall the definition of the Hellinger distance between two probability measures, ν_1 and ν_2 , dominated by the same measure ν with Radon–Nikodym derivatives f_1 and f_2 , respectively:

$$d_H(\nu_1, \nu_2) = \left\{ \int \left(\sqrt{f_1} - \sqrt{f_2} \right)^2 d\nu \right\}^{1/2}. \tag{18}$$

We are now able to illustrate our next finding.

Theorem 7. *The posterior \mathbb{P}^z given in Proposition 6 is Lipschitz continuous with respect to the Hellinger distance in the data \mathbf{z} : for any $r > 0$, there exists a strictly positive constant $m(r)$ such that for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{k,n}$, with $\max\{\|\text{vec}(\mathbf{z}_1)\|, \|\text{vec}(\mathbf{z}_2)\|\} \leq r$,*

$$d_H(\mathbb{P}^{\mathbf{z}_1}, \mathbb{P}^{\mathbf{z}_2}) \leq m(r) \|\text{vec}(\mathbf{z}_1) - \text{vec}(\mathbf{z}_2)\|. \tag{19}$$

4. Approximations on L^p spaces

This section suggests a method of approximation of k -variate isotropic random fields \mathbb{Z} on \mathbb{S}^d , and studies the level of its accuracy in a L^p sense.

The following Karhunen-Loeve expansion, proved in [Ma \(2016\)](#), see also [Lu, Leonenko and Ma \(2020\)](#), plays a fundamental role in this section. Let $\{\mathbf{V}_n\}_{n=0}^\infty$ be a sequence of independent k -variate random vectors with $\mathbb{E}\mathbf{V}_n = \mathbf{0}$ and

$$\text{cov}(\mathbf{V}_n, \mathbf{V}_n) = \mathbb{E}(\mathbf{V}_n \mathbf{V}_n') = \frac{2n + d - 1}{d - 1} \mathbb{I}(k) =: \alpha_n^2 \mathbb{I}(k).$$

Let \mathbf{U} be a $(d+1)$ -dimensional random vector, uniformly distributed on \mathbb{S}^d which is independent of $\{\mathbf{V}_n\}_{n=0}^\infty$. Finally, let $\{\mathbf{A}_n\}_{n=0}^\infty$ be a sequence of $k \times k$ positive definite matrices. It can be easily shown that, if the series $\sum_{n=0}^\infty \mathbf{A}_n C_n^{(d-1)/2}(1)$ converges, then

$$\mathbb{Z}(\mathbf{x}) = \sum_{n=0}^\infty \mathbf{A}_n^{1/2} \mathbf{V}_n C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}), \quad \mathbf{x} \in \mathbb{S}^d, \tag{20}$$

is a k -variate isotropic random field on \mathbb{S}^d , with $\mathbf{0}$ -mean and covariance matrix function of the form

$$\mathbb{C}(\theta) = \sum_{n=0}^\infty \mathbf{A}_n C_n^{(d-1)/2}(\cos \theta), \quad \theta \in [0, \pi],$$

i.e., of the form (7) with $\mathbf{A}_n = \mathbf{B}_n / C_n^{(d-1)/2}(1)$.

Observe that the realizations of the random field defined in (20) will be constant across hyperplanes that intersect the sphere (i.e., constant across arcs over the spherical surface). The vector \mathbf{U} determines the orientation of these arcs. Thus, a more flexible variant of the random field constructed above, that preserves its mean and covariance functions, is given by

$$\mathbb{Z}(\mathbf{x}) = \sum_{n=0}^{\infty} \mathbf{A}_n^{1/2} \mathbf{V}_n C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n), \quad \mathbf{x} \in \mathbb{S}^d, \quad (21)$$

where $\mathbf{U}_0, \mathbf{U}_1, \dots$ is a sequence of independent random vectors, being uniformly distributed on \mathbb{S}^d , and being independent of the vectors $\mathbf{V}_0, \mathbf{V}_1, \dots$. The proof of this more general assertion follows the same arguments reported in Ma (2016) (we provide more details in Appendix A.1).

Remark 8. Notice that both expansions, (20) and (21), generate random fields with the desired covariance structure, however their finite-dimensional distributions are clearly non-Gaussian. A common strategy to achieve approximately Gaussian random fields is to overlap several independent realizations (properly rescaled). This procedure is justified by the central limit theorem (for more details, see Emery and Lantuéjoul, 2006 and Alegría, Emery and Lantuéjoul, 2020, and the references therein).

For $R \in \mathbb{N}$ we define the truncated series

$$\mathbb{Z}^R(\mathbf{x}) = \sum_{n=0}^R \mathbf{A}_n^{1/2} \mathbf{V}_n C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n), \quad \mathbf{x} \in \mathbb{S}^d, \quad (22)$$

that is the truncated version of the random field, \mathbb{Z} , whose expansion has been defined at (21). For justifying the approximation of \mathbb{Z} by \mathbb{Z}^R , we have to show that \mathbb{Z}^R converges to \mathbb{Z} , when $R \rightarrow \infty$ in several norms counting the risk simultaneously in stochastic and analytic way.

Precisely, the following result characterizes the rate of convergence of the error $\mathbb{Z} - \mathbb{Z}^R$ in L^p and \mathbb{P} -almost sure sense in terms of the decay of the sequence $\{\|\mathbf{A}_n\|_F\}_{n \in \mathbb{N}}$.

Theorem 9. *Let \mathbb{Z} be a k -variate isotropic zero mean Gaussian random field on \mathbb{S}^d given by (21). Assume that the trace of the matrix \mathbf{A} decays algebraically with order $d-1+\varepsilon$; i.e., there exist $c_0 > 0$ and $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$,*

$$\text{trace}(\mathbf{A}_n) \leq c_0 n^{-d+1-\varepsilon}, \quad (23)$$

for some $\varepsilon > 0$. Then, the series of truncated random fields $\{\mathbb{Z}^R\}_{R \in \mathbb{N}}$ converges to the random field \mathbb{Z} (either) in the following sense:

1. in $L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))$, and the truncated error is bounded by

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq cR^{-\varepsilon/2}, \quad (24)$$

for $R \geq N_0$, where $c = c(d, \varepsilon) > 0$;

2. in $L^p(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))$ for every $p > 0$. Moreover, there exists a constant $c = c(d, \varepsilon, p) > 0$ such that the truncated error to be bounded by

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^p(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq cR^{-\varepsilon/2}, \tag{25}$$

for $R \geq N_0$, where $c = c(d, \varepsilon) > 0$;

3. \mathbf{P} -almost surely, and the truncated error is asymptotically bounded by

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \leq R^{-\gamma}, \quad \mathbf{P}\text{-a.s.}, \tag{26}$$

for $\gamma < \varepsilon/2$.

4.1. Numerical illustrations

The truncated expansion (22) can be used to simulate zero-mean random fields with a prescribed second order dependency structure. The experiments developed in this section are the multivariate versions of similar numerical studies reported by [Lang and Schwab \(2015\)](#) for univariate random fields.

In this section, we restrict ourselves to the bivariate case, where the (i, j) th entry of the Schoenberg’s matrix \mathbf{B}_n is given by

$$(\mathbf{B}_n)_{i,j} = \sigma_{i,i}\sigma_{j,j}\rho_{i,j}(b_n)_{i,j}, \quad i, j = 1, 2.$$

Here, for each i, j , $\{(b_n)_{i,j}\}$ is a parametric sequence such that $\sum_{n=0}^{\infty} (b_n)_{i,j} = 1$ (see the examples below), whereas $\sigma_{i,i}^2$ is the variance of the i th component of the multivariate random field, and $\rho_{i,j}$ is the collocated correlation coefficient. Thus, we can write $\rho_{i,j}$ in the equation above because $\rho_{i,i} = 1$ by construction. Throughout, we assume that $\sigma_{i,i} = 1$, for $i = 1, 2$. Thus, both components of the random field will have unit variance. Our numerical findings are based on the following two special examples for the sequence $\{(b_n)_{i,j}\}$.

- \mathcal{F} family. [Bevilacqua, Diggle and Porcu \(2019\)](#) have recently proposed a bivariate structure on the basis of the so-called \mathcal{F} family of functions ([Alegría et al., 2019b](#)). Precisely, this model is defined as

$$(b_n^{\mathcal{F}})_{i,j} = \frac{B(\alpha_{ij}, \nu_{ij} + \tau_{ij})}{B(\alpha_{ij}, \nu_{ij})} \frac{(\tau_{ij})_n (\alpha_{ij})_n}{n! (\alpha_{ij} + \tau_{ij} + \nu_{ij})_n}, \quad i, j = 1, 2, \tag{27}$$

where $\tau_{i,j}$, $\alpha_{i,j}$ and $\nu_{i,j}$ are positive parameters, and $B(\cdot, \cdot)$ is the Beta function. We fix $\tau_{ij} = 1$, for all $i, j = 1, 2$, since we are essentially interested in the smoothness and range of the associated random field, which are governed by the parameters ν_{ij} and α_{ij} , respectively (see [Alegría et al., 2019b](#)). The technical condition

$$|\rho_{12}| \leq B(\alpha_{12}, \nu_{12}) / \sqrt{B(\alpha_{11}, \nu_{11})B(\alpha_{22}, \nu_{22})},$$

stated in [Bevilacqua, Diggle and Porcu \(2019\)](#), is fundamental to ensure the validity of this model. We consider the following scenarios:

- **Scenario F1.** A random field with each component having the same smoothness and range: $\nu_{ij} = 2$, $\alpha_{ij} = 4$, for all $i, j = 1, 2$. The cross correlation coefficient is $\rho_{12} = 0.85$, which means that both fields are strongly positively correlated.
- **Scenario F2.** Same scenario as in **F1**, but $\nu_{ij} = 4$, $\alpha_{ij} = 8$, for all $i, j = 1, 2$. The sample paths are expected to be smoother in comparison to Scenario I. The cross correlation coefficient is set to $\rho_{12} = -0.85$, which means that both fields are strongly negatively correlated.
- **Scenario F3.** A random field with all components having the same smoothness, but different values of α : $\nu_{ij} = 4$, for all $i, j = 1, 2$, and $\alpha_{11} = 8$, $\alpha_{22} = 4$, and $\alpha_{12} = 6$. Here, the range is smaller for the first random field, since an increase in α corresponds to decreasing the correlation range. The cross correlation coefficient is chosen as $\rho_{12} = 0.85$. Under this choice both fields are strongly positively correlated.
- **Scenario F4.** A random field with all components having the same range: $\alpha_{ij} = 8$, for all $i, j = 1, 2$, and different degrees of smoothness: $\nu_{11} = 2$, $\nu_{22} = 4$, and $\nu_{12} = 3$. Here, the second random field is smoother than the first one. The cross correlation coefficient is chosen as $\rho_{12} = 0.85$. Under this choice both fields are strongly positively correlated.
- **Multiquadric family.** The bivariate multiquadric model is characterized by the sequences (Ma, 2012; Bevilacqua, Diggle and Porcu, 2019)

$$(b_n^{\mathcal{M}})_{i,j} = (1 - \zeta_{ij})\zeta_{ij}^n, \quad i, j = 1, 2, \quad (28)$$

where $\zeta_{ij} \in (0, 1)$ are parameters. This model is admissible provided that (Bevilacqua, Diggle and Porcu, 2019)

$$|\rho_{12}| \leq \sqrt{(1 - \zeta_{11})(1 - \zeta_{22})}/(1 - \zeta_{12}).$$

Unlike the \mathcal{F} family, the members of the multiquadric class are infinitely differentiable at the origin, so they always induce random fields with smooth realizations. For this parametric class, we look at the following scenarios:

- **Scenario M1.** We consider $\zeta_{11} = 0.8$, $\zeta_{22} = 0.7$ and $\zeta_{12} = 0.65$, which indicates that the range of the first random field component is smaller. The cross correlation coefficient is $\rho_{12} = 0.65$, which means that both fields are positively correlated.
- **Scenario M2.** Same values for ζ_{ij} as in **M1**, but the cross correlation coefficient is $\rho_{12} = -0.65$, which means that both fields are negatively correlated.
- **Scenario M3.** We consider $\zeta_{11} = 0.5$, $\zeta_{22} = 0.4$ and $\zeta_{12} = 0.35$. Again, it indicates that the range of the first random field component

is smaller. The cross correlation coefficient is $\rho_{12} = 0.8$, which means that both fields are strongly positively correlated.

- **Scenario M4.** Same values for ζ_{ij} as in **M3**, but the cross correlation coefficient is $\rho_{12} = -0.8$, which means that both fields are strongly negatively correlated.

Figures 1 and 2 show simulated bivariate random fields on \mathbb{S}^2 , from the \mathcal{F} and multiquadric families, respectively, over a grid of size 600×300 , with $R = 200$. The simulations match the theoretical aspects in terms of smoothness, ranges, and cross-correlation.

We turn to a numerical validation of the truncation error derived above. We simulate random fields over 450 sites (these locations were generated through a uniform sampling over the spherical surface), under scenarios **F1–F4**. We focus on the \mathcal{F} family since it has a parameter that controls the asymptotic decay of the Schoenberg’s sequence. Indeed, the asymptotic behavior $(b_n^{\mathcal{F}})_{i,j} \sim n^{-(1+\nu_{i,j})}$ (Alegría et al., 2019b) implies that the condition (23) is satisfied with $\epsilon = \min\{\nu_{1,1}, \nu_{2,2}\}$. We take as the “true” random field the truncated expansion (22) with $R = 500$. We gradually truncate the expansion at different values of R and measure the L^2 discrepancy between the truncated and true fields. For each scenario, we repeat this experiment 100 times. Observe that the uniformly distributed spatial locations together with the independent repetitions of the experiment allow us to approximate the integrals involved in the L^2 error by means of a Monte Carlo argument. Figure 3 displays the results. We observe that the empirical error matches the theoretical one for each scenario (the error decays algebraically with order $R^{-\min\{\nu_{1,1}, \nu_{2,2}\}/2}$).

Although we have paid attention to the two parametric models described above, additional parametric families of Schoenberg’s coefficients can be employed, (see, e.g., Ma, 2012 and Guinness and Fuentes, 2016).

5. NCEP/NCAR bivariate flux reanalysis

In this section, we apply the proposed modeling framework presented in Section 3 to a bivariate meteorological dataset. Recent climatic changes have affected various atmospheric constituents and downward solar radiation flux (see, e.g., Forster et al., 2007; Wild, 2009). Downward solar radiation flux (DSRF throughout) is an important component of the Earth’s climate system, affecting, among other things, surface temperature and the hydrologic cycle. DSRF is affected by atmospheric pressures at various levels because these directly affect cloud cover.

In this data illustration, we consider a bivariate spatial data set from two 2019 NCEP/NCAR Flux Reanalyses (Kalnay et al., 1996): atmospheric pressure at high cloud bottom and DSRF averaged over 2019. Of course, by averaging over time, we average over possibly important temporal patterns in these data. We plot these data in Figure 4. These data show apparent negative correlation and strong spatial patterns.

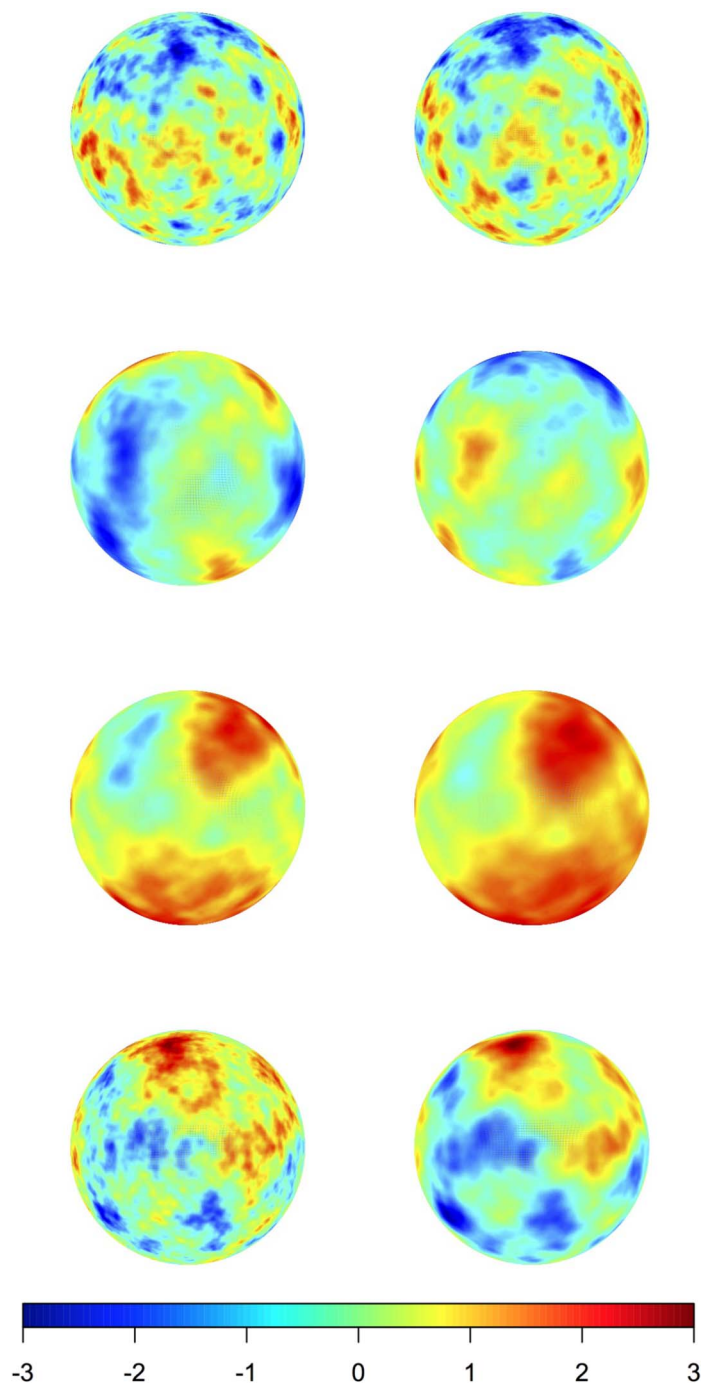


FIG 1. Simulated bivariate random fields on \mathbb{S}^2 , over a grid of size 600×300 , with $R = 200$, for Scenarios **F1–F4** (from top to bottom).

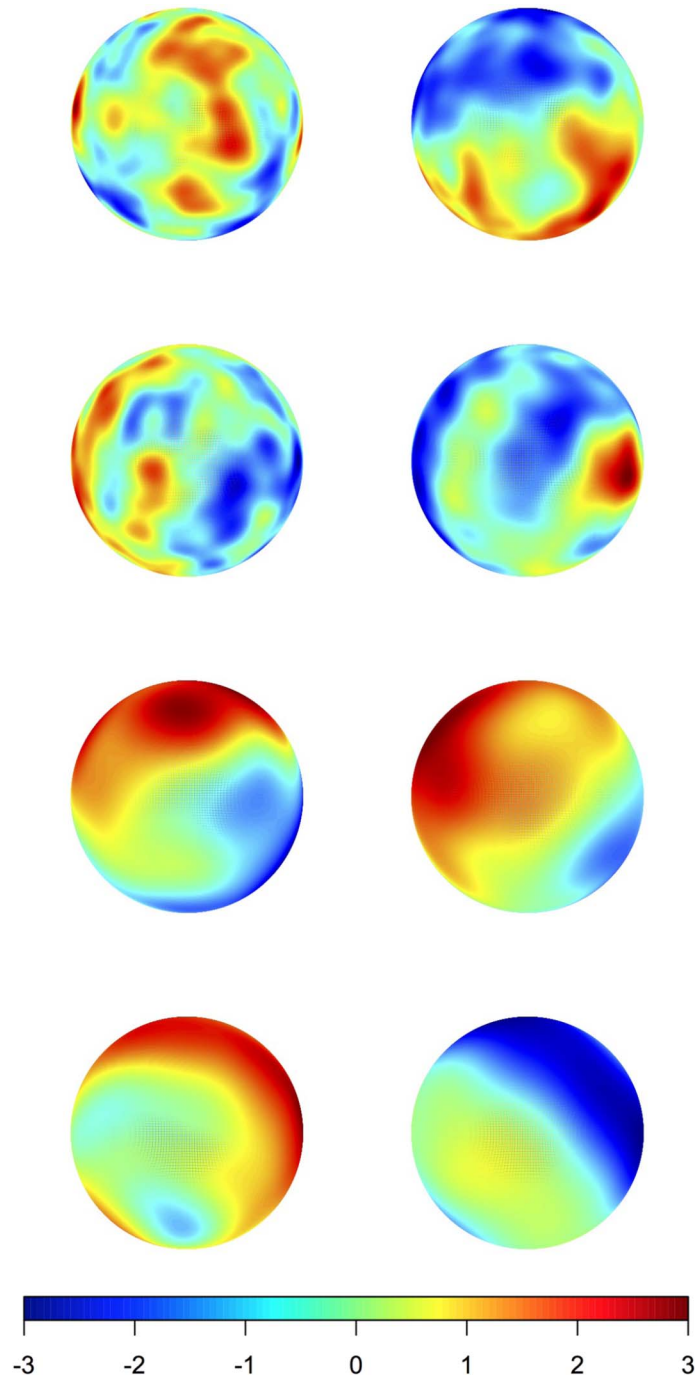


FIG 2. Simulated bivariate random fields on \mathbb{S}^2 , over a grid of size 600×300 , with $R = 200$, for Scenarios $M1$ – $M4$ (from top to bottom).

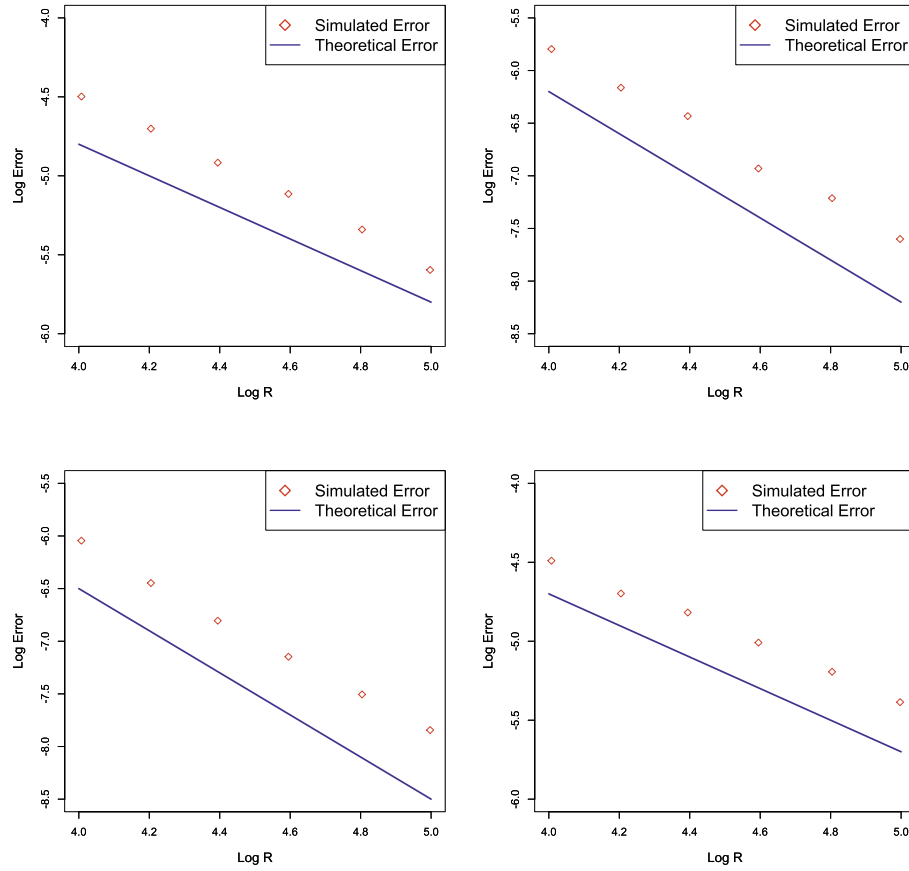


FIG 3. Convergence rates in terms of R in a logarithmic scale, for Scenarios **F1–F4** (from top-left to bottom-right). The average empirical convergence rates, based on 100 independent realizations, are given by the red squares, whereas the theoretical convergence rates are given by the solid blue lines.

For scalability, we thin these data to $n_s = 1568$ locations such that each remaining observation represents approximately the same area. Thus, we thin more aggressively in polar regions than in equatorial regions. For simplicity in modeling our data, we subtract off the mean and divide by the standard deviation of each quantity so that the sample mean and variance are 0 and 1, respectively.

To specify a covariance matrix function of the form (7), we use independent prior distributions for the elements of $\tilde{\mathbb{D}}_n$ that follow assumptions **A.1–A.5**. We assume a prior distribution of the form

$$\tilde{\mathbb{D}}_n = \begin{pmatrix} \sigma_{1n} & 0 \\ a_n & \sigma_{2n} \end{pmatrix}, \quad (29)$$

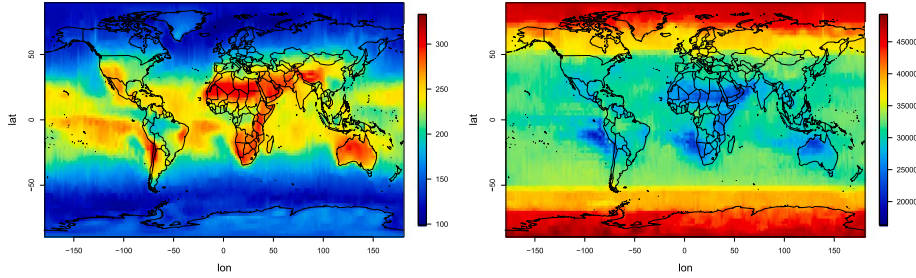


FIG 4. Annual average of two NCEP/NCAR reanalysis data sets from 2019: (Left) Atmospheric pressure at high cloud bottom (Right) DSRF.

where

$$\begin{aligned}
 \sigma_{1n} &\sim \text{Gamma} \left(1, \frac{\sqrt{2}(n+1)}{s_1} \right), \\
 a_n &\sim \mathcal{N} \left(0, \frac{s_1^2}{(n+1)^2} \right), \\
 \sigma_{2n} &\sim \text{Gamma} \left(1, \frac{\sqrt{2}(n+1)}{s_2} \right),
 \end{aligned}
 \tag{30}$$

for $n = 0, 1, 2, \dots$ and where \mathcal{N} denotes normal distribution. Here, we assume that all parameters are *a priori* independent; thus, Assumptions **A.2** and **A.3** are met. Assumption **A.1** is met because $\tilde{\mathbb{D}}_n$ is lower triangular. Because $\mathbb{E}(\sigma_{1n}^2) = \mathbb{E}(a_n^2) = s_1^2/(n+1)^2$ with $\sum_{n=0}^{\infty} s_1^2/(n+1)^2 = s_1^2\pi^2/6$, Assumption **A.4** is satisfied. Given the prior distributions proposed, Assumption **A.5** is also satisfied, although the Gamma is not always defined at 0.

Intuitively, the prior distribution provides increased shrinkage or penalization for higher-order terms, effectively pushing these parameters to 0. This type of prior distribution allows us to fit models with many Gegenbauer polynomial terms without many of the issues associated with including many parameters in a model (e.g., overfitting and high parameter uncertainty). Moreover, this prior distribution on $\tilde{\mathbb{D}}_n$ yields positive definite $\tilde{\mathbb{B}}_n$ which, when combined with Gegenbauer polynomials of $\cos(\theta)$, yield positive definite matrix functions with the great-circle distance as a metric.

We add two practical considerations. First, to account for variation over a very short distance in each outcome, we include nugget effects τ_1^2 and τ_2^2 . Second, although this model is only guaranteed to be positive definite using all terms in (7), for practical reasons, we focus on finite truncations of (7). We explore model fit as a function of the truncation number (*i.e.*, the number of included Gegenbauer polynomials included $n = 0, \dots, N$). The nugget effects τ_1^2 and τ_2^2 ensure positive definiteness. We assume *a priori* that these nugget terms come from a uniform distribution between 10^{-4} and 1. The lower bound of 10^{-4} assures that the joint covariance matrix is positive definite. The upper bound of 1 is set because we have scaled both variables to have a variance of 1.

Let \mathbf{Y}_1 be atmospheric pressures at high cloud bottom and \mathbf{Y}_2 be downward solar fluxes. The joint model for \mathbf{Y}_1 and \mathbf{Y}_2 is

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}, \sum_{n=0}^N \tilde{\mathbb{B}}_n \otimes \mathbb{C}_n + \text{diag}(\tau_1^2, \tau_2^2) \otimes \mathbb{I}(n_s) \right), \quad (31)$$

where \otimes is the Kronecker product, \mathbb{C}_n is a $n_s \times n_s$ matrix of the n th order Gegenbauer polynomial evaluated at the cosine of great circle distances between locations on the sphere, and $\text{diag}(\tau_1^2, \tau_2^2)$ is a 2×2 diagonal matrix. Because of the clear trends in latitude, the absolute value of latitude could be incorporated as a covariate in this model. In this illustration, however, we present a simple model featuring our nonparametric covariance model.

We fit the model using adaptive Markov chain Monte Carlo (MCMC) (see [Roberts and Rosenthal, 2009](#), for a discussion on adaptive MCMC). Specifically we do multivariate adaptive Metropolis-Hastings updates on the log-scale for σ_{1n} and σ_{2n} , and on the untransformed scale for a_n ([Haario, Saksman and Tamminen, 2001](#)). We update each of these as a unique step. We jointly propose candidate σ_{1n} for $n = 0, 1, \dots, N$ from a multivariate log-normal distribution using the natural log of the current values of σ_{1n} as μ . We propose values for σ_{2n} similarly. For a_n , we sample candidate values from a multivariate Normal centered at its current value. We tune the proposal covariance using the covariance of previous samples of a_n and the natural log of σ_{1n} and σ_{2n} , as well as adding a small number to the diagonal of these covariance matrices ([Haario, Saksman and Tamminen, 2001](#)). We update τ_1^2 and τ_2^2 univariately using an adaptive Metropolis-Hastings algorithm using a log-Normal random walk, where the candidate variance is turned to obtain an acceptance rate between 0.2 and 0.5, a rule presented in [Roberts, Gelman and Gilks \(1997\)](#).

We fit the model using $N = 2, 4, 6, 8, 10, 15, 20, 30$ Gegenbauer polynomials using $s_1 = s_2 = 1$ or $s_1 = s_2 = 100$ as prior distributions, giving 16 models in total. For each model, we calculate the mean log-likelihood and deviance information criterion (DIC). Using these values, we explore how model fit changes with the number of Gegenbauer polynomials included and depending on the variance of the prior distribution (*i.e.*, the degree of shrinkage or penalization).

In [Figure 5](#), we plot the log-likelihood averaged over all post-burn-in iterations of our MCMC sampler. While adding additional Gegenbauer polynomials increases the log-likelihood, the improvements diminish as the number of terms increase. We also plot the improvement in DIC (smaller is better) using the lower variance prior distribution in [Figure 5](#). We find that lower variance prior distribution has lower (better) DIC for all models. However, the relative improvement decreases as the number of terms increases, likely because the higher variance model still has heavy shrinkage for high-order polynomials. Here, we do not make any claims about how the number of Gegenbauer polynomials affects predictive performance.

For the model with $N = 30$ Gegenbauer polynomials with higher prior shrinkage $s_1 = s_2 = 1$, we summarize the behavior of the resulting covariance matrix function. At the estimated posterior mean, we plot $C_{11}(\theta)$, $C_{12}(\theta)$, and $C_{22}(\theta)$

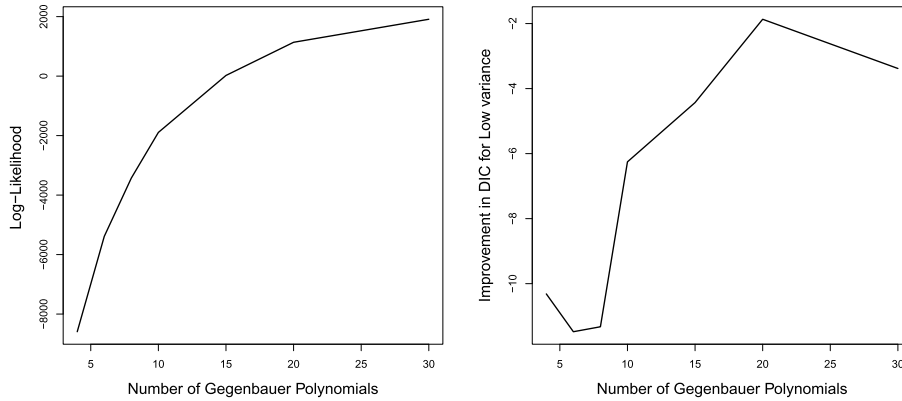


FIG 5. (Left) Mean log-likelihood as a function of the number of Gegenbauer polynomial terms used. (Right) Improvement in DIC using low variance (high shrinkage) prior distribution.

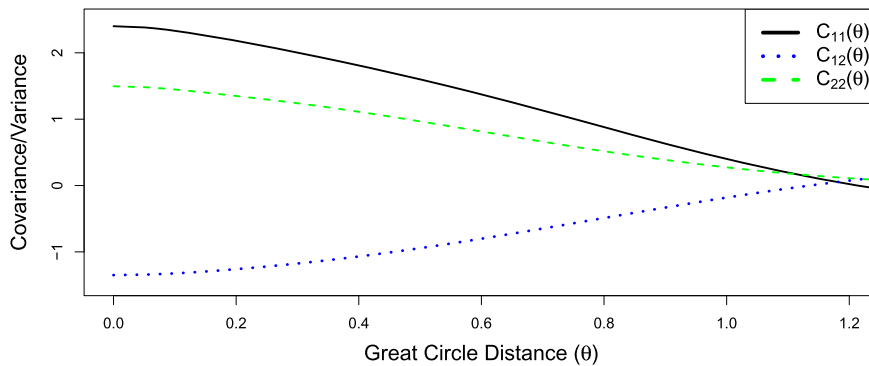


FIG 6. Covariance matrix function of θ at the posterior mean.

in Figure 6. Defining the effective range as the distance at which the covariance (or cross-covariance) is 0.05% its most extreme level, we find that the estimated effective ranges of $C_{11}(\theta)$, $C_{12}(\theta)$, and $C_{22}(\theta)$ are 7250, 6880, and 8060 km. These long effective ranges are unsurprising given the relatively smooth maps we observe in Figure 4.

In this section, we model two negatively correlated NCEP/NCAR reanalysis datasets, downward solar flux and atmospheric pressure and high cloud bottom, using a Bayesian nonparametric covariance matrix function constructed through Gegenbauer polynomials to model spatial cross-covariance. In our analysis, we focus on the degree of shrinkage imposed by our prior distribution and the number of Gegenbauer polynomials in the cross-covariance function. We only considered models with a fixed and finite number of Gegenbauer polynomials. While it is not possible to use an infinite number of Gegenbauer polynomial, future extensions of this model could allow the number of terms to be random. If

the number of components comes from a Poisson prior distribution, then model fitting would require reversible jump MCMC (Green, 1995) or a point process representation of the Poisson prior distribution (Stephens, 2000). Alternatively, scaling the prior distributions of $\tilde{\mathbb{D}}_n$ using stick-breaking weights (Sethuraman, 1994) may be a possible route to learning the number of polynomials needed.

One of the most fundamental limitations of our analysis is that our covariance model is isotropic and stationary. Therefore, the estimated bivariate cross-covariance function represents the estimated global covariance pattern as function of distance, not accounting for regional differences. However, because global meteorological data over the Earth often exhibit anisotropies (Sahoo, Guinness and Reich, 2019), the estimated covariance functions may not be representative of local covariance patterns. Thus, the results must be interpreted with this limitation in mind. We discuss some extensions of this framework in Section 6 that may extend the model presented here.

6. Conclusion

We have provided a Bayesian nonparametric modeling framework for estimating matrix valued covariance functions. Practical implementation of the estimation techniques proposed in this paper require truncation of the Schoenberg expansion. Thus, we studied approximations of k -variate isotropic random fields and their accuracy in a L^p sense, both theoretically and through simulation. In addition, we have explored the effect of the prior distribution and truncation on the model fit using a global climate reanalysis dataset.

We have worked under the assumption of geodesic isotropy, which might be questionable for some global data sets where local geographic conditions lead to anisotropy. In future analyses, we could explore approaches for incorporating anisotropy or nonstationarity within our proposed framework. Climate data also typically exhibit axial symmetry; that is, the covariance function is stationary by longitude and nonstationary by latitude (Jones, 1963; Stein, 2007). Extending this paper to the case of axial symmetry will require considerable technical work, as spectral representations for the axially symmetric case are extremely complicated.

In Section 5, we focus our analysis on the how the number of Gegenbauer polynomial terms and prior shrinkage affect fit and model complexity. In future work, one could explore more thoroughly the various prior formulations on matrices \mathbf{B}_n in (7) and their resulting posteriors. As a part of this study, it is possible that there are some analogs between specific prior distributions on \mathbf{B}_n and some closed-form representations presented in Guinness and Fuentes (2016) and Ma (2012) with prior distributions dependent on the specification of \mathbf{B}_n . In addition, we suggest exploring this framework in non-Gaussian settings (as in Terdik, 2015) and for tensor-valued quantities (see, e.g., Leonenko and Malyarenko, 2017).

One very promising application for this modeling framework is the study of Cosmic Microwave Background (CMB) radiation, including temperature (often

called \mathbf{T}) as a function of redshift and polarizations (both electric and magnetic modes, called \mathbf{E} and \mathbf{B} modes) (see, for example, [Fixsen, 2009](#); [Durrer, 2020](#)). These data provide an application where the flexible models presented here could provide insight into early universe cosmology ([Cabella and Marinucci, 2009](#); [Marinucci and Peccati, 2011](#)). However, because these datasets are generally large, one would likely need to consider low-rank (see, e.g., [Higdon, 1998](#); [Banerjee et al., 2008](#); [Cressie and Johannesson, 2008](#)) or directed acyclic graphical models (see [Katzfuss and Guinness, 2021](#), for a recent review).

Appendix A: Proofs

A.1. An alternative Karhunen-Loeve expansion

The goal of this appendix section is to justify that the expansion (21) produces a random field with the covariance function (7). Firstly, it is clear that such a random field has zero-mean (see [Ma, 2016](#) for details). Secondly, observe that the convergence of the series $\sum_{n=0}^{\infty} \mathbf{A}_n C_n^{(d-1)/2}(1)$ is fundamental to ensure that the random field (21) has finite variance, so the covariance function is well-defined. We proceed to calculate the covariance function of the random field defined in (21):

$$\begin{aligned} & \mathbb{E}(\mathbb{Z}(\mathbf{x}_1)\mathbb{Z}(\mathbf{x}_2)') \\ &= \mathbb{E}\left(\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbf{A}_n^{1/2}\mathbf{V}_n\mathbf{V}_m'(\mathbf{A}_m^{1/2})'C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{U}_n)C_m^{(d-1)/2}(\mathbf{x}_2'\mathbf{U}_m)\right) \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbf{A}_n^{1/2}\mathbb{E}(\mathbf{V}_n\mathbf{V}_m')(\mathbf{A}_m^{1/2})'\mathbb{E}\left(C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{U}_n)C_m^{(d-1)/2}(\mathbf{x}_2'\mathbf{U}_m)\right) \\ &= \sum_{n=0}^{\infty}\alpha_n^2\mathbf{A}_n\mathbb{E}\left(C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{U}_n)C_n^{(d-1)/2}(\mathbf{x}_2'\mathbf{U}_n)\right) \end{aligned}$$

In the previous equalities, we used the independence between the \mathbf{V}_n 's and \mathbf{U}_n 's, as well as the fact that $\mathbb{E}(\mathbf{V}_n\mathbf{V}_m') = \delta_{n=m}\alpha_n^2\mathbb{I}(k)$. Finally, the integral equation ([Ma, 2016](#))

$$\frac{1}{\omega_d}\int_{\mathbb{S}^d}C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{u})C_m^{(d-1)/2}(\mathbf{x}_2'\mathbf{u})d\mathbf{u}=\frac{\delta_{n=m}}{\alpha_n^2}C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{x}_2), \tag{32}$$

where $\omega_d = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$ is the surface area of \mathbb{S}^d , implies that

$$\mathbb{E}(\mathbb{Z}(\mathbf{x}_1)\mathbb{Z}(\mathbf{x}_2)')=\sum_{n=0}^{\infty}\mathbf{A}_n C_n^{(d-1)/2}(\mathbf{x}_1'\mathbf{x}_2),$$

as expected.

A.2. Proof of Proposition 1

If $j = l$, by condition **A.4** we have that

$$\sum_{n=0}^{\infty} \mathbb{E}((\tilde{\mathbb{D}}_n)_{j,r})^2 < \infty.$$

By the monotone convergence theorem, series and expectation can be swapped and therefore:

$$\mathbb{E} \left\{ \sum_{n=0}^{\infty} ((\tilde{\mathbb{D}}_n)_{j,r})^2 \right\} < \infty, \tag{33}$$

which in turn implies that $\sum_{n=0}^{\infty} ((\tilde{\mathbb{D}}_n)_{j,r})^2 < \infty$ almost surely since any random variable with finite expectation is finite almost surely. If $j \neq l$, for the same reason, it is sufficient to verify that

$$\sum_{n=0}^{\infty} \mathbb{E} \left\{ |(\tilde{\mathbb{D}}_n)_{j,r}(\tilde{\mathbb{D}}_n)_{l,r}| \right\} < \infty. \tag{34}$$

By condition **A.3**, $(\tilde{\mathbb{D}}_n)_{j,r}$ and $(\tilde{\mathbb{D}}_n)_{l,r}$ are independent, and therefore it is sufficient to verify that

$$\sum_{n=0}^{\infty} \mathbb{E} \left\{ |(\tilde{\mathbb{D}}_n)_{j,r}| \right\} \mathbb{E} \left\{ |(\tilde{\mathbb{D}}_n)_{l,r}| \right\} < \infty.$$

This is true by condition **A.4** and Jensen inequality. Indeed,

$$\mathbb{E} \left\{ |(\tilde{\mathbb{D}}_n)_{j,r}| \right\} \mathbb{E} \left\{ |(\tilde{\mathbb{D}}_n)_{l,r}| \right\} \leq \sqrt{\mathbb{E}((\tilde{\mathbb{D}}_n)_{j,r})^2 \mathbb{E}((\tilde{\mathbb{D}}_n)_{l,r})^2} = d_n. \quad \square$$

A.3. Proof of Proposition 2

Denote by Θ the collection of all arrays $\{(\mathbb{D}_n)_{j,l} : j, l = 1, \dots, k, n = 0, 1, 2, \dots\}$ such that $(\mathbb{D}_n)_{j,j} \geq 0$ for $j = 1, \dots, k$ and $(\mathbb{D}_n)_{j,l} = 0$ for $j = 1, \dots, l - 1, l = 2, \dots, k$ and $n = 0, 1, \dots$. Moreover, denote by $\tilde{\mathbb{D}}$ the random array $\{(\tilde{\mathbb{D}}_n)_{j,l} : j, l = 1, \dots, k, n = 0, 1, 2, \dots\}$. Let $\mathcal{T} : \Theta \rightarrow \mathcal{A}$ be the function defined through (10) and (11) that maps each element of Θ into the corresponding function K in \mathcal{A} . By (14) and (13), $\tilde{K} = \mathcal{T}(\tilde{\mathbb{D}})$ and by Lemma 10 in the Appendix, \mathcal{T} is continuous and therefore measurable. Hence, measurability of $\tilde{\mathbb{D}}$ implies that \tilde{K} is in turn measurable as a function from Ω into \mathcal{A} and therefore a random variable. \square

A.4. Proof of Theorem 3

In what follows, we let denote $\|f\|_{\infty} = \sup_E |f|$ for any bounded function $f : E \rightarrow \mathbb{R}$.

Lemma 10. For every K in \mathcal{A} , satisfying (10) and (11), and every $\varepsilon > 0$, there is an integer $m \geq 1$ and $\delta > 0$ such that:

$$\begin{aligned} & \left\{ \omega \in \Omega : \left\| \tilde{K}(\omega) - K \right\|_{\infty} \leq \varepsilon \right\} \\ & \supset \left\{ \omega \in \Omega : \left| (\tilde{\mathbb{D}}_n)_{j_1, j_2}(\omega) - (\mathbb{D}_n)_{j_1, j_2} \right| \leq \delta, j_1, j_2 = 1, \dots, k, n = 0, \dots, m \right\}. \end{aligned} \tag{35}$$

Proof. In the rest of the proof, let $\tilde{K}, \tilde{\mathbb{D}}_n, \tilde{\mathbf{B}}_n$ stand for $\tilde{K}(\omega), \tilde{\mathbb{D}}_n(\omega)$ and $\tilde{\mathbf{B}}_n(\omega)$, respectively. We aim at proving that for every $\varepsilon > 0$ there is an integer $m \geq 1$ and $\delta > 0$ such that if $\left| (\tilde{\mathbb{D}}_n)_{j_1, j_2} - (\mathbb{D}_n)_{j_1, j_2} \right| \leq \delta$ for every $j_1, j_2 = 1, \dots, k$, and every $n = 0, \dots, m$ then $\left\| \tilde{K} - K_0 \right\|_{\infty} \leq \varepsilon$.

To this aim, note that:

$$\begin{aligned} & \left\| \tilde{K} - K \right\|_{\infty} \\ & \leq \sup_{\mathbf{x}_1, \mathbf{x}_2, j_1, j_2} \sum_{n=0}^m |c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2))| \left| \sum_{r=1}^k (\tilde{\mathbb{D}}_n)_{j_1, r} (\tilde{\mathbb{D}}_n)_{j_2, r} - \sum_{r=1}^k (\mathbb{D}_n)_{j_1, r} (\mathbb{D}_n)_{j_2, r} \right| \\ & + \sup_{\mathbf{x}_1, \mathbf{x}_2, j_1, j_2} \left| \sum_{n=m+1}^{\infty} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)) \sum_{r=1}^k (\tilde{\mathbb{D}}_n)_{j_1, r} (\tilde{\mathbb{D}}_n)_{j_2, r} \right| \\ & + \sup_{\mathbf{x}_1, \mathbf{x}_2, j_1, j_2} \left| \sum_{n=m+1}^{\infty} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)) \sum_{r=1}^k (\mathbb{D}_n)_{j_1, r} (\mathbb{D}_n)_{j_2, r} \right| \end{aligned} \tag{36}$$

The series (10) converges uniformly by the Weierstrass M-test being $\|c_n(d, \cdot)\|_{\infty} \leq 1$. Clearly, the same is true for \tilde{K} . Hence, there is an integer $m \geq 1$ such that:

$$\begin{aligned} & \sup_{\mathbf{x}_1, \mathbf{x}_2, j_1, j_2} \left| \sum_{n=m+1}^{\infty} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)) \sum_{r=1}^k (\tilde{\mathbb{D}}_n)_{j_1, r} (\tilde{\mathbb{D}}_n)_{j_2, r} \right| \leq \varepsilon/3, \\ & \sup_{\mathbf{x}_1, \mathbf{x}_2, j_1, j_2} \left| \sum_{n=m+1}^{\infty} c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2)) \sum_{r=1}^k (\mathbb{D}_n)_{j_1, r} (\mathbb{D}_n)_{j_2, r} \right| \leq \varepsilon/3. \end{aligned}$$

Therefore, (36) implies:

$$\begin{aligned} & \left\| \tilde{K} - K \right\|_{\infty} \\ & \leq \sup_{j_1, j_2} \sum_{n=0}^m \left| \sum_{r=1}^k (\tilde{\mathbb{D}}_n)_{j_1, r} \{ (\tilde{\mathbb{D}}_n)_{j_2, r} - (\mathbb{D}_n)_{j_2, r} \} + \sum_{r=1}^k \{ (\tilde{\mathbb{D}}_n)_{j_1, r} - (\mathbb{D}_n)_{j_1, r} \} (\mathbb{D}_n)_{j_2, r} \right| \\ & + 2\varepsilon/3 \\ & \leq \delta \max_{j_1, j_2 \in \{1, \dots, k\}} \sum_{n=0}^m \left\{ \sum_{r=1}^k |(\tilde{\mathbb{D}}_n)_{j_1, r}| + \sum_{r=1}^k |(\mathbb{D}_n)_{j_2, r}| \right\} + 2\varepsilon/3 \end{aligned}$$

$$\leq \delta \max_{j_1, j_2 \in \{1, \dots, k\}} \sum_{n=0}^m \left\{ \sum_{r=1}^k |(\mathbb{D}_n)_{j_1, r}| + \sum_{r=1}^k |(\mathbb{D}_n)_{j_2, r}| \right\} + (m + 1) k \delta^2 + 2\varepsilon/3,$$

which is smaller than ε if δ is enough small and the proof is complete. □

Proof of Theorem 3. The event appearing on the right hand side of (35) is equal to

$$\bigcap_{n=0}^m \bigcap_{j_1=1}^k \bigcap_{j_2=1}^k \left\{ \omega \in \Omega : \left| (\tilde{\mathbb{D}}_n)_{j_1, j_2}(\omega) - (\mathbb{D}_n)_{j_1, j_2} \right| \leq \delta \right\},$$

where the event in the intersections are independent since the elements of the random array $\{(\tilde{\mathbb{D}}_n)_{j_1, j_2} : n = 0, 1, \dots, j_1, j_2 = 1, \dots, k\}$ are independent. Thanks to Lemma 10, the proof is completed noting each event in the intersection has positive probability. Indeed, for each $n = 0, 1, \dots$, we assume that the support of $(\tilde{\mathbb{D}}_n)_{1,1}$ is the whole $[0, \infty)$ and the support of $(\tilde{\mathbb{D}}_n)_{2,1}$ is the whole \mathbb{R} . □

A.5. Proof of Proposition 4

The proof is obtained by swapping series and expectation in (14). Such swapping is allowed since we can verify that:

$$\sum_{n=0}^{\infty} \mathbb{E} \left(\left| (\tilde{\mathbf{B}}_n)_{j_1, j_2} \right| \right) |c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2))| < \infty$$

Indeed, $|c_n(d, \theta(\mathbf{x}_1, \mathbf{x}_2))| \leq 1$, and

$$\sum_{n=0}^{\infty} \mathbb{E} \left(\left| (\tilde{\mathbf{B}}_n)_{j_1, j_2} \right| \right) \leq \sum_{r=1}^k \sum_{n=0}^{\infty} \mathbb{E} \left(\left| (\tilde{\mathbb{D}}_n)_{j_1, r} (\tilde{\mathbb{D}}_n)_{j_2, r} \right| \right),$$

which is finite due to (33) if $j_1 = j_2$ and due to (34) if $j_1 \neq j_2$.

A.6. Proof of Proposition 5

The thesis follows combining condition A.5 with Theorem 3.5 by Guella and Menegatto (2018).

A.7. Proof of Proposition 6

The proof is based on the application of Bayes’ theorem given in a general version by Schervish (1995) in its Theorem 1.31. In particular, such theorem ensures the existence of the conditional distribution of \mathbb{D} given $\text{vec}(\mathbb{Z}(\mathbf{x}_1), \dots, \mathbb{Z}(\mathbf{x}_n))$. The theorem can be applied since the conditional distribution of $\text{vec}(\mathbb{Z}(\mathbf{x}_1), \dots,$

$\mathbb{Z}(\mathbf{x}_n)$ given \mathbb{D} is dominated by a measure which does not depend on \mathbb{D} , namely the kn -dimensional Lebesgue measure. Indeed, such conditional distribution is Gaussian with density $f(\text{vec}(\mathbf{z}); \mathbb{D})$ and its covariance matrix $\Sigma_n(\mathbb{D})$ is positive definite since the covariance function \tilde{K} given in (14) is strictly positive definite almost surely due to Proposition 5. This also ensures that the multivariate Gaussian density $f(\text{vec}(\mathbf{z}); \mathbb{D})$ appearing in (16) is well defined. Moreover, being the multivariate Gaussian density strictly positive and bounded, $M(\mathbf{z})$ is always positive and finite so that the posterior is always uniquely determined and well defined. \square

A.8. Proof of Theorem 7

Proposition 6 reveals the form of the posterior of the parameter $\tilde{\mathbb{D}}$. For convenience, denote $W = \tilde{\mathbb{D}}$ and $w = \mathbb{D}$.

Given a matrix A $k \times k$, it is clear that the quadratic form $z^T Az$ is Lipschitz continuous being

$$|z^T Az| \leq \|A\| \|z\|^2, \tag{37}$$

where $\|A\|$ is the induced operator norm:

$$\|A\| = \sup\{\|Az\| : z \in \mathbb{R}^k \text{ with } \|z\| = 1\} = \sup\{\|Az\|/\|z\| : z \in \mathbb{R}^k \setminus \{0\}\}.$$

Taking $A = \Sigma_k^{-1}(W)$, this implies three facts.

Fix $r > 0$ and let $z_1, z_2 \in \mathcal{M}_{k,n}$ with $\max\{\|\text{vec}(z_1)\|, \|\text{vec}(z_2)\|\} \leq r$. First, for some constant $c_1 > 0$,

$$\left| \sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)} \right| \leq c_1(r) \|\text{vec}(z_1 - z_2)\|. \tag{38}$$

Indeed, using (37), for some constants $c_0, c_3(r) > 0$,

$$\begin{aligned} & \left| \sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)} \right| \\ &= c_0 e^{-\text{vec}(z_1)^T (\Sigma_k(W))^{-1} \text{vec}(z_1)/4} \times \\ & \times \left| 1 - e^{(\text{vec}(z_1)^T - \text{vec}(z_2)^T) (\Sigma_k(W))^{-1} (\text{vec}(z_1) - \text{vec}(z_2))/4} \right| \\ & \leq c_3(r) (\text{vec}(z_1)^T - \text{vec}(z_2)^T) (\Sigma_k(W))^{-1} (\text{vec}(z_1) - \text{vec}(z_2)) \\ & \leq c_3(r) \|\text{vec}(z_1) - \text{vec}(z_2)\|^2 \leq 2rc_3(r) \|\text{vec}(z_1) - \text{vec}(z_2)\|. \end{aligned} \tag{39}$$

Second, if $z \in \mathcal{M}_{k,n}$ and $\|\text{vec}(z)\| \leq r$, then

$$M(z) > k(r) \tag{40}$$

for some $k(r) > 0$. This is obtain applying directly (37) with $A = \Sigma_k^{-1}$.

Third, $\sqrt{M(z)}$ is Lipschitz as well, being

$$\left| \sqrt{M(z_1)} - \sqrt{M(z_2)} \right| \leq c_2(r) \|\text{vec}(z_1 - z_2)\|, \tag{41}$$

for some constant $c_2(r)$. This can be obtained since by a similar argument used in (39) one has that $|f(\text{vec}(z_1); W) - f(\text{vec}(z_2); W)| \leq c_1(r)\|z_1 - z_2\|$ so that $|M(z_1) - M(z_2)| \leq c_1(r)\|z_1 - z_2\|$, and moreover by (40),

$$\left| \sqrt{M(z_1)} - \sqrt{M(z_2)} \right| = \frac{|M(z_1) - M(z_2)|}{\sqrt{M(z_1)} + \sqrt{M(z_2)}} \leq \frac{|M(z_1) - M(z_2)|}{2k(r)}.$$

The inequalities (38), (40) and (41) allow to prove the theorem noting that if W is equipped with its prior distribution, then:

$$\begin{aligned} & d_H^2(\mathbb{P}^{z_1} - \mathbb{P}^{z_2}) \\ &= \mathbb{E} \left\{ \frac{\sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)}}{\sqrt{M(z_1)}} + \sqrt{f(\text{vec}(z_2); W)} \left(\frac{1}{\sqrt{M(z_1)}} - \frac{1}{\sqrt{M(z_2)}} \right) \right\}^2 \\ &\leq \mathbb{E} \left\{ \frac{\sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)}}{\sqrt{M(z_1)}} \right\}^2 + \left(\frac{1}{\sqrt{M(z_1)}} - \frac{1}{\sqrt{M(z_2)}} \right)^2 \mathbb{E} \{f(\text{vec}(z_2); W)\} \\ &\leq \frac{1}{M(z_1)} \mathbb{E} \left\{ \sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)} \right\}^2 + \left(\frac{(\sqrt{M(z_1)} - \sqrt{M(z_2)})^2}{M(z_1)M(z_2)} \right). \end{aligned}$$

In order to understand the first inequality, note that $f(\text{vec}(z_1); W) \leq f(\text{vec}(z_2); W)$ implies that $M(z_1) \leq M(z_2)$ and therefore $(\sqrt{f(\text{vec}(z_1); W)} - \sqrt{f(\text{vec}(z_2); W)})(M(z_2) - M(z_1)) \leq 0$ for any pair (z_1, z_2) . \square

A.9. Proof of Theorem 9

By (2) and (3) we extract

$$\begin{aligned} & \|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)}^2 \\ &= \langle \mathbb{Z} - \mathbb{Z}^R, \mathbb{Z} - \mathbb{Z}^R \rangle_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \\ &= \sum_{n, m > R} \langle \mathbf{A}_n^{1/2} \mathbf{V}_n C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n), \mathbf{A}_m^{1/2} \mathbf{V}_m C_m^{(d-1)/2}(\mathbf{x}'\mathbf{U}_m) \rangle_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \\ &= \sum_{n, m > R} \int_{\mathbb{S}^d} \langle \mathbf{A}_n^{1/2} \mathbf{V}_n C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n), \mathbf{A}_m^{1/2} \mathbf{V}_m C_m^{(d-1)/2}(\mathbf{x}'\mathbf{U}_m) \rangle_F d\mu(\mathbf{x}) \\ &= \sum_{n, m > R} \text{trace}(\mathbf{A}_n^{1/2} \mathbf{V}_n \mathbf{V}_m' (\mathbf{A}_m^{1/2})') \int_{\mathbb{S}^d} C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n) C_m^{(d-1)/2}(\mathbf{x}'\mathbf{U}_m) d\mu(\mathbf{x}) \end{aligned}$$

By Fubini-Tonelli Theorem, the mixed-norm (5) takes the form

$$\begin{aligned} & \|\mathbb{Z} - \mathbb{Z}^R\|_{L^2\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k)}^2 = \mathbb{E}(\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)}^2) \\ &= \sum_{n, m > R} \text{trace}(\mathbf{A}_n^{1/2} \mathbb{E}(\mathbf{V}_n \mathbf{V}_m') (\mathbf{A}_m^{1/2})') \\ & \quad \times \mathbb{E} \left(\int_{\mathbb{S}^d} C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n) C_m^{(d-1)/2}(\mathbf{x}'\mathbf{U}_m) d\mu(\mathbf{x}) \right) \end{aligned}$$

$$= \sum_{n>R} \alpha_n^2 \text{trace}(\mathbf{A}_n) \mathbb{E} \left(\int_{\mathbb{S}^d} C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n) C_n^{(d-1)/2}(\mathbf{x}'\mathbf{U}_n) d\mu(\mathbf{x}) \right)$$

where for the second equality we used the independence between the \mathbf{V}_n 's and \mathbf{U}_n 's, whereas for the third equality we used the fact that $\mathbb{E}(\mathbf{V}_n \mathbf{V}'_m) = \delta_{n=m} \alpha_n^2 \mathbb{I}(k)$. The identity (32) implies that

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}^2 = \sum_{n>R} \text{trace}(\mathbf{A}_n) C_n^{(d-1)/2}(1).$$

Condition (23), coupled with the relationship $C_n^{(d-1)/2}(1) = \binom{n+d-2}{n} \sim n^{d-2}$, allows us to conclude

$$\begin{aligned} \|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}^2 &\leq c \sum_{n>R} n^{-1-\varepsilon} \\ &\leq c \int_R^\infty \frac{dx}{x^{1+\varepsilon}} = cR^{-\varepsilon}, \end{aligned}$$

since $\varepsilon > 0$.

To show (25), we begin with the case $p < 2$. Applying the Hölder inequality on a probability space for the index $2/p > 1$ directly gives

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^p(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq \|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq cR^{-\varepsilon/2}.$$

We consider now $p > 2$. We can choose $\nu = \nu_p \in \mathbb{N}$ such that $2(\nu - 1) < p \leq 2\nu$. Similarly to the case $p < 2$, the Hölder inequality for the index $2\nu/p$ yields

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^p(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq \|\mathbb{Z} - \mathbb{Z}^R\|_{L^{2\nu}(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}. \tag{42}$$

By (Da Prato and Zabczyk, 1992, Corollary 2.17) there exists a constant $c = c_\nu > 0$ such that

$$\|\mathbb{Z} - \mathbb{Z}^R\|_{L^{2\nu}(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))} \leq c_\nu \|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}. \tag{43}$$

The combination of (42), (43) and (24) leads to (25).

For the last claim, we have to prove that

$$\mathbf{P}(\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \geq R^{-\gamma}, \text{ for infinity many } R \in \mathbb{N}) = 0. \tag{44}$$

This will be a consequence of the Borel-Cantelli lemma (Achim, 2008, Theorem 2.7). It suffices to prove that

$$\sum_{R=1}^\infty \mathbf{P}(\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \geq R^{-\gamma}) < \infty. \tag{45}$$

Let $p > 0$. By Chebyshev's inequality, (25) and using the mixed-norm (4), we derive

$$\mathbf{P}(\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)} \geq R^{-\gamma}) \leq R^{\gamma p} \mathbb{E}(\|\mathbb{Z} - \mathbb{Z}^R\|_{L^2(\mathbb{S}^d; \mathbb{R}^k)}^p)$$

$$\begin{aligned}
&= R^{\gamma p} \|\mathbb{Z} - \mathbb{Z}^R\|_{L^p(\Omega, L^2(\mathbb{S}^d; \mathbb{R}^k))}^p \\
&\leq cR^{-(\varepsilon/2-\gamma)p}.
\end{aligned} \tag{46}$$

We choose now $p > 1/(\varepsilon/2 - \gamma)$, this is allowed since $\gamma < \varepsilon/2$. For this choice we obtain

$$\sum_{R=1}^{\infty} R^{-(\varepsilon/2-\gamma)p} < \infty,$$

which together with (46) leads to (45) and completes the proof. \square

Acknowledgments

Alfredo Alegría and Emilio Porcu are supported by the National Agency for Research and Development, Chile, through grants CONICYT-ANID/FONDECYT/INICIACIÓN/No. 11190686 (A. Alegría) and CONICYT-ANID/FONDECYT/REGULAR/No. 1170290 (E. Porcu).

References

- ACHIM, K. (2008). *Probability theory. A comprehensive course*. Universitext. Springer-Verlag London, Ltd., London. [MR2372119](#)
- ALEGRÍA, A. (2020). Cross-dimple in the cross-covariance functions of bivariate isotropic random fields on spheres. *Stat* **9** e301. [MR4156481](#)
- ALEGRÍA, A., EMERY, X. and LANTUÉJOUL, C. (2020). The turning arcs: a computationally efficient algorithm to simulate isotropic vector-valued Gaussian random fields on the d -sphere. *Statistics and Computing* **30** 1403–1418. [MR4142746](#)
- ALEGRÍA, A., PORCU, E., FURRER, R. and MATEU, J. (2019a). Covariance functions for multivariate Gaussian fields. *Stochastic Environmental Research Risk Assessment* **33** 1593–1608.
- ALEGRÍA, A., CUEVAS, F., DIGGLE, P. J. and PORCU, E. (2019b). A new class of covariance functions of random fields on spheres. *Technical Report, University of Aalborg*.
- BANERJEE, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics* **61** 617–625. [MR2140936](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 825–848. [MR2523906](#)
- BEVILACQUA, M., DIGGLE, P. and PORCU, E. (2019). Families of covariance functions for bivariate random fields over spheres. *Spatial Statistics*. Under Second Review. [MR4181144](#)
- CABELLA, P. and MARINUCCI, D. (2009). Statistical challenges in the analysis of cosmic microwave background radiation. *The Annals of Applied Statistics* **3** 61–95. [MR2668700](#)

- CASTILLO, I., KERKYACHARIAN, G. and PICARD, D. (2014). Thomas Bayes' walk on manifolds. *Probab. Theory Related Fields* **158** 665–710. [MR3176362](#)
- CHILES, J.-P. and DELFINER, P. (2009). *Geostatistics: modeling spatial uncertainty* **497**. John Wiley & Sons. [MR2850475](#)
- CHOPIN, N., ROUSSEAU, J. and LISEO, B. (2013). Computational aspects of Bayesian spectral density estimation. *Journal of Computational and Graphical Statistics* **22** 533–557. [MR3173730](#)
- COMBES, V., HORMÁZABAL, S. and DI LORENZO, E. (2017). Interannual variability of the subsurface eddy field in the Southeast Pacific. *Journal of Geophysical Research-Oceans* **120** 2769–2783.
- CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 209–226. [MR2412639](#)
- DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic equations in infinite dimensions. Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. [MR1207136](#)
- DAI, F. and XU, Y. (2013). *Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer Monographs in Mathematics*. Springer, New York. [MR3060033](#)
- DASHTI, M. and STUART, A. (2016). The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification* (R. Ghanem, D. Higdon and H. Owhadi, eds.) 1–118. [MR3838441](#)
- DETTE, H., KONSTANTINOVA, M., SCHORNING, K. and GÖSMANN, J. (2019). Optimal designs for regression with spherical data. *Electronic Journal of Statistics* **13** 361–390. [MR3910487](#)
- DI LORENZO, E., COMBES, V., KEISTER, J., STRUB, P., ANDREW, T., PETER, F., MARCK, O., FURTADO, J. ., BRACCO, A., BOGRAD, S., PETERSON, W., SCHWING, F., TAGUCHI, B., HORMÁZABAL, S. and PARADA, C. (2014). Synthesis of Pacific Ocean Climate and ecosystem dynamics. *Oceanography* **26** 68–81.
- DU, J., MA, C. and LI, Y. (2013). Isotropic variogram matrix functions on spheres. *Math. Geosci.* **45** 341–357. [MR3107159](#)
- DUAN, J. A., GUINDANI, M. and GELFAND, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika* **94** 809–825. [MR2416794](#)
- DURRER, R. (2020). *The cosmic microwave background*. Cambridge University Press.
- EDWARDS, M., CASTRUCCIO, S. and HAMMERLING, D. (2019). A multivariate Global Spatio-Temporal Stochastic Generator for Climate Ensembles. *Journal of Agricultural, Biological and Environmental Sciences*. [MR3996454](#)
- EMERY, X. and LANTUÉJOUL, C. (2006). Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences* **32** 1615–1628.
- EMERY, X. and PORCU, E. (2019). Simulating isotropic vector-valued Gaussian random fields on the sphere through finite harmonics approximations. *Stochastic Environmental Research and Risk Assessment* **33** 1659–1667.
- FIXSEN, D. (2009). The temperature of the cosmic microwave background. *The*

- Astrophysical Journal* **707** 916.
- FORSTER, P., RAMASWAMY, V., ARTAXO, P., BERNTSEN, T., BETTS, R., FAHEY, D. W., HAYWOOD, J., LEAN, J., LOWE, D. C. and MYHRE, G. (2007). Changes in atmospheric constituents and in radiative forcing. Chapter 2. In *Climate Change 2007. The Physical Science Basis*.
- GELFAND, A. E., KOTTAS, A. and MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100** 1021–1035. [MR2201028](#)
- GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. [MR3102554](#)
- GORSKI, K. M., HIVON, E., BANDAY, A. J., WANDEL, B. D., HANSEN, F. K., REINECKE, M. and BARTELMANN, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal* **622** 759.
- GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. [MR1380810](#)
- GUELLA, J. C. and MENEGATTO, V. A. (2018). Schoenberg’s theorem for positive definite functions on products: a unifying framework. *J. Fourier Anal Appl.* To appear. [MR3977123](#)
- GUINNESS, J. and FUENTES, M. (2016). Isotropic covariance functions on spheres: some properties and modeling considerations. *Journal of Multivariate Analysis* **143** 143–152. [MR3431424](#)
- HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. [MR1828504](#)
- HANNAN, E. J. (2009). *Multiple Time Series*. *Wiley Series in Probability and Statistics*. Wiley. [MR0279952](#)
- HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics* **5** 173–190.
- HOLBROOK, A., LAN, S., VANDENBERG-RODES, A. and SHAHBABA, B. (2018). Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: with application to Bayesian spectral density estimation. *Journal of Statistical Computation and Simulation* **88** 982–1002. [MR3750634](#)
- JONA-LASINIO, G., GELFAND, A. and JONA-LASINIO, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics* **6** 1478–1498. [MR3058672](#)
- JONES, R. H. (1963). Stochastic processes on a sphere. *Ann. Math. Statist.* **34** 213–218. [MR0170378](#)
- KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G. and WOOLLEN, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77** 437–472.
- KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. [MR4194207](#)
- KERKYACHARIAN, G., OGAWA, S., PETRUSHEV, P. and PICARD, D. (2018). Regularity of Gaussian processes on Dirichlet spaces. *Constr. Approx.* **47**

- 277–320. [MR3769279](#)
- LANG, A. A. and SCHWAB, C. (2015). Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.* **25** 3047–3094. [MR3404631](#)
- LEONENKO, N. and MALYARENKO, A. (2017). Matérn class tensor-valued random fields and beyond. *Journal of Statistical Physics* **168** 1276–1301. [MR3691251](#)
- LU, T., LEONENKO, N. and MA, C. (2020). Series representations of isotropic vector random fields on balls. *Statist. Probab. Lett.* **156** 108583, 7. [MR3996834](#)
- LU, T. and MA, C. (2020). Isotropic covariance matrix functions on compact two-point homogeneous spaces. *J. Theoret. Probab.* **33** 1630–1656. [MR4125969](#)
- MA, C. (2012). Stationary and isotropic vector random fields on spheres. *Math. Geosci.* **44** 765–778. [MR2956272](#)
- MA, C. (2015). Isotropic covariance matrix functions on all spheres. *Math. Geosci.* **47** 699–717. [MR3368160](#)
- MA, C. (2016). Stochastic representations of isotropic vector random fields on spheres. *Stoch. Anal. Appl.* **34** 389–403. [MR3488255](#)
- MARINUCCI, D. and PECCATI, G. (2011). *Random fields on the sphere: representation, limit theorems and cosmological applications* **389**. Cambridge University Press. [MR2840154](#)
- MASTRANTONIO, G., LASINIO, G. J., POLLICE, A., CAPOTORTI, G., TEODONIO, L., GENOVA, G. and BLASI, C. (2019). A hierarchical multivariate spatio-temporal model for clustered climate data with annual cycles. *Annals of Applied Statistics* **13** 797–823. [MR3963553](#)
- MÜELLER, P., QUINTANA, F. A. and PAGE, G. (2018). Nonparametric Bayesian inference in applications. *Statistical Methods & Applications* **27** 175–206. [MR3807363](#)
- NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. and SAIN, S. (2015). A multi-resolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* **24** 579–599. [MR3357396](#)
- PORCU, E., ALEGRÍA, A. and FURRER, R. (2018). Modeling spatially global and temporally evolving data. *International Statistical Review* **86** 344–377. [MR3852415](#)
- PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association* **111** 888–898. [MR3538713](#)
- PORCU, E., BISSIRI, P. G., TAGLE, F. and QUINTANA, F. (2019). Nonparametric Bayesian Modeling and Estimation of Spatial Covariance Functions for Global Data. Technical Report.
- REICH, B. J. and FUENTES, M. (2012). Nonparametric Bayesian models for a spatial covariance. *Statistical Methodology* **9** 265–274. Special Issue on Astrostatistics + Special Issue on Spatial Statistics.
- REINSEL, G., TIAO, G., WANG, M., LEWIS, R. and NYCHKA, D. (1981). Sta-

- tistical analysis of stratospheric ozone data for the detection of trends. *Atmospheric Environment (1967)* **15** 1569–1577.
- ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7** 110–120. [MR1428751](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* **18** 349–367. [MR2749836](#)
- SAHOO, I., GUINNESS, J. and REICH, B. J. (2019). A test for isotropy on a sphere using spherical harmonic functions. *Statistica Sinica* **29** 1253–1276. [MR3932517](#)
- SCHERVISH, M. J. (1995). *Theory of statistics. Springer Series in Statistics.* Springer-Verlag, New York. [MR1354146](#)
- SCHMIDT, A. M. and O'HAGAN, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 743–758. [MR1998632](#)
- SCHOENBERG, I. J. (1942). Positive definite functions on spheres. *Duke Math. J.* **9** 96–108. [MR0005922](#)
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* 639–650. [MR1309433](#)
- SHIROTA, S. and GELFAND, A. E. (2017). Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics* **11** 481–503. [MR3693535](#)
- STEIN, M. L. (2007). Spatial variation of total column ozone on a global scale. *The Annals of Applied Statistics* **1** 191–210. [MR2393847](#)
- STEPHENS, M. (2000). Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. *Annals of Statistics* 40–74. [MR1762903](#)
- STUART, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta Numerica* **19** 451–559. [MR2652785](#)
- TERDIK, G. (2015). Angular spectra for non-Gaussian isotropic fields. *Brazilian Journal of Probability and Statistics* **29** 833–865. [MR3397396](#)
- WANG, F. and GELFAND, A. E. (2014). Modeling space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association* **109** 1565–1580. [MR3293610](#)
- WHITE, P. A. and PORCU, E. (2019). Nonseparable covariance models on circles cross time: A study of Mexico City ozone. *Environmetrics* **30** e2558. [MR3999535](#)
- WILD, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres* **114**.
- YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results.* Springer, New York. [MR0893393](#)
- ZHENG, Y., ZHU, J. and ROY, A. (2010). Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika* **97** 238–245. [MR2594432](#)