Covariance models on the surface of a sphere: when does it matter?

Jaehong Jeong and Mikyoung Jun

Received 8 April 2015; Accepted 12 May 2015

There is a growing interest in developing covariance functions for processes on the surface of a sphere because of the wide availability of data on the globe. Utilizing the one-to-one mapping between the Euclidean distance and the great circle distance, isotropic and positive definite functions in a Euclidean space can be used as covariance functions on the surface of a sphere. This approach, however, may result in physically unrealistic distortion on the sphere especially for large distances. We consider several classes of parametric covariance functions on the surface of a sphere, defined with either the great circle distance or the Euclidean distance, and investigate their impact upon spatial prediction. We fit several isotropic covariance models to simulated data as well as real data from National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis on the sphere. We demonstrate that covariance functions originally defined with the Euclidean distance may not be adequate for some global data. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: covariance function; Euclidean distance; geopotential height; great circle distance; process on a sphere

1 Introduction

In geophysical and environmental sciences, data often come in a global scale. Covariance functions for global data sets need to be positive definite on the surface of a sphere, and the distance computation is important in spatial modelling. For an integer $d \geq 1$, define $S_d^2 = \{ x \in \mathbb{R}^{d+1} : ||x|| = r \}$ to be a $(d$-dimensional) sphere with radius $r$, where $||x||$ is the Euclidean norm of $x \in \mathbb{R}^{d+1}$. We also define the great circle distance on $S_d^2$ by $\theta(x, y) = r \times \arccos(\langle x, y \rangle)$ where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^{d+1}$.

We consider the surface of Earth as the spatial domain. Let $S^2_\mathbb{R}$ denote the surface of Earth, where $R$ denotes Earth’s radius (we approximate Earth as a perfect sphere). The chordal distance between the two points, $s_1 = (L_1, l_1)$ and $s_2 = (L_2, l_2)$, on $S^2_\mathbb{R}$ ($L$ and $l$ denote latitude and longitude, respectively) is given by

$$\text{ch}(s_1, s_2) = 2R \left[ \sin^2 \left( \frac{(L_1 - L_2)}{2} \right) + \cos L_1 \cos L_2 \sin^2 \left( \frac{(l_1 - l_2)}{2} \right) \right]^{1/2}.$$  

The great circle distance between the two locations then is given by $\theta = \theta(s_1, s_2) = 2R \times \arcsin (\text{ch}(s_1, s_2)/(2R))$. The chordal distance is simply the Euclidean distance penetrating the spatial domain on the surface of Earth and producing a straight-line approximation to the great circle distance (Banerjee, 2005).

Yadrenko (1983) pointed out that any covariance function in $\mathbb{R}^3$ can be considered as a covariance function for processes on $S^2_\mathbb{R}$ using the chordal (i.e. Euclidean) distance. This construction can provide a rich class of covariance functions on $S^2_\mathbb{R}$ (Gneiting, 1999). As argued by Gneiting (2013), the great circle distance is a physically most natural
distance metric for processes on a sphere. However, literature on covariance modelling using the great circle distance on the surface of a sphere is scarce because of its mathematical challenge. Some efforts have been made in examining the validity of several parametric covariance functions on the surface of a sphere (Huang et al., 2011; Gneiting, 2013) and in developing valid parametric covariance functions with the great circle distance from various constructional approaches (Du & Ma, 2012; Du et al., 2013; Gneiting, 2013; Guinness & Fuentes, 2013; Jeong & Jun, 2015).

Although Huang et al. (2011) and Gneiting (2013) studied validity of covariance functions defined with either the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well.

In this paper, we consider several positive definite functions on \( \mathbb{R}^{d+1} \) and \( S_d^2 \) for \( d = 1, 2 \) and compare them in simulation studies and real-data applications. The rest of the paper is organized as follows. In Section 2, we discuss some characteristics of covariance functions on a sphere. Then we present two simulation studies on \( S^1 \) and \( S^2 \) in Section 3. Section 4 illustrates real application results to geopotential height data set. Finally, Section 5 concludes the paper with discussion.

## 2 Characteristics of covariance functions on a sphere

We first review some known results on covariance functions in the Euclidean space as well as those on the surface of a sphere. A function \( f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) is called positive definite if

\[
\sum_{i,j=1}^n c_i c_j f(x_i, x_j) \geq 0 \tag{1}
\]

for all finite \( n \in \mathbb{N} \), all distinct points \( x_1, \ldots, x_n \in \mathbb{R}^d \), and all real \( c_1, \ldots, c_n \). A function is strictly positive definite when the inequality in (1) is strict (unless \( c_1 = c_2 = \cdots = c_n = 0 \)). For a real random field \( Z \) in \( \mathbb{R}^d \) with \( E(Z(x))^2 < \infty \) for all \( x \in \mathbb{R}^d \), the covariance function \( K(x, y) = \text{cov}(Z(x), Z(y)) \) must satisfy the condition in (1). The random field \( Z \) is called weakly stationary if its mean function is constant, it has finite second moments and its covariance function can be written as \( \text{cov}(Z(x), Z(y)) = K(x - y) \) for all \( x, y \in \mathbb{R}^d \) and a positive definite function \( K \), that is, the covariance function of \( Z \), depends on \( x \) and \( y \) only through \( x - y \). Furthermore, if its covariance function satisfies \( \text{cov}(Z(x), Z(y)) = \varphi(||x - y||) \) for a positive definite function \( \varphi \), then the random field \( Z \) is weakly isotropic. An isotropic property for processes in \( \mathbb{R}^d \) can be thought as an invariance property under translation and rotation (Stein, 1999).

The covariance function of a random field and the smoothness of its realization are related to mean square properties of the random field. The random field \( Z \) is called mean square continuous at \( x \) if

\[
E(Z(y) - Z(x))^2 \rightarrow 0 \quad \text{as} \quad y \rightarrow x.
\]

For weakly stationary random field, mean square continuity is equivalent to continuity of the covariance function at the origin, but it does not imply continuity of its realization (Stein, 1999). Moreover, a random field \( Z \) on \( \mathbb{R} \) with finite second moments is mean square differentiable at \( t \) if there exists \( Z'(t) = \lim_{h_n \rightarrow 0} (Z(t + h_n) - Z(t))/h_n \) in \( L^2 \) for sequences \( h_n \rightarrow 0 \). The smoothness of a random field can be determined through the number of mean square derivatives. Gneiting (2013) defined the class of \( \Phi_d \) with the correlation functions of mean square continuous, stationary and isotropic random fields in \( \mathbb{R}^d \). Every positive definite function \( \varphi : [0, \infty) \rightarrow \mathbb{R} \) with \( \varphi(0) = 1 \) is the correlation of an isotropic process, and the members of \( \Phi_2 \) and \( \Phi_3 \) are the cornerstones for covariance models for spatial data in a planar domain (Gneiting, 2013). An isotropic property on a sphere means that the covariance
function depends on distance only. That is, a random field $Z$ on $S^2$ is called isotropic if its covariance function satisfies $\text{cov}(Z(\mathbf{x}), Z(\mathbf{y})) = \psi(\mathbf{t})$ for all $\mathbf{x}, \mathbf{y} \in S^2$. We then similarly define $\Psi_d$, the class of continuous, isotropic covariance functions $\psi: [0, \pi \times r] \rightarrow \mathbb{R}$ on $S^2$.

Because a sphere can be viewed as a subset of the Euclidean space, valid covariance functions on $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ can be restricted to $S^2 \times S^2$. This implies that valid correlation functions on $\mathbb{R}^d$ is called isotropic if its covariance function satisfies $\text{cov}([Z(\mathbf{x})], [Z(\mathbf{y})]) = \psi(\mathbf{t})$. We then similarly define $\Psi_{d+1}$, the class of continuous, isotropic covariance functions $\psi: [0, \pi \times r] \rightarrow \mathbb{R}$ on $S^2$.

We now list several parametric classes of covariance functions defined on the surface of a sphere or defined originally in the Euclidean space (then can be used through the projection). Some of the models are used in the simulation and data examples. We only consider isotropic covariance models on a sphere. Functions in the class $\Psi_{d+1}$ are characterized in terms of an infinite sum of Gegenbauer polynomials with non-negative coefficients and cosine of the great circle distance $\theta$. The class $\Phi_{d+1}$ is the set of functions $\phi$ that are absolutely integrable in $S^2$, that is, a function $\phi$ is in $\Phi_{d+1}$ if

$$
\int_{S^2} |\phi(\mathbf{t})| d\mathbf{t} < \infty.
$$

We also focus on the fact that on the sphere, the correlation between two points with large distance apart may not necessarily be small (compared with the correlation between nearby two points). In fact, if there is a wave travelling around the sphere, two points nearly maximum possible distance apart may be perfectly positively (or negatively) correlated, which cannot happen in the Euclidean space.

We also focus on the fact that there are some fundamental differences between covariance models originally defined on the surface of a sphere and those in the Euclidean space. For instance, there exists a lower bound on isotropic correlation function in the Euclidean space. A function $\varphi$ is an isotropic correlation function in the Euclidean space if and only if it has the form, $\varphi(t) = \int_0^\infty \Lambda_d(u) dG(u)$, where $\int_0^\infty dG(u) = 1$ and $G$ is non-decreasing. Note that

$$
\Lambda_d(r) = 2^{(d-2)/2} \Gamma(d/2) r^{(d-2)/2} J_{(d-2)/2}(r)
$$

(Stein, 1999). This implies that valid correlation functions on $S^2$ constructed through the mapping described earlier from $\varphi$ cannot have values less than $\inf_{s \geq 0} \Lambda_d(s)$ or $-0.218$. In particular, the Matérn class yields non-negative correlations only. Although the importance of the Matérn family is highlighted by Stein (1999) because of its flexibility with regard to the local behaviour of the processes, it might not be appropriate in applications where there is a significantly negative spatial correlation. In fact, many of the isotropic covariance functions in $\mathbb{R}^d$ used in the literature take non-negative values only.

We now list several parametric classes of covariance functions defined on the surface of a sphere or defined originally in the Euclidean space (then can be used through the projection). Some of the models are used in the simulation and data examples. We only consider isotropic covariance models on a sphere. Functions in the class $\Psi_{d+1}$ are characterized in terms of an infinite sum of Gegenbauer polynomials with non-negative coefficients and cosine of the great circle distance $\theta$. The class $\Phi_{d+1}$ is the set of functions $\phi$ that are absolutely integrable in $S^2$, that is, a function $\phi$ is in $\Phi_{d+1}$ if

$$
\int_{S^2} |\phi(\mathbf{t})| d\mathbf{t} < \infty.
$$

We also focus on the fact that on the sphere, the correlation between two points with large distance apart may not necessarily be small (compared with the correlation between nearby two points). In fact, if there is a wave travelling around the sphere, two points nearly maximum possible distance apart may be perfectly positively (or negatively) correlated, which cannot happen in the Euclidean space.

We now list several parametric classes of covariance functions defined on the surface of a sphere or defined originally in the Euclidean space (then can be used through the projection). Some of the models are used in the simulation and data examples. We only consider isotropic covariance models on a sphere. Functions in the class $\Psi_{d+1}$ are characterized in terms of an infinite sum of Gegenbauer polynomials with non-negative coefficients and cosine of the great circle distance $\theta$. The class $\Phi_{d+1}$ is the set of functions $\phi$ that are absolutely integrable in $S^2$, that is, a function $\phi$ is in $\Phi_{d+1}$ if

$$
\int_{S^2} |\phi(\mathbf{t})| d\mathbf{t} < \infty.
$$

We also focus on the fact that on the sphere, the correlation between two points with large distance apart may not necessarily be small (compared with the correlation between nearby two points). In fact, if there is a wave travelling around the sphere, two points nearly maximum possible distance apart may be perfectly positively (or negatively) correlated, which cannot happen in the Euclidean space.

We now list several parametric classes of covariance functions defined on the surface of a sphere or defined originally in the Euclidean space (then can be used through the projection). Some of the models are used in the simulation and data examples. We only consider isotropic covariance models on a sphere. Functions in the class $\Psi_{d+1}$ are characterized in terms of an infinite sum of Gegenbauer polynomials with non-negative coefficients and cosine of the great circle distance $\theta$. The class $\Phi_{d+1}$ is the set of functions $\phi$ that are absolutely integrable in $S^2$, that is, a function $\phi$ is in $\Phi_{d+1}$ if

$$
\int_{S^2} |\phi(\mathbf{t})| d\mathbf{t} < \infty.
$$

We also focus on the fact that on the sphere, the correlation between two points with large distance apart may not necessarily be small (compared with the correlation between nearby two points). In fact, if there is a wave travelling around the sphere, two points nearly maximum possible distance apart may be perfectly positively (or negatively) correlated, which cannot happen in the Euclidean space.
For $d \geq 1$, the class $\Psi_d$ consists of the functions of the form

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

with non-negative coefficients $b_{n,d}$ such that $\sum_{n=0}^{\infty} b_{n,d} = 1$ and the Gegenbauer polynomial of degree $n$, $c_n^{(d-1)/2}$ (Schoenberg, 1942; Chen et al., 2003). Moreover, the class $\Psi_\infty$ consists of the functions with the following form

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

(3)

with non-negative coefficients $b_n$ such that $\sum_{n=0}^{\infty} b_n = 1$. The infinite sum is strictly positive definite on $S^d_r$ when the coefficients $b_{n,d}$ and $b_n$ are strictly positive for infinitely many odd and infinitely many even integers $n$, and only a few closed forms, such as the multiquadratic family, for such infinite sums are known in general. The multiquadratic covariance function (Gneiting, 2013) is defined by

$$
\psi(\theta) = \sigma^2 (1 - r)^{2\nu} / \{1 + \tau^2 - 2\tau \cos(\theta/r)\}^{\nu}, \quad \theta \in [0, \pi \times r]
$$

from a standard Taylor series of (3), when $\sigma^2 > 0$, $c > 0$ and $r \in (0, 1)$. The Matérn class, given as

$$
\varphi(t) = \sigma^2 2^{\nu-1} \Gamma(\nu)^{-1} (t/\alpha)^\nu K_\nu(t/\alpha), \quad t \geq 0,
$$

(4)

where the parameters, $\sigma^2, \alpha, \nu > 0$, are marginal variance, spatial range and smoothness parameters, respectively, is positive definite in $\mathbb{R}^d$ for any $d \in \mathbb{N}$ with the Euclidean distance (Stein, 1999). Here, $K_\nu$ is the modified Bessel function of the second kind of order $\nu$.

Gneiting (2013) showed that completely monotone functions (that have derivatives $\varphi^{(k)}$ of all orders with $(-1)^k \varphi^{(k)}(t) \geq 0$ for all non-negative integers $k$ and all positive $t$) including the power exponential, Matérn and generalized Cauchy families are positive definite, through the restriction of a function $\varphi : [0, \infty) \rightarrow \mathbb{R}$ to the interval $[0, \pi \times r]$:

$$
\psi = \varphi_{[0, \pi \times r]} \text{ under applicable conditions, on } S^d_r \text{ of any dimension. One necessary condition of the membership in the class } \Psi_d \text{ is that either the fractal index or the smoothness parameter requires to satisfy } \beta \in (0, 1) \text{ or } \nu \in (0, 0.5), \text{ respectively.}
$$

The powered exponential family defined by

$$
\psi(\theta) = \sigma^2 \exp\{-\{\theta/(cr)\}^\beta\}, \quad \theta \in [0, \pi \times r]
$$

where $\sigma^2 > 0$, $c > 0$ is valid on any dimensional $S^d_r$ if $\beta \in (0, 1)$, and the Matérn family requires $\nu \in (0, 0.5]$ similarly. Jeong & Jun (2015) compare Matérn class with great circle distance to that with the Euclidean distance for spatial data on the surface of a sphere.

Compactly supported members of the class $\Phi_3$ may be valid on $S^d_r$ through $\psi = \varphi_{[0, \pi \times r]}$ (Gneiting, 2013). The spherical and the Wendland’s functions on $\Phi_3$, remain valid with direct substitution of the Euclidean distance by the great circle distance on $\Psi_d$, $d = 1, 2, 3$. In particular, the $C^4$-Wendland covariance function (Wendland, 1995), defined as

$$
\psi(\theta) = \sigma^2 (1 + (\theta \tau)/(cr) + \theta^2 (\tau^2 - 1)/(3c^2 \tau^2)) \{1 - \theta/(cr)\}^c, \quad \theta \in [0, \pi \times r],
$$

(5)

where $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter, has four derivatives at the origin and thus may be suitable for smooth data on $S^d_r$.

The following covariance function (defined in the Euclidean space) models a hole effect,

$$
\varphi(t) = \sigma^2 (\alpha/t) \sin(t/\alpha), \quad t > 0,
$$

(6)

with $\varphi(0) = \sigma^2$, and it is called the wave covariance function. This function is able to deal with the situation where the correlation between two points far apart is bigger (in magnitude) than the correlation between two points closer, or
there is an oscillating pattern in the correlation function. This belongs to the class $\Phi_d$ for $d = 1, 2, 3$, but it is not valid on $S^2$ (Huang et al., 2011). Thus, this function can only be applied to the data on the sphere through the projection from the Euclidean space.

Schoenberg (1942) noted that the class $\Psi_d$ enjoys the useful closure properties. The class $\Psi_d$ is convex, closed under products and closed under limits, which are continuous. For example, if $\psi_1(\theta), \psi_2(\theta) \in \Psi_d$, then $\lambda \psi_1(\theta) + (1 - \lambda) \psi_2(\theta) \in \Psi_d$ for every $\lambda \in [0, 1]$ and $\psi_1(\theta) \times \psi_2(\theta) \in \Psi_d$. Moreover, if $\psi_n(\theta) \in \Psi_d$, $\psi_n(\theta) \to \psi(\theta)$ as $n \to \infty$, and $\psi(\theta)$ is continuous, then also $\psi(\theta) \in \Psi_d$. These closure properties of the class $\Psi_d$ offer additional flexibility to model negative correlations. If we use convex sums or products of parametric families of correlation functions on $S^d$, in terms of the great circle distance with a Legendre function of the form $\psi(\theta) = C_{n}^{d(1-1/2)} \cos(\theta/r)/C_{n}^{d(1-1/2)}$, including the case $\cos(\theta/r)$ for $d = 1, 2, 3$, we can easily model negative correlations (Gneiting, 2013).

The sine-power covariance function is defined by

$$
\psi(\theta) = \sigma^2 \{1 - \left(\sin \frac{\theta}{2}\right)^\beta\}, \quad \theta \in [0, \pi \times r],
$$

where the parameter $\beta \in (0, 2]$ corresponds to the fractal index and controls the smoothness of the process. This covariance function operates directly on a sphere (belongs to the class $\Psi_d$ for all dimensions). On the other hand, the cosine function, $\varphi(t) = \cos(nt)$, where $t \geq 0$, belongs to the class $\Phi_1$ only for any value of $n > 0$. On $S^d$, $
\psi(\theta) = \cos(n\theta/r)$ for $\theta \in [0, \pi \times r]$, is non-strictly positive definite for all dimensions when $n = 1$. For non-integer values of $n > 0$, $\psi(\theta)$ is not valid on $S^1$ and for integer values of $n \geq 2$, it is non-strictly positive definite only on $S^1$ but not on $S^2$ (Gneiting, 2013).

When there is no significantly negative correlation, and when the spatial range is small, we do not expect that using the great circle distance or the Euclidean distance may make a significant difference. Moreover, when the prediction location is surrounded by enough number of estimation locations, we do not expect significant difference between models defined on the sphere and models defined in the Euclidean space (and used through projection). Therefore, for the rest of the paper, we mainly consider the situations where either spatial range is large or when there are not many estimation locations close to the prediction locations. Such situation may arise in real application, for example, when we are interested in predicting over the ocean with most observations on the land or when we are interested in predicting near the poles with not many observations near the poles. We also consider cases where there are significantly negative correlations for large distance lags.

### 3 Simulation studies

We present two simulation studies on $S^1$ and $S^2$. In the first example (Section 3.1), the truth is generated from an exponential covariance model with the great circle distance. In the second example (Section 3.2), the truth is generated from oscillating Matérn covariance model implemented in R-INLA package (Martins et al., 2013).

#### 3.1. Example on $S^1$

We simulated mean zero Gaussian random fields on $S^1$ using exponential covariance models with various range parameters using the great circle distance. The exponential covariance function on $S^d$ is defined by

$$
\psi(\theta) = \sigma^2 \exp(-\theta/\alpha), \quad \theta \in [0, \pi \times r],
$$

where $\sigma^2 > 0$ is a variance and $\alpha > 0$ is a spatial range parameter. We then compared fitted results using the true covariance model as well as exponential covariance model with the Euclidean distance, that is, $\varphi(t) = \sigma^2 \exp(-t/\alpha)$, $t > 0$. Exponential covariance functions belong to both $\Phi_2$ and $\Psi_1$, so they are valid on the surface of a
sphere regardless of distance. We set the marginal variance \( \sigma^2 = 1 \) and considered the various spatial ranges \( \alpha = 2\pi, 1.5\pi, \pi/1.5, \pi/2, \pi/4 \). We randomly selected 100 locations (angles) for parameter estimation from \( (\pi/2, 3\pi/2) \) on a unit circle and 10 fixed and equally spaced locations for prediction from \( [0, \pi/2] \) as in Figure 1. We compared the covariance models in terms of prediction using the root mean square error as well as the two popular scoring rules (Gneiting & Raftery, 2007), the mean absolute error and the continuous ranked probability score (CRPS). We repeated this experiment 100 times: the sampling locations are different each time, and the prediction locations are fixed.

Figure 2(a)–(d) presents mean absolute error and CRPS from the two models, the exponential models using great circle and chordal distances, displayed against 10 fixed prediction locations, for the true spatial range values that are \( 2\pi \) and \( \pi/4 \). When \( \alpha = 2\pi \), we observe that the exponential model using the great circle distance performed significantly better than that using the chordal distance. Except for the prediction locations that are relatively close to sampling locations, there are considerable differences between two models in prediction errors. On the other hand, when \( \alpha = \pi/4 \), there is no significant difference between two models in both prediction errors, which agrees with findings of Guinness & Fuentes (2013) and Jeong & Jun (2015).

### 3.2. Example on \( S^2_R \)

We consider mean zero Gaussian random fields on the surface of Earth (with radius \( R = 6371 \) km) with a Matérn covariance function from oscillating stochastic partial differential equations models with the \texttt{spde1} class from the \texttt{R-INLA} version 0.0-1413638221. We set \( \sigma = 1, \kappa = 0.5, \tau = \kappa \times \sigma/\sqrt{4\pi}, \nu = 2 \) and \( \theta_{osc} = 0.3 \). Here \( \kappa \) is the spatial scale parameter, \( \tau \) controls the variance with \( \sigma \), \( \nu \) controls the smoothness of the process and \( \theta_{osc} \) controls the strength of oscillation. We subtracted the constant mean (average over all locations) to have mean zero residual fields as in Figure 3(a). There are 128 longitude points and 64 latitude points, and the size of the data is 8192. We randomly selected 300 locations where values are smaller than 0 for parameter estimation and 100 locations where values are larger than 1 for prediction. We repeated this procedure 100 times: all locations are different each time. It is clear from Figure 3(a) and (b) that values at large distance lags are negatively correlated. We compared fitted
Figure 2. (Simulation example on $S^1$) (a) Mean absolute error (MAE) and (b) mean continuous ranked probability score (CRPS), averaged over 100 replications displayed against prediction locations ($\theta$). The true spatial range is $\alpha = 2\pi$. (c) and (d) are the same as (a) and (b), except that $\alpha = \pi/4$. Triangles and circles represent the values of prediction errors for the exponential models using the great circle and Euclidean distances, respectively.

Figure 3. (Simulation example on $S^2$) (a) A realization of residual fields. (a) shows sampling locations (○) and prediction locations (×). (b) Empirical semivariogram values for selected locations versus the great circle distance. For (b), dotted line represents sample variance.
Table I. (Simulation example on $S^2_R$) Sample means and standard deviations of parameter estimates, maximum log-likelihood values and prediction errors (100 cases).

| Model | MC            | MG            | C             | H             |
|-------|---------------|---------------|---------------|---------------|
| $\hat{\sigma}^2$ | 6.654 (1.237) | 3.260 (0.448) | 26.033 (7.163) | 7.496 (1.711) |
| $\hat{\alpha}$ or $\hat{\lambda}$ | 7918.459 (1669.169) | 49423.925 (7016.223) | 0.999 (0.006) | 58.499 (56.164) |
| $\hat{\gamma}$ or $\hat{\beta}$ | 1.037 (0.054) | 0.500 (0.000) | 1.820 (0.038) | — |
| Max.loglik | 356.505 (11.217) | 298.233 (6.923) | 354.384 (10.533) | $-430.486$ (38.393) |
| RMSE | 1.123 (0.179) | 0.975 (0.085) | 0.874 (0.126) | 1.961 (0.369) |
| MAE | 0.903 (0.177) | 0.749 (0.065) | 0.717 (0.118) | 1.751 (0.337) |
| CRPS | 0.639 (0.096) | 0.573 (0.058) | 0.514 (0.055) | 1.166 (0.213) |

For model C, $\lambda \in (0, 1)$ is a weighting parameter.

results from a Matérn covariance model with the Euclidean distance (call it, MC) to those from a Matérn covariance model with the great circle distance (call it, MG). Moreover, to deal with negative correlations, we use convex sum of valid covariance functions with the great circle distance (call it, C):

$$\psi(\theta) = \sigma^2 \left[ \lambda \left(1 - \left(\frac{\sin \theta}{2R}\right)^\beta \right) + (1 - \lambda) \cos(\theta/R) \right], \quad \beta \in (0, 2] \text{ and } \lambda \in (0, 1).$$

We also consider the hole-effect model with the Euclidean distance (call it, H) defined in (6), a model defined with the Euclidean distance, for a comparison.

Table I contains results of parameter estimation and prediction for the various models considered. From Table I, we observe that both MC and MG have large estimates of the spatial range parameter and MG has smaller prediction errors than MC. Although MC has larger maximum log-likelihood values than MG, MC leads to poor prediction, possibly because of the large estimate of the spatial range. Note that the best model in terms of prediction errors is C. For C, the estimate of $\lambda$ is close to 1, and resulting estimated model is dominated by the sine-power model. Nevertheless, its correlation function allows much smaller correlation values for large distance lag, compared with the models for the Euclidean space. On the other hand, for H, although it allows negative correlations unlike Matérn covariance model, it resulted in poor model fit and spatial prediction. From Figure 4(a)–(d), we observe clear differences of prediction errors between models using the great circle and Euclidean distances when the prediction locations are relatively far away from their nearest sampling locations. Overall, C outperforms MC and MG in prediction.

4 Application

4.1. Data and mean structure

We consider geopotential height data on a global scale. The geopotential height approximates the actual height of a surface pressure at a certain level above mean sea level. It is an adjustment to geometric height using the variation of gravity with elevation and latitude. The study of the geopotential height might be important in learning abnormal weather phenomena. According to Hafez & Almazroui (2014), the geopotential height at level 500 hPa plays a dominant role in controlling weather and climate conditions. Moreover, it became evident that the variability of global geopotential height is clearly impacted by global warming and climatic indices over the last several decades (Marshall, 2002; Zhu et al., 2002; Hafez, 2012; Hafez & Almazroui, 2014).
Stat Covariance models on a sphere

The ISI’s Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.84

Figure 4. (Simulation example on $S^2$) Boxplots of differences of (a) absolute error (AE) and (c) continuous ranked probability score (CRPS) values from MC and MG, displayed against minimum great circle distance between a prediction location and its nearest sampling location. (b) and (d) are same as (a) and (c), except that selected models are MG and C. For (a)–(d), red circles represent average values for each bin.

Figure 5. (Data example I) Square root of the geopotential height at level 500 hPa (unit: $\sqrt{\text{m}}$).

Data sets are obtained at level 500 hPa from NCEP/NCAR reanalysis project (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html). The NCEP/NCAR reanalysis 1 project uses a state-of-the-art analysis/forecast system to perform data assimilation using past data from 1948 to the present (see Kalnay et al. 1996 for more detailed information). The output values are given on regular grids, and there are 144 longitude points and...
73 latitude points (the size of the data is 10,512). We used boreal summer geopotential height in the Northern hemisphere (June, July and August), and for each grid, we computed a pointwise mean as the average over 2014. The unit for the geopotential height is in metre, and the unit for the distance is in kilometre. For variance stabilization, we took a square root transformation of the data.

We decompose the data into its mean structure (large-scale variation) and the residual (for small-scale spatial variation). Figure 5 suggests clear large-scale spatial structure depending on latitude. Thus, we modelled the mean structure through simple harmonic regression:

\[ m(L) = a_0 + a_1 \cos(L \times \pi/90^\circ) + a_2 \sin(L \times \pi/90^\circ). \]  

We consider two cases: one is with a constant (unknown) mean (that is, \( a_1 = a_2 = 0 \) in (8)), and the other is given by (8). In both cases, we first estimate the mean structure using regression and then work with the residual to fit the covariance structure.

### 4.2. Example I: horizontal directional sampling design for prediction

Figure 6(a) shows the map of residuals after subtracting the constant mean, and Figure 6(b) shows the empirical semivariogram of residuals. The semivariogram clearly shows negative covariances for large distances. To save computational burden, we randomly selected 600 locations near the red region for parameter estimation and selected 200 locations over the blue region for prediction as in Figure 6(a). We repeated this process 100 times, and for each time, all locations are randomly sampled and thus different. We compared the covariance models, MC, MG and the convex sum model, C, with \( C^4 \)-Wendland covariance function (5) and cosine function. As shown in Table II, MG gives poor model fit and prediction. Although MC obtains the largest maximum log-likelihood value, it leads to poor prediction compared with C. This is expected as the Matérn model is not able to produce negative correlations. Similarly to the simulation example of Section 3.2, C is the best model in terms of prediction.

![Figure 6](image-url)
Table II. (Data example I - constant mean) Sample means and standard deviations of parameter estimates, maximum log-likelihood values and prediction errors (100 cases).

| Model | MC | MG | C |
|-------|----|----|---|
| $\hat{\sigma}^2$ | 1.410 (0.087) | 0.800 (0.115) | 2.240 (0.377) |
| $\hat{\lambda}$ or $\hat{\lambda}$ | 2128.225 (114.514) | 136377.194 (18394.328) | 0.811 (0.108) |
| $\hat{\nu}$ or $\hat{c}$ | 2.605 (0.072) | 0.500 (0.000) | 2.931 (0.084) |
| $\hat{r}$ | — | — | 7.826 (0.302) |
| Max. loglik | 2167.401 (20.798) | 1413.049 (11.086) | 2160.817 (20.348) |
| RMSE | 3.972 (0.137) | 6.939 (0.086) | 3.292 (0.186) |
| MAE | 3.853 (0.142) | 6.840 (0.090) | 3.168 (0.193) |
| CRPS | 3.427 (0.145) | 6.767 (0.089) | 2.650 (0.196) |

For model C, $c \in (0, \pi]$ is a support parameter and $r \geq 6$ is a shape parameter.

Figure 7. (Data example I - the constant mean) Boxplots of differences of (a) absolute error (AE) and (b) continuous ranked probability score (CRPS) values from MC and C, displayed against minimum great circle distance between a prediction location and its nearest sampling location. Red circles represent average values in each bin.

Figure 7(a) and (b) shows boxplots of differences of absolute error (AE) and CRPS values between MC and C, displayed against minimum great circle distance between a prediction location and its nearest sampling location. For all distances, C outperforms MC. Moreover, the differences of two prediction errors between two models increase as minimum distances between the prediction locations and their nearest sampling locations increase. On the other hand, Figure 8(a) presents residual fields after removing mean structure by using simple harmonic regression depending on latitude as in (8). We fitted $C^4$-Wendland covariance functions with the Euclidean distance (WC) and the great circle distance (WG) in addition to the covariance models considered previously with constant mean structure. For WC, the covariance function is defined by

$$\varphi(t) = \sigma^2 \left\{ 1 + (t\tau)/(cR) + t^2(\tau^2 - 1)/(3c^2R^2) \right\} \left\{ 1 - t/(cR) \right\}^{\frac{1}{2}},$$
Figure 8. (Data example I) (a) A realization of residual fields after removing mean structure through simple harmonic regression depending on latitude. (a) shows sampling locations (○) and prediction locations (×). (b) Empirical semivariogram values for selected locations displayed against the great circle distance. For (b), dotted line represents sample variance.

Table III. (Data example I - simple harmonic regression depending on latitude) Sample means and standard deviations of parameter estimates, maximum log-likelihood values and prediction errors (100 cases).

| Model | MC     | MG     | C       | WG      | WC      |
|-------|--------|--------|---------|---------|---------|
| $\hat{\sigma}^2$ | 0.776 (0.076) | 0.100 (0.001) | 0.820 (0.053) | 0.706 (0.036) | 0.770 (0.139) |
| $\hat{\lambda}$ or $\hat{\lambda}$ | 1770.346 (121.073) | 10754.167 (91.314) | 0.798 (0.056) | — | — |
| $\hat{\psi}$ or $\hat{c}$ | 2.699 (0.088) | 0.500 (0.000) | 2.887 (0.074) | 2.873 (0.020) | 9.794 (22.488) |
| $\hat{\tau}$ | — | — | 9.426 (0.225) | 9.237 (0.198) | 29.986 (68.228) |
| Max.loglik | 2171.231 (21.331) | 1242.183 (6.178) | 2168.634 (6.178) | 2168.689 (20.324) | 2168.653 (20.271) |
| RMSE | 2.441 (0.137) | 2.351 (0.057) | 2.177 (0.085) | 1.988 (0.074) | 2.184 (0.357) |
| MAE | 2.314 (0.142) | 2.217 (0.065) | 2.047 (0.090) | 1.856 (0.079) | 2.053 (0.365) |
| CRPS | 1.952 (0.135) | 2.100 (0.064) | 1.677 (0.084) | 1.493 (0.072) | 1.689 (0.351) |

For models C and WG, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.

where $\tau \geq 6$, $c > 0$ and $t > 0$. The sampling and prediction locations remained the same as in the previous example. From Table III, all models have comparable maximum log-likelihood values except MG. This may be due to the fact that the residual field is fairly smooth. C and WG give better prediction performances than MC and WC, respectively. Note that WC has much larger sample standard deviations of estimates for support and shape parameters than WG.

When comparing the two mean structures, prediction errors for the constant mean case are larger than those for the mean given by (8). Because the geopotential height data mainly show large scale, smooth, variation depending on latitude, the mean structure using simple harmonic regression resulted in improved prediction. Overall, the covariance functions of the class $\Psi_d$ performed better than those of the class $\Phi_{d+1}$ regardless of mean structures.
4.3. Example II: vertical directional sampling design for prediction

We entertain the same set of covariance models with the mean structures as in Section 4.2. However, we changed a sampling design for prediction. We randomly selected 600 locations where longitude is less than $0^\circ$ for parameter estimation and selected 200 locations where that is greater than $0^\circ$ for prediction as in Figure 9(a). The empirical semivariogram in Figure 9(b) shows that there is not much non-negative covariance values, unlike the previous example.

![Semivariogram](image)

**Figure 9.** (Data example II) (a) A realization of residual fields after subtracting the constant mean. (a) shows sampling locations ($\circ$) and prediction locations ($\times$). (b) Empirical semivariogram values for selected locations displayed against the great circle distance. For (b), dotted line represents sample variance.

| Model | MC | MG | C | WG | WC |
|-------|----|----|---|----|----|
| $\sigma^2$ | 2.062 (0.278) | 0.311 (0.005) | 3.621 (0.315) | 3.545 (0.302) | 3.249 (0.208) |
| $\lambda$ or $\lambda_0$ | 1220.475 (158.687) | 10527.394 (87.840) | 0.934 (0.084) | — | — |
| $c$ or $\delta$ | 3.412 (0.229) | 0.500 (0.000) | 2.939 (0.168) | 3.026 (0.061) | 3.018 (0.957) |
| $c$ | — | — | 8.539 (0.648) | 8.688 (0.377) | 8.783 (2.826) |
| Max. loglik | 1959.170 (40.787) | 921.716 (12.424) | 1923.248 (31.163) | 1923.297 (31.593) | 1924.991 (32.035) |
| RMSE | 1.167 (0.079) | 1.610 (0.063) | 1.107 (0.118) | 1.042 (0.090) | 1.068 (0.097) |
| MAE | 0.929 (0.075) | 1.340 (0.063) | 0.855 (0.105) | 0.804 (0.086) | 0.832 (0.090) |
| CRPS | 0.606 (0.054) | 1.167 (0.059) | 0.558 (0.064) | 0.527 (0.053) | 0.541 (0.057) |

For models C and WG, $c \in (0, \pi)$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.
Figure 10. (Data example II - the constant mean) Boxplots of differences of (a) absolute error (AE) and (b) continuous ranked probability score (CRPS) values from WG and WC, displayed against minimum great circle distance between a prediction location and its nearest sampling location. Red circles represent average values in each bin.

With a constant mean structure, Table IV shows similar results as in Section 4.2. All models except MG are similar in terms of maximum log-likelihood values. Both C and WG have better performance in terms of prediction than MC and WC, respectively. Although results in Table IV show that models, C and WG, do not outperform MC and WC as significantly as in Section 4.2, there still exist some improvements in terms of prediction with the models defined with the great circle distance. Figure 10(a) and (b) shows that WG has smaller AE and CRPS values than WC, when the prediction locations are relatively far away from their nearest sampling locations.

When we consider the mean structure in (8), there is no significant difference in terms of prediction errors between C⁴-Wendland models using the great circle distance and the Euclidean distance. However, both convex sum and C⁴-Wendland models using the great circle distance perform better than Matérn covariance model using the Euclidean distance in prediction. It is expected that the vertical directional design has smaller prediction errors than the horizontal directional design in Section 4.2 (Tables II and IV) because of large-scale variation depending strongly on latitude.

5 Conclusion

We have considered several classes of isotropic covariance functions with either the great circle distance or the Euclidean distance and compared them in terms of parameter estimation and spatial prediction. We have shown that when the true spatial range is large, the prediction performance of covariance models defined on the sphere using the great circle distance (that is, $\psi(\theta)$ on $S^2$) is better than that of the functions projected from the Euclidean space. Moreover, when the data show significantly negative correlations at large distance lags, isotropic covariance models in the class $\hat{\Phi}_3$ are not adequate, and there is a substantial difference between covariance models from the classes $\Psi_2$ and $\Phi_3$ in prediction. In the analysis of geopotential height data, we showed that distortion of the Euclidean distance might lead to poor prediction when the prediction locations are far away from sampling locations.
Acknowledgements

Mikyoung Jun’s research was supported by NSF grant DMS-1208421. This publication is based in part on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST). The NCEP/NCAR reanalysis data used in this work was provided by the NOAA/ESRL PSD, Boulder, Colorado, USA. The authors would like to thank Peter Guttorp for the discussion about this problem and Ramalingam Saravanan for suggesting to look at geopotential height data. We also thank David Bolin and Håvard Rue for their help to generate an oscillating Matérn models from the R-INLA package.

References

Banerjee, S (2005), ‘On geodetic distance computations in spatial modeling’, Biometrics, 61(2), 617–625.
Chen, D, Menegatto, VA & Sun, X (2003), ‘A necessary and sufficient condition for strictly positive definite functions on spheres’, Proceedings of the American Mathematical Society, 131(9), 2733–2740.
Du, J & Ma, C (2012), ‘Variogram matrix functions for vector random fields with second-order increments’, Mathematical Geosciences, 44(4), 411–425.
Du, J, Ma, C & Li, Y (2013), ‘Isotropic variogram matrix functions on spheres’, Mathematical Geosciences, 45(3), 341–357.
Gneiting, T (1999), ‘Correlation functions for atmospheric data analysis’, Quarterly Journal of the Royal Meteorological Society, 125(559), 2449–2464.
Gneiting, T (2013), ‘Strictly and non-strictly positive definite functions on spheres’, Bernoulli, 19(4), 1327–1349.
Gneiting, T & Raftery, AE (2007), ‘Strictly proper scoring rules, prediction, and estimation’, Journal of the American Statistical Association, 102(477), 359–378.
Guinness, J & Fuentes, M (2013), ‘Covariance functions for mean square differentiable processes on spheres’, preprint.
Hafez, Y (2012), ‘Blocking Systems Persist over North Hemisphere and Its Role in Extreme Hot Waves over Russia During Summer 2010’, Atmospheric Model Applications in Yucel, I (ed), InTech, 137–156. Available from: http://www.intechopen.com/books/atmospheric-model-applications/blocking-systems-persist-over-north-hemisphere-and-its-role-in-extreme-hot-waves-over-russia-during-s. DOI: 10.5772/33810.
Hafez, YY & Almazroui, M (2014), ‘Recent study of anomaly of global annual geopotential height and global warming’, Atmospheric and Climate Sciences, 4, 347–357.
Huang, C, Zhang, H & Robeson, SM (2011), ‘On the validity of commonly used covariance and variogram functions on the sphere’, Mathematical Geosciences, 43(6), 721–733.
Jeong, J & Jun, M (2015), ‘A class of Matérn-like covariance functions for smooth processes on a sphere’, Spatial Statistics, 11, 1–18.
Kalnay, E, Kanamitsu, M, Kistler, R, Collins, W, Deaven, D, Gandin, L, Iredell, M, Saha, S, White, G & Woollen, J et al (1996), ‘The NCEP/NCAR 40-year reanalysis project’, Bulletin of the American Meteorological Society, 77(3), 437–471.
Marshall, GJ (2002), ‘Trends in antarctic geopotential height and temperature: a comparison between radiosonde and NCEP-NCAR reanalysis data’, Journal of Climate, 15(6), 659–674.

Martins, TG, Simpson, D, Lindgren, F & Rue, H (2013), ‘Bayesian computing with INLA: new features’, Computational Statistics & Data Analysis, 67, 68–83.

Miller, KS & Samko, SG (2001), ‘Completely monotonic functions’, Integral Transforms and Special Functions, 12(4), 389–402.

Schoenberg, IJ (1942), ‘Positive definite functions on spheres.’, Duke Mathematical Journal, 9, 96–108.

Stein, ML (1999), Interpolation of Spatial Data: Some Theory for Kriging, Springer Verlag, New York.

Wendland, H (1995), ‘Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree’, Advances in Computational Mathematics, 4(1), 389–396.

Yadrenko, MI (1983), Spectral Theory of Random Fields, Optimization Software, Inc., New York.

Yaglom, AM (1987), Correlation Theory of Stationary and Related Random Functions: Vol.: 1: Basic Results, Springer-Verlag, New York.

Zhu, J, Wang, S & Zhang, X (2002), ‘Global warming mode of atmospheric circulation’, Atmospheric Science Letters, 3(1), 1–13.