Covariance functions for mean square differentiable processes on spheres

Joseph Guinness and Montserrat Fuentes

Abstract

Many applications in spatial statistics involve data observed over large regions on the Earth's surface. There is a large statistical literature devoted to covariance functions capable of modeling the degree of smoothness in data on Euclidean spaces. We adapt some of this work to covariance functions for processes on spheres, where the natural distance is great circle distance. In doing so, we define the notion of mean square differentiable processes on spheres and give necessary and sufficient conditions for an isotropic covariance function on a sphere to correspond to an m times mean square differentiable process. These conditions imply that if a process on a Euclidean space is restricted to a sphere of lower dimension, the process will retain its mean square differentiability properties. The restriction requires the covariance function to take Euclidean distance as its argument. To address the issue of whether covariance functions using Euclidean distance result in poorly fitting models, we introduce an analog to the Matérn covariance function that is valid on spheres with great circle distance metric, as the usual Matérn that is only generally valid with Euclidean distance. These covariance functions are compared with several others in applications involving satellite and climate model data.

1 Introduction

When modeling dependent spatial data, the choice of an appropriate covariance function is crucial for producing accurate predictions and estimating prediction uncertainties. Statistical theory (Stein, 1999; Zhang, 2004) shows that when the goal is interpolation of highly dependent data packed tightly on a compact domain, it is of utmost importance to correctly specify the local properties of the process, which are determined by the behavior of the covariance function near the origin. In recent years the Matérn family of covariance functions has gained widespread popularity in spatial statistics (Guttorp and Gneiting, 2006) due partly to its ability to control the local behavior of the process. Specifically, let $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$, be a random field. The isotropic Matérn covariance function is given by

$$M(\|\boldsymbol{h}\|) = \mathrm{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})) = \frac{\sigma^2 \pi^{1/2}}{2^{\nu-1}\Gamma(\nu+1/2)\alpha^{2\nu}} \mathcal{K}_\nu(\alpha\|\boldsymbol{h}\|)(\alpha\|\boldsymbol{h}\|)^\nu, \qquad (1)$$

where $\sigma^2, \alpha, \nu > 0$, and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind. We say that M is isotropic because it depends on the locations $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$ only through the Euclidean distance $\|\boldsymbol{h}\|$ between them. The parameter of interest here is ν, which controls the smoothness of the process, defined in terms of its mean square differentiability.

In environmental statistics, we often encounter data observed at locations on the surface of the Earth, for example observations from satellites or output from climate models, so it is important to adapt the existing theory and practical recommendations for fitting spatial models on Euclidean spaces to spheres, where the natural definition of distance between two points is the great circle distance (also known as geodesic distance) rather than the Euclidean distance (also known as chordal distance). We adapt some of the basic results for mean square differentiable processes on Euclidean spaces to spheres. Specifically, in Section 3 we define the notion of an m times mean square differentiable process on a sphere, and we give necessary and sufficient conditions for mean square differentiability when the covariance function is expressed either as a Fourier series or as an explicit function of great circle distance.

After reviewing some existing theory for isotropic covariance functions on spheres, we define in Section 2 an analog to the Matérn covariance function, which we call the circular Matérn, that is valid on spheres with great circle distance as the argument. Indeed, while the Matérn is a valid positive definite function on Euclidean spaces of any dimension, in some cases it is not positive definite on spheres equipped with the great circle distance metric (Gneiting, 2011). Furthermore, its validity on spheres is tied to the value of the smoothness parameter, so the feature for which the Matérn family has become popular is handcuffed when applied to spheres. The circular Matérn, however, is always valid on spheres of three or fewer dimensions and retains many important connections to the Matérn family. For example, it has three parameters for controlling the variance, range, and smoothness, and it has a spectral representation of the same form as the Matérn. In Section 4 we show that the circular Matérn has a simple closed form when ν is a half-integer, and we develop methods for efficiently computing it when ν takes on arbitrary values.

Of course, spheres are subsets of Euclidean spaces, so a covariance function that is valid on a Euclidean space can be applied to a sphere of lower dimension if the Euclidean distance is used. More formally, for $d \geq 1$, define the d-sphere as $\mathbb{S}^d = \{\boldsymbol{x} \in \mathbb{R}^{d+1} : \|\boldsymbol{x}\| = 1\}$ and the great circle distance function as $\theta(\boldsymbol{x}, \boldsymbol{y}) = \arccos(\langle \boldsymbol{x}, \boldsymbol{y} \rangle)$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^d$, where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^d$. The Euclidean distance between two points on a sphere can be expressed in terms of great circle distance as $\|\boldsymbol{x} - \boldsymbol{y}\| = 2\sin(\theta(\boldsymbol{x}, \boldsymbol{y})/2)$, so if K is a valid isotropic covariance function on $\mathbb{R}^{d+1}$, then $\psi(\theta) = K(2\sin(\theta/2))$ is a valid covariance function on $\mathbb{S}^d$ with great circle distance metric (Yadrenko, 1983; Yaglom, 1987). Stated more simply, this approach starts with a valid process on $\mathbb{R}^{d+1}$ and restricts it to the sphere $\mathbb{S}^d$, so while the process on $\mathbb{S}^d$ is trivially valid, we must use the chordal distance in calculations of the covariance. In what follows, we refer to the Matérn with chordal distance, $\varphi(\theta) = M(2\sin(\theta/2))$, as the chordal Matérn covariance function.

When the great circle distance is small compared to π, the chordal distance and the great circle distance are approximately equal, so it is reasonable to expect that the choice of distance function–either chordal or great circle distance–will not significantly distort the local properties of the process or the covariance function. However, Gneiting (2011) noted that since the two distance metrics diverge as they increase, a covariance function defined in terms of chordal distance may lead to physically unrealistic models. It is important to understand the implications of defining a covariance function in terms of chordal distance and whether such covariance functions will poorly model data observed over large regions on a sphere. There have been some efforts to compare the two distance metrics, most notably Banerjee (2005), which fits parametric spatial covariance functions to data observed at locations on the Earth. The results there suggested that using the chordal versus great circle distance may produce slightly different model estimates. However, the observation region for those data was quite small compared to the entire globe–less than $5°\times5°$ latitude $\times$ longitude–and the study considered the Matérn with great circle distance, which is not generally a positive definite function on a sphere, so a more thorough investigation is warranted.

The discussion thus far raises an important question for modeling geostatistical data on spheres: in practice, what are the relative modeling benefits of a covariance function defined in terms of the more natural great circle distance versus a covariance function with the ability to flexibly control the smoothness of the process? Until now, given a set of data, one could have compared the fit of the chordal Matérn to the fit of any one of the existing covariance functions defined in terms of great circle distance. However, it was not possible to say whether a covariance function possessing both qualities–great circle distance and flexibility of smoothness–would fit better still. The development of the circular Matérn allows us to make that comparison. In Section 5, we fit several covariance functions, including the chordal and circular Matérn, to two distinctly different datasets: one contains irregularly spaced satellite observations of total column ozone, and the other contains gridded temperature values from the output of a climate model. For these data, we do not find any large discrepancies between the chordal and circular Matérn; however, they both generally outperform the alternative covariance functions that we considered.

2 The circular Matérn covariance function

Let us first consider the class of real, continuous, and even functions on $[-\pi, \pi]$, denoted here by $C[-\pi, \pi]$. Each member of this class, $\psi \in C[-\pi, \pi]$, may be associated in a one-to-one correspondence with an infinite sequence of Fourier coefficients $\{f_k\}_{k\in\mathbb{Z}}$ with the usual relationship

$$\psi(\theta) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} f_k \exp(ik\theta). \tag{2}$$

where ψ real and even is equivalent to f_k real and even. We outline the conditions on f_k under which ψ is a valid covariance function on a sphere. Following Gneiting (2011), a function $h : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$ is positive definite on the d-sphere if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k h(\boldsymbol{x}_j, \boldsymbol{x}_k) \geq 0 \tag{3}$$

for all n, locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{S}^d$, and constants $c_1, \ldots, c_n \in \mathbb{R}$. A function is strictly positive definite if the inequality in (3) is strict whenever at least one of the constants is non-zero. Positive definiteness ensures that the variance of all linear combinations of observations is positive. The function h is isotropic if there exists a function $\psi : [0, \pi] \to \mathbb{R}$ such that

$$h(\boldsymbol{x}, \boldsymbol{y}) = \psi(\theta(\boldsymbol{x}, \boldsymbol{y})) \quad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^d.$$

Clearly, a continuous isotropic covariance function ψ may be considered a member of $C[-\pi, \pi]$ if we define $\psi(-\theta) = \psi(\theta)$, and thus every isotropic covariance function may be associated with a set of real and symmetric Fourier coefficients and expressed as in (2)

Schoenberg (1942) showed that f_k positive and summable implies that the associated covariance function is positive definite on $\mathbb{S}^1$. More recently, Gneiting (2011) demonstrated that if, in addition to f_k being positive and summable, $f_k - f_{k+2} \geq 0$ for all $k \geq 0$, the associated covariance function is positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$ as well, with $f_k - f_{k+2} > 0$ for all $k \geq 0$ implying strict positive definiteness. Stronger monotonicity conditions on f_k imply positive definiteness on higher-order spheres, and we refer the reader to Gneiting (2011) for details.

These results make it easy to generate covariance functions that are positive definite on spheres. We propose a new three-parameter covariance function that we call the *circular Matérn* covariance function:

$$\psi(\theta) = \frac{\sigma^2}{2\pi}\sum_{k\in\mathbb{Z}} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^{\nu+1/2}}, \tag{4}$$

where $\sigma^2, \alpha, \nu > 0$ are the three parameters. Clearly, the circular Matérn is positive definite on $\mathbb{S}^1$ since $f_k = (\alpha^2 + k^2)^{-\nu-1/2}$ is positive and summable, and further, it is strictly positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$ since f_k is strictly monotone decreasing for $k \geq 0$. The circular Matérn uses great circle distance explicitly as its argument, whereas the chordal Matérn must use chordal distance as its argument to ensure that it is valid on a sphere. The connection to the Matérn covariance function should be clear to those familiar with its spectral density, which is proportional to $(\alpha^2 + \omega^2)^{-\nu-1/2}$. In Section 3, we show that the parameter ν in the circular Matérn controls the mean square differentiability of the associated process on a sphere, which is analogous to the parameter ν that appears in the Matérn. The circular Matérn has a closed form expression in terms of polynomials and hyperbolic functions when ν is a half-integer, which is analogous to the Matérn as well, since it has a closed form expression in terms of polynomials and exponential functions when ν is a half-integer. The details are given in Section 4, as well as a discussion of efficient methods of computing the circular Matérn when ν is an arbitrary positive value.

It is important to note that the assumption of isotropy is often not justifiable for data on the surface of the Earth. However, many existing methods for generating anisotropic covariance functions arise from making modifications to isotropic covariance functions. This includes deformation approaches (Sampson and Guttorp, 1992; Anderes and Stein, 2008), partial differential equation approaches (Jun and Stein, 2007), and convolution approaches (Higdon, 1998; Paciorek and Schervish, 2006; Fuentes, 2002). Thus, even if one is ultimately interested in anisotropic models, careful study of the properties of isotropic models remains vital.

3 Mean square differentiability

As discussed in the introduction, correctly specifying the local properties of a process is important when one is interested in interpolating spatial data and providing accurate estimates of prediction uncertainty. The local properties, or the smoothness, of a process can be defined formally with respect to the number of mean square derivatives it possesses. For isotropic processes on $\mathbb{S}^2$, Hitczenko and Stein (2012) gave conditions for mean square differentiability when the covariance function is expressed in terms of its spherical harmonic representation. Here, we give conditions that may be used to determine the number of mean square derivatives directly from the covariance function or from its Fourier series representation, and the results apply to spheres of arbitrary dimension.

Let $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{S}^d$ be a stochastic process on the d-sphere with isotropic covariance function ψ, and let be $\mathcal{H}_\psi$ be the Hilbert space of linear combinations of Z with finite variance. Thus $\mathcal{H}_\psi$ is the set of all random variables with finite variance that can be expressed as $\sum_{k=1}^n a_k Z(\boldsymbol{x}_k)$ with $a_k \in \mathbb{R}$ and n possibly infinite. Derivatives of random or deterministic functions on Euclidean spaces must be defined with respect to a direction along a straight line. On spheres, the analog of a straight line is a geodesic or a great circle, so to study the derivatives of random functions on spheres, we must define the notion of a great circle, which is the intersection of $\mathbb{S}^d$ with any plane that passes through the origin. A sphere contains infinitely many great circles. Examples can be described using the analogy of the Earth as a sphere; lines of longitude and the equator lie along great circles, whereas non-equatorial lines of latitude do not.

Suppose $\mathbb{X}$ is the collection of all the points along one great circle. The great circle $\mathbb{X}$ is isometrically isomorphic to $\mathbb{S}^1$ since rotations in Euclidean spaces are isometric isomorphisms, and $\mathbb{S}^1$ is clearly isometrically isomorphic to $[0, 2\pi)$ if we define distance in $[0, 2\pi)$ to be $d(\phi_1, \phi_2) = \min(|\phi_1 - \phi_2|, 2\pi - |\phi_1 - \phi_2|)$. Thus there there is a distance-preserving mapping $\phi : \mathbb{X} \to [0, 2\pi)$ that associates each point on a great circle with an angle. In Figure 1, we show an example with $\phi(\boldsymbol{x}_0) = 0$ and a "clockwise" orientation.

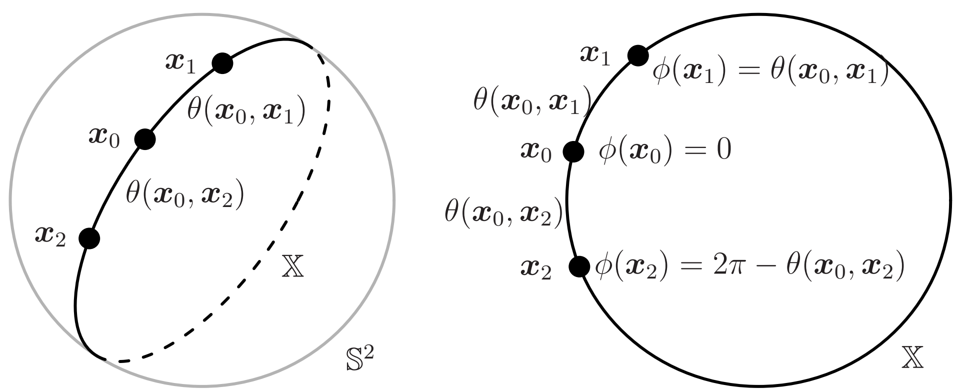

Figure 1: Illustration of a great circle $\mathbb{X}$ on the two-sphere $\mathbb{S}^2$ and the definition of ϕ.

Next, for some choice of ϕ, we define $Z_{\mathbb{X}}(\phi(\boldsymbol{x})) = Z(\boldsymbol{x})$ to be the restriction of Z to $\mathbb{X}$. Then we say that $Z_{\mathbb{X}}$ is mean square differentiable at $\boldsymbol{x}$ if the limit

$$Z_{\mathbb{X}}^{(1)}(\phi(\boldsymbol{x})) = \lim_{\varepsilon \to 0} \frac{Z_{\mathbb{X}}(\phi(\boldsymbol{x}) + \varepsilon) - Z_{\mathbb{X}}(\phi(\boldsymbol{x}))}{\varepsilon}$$

exists in $\mathcal{H}_\psi$, and we say that Z is mean square differentiable at $\boldsymbol{x}$ if $Z_{\mathbb{X}}^{(1)}(\phi(\boldsymbol{x}))$ exists for every $\mathbb{X}$ that contains $\boldsymbol{x}$. The entire process Z is mean square differentiable if Z is mean square differentiable at $\boldsymbol{x}$ for every $\boldsymbol{x} \subset \mathbb{S}^d$. Clearly, if the process is isotropic, mean square differentiability at one point along one great circle implies mean square differentiability of the entire process. To define higher order differentiability, we say that Z is $m+1$ times mean square differentiable at $\boldsymbol{x}$ if Z is m times mean square differentiable, and

$$Z_{\mathbb{X}}^{(m+1)}(\phi(\boldsymbol{x})) = \lim_{\varepsilon \to 0} \frac{Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}) + \varepsilon) - Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}))}{\varepsilon}$$

exists for every $\mathbb{X}$ that contains $\boldsymbol{x}$. Then Z is $m+1$ times mean square differentiable if Z is $m+1$ times mean square differentiable at $\boldsymbol{x}$ for every $\boldsymbol{x} \subset \mathbb{S}^d$. The following theorem adapts results for mean square differentiable processes on Euclidean spaces to processes on spheres.

Theorem 1. *If an isotropic process Z on $\mathbb{S}^d$ has covariance function*

$$\psi(\theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} f_k \exp(ik\theta),$$

then the following statements are equivalent:

1. *Z is m times mean square differentiable.*
2. *$\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$.*
3. *$\psi^{2m}(0)$ exists and is finite.*

Proof. 1 $\Leftrightarrow$ 2: Because Z is isotropic, to prove that 1 is equivalent to 2 it suffices to show that $Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}))$, as defined above, exists for some $\boldsymbol{x}$ in some great circle $\mathbb{X}$ if and only if $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$. Define Y to be a process on the real line such that $Y(\phi(\boldsymbol{x}) + 2\pi k) = Z_{\mathbb{X}}(\phi(\boldsymbol{x}))$ for $k \in \mathbb{Z}$. Stein (1999) defines a process Y on the real line to be mean square differentiable at ϕ if $\lim_{\varepsilon \to 0}(Y(\phi + \varepsilon) - Y(\phi))/\varepsilon$ exists in the Hilbert space of linear combinations of Y, with higher order derivatives defined similarly as above. Therefore, $Z_{\mathbb{X}}$ is m times mean square differentiable at $\boldsymbol{x}$ if and only if Y is at $\phi(\boldsymbol{x})$. Embedding a restriction of Z in a process on the real line allows us to use classical results on mean square differentiable processes on the real line. Due to Bochner's theorem, isotropic covariance functions on the real line can be expressed as

$$\mathrm{Cov}(Y(\phi_1), Y(\phi_2)) = \int_{\mathbb{R}} e^{i\omega(\phi_1 - \phi_2)} dF(\omega),$$

where $F(\omega)$ is called the spectral measure. Stein (1999) shows that if a process on the real line has spectral measure F, then it is m times mean square differentiable if and only if $\int_{\mathbb{R}} \omega^{2m} dF(\omega) < \infty$. The process Y that we constructed has covariance function

$$\mathrm{Cov}(Y(\phi_1), Y(\phi_2)) = \sum_{k=-\infty}^{\infty} e^{ik(\phi_1 - \phi_2)} f_k,$$

Thus, its spectral measure is $F(\omega) = \sum_{k \le \omega} f_k$, and

$$\int_{\mathbb{R}} \omega^{2m} dF(\omega) = \sum_{k=-\infty}^{\infty} k^{2m} f_k.$$

Hence Y, and therefore Z, is m times mean square differentiable if and only if $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$.

$2 \Leftrightarrow 3$: The finiteness of the $2m$'th moment of a positive finite measure is equivalent to the existence and finiteness of the $2m$'th derivative of its characteristic function (Chung, 2001, Theorem 6.4.1), so $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$ is equivalent to $\psi^{2m}(0)$ existing and finite.

□

Theorem 1 states that the differentiability properties of a process on $\mathbb{S}^d$ can be determined by the Fourier coefficients of the covariance function, which is expected since the derivatives are defined with respect to the process along great circles. The following corollary follows directly from Theorem 1 by setting $f_k = \sigma^2(\alpha^2 + k^2)^{-\nu-1/2}$.

Corollary 1. *If a process Z on $\mathbb{S}^3$ has circular Matérn covariance function, then Z is m times mean square differentiable if and only if $\nu > m$.*

Since it is a common practice to restrict a process on a Euclidean space to a sphere of lower dimension, it is of interest to know how this operation affects the mean square differentiability properties of the process on a sphere. The following corollary follows easily from Theorem 1 and gives credence to our heuristic assertion that restricting a process to a sphere does not distort the local properties of the process.

Corollary 2. *If $K(h)$ is the covariance function for an isotropic, m times mean square differentiable process on $\mathbb{R}^{d+1}$, then $\psi(\theta) = K(2\sin(\theta/2))$ is the covariance function for an isotropic, m times mean square differentiable process on $\mathbb{S}^d$.*

Proof. Since K is the covariance function for an m times mean square differentiable process on $\mathbb{R}^{d+1}$, the derivatives $K^{(j)}(0)$ exist and are finite for all $j \leq 2m$. The function $\psi(\theta) = K(2\sin(\theta/2))$ is always the covariance function for an isotropic process on $\mathbb{S}^d$. According to Theorem 1, to prove that the resulting process is m times mean square differentiable on $\mathbb{S}^d$, we must show that $\psi^{(2m)}(0)$ exists and is finite. Writing $f(\theta) = 2\sin(\theta/2)$, by Faà di Bruno's formula for derivatives of composite functions (Johnson, 2002),

$$\psi^{(2m)}(\theta) = \sum \frac{(2m)!}{b_1! b_2! \cdots b_{2m}!} K^{(j)}(f(\theta)) \left(\frac{f'(\theta)}{1!}\right)^{b_1} \left(\frac{f''(\theta)}{2!}\right)^{b_2} \cdots \left(\frac{f^{(2m)}(\theta)}{(2m)!}\right)^{b_{2m}},$$

where the sum is over all nonnegative integer solutions $b_1, \ldots, b_{2m}$ of $b_1 + 2b_2 + \cdots + 2mb_{2m} = 2m$, with $j = b_1 + \cdots + b_{2m}$. The largest value of j that appears in the sum is when $b_1 = 2m$ and $b_i = 0$ for all $i \neq 1$, which gives $j = 2m$. Since $K^{(j)}(0)$ exists and is finite for all $j \leq 2m$, $\psi^{(2m)}(0)$ must exist and be finite as well, since powers of the derivatives of $f(\theta) = 2\sin(\theta/2)$ are analytic, and the sum has finitely many terms. □

4 Computing the circular Matérn

The development of the circular Matérn covariance function allows us to address the issue of whether covariance functions defined directly in terms of great circle distance produce more realistic models than do covariance functions defined in terms of chordal distance, when both covariance functions possess the ability to flexibly control the smoothness of the process. In order to use the circular Matérn in practice, we must describe methods for computing the function, since its definition is given in terms of an infinite sum.

4.1 Closed-form for half-integer smoothness

When the smoothness parameter ν is a half-integer, the circular Matérn is given by the formula

$$\psi_n(\theta) = \frac{a}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^n},$$

where for simplicity, we set $\sigma^2 = a = 2\alpha \sinh(\alpha\pi)$. The function $\psi_n(\theta)$, $\theta \in [0, 2\pi]$, has the closed form

$$\psi_n(\theta) = \sum_{k=0}^{n-1} a_{nk} (\alpha(\theta - \pi))^k \mathrm{hyp}^{(k)}(\alpha(\theta - \pi)),$$

where $\mathrm{hyp}^{(k)}(t)$ is simply $\cosh(t)$ when k is even and $\sinh(t)$ when k is odd, so for example

$$\begin{aligned}
\psi_1(\theta) =& a_{10} \cosh(\alpha(\theta - \pi)), \\
\psi_2(\theta) =& a_{20} \cosh(\alpha(\theta - \pi)) + a_{21}(\alpha(\theta - \pi)) \sinh(\alpha(\theta - \pi)), \\
\psi_3(\theta) =& a_{30} \cosh(\alpha(\theta - \pi)) + a_{31}(\alpha(\theta - \pi)) \sinh(\alpha(\theta - \pi)) + \\
& a_{32}(\alpha(\theta - \pi))^2 \cosh(\alpha(\theta - \pi)), \\
\psi_4(\theta) =& \dots
\end{aligned}$$

We now give explicit expressions for the constant coefficients a_{nk}, which depend on α. First,

$$a_{n,n-1} = \left[(-2\alpha^2)^{n-1}(n-1)!\right]^{-1}.$$

For $r = 0, \dots, n-2$ and $k = 0, \dots, n-1$, we define

$$h_{rk} = \sum_{j=0}^{2r+1} \binom{2r+1}{j} (k)_j \, (\alpha\pi)^{k-j} \mathrm{hyp}^{(k-j+1)}(\alpha\pi),$$

where $(k)_j$ is the falling factorial and equals 1 if $j = 0$ and equals $(k)(k-1)\cdots(k-j+1)$ if $j > 0$. We define the matrix $\boldsymbol{H}_{n-1}$ to be the $(n-1)\times(n-1)$ matrix with $(r+1, k+1)$th entry h_{rk}, and the $(n-1)\times 1$ vector $\boldsymbol{h}_{n-1}$ to have $(r+1)$th entry $h_{r,n-1}$. Then the vector of coefficients $\boldsymbol{a}_n = (a_{n0}, \ldots, a_{n,n-1})'$ is given by the formula

$$\boldsymbol{a}_n = \begin{bmatrix} (a_{n0}, \ldots, a_{n,n-2})' \\ a_{n,n-1} \end{bmatrix} = \begin{bmatrix} -a_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{h}_{n-1} \\ a_{n,n-1} \end{bmatrix}.$$

A proof of this result is given in Appendix A.

4.2 Approximation for arbitrary smoothness

In most cases, the smoothness of the process is not known *a priori*, so it is desirable to have methods for estimating the smoothness from the data. This generally requires computing exactly or approximating the covariance function with arbitrary values of ν. As far as we know, the circular Matérn does not have a closed form expression in terms of elementary or special functions of θ when ν is not a half-integer. As a result, we are forced to resort to an approximation, but in this case we show that there is a computationally efficient approximation with good theoretical properties.

According to the Poisson summation formula (Zwillinger, 2003), the circular Matérn can always be written as

$$\psi(\theta) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{e^{ik\theta}}{(\alpha^2+k^2)^{\nu+1/2}} = \sum_{n=-\infty}^{\infty} M(\theta + 2\pi n), \tag{5}$$

where M is the continuous Fourier transform of $f(\omega) = (\alpha^2+\omega^2)^{-\nu-1/2}$, so that $M(\theta)$ is proportional to $\mathcal{K}_\nu(\alpha\theta)(\alpha\theta)^\nu$ with proportionality constant depending on α and ν. As an approximation to ψ, we may truncate the right hand sum in (5), obtaining

$$C_N(\theta) = \sum_{n=-N}^{N} M(\theta + 2\pi n).$$

This approximation should be numerically sufficient in most cases when α is not too small, and ν is not too large, i.e. when neither the range nor the smoothness are too large. For example, when $\nu = 1/2$, M decreases exponentially with rate α.

However, the truncated approximation does not carry with it any theoretical guarantees of positive definiteness or "closeness" to ψ. If we add an even polynomial to $C_N(\theta)$, such theoretical results are possible. To this end, we propose the approximation

$$\widetilde{\psi}(\theta) = C_N(\theta) + p_{2d}(\theta),$$

where $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ is chosen so that $R(\theta) := \psi(\theta) - \widetilde{\psi}(\theta)$ is $2d$ times continuously differentiable on the unit circle $\mathbb{T}$. Controlling the derivatives of the difference

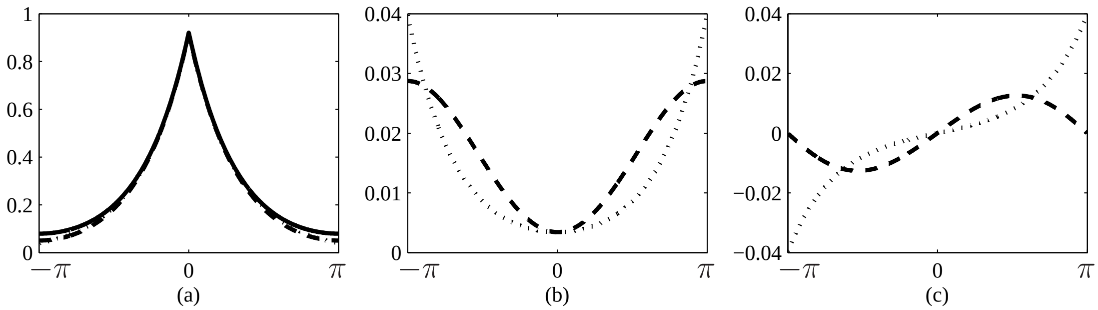

Figure 2: (a) $\psi(\theta)$ with $\nu = 1/2$, $\alpha = 1$ (solid line), along with $\widetilde{\psi}(\theta)$ (dashed line) and $C_0(\theta)$ (dotted line), (b) $R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$ (dashed line) and $R_0(\theta) = \psi(\theta) - C_0(\theta)$ (dotted line), (c) $R^{(1)}(\theta) = \psi^{(1)}(\theta) - \widetilde{\psi}^{(1)}(\theta)$ (dashed line) and $R_0^{(1)}(\theta) = \psi^{(1)}(\theta) - C_0^{(1)}(\theta)$ (dotted line).

$R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$ between a covariance function and an approximation to it is central to proving results relating to both the positive definiteness of the approximation and the extent to which $\widetilde{\psi}$ is a good approximation to ψ, where we define the approximation to be good if the resulting Gaussian measures are equivalent.

Defining $R_N(\theta) := \psi(\theta) - C_N(\theta)$, we rewrite $R(\theta) = R_N(\theta) - p_{2d}(\theta)$. It can be shown (Lemma 2 in Appendix B), that $R_N(\theta)$ is infinitely continuously differentiable on $[-\pi, \pi]$. However, the derivatives at $-\pi$ and π are not necessarily equal to each other, so R_N is not infinitely continuously differentiable on the unit circle. Specifically, R_N is an even function, so its even derivatives are continous on the unit circle, but its odd derivatives may be discontinuous at one point on the unit circle. However, if we choose p_{2d} so that its odd derivatives at $-\pi$ and π up to order $2d-1$ match those of R_N, then R will be $2d$ times continuously differentiable on the unit circle, since p_{2d} is also an even and infinitely continuously differentiable function on $[-\pi, \pi]$. In Figure 2, we show an example with $d = 1$ and $N = 0$.

In general, in order for the odd derivatives of p_{2d} at $-\pi$ and π to match those of R_N, we set $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$, with $a_{2d} = R_N^{(2d-1)}(\pi)/((2d)!\pi)$, and proceeding recursively, set

$$a_{2(d-j)} = \frac{1}{\pi(2(d-j))!}\left[R_N^{(2(d-j)-1)}(\pi) - \sum_{k=0}^{j-1} \frac{(2(d-k))!}{(2(j-k)+1)!} a_{2(d-k)} \pi^{2(j-k)+1}\right] \quad (6)$$

for $j = 1, \ldots, d-1$. The value of a_0 does not affect the differentiability. The computation of these odd derivatives appears at first to be a daunting task since R_N is an infinite sum. However, when $\theta = \pi$, this sum can be rewritten as

$$R_N(\pi) = M(-\pi(2N+1)) + \sum_{n=1}^{\infty} M(-\pi(2N+1+2n)) + M(\pi(2N+1+2n)).$$

Since M is even, when R_N is differentiated an odd number of times, each term in the sum will be zero, and we can compute any odd derivative of R_N evaluated at π by simply computing the odd derivative of M evaluated at $-\pi(2N+1)$.

The following theorem establishes positive definiteness of the approximate covariance function.

Theorem 2. *If $d > \nu$, there exists an even polynomial $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ and a finite integer N for which $\widetilde{\psi}(\theta) = C_N(\theta) + p_{2d}(\theta)$ is positive definite on $\mathbb{S}^1$, and if $d > \nu + 1/2$, there exists even polynomial and finite integer N for which $\widetilde{\psi}$ is positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$.*

Proof. To prove that $\widetilde{\psi}(\theta)$ is positive definite on $\mathbb{S}^1$, it is sufficient to show that $\widetilde{f}_k = \int_{\mathbb{T}} \widetilde{\psi}(\theta) e^{-ik\theta} d\theta > 0$ for all $k \in \mathbb{Z}$, and $\sum_{k\in\mathbb{Z}} \widetilde{f}_k < \infty$. Using $R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$, $\widetilde{f}_k$ can be expressed as

$$\widetilde{f}_k = \int_{\mathbb{T}} (\psi(\theta) - R(\theta)) e^{-ik\theta} d\theta = f_k - \varepsilon_k,$$

where f_k and ε_k are the Fourier coefficients for ψ and R, respectively. Defining a_{2k} as in (6) for $k = 1, \ldots, d$, choose N large enough so that $\sup_{\theta\in[-\pi,\pi]} R_N^{(j)}(\theta) < (1/2)(\alpha^2+1)^{-\nu-1/2}$ for both $j = 2d$ and $j = 2d-1$, and so that

$$\int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < f_0^2,$$

which are all possible due to Lemma 2 in Appendix B. Then $R(\theta) = R_N(\theta) - p_{2d}(\theta)$ is $2d$ times continuously differentiable on $\mathbb{T}$, and $R^{(2d)}$ is differentiable everywhere on $\mathbb{T}$ except at π. Furthermore, $R^{(2d)}(\theta)$ is bounded above by $(\alpha^2+1)^{-\nu-1/2}$ because $p_{2d}^{(2d)}(\theta) = R_N^{(2d-1)}(\pi)$, and N was chosen so that $R_N^{(2d-1)}(\theta)$ and $R_N^{(2d)}(\theta)$ were both bounded above by $(1/2)(\alpha^2+1)^{-\nu-1/2}$.

Using Lemma 9.5 in Körner (1988), the differentiability properties of R imply that $|\varepsilon_k| < A|k|^{-2d-1}$ for $k \neq 0$, where $A = (\alpha^2+1)^{-\nu-1/2}$. One can check that $f_k = (\alpha^2+k^2)^{-\nu-1/2} \geq |k|^{-2\nu-1}(\alpha^2+1)^{-\nu-1/2}$ for $k \neq 0$. Therefore, if $d > \nu$, then $|\varepsilon_k| < f_k$ for every $k \neq 0$, and thus $\widetilde{f}_k > 0$ for every $k \neq 0$.

The sum

$$\sum_{k\in\mathbb{Z}} |\varepsilon_k|^2 = \int_{\mathbb{T}} |R(\theta)|^2 < \int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < f_0^2,$$

which implies that $|\varepsilon_0| < f_0$, so that $|\varepsilon_k| < f_k$ for every k, and thus $\widetilde{f}_k > 0$ for every k. Finally, $|\varepsilon_k| < f_k$ for every k, and $\sum f_k < \infty$, imply that $\sum \widetilde{f}_k = \sum f_k - \varepsilon_k < \infty$. Therefore, $\widetilde{\psi}$ is positive definite on the circle.

To prove that $\widetilde{\psi}$ is positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$, it is sufficient to show that $\widetilde{f}_k - \widetilde{f}_{k+1} > 0$ for every $k \geq 0$. By the generalized binomial theorem, when $k > \max(1, \alpha^2)$,

$$\begin{aligned} f_k &= k^{-2\nu-1} + c_1 k^{-2\nu-2} + o(k^{-2\nu-2}), \\ f_{k+1} &= (k+1)^{-2\nu-1} + c_2(k+1)^{-2\nu-2} + o(k^{-2\nu-2}), \\ &= k^{-2\nu-1} + c_3 k^{-2\nu-2} + o(k^{-2\nu-2}), \end{aligned}$$

with f_k monotonically decreasing implying that $c_0 = c_1 - c_3 > 0$. So we have $f_k - f_{k+1} = c_0 k^{-2\nu-2} + d_k$, where $d_k = o(k^{-2\nu-2})$. Choose $\delta > 0$ such that $2\delta < c_0$. Then

$$f_k - f_{k+1} > (c_0 - \delta)k^{-2\nu-2} + d_k,$$

and there exists $k_0 < \infty$ such that $|d_k| < \delta k^{-2\nu-2}$ for all $k > k_0$. Therefore $f_k - f_{k+1} > (c_0 - 2\delta)k^{-2\nu-2}$ for all $k > k_0$, with $c_0 - 2\delta > 0$.

Choose N large enough that $\sup_{\theta \in [-\pi,\pi]} R_N^{(j)}(\theta) < (1/4)(c_0 - 2\delta)$ for $j = 2d$ and $j = 2d - 1$ and so that

$$\int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < \frac{1}{4}\left(f_k - f_{k+1}\right)^2$$

for every $k \leq k_0$. Then we can bound $|\varepsilon_k| < (1/2)(c_0 - 2\delta)|k|^{-2d-1}$ for $k \neq 0$ as before, and thus $|\varepsilon_k - \varepsilon_{k+1}| < (c_0 - 2\delta)k^{-2d-1}$ for all $k > k_0$. Therefore, if $d > \nu + 1/2$, $|\varepsilon_k - \varepsilon_{k+1}| < f_k - f_{k+1}$ for every $k > k_0$.

Since $\sum_{j \in \mathbb{Z}} |\varepsilon_j|^2 < (1/4)(f_k - f_{k+1})^2$ for every $k \leq k_0$, we have $|\varepsilon_j| < (1/2)(f_k - f_{k+1})$ for every $k \leq k_0$ and for every j. Therefore, $|\varepsilon_k - \varepsilon_{k+1}| < f_k - f_{k+1}$ for every $k \leq k_0$. Thus $\widetilde{f}_k - \widetilde{f}_{k+1} > 0$ for every k.

□

The following theorem asserts that it is possible to construct the approximation so that not only is it positive definite, but it well approximates the true covariance function, in that the two Gaussian measures are equivalent. We denote the Gaussian measure on $\mathbb{S}^1$ with mean 0 and covariance function ψ by $G(0, \psi)$.

Theorem 3. *If $d > \nu + 1/4$, there exists an even polynomial $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ and a finite integer N for which the Gaussian measures $G(0, \psi)$ and $G(0, \widetilde{\psi})$ on $\mathbb{S}^1$ are equivalent.*

Proof. Using results in Stein (1999), the two Gaussian measures $G(0, \psi)$ and $G(0, \widetilde{\psi})$ are equivalent if $f_k = O(\widetilde{f}_k)$, $\widetilde{f}_k = O(f_k)$, and the following sum is finite:

$$\sum_{k \in \mathbb{Z}} \frac{\left(f_k - \widetilde{f}_k\right)^2}{f_k^2} = \sum_{k \in \mathbb{Z}} \frac{\varepsilon_k^2}{f_k^2}.$$

In the proof of Theorem 2, we showed that it is possible to choose N large enough so that $|\varepsilon_k| < A|k|^{-2d-1}$ for all $k \neq 0$ and that $f_k > A|k|^{-2\nu-1}$. Therefore, if $d > \nu + 1/4$, the Fourier coefficients are of the same order, and $\varepsilon_k^2/f_k^2 < |k|^{-(1+\delta)}$ for some $\delta > 0$ and for all $k \neq 0$, so the sum converges. □

We conjecture that it is possible to choose d and N large enough so that the the Gaussian measures are equivalent on $\mathbb{S}^2$ and $\mathbb{S}^3$, but a proof of that would involve describing conditions for equivalence of Gaussian measures on higher order spheres, which is beyond the scope of this paper.

5 Application to satellite and climate model data

Both Gneiting (2011) and Huang et al. (2011) list several covariance functions that are known to be valid on $\mathbb{S}^2$. In Table 1, we give the functional form and parameters for those covariance functions and for the circular and chordal Matérn discussed herein. We investigate the performance of these covariance functions on two datasets, where performance is judged based on two criteria: (1) the value of the Gaussian loglikelihood function at its maximum when fit to each dataset, and (2) the width and coverage of 90% prediction intervals for the process at held-out spatial locations. The first criterion assesses whether the covariance model gives an accurate representation of the process, and the second evaluates the accuracy of predictions and prediction uncertainty.

The first dataset we consider contains values of total column ozone derived from observations made by the Ozone Monitoring Instrument on board NASA's Aura Satellite. Aura follows a nearly sun-synchronous orbit with a period of roughly 100 minutes. We consider observations from a single orbit that encompass a wide range of latitudes over a longitude band of roughly 23 degrees near the equator. The data are plotted against latitude in Figure 3. Since all of the data are collected within a 50-minute window, we ignore the time dimension in the data and proceed as if they were collected simultaneously. We fit the covariance models to 1000 of these ozone values and for prediction purposes hold out an additional 1000 values at distinct locations. All of the data are publicly available on the web; we downloaded them using the Simple Subset Wizard (`http://disc.sci.gsfc.nasa.gov/SSW/`) with keyword OMDOAO3 for the date of March 19, 2012.

The second dataset contains 10 meter height surface temperature outputs from a single run of the Community Climate System Model Version 4 (CCSM4). We consider a spatial field consisting of a single year's average temperature, corresponding to year 50 of this particular run of the model. The values are plotted against latitude in Figure 4. The temperature values from the climate model output tend to be smoother as a function of spatial location than are the total column ozone values. Again, we consider 1000 temperature values and hold out an additional 1000 values. The locations are regularly-spaced over only the oceans.

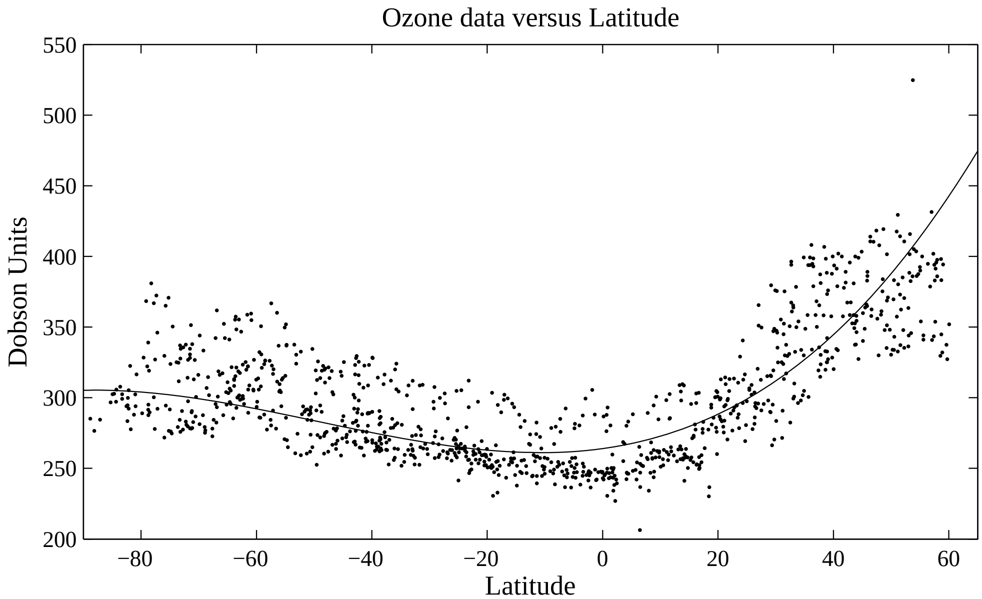

Figure 3: Total column ozone observations as a function of latitude. The smooth curve is a generalized least squares estimate of a cubic mean function, assuming the covariance function is the REML estimate of the circular Matérn.

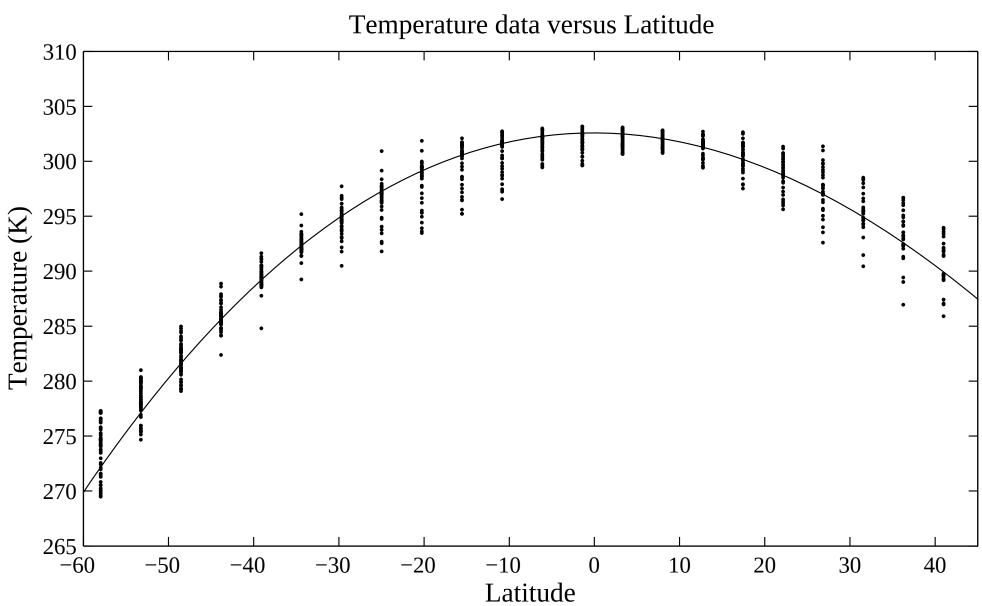

Figure 4: Climate model temperature values as a function of latitude. The smooth curve is a generalized least squares estimate of a cubic mean function, assuming the covariance function is the REML estimate of the circular Matérn.

Name	Expression	Parameter Values
Circular Matérn	$\sum_{k=-\infty}^{\infty} \frac{\exp(ik\theta)}{(\alpha^2+k^2)^{\nu+1/2}}$	$\alpha, \nu > 0$
Chordal Matérn	$\left(\alpha 2 \sin\left(\frac{\theta}{2}\right)\right)^{\nu} \mathcal{K}_{\nu}\left(\alpha 2 \sin\left(\frac{\theta}{2}\right)\right)$	$\alpha, \nu > 0$
Matérn	$(\alpha\theta)^{\nu}\mathcal{K}_{\nu}(\alpha\theta)$	$\alpha > 0; \nu \in (0, \frac{1}{2}]$
Powered Exponential	$\exp\left(-(\alpha\theta)^{\nu}\right)$	$\alpha > 0; \nu \in (0, 1]$
Generalized Cauchy	$(1 + (\alpha\theta)^{\nu})^{-\tau/\nu}$	$\alpha, \tau > 0; \nu \in (0, 1]$
Multiquadric	$(1-\tau)^{2\alpha}/(1+\tau^2-2\tau\cos(\theta))^{\alpha}$	$\alpha > 0; \tau \in (0, 1)$
Sine Power	$1 - \left(\sin\frac{\theta}{2}\right)^{\nu}$	$\nu \in (0, 2)$
Spherical	$\left(1 + \frac{\alpha\theta}{2}\right)(1-\alpha\theta)_+^2$	$\alpha > 0$
Askey	$(1-\alpha\theta)_+^{\tau}$	$\alpha > 0; \tau \geq 2$
C^2-Wendland	$(1+\tau\alpha\theta)(1-\alpha\theta)_+^{\tau}$	$\alpha \geq \frac{1}{\pi}; \tau \geq 4$
C^4 Wendland	$\left(1+\tau\alpha\theta+\frac{\tau^2-1}{3}(\alpha\theta)^2\right)(1-\alpha\theta)_+^{\tau}$	$\alpha \geq \frac{1}{\pi}; \tau \geq 6$

Table 1: List of covariance functions

For each covariance function $\varphi(\theta)$ listed in Table 1, we fit the covariance function $\gamma\mathbf{1}_{\theta=0} + \sigma^2\varphi(\theta)$, so in addition to the parameters listed in Table 1, we include the possibility of a nugget term and a multiplicative term. The models are fit using residual maximum likelihood (REML), assuming an unknown mean function that is cubic in latitude. In Table 2, we report the maximum residual loglikelihood for each dataset and for each covariance family, relative to that of the best-fitting covariance function.

Conditional on the fitted covariance functions for each covariance family and each dataset, we compute best linear unbiased predictions (BLUPs) $\widehat{Z}(\boldsymbol{x}_0)$ of the process $Z(\boldsymbol{x}_0)$ at each held-out location, $\boldsymbol{x}_0$. For each prediction, we compute the mean square prediction error $E(\widehat{Z}(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0))^2$ and form 90% prediction intervals based on a Gaussian assumption. The details of the computation of the BLUPs and their mean square errors can be found in Stein (1999, Sec. 1.5). In Table 2, we report the average width of the prediction intervals and their empirical coverage for each covariance function.

We see that for the ozone data, which is rougher, many of the covariance families perform nearly equally well in terms of residual maximum likelihood and prediction. The circular and chordal Matérn both return estimates of $\widehat{\nu} = 0.419$ rounded to three decimals. Thus it is not surprising that the covariance families that perform well all can be made linear at the origin. On the contrary, those that perform poorly on the ozone data–the Multiquadric, the C^2-Wendland, and the C^4-Wendland–are

	Ozone Data			Climate Model Data		
Cov. Family	Δloglik	width	coverage	Δloglik	width	coverage
Circular Matérn	0.00	42.24	90.90	0.00	0.90	89.30
Euclidean Matérn	−0.00	42.24	90.90	−0.05	0.90	89.30
Matérn	−0.00	42.24	90.90	−119.27	1.48	94.40
Powered Exponential	−0.02	42.24	90.90	−119.27	1.48	94.40
Generalized Cauchy	−0.05	42.24	91.00	−119.83	1.48	94.40
Multiquadric	−8.93	43.14	91.10	−4.47	0.95	90.10
Sine Power	−1.24	42.22	91.20	−17.53	0.98	89.80
Spherical	−2.14	41.82	90.80	−121.84	1.48	94.70
Askey	−1.58	41.99	90.80	−113.66	1.47	94.40
C^2-Wendland	−15.27	43.12	91.20	−0.10	0.91	89.30
C^4-Wendland	−21.04	43.37	91.20	−8.41	0.99	90.00

Table 2: Residual maximum loglikelihoods and average widths and empirical coverage probabilities of 90% prediction intervals over all held-out sites $\boldsymbol{x}_0$. The loglikelihood is reported relative to the best-fitting model for each dataset, which is the circular Matérn in both cases.

always differentiable at the origin. For those three covariance functions, the empirical coverage probabilities are too large.

Except for the circular and chordal Matérn, all of the covariance families that performed well on the ozone data perform poorly on the climate model data, with some performing extremely poorly. The circular Matérn and the chordal Matérn return $\widehat{\nu} = 1.459$ and $\widehat{\nu} = 1.457$, respectively, which suggests that the climate model data can be modeled by a process that is once but not twice mean square differentiable. Here, the covariance families that can be made differentiable at the origin perform well, while those that are always linear at the origin perform particularly poorly in terms of loglikelihood and prediction. These covariance functions overestimate the mean square prediction error, leading to overly conservative prediction intervals. It is also interesting to note that the C^2-Wendland outperforms the C^4-Wendland in terms of loglikelihood; the C^4-Wendland is too smooth since it possesses four continuous derivatives at the origin, while the C^2-Wendland has just two.

The chordal and circular Matérn fit both datasets reasonably well in terms of Gaussian loglikelihoods, whereas all of the other covariance functions fit poorly to at least one of the datasets. For these datasets, we are not able to detect any large improvement of the circular Matérn over the chordal Matérn; the circular Matérn gives slightly higher residual maximum likelihoods over the chordal Matérn, but we do not interpret this difference to be practically significant. The evidence here suggests that having a covariance function with flexible smoothness is more beneficial than having a covariance function that is expressed in terms of great circle distance explicitly.

Furthermore, a covariance with both qualities–flexible smoothness and expressed in terms of great circle distance–does not perform any better than one with only flexible smoothness. Regardless, the circular Matérn provides a reasonable alternative to the chordal Matérn.

6 Conclusions

We have adapted some of the results for mean square differentiable processes on Euclidean spaces to processes on spheres and provided simple conditions for an isotropic covariance function to correspond to an m times mean square differentiable process on a sphere. Using these results, we showed that when we restrict an m times mean square differentiable process on $\mathbb{R}^{d+1}$ to the sphere $\mathbb{S}^d$, the result is an m times mean square differentiable process on $\mathbb{S}^d$. We have identified an analog to the Matérn covariance function that is valid on spheres of three dimensions or fewer with great circle distance as its argument and proved that it has a simple closed form when its smoothness parameter ν is a half-integer. We also give computationally efficient methods for computing it when ν takes on arbitrary values. The development of the circular Matérn allows a comparison with the chordal Matérn, which makes it possible to investigate the benefits of having a covariance function expressed in terms of great circle distance. For the data presented in Section 5, we do not detect any significant benefit.

While the approximation for the circular Matérn is computationally efficient, in that it requires only a finite number of operations, it is still noticeably slower than the computation of the chordal Matérn. For very large datasets, where the computational bottleneck is the $O(n^3)$ Cholesky factorization of the covariance matrix, this may not be a problem. However, for moderately-sized datasets like those we consider here, or for estimation methods that don't require a full Cholesky factorization, the slowdown associated with computing the entries of the covariance matrix may be substantial.

References

Ethan B. Anderes and Michael L. Stein. Estimating deformations of isotropic Gaussian random fields on the plane. *Annals of Statistics*, 36(2):719–741, 2008.

Sudipto Banerjee. On geodetic distance computations in spatial modeling. *Biometrics*, 61(2):617–625, 2005.

Kai Lai Chung. *A Course in Probability Theory*. Academic Press, 3 edition, 2001.

Digital Library of Mathematical Functions, 2012. URL `http://dlmf.nist.gov/`. Release 1.0.5 of 2012-10-01.

Montserrat Fuentes. Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1):197–210, 2002.

Tilmann Gneiting. Strictly and non-strictly positive definite functions on spheres. *Preprint*, arXiv:1111.7077, 2011.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Elsevier, 7 edition, 2007.

Peter Guttorp and Tilmann Gneiting. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, 93(4):989–995, 2006.

David Higdon. A process convolution approach to modelling temperatures in the North Atlantic ocean. *Environmental and Ecological Statistics*, 5:173–190, 1998.

Marcin Hitczenko and Michael L. Stein. Some theory for anisotropic processes on the sphere. *Statistical Methodology*, 9(2):211–227, 2012.

Chunfeng Huang, Haimeng Zhang, and Scott M. Robeson. On the validity of commonly used covariance and variogram functions on the sphere. *Mathematical Geosciences*, 43:721–722, 2011.

Warren P. Johnson. The curious history of Faà di Bruno's formula. *The American Mathematical Monthly*, 109(3):217–234, 2002.

Mikyoung Jun and Michael L. Stein. An approach to producing space-time covariance functions on spheres. *Technometrics*, 49(4):468–479, 2007.

Christopher J. Paciorek and Mark J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506, 2006.

Paul D. Sampson and Peter Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87 (417):108–119, 1992.

I. J. Schoenberg. Positive definite functions on spheres. *Duke Mathematical Journal*, 9(1):96–108, 1942.

Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.

Mikhail I. Yadrenko. *Spectral Theory of Random Fields*. Optimization Software, 1983.

A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions I*. Springer-Verlag, 1987.

Hao Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.

Daniel Zwillinger. *Standard Mathematial Tables and Formulae*. Chapman and Hall, 31 edition, 2003.

Appendices to supplement "Covariance functions for mean square differentiable processes on spheres," by Guinness and Fuentes

A Proof for closed form of the circular Matérn

When the smoothness parameter ν is a half-integer, the circular Matérn is given by the formula

$$\psi_n(\theta) = \frac{a}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^n},$$

where $a = 2\alpha \sinh(\alpha\pi)$. The function $\psi_n(\theta)$, $\theta \in [0, 2\pi]$, has the closed form

$$\psi_n(\theta) = \sum_{k=0}^{n-1} a_{nk} (\alpha(\theta - \pi))^k \mathrm{hyp}^{(k)}(\alpha(\theta - \pi)),$$

where $\mathrm{hyp}^{(k)}(t)$ is simply $\cosh(t)$ when k is even and $\sinh(t)$ when k is odd. The coefficients $\boldsymbol{a}_n = (a_{n0}, \ldots, a_{n,n-1})'$ are given by

$$a_{n,n-1} = \left[(-2\alpha^2)^{n-1}(n-1)!\right]^{-1}.$$

and

$$\boldsymbol{a}_n = \begin{bmatrix} (a_{n0}, \ldots, a_{n,n-2})' \\ a_{n,n-1} \end{bmatrix} = \begin{bmatrix} -a_{n,n-1} \boldsymbol{H}_{n-1}^{-1} \boldsymbol{h}_{n-1} \\ a_{n,n-1} \end{bmatrix},$$

where $\boldsymbol{H}_{n-1}$ and $\boldsymbol{h}_{n-1}$ are defined in Section 4.1.

Proof: The function $\psi_n(\theta)$ satisfies the following inhomogeneous differential equation with constant coefficients:

$$\sum_{m=0}^{n-1} c_m \psi_n^{(2m)}(\theta) = \psi_1(\theta),$$

with

$$c_m = \binom{n-1}{m} (-1)^m \alpha^{2(n-1-m)}.$$

We know that $\psi_1(\theta) = \cosh(\alpha(\theta - \pi))$ (Gradshteyn and Ryzhik, 2007, Equation 1.445.2). Making the substitution $t = \theta - \pi$, the differential equation has general solution

$$\psi_n(t + \pi) = \sum_{k=0}^{2(n-1)} (b_{k1}(\alpha t)^k e^{\alpha t} + b_{k2}(\alpha t)^k e^{-\alpha t}),$$

which can be solved by the method of undetermined coefficients. Symmetry conditions on $\psi_n(t+\pi)$ around $t=0$ require that $b_{k1}=b_{k2}$ if k is even, and $b_{k1}=-b_{k2}$ when k is odd, so the general solution can be rewritten as

$$\psi_n(t+\pi)=\sum_{k=0}^{2(n-1)} b_k(\alpha t)^k \mathrm{hyp}^{(k)}(\alpha t).$$

The $(2m)$th derivative of ψ_n is thus given by

$$\psi_n^{(2m)}(t+\pi)=\sum_{k=0}^{2(n-1)} b_k\alpha^{2m}\sum_{j=0}^{2m}\binom{2m}{j}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t),$$

and the differential equation is

$$\sum_{m=0}^{n-1} c_m\sum_{k=0}^{2(n-1)} b_k\alpha^{2m}\sum_{j=0}^{2m}\binom{2m}{j}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)=\cosh(\alpha t). \tag{7}$$

To evaluate the left hand side of (7), we exchange the order of addition to arrive at

$$\sum_{k=0}^{2(n-1)}\sum_{j=0}^{2(n-1)}\sum_{m=\lceil j/2\rceil}^{n-1} c_m b_k\alpha^{2m}\binom{2m}{j}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)=\cosh(\alpha t), \tag{8}$$

where $\lceil\cdot\rceil$ is the ceiling function. To study (8), we proceed by fixing k and summing over j and m. We can ignore the terms for which $j>k$ because in those cases $(k)_j=0$. For $k=0,\dots,2(n-1)$, we define

$$\begin{aligned}
p_k(t) :=& \sum_{j=0}^{k}\sum_{m=\lceil j/2\rceil}^{n-1} c_m\alpha^{2m}\binom{2m}{j}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\\
=& \sum_{j=0}^{k}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\sum_{m=\lceil j/2\rceil}^{n-1}\alpha^{2m}\alpha^{2(n-1-m)}(-1)^m\binom{n-1}{m}\binom{2m}{j}\\
=& \sum_{j=0}^{k}\frac{1}{j!}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\alpha^{2(n-1)}\sum_{m=\lceil j/2\rceil}^{n-1}(-1)^m\binom{n-1}{m}(2m)\cdots(2m-j+1)\\
=& \sum_{j=0}^{k}\frac{1}{j!}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\alpha^{2(n-1)}\sum_{m=\lceil j/2\rceil}^{n-1}(-1)^m\binom{n-1}{m}P_j(m),
\end{aligned}$$

where $P_j(m)=(2m)(2m-1)\cdots(2m-j+1)$ is a jth order polynomial in m that equals zero when $m=0,\dots,\lceil j/2\rceil-1$, so we can allow the sum over m to run from 0 to $n-1$ in

$$p_k(t)=\sum_{j=0}^{k}\frac{1}{j!}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\alpha^{2(n-1)}\sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}P_j(m).$$

Since P_j is a jth order polynomial, the sum over m is zero when $j < n-1$ (Gradshteyn and Ryzhik, 2007, Equation 0.154.3), and hence $p_k(t)$ is zero when $k < n-1$. When $j = n-1$,

$$\begin{aligned}\sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}P_j(m) &= \sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}\left((2m)^{n-1}+Q_{n-2}(m)\right)\\ &= \sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}(2m)^{n-1}\\ &= 2^{n-1}(-1)^{n-1}(n-1)!\\ &\neq 0, \qquad (9)\end{aligned}$$

where the second equality is due to the fact that Q_{n-2} is a polynomial of degree $n-2$, and the third equality follows from Gradshteyn and Ryzhik (2007, Equation 0.154.4). If we rewrite $p_k(t)$ for $k \geq n-1$ as

$$p_k(t) = \sum_{j=n-1}^{k}\frac{1}{j!}(k)_j(\alpha t)^{k-j}\mathrm{hyp}^{(k-j)}(\alpha t)\alpha^{2(n-1)}\sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}P_j(m), \qquad (10)$$

we can now see that the set of functions $p_k(t)$ for $k = n-1, \ldots 2(n-1)$ are linearly independent, since $p_k(t)$ can be written as

$$p_k(t) = \sum_{l=0}^{k-(n-1)} w_{kl}(\alpha t)^l\mathrm{hyp}^{(l)}(\alpha t),$$

where w_{kl} are constants for which $w_{k,k-(n-1)} \neq 0$ by (9). The differential equation in 7 is now reduced to

$$\sum_{k=n-1}^{2(n-1)} b_k p_k(t) = \cosh(\alpha t).$$

with the set of functions $p_k(t)$ linearly independent. When $k = n-1$, we have

$$\begin{aligned}p_{n-1}(t) &= b_{n-1}\cosh(\alpha t)\alpha^{2(n-1)}\sum_{m=0}^{n-1}(-1)^m\binom{n-1}{m}\left((2m)^{n-1}+Q_{n-2}(m)\right)\\ &= b_{n-1}\cosh(\alpha t)\alpha^{2(n-1)}2^{n-1}(-1)^{n-1}(n-1)!,\end{aligned}$$

which is proportional to $\cosh(\alpha t)$. Since the differential equation is set equal to $\cosh(\alpha t)$ and the p_k's are linearly independent for $k \geq n-1$, we have determined that

$$\frac{1}{b_{n-1}} = \left(-2\alpha^2\right)^{n-1}(n-1)!.$$

We now have the equation

$$\sum_{k=n}^{2(n-1)} b_k p_k(t) = 0. \tag{11}$$

Since the p_k's are linearly independent, (11) implies that $b_k = 0$ for all $k > n-1$. However, the other coefficients $b_0, \dots, b_{n-2}$ must be determined by enforcing the initial conditions implied by ψ_n. We know that the odd derivatives of ψ_n up to order $2(n-1)-1$ must be zero when evaluated at 0 and 2π. This gives us $n-1$ initial conditions. The $(2r+1)$th derivative of ψ_n evaluated at $\theta = 2\pi$ (or $t = \pi$) is

$$\psi_n^{(2r+1)}(2\pi) = \sum_{k=0}^{n-1} b_k \sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{k-j} (k)_j \mathrm{hyp}^{(k-j+1)}(\alpha\pi).$$

Setting this equation equal to zero gives us the $n-1$ equations corresponding to $r = 0, \dots, n-2$:

$$\begin{aligned} &\sum_{k=0}^{n-2} \left[\sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{k-j} (k)_j \mathrm{hyp}^{(k-j+1)}(\alpha\pi) \right] b_k \\ &= -b_{n-1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{n-1-j} (n-1)_j \mathrm{hyp}^{(n-j)}(\alpha\pi). \end{aligned}$$

With h_{rk}, H_{n-1}, and $\boldsymbol{h}_{n-1}$ defined as above, the solution of this system of equations is

$$(b_0, \dots, b_{n-2})' = -b_{n-1} H_{n-1}^{-1} \boldsymbol{h}_{n-1}.$$

B Proofs of lemmas

Lemma 1. *For $\alpha > 0$, $\nu, \mu \in \mathbb{R}$, $j \in \mathbb{Z}$, the sequence of functions*

$$S_n(\theta) = \sum_{0<|k|\le n} \mathrm{sgn}(k)^j \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\mu$$

is uniformly convergent on $[-\pi, \pi]$.

Proof. Using the Cauchy criterion, our aim is to show that for every $\varepsilon > 0$, there exists N such that for every $n, m \ge N$, $\theta \in [-\pi, \pi]$ implies that $|S_n(\theta) - S_m(\theta)| \le \varepsilon$. First we see that if $n, m \ge N$, then

$$|S_n(\theta) - S_m(\theta)| \le \sum_{|k|>N} \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\mu$$

since $\mathcal{K}_\nu(x) > 0$ for $x > 0$ (Digital Library of Mathematical Functions, 2012, 10.37). Using the fact that $\mathcal{K}_\nu(x)$ is decreasing in x (Digital Library of Mathematical Functions, 2012, 10.37, 10.27.3), and that for $\theta \in [-\pi, \pi]$, $2\pi|k| - \pi \leq |\theta + 2\pi k| \leq 2\pi|k| + \pi$, we obtain $\mathcal{K}_\nu(\alpha|\theta + 2\pi k|) \leq \mathcal{K}_\nu(\alpha(2\pi|k| - \pi))$, and $(\alpha|\theta + 2\pi k|)^\mu \leq (\alpha(2\pi|k| + \pi)^\mu$ for $\mu \geq 0$. We assume $\mu \geq 0$ because if $\mu < 0$, the summand is eventually bounded by $\mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^0$. Therefore

$$|S_n(\theta) - S_m(\theta)| \leq 2 \sum_{k>N} \mathcal{K}_\nu(\alpha(2\pi k - \pi))(\alpha(2\pi k + \pi))^\mu,$$

so if this sum converges, we can always find N such that $|S_n(\theta) - S_m(\theta)| < \varepsilon$. The modified Bessel function of the second kind has the property that $K_\nu(z) \sim (\pi/2)^{1/2} z^{-1/2} e^{-z}$. This implies that there exists a positive constant A and integer M such that for all $k > M$,

$$\mathcal{K}_\nu(\alpha(2\pi k - \pi)) \leq A\sqrt{\frac{\pi}{2}}(\alpha(2\pi k - \pi))^{-1/2} e^{-\alpha(2\pi k - \pi)}.$$

Thus

$$\sum_{k>M} \mathcal{K}_\nu(\alpha(2\pi k - \pi))(\alpha(2\pi k + \pi))^\mu \leq$$

$$A\sqrt{\frac{\pi}{2}} e^{2\pi\alpha} \sum_{k>M} \frac{(2\pi k + \pi)^{1/2}}{(2\pi k - \pi)^{1/2}} (\alpha(2\pi k + \pi))^{\mu - 1/2} e^{-\alpha(2\pi k + \pi)},$$

which clearly converges. □

Lemma 2. *For every $N, j \in \mathbb{Z}^+$, $R_N^{(j)}(\theta)$ is continuous on $[-\pi, \pi]$, and for every j, $R_N^{(j)}(\theta)$ converges to zero uniformly on $[-\pi, \pi]$ as $N \to \infty$.*

Proof. We first need to show that $R_N(\theta)$ can be repeatedly differentiated term-by-term so that we can write down expressions for $R_N^{(j)}(\theta)$. Formally, we define

$$R_{N,n}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\nu.$$

Following Theorem 7.17 (Rudin, 1976), suppose the following conditions hold for a sequence of functions f_n:

(a) $f_n(\theta)$ is differentiable on $[-\pi, \pi]$ for each n,

(b) $f_n(\theta_0)$ converges for some $\theta_0 \in [-\pi, \pi]$,

(c) $f_n^{(1)}(\theta)$ converges uniformly on $[-\pi, \pi]$.

Then f_n converges uniformly on $[-\pi, \pi]$ to a function f, and $f^{(1)}(\theta) = \lim_{n\to\infty} f_n^{(1)}(\theta)$ for $\theta \in [-\pi, \pi]$, i.e. the limit of the derivatives of a sequence of functions is equal to the derivative of the limit.

We set $f_n(\theta) = R_{N,n}(\theta)$ and check the conditions of the theorem. If $k \neq 0$, $|\theta + 2\pi k| > 0$ for all $\theta \in [-\pi, \pi]$. Since the modified Bessel function of the second kind and polynomials are both differentiable away from 0, it follows that $R_{N,n}(\theta)$ is also differentiable for each n, so (a) holds. According to Lemma 1, $R_{N,n}$ is uniformly convergent on $[-\pi, \pi]$ as $n \to \infty$, so (b) holds as well. The derivative of the modified Bessel function can be expressed as

$$\mathcal{K}_\nu^{(1)}(x) = -\frac{1}{2}\left(\mathcal{K}_{\nu-1}(x) + \mathcal{K}_{\nu+1}(x)\right) \tag{12}$$

(Watson, 1966), so the derivative of $R_{N,n}(\theta)$ is given by

$$R_{N,n}^{(1)}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \frac{\text{sgn}(k)\alpha}{2}\left(-\mathcal{K}_{\nu-1}(\alpha|\theta+2\pi k|) - \mathcal{K}_{\nu+1}(\alpha|\theta+2\pi k|)\right)(\alpha|\theta+2\pi k|)^\nu$$
$$+\alpha\nu\mathcal{K}_\nu(\alpha|\theta+2\pi k|)(\alpha|\theta+2\pi k|)^{\nu-1},$$

which consists of three terms, each of which can be written in the form in Lemma 1. Therefore $R_{N,n}^{(1)}(\theta)$ converges uniformly, so (c) holds and thus $R_N(\theta)$ can be differentiated once term-by-term.

In general, using repeated applications of (12), the j'th derivative of $R_{N,n}$ can be expressed as

$$R_{N,n}^{(j)}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \sum_{l=0}^{j} \text{sgn}(k)^j \alpha^j \binom{j}{l} \left[(-2)^{-j+l} \sum_{m=0}^{j-l} \binom{j-l}{m} \mathcal{K}_{\nu-j+l+2m}(\alpha|\theta+2\pi k|)\right] \times$$
$$\left[\frac{\Gamma(\nu+1)}{\Gamma(\nu+1-l)}(\alpha|\theta+2\pi k|)^{\nu-l}\right]. \tag{13}$$

Since the number of terms is finite, we can exchange the order of summation so that $R_{N,n}(\theta)$ can be written as a finite number of sums of the form given in Lemma 1. Proceeding inductively, we assume that R_N can be differentiated j times term-by-term, with derivative given by the limit as $n \to \infty$ of the expression in (13). Then to complete the induction we must show that $R_N^{(j)}$ can be differentiated term-by-term, which amounts to establishing the conditions of Theorem 7.17 with $f_n = R_{N,n}^{(j)}$. Differentiability holds again due to differentiability of polynomials and the modified Bessel function of the second kind. Convergence at a point holds again because of the form of $R_{N,n}^{(j)}(\theta)$ and Lemma 1. Uniform convergence of the derivative also holds because the $(j+1)$'th derivative of $R_{N,n}$ can also be written as a finite number of sums of the form in Lemma 1. An additional consequence of Theorem 7.17 is that $R_{N,n}^{(j)}$ converges uniformly on $[-\pi, \pi]$. Therefore we have shown that $R_N(\theta)$ can be differentiated term-by-term an arbitrary number of times, and the convergence of the j'th derivative (as $n \to \infty$) is uniform on $[-\pi, \pi]$. Continuity of the derivatives follows from the fact that each derivative is differentiable.

The uniform convergence of the derivatives allows us to easily show that $R_N^{(j)}(\theta)$ converges to zero uniformly on $[-\pi, \pi]$ as $N \to \infty$. Suppose that $N > M$. Then we can write $R_N^{(j)}(\theta) = R_M^{(j)}(\theta) - R_{M,N}^{(j)}(\theta)$. We have just shown that $R_{M,N}^{(j)}$ converges uniformly to $R_M^{(j)}$ as $N \to \infty$, which means that for every $\varepsilon > 0$, we can find N_0 such that $N > N_0$ implies $|R_{M,N}^{(j)} - R_M^{(j)}| < \varepsilon$, which in turn implies that $|R_N^{(j)}(\theta)| < \varepsilon$ for every $N > N_0$. □