

# Nonparametric Bayesian Modeling and Estimation of Spatial Correlation Functions for Global Data

Emilio Porcu<sup>\*,†</sup>, Pier Giovanni Bissiri<sup>‡</sup>, Felipe Tagle<sup>§</sup>, Rubén Soza<sup>¶,||</sup>,  
and Fernando A. Quintana<sup>\*\*††</sup>

**Abstract.** We provide a nonparametric spectral approach to the modeling of correlation functions on spheres. The sequence of Schoenberg coefficients and their associated covariance functions are treated as random rather than assuming a parametric form. We propose a stick-breaking representation for the spectrum, and show that such a choice spans the support of the class of geodesically isotropic covariance functions under uniform convergence. Further, we examine the first order properties of such representation, from which geometric properties can be inferred, in terms of Hölder continuity, of the associated Gaussian random field. The properties of the posterior, in terms of existence, uniqueness, and Lipschitz continuity, are then inspected. Our findings are validated with MCMC simulations and illustrated using a global data set on surface temperatures.

**Keywords:** correlation function, great-circle distance, mean square differentiability, nonparametric Bayes, spheres.

## 1 Introduction

There has been an increasing interest in the modeling, inference and prediction of random processes defined continuously over a large portion or the entire planet Earth. Global data sets have become ubiquitous, and along with their increasing complexity, new scientific questions and challenges have emerged involving several branches of applied mathematics, statistics, computer sciences and machine learning.

### 1.1 Context and State of the Art

This paper deals with nonparametric Bayesian modeling, inference and prediction, for Gaussian random fields defined continuously over the two-dimensional sphere embedded in the three dimensional Euclidean space. Although more sophisticated settings may be considered (see, *e.g.*, Lindgren et al., 2011, with their approach on manifolds),

---

\*Trinity College Dublin, [porcue@tcd.ie](mailto:porcue@tcd.ie)

†Millennium Nucleus Center for the Discovery of Structures in Complex Data

‡Università di Bologna, [piergiovanni.bissiri@unibo.it](mailto:piergiovanni.bissiri@unibo.it)

§University of Notre Dame, [ftagleso@nd.edu](mailto:ftagleso@nd.edu)

¶Pontificia Universidad Católica de Chile, [rsozac@mat.uc.cl](mailto:rsozac@mat.uc.cl)

||Millennium Nucleus Center for the Discovery of Structures in Complex Data

\*\*Pontificia Universidad Católica de Chile, [quintana@mat.uc.cl](mailto:quintana@mat.uc.cl)

††Millennium Nucleus Center for the Discovery of Structures in Complex Data

we choose the sphere as the reference domain for the stochastic processes of interest as it is a reasonable approximation to the true geometry of the Earth (Diggle and Ribeiro, 2007). We take explicit advantage of the assumption of Gaussianity, which implies that the underlying process is uniquely characterized by its second order properties: the mean, and the covariance function (Stein, 1999). More formally, let the unit sphere  $\mathbb{S}^2$  be defined as the set of points in  $\mathbb{R}^3$  with unit Euclidean distance. We consider Gaussian random fields  $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{S}^2\}$ , with constant mean, and with covariance function  $C(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \text{corr}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\}$ , for  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^2$ . Here,  $\sigma^2$  is the variance and *corr* denotes correlation function. In this paper, estimation of the variance plays a minor role, and most attention will be given to the correlation function.

In particular, we focus on covariance functions that are geodesically isotropic (Porcu et al., 2018a): the function  $C$  above depends exclusively on the geodesic distance, that is, the shortest arc joining two points on the spherical shell. Mathematical details will be given subsequently. The nonparametric Bayesian approach to the function  $C$  has three basic ingredients: (a) the function  $C$  does not belong to any parametric family and is instead treated as an unknown function, to be estimated from data; (b) since  $C$  must be positive definite (details are given below) it is convenient to resort to spectral techniques, for which the mathematical restrictions are less severe; (c) a prior needs to be specified for the corresponding spectral density.

To the knowledge of the authors, random fields on spheres have not been considered so far in the Bayesian nonparametric literature, unlike random fields on planar surfaces: see, among others, Schmidt and O’Hagan (2003); Gelfand et al. (2005); Duan et al. (2007); Zheng et al. (2010); Reich and Fuentes (2012); Chopin et al. (2013); Holbrook et al. (2018); Müller et al. (2018).

## 1.2 Random Fields on Spheres: Literature Review

The literature on random fields on spheres, manifolds, or product spaces involving spheres with locally compact groups, has become quite extensive, prompting several recent reviews. Gneiting (2013) provides a comprehensive treatment of covariance functions on spheres, Jeong et al. (2017) offers a review of geostatistical approaches to spheres, and Porcu et al. (2018a) surveys the state-of-the-art of covariance functions on spheres and spheres cross time. Further, the essays in Gneiting (2013) and Porcu et al. (2018a) report collections of open problems pertaining to the mathematical, statistical and computer science communities.

A common assumption in spatial statistics is that the covariance function has the property of being isotropic, that is, the covariance between any pair of random variables observed at two different points of the spatial domain depends exclusively on the distance between the points. A particularly relevant issue when modeling Gaussian fields on spheres is that their covariance function should depend on the appropriate metric, in this case, the great circle (or geodesic) distance. As an alternative, one can build covariance functions based on chordal distance, and we refer the reader to Banerjee (2005), Gneiting (2013), Porcu et al. (2016) and Porcu et al. (2018b) for constructive criticism about the use of such a metric on the sphere.

Covariance functions are positive definite and there is a well-established theory about positive definite functions on  $d$ -dimensional spheres of  $\mathbb{R}^{d+1}$ . The reader is referred to Schoenberg (1942) and Gneiting (2013); and to the generalizations obtained by Berg and Porcu (2017) and Porcu et al. (2016).

Spectral representations of covariance functions on spheres allow for the understanding of the properties of a Gaussian field with a geodesically isotropic covariance. For instance, Lang and Schwab (2013) show that the rate of decay of the spectrum related to the function  $C$  determines the regularity properties of the associated Gaussian field in terms of interpolation spaces and Hölder continuity of sample paths. Spectral analysis on spheres became recently useful in contexts as diverse as spatial statistics (Guinness and Fuentes, 2016), equivalence of Gaussian measures and infill asymptotics (Arafat et al., 2018), approximation theory (Menegatto et al., 2006; Beatson et al., 2014; Ziegel, 2014; Massa et al., 2017) and spatial point processes (Møller et al., 2018).

Statistical methods based on spectral techniques have been explored from the more practical perspective of overcoming the computational burden associated with classical methods based on covariance functions. Most of these techniques are based on spectral representations for processes defined on a lattice and we refer again to Porcu et al. (2018a), and references therein, for a comprehensive review.

### 1.3 Our contribution

For a clearer exposition, our contribution can be categorized as follows:

- (a) **Spectral analysis and priors on  $d$ -dimensional spheres.** For the sake of completeness, we work on  $d$ -dimensional spheres (including the Hilbert sphere, for which details are given below). We choose a stick-breaking representation for the spectral expansion related to geodesically isotropic covariance functions.
- (b) **Provide the topological support for the prior.** The spectral expansion of the function  $C$  becomes, under our setting, a random sequence. Thus, the function  $C$  inherits this fact and becomes a random object for which a topological support must be provided. We do so under the topology of uniform convergence.
- (c) **First order properties and Hölder continuity.** We choose a special case of a stick-breaking representation, called GEM (Pitman and Yor, 1997, details below) that allows for the inspection of the first order properties of the random spectral sequence. As a corollary, we deduce the geometric properties of a Gaussian field with covariance function of this nature, in terms of Hölder continuity.
- (d) **Posterior inspection.** We show that the posterior for the above random spectral sequence exists, it is absolutely continuous, and admits a closed form. Finally, we show that such a posterior is Lipschitz continuous with respect to the Hellinger distance.

Our findings are validated through both a simulation study as well as through data analysis. Specifically, the outline of the paper is as follows: Section 2 contains the mathemat-

ical background needed to understand our proposed methodology. Section 3 describes our spectral nonparametric approach together with the prior specification. Section 4 presents the posterior analysis for a single realization of the Gaussian process under the proposed prior. This is later extended to the case of multiple samples in Section 5. Section 6 provides a simulation study that illustrates some practical aspects of our approach. Section 7 reports on the application of our method to a dataset of global temperatures provided by the National Center for Atmospheric Research (NCAR). A discussion finishes the paper in Section 8. Mathematical proofs and other details are given in a Supplementary Materials file (Porcu et al., 2020).

## 2 Mathematical Background

### 2.1 Random Fields and Correlation Functions on $d$ -Dimensional Spheres

We start with some notation. For a positive integer  $d$ ,  $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1}, \|\mathbf{x}\| = 1\}$  denotes the  $d$ -dimensional unit sphere embedded in  $\mathbb{R}^{d+1}$ , with  $\|\cdot\|$  being the Euclidean distance. We shall also refer to the Hilbert sphere  $\mathbb{S}^\infty = \{\mathbf{x} \in \mathbb{R}^{\mathbb{N}}, \|\mathbf{x}\| = 1\}$ . The great circle distance  $\theta : \mathbb{S}^d \times \mathbb{S}^d \rightarrow [0, \pi]$  is the continuous mapping defined as

$$\theta(\mathbf{x}_1, \mathbf{x}_2) = \arccos(\mathbf{x}_1^\top \mathbf{x}_2) \in [0, \pi],$$

for  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$ , where  $\top$  is the transpose operator. Throughout, we equivalently use  $\theta$  or  $\theta(\mathbf{x}_1, \mathbf{x}_2)$  to denote this distance, provided no ambiguity arises.

We consider stationary Gaussian random fields  $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$ , with constant mean, and with covariance function  $C(\mathbf{x}_1, \mathbf{x}_2) = \text{cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\}$ , for  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$ . Since the variance of any linear combination of a finite realization must be non-negative, the covariance function cannot be an arbitrary mapping. In other words, for any positive integer  $n$ ,  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{S}^d$  and  $\{c_1, \dots, c_n\} \subset \mathbb{R}$ , we have the condition

$$\text{var} \left( \sum_{i=1}^n c_i Z(\mathbf{x}_i) \right) = \sum_{i,j=1}^n c_i c_j C(\mathbf{x}_i, \mathbf{x}_j) \geq 0. \quad (1)$$

The mappings  $C$  that satisfy (1) are called positive definite, or strictly positive definite if the inequality is strict for any non zero vector  $(c_1, \dots, c_n)^\top$ . A discussion of these and related topics can be found in Menegatto et al. (2006) and references therein. If, in addition,

$$C(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \psi(\theta(\mathbf{x}_1, \mathbf{x}_2)), \quad \mathbf{x}_i \in \mathbb{S}^d, \quad i = 1, 2, \quad (2)$$

for some mapping  $\psi : [0, \pi] \rightarrow \mathbb{R}$  such that  $\psi(0) = 1$  and  $\sigma^2 > 0$  denoting the variance of  $Z(\mathbf{x})$ , then  $C$  is called a geodesically isotropic covariance function by Porcu et al. (2018a), with  $\psi(\theta)$  a correlation function.

Gneiting (2013) calls  $\Psi_d$  the class of continuous functions  $\psi : [0, \pi] \rightarrow \mathbb{R}$  with  $\psi(0) = 1$  such that the positive definite function  $C$  defined on  $\mathbb{S}^d \times \mathbb{S}^d$  given by (2) is positive definite. For the remainder of the paper, we do not distinguish between positive and

strict positive definiteness, unless otherwise specified. We also define  $\Psi_\infty = \bigcap_{d=1}^\infty \Psi_d$ , with the inclusion relation

$$\Psi_1 \supset \Psi_2 \supset \cdots \supset \Psi_d \supset \cdots \supset \Psi_\infty \quad (3)$$

being strict.

## 2.2 Spectral Representations on Spheres

Spectral representations for positive definite functions on spheres, being the equivalent of Bochner and Schoenberg's theorems in Euclidean spaces (see Daley and Porcu, 2013, and references therein) are available thanks to Schoenberg (1942), who shows that a mapping  $\psi : [0, \pi] \rightarrow \mathbb{R}$  belongs to the class  $\Psi_d$  if and only if it can be uniquely written as

$$\psi(\theta) = \sum_{n=0}^{\infty} b_{n,d} c_n^{(d-1)/2}(\cos \theta), \quad \theta \in [0, \pi], \quad (4)$$

where  $c_n^\lambda$  denotes the normalized  $\lambda$ -Gegenbauer polynomial of degree  $n$  (Abramowitz and Stegun, 1964), and  $\{b_{n,d}\}_{n=0}^\infty$  is a probability mass sequence. Note that in the cases  $d = 1$  (the unit circle) and  $d = 2$  (the unit sphere of  $\mathbb{R}^3$ ) the Gegenbauer polynomials simplify to Tchebycheff and Legendre polynomials (Abramowitz and Stegun, 1964), respectively.

Characterization of the class  $\Psi_\infty$  is also available thanks to Schoenberg (1942):  $\psi$  belongs to the class  $\Psi_\infty$  if and only if

$$\psi(\theta) = \sum_{n=0}^{\infty} b_n (\cos \theta)^n, \quad \theta \in [0, \pi], \quad (5)$$

with  $\{b_n\}_{n=0}^\infty$  again a probability mass sequence. We follow Daley and Porcu (2013) and denote the  $d$ -Schoenberg sequences of coefficients in (4) by  $\{b_{n,d}\}_{n=0}^\infty$ , to emphasize the dependence on the index  $d$  in the class  $\Psi_d$ . Analogously,  $\{b_n\}_{n=0}^\infty$  are called Schoenberg coefficients throughout. Fourier inversion allows for an explicit representation of the sequences  $\{b_{n,d}\}$ . Specifically, for a fixed positive integer  $d$  and any  $n = 0, 1, \dots$ ,

$$b_{n,d} = \kappa(n, d) \int_0^\pi \psi(\theta) c_n^{(d-1)/2}(\cos \theta) (\sin \theta)^{d-1} d\theta, \quad \psi \in \Psi_d, \quad (6)$$

where  $\kappa(n, d)$  is a positive constant (see Berg and Porcu, 2017). Moreover, arguments in Corollary 2 of Gneiting (2013) show that, for any  $d \geq 2$  and  $n = 0, 1, \dots$ ,

$$\begin{aligned} b_{0,3} &= b_{0,1} - \frac{1}{2}b_{2,1}, \\ b_{n,3} &= \frac{1}{2}(n+1)(b_{n,1} - b_{n+2,1}), \quad n \geq 1, \\ b_{n,d+2} &= \frac{(n+d-1)(n+d)}{d(2n+d-1)}b_{n,d} - \frac{(n+1)(n+2)}{d(2n+d+3)}b_{n+2,d}. \end{aligned} \quad (7)$$

| Family            | Analytic expression  | Parameters range              |
|-------------------|--|-------------------------------|
| Negative Binomial | $\psi(\theta) = \left(\frac{1-\delta}{1-\delta \cos \theta}\right)^\tau$ | $\delta \in (0, 1), \tau > 0$ |
| Multiquadric      | $\psi(\theta) = \left(\frac{(1-p)^2}{1+p^2-2p \cos \theta}\right)^\tau$  | $p \in (0, 1), \tau > 0$      |
| Sine Series       | $\psi(\theta) = \frac{1}{2}e^{\cos \theta - 1}(1 + \cos \theta)$         |                               |
| Sine Power        | $\psi(\theta) = 1 - 2^{-\gamma}(1 - \cos \theta)^{\gamma/2}$             | $\gamma \in (0, 2]$           |
| Poisson           | $\psi(\theta) = \exp(\lambda(\cos \theta - 1))$                          | $\lambda > 0$                 |

Table 1: Parametric families of members of the classes  $\Psi_\infty$ , with Schoenberg coefficients available in closed form.

Parametric families of some members of the classes  $\Psi_\infty$  are listed in Table 1. They are all obtained by evaluating the probability generating function associated to a probability mass sequence. The Schoenberg sequences of the third and fifth entries are provided in Porcu et al. (2016). For the Sine Power family, the Schoenberg coefficients have been obtained by Soubeyrand et al. (2008). Other parametric families whose Schoenberg coefficients are not available in closed-form are listed in Gneiting (2013).

The first entry in Table 1 is the Negative Binomial family. Its Schoenberg sequence is given by

$$b_n(\delta, \tau) = \binom{n + \tau - 1}{n} \delta^n (1 - \delta)^\tau, \quad \delta \in (0, 1), \tau > 0, \quad (8)$$

for  $n = 0, 1, \dots$ , and corresponds to the probability mass function of the Negative Binomial distribution. The Multiquadric family – the second entry in Table 1 – is obtained through the same Schoenberg sequence but with the change of variable  $\delta = 2p/(1 + p^2)$ , for  $p \in (0, 1)$ . Note that an explicit form of the  $d$ -Schoenberg coefficients can be calculated using Theorem 4.2(b) in Møller et al. (2018).

### 3 Nonparametric Spectral Modeling of Covariance Functions on Spheres

Although applications usually involve cases with  $d = 1$  and  $d = 2$ , the following exposition applies to the general case of the  $d$ -dimensional spheres  $\mathbb{S}^d$ , where  $d$  can be either finite or infinite.

Equations (4) and (6) establish a one-to-one mapping between the class  $\Psi_d$  and the infinite-dimensional simplex

$$\Delta_\infty = \left\{ \{t_n\}_{n \geq 0} : t_n \geq 0, \sum_{n=0}^{\infty} t_n = 1 \right\}, \quad (9)$$

to which the  $d$ -Schoenberg coefficients necessarily belong to. Therefore, defining a (prior) probability measure on the space  $\Psi_d$  (for a given dimension  $d$ ) is equivalent to defining a probability measure on  $\Delta_\infty$ .

It might be worth stressing that, for any  $d \in \mathbb{N}$  and any element  $\psi \in \Psi_d \setminus \Psi_{d+1}$ , there exist  $\{b_{n,1}\}_{n=0}^\infty, \{b_{n,2}\}_{n=0}^\infty, \dots, \{b_{n,d-1}\}_{n=0}^\infty, \{b_{n,d}\}_{n=0}^\infty$  Schoenberg sequences associated to  $\psi$  by a series representation of the type (4), all of which belong to  $\Delta_\infty$ . These  $d$ -Schoenberg sequences are mutually related by recursive relations (Gneiting, 2013) and projection operators (Møller et al., 2018).

For any element  $\psi \in \Psi_d$ , we assume that a spatial covariance is of the form  $C(\theta) = \sigma^2 \psi(\theta)$ , with  $\sigma^2$  denoting the variance. Setting aside the case of  $\sigma^2$  for now, we focus on the construction and study of properties of a prior distribution for the class of covariance functions on a sphere. Let us consider a random sequence  $B_d = \{B_{n,d}\}_{n=0}^\infty$ , where  $B_{n,d} \geq 0$ , for every  $n$ , and  $\sum_{n=0}^\infty B_{n,d} = 1$ , almost surely. We call  $B_d$  a  $d$ -Schoenberg random sequence, and, accordingly,  $B_\infty = \{B_{n,\infty}\}_{n=0}^\infty$  a Schoenberg random sequence.

Without loss of generality, we assume throughout that every random variable under consideration is defined on the same probability space  $(\Omega, \mathcal{F}, P)$ . The fact that the  $d$ -Schoenberg sequence is now treated as a random element  $\{B_{n,d}\}_{n=0}^\infty$  makes it possible to define a  $\Psi_d$ -valued random object  $\psi^{(B_d)}$  through the usual relation

$$\psi^{(B_d)}(\theta) = \sum_{n=0}^{\infty} B_{n,d} c_{n,d}(\cos \theta). \quad (10)$$

We now equip the class  $\Psi_d$  with an appropriate sigma-algebra,  $\mathcal{A}_d$ , so that  $\psi^{(B_d)}$  is properly defined as a random element (namely a measurable function) from  $(\Omega, \mathcal{F}, P)$  into  $(\Psi_d, \mathcal{A}_d)$ . To complete this step, it becomes critical to equip the class  $\Psi_d$  with the topology induced by the uniform (or sup) norm, that is, the norm given by

$$\|\psi\|_\infty = \sup_{\theta \in [0, \pi]} \psi(\theta),$$

or equivalently, by uniform convergence. Also, we let  $\mathcal{A}_d$  be the corresponding Borel sigma-field, namely the sigma-algebra generated by the open sets in the topology of uniform convergence. To justify that such a choice is appropriate we argue as follows. The simplex  $\Delta_\infty$  is naturally equipped with the product topology and the corresponding Borel sigma-field. Let  $\mathcal{T} : \Delta_\infty \rightarrow \Psi_d$  be the function that maps each element  $b_d$  ( $b$ ) of  $\Delta_\infty$  into the function in  $\Psi_d$  ( $\Psi_\infty$ ) whose  $d$ -Schoenberg sequence is  $b_d$  ( $b$ ). So, by (10),  $\psi^{(B_d)} = \mathcal{T}(B_d)$  and by Lemma 2 in Part C of the Supplementary Materials File,  $\mathcal{T}$  is continuous and therefore measurable. Hence, measurability of  $\{B_{n,d}\}_{n=0}^\infty : \Omega \rightarrow \Delta_\infty$  implies that  $\psi^{(B_d)}$  is in turn measurable as a function from  $\Omega$  into  $\Psi_d$ . Analogous remarks can be made about  $\psi^B$  as a function from  $\Omega$  into  $\Psi_\infty$ .

### 3.1 Defining a Prior for $d$ -Schoenberg Sequences

We can now take advantage of the above considerations, together with the fact that Fourier pairs are unique, to assign a prior distribution to the random element  $\psi^{(B_d)}$ .

We do this by defining a prior for the sequence  $B_d = \{B_{n,d}\}_{n=0}^\infty$ . A natural choice is to consider a discrete parametric distribution on the natural numbers, and place a prior on its parameters, but this would imply making strong assumptions on the behavior of the sequence of the  $d$ -Schoenberg coefficients that would be difficult to justify. For this reason it is worth considering a Bayesian nonparametric approach. Several options are available for this task. For instance, any element of the class of proper species sampling priors (Pitman, 1996) provides a sequence supported in  $\Delta_\infty$ . However, this is far too general to be of practical use. Therefore, we opt for adopting a prior distribution for the sequence  $B_d$  using a *stick-breaking representation* (see e.g. Ishwaran and James, 2001):

$$\begin{aligned} B_{0,d} &= W_{0,d}, \\ B_{j,d} &= W_{j,d} \prod_{n=0}^{j-1} (1 - W_{n,d}), \quad j \geq 1, \end{aligned} \tag{11}$$

where  $\{W_{j,d}\}_{j=0}^\infty$  is a sequence of independent random variables with values in the unit interval  $[0, 1]$  such that the support of  $W_{j,d}$  is the whole unit interval and

$$\sum_{j=0}^{\infty} \log(\mathbb{E}(1 - W_{j,d})) = -\infty. \tag{12}$$

It is not difficult to verify that under (11) and (12),  $\sum_{n=0}^\infty B_{n,d} = 1$  almost surely (see subsequent Lemma 1). Ishwaran and James (2001) show that such a convergence is guaranteed even if condition (12) is replaced with a weaker condition.

The representation in (11) combined with condition (12) are central to providing the topological support of the (prior) distribution of  $\psi^{(B_d)}$  given by (11) with respect to the topology of uniform convergence. Recall that the support of the random variable  $\psi^{(B_d)}$  is the set of all elements  $\psi$  within the class  $\Psi_d$  such that the (prior) probability that  $\psi^{(B_d)}$  belongs to  $U(\psi)$  is positive for every open ball  $U(\psi)$  of  $\psi$  in the topology of the uniform convergence. For clarity, we henceforth denote by  $\text{supp}(X)$  the support of a random variable  $X$ . The following theorem shows that under minimal conditions the support of  $\psi^{(B_d)}$  is full, that is, it coincides with the whole class  $\Psi_d$ , in the sup norm.

**Theorem 1.** *Let  $d$  be a positive integer or  $d = \infty$ . Let  $\psi^{(B_d)}$  be a random element with values in  $\Psi_d$  and defined according to (10). Let the sequence  $\{B_{n,d}\}_{n=0}^\infty$  admit the stick-breaking representation depicted through (11), with  $\{W_{n,d}\}_{n=0}^\infty$  a sequence of independent random variables such that  $\text{supp}(W_{n,d}) = [0, 1]$  for every  $n = 0, 1, \dots$  and (12) is satisfied.*

*Then,  $\text{supp}(\psi^{(B_d)}) = \Psi_d$  in the sup norm.*

Theorem 1 shows clearly why the approach proposed in this paper falls within the framework of Bayesian nonparametric approaches.

The following result illustrates a useful property of the proposed construction.

**Proposition 2.** *Let the hypotheses of Theorem 1 hold true. Additionally, if for every  $j = 0, 1, 2, \dots$ ,  $P(W_{j,d} \in \{0, 1\}) = 0$ , then for every positive integer  $N$ , and every finite*



collection of pairwise distinct points  $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{S}^d$  and constants  $c_1, \dots, c_N \in \mathbb{R}$ , where  $c_j \neq 0$ , for some  $j$ , the following holds almost surely:

$$\sum_{i=1}^N \sum_{j=1}^N c_i c_j \psi^{(B_d)}(\theta(\mathbf{x}_i, \mathbf{x}_j)) > 0.$$

That is,  $\psi^{(B_d)}$  is almost surely strictly positive definite.

### 3.2 First Order Properties and Hölder Continuity Under GEM specifications

So far we have provided a completely general construction for random elements  $\psi^{(B)}$  having the whole class  $\Psi_d$  as support. A deeper study of these objects needs a parametric specification for the marginal distributions of the independent random variables  $W_{0,d}, W_{1,d}, \dots$  appearing in the stick-breaking representation (11) that would induce a parametric specification on the  $d$ -Schoenberg random sequence  $B_d$ . By Theorem 1, it is only necessary that the support of each  $W_{n,d}$  be the unit interval. A well known probability distribution for  $B_d$  is obtained through the stick-breaking representation (11) that arises when  $W_{j,d}$  follows a Beta distribution with parameters  $(1 - \alpha, \vartheta + (j + 1)\alpha)$ , with  $0 \leq \alpha < 1$  and  $\vartheta > -\alpha$ . Such distribution is known in the probability literature as the Griffiths-Engen-McCloskey (GEM) distribution and is related to the Pitman-Yor process (Pitman, 1996; Pitman and Yor, 1997), also known as two parameter Poisson Dirichlet process. This particular specification implies increased flexibility than, e.g. the well known Ferguson-Dirichlet process (Freedman, 1963; Ferguson, 1973, 1974), while still retaining some tractability. For further reference, we use the notation  $B_d \sim \text{GEM}(\alpha, \vartheta)$ . In particular, when  $\alpha = 0$  the Ferguson-Dirichlet process is obtained. For a detailed account about these distributions, the reader is referred to Pitman (2006).

Theorem 1 together with the GEM specification shows that the support of any random object  $\psi^{(B)}$  defined according to (10) is the whole class  $\Psi_d$ . The next result provides the first order properties of the  $d$ -Schoenberg random sequence  $B_d$  when coupled with the stick-breaking representation under the GEM specification. Denote by  $x^{(m)}$  the rising factorial, that is,  $x^{(m)} = \Gamma(x + m)/\Gamma(x) = x(x + 1) \dots (x + m - 1)$ , for every integer  $m \geq 1$  and  $x^{(0)} = 1$ , for every  $x > 0$ .

**Proposition 3.** *Assume the prior distribution of the  $d$ -Schoenberg random sequence  $B_d = \{B_{n,d}\}_{n=0}^\infty$  to be specified as  $\text{GEM}(\alpha, \vartheta)$ . For a fixed  $n \in \mathbb{N}_0$ , denote by  $\mu_d(\alpha, \vartheta, n)$  the corresponding expectation of  $B_{n,d}$ . Then,*

$$\mu_d(0, \vartheta, n) = \frac{1}{1 + \vartheta} \left( \frac{\vartheta}{1 + \vartheta} \right)^n,$$

and

$$\mu_d(\alpha, \vartheta, n) = \frac{1 - \alpha}{1 + \vartheta} \frac{(1 + \vartheta/\alpha)^{(n)}}{(1 + (1 + \vartheta)/\alpha)^{(n)}}, \quad \text{if } \alpha > 0.$$

Some comments are in order. Proposition 3 shows that the mapping  $n \mapsto \mu_d(\alpha, \vartheta, n)$  is the probability mass function of the beta geometric distribution with parameters  $(1 - \alpha)/\alpha$  and  $(1 + \vartheta)/\alpha$ . In other words, using the notation introduced in Section 2, and denoting by  $g_{u,v}$  the beta density with parameters  $u$  and  $v$ ,

$$\mu_d(\alpha, \vartheta, n) = \int_{[0,1]} b_n(\delta, 1) g_{(1-\alpha)/\alpha, (1+\vartheta)/\alpha}(\delta) d\delta, \quad (13)$$

where  $b_n(\delta, \tau)$  is the  $d$ -Schoenberg sequence associated with the negative binomial family, as reported in (8). Thus, it becomes apparent that  $\mu_d$  does not depend on the integer  $d$  that indexes the class  $\Psi_d$ .

Proposition 3 shows even more: the prior distribution of the random element  $\psi^{(B_d)}$  is centered at the (deterministic) element of  $\Psi_d$  whose  $d$ -Schoenberg coefficient are given by (13), namely:

$$\mathbb{E}(\psi^{(B_d)}(\theta)) = \sum_{n=0}^{\infty} \mu_d(\alpha, \vartheta, n) c_n^{(d-1)/2}(\cos \theta), \quad \theta \in [0, \pi].$$

In particular, if  $d = \infty$ , then arguments in Proposition 3.4 in Berg and Porcu (2017) and the application of bounded convergence show that the prior of the random element  $\psi^{(B)}(\theta)$  has expected value

$$\mathbb{E}(\psi^{(B_d)}(\theta)) = \int_{[0,1]} \mathcal{N}_{\delta,1}(\theta) g_{(1-\alpha)/\alpha, (1+\vartheta)/\alpha}(\delta) d\delta, \quad \theta \in [0, \pi],$$

where  $\mathcal{N}_{\delta,\tau}(\theta) = \left(\frac{1-\delta}{1-\delta \cos \theta}\right)^\tau$  is the negative binomial family as reported in the first entry of Table 1 and belonging to the class  $\Psi_\infty$ .

The behavior of the correlation function at the origin is fundamental in determining the geometric properties of the associated Gaussian random field (Diggle and Ribeiro, 2007). The framework proposed in this section allows to provide results along these lines. Since this paper deals with correlation functions  $\psi(\theta)$ , for  $\psi : [0, \pi] \rightarrow \mathbb{R}$ , derivatives at the origin of  $\psi$  can be seen as derivatives of the even extension  $\tilde{\psi}$  of  $\psi$  to the interval  $[-\pi, \pi]$ .

For the remainder of the paper, for any pair of real-valued sequences  $\{a_n\}$  and  $\{d_n\}$ , we write  $a_n \sim d_n$  ( $a_n$  is asymptotically equivalent to  $d_n$ ) when  $a_n/d_n \rightarrow 1$  as  $n \rightarrow \infty$ . Also, we define the Digamma function  $F$  (Abramowitz and Stegun, 1964) through

$$F(x) = \frac{d}{dx} \log \Gamma(x), \quad x > 0.$$

**Proposition 4.** *Assume that the  $d$ -Schoenberg random sequence  $B_d = \{B_{n,d}\}_{n=0}^\infty$  has prior specified as  $\text{GEM}(\alpha, \vartheta)$ . Then,  $P$ -almost surely,*

1. *If  $0 < \alpha < 1$ , then there exists a random variable  $Z$ , with  $0 < Z < \infty$ , such that*

$$B_j \sim Z(1 - \alpha)\Gamma(1 - \alpha)^{1-\alpha} j^{-1/\alpha}, \quad j \rightarrow \infty.$$

2. If  $\alpha = 0$ , then

$$B_j \sim e^{-(F(1+\vartheta)-F(\vartheta))j}, \quad j \rightarrow \infty.$$

Proposition 4 implies actually much more. For convenience, we recall a recent result relating Hölder continuity and  $d$ -Schoenberg sequences.

**Proposition 5** (Lang and Schwab, 2013). *Let  $d$  be a positive integer and let  $\psi$  be a member of the class  $\Psi_d$  with  $d$ -Schoenberg sequence  $b_d = \{b_{n,d}\}_{n=0}^\infty$  uniquely determined according to (6). If*

$$\sum_{n=0}^{\infty} b_{n,d} n^\beta < \infty,$$

for some  $\beta > 0$ , then there exists a constant  $k_\beta$  such that for all  $\theta \in [0, \pi]$ ,

$$|\psi(\theta) - 1| \leq k_\beta \theta^\beta.$$

We can now combine Propositions 4 and 5 to assert the following:

**Proposition 6.** *For  $d \in \mathbb{N}$ , assume that the  $d$ -Schoenberg random sequence  $B_d = \{B_{n,d}\}_{n=0}^\infty$  has GEM( $\alpha, \vartheta$ ) distribution.*

1. *If  $0 < \alpha < 1$ , then for any  $\beta \in (0, \alpha/(1-\alpha))$  there exists a constant  $k_\beta$  such that for all  $\theta \in [0, \pi]$ ,*

$$|\psi^{(B)}(\theta) - 1| \leq k_\beta \theta^\beta, \quad (14)$$

*almost surely.*

2. *If  $\alpha = 0$ , then (14) holds true, almost surely, for any  $\beta > 0$ .*

Proposition 6 has precise consequences in terms of sample path properties of Gaussian random fields. In fact, the Hölder continuity of the covariance function is directly related to the P-almost sure continuity of the associated Gaussian field. This is a direct consequence of the Kolmogorov-Chentsov theorem, which analogue on the sphere has been provided by Lang and Schwab (2015).

## 4 Existence of the Posterior and Lipschitz Continuity

The specification of the prior in Section 3.1 allows us to obtain further results. In particular, let  $d$  be a positive integer. Suppose the Gaussian process  $Z$  on  $\mathbb{S}^d$  is sampled at locations  $(\mathbf{x}_1, \dots, \mathbf{x}_N)^\top$ . Call  $\mathbf{Z}$  the  $n$ -dimensional random vector  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_N))^\top$  and let  $\mathbf{z} := (z(\mathbf{x}_1), \dots, z(\mathbf{x}_N))^\top$  be one such realization. Let  $Z$  have unknown correlation  $\psi^{(B_d)}$ , being uniquely determined through the  $d$ -Schoenberg random sequence  $B_d$ . For simplicity we assume that the variance of the marginal distributions of the process is known and fixed. Then, conditionally on the event  $\{B_d = b_d\}$ ,  $Z$  is a Gaussian process with correlation function  $\psi^{(b_d)}(\theta)$ ,  $\theta \in [0, \pi]$ . We note that joint distribution of  $B_d$  and  $Z$  can be specified through the prior distribution of  $B_d$  together with the conditional distribution of  $Z$  given  $B_d$ .

For any set  $A$ , let  $\mathbb{I}_A$  denote its corresponding indicator function, namely  $\mathbb{I}_A(x) = 1$  if  $x$  belongs to  $A$  and  $\mathbb{I}_A(x) = 0$  otherwise. It is worth noting that the joint distribution of  $Z$  and  $B_d$  can be obtained by setting:

$$Z(\mathbf{x}) = \mu + \sigma^2 \sum_{n=0}^{\infty} \mathbb{I}_{\{M=n\}} Z_n(\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^d, \quad (15)$$

where  $\mu = \mathbb{E}(Z(\mathbf{x}))$ , *i.e.*,  $\mu$  is the (constant) mean of the process and  $\sigma^2$  is the variance. Here, the discrete random variable  $M$  and the processes  $\{Z_n(\mathbf{x}) : \mathbf{x} \in \mathbb{S}^d, n = 0, 1, \dots\}$  are conditionally independent given  $B_d$ . Also, we let  $P(M = n | B_d) = B_{n,d}$  and for  $n = 0, 1, \dots$ ,  $\{Z_n(\mathbf{x}) : \mathbf{x} \in \mathbb{S}^d\}$  are zero mean Gaussian processes independent of  $B_d$  with

$$\mathbb{E}(Z_n(\mathbf{x})Z_m(\mathbf{y})) = c_n^{(d-1)/2}(\cos \theta(\mathbf{x}, \mathbf{y}))\delta_{n=m}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^d, \quad n, m = 0, 1, \dots,$$

with  $\delta$  denoting the Kronecker delta. The assumption above on the covariance between  $Z_n$  and  $Z_m$  is certainly correct because, for every  $n \in \mathbb{N}_0$  and for any positive integer  $d$ , the  $n$ -th normalized Gegenbauer polynomial of order  $(d-1)/2$  is positive definite on  $\mathbb{S}^d \times \mathbb{S}^d$ ; see Berg and Porcu (2017) for technical details.

We now have the ingredients to establish the posterior distribution of  $B_d$ , namely the conditional distribution of  $B_d$  given  $\mathbf{Z} = \mathbf{z}$ , provided such a distribution exists. We denote the posterior of  $B_d$  by  $\mathbb{P}^{\mathbf{z}}$  and the prior by  $\mathbb{P}$ .

**Proposition 7.** *Let  $\psi^{(B_d)}$  be a random element with values in  $\Psi_d$ , defined according to (10). Let the sequence  $\{B_{n,d}\}_{n=0}^{\infty}$  admit the stick-breaking representation depicted through (11), with  $\{W_{n,d}\}_{n=0}^{\infty}$  being a sequence of independent random variables such that  $P(W_{n,d} \in (0, 1)) = 1$  for every  $n = 0, 1, \dots$  and (12) is satisfied.*

*Then, the posterior  $\mathbb{P}^{\mathbf{z}}$  of  $B_d$  exists, is unique and absolutely continuous with respect to  $\mathbb{P}$ , with uniquely determined density evaluated at  $b_d = \{b_{n,d}\}_{n=0}^{\infty} \in \Delta_{\infty}$ :*

$$\frac{f(\mathbf{z}; b_d)}{M(\mathbf{z})}, \quad (16)$$

where  $f(\mathbf{z}; b_d)$  is the  $N$ -dimensional centered multivariate Gaussian density evaluated at  $\mathbf{z}$  with covariance matrix

$$\Sigma_N(b_d) := \sigma^2 \sum_{n=0}^{\infty} b_{n,d} \Sigma_n^{(N)}, \quad (17)$$

where

$$\Sigma_n^{(N)} = \left( c_n^{(d-1)/2}(\cos \theta(\mathbf{x}_j, \mathbf{x}_k)) \right)_{j,k=1}^N,$$

and where

$$M(\mathbf{z}) = \mathbb{E}(f(\mathbf{z}; B_d)).$$

It is also useful to examine whether the posterior  $\mathbb{P}^z$  varies smoothly as a function of  $z$ , as it relates to the robustness of the posterior to small changes in the observed data. Using continuity in the Hellinger metrics as a criterion, if it is established then we say that the estimation task is *well-posed* (see Stuart, 2010; Dashti and Stuart, 2017).

In order to address the well-posedness of the posterior, recall the definition of the Hellinger distance between two probability measures,  $\nu_1$  and  $\nu_2$ , dominated by the same measure  $\nu$  with Radon–Nikodym derivatives  $f_1$  and  $f_2$ , respectively:

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

We are now able to state our next finding.

**Theorem 8.** *Under the same conditions as in Proposition 7, it is true that the posterior  $\mathbb{P}^z$  is Lipschitz continuous with respect to the Hellinger distance in the data  $z$ : for any  $r > 0$ , there exists a strictly positive constant  $m(r)$  such that for all  $z_1, z_2 \in \mathbb{R}^N$  with  $\max\{\|z_1\|, \|z_2\|\} \leq r$ ,*

$$d_H(\mathbb{P}^{z_1}, \mathbb{P}^{z_2}) \leq m(r) \|z_1 - z_2\|. \quad (19)$$

## 5 Statistical Model

For the remainder of this article, we will assume a statistical model based on a realization (possibly with replicates) of a Gaussian process  $Z$  on  $\mathbb{S}^d$ , sampled at locations  $(\mathbf{x}_1, \dots, \mathbf{x}_N)^\top$ . We work under the setting outlined in Proposition 7, *i.e.*,  $Z$  is assumed to be a Gaussian process with unit variance and correlation function  $\psi^{(B_d)}$ , a random element with values in  $\Psi_d$ , as previously defined in (10). For the upcoming developments, we consider the case where  $m$  replicates of this process is available. We denote these by  $\mathbf{Z}^1, \dots, \mathbf{Z}^m$ . Write  $\mathbf{Z}^{(m)} = (\mathbf{Z}^1, \dots, \mathbf{Z}^m)$ . The model then states that  $\mathbf{Z}^1, \dots, \mathbf{Z}^m$  are conditionally iid given  $B_d = b_d$  with multivariate Gaussian distribution with covariance matrix  $\Sigma_N(b_d)$ , as defined in (17). In turn, the sequence  $B_d$  is modeled as generated from a GEM( $\alpha, \vartheta$ ) sequence, *i.e.*, as in (11), with  $W_{j,d} \sim \text{Beta}(1 - \alpha, \vartheta + (j + 1)\alpha)$ , where  $0 \leq \alpha < 1$  and  $\vartheta > -\alpha$ . Finally, we assume a prior distribution on  $\alpha$  and  $\vartheta$ . In summary, we work with the following model:

$$\mathbf{Z}^1, \dots, \mathbf{Z}^m \mid B_d = b_d \stackrel{\text{iid}}{\sim} \mathcal{N}_N(\mathbf{0}_N, \Sigma_N(b_d)), \quad (20)$$

$$B_d \mid \alpha, \vartheta \sim \text{GEM}(\alpha, \vartheta), \quad (21)$$

$$\vartheta \mid \alpha \sim \mathcal{LTN}(-\alpha, 0, A^2), \quad (22)$$

$$\alpha \sim \text{Beta}(\alpha_0, \beta_0), \quad (23)$$

where  $\mathbf{0}_N$  denotes the zero-vector in  $\mathbb{R}^N$ ,  $\mathcal{LTN}(\theta_1, \theta_2, \theta_3^2)$  denotes the univariate normal distribution with mean  $\theta_2$  and variance  $\theta_3^2$ , but truncated to the interval  $(\theta_1, \infty)$ , that is, with probability density function given by

$$p(x; \theta_1, \theta_2, \theta_3^2) = \frac{1}{\left[1 - \Phi\left(\frac{\theta_1 - \theta_2}{\theta_3}\right)\right] \sqrt{2\pi\theta_3^2}} \exp\left\{-\frac{(x - \theta_2)^2}{2\theta_3^2}\right\}, \quad x \geq \theta_1,$$

and  $\alpha_0$ ,  $\beta_0$  and  $A^2$  are user-supplied positive constants. Note that the prior on GEM parameters,  $p(\alpha, \vartheta)$  is specified by means of a conditional and a marginal distribution. This same strategy was employed in Jara et al. (2010). In all simulation experiments and data analyses described below we assume  $\sigma = 1$  by which we simply imply that the data are assumed to be standardized so that the correlation function and the Schoenberg sequence  $b_d$  are our main inferential targets. We note also that extending the previous model to the case of unknown variance is straightforward, only requiring an extra prior distribution on  $\sigma^2$  (or  $\sigma$ ). We omit the corresponding details.

It is also straightforward to show that the theoretical results developed earlier in Section 4 for the case of a single realization of the Gaussian process (*i.e.*,  $m = 1$ ) remain valid in the more general case (*i.e.*,  $m > 1$ ), with some obvious adjustments, e.g. as in the case of Proposition 7 and Theorem 8, which we use here without proof.

## 6 Simulation Study

We have so far studied properties of the proposed class of nonparametric priors for a correlation function over the sphere. We also discussed in Section 4 the case of the posterior distribution corresponding to a sample from a Gaussian process having a correlation function in  $\Psi_d$ , which can then be represented by means of (11). We now study through simulation the performance of various aspects of the model, in particular, estimation of correlation functions in  $\mathbb{S}^d$  and prediction.

Despite the generality of the results and model stated above, and for the sake of simplicity, the simulation experiments we present next are performed over the two dimensional sphere  $\mathbb{S}^2$  embedded in  $\mathbb{R}^3$ . We focus on two aspects: on the one hand, we explore the efficiency of Markov Chain Monte Carlo (MCMC) techniques from simulations drawn from a  $\text{GEM}(\alpha, \vartheta)$  distribution. We consider the case of a varying number of points on the sphere. On the other, we examine the flexibility of the proposed model in reproducing the second-order structure of common parametric forms. In particular, we consider two parametric classes of geodesically isotropic correlation functions. First, the exponential correlation function  $\psi_E$ , given by

$$\psi_E(\theta; \phi) = \exp\left(-\frac{\theta}{\phi}\right), \quad \theta \in [0, \pi], \quad (24)$$

where  $\phi > 0$  is a positive scaling parameter. Arguments in Gneiting (2013) show that  $\psi_E$  belongs to the class  $\Psi_\infty$  for all  $\phi > 0$ . Secondly is the multiquadric correlation  $\psi_M(\cdot; p, \tau)$ , corresponding to the second entry in Table 1, where the parameter restriction is reported in the same table.

In all the simulation scenarios discussed below, we adopt the model defined as (20) through (23) with  $\alpha_0 = 1$ ,  $\beta_0 = 5$  and  $A^2 = 10^5$ , reflecting a prior that assigns higher mass to values of  $\alpha$  closer to 0 (*i.e.*, the Dirichlet process) and a vague specification for  $\vartheta$ .

### 6.1 Simulation I

In this first simulation experiment, we explore some inferential aspects of the Schoenberg sequence  $b_2$  for data drawn from the  $\text{GEM}(\alpha, \vartheta)$  prior. In particular, we consider estimation of the Schoenberg coefficients, and most importantly, of the corresponding correlation function  $\psi^{(b_2)}(\theta)$ ,  $\theta \in [0, \pi]$ . We assess the efficiency of MCMC estimation technique, varying the number of points on the sphere from which synthetic realizations are drawn.

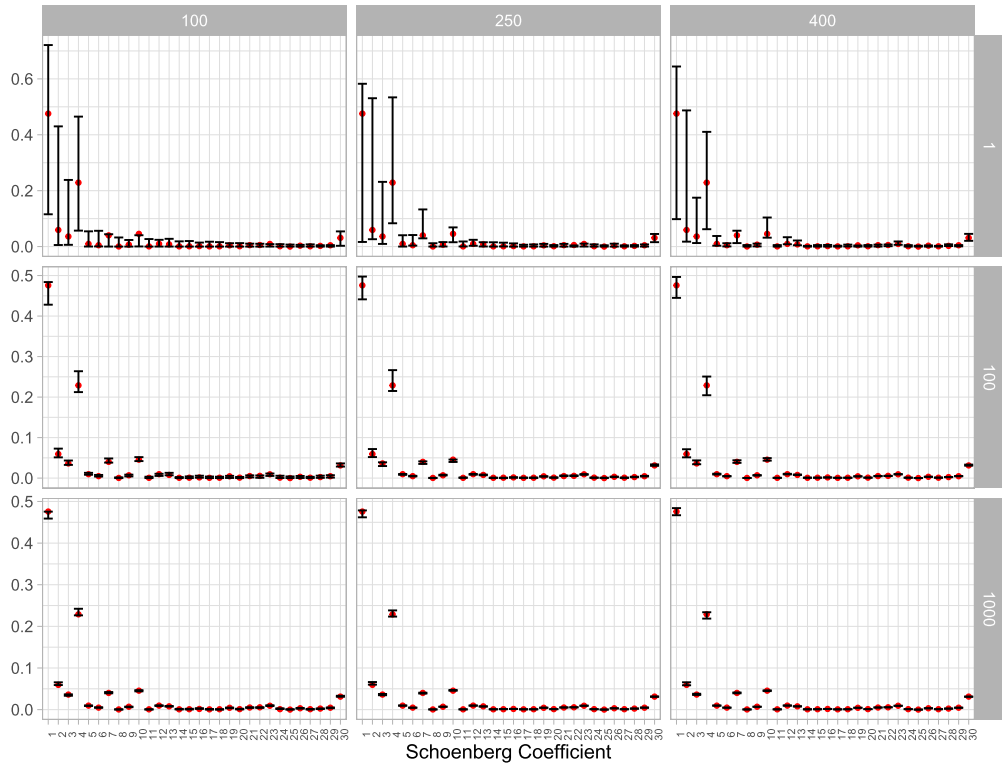


Figure 1: Posterior distribution of the 30 terms of the truncated Schoenberg sequences. The red dots correspond to the true simulation value, while the bars represent the posterior 95% credibility intervals. The rows indicate the different values of  $m \in \{1, 100, 1000\}$  while the columns the different values of  $N \in \{100, 250, 400\}$ .

The spatial design mimics an infill asymptotics setting on the sphere, where an increasing number of points is considered over a compact set. For a recent account of infill asymptotics on  $d$ -dimensional spheres, the reader is referred to Arafat et al. (2018). Specifically, we generate 3 increasing sets consisting of  $N$  randomly chosen points over the unit sphere, with  $N = 100, 250, 400$ . For each  $N$ , we fix and assume that the true GEM distribution has a fixed value of  $\vartheta = 3$  and  $\alpha = 0.3$ , from which a random realization is drawn. Although a realization of GEM distribution entails a countably

infinite sequence in  $(0, \infty)$ , conducting inference necessarily calls for a truncation of the sequence, details of which are provided in Ishwaran and James (2001). Here we consider a truncation up to and including the 30-th term. The obtained covariance matrix  $\tilde{\Sigma}_N(b_d)$  is the analogue of the matrix  $\Sigma_N(b_d)$  as defined in (17), and obtained by replacing the series with the corresponding partial sum. Inference is performed on  $m \in \{1, 100, 1000\}$  independent spatial replicates of the Gaussian random field  $Z$ .

In particular, inference is based on Hamiltonian Monte Carlo, implemented in the Stan modeling language (Carpenter et al., 2017) within the R package `rstan`. A single chain was considered for each of the experiments, of length 2000, the first 1000 iterations of which were discarded as warm-up. For each of the noted values of  $N$ , we fit model (20)–(23).

Figure 1 compares the posterior samples for each of the 30 Schoenberg coefficients in the truncated representation for each of the  $N$  values (columns) and each of the number  $m$  of realizations (rows). Note how the posterior distributions of the coefficients progressively tend to become more concentrated, with the exception of the last one, a side effect of the truncation. In all cases, the posterior distributions follow the true values, with levels of posterior concentration that generally increase with  $N$  and particularly with  $m$ . Indeed, ranging from  $m = 1$  to  $m = 1000$  produces a substantial reduction in the posterior variability for each of these coefficients.

Of fundamental importance for this work is the correlation function. Figure 2 shows the (truncated) correlation functions corresponding to each of the simulation scenarios, evaluated at a grid of 60 equally spaced values over  $[0, \pi]$ . The true values are in black and 95% posterior credibility bands are in red. All dots are joined to facilitate graphical display. As expected, the true collection of correlation points are contained in the corresponding credibility bands, and increasing  $m$  has the effect of reducing the width of these bands so that for  $m = 1000$  they are barely distinguishable from the true values.

In closing this first simulation study we mention that we carried out a number of alternative computational experiments with different combinations of GEM parameter values as data generating distribution (including  $\alpha = 0$  and several values for  $\vartheta$ ). We also tried different specifications of the prior, including the case when both  $\alpha$  and  $\vartheta$  are known. We found a general agreement in the results in the sense that the inference for  $b_2$  and  $\psi^{(b_2)}(\theta)$  for  $\theta \in [0, \pi]$  produced results of similar quality to those shown here (data not shown).

## 6.2 Simulation II

In this second simulation experiment we compare our model with that corresponding to the exponential correlation function. Specifically, we simulate data from a zero mean Gaussian random field,  $Z$ , defined over the two dimensional sphere  $\mathbb{S}^2$ , assumed to be continuous in the mean-square sense (thus, no nugget effect is considered in the covariance function), with unit variance, and with exponential correlation function  $\psi_E$  as defined in (24). Thus,  $Z$  is continuous but not mean-square differentiable because  $\psi_E$  is not differentiable (under an even extension) at the origin. We consider two cases for the scaling parameter,  $\phi = 100, 1000$  km (see Figure 3). For each value of  $\phi$ , we



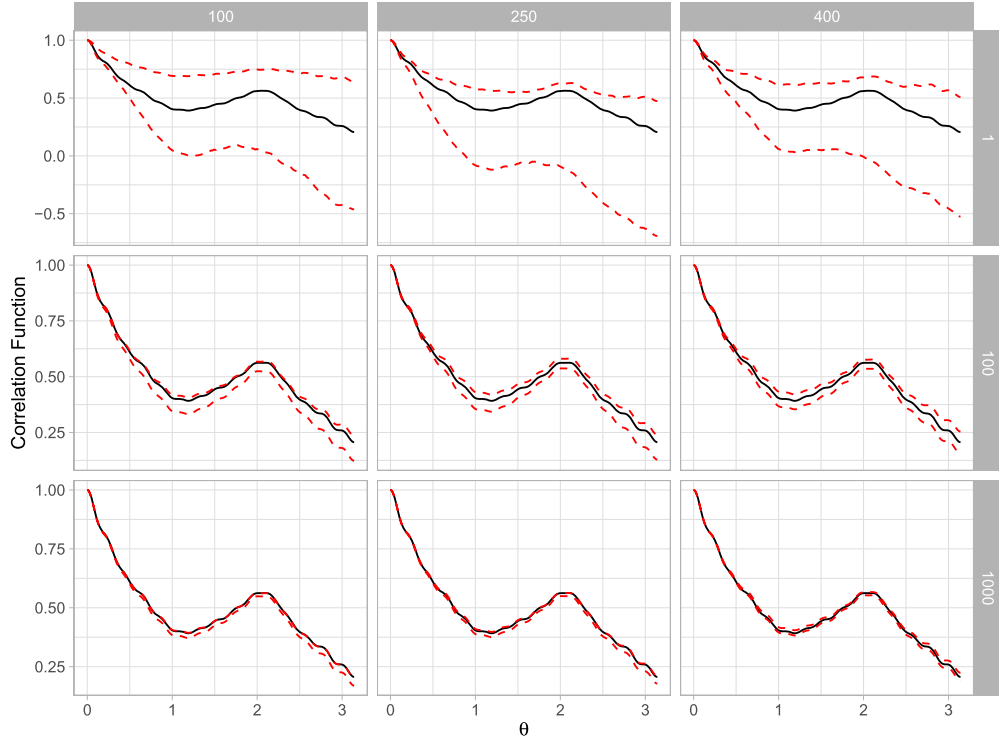


Figure 2: Posterior distribution of the estimated correlation function for each of the simulation scenarios. The black curve corresponds to the true correlation function evaluated at a grid (dots are joined for display). The red curves correspond to point-wise 95% credibility bands (also joined for display).

randomly generate 100 points, simulate 50 independent realization of the associated random field, and then fit both exponential correlation functions  $\psi_E$ , using a vague uniform prior for the scaling parameter as  $\phi \sim U(0, 10^8)$ , and the nonparametric model

$$\psi_{\text{BNP}}(\theta; K) = \sum_{n=0}^K b_{n,2} c_{n,d}^{(d-1)/2}(\cos(\theta)), \quad \theta \in [0, \pi], \quad (25)$$

with coefficients  $b_{n,2}$  based on the GEM specification. Specifically, we consider 3 levels of truncation,  $K = 30, 40, 50$ . We repeat this 50 times and compute the average predictive scores, shown in Table 2. Such prediction scores are the root mean square error and the continuous ranked probability score. Letting  $\hat{Z}_{-i}(\mathbf{x}_i)$  denote the *drop one prediction*, namely the best prediction based on all data  $Z(\mathbf{x}_j)$  for  $j \neq i$ , the *root mean square error* (RMSE) is:

$$\left\{ \frac{1}{N} \sum_{i=1}^N (Z(\mathbf{x}_i) - \hat{Z}_{-i}(\mathbf{x}_i))^2 \right\}^{1/2}.$$

Letting  $F_i(y) = P(Z(\mathbf{x}_i) \leq y \mid Z(\mathbf{x}_j), j \neq i)$  be the *predictive cumulative density function*, the *continuous ranked probability score* (CRPS) is

$$\left\{ \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} (\hat{F}_i(y) - \mathbb{I}_{\{Z(\mathbf{x}_i) \leq y\}})^2 \right\}^{1/2}.$$

Both RMSE and CRPS thus correspond to measures of predictive accuracy.

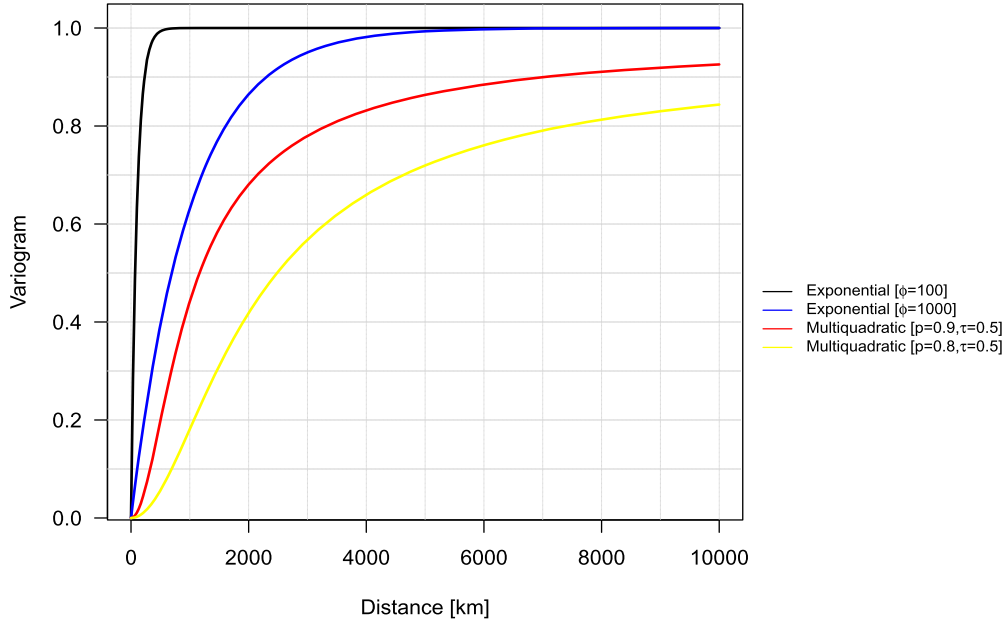


Figure 3: Variograms for selected models.

In both cases we observe that the level of agreement between the Bayesian Nonparametric (BNP) and parametric models increases with the level of truncation, specially for the case where the scaling parameter equals 1000. Nevertheless, the lack-of-smoothness at the origin, a well-known feature of the exponential correlation function, complicates the approximation. This explains why the model used to generate the data provides better predictions than the BNP, although by a modest margin that decreases with  $K$ .

Analogously, we repeat the experiment but now with data generated from a multi-quadratic correlation function  $\psi_M(\cdot; p, \tau)$ , with  $\tau = 0.5$  and  $p = 0.8, 0.9$  (see Figure 3), and using independent priors for  $p$  and  $\nu$  given by  $p \sim U(0, 1)$  and  $\tau \sim U(0, 10^8)$ . The average predictive scores are presented in Table 3.

For both values of  $p$ , we find that even the more parsimonious  $\psi_{\text{BNP}}(\cdot; 30)$  performs comparably to what is obtained using the true data generating model. Indeed, for  $p = 0.8$ , increasing the number of coefficients yields no meaningful improvement in

| Model                         | $\phi = 100$            |                         | $\phi = 1000$           |                         |
|-------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|                               | RMSE                    | CRPS                    | RMSE                    | CRPS                    |
| $\psi_{\text{BNP}(\cdot;30)}$ | 1.365<br>(0.961, 2.462) | 0.874<br>(0.727, 1.166) | 1.005<br>(0.843, 1.248) | 0.751<br>(0.690, 0.851) |
| $\psi_{\text{BNP}(\cdot;40)}$ | 1.205<br>(0.917, 1.795) | 0.822<br>(0.716, 0.953) | 0.969<br>(0.836, 1.119) | 0.738<br>(0.683, 0.802) |
| $\psi_{\text{BNP}(\cdot;50)}$ | 1.154<br>(0.945, 1.561) | 0.808<br>(0.726, 0.907) | 0.934<br>(0.807, 1.097) | 0.725<br>(0.674, 0.791) |
| $\psi_{\text{E}}$             | 1.021<br>(0.874, 1.162) | 0.760<br>(0.705, 0.809) | 0.898<br>(0.773, 1.054) | 0.709<br>(0.655, 0.771) |

Table 2: Prediction scores for each model (50 cases), with the true model having an exponential correlation function with  $\phi = 100, 1000$ . Here,  $\psi_{\text{BNP}(\cdot; \cdot)}$  is defined in (25) and  $\psi_{\text{E}}$  in (24).

| Model                          | $p = 0.8$               |                         | $p = 0.9$               |                         |
|--------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|                                | RMSE                    | CRPS                    | RMSE                    | CRPS                    |
| $\psi_{\text{BNP}(\cdot; 30)}$ | 0.471<br>(0.400, 0.560) | 0.477<br>(0.440, 0.522) | 0.703<br>(0.620, 0.835) | 0.620<br>(0.580, 0.691) |
| $\psi_{\text{BNP}(\cdot; 40)}$ | 0.471<br>(0.400, 0.559) | 0.477<br>(0.436, 0.521) | 0.694<br>(0.586, 0.805) | 0.613<br>(0.561, 0.678) |
| $\psi_{\text{BNP}(\cdot; 50)}$ | 0.471<br>(0.400, 0.559) | 0.477<br>(0.436, 0.521) | 0.692<br>(0.588, 0.803) | 0.611<br>(0.560, 0.669) |
| $\psi_{\text{M}}$              | 0.471<br>(0.396, 0.554) | 0.477<br>(0.433, 0.52)  | 0.692<br>(0.598, 0.798) | 0.612<br>(0.565, 0.670) |

Table 3: Average prediction scores for each model (50 cases), with data being generated from a multiquadric correlation function  $\psi_{\text{M}}$  with  $\tau = 0.5$  and  $p = 0.8, 0.9$ .

the predictive performance. This is to be expected because unlike the exponential family case illustrated earlier, the multiquadric correlation function is a better-behaved element of  $\Psi_{\infty}$  (recall Table 1).

## 7 Data Illustration: Surface Temperature

Lastly, we provide an illustration of the proposed nonparametric device to surface temperature data provided by the publicly available Large Ensemble Project (LENS) (Kay et al., 2015), developed at NCAR. The dataset consists of over 30 simulations from the Community Earth System Model (CESM) version 1, covering the years 1920 to 2100. Historical forcing is used from 1920 to 2005, and the representative concentration pathway 8.5 (RCP8.5) onwards until 2100. The horizontal resolution is approximately  $1^{\circ}$ , with 288 longitude points, and 192 latitude points, for a total of 55,296 points at each time-step. The chaotic nature of the climate system ensures that each simulation resembles an independent realization from a common data generating mechanism, by introducing small round-off level differences in the initial atmospheric conditions. We use Boreal summer averages – averages for the months of June, July and August (JJA)

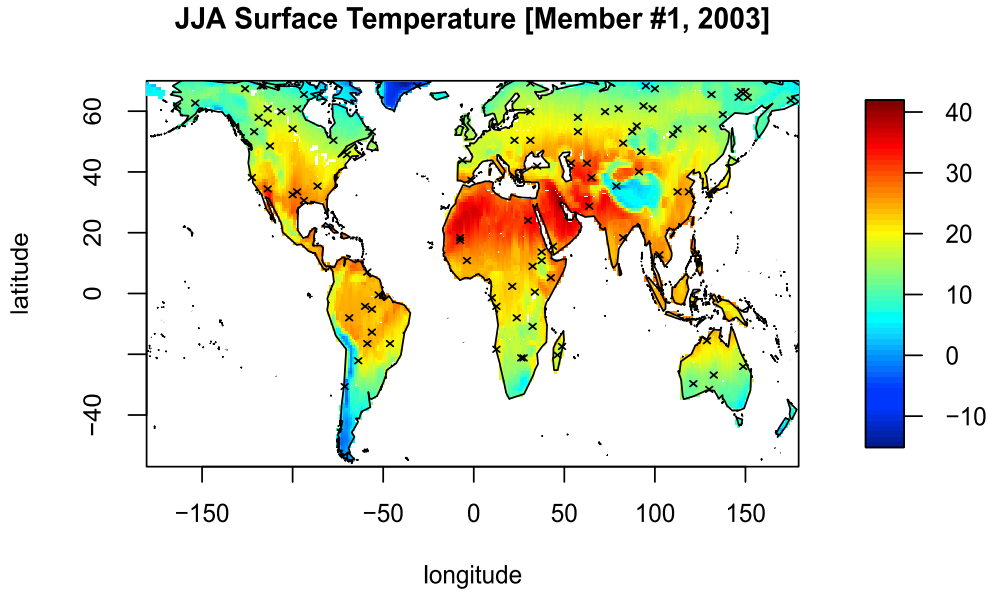


Figure 4: Surface temperature in degrees Celsius for the mean over the months of June, July and August (JJA), for the first member of the LENS ensemble, for the year 2003. The random set of points used for the fitting are denoted by crosses.

– for the years 2003 to 2005, for the first 30 ensemble members. As an example, Figure 4 shows the 2003 JJA mean of the first ensemble member. This gives us a set of 30 means for each of the three JJA months and we assume the total of  $m = 90$  JJA means to be i.i.d. spatial replicates. Given the different surface temperature characteristics between land and ocean, we focus only on land regions, and only within the latitude band  $57^{\circ}\text{S}$ – $70^{\circ}\text{N}$  to avoid potential distortions near the poles, thus reducing the number of points to 12,037. For this study,  $N = 100$  points were chosen from these at random, whose locations are depicted in Figure 4. Our main goal in this analysis is to illustrate the application of our model compared to other parametric alternatives, and to estimate the spatial correlation function.

A preliminary step to the modeling of the correlation structure is the removal of the mean structure from the JJA averages. Previous studies, such as Stein (2007) and Jun et al. (2008) for ozone concentration, and Jeong and Jun (2015) for sea-level pressure, have used spherical harmonics for this purpose. We initially considered their use, but found that they had difficulty in capturing localized features, for instance, over the Qinghai-Tibetan plateau and its surroundings, despite considering degrees  $\ell = 0, \dots, 12$ , and orders  $0 \leq m \leq \ell$ . Instead, standardizing the residuals across the 90 realizations worked well in achieving a mean close to zero. Figure 5 shows the mean surface and the residual surface of the realization in Figure 4.

To examine the sensitivity of the predictive performance of the BNP model to the degree of truncation, we consider  $K = 30, 40, 50$  as in the simulation study in Section 6.

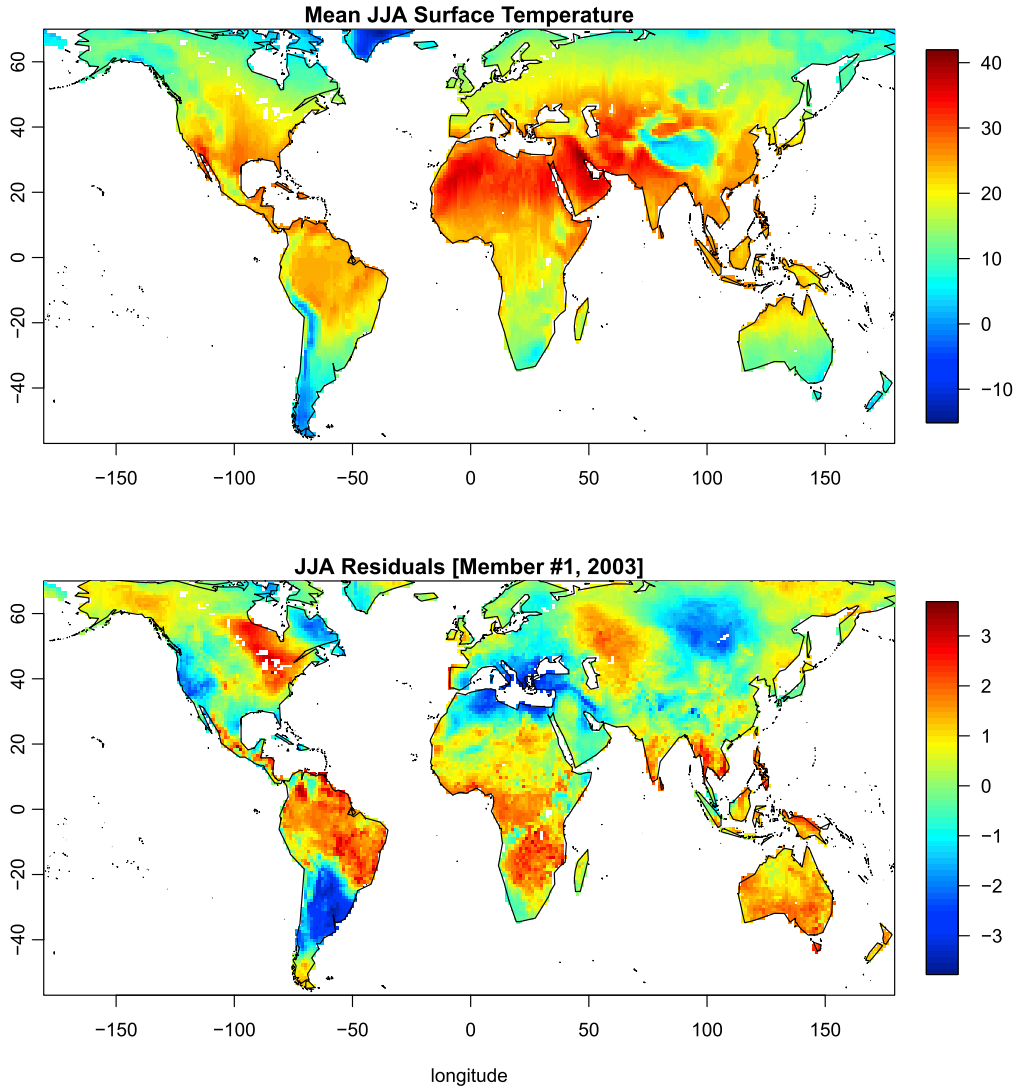


Figure 5: (top) Mean JJA surface temperature for the 30 ensembles, over the years 2003–2005; (bottom) JJA surface temperature residuals from the first ensemble member for the year 2003.

To further test the model’s performance, we added this time the case  $K = 100$ . The analysis is again based on model (20) through (23) with the same prior choices as in Section 6, *i.e.*,  $\alpha_0 = 1$ ,  $\beta_0 = 5$  and  $A^2 = 10^5$ .

In order to assess the performance relative to parametric alternatives, we fit the exponential correlation function  $\psi_E$  as in (24), the multiquadric correlation function  $\psi_M$

| Model                          | $E(\phi   \mathbf{Z})$   | $E(p   \mathbf{Z})$         | $E(\tau   \mathbf{Z})$  | $E(\nu   \mathbf{Z})$   |
|--------------------------------|--------------------------|-----------------------------|-------------------------|-------------------------|
| $\psi_E$                       | 1672<br>(1615.9, 1738.8) | –<br>–                      | –<br>–                  | –<br>–                  |
| $\psi_M$                       | –<br>–                   | 0.928<br>(0.925, 0.932)     | 0.443<br>(0.415, 0.471) | –<br>–                  |
| $\psi_{LM}$                    | 5.283<br>(4.785, 5.869)  | –<br>–                      | –<br>–                  | 0.197<br>(0.191, 0.202) |
|                                | $E(\alpha   \mathbf{Z})$ | $E(\vartheta   \mathbf{Z})$ |                         |                         |
| $\psi_{\text{BNP}(\cdot;30)}$  | 0.049<br>(0.001, 0.160)  | 11.147<br>(6.505, 16.412)   |                         |                         |
| $\psi_{\text{BNP}(\cdot;40)}$  | 0.051<br>(0.001, 0.174)  | 11.744<br>(6.774, 16.601)   |                         |                         |
| $\psi_{\text{BNP}(\cdot;50)}$  | 0.113<br>(0.006, 0.289)  | 11.760<br>(4.858, 18.316)   |                         |                         |
| $\psi_{\text{BNP}(\cdot;100)}$ | 0.320<br>(0.189, 0.437)  | 6.321<br>(1.81, 12.584)     |                         |                         |

Table 4: Posterior means and credibility intervals in parentheses, for the exponential, multiquadric and Legendre-Matérn correlation functions fit to LENS JJA surface temperature means. The parameter  $\phi$  corresponds to the scale parameter of the exponential and Legendre-Matérn, the parameters  $p$  and  $\tau$  are those denoted in Table 1 for the multiquadric, and  $\nu$  refers to the smoothness parameter of the Legendre-Matérn. The table also includes posterior means and credibility intervals for GEM parameters.

defined in second entry in Table 1, and the Legendre-Matérn correlation function (Guinness and Fuentes, 2016) with truncation after the first 50 terms. Finally, we consider two elements of the class  $\Psi_\infty$  that do not admit a  $d$ -Schoenberg expansion in closed form: the power exponential model (see Gneiting, 2013, for details) and the generalized Cauchy class (Gneiting and Schlather, 2004). Table 4 shows the parameter estimates (posterior means) and their associated 95% credibility intervals. Note that only the results for the exponential correlation,  $\psi_E$ , the multiquadric,  $\psi_M$ , and the Legendre-Matérn,  $\psi_{LM}$  are shown, as these were the best performing.

Table 5 reports the average predictive scores RMSE and CRPS, across the 90 spatial replicates, for the three versions of the nonparametric model  $\psi_{\text{BNP}}$  together with the other parametric models.

As expected, the additional flexibility of  $\psi_{\text{BNP}(\cdot;100)}$ , afforded by the larger number of Schoenberg coefficients, allows it to outperform the other  $\psi_{\text{BNP}(\cdot;K)}$  models,  $K = 30, 40, 50$ . The posterior means and 95% credibility intervals for  $\alpha$  on the one hand clearly suggest a departure from the simplistic case of the Dirichlet process (*i.e.*, when  $\alpha = 0$ ), especially when  $K = 100$ , as seen from the last two rows in Table 4. In fact, when  $K = 100$  this posterior distribution is clearly shifted to the right compared to the other cases. On the other hand, the posterior of  $\vartheta$  is concentrated on fairly large values, suggesting that all the terms in the truncated mixture are important to achieve the approximation required by the data. When  $K = 100$ , this pos-

| Model                           | RMSE                 | CRPS                 |
|---------------------------------|----------------------|----------------------|
| $\psi_{\text{BNP}}(\cdot; 30)$  | 0.709 (0.580, 0.913) | 0.620 (0.562, 0.702) |
| $\psi_{\text{BNP}}(\cdot; 40)$  | 0.659 (0.522, 0.786) | 0.597 (0.540, 0.655) |
| $\psi_{\text{BNP}}(\cdot; 50)$  | 0.647 (0.515, 0.813) | 0.590 (0.538, 0.651) |
| $\psi_{\text{BNP}}(\cdot; 100)$ | 0.640 (0.503, 0.788) | 0.585 (0.527, 0.659) |
| $\psi_{\text{E}}$               | 0.650 (0.532, 0.806) | 0.591 (0.542, 0.656) |
| $\psi_{\text{M}}$               | 0.654 (0.535, 0.805) | 0.594 (0.542, 0.651) |
| $\psi_{\text{LM}}$              | 0.659 (0.539, 0.810) | 0.597 (0.546, 0.652) |

Table 5: Average predictive scores over the 90 spatial replicates, for the nonparametric model  $\psi_{\text{BNP}}(\cdot; K)$  with truncation levels  $K = 30, 40, 50, 100$ , and the chosen parametric models. 95% central range of values indicated in parentheses.

terior is somewhat shifted to the left compared to  $K = 30, 40, 50$ . It is also interesting to note that the model performs better than the parametric counterparts in terms of both RMSE and CRPS. With respect to the parametric models, the Legendre-Matérn has  $E(\nu | \mathbf{Z}) = 0.197$  suggesting that the process is not mean-square differentiable, and thus explaining the better fit of the exponential and the Legendre-Matérn; whereas the multiquadric, which is twice differentiable at the origin, underperforms. In terms of predictive performance, we find that the exponential model slightly outperforms the other two. We point out that the lack of smoothness exhibited by the data, and empirically corroborated by the posterior summaries in Table 4, makes it generally harder for a mixture of smooth functions to produce good approximations as required by the available data. In practice, this is reflected in the need to employ an increased number of mixture terms compared to other datasets that do not exhibit this problem. For most problems, a truncation at 50 mixture terms would be more than adequate (Ishwaran and James, 2001), but interestingly, the results exhibited in Table 5 show that for these data we actually need to “get to the limit” and use more than  $K = 50$  terms to obtain a clearly better performance with respect to competitors.

Finally, we consider estimation of the spatial correlation function using the proposed model and also the parametric alternatives mentioned earlier. Figure 6 shows the empirical spatial correlation function obtained from package `ncf` available for the R software. All the curves are joined to help graphical display. The posterior mean of the spatial correlation functions for several of the methods considered earlier (exponential, Legendre-Matérn, multiquadratic) are also included for comparison. Recall that the exponential correlation model (dotted line) gave the best parametric estimate according to Table 5. Even though the exponential correlation estimate generally follows the empirical curve, it cannot reproduce the ups and downs observed in the solid curve. Considering now the BNP estimate, the dashed line corresponds to the posterior mean of the correlation function obtained from  $\psi_{\text{BNP}}(\cdot; 100)$  (estimates are virtually identical to the case of  $\psi_{\text{BNP}}(\cdot; 50)$ , not shown here). Observe how the BNP estimate closely follows the empirically computed function, thus denoting the flexibility of the nonparametric construction and model we have studied here.

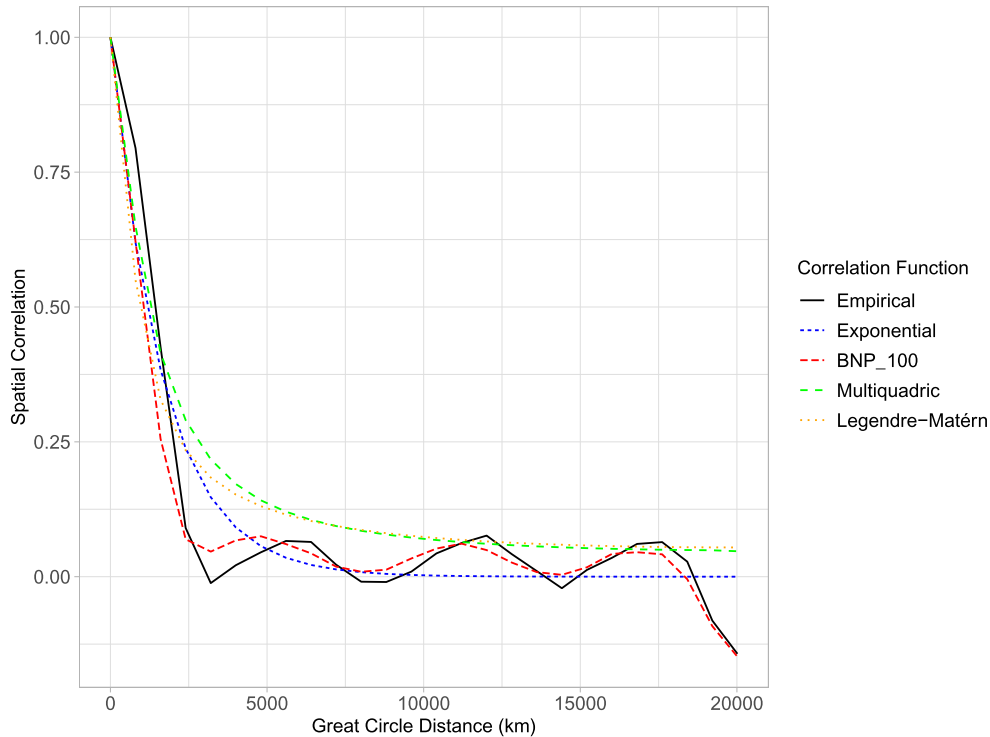


Figure 6: Estimated correlation functions under the various models considered in this analysis. The solid curve represents the empirical spatial correlation function, estimated using package `ncf` from R.

## 8 Conclusion and Discussion

This paper provides a novel perspective on covariance functions on spheres. We have shown that a nonparametric Bayesian approach offers a valid alternative to parametric forms from the point of view of statistical accuracy and optimal linear prediction. Our theoretical findings have been supported by a satisfactory simulation study and by a data analysis on global surface temperatures.

This work poses a series of challenges that may be of interest to pursue in the future. One of them is the nonparametric Bayesian modeling of space-time covariance functions, for which the spectral representation becomes quite involved. Surely, it would be worthwhile to extend the proposed approach to multivariate random fields, for which modeling marginal and cross spectra will become necessary.

The approach proposed in this paper has an interesting connection with integration methods that have been recently proposed in the machine learning community, and a good contribution would be to show the advantages of our approach within that community.



## Supplementary Material

Supplementary Materials File to “Nonparametric Bayesian Modeling and Estimation of Spatial Covariance Functions for Global Data” (DOI: [10.1214/20-BA1228SUPP](https://doi.org/10.1214/20-BA1228SUPP); .pdf).

## References

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. Courier Corporation. [MR0208797](#). 5, 10
- Arafat, A., Porcu, E., Bevilacqua, M., and Mateu, J. (2018). “Equivalence and Orthogonality of Gaussian Measures on Spheres.” *Journal of Multivariate Analysis*, 267: 306–318. [MR3830648](#). doi: <https://doi.org/10.1016/j.jmva.2018.05.005>. 3, 15
- Banerjee, S. (2005). “On Geodetic Distance Computations in Spatial Modeling.” *Biometrics*, 61: 617–625. [MR2140936](#). doi: <https://doi.org/10.1111/j.1541-0420.2005.00320.x>. 2
- Beatson, R. K., zu Castell, W., and Xu, Y. (2014). “Pólya Criterion for (strict) Positive Definiteness on the sphere.” *IMA Journal of Numerical Analysis*, 34: 550–568. [MR3194799](#). doi: <https://doi.org/10.1093/imanum/drt008>. 3
- Berg, C. and Porcu, E. (2017). “From Schoenberg coefficients to Schoenberg functions.” *Constructive Approximation*, 45: 217–241. [MR3619442](#). doi: <https://doi.org/10.1007/s00365-016-9323-9>. 3, 5, 10, 12
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). “Stan: A probabilistic programming language.” *Journal of statistical software*, 76(1). 16
- Chopin, N., Rousseau, J., and Liseo, B. (2013). “Computational Aspects of Bayesian Spectral Density Estimation.” *Journal of Computational and Graphical Statistics*, 22(3): 533–557. [MR3173730](#). doi: <https://doi.org/10.1080/10618600.2013.785293>. 2
- Daley, D. J. and Porcu, E. (2013). “Dimension Walks and Schoenberg Spectral Measures.” *Proceedings of the American Mathematical Society*, 141: 1813–1824. [MR3168486](#). doi: <https://doi.org/10.1090/S0002-9939-2014-11894-6>. 5
- Dashti, M. and Stuart, A. M. (2017). “The Bayesian approach to inverse problems.” In *Handbook of uncertainty quantification. Vol. 1, 2, 3*, 311–428. Springer, Cham. [MR3839555](#). 13
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer. [MR2293378](#). 2, 10
- Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). “Generalized Spatial Dirichlet Process Models.” *Biometrika*, 94(4): 809–825. [MR2416794](#). doi: <https://doi.org/10.1093/biomet/asm071>. 2

- Ferguson, T. S. (1973). “A Bayesian analysis of some nonparametric problems.” *Annals of Statistics*, 1: 209–230. [MR0350949](#). 9
- Ferguson, T. S. (1974). “Prior distributions on spaces of probability measures.” *Annals of Statistics*, 2: 615–629. [MR0438568](#). 9
- Freedman, D. A. (1963). “On the asymptotic behavior of Bayes’ estimates in the discrete case.” *Annals of Mathematical Statistics*, 34: 1386–1403. [MR0158483](#). doi: <https://doi.org/10.1214/aoms/1177703871>. 9
- Gelfand, A., Kottas, A., and MacEachern, S. N. (2005). “Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing.” *Journal of the American Statistical Association*, 100(471): 1021–1035. [MR2201028](#). doi: <https://doi.org/10.1198/016214504000002078>. 2
- Gneiting, T. (2013). “Strictly and Non-Strictly Positive Definite Functions on Spheres.” *Bernoulli*, 19(4): 1327–1349. [MR3102554](#). doi: <https://doi.org/10.3150/12-BEJSP06>. 2, 3, 4, 5, 6, 7, 14, 22
- Gneiting, T. and Schlather, M. (2004). “Stochastic Models That Separate Fractal Dimension and the Hurst Effect.” *SIAM Review*, 46: 269–282. [MR2114455](#). doi: <https://doi.org/10.1137/S0036144501394387>. 22
- Guinness, J. and Fuentes, M. (2016). “Isotropic Covariance Functions on Spheres: some Properties and Modeling Considerations.” *Journal of Multivariate Analysis*, 143: 143–152. [MR3431424](#). doi: <https://doi.org/10.1016/j.jmva.2015.08.018>. 3, 22
- 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(5): 982–1002. [MR3750634](#). doi: <https://doi.org/10.1080/00949655.2017.1416470>. 2
- Ishwaran, H. and James, L. F. (2001). “Gibbs Sampling Methods for Stick-Breaking Priors.” *Journal of the American Statistical Association*, 96(453): 161–173. [MR1952729](#). doi: <https://doi.org/10.1198/016214501750332758>. 8, 16, 23
- Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). “Bayesian semiparametric inference for multivariate doubly-interval-censored data.” *The Annals of Applied Statistics*, 4(4): 2126–2149. [MR2829950](#). doi: <https://doi.org/10.1214/10-AOAS368>. 14
- Jeong, J., Castruccio, S., Crippa, P., and Genton, M. G. (2017). “Statistics-Based Compression of Global Wind Fields.” Available at arXiv: [1702.01995v3](https://arxiv.org/abs/1702.01995v3). 2
- Jeong, J. and Jun, M. (2015). “A class of Matérn-like covariance functions for smooth processes on a sphere.” *Spatial Statistics*, 11: 1–18. [MR3311853](#). doi: <https://doi.org/10.1016/j.spasta.2014.11.001>. 20
- Jun, M., Stein, M. L., et al. (2008). “Nonstationary covariance models for global data.” *The Annals of Applied Statistics*, 2(4): 1271–1289. [MR2655659](#). doi: <https://doi.org/10.1214/08-AOAS183>. 20

- Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., et al. (2015). “The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability.” *Bulletin of the American Meteorological Society*, 96(8): 1333–1349. 19
- Lang, A. and Schwab, C. (2013). “Isotropic Random Fields on the Sphere: Regularity, Fast Simulation and Stochastic Partial Differential Equations.” *Annals of Applied Probability*, 25: 3047–3094. MR3404631. doi: <https://doi.org/10.1214/14-AAP1067>. 3, 11
- Lang, A. and Schwab, C. (2015). “Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations.” *The Annals of Applied Probability*, 25: 3047–3094. MR3404631. doi: <https://doi.org/10.1214/14-AAP1067>. 11
- Lindgren, F., Rue, H., and Lindström, J. (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. MR2853727. doi: <https://doi.org/10.1111/j.1467-9868.2011.00777.x>. 1
- Massa, E., Peron, A., and Porcu, E. (2017). “Positive Definite Functions on Complex Spheres, and their Walks through Dimensions.” *SIGMA*, 13: 088. MR3720944. doi: <https://doi.org/10.3842/SIGMA.2017.088>. 3
- Menegatto, V. A., Oliveira, C. P., and Peron, A. P. (2006). “Strictly Positive Definite Kernels on Subsets of the Complex Plane.” *Computer Mathematics and Its Applications*, 51: 1233–1250. MR2235825. doi: <https://doi.org/10.1016/j.camwa.2006.04.006>. 3, 4
- Møller, J., Nielsen, M., Porcu, E., and Rubak, E. (2018). “Determinantal Point Process Models on the Sphere.” *Bernoulli*, 24(2): 1171–1201. Available at arXiv: 1607.03675. MR3706791. doi: <https://doi.org/10.3150/16-BEJ896>. 3, 6, 7
- Müller, P., Quintana, F. A., and Page, G. (2018). “Nonparametric Bayesian inference in applications.” *Statistical Methods & Applications*, 27(2): 175–206. MR3807363. doi: <https://doi.org/10.1007/s10260-017-0405-z>. 2
- Pitman, J. (1996). “Some developments of the Blackwell–MacQueen urn scheme.” In *Statistics, Probability and Game Theory*, volume 30 of *IMS Lecture Notes Monogr. Ser.*, 245–267. Hayward, CA: Inst. Math. Statist. MR1481784. doi: <https://doi.org/10.1214/lnms/1215453576>. 8, 9
- Pitman, J. (2006). *Combinatorial Stochastic Processes*. Berlin: Springer. 9
- Pitman, J. and Yor, M. (1997). “The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator.” *Annals of Probability*, 25(2): 855–900. MR1434129. doi: <https://doi.org/10.1214/aop/1024404422>. 3, 9
- Porcu, E., Alegria, A., and Furrer, R. (2018a). “Modeling Temporally Evolving and

- Spatially Globally Dependent Data.” *International Statistical Review*, 86(2): 344–377. MR3852415. doi: <https://doi.org/10.1111/insr.12266>. 2, 3, 4
- 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(514): 888–898. MR3538713. doi: <https://doi.org/10.1080/01621459.2015.1072541>. 2, 3, 6
- Porcu, E. Bissiri, P. G., Tagle, F., Soza, R., and Quintana, F. A. (2020). “Supplementary Materials File to “Nonparametric Bayesian Modeling and Estimation of Spatial Covariance Functions for Global Data”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/20-BA1228SUPP>. 4
- Porcu, E., Castruccio, S., Alegria, A., and Crippa, P. (2018b). “Axially Symmetric Models for Global Data: a Journey between Geostatistics and Stochastic Generators.” *Technical Report, Newcastle University*. 2
- Reich, B. J. and Fuentes, M. (2012). “Nonparametric Bayesian models for a spatial covariance.” *Statistical Methodology*, 9(1): 265–274. Special Issue on Astrostatistics + Special Issue on Spatial Statistics. URL <http://www.sciencedirect.com/science/article/pii/S1572312711000104> 2
- 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(3): 743–758. MR1998632. doi: <https://doi.org/10.1111/1467-9868.00413>. 2
- Schoenberg, I. J. (1942). “Positive definite functions on spheres.” *Duke Mathematical Journal*, 9(1): 96–108. MR0005922. doi: <https://doi.org/10.1215/S0012-7094-42-00908-6>. 3, 5
- Soubeyrand, S., Enjalbert, J., and Sache, I. (2008). “Accounting for roughness of circular processes: Using Gaussian random processes to model the anisotropic spread of airborne plant disease.” *Theoretical Population Biology*, 73: 92–103. 6
- Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York. MR1697409. doi: <https://doi.org/10.1007/978-1-4612-1494-6>. 2
- Stein, M. L. (2007). “Spatial variation of total column ozone on a global scale.” *The Annals of Applied Statistics*, 1(1): 191–210. MR2393847. doi: <https://doi.org/10.1214/07-AOAS106>. 20
- Stuart, A. M. (2010). “Inverse problems: a Bayesian perspective.” *Acta Numerica*, 19: 451–559. MR2652785. doi: <https://doi.org/10.1017/S0962492910000061>. 13
- Zheng, Y., Zhu, J., and Roy, A. (2010). “Nonparametric Bayesian inference for the spectral density function of a random field.” *Biometrika*, 97(8): 238–245. MR2594432. doi: <https://doi.org/10.1093/biomet/asp066>. 2
- Ziegel, J. (2014). “Convolution Roots and Differentiability of Isotropic Positive Definite

Functions on Spheres.” *Proceedings of the American Mathematical Society*, 142: 2053–2077. MR3182025. doi: <https://doi.org/10.1090/S0002-9939-2014-11989-7>. 3

### **Acknowledgments**

We thank all the reviewers and Editors for comments that greatly improved the current version of this manuscript. Emilio Porcu is partially supported by Proyecto Fondecyt number 1170290 from the Chilean Commission for Scientific and Technological Research. Fernando A. Quintana is partially supported by Proyecto Fondecyt number 1180034 from the Chilean Commission for Scientific and Technological Research. Fernando A. Quintana, Rubén Soza and Emilio Porcu are supported by Iniciativa Científica Milenio – Minecon Núcleo Milenio MIDAS.