



Locally Anisotropic Nonstationary Covariance Functions on the Sphere

Jian CAO, Jingjie ZHANG, Zhuoer SUN, and Matthias KATZFUSS

Rapid developments in satellite remote-sensing technology have enabled the collection of geospatial data on a global scale, hence increasing the need for covariance functions that can capture spatial dependence on spherical domains. We propose a general method of constructing nonstationary, locally anisotropic covariance functions on the sphere based on covariance functions in $\mathbb{R}^3$. We also provide theorems that specify the conditions under which the resulting correlation function is isotropic or axially symmetric. For large datasets on the sphere commonly seen in modern applications, the Vecchia approximation is used to achieve higher scalability on statistical inference. The importance of flexible covariance structures is demonstrated numerically using simulated data and a precipitation dataset.

Supplementary materials accompanying this paper appear online.

Key Words: Axial symmetry; Local anisotropy; Nonstationarity; Global data; Vecchia approximation.

1. INTRODUCTION

Traditionally, geostatistical analysis relied on approximating small or regional spatial domains as flat subsets of $\mathbb{R}^2$. However, since the deployment of satellites in the collection of global data, there is an increasing demand for covariance functions that are valid on spheres. In this paper, we aim to propose a new family of spherical covariance functions, defined over the unit 2-sphere $\mathbb{S} = \{\tilde{\mathbf{s}} \in \mathbb{R}^3 : \|\tilde{\mathbf{s}}\| = 1\}$, which are able to capture nonstationary features commonly observed in geostatistical datasets.

For processes defined over $\mathbb{S}$, two different distance measures are commonly used, namely the great-arc (or great-circle) distance, which measures the distance “going along the surface of the sphere”, and the Euclidean or chordal distance, which “pierces through the sphere.”

J. Cao · M. Katzfuss (✉) Department of Statistics and Institute of Data Science, Texas A&M University, College Station, TX, USA (E-mail: katzfuss@gmail.com).

J. Zhang · Z. Sun Department of Statistics, Texas A&M University, College Station, TX, USA.

© 2023 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Volume 29, Number 2, Pages 212–231

<https://doi.org/10.1007/s13253-023-00573-y>

The relationship between great-arc distance and chordal distance on $\mathbb{S}$ is given by:

$$r = 2 \sin(\theta/2), \quad (1)$$

where r is the chordal distance between two points on $\mathbb{S}$, and θ is the corresponding central angle, proportional to the great-arc distance. Although a function of chordal distance is naturally a function of great-circle distance, finding a valid correlation function directly based on great-arc distance is not trivial (e.g., [Jones 1963](#)) due to the curvature of $\mathbb{S}$. Most well-known covariance functions are valid (i.e., positive definite) on $\mathbb{R}^d$, for $d \geq 1$, under Euclidean or Mahalanobis distance, yet may become invalid under great-arc distance (e.g., [Gneiting 2013](#)). For example, the Matérn covariance function is only valid under great-arc distance if its smoothness is no greater than 0.5 ([Gneiting 2013](#)). [Huang et al. \(2011\)](#) summarized the validity of commonly used covariance functions under great-arc distance, focusing mostly on isotropic covariance functions. [Gneiting \(2013\)](#) further developed characterizations and constructions of isotropic positive-definite functions on spheres and proved that subject to a natural support condition, many isotropic positive-definite functions on the Euclidean space $\mathbb{R}^3$ allow for the direct substitution of the chordal distance by the great-arc distance on the sphere. [Bissiri et al. \(2020\)](#) proposed conditions for strict positive-definiteness of axially symmetric covariance functions.

A comprehensive review of modeling spherical processes using the great-arc distance was provided in [Jeong et al. \(2017\)](#) and [Blake et al. \(2022\)](#). Here, we provide a non-exhaustive list of representative works. In the category of isotropic spherical covariance functions, [Ma \(2015\)](#) proposed a construction based on Gegenbauer polynomials from [Ma \(2012\)](#). Similarly, [Du et al. \(2013\)](#) designed isotropic variogram functions on spheres using an infinite sum of the products of positive-definite matrices and ultraspherical polynomials. [Alegría et al. \(2021\)](#) proposed the $\mathcal{F}$ family of isotropic covariance functions that parameterizes the differentiability of the Gaussian field. Another important class of spherical kernels are nonstationary but include axial symmetry as a special case. These kernels may be based on differential operators ([Jun and Stein 2007, 2008; Jun 2014](#)), on spherical harmonic representations ([Stein 2007; Emery et al. 2019](#)), on kernel convolution ([Heaton et al. 2014](#)), on stochastic partial differential equations ([Lindgren et al. 2011](#)), or on multi-step spectral methods ([Castruccio and Stein 2013; Castruccio and Genton 2014, 2016](#)). [Blake et al. \(2022\)](#) proposed adaptations of three existing families of stationary kernels, namely the Stieltjes construction ([Menegatto 2020](#)), the $\mathcal{F}$ family ([Alegría et al. 2021](#)), and the spectral adaptive approach ([Emery et al. 2021](#)) to achieve general nonstationary covariance kernels without an explicit formulation for axial symmetry. These kernels typically involve more intricate designs and hence less interpretability than chordal-distance-based kernels.

There are several works in the literature comparing the great-arc and the chordal distances in terms of parameter estimation and prediction. [Guinness and Fuentes \(2016\)](#) and [Jeong and Jun \(2015a\)](#) found that the Matérn class with Euclidean distance often performs better than kernels using the great-arc distance. While [Jeong and Jun \(2015b\)](#) and [Alegría et al. \(2021\)](#) found situations where great-arc distance could outperform chordal Matérn kernels, these situations typically assume that a) correlation between locations far apart is still significant and b) there are only a small number of nearby observed locations for prediction locations,

which is typically not the case for the simulated or satellite datasets considered here. Hence, in the present paper, we follow the idea of [Yaglom \(1987\)](#), restricting a valid covariance function in $\mathbb{R}^3$ to $\mathbb{S}$ under the chordal distance, and aim to construct a family of interpretable covariance functions that offer sufficient flexibility to capture the heterogeneous distortion of the tangent planes at different locations on the sphere.

Our covariance construction is based on the locally anisotropic covariance functions for Euclidean space proposed in [Paciorek and Schervish \(2006\)](#). Similar ideas were also discussed in [Katzfuss \(2011\)](#) and [Knapp \(2012\)](#). We will introduce the properties of our general covariance parameterization and specific parameterizations that lead to isotropic or axially symmetric covariance structures to suit various geostatistical applications. For large datasets on the sphere (e.g., with more than 10^4 points), straightforward computation of the Gaussian log-likelihood is too expensive for statistical inference, for which we use the Vecchia approximation ([Vecchia 1988](#)) of Gaussian processes (GPs) in our numerical studies. Our main contributions are: a novel parameterization of the locally anisotropic matrix proposed in [Paciorek and Schervish \(2006\)](#), an exploration of the special cases of the covariance class that ensure isotropy and axial symmetry, and a scalable extension using the Vecchia approximation.

The remainder of this article is organized as follows. Section 2 reviews a nonstationary correlation function on $\mathbb{R}^d$. In Sect. 3, we construct classes of nonstationary covariance functions on the sphere and provide theorems that specify the conditions for isotropic and axially symmetric covariance structures. Section 4 reviews the Vecchia approximation for large spatial datasets. In Sects. 5 and 6, we use simulated data and a precipitation dataset from a physical model to highlight the advantage of our flexible nonstationary covariance structure. Section 7 concludes. Proofs are provided in the Appendix. Code can be found at Github (<https://github.com/katzfuss-group/sphere-local-aniso-cov>).

2. LOCALLY ANISOTROPIC COVARIANCE FUNCTIONS

In this section, we briefly review an intuitive construction for nonstationarity proposed in [Paciorek and Schervish \(2006\)](#) based on any isotropic correlation function, denoted by ρ , in $\mathbb{R}^d$ for all $d \in \mathbb{N}$. Specifically, the nonstationary correlation function is composed as:

$$\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j) = c(\mathbf{s}_i, \mathbf{s}_j) \rho(q(\mathbf{s}_i, \mathbf{s}_j)), \quad (2)$$

$$q(\mathbf{s}_i, \mathbf{s}_j) = \{2(\mathbf{s}_i - \mathbf{s}_j)'(\boldsymbol{\Sigma}(\mathbf{s}_i) + \boldsymbol{\Sigma}(\mathbf{s}_j))^{-1}(\mathbf{s}_i - \mathbf{s}_j)\}^{1/2}, \quad (3)$$

$$c(\mathbf{s}_i, \mathbf{s}_j) = |\boldsymbol{\Sigma}(\mathbf{s}_i)|^{1/4} |\boldsymbol{\Sigma}(\mathbf{s}_j)|^{1/4} |(\boldsymbol{\Sigma}(\mathbf{s}_i) + \boldsymbol{\Sigma}(\mathbf{s}_j))/2|^{-1/2}, \quad (4)$$

where the positive-definite $d \times d$ matrix $\boldsymbol{\Sigma}(\mathbf{s}_i)$ is the local anisotropy matrix describing the spatially varying rotation and scaling, $q(\mathbf{s}_i, \mathbf{s}_j)$ is the Mahalanobis distance with respect to the average anisotropy matrix $(\boldsymbol{\Sigma}(\mathbf{s}_i) + \boldsymbol{\Sigma}(\mathbf{s}_j))/2$, and $c(\mathbf{s}_i, \mathbf{s}_j)$ is the normalization term. The anisotropy matrix $\boldsymbol{\Sigma}(\mathbf{s})$ at each location needs to be a positive-definite matrix in $\mathbb{R}^{d \times d}$, to which we assign the local rotational and scaling effects at the location $\mathbf{s}$. Hence, spatially varying anisotropy is achieved in the covariance structure represented by Eqs. (2), (3), and (4).

This nonstationarity design combines different local anisotropic correlation structures into a valid global correlation function, leading to greater model expressiveness. Under the assumption that the anisotropy matrix $\Sigma(\mathbf{s})$ varies smoothly across the domain, the differentiability of ρ_{NS} follows that of the underlying isotropic correlation function ρ at zero. When ρ is chosen as the isotropic Matérn covariance function, one may also vary the smoothness parameter across the domain to achieve different local smoothness; see Sect. 3.3 for a detailed description.

3. CLASSES OF NONSTATIONARY COVARIANCE FUNCTIONS ON THE SPHERE

Given the intuitive construction for nonstationarity in Sect. 2, an important question is how to parameterize the anisotropy matrix $\Sigma(\mathbf{s})$ to better capture the local covariance structure, which can be largely problem dependent. In this section, we consider one general parameterization of $\Sigma(\mathbf{s})$ that represents local rotation and scaling on a tangent plane of the sphere and probe the conditions for achieving the special cases of isotropic and axially symmetric covariance functions.

3.1. CONSTRUCTION OF THE COVARIANCE FUNCTIONS

Without loss of generality, assume that $\mathbb{S}$ is centered at the origin in $\mathbb{R}^3$, $(0, 0, 0)$, and that the intersection of the prime meridian and the equator, denoted by $\mathbf{c} := (0, 0)$ (0° longitude, 0° latitude), is located on the x -axis (i.e., it has the Euclidean coordinates $\tilde{\mathbf{c}} := (1, 0, 0)$). Figure 1 shows the part of the sphere that lies in the first (positive) octant of a Cartesian coordinate system, including the origin and $\mathbf{c}$. The Euclidean coordinates $\tilde{\mathbf{s}} = (x, y, z)$ of any point $\mathbf{s} = (l, L)$ with longitude l and latitude L on $\mathbb{S}$ are given by:

$$x = \cos(L) \cos(l), \quad y = \cos(L) \sin(l), \quad z = \sin(L).$$

For d -dimensional Euclidean space, we can parameterize $\Sigma(\mathbf{s})$ using d scaling parameters and $d - 1$ rotation parameters (see, e.g., Banerjee et al. 2008). Although $\mathbb{S}$ “lives” in $\mathbb{R}^3$, the surface of $\mathbb{S}$ is (locally) a two-dimensional space (i.e., the tangent plane) at any point $\mathbf{s} \in \mathbb{S}$, indicating a parameterization with only two local scaling and one local rotation parameters. Consider the tangent plane at $\mathbf{c}$, which is the (y, z) -plane, as shown in Fig. 1. We use $\gamma_1(\mathbf{c}) > 0$ and $\gamma_2(\mathbf{c}) > 0$ as the scaling parameters in the y and z -directions, respectively, and $\kappa(\mathbf{c}) \in [0, \pi/2)$ as the rotation parameter, whose collective effect is shown in Fig. 2. Specifically, the scaling matrix at $\mathbf{c}$ is a diagonal matrix $\mathbf{D}(\boldsymbol{\gamma}) := \text{diag}\{1, \gamma_1, \gamma_2\}$ and the rotation matrix that rotates the (y, z) -plane at $\tilde{\mathbf{c}}$ about the x -axis is given by:

$$\mathcal{R}_x(\kappa) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \kappa & -\sin \kappa \\ 0 & \sin \kappa & \cos \kappa \end{pmatrix}.$$

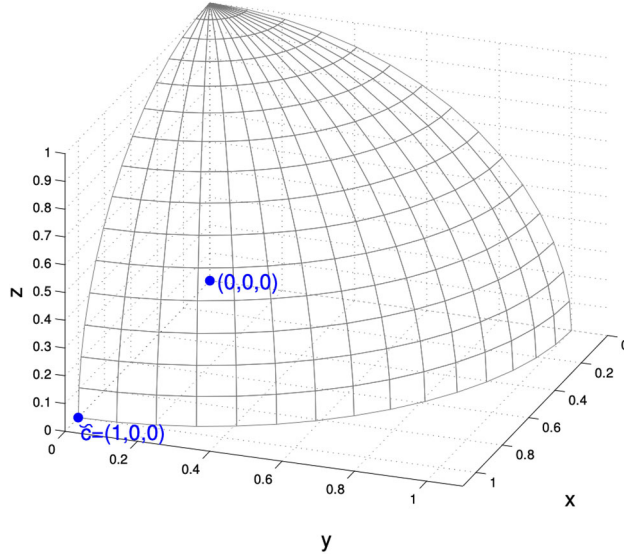


Figure 1. The part of a unit sphere centered at the origin that lies in the first octant of the Cartesian coordinate system, where all coordinates are positive. The origin and the point $\mathbf{c}$ referred to in the text are shown in blue .

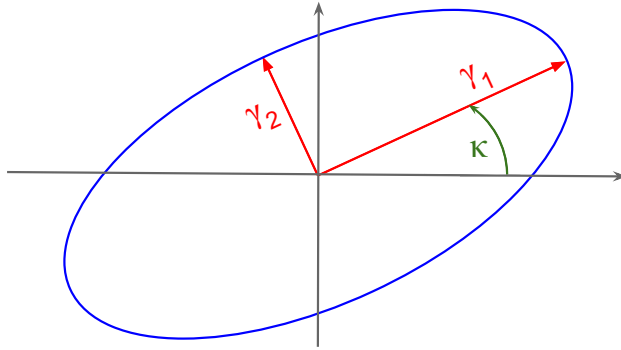


Figure 2. Illustration of the scaling parameters γ_1 and γ_2 and the rotation parameter κ at reference point $\tilde{\mathbf{c}}$, along with an oval representing the resulting correlation contour .

The anisotropy matrix at the reference point $\tilde{\mathbf{c}}$ (spherical coordinates $\mathbf{c}$) is then given by:

$$\tilde{\Sigma}(\mathbf{c}) := \mathcal{R}_x(\kappa(\mathbf{c})) \mathbf{D}(\boldsymbol{\gamma}(\mathbf{c})) \mathcal{R}_x(\kappa(\mathbf{c}))'. \quad (5)$$

Consider an arbitrary location $\mathbf{s}$ on the $\mathbb{S}$, whose Euclidean coordinates are denoted by $\tilde{\mathbf{s}}$. The way we construct the anisotropy matrix for $\mathbf{s}$, denoted by $\Sigma(\mathbf{s})$, is to left and right multiply (5) by the rotation matrix $\mathcal{R}(\mathbf{s})$ such that $\tilde{\mathbf{s}}' \Sigma(\mathbf{s})^{-1} \tilde{\mathbf{s}} = \tilde{\mathbf{s}}' (\mathcal{R}(\mathbf{s}) \tilde{\Sigma}(\mathbf{s}) \mathcal{R}(\mathbf{s})')^{-1} \tilde{\mathbf{s}} = \tilde{\mathbf{c}}' \tilde{\Sigma}(\mathbf{s})^{-1} \tilde{\mathbf{c}}$. Here, $\tilde{\Sigma}(\mathbf{s})$ is parameterized by $\kappa(\mathbf{s})$ and $\boldsymbol{\gamma}(\mathbf{s})$, similar to (5), denoting the local rotation and scaling at $\mathbf{s}$ measured at the reference point $\mathbf{c}$. To rotate $\mathbf{s}$ to $\mathbf{c}$, we can rotate $\tilde{\mathbf{s}}$ about the y -axis and z -axis by $-L$ and $-I$, respectively, which is equivalent to left multiplication of

$\tilde{\mathbf{c}}$ by $\mathcal{R}_y(-L)\mathcal{R}_z(-l)$, where

$$\mathcal{R}_y(\theta) := \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \quad \text{and} \quad \mathcal{R}_z(\theta) := \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence, $\mathcal{R}(\mathbf{s})$ and $\Sigma(\mathbf{s})$ are defined as $\mathcal{R}_z(l)\mathcal{R}_y(L)$ and $\mathcal{R}_z(l)\mathcal{R}_y(L)\tilde{\Sigma}(\mathbf{s})\mathcal{R}_y(L)'\mathcal{R}_z(l)'$, respectively. Therefore, at an arbitrary location $\mathbf{s}$, we have:

$$\tilde{\mathbf{s}}'\Sigma(\mathbf{s})^{-1}\tilde{\mathbf{s}} = \tilde{\mathbf{c}}'\tilde{\Sigma}(\mathbf{s})^{-1}\tilde{\mathbf{c}} = \tilde{\mathbf{c}}'\left(\mathcal{R}_x(\kappa(\mathbf{s}))\mathbf{D}(\boldsymbol{\gamma}(\mathbf{s}))\mathcal{R}_x(\kappa(\mathbf{s}))'\right)^{-1}\tilde{\mathbf{c}}.$$

The anisotropy matrix $\Sigma(\mathbf{s})$ achieves nonstationarity through introducing local rotation and scaling. One can further increasing the nonstationarity through assuming heterogeneous variances in the domain $\sigma^2(\mathbf{s}) > 0$, with which the covariance function between two locations $\mathbf{s}_i$ and $\mathbf{s}_j$ amounts to:

$$C(\mathbf{s}_i, \mathbf{s}_j) = \sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j)\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j),$$

where $\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j)$ is defined by (2) to (4) through the anisotropy matrix $\Sigma(\mathbf{s})$.

3.2. PROPERTIES

The approach above provides a parameterization of $\Sigma(\mathbf{s})$ in terms of two spatially varying ranges, $\boldsymbol{\gamma}(\mathbf{s})$, and one spatially varying rotation, $\kappa(\mathbf{s})$, which can in turn be parameterized in suitable ways for different applications. In this section, we provide conditions on $\boldsymbol{\gamma}(\mathbf{s})$ and $\kappa(\mathbf{s})$ such that the resulting correlation function in (2) is isotropic or axially symmetric; proofs of the theorems are included in Appendix A.

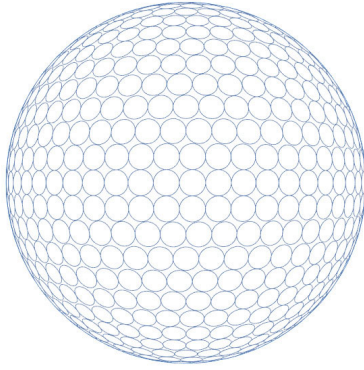
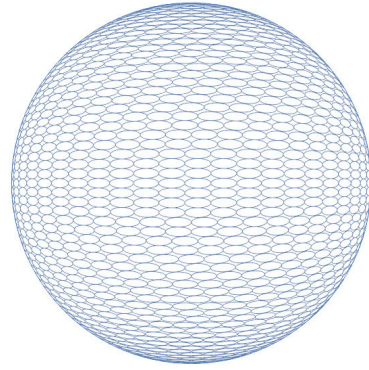
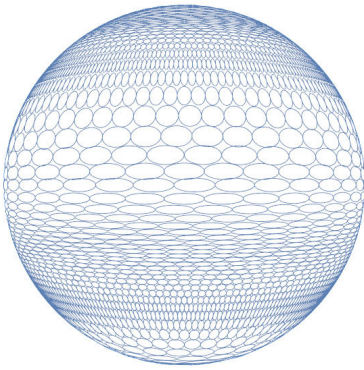
An isotropic covariance function is a function of only distance between two locations. Due to the one-to-one relationship between the great-arc and the chordal distances in (1), isotropic covariance functions on the sphere are isotropic with respect to both chordal and great-arc distance. By adding constraints to the scaling parameters $\gamma_1(\mathbf{s})$ and $\gamma_2(\mathbf{s})$, the correlation function ρ_{NS} in (2) can achieve isotropy on the sphere:

Theorem 1. *The correlation function ρ_{NS} in (2) is isotropic (i.e., depends only on distance) if $\gamma_1(\mathbf{s}) = \gamma_2(\mathbf{s}) \equiv \gamma$ is constant.*

A subclass of covariance functions that are specifically useful for spherical domains are axially symmetric covariance functions (e.g., Stein 2007), under which the correlation between a pair of locations on the sphere depends on longitudes only through the longitude difference:

Definition. A covariance function $C : \mathbb{S} \times \mathbb{S} \rightarrow \mathbb{R}$ is called axially symmetric if there exists a function C_A such that

$$C(\mathbf{s}_i, \mathbf{s}_j) = C_A(l_i - l_j, L_i, L_j),$$

(a) Isotropy ($\gamma_1 = \gamma_2$)(b) “Anisotropy” ($\kappa = 0; \gamma_1 > \gamma_2$)

(c) Axial symmetry

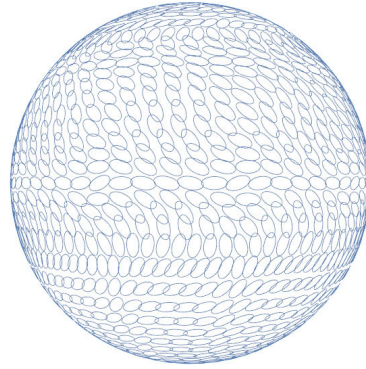
(d) General nonstationarity with spatially varying rotation $\kappa(\cdot)$

Figure 3. Illustration of special cases of the nonstationary correlation functions in (2) via correlation contours .

where $\mathbf{s}_i = (l_i, L_i)$ and $\mathbf{s}_j = (l_j, L_j)$ with longitudes l_i and l_j and latitudes L_i and L_j on $\mathbb{S}$.

Axially symmetric covariance functions can be also obtained based on the theorem below:

Theorem 2. *The correlation function ρ_{NS} in (2) is axially symmetric if $\kappa(\mathbf{s}) \equiv 0$ and $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are functions of latitude only (i.e., they do not depend on longitude).*

Special cases of our general nonstationary covariance function that include isotropic, anisotropic, axially symmetric, and general nonstationary parameterizations are visualized in Fig. 3.

3.3. EXAMPLE: A NONSTATIONARY MATÉRN COVARIANCE ON THE SPHERE

The Matérn correlation function is highly popular in geospatial analysis. It is valid in $\mathbb{R}^d$ for any $d \in \mathbb{N}$ and given by

$$\mathcal{M}_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} r^\nu \mathcal{K}_\nu(r), \quad r \geq 0,$$

where $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order $\nu > 0$. The standard deviation, denoted by $\sigma(\cdot)$, and the smoothness parameter $\nu(\cdot)$ in the Matérn can also vary over space (Stein 2005). Hence, we can obtain a highly flexible Matérn covariance on the sphere of the form

$$\mathcal{M}_{NS}(\mathbf{s}_i, \mathbf{s}_j) = \sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j)c(\mathbf{s}_i, \mathbf{s}_j)\mathcal{M}_{(\nu(\mathbf{s}_i)+\nu(\mathbf{s}_j))/2}(q(\mathbf{s}_i, \mathbf{s}_j)), \quad (6)$$

where c, q are as in (3)–(4), and $\Sigma(\mathbf{s})$ can be parameterized in terms of spatially varying scales $\boldsymbol{\nu}(\mathbf{s})$ and rotation $\rho(\mathbf{s})$ as in Sect. 3.1.

Guinness and Fuentes (2016) showed that the local smoothness properties of the Matérn covariance are preserved when restricting a process in Euclidean space to the sphere. Specifically, a GP with covariance function $\mathcal{M}_{NS}$ has m mean square derivatives at $\mathbf{s}$ if and only if $\nu(\mathbf{s}) > m$.

4. VECCHIA APPROXIMATION

For many modern large datasets, including those on the sphere, direct application of GPs is too computationally expensive, as the cost scales cubically in the number of data points. The approximation proposed by Vecchia (1988) has become highly popular in recent years, which has linear computational complexity and straightforward parallel features while maintaining high accuracy measured by the Kullback–Leibler (KL) divergence from the true process (e.g., Guinness 2018; Katzfuss and Guinness 2021). Based on a given ordering of the observations, the Vecchia approximation replaces the high-dimensional joint multivariate normal density with a product of univariate conditional normal densities, in which each variable conditions only on a small subset of previous observations in the ordering, amounting to an ordered conditionally independent approximation.

Denote $h(i) = \{1, 2, \dots, i-1\}$ with $h(1) = \emptyset$ and $\mathbf{y}_{h(i)} = (y_1, \dots, y_{i-1})'$. Consider a GP $y(\cdot) \sim GP(0, C)$ on a spatial region $\mathcal{D}$ with zero mean and covariance function C . The joint density of the observation $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is given by

$$f(\mathbf{y}) = \prod_{i=1}^n f(y_i | \mathbf{y}_{h(i)}).$$

The Vecchia approximation replaces $h(i)$ with a subset $g(i) \subset h(i)$, where $g(i)$ is usually chosen to select those indices corresponding to the m observations nearest in distance to the i th observation. This leads to the following approximation of the joint density:

$$\hat{f}(\mathbf{y}) = \prod_{i=1}^n f(y_i | \mathbf{y}_{g(i)}). \quad (7)$$

The Vecchia approximation ensures computational feasibility for large spatial datasets. The choices for ordering the locations and selecting the conditioning sets $\{g(i)\}_{i=1}^n$ are

typically based on distance or estimated correlation (Katzfuss and Guinness 2021). Here, we will use maximum-minimum-distance ordering (Guinness 2018) and nearest-neighbor conditioning in the numerical studies, both based on the chordal distances. Correlation-based ordering and conditioning that takes into account the potential nonstationary structure is also possible (Katzfuss et al. 2022; Kang and Katzfuss 2021). Aside from likelihood-based parameter inference based on the Vecchia likelihood in (7), the Vecchia approximation can also be applied to unknown locations in order to obtain accurate approximations of posterior predictive distributions (e.g., Katzfuss et al. 2020a). In the case of noisy data, the Vecchia approximation can be applied to the latent (noise-free) GP as before and then combined with an incomplete-Cholesky decomposition of the posterior precision matrix to preserve the low computational complexity (Schäfer et al. 2021). We implemented Vecchia inference based on our new covariance function by extending the R package GPvecchia (Katzfuss et al. 2020b).

5. NUMERICAL STUDY

We perform simulations to demonstrate the improvement of posterior inference gained from adopting a more flexible covariance structure. Specifically, we simulate GPs that are isotropic, axially symmetric, and general nonstationary based on different parameterizations of the covariance structure introduced in (6). The scaling parameters $\gamma_1(\mathbf{s})$ and $\gamma_2(\mathbf{s})$ are parameterized as:

$$\gamma_1(\mathbf{s}) = \exp(\beta_{10} + \beta_{11} \sin(l) + \beta_{12} L), \quad (8)$$

$$\gamma_2(\mathbf{s}) = \exp(\beta_{20} + \beta_{21} \sin(l) + \beta_{22} L), \quad (9)$$

where $\mathbf{s} = (l, L)$ with longitude l and latitude L . Based on Theorems 1 and 2, isotropic, axially symmetric, and general nonstationary covariance structures are constructed as follows:

Isotropy: According to Theorem 1, the correlation function ρ_{NS} in (2) is isotropic if $\gamma_1(\mathbf{s}) = \gamma_2(\mathbf{s}) \equiv \gamma$ is constant, and so we set the parameters in (8)–(9) as

$$\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (-0.5, 0, 0),$$

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (-0.5, 0, 0).$$

Axial symmetry: According to Theorem 2, the correlation function ρ_{NS} in (2) is axially symmetric if $\kappa(\mathbf{s}) \equiv 0$ and $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are functions of latitude only and we set the parameters as

$$\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12}) = (-0.5, 0, 1.44),$$

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}) = (-3.2, 0, 1.44).$$

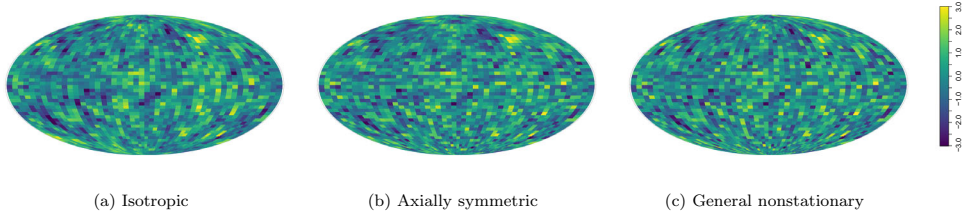


Figure 4. Realizations of the isotropic, axially symmetric, and general nonstationary GPs over a 50×50 grid on the sphere .

General nonstationarity: A more general nonstationary covariance function can be obtained by setting

$$\begin{aligned}\kappa &= 0.8, \\ \beta_1 &= (\beta_{10}, \beta_{11}, \beta_{12}) = (-0.5, -1.2, 1.44), \\ \beta_2 &= (\beta_{20}, \beta_{21}, \beta_{22}) = (-3.2, -0.3, 1.44).\end{aligned}$$

Each dataset is generated on a regular latitude/longitude grid of size $50 \times 50 = 2,500$ on the sphere and then split into training and testing subsets, under two different scenarios: (a) a random sampling of 20% of all locations as the test dataset; (b) ten randomly selected regions, each with a longitudinal band width of 0.4 and a latitudinal band width of 0.2, that sum up to approximately 20% of all locations as the test dataset. The training dataset is then modeled by the three progressively more flexible covariance structures:

Isotropy: unknown β_{10}, β_{20} , with $\beta_{10} = \beta_{20}$ and fixed $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = \kappa = 0$.

Axial symmetry: unknown $\beta_{10}, \beta_{12}, \beta_{20}, \beta_{22}$, with fixed $\beta_{11} = \beta_{21} = \kappa = 0$.

General nonstationarity: unknown $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}, \kappa$.

Realizations of the isotropic, axially symmetric, and general nonstationary GPs are shown in Fig. 4.

The prediction scores for the nine combinations of ‘true’ and ‘assumed’ covariance structures are summarized in Table 1 that include the mean absolute error (MAE), the root mean squared error (RMSE), the continuous ranked probability score (CRPS), and the energy score. MAE and RMSE consider only point predictions, CRPS evaluates marginal predictive distributions, and the energy score evaluates the joint predictive distribution for the entire test set. The variance and smoothness parameters of the Matérn covariance function $\mathcal{M}_\nu(r)$ are $\sigma(s) = 1$ and $\nu(s) = 0.5$, respectively, both considered as known. We ran an adaptive MCMC algorithm (Vihola 2012) for 5,000 iterations for Bayesian inference on the unknown parameters. We applied the Vecchia likelihood approximation in (7), which uses maximum–minimum-distance ordering and nearest-neighbor conditioning with a conditioning-set size of ten. To obtain the posterior predictive distribution at the testing locations, we also used the Vecchia approximation (i.e., ordinary kriging with ten nearest neighbors), based on MCMC

Table 1. Prediction scores (lower is better), each averaged over five simulated datasets, for the nine different combinations of the true and assumed covariance structures

	Random				Region			
	MAE	RMSE	CRPS	Energy	MAE	RMSE	CRPS	Energy
<i>True model— isotropic</i>								
Isotropic	0.569	0.728	0.563	16.1	0.716	0.904	0.710	18.8
Axially symmetric	0.567	0.727	0.556	15.9	0.716	0.904	0.705	18.6
Nonstationary	0.568	0.728	0.551	15.7	0.716	0.904	0.698	18.4
<i>True model—axially symmetric</i>								
Isotropic	0.754	0.961	0.751	21.3	0.768	0.968	0.761	20.1
Axially symmetric	0.637	0.834	0.621	18.1	0.741	0.932	0.732	19.2
Nonstationary	0.637	0.835	0.616	18.0	0.741	0.931	0.727	19.1
<i>True model—nonstationary</i>								
Isotropic	0.734	0.938	0.730	20.8	0.777	0.973	0.773	20.3
Axially symmetric	0.688	0.883	0.671	19.2	0.761	0.953	0.752	19.6
Nonstationary	0.681	0.874	0.659	18.8	0.754	0.943	0.739	19.3

Test sets are selected as random locations (random) or regions (region) that amount to 20% of the total dataset

samples of the unknown parameters after a burn-in of size one thousand; the resulting predictive distribution is a mixture of Gaussians.

For datasets generated by isotropic GPs, the prediction scores are very similar under the assumptions of isotropic, axially symmetric, and nonstationary covariance structures. The difference becomes more pronounced when the datasets are generated from an axially symmetric GP, where the axially symmetric and the nonstationary structures have similar performance. Specifically, both the axially symmetric and the nonstationary structures have a 13% improvement in RMSE and a 17% improvement in CRPS for random test locations relative to the baseline isotropic structure, highlighting the improvement in point prediction accuracy and uncertainty quantification. Similar conclusions can be drawn for the unknown-region scenario, although the performance difference becomes less significant. Furthermore, in modeling the general nonstationary GPs, there is a uniform improvement in all prediction scores when switching from the axially symmetric to the general nonstationary covariance structure. Notice that the improvement of general nonstationary over isotropic is smaller when the true model is general nonstationary than when the true model is axially symmetric, because the extra task of training β_{11} , β_{21} , and κ made the optimization more challenging overall. The prediction accuracy was barely affected when using a model more flexible than the truth (i.e., fitting a general nonstationary model to axially symmetric data), but significant accuracy loss may happen when the covariance kernel lacks expressiveness (i.e., fitting an isotropic model to axially symmetric data).

6. APPLICATION TO PRECIPITATION DATA

Precipitation data are often collected by sensors at irregularly spaced weather-monitoring stations, necessitating statistical inference in regions with insufficient sensor coverage or temporal misalignment. Accurate modeling of nonstationary dependence is often crucial, for

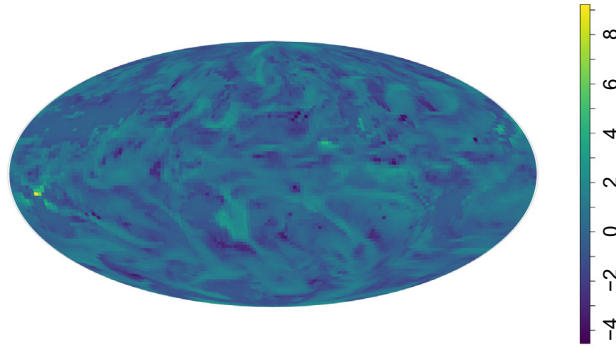


Figure 5. Visualization of the precipitation dataset after preprocessing .

Table 2. Prediction scores, using the precipitation data, of isotropic, axially symmetric, and general non-stationary structures

	Random				Region			
	MAE	RMSE	CRPS	Energy	MAE	RMSE	CRPS	Energy
Isotropic	0.383	0.549	0.376	27.1	0.756	0.967	0.749	43.1
Axially symmetric	0.268	0.426	0.264	21.6	0.654	0.834	0.649	37.7
Nonstationary	0.269	0.428	0.263	21.6	0.651	0.832	0.620	35.7

Test sets are selected as random locations (random) or regions (region) that amount to 20% of the total dataset

example for quantifying the probability of flooding due to the total rainfall in a catchment area exceeding a particular threshold. In this section, we examine the inference improvements that can be achieved by increasing the flexibility of the covariance kernel function. Using the same three types of covariance structures used in Sect. 5, we model a precipitation dataset from the Community Earth System Model (CESM) Large Ensemble Project (Kay et al. 2015). After subsetting, the dataset contains precipitation rates (m/s) on July 1, 401, on a roughly a 2° resolution in terms of longitude and latitude, totaling $144 \times 96 = 13,824$ locations on a spherical grid. We consider the standardized log-precipitation anomalies (i.e., log-precipitation standardized across time) shown in Fig. 5, which does not indicate any distinct mean structure. 20% of the dataset is used for testing, selected either as random locations or as random regions, similar to Sect. 5.

Model parameters have the same initializations under the three model assumptions, and the smoothness parameter ν for the CESM dataset is fixed at 2.5 due to the increased smoothness compared with the simulated datasets in Fig. 4. We also summarize the performance (e.g., scores) of posterior inference in Table 2. The isotropic covariance structure is significantly out-performed by the axially symmetric and the nonstationary structures while the difference between the latter two is indistinguishable. To visualize the difference between the trained kernel functions, we plotted the 0.7-correlation contours centered at evenly distributed locations on the sphere in Fig. 6.

The contours from the axially symmetric and nonstationary kernels closely resemble each other, and they are both significantly bigger than those of the isotropic kernel. This indicates

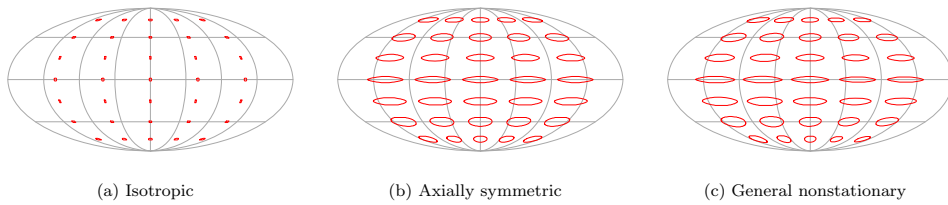


Figure 6. 0.7-correlation contours of the trained isotropic, axially symmetric, and general nonstationary GPs on the sphere .

that using an isotropic covariance structure can be insufficient for practical modeling and a more flexible kernel can capture the correlation structure more precisely, hence producing more accurate inference. For the axially symmetric and nonstationary kernels, elongated ovals indicate strong dependence along the equator, while the dependence is weaker and more isotropic near the poles. The decay of correlation from a location of interest can be inferred in a similar manner as in Fig. 2 along different directions.

7. CONCLUSIONS

We proposed an extension of [Paciorek and Schervish \(2006\)](#) for constructing nonstationary, locally anisotropic covariance functions on the sphere based on isotropic covariance functions in $\mathbb{R}^3$. Special parameterizations of the nonstationary covariance function that amount to isotropic and axially symmetric covariance structures were also discussed. Axially symmetric covariance functions are widely used in geospatial analysis of global data, and their advantages over isotropic covariance structures were demonstrated with both simulated Gaussian random fields and a CESM precipitation dataset. The extra flexibility of our nonstationary parameterization also improved posterior inference compared with the axially symmetric structure, although the improvement was less significant. For large datasets on the sphere, straightforward computation of Gaussian probabilities is typically too computationally expensive, and so the Vecchia family of approximations based on nearest neighbors is often used. This nearest-neighbor approach further reduces any potentially advantages of great-arc-distance-based kernels to outperform the chordal-distance-based kernels, which typically rely on long-range dependence and lack of close-by observed locations ([Jeong and Jun 2015b](#); [Alegría et al. 2021](#)).

The locally anisotropic nature of our approach ensures an analytical and interpretable expression for the location correlation structure, but it also limits the dependence that can be expressed; for example, dependence along a curved ridge is difficult to capture with our approach. The flexibility in our construction of the anisotropy matrix $\Sigma(\mathbf{s})$ based on (5), (8), and (9) could be increased by modeling $\log(\gamma_1(\mathbf{s}))$ and $\log(\gamma_2(\mathbf{s}))$ as Gaussian processes.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

A PROOFS

Proof of Theorem 1. If $\gamma_1(\mathbf{s}) = \gamma_2(\mathbf{s}) \equiv \gamma$ is constant, then

$$\begin{aligned} \mathbf{D}(\gamma) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \\ \tilde{\Sigma}(s) &= \mathcal{R}_x(\kappa(s))\mathbf{D}(\gamma(\mathbf{s}))\mathcal{R}_x(\kappa(s))' \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \kappa(s) & -\sin \kappa(s) \\ 0 & \sin \kappa(s) & \cos \kappa(s) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \kappa(s) & \sin \kappa(s) \\ 0 & -\sin \kappa(s) & \cos \kappa(s) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{pmatrix}. \end{aligned}$$

To compute $\Sigma(\mathbf{s}) = \mathcal{R}_z(l)\mathcal{R}_y(L)\tilde{\Sigma}(s)\mathcal{R}_y(L)'\mathcal{R}_z(l)'$, we first compute

$$\begin{aligned} \mathbf{A} := \mathcal{R}_z(l)\mathcal{R}_y(L) &= \begin{pmatrix} \cos(l) - \sin(l) & 0 \\ \sin(l) & \cos(l) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(L) & 0 & \sin(L) \\ 0 & 1 & 0 \\ -\sin(L) & 0 & \cos(L) \end{pmatrix} \\ &= \begin{pmatrix} \cos(l)\cos(L) - \sin(l)\sin(L) & \sin(L) \\ \sin(l)\cos(L) & \cos(l)\sin(L) \\ -\sin(L) & 0 & \cos(L) \end{pmatrix}, \\ \Sigma(\mathbf{s}) = \mathbf{A}\tilde{\Sigma}(\mathbf{s})\mathbf{A}' &= \begin{pmatrix} (1-\gamma)x^2 + \gamma & (1-\gamma)xy & (1-\gamma)xz \\ (1-\gamma)xy & (1-\gamma)y^2 + \gamma & (1-\gamma)yz \\ (1-\gamma)xz & (1-\gamma)yz & (1-\gamma)z^2 + \gamma \end{pmatrix} \\ &= (1-\gamma) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \begin{pmatrix} x & y & z \end{pmatrix} + \gamma \mathbf{I}_3 \\ &= (1-\gamma)\tilde{\mathbf{s}}\tilde{\mathbf{s}}' + \gamma \mathbf{I}_3, \end{aligned}$$

where $x = \cos(L)\cos(l)$, $y = \cos(L)\sin(l)$, $z = \sin(L)$ are the (x, y, z) -coordinates of a three-dimensional Cartesian coordinate system. Then

$$\begin{aligned} |\Sigma(\mathbf{s})| &= \det\{(1-\gamma)\tilde{\mathbf{s}}\tilde{\mathbf{s}}' + \gamma \mathbf{I}_3\} \\ &= \gamma^3 \cdot \det\left\{\mathbf{I}_3 + \frac{1-\gamma}{\gamma}\tilde{\mathbf{s}}\tilde{\mathbf{s}}'\right\} \\ &= \gamma^3 \cdot \det\left\{1 + \frac{1-\gamma}{\gamma}\tilde{\mathbf{s}}'\tilde{\mathbf{s}}\right\} \\ &= \gamma^3 \cdot \det\left\{1 + \frac{1-\gamma}{\gamma} \cdot 1\right\} \\ &= \gamma^2 \end{aligned}$$

does not depend on $\mathbf{s}$. And for $i \neq j$,

$$(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1} = \left((1 - \gamma)\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i' + (1 - \gamma)\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + 2\gamma\mathbf{I}_3 \right)^{-1}.$$

WLOG, we ignore the constant coefficients inside the inverse, and then

$$\begin{aligned} (\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1} &= \left(\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i' + \tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3 \right)^{-1} \\ &= (\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1} - (\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1}\tilde{\mathbf{s}}_i \\ &\quad [1 + \tilde{\mathbf{s}}_i'(\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1}\tilde{\mathbf{s}}_i]^{-1}\tilde{\mathbf{s}}_i'(\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1}. \end{aligned}$$

Let $\mathbf{B} := (\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1}$, and so

$$\begin{aligned} (\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1} &= \mathbf{B} - \mathbf{B}\tilde{\mathbf{s}}_i(1 + \tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_i)^{-1}\tilde{\mathbf{s}}_i'B = \mathbf{B} - \frac{\mathbf{B}\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i'\mathbf{B}}{1 + \tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_i}, \\ q^2(\mathbf{s}_i, s_j) &\propto (\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) \\ &\propto (\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\mathbf{B}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) - \frac{1}{1 + \tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_i}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\mathbf{B}\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i'\mathbf{B}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j). \end{aligned}$$

So, computation of $q(\mathbf{s}_i, s_j)$ only involves terms $\tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_i$, $\tilde{\mathbf{s}}_j'B\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_j$. Because

$$\mathbf{B} = (\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \mathbf{I}_3)^{-1} = \mathbf{I}_3 - \tilde{\mathbf{s}}_j(1 + \tilde{\mathbf{s}}_j'\tilde{\mathbf{s}}_j)^{-1}\tilde{\mathbf{s}}_j' = \mathbf{I}_3 - \frac{1}{2}\tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j',$$

we have

$$\begin{aligned} \tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_i &= \tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_i - \frac{1}{2}(\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j)^2 = 1 - \frac{1}{2}(\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j)^2 \\ \tilde{\mathbf{s}}_j'B\tilde{\mathbf{s}}_j &= \tilde{\mathbf{s}}_j'\tilde{\mathbf{s}}_j - \frac{1}{2}(\tilde{\mathbf{s}}_j'\tilde{\mathbf{s}}_j)^2 = 1 - \frac{1}{2} = \frac{1}{2} \\ \tilde{\mathbf{s}}_i'B\tilde{\mathbf{s}}_j &= \tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j - \frac{1}{2}(\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j)(\tilde{\mathbf{s}}_j'\tilde{\mathbf{s}}_j) = \tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j - \frac{1}{2}\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j = \frac{1}{2}\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j. \end{aligned}$$

Further,

$$\tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_j = \left[(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) - \tilde{\mathbf{s}}_i'\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j'\tilde{\mathbf{s}}_j \right] / 2 = \left[(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) - 2 \right] / 2.$$

So, $q(\mathbf{s}_i, \mathbf{s}_j)$ just depends on the distance $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$. For the normalization term $c(\mathbf{s}_i, \mathbf{s}_j)$, since we have proved that $|\Sigma(\mathbf{s}_i)| = |\Sigma(\mathbf{s}_j)| \equiv \gamma^2$,

$$\begin{aligned}
 c(\mathbf{s}_i, \mathbf{s}_j) &= |\Sigma(\mathbf{s}_i)|^{1/4} |\Sigma(\mathbf{s}_j)|^{1/4} |(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))/2|^{-1/2} \\
 &\propto |(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}|^{1/2} \\
 &\propto \left(|(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}| \cdot |(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)| / [(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)] \right)^{1/2} \\
 &\propto \left(|(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}| \cdot \det\{(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\} / [(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)] \right)^{1/2} \\
 &\propto \left(\det\{(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\} / [(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)] \right)^{1/2} \\
 &\propto \left(\det\{(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)\} / [(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)] \right)^{1/2} \\
 &\propto \left\{ \frac{q^2(\mathbf{s}_i, \mathbf{s}_j)}{(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)} \right\}^{1/2}.
 \end{aligned} \tag{10}$$

We have proved that $q(\mathbf{s}_i, \mathbf{s}_j)$ just depends on $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$, so $c(\mathbf{s}_i, \mathbf{s}_j)$ also only depends on the distance $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$.

Overall, we can show

$$\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j) = c(\mathbf{s}_i, \mathbf{s}_j) \rho(q(\mathbf{s}_i, \mathbf{s}_j))$$

only depends on the distance $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$, where $\rho(q)$ is a valid isotropic correlation function. So, $\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j)$ is isotropic. $\square$

Proof of Theorem 2. If $\kappa(\mathbf{s}) \equiv 0$ and $\gamma_1(\cdot)$, $\gamma_2(\cdot)$ depend on L only, then $\mathcal{R}_x(\kappa(s)) \equiv \mathcal{R}_x(0) = \mathbf{I}_3$. Then

$$\tilde{\Sigma}(s) = \mathbf{D}(\gamma(\mathbf{s})) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma_1(L) & 0 \\ 0 & 0 & \gamma_2(L) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma_1(L) & 0 \\ 0 & 0 & \gamma_1(L) \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \gamma_2(L) - \gamma_1(L) \end{pmatrix}.$$

Due to the results in Theorem 1, we have

$$\begin{aligned}
 \Sigma(\mathbf{s}) &= \mathcal{R}_z(l) \mathcal{R}_y(L) \tilde{\Sigma}(s) \mathcal{R}_y(L)' \mathcal{R}_z(l)' \\
 &= (1 - \gamma_1(L)) \tilde{\mathbf{s}} \tilde{\mathbf{s}}' + \gamma_1(L) \mathbf{I}_3 + (\gamma_2(L) - \gamma_1(L)) \tilde{\mathbf{s}}^* (\tilde{\mathbf{s}}^*)',
 \end{aligned}$$

where

$$\tilde{\mathbf{s}}^* = \begin{pmatrix} \cos(l) \sin(L) \\ \sin(l) \sin(L) \\ \cos(L) \end{pmatrix} \quad (\tilde{\mathbf{s}}^*)'(\tilde{\mathbf{s}}^*) = 1.$$

Thus,

$$\begin{aligned}
|\Sigma(\mathbf{s})| &= \det\{(1 - \gamma_1(L)) \tilde{\mathbf{s}}\tilde{\mathbf{s}}' + \gamma_1(L)\mathbf{I}_3 + (\gamma_2(L) - \gamma_1(L))\tilde{\mathbf{s}}^*(\tilde{\mathbf{s}}^*)'\} \\
&= \gamma_1(L)^3 \cdot \det\left\{\frac{1 - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}}\tilde{\mathbf{s}}' + \frac{\gamma_2(L) - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}}^*(\tilde{\mathbf{s}}^*)' + \mathbf{I}_3\right\} \\
&= \gamma_1(L)^3 \cdot \det\left\{\left(\frac{1 - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}} \quad \frac{\gamma_2(L) - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}}^*\right)\begin{pmatrix} \tilde{\mathbf{s}}' \\ (\tilde{\mathbf{s}}^*)' \end{pmatrix} + \mathbf{I}_3\right\} \\
&= \gamma_1(L)^3 \cdot \det\left\{\begin{pmatrix} \tilde{\mathbf{s}}' \\ (\tilde{\mathbf{s}}^*)' \end{pmatrix}\left(\frac{1 - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}} \quad \frac{\gamma_2(L) - \gamma_1(L)}{\gamma_1(L)}\tilde{\mathbf{s}}^*\right) + \mathbf{I}_2\right\} \\
&= \gamma_1(L)^3 \cdot \begin{vmatrix} \frac{1 - \gamma_1(L)}{\gamma_1(L)} + 1 & 2 \sin(L) \cos(L) \\ 2 \sin(L) \cos(L) & \frac{\gamma_2(L) - \gamma_1(L)}{\gamma_1(L)} + 1 \end{vmatrix} \\
&= \gamma_1(L)\gamma_2(L) - 4\gamma_1(L)^3 \sin^2(L) \cos^2(L)
\end{aligned}$$

only depend on L . WLOG, ignore $\gamma_1(L)$, $\gamma_2(L)$ again (they only depend on L),

$$\begin{aligned}
&(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1} \\
&= \left[\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i' + \tilde{\mathbf{s}}_i^*(\tilde{\mathbf{s}}_i^*)' + \tilde{\mathbf{s}}_j\tilde{\mathbf{s}}_j' + \tilde{\mathbf{s}}_j^*(\tilde{\mathbf{s}}_j^*)' + \mathbf{I}_3\right]^{-1} \\
&= \left[\begin{pmatrix} \tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i^* \end{pmatrix}\begin{pmatrix} \tilde{\mathbf{s}}_i' \\ (\tilde{\mathbf{s}}_i^*)' \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{s}}_j & \tilde{\mathbf{s}}_j^* \end{pmatrix}\begin{pmatrix} \tilde{\mathbf{s}}_j' \\ (\tilde{\mathbf{s}}_j^*)' \end{pmatrix} + \mathbf{I}_3\right]^{-1} \\
&= \mathbf{V}^{-1} - \mathbf{V}^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i^* \end{pmatrix}\left[\mathbf{I}_2 + \begin{pmatrix} \tilde{\mathbf{s}}_i' \\ (\tilde{\mathbf{s}}_i^*)' \end{pmatrix}\mathbf{V}^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i^* \end{pmatrix}\right]^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i' \\ (\tilde{\mathbf{s}}_i^*)' \end{pmatrix}\mathbf{V}^{-1} \\
&= \mathbf{V}^{-1} - \mathbf{V}^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i^* \end{pmatrix}\left[\mathbf{I}_2 + \begin{pmatrix} \tilde{\mathbf{s}}_i'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i^* \\ (\tilde{\mathbf{s}}_i^*)'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i & (\tilde{\mathbf{s}}_i^*)'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i^* \end{pmatrix}\right]^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i' \\ (\tilde{\mathbf{s}}_i^*)' \end{pmatrix}\mathbf{V}^{-1},
\end{aligned}$$

where

$$\mathbf{V} = \begin{pmatrix} \tilde{\mathbf{s}}_j & \tilde{\mathbf{s}}_j^* \end{pmatrix}\begin{pmatrix} \tilde{\mathbf{s}}_j' \\ (\tilde{\mathbf{s}}_j^*)' \end{pmatrix} + \mathbf{I}_3.$$

Then

$$\begin{aligned}
q^2(\mathbf{s}_i, \mathbf{s}_j) &\propto (\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) \\
&\propto (\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\mathbf{V}^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) - (\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'\mathbf{V}^{-1}\begin{pmatrix} \tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i^* \end{pmatrix} \\
&\quad \left[\mathbf{I}_2 + \begin{pmatrix} \tilde{\mathbf{s}}_i'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i & \tilde{\mathbf{s}}_i'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i^* \\ (\tilde{\mathbf{s}}_i^*)'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i & (\tilde{\mathbf{s}}_i^*)'\mathbf{V}^{-1}\tilde{\mathbf{s}}_i^* \end{pmatrix}\right]^{-1} \\
&\quad \begin{pmatrix} \tilde{\mathbf{s}}_i' \\ (\tilde{\mathbf{s}}_i^*)' \end{pmatrix}\mathbf{V}^{-1}(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j).
\end{aligned}$$

Because

$$\begin{aligned} \mathbf{V}^{-1} &= \mathbf{I}_3 - \begin{pmatrix} \tilde{\mathbf{s}}_j & \tilde{\mathbf{s}}_j^* \end{pmatrix} \left[\mathbf{I}_2 + \begin{pmatrix} 1 & \tilde{\mathbf{s}}_j' \tilde{\mathbf{s}}_j^* \\ (\tilde{\mathbf{s}}_j)' \tilde{\mathbf{s}}_j & 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} \tilde{\mathbf{s}}_j' \\ (\tilde{\mathbf{s}}_j^*)' \end{pmatrix} \\ &= \mathbf{I}_3 - \frac{1}{4 - (\tilde{\mathbf{s}}_j' \tilde{\mathbf{s}}_j^*)^2} \begin{pmatrix} \tilde{\mathbf{s}}_j & \tilde{\mathbf{s}}_j^* \end{pmatrix} \left[\begin{pmatrix} 2 & -\tilde{\mathbf{s}}_j' \tilde{\mathbf{s}}_j^* \\ -(\tilde{\mathbf{s}}_j)' \tilde{\mathbf{s}}_j & 2 \end{pmatrix} \right]^{-1} \begin{pmatrix} \tilde{\mathbf{s}}_j' \\ (\tilde{\mathbf{s}}_j^*)' \end{pmatrix}, \end{aligned}$$

we can figure out that the computation of $q^2(\mathbf{s}_i, \mathbf{s}_j)$ only involves the following types of terms

$$\left\{ \begin{array}{l} \tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_i = 1 \\ \tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_i^* = \tilde{\mathbf{s}}_i' \cdot \left[(\tan(s_{i2})) \tilde{\mathbf{s}}_i + \left(0, 0, \cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right)' \right] \\ \quad = \tan(s_{i2}) + \sin(s_{i2}) \left[\cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right] \\ (\tilde{\mathbf{s}}_i^*)' \tilde{\mathbf{s}}_i^* = 1 \\ (\tilde{\mathbf{s}}_i^*)' \tilde{\mathbf{s}}_j^* \\ \quad = \left[(\tan(s_{i2})) \tilde{\mathbf{s}}_i + \left(0, 0, \cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right)' \right]' \\ \quad \cdot \left[(\tan(s_{j2})) \tilde{\mathbf{s}}_j + \left(0, 0, \cos(s_{j2}) - \frac{\sin^2(s_{j2})}{\cos(s_{j2})} \right)' \right] \\ \quad = \tan(s_{i2}) \tan(s_{j2}) (\tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j) + \tan(s_{i2}) \sin(s_{i2}) \left[\cos(s_{j2}) - \frac{\sin^2(s_{j2})}{\cos(s_{j2})} \right] \\ \quad \quad + \tan(s_{j2}) \sin(s_{j2}) \left[\cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right] \\ \quad \quad + \left[\cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right] \left[\cos(s_{j2}) - \frac{\sin^2(s_{j2})}{\cos(s_{j2})} \right] \\ \tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j^* = \tan(s_{j2}) (\tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j) + \sin(s_{i2}) \left[\cos(s_{j2}) - \frac{\sin^2(s_{j2})}{\cos(s_{j2})} \right] \\ \tilde{\mathbf{s}}_j' \tilde{\mathbf{s}}_i^* = \tan(s_{i2}) (\tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j) + \sin(s_{j2}) \left[\cos(s_{i2}) - \frac{\sin^2(s_{i2})}{\cos(s_{i2})} \right] \\ \tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j = [(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j) - 2] / 2. \end{array} \right.$$

We can change the index i to j for the first three terms, and they are still valid. Thus, these values only depend on s_{i2} , s_{j2} and $\tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j$, and $\tilde{\mathbf{s}}_i' \tilde{\mathbf{s}}_j$ can be expressed in terms of the distance $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$. The computation of $q^2(\mathbf{s}_i, \mathbf{s}_j)$ only depends on the distance $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$ and the longitudes s_{i2} , s_{j2} . Similar to (10) in the proof of Theorem 1, we can also show that $c(\mathbf{s}_i, \mathbf{s}_j)$ is a function of $(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)'(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j)$, s_{i2} and s_{j2} . Then $\rho_{NS}(\mathbf{s}_i, \mathbf{s}_j) = c(\mathbf{s}_i, \mathbf{s}_j) \rho(q(\mathbf{s}_i, \mathbf{s}_j)) := \rho_A(\tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j, s_{i2}, s_{j2})$, so it is axially symmetric. $\square$

REFERENCES

- Alegria A, Cuevas-Pacheco F, Diggle P, Porcu E (2021) The f-family of covariance functions: a matérn analogue for modeling random fields on spheres. *Spat Stat* 43:100512
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. *J R Stat Soc Ser B* 70(4):825–848
- Bissiri PG, Peron AP, Porcu E (2020) Strict positive definiteness under axial symmetry on the sphere. *Stoch Environ Res Risk Assess* 34:723–732
- Blake LR, Porcu E, Hammerling DM (2022) Parametric nonstationary covariance functions on spheres. *Statistics* 11(1):e468
- Castruccio S, Genton MG (2014) Beyond axial symmetry: an improved class of models for global data. *Statistics* 3(1):48–55
- Castruccio S, Genton MG (2016) Compressing an ensemble with statistical models: an algorithm for global 3d spatio-temporal temperature. *Technometrics* 58(3):319–328
- Castruccio S, Stein ML et al (2013) Global space-time models for climate ensembles. *Ann Appl Stat* 7(3):1593–1611
- Du J, Ma C, Li Y (2013) Isotropic variogram matrix functions on spheres. *Math Geosci* 45(3):341–357
- Emery X, Porcu E, Bissiri PG (2019) A semiparametric class of axially symmetric random fields on the sphere. *Stoch Environ Res Risk Assess* 33:1863–1874
- Emery X, Arroyo D, Mery N (2021) Twenty-two families of multivariate covariance kernels on spheres, with their spectral representations and sufficient validity conditions. *Stoch Environ Res Risk Assess* 66:1–21
- Gneiting T (2013) Strictly and non-strictly positive definite functions on spheres. *Bernoulli* 19(4):1327–1349
- Guinness J (2018) Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* 60(4):415–429
- Guinness J, Fuentes M (2016) Isotropic covariance functions on spheres: some properties and modeling considerations. *J Multivar Anal* 143:143–152
- Heaton M, Katzfuss M, Berrett C, Nychka D (2014) Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics* 25(1):2–15
- Huang C, Zhang H, Robeson SM (2011) On the validity of commonly used covariance and variogram functions on the sphere. *Math Geosci* 43(6):721–733
- Jeong J, Jun M (2015a) A class of matérn-like covariance functions for smooth processes on a sphere. *Spat Stat* 11:1–18
- Jeong J, Jun M (2015b) Covariance models on the surface of a sphere: When does it matter? *Statistics* 4(1):167–182
- Jeong J, Jun M, Genton MG (2017) Spherical process models for global spatial statistics. *Stat Sci* 32(4):501
- Jones R (1963) Stochastic processes on a sphere. *Ann Math Stat* 34(1):213–218
- Jun M (2014) Matérn-based nonstationary cross-covariance models for global processes. *J Multivar Anal* 128:134–146
- Jun M, Stein ML (2007) An approach to producing space-time covariance functions on spheres. *Technometrics* 49(4):468–479
- Jun M, Stein ML (2008) Nonstationary covariance models for global data. *Ann Appl Stat* 2(4):1271–1289
- Kang M, Katzfuss M (2021) Correlation-based sparse inverse Cholesky factorization for fast Gaussian-process inference. [arXiv:2112.14591](https://arxiv.org/abs/2112.14591)
- Katzfuss M (2011) Hierarchical spatial and spatio-temporal modeling of massive datasets, with application to global mapping of CO₂. PhD thesis, The Ohio State University
- Katzfuss M, Guinness J (2021) A general framework for Vecchia approximations of Gaussian processes. *Stat Sci* 36(1):124–141
- Katzfuss M, Guinness J, Gong W, Zilber D (2020a) Vecchia approximations of Gaussian-process predictions. *J Agric Biol Environ Stat* 25(3):383–414

- Katzfuss M, Jurek M, Zilber D, Gong W, Guinness J, Zhang J, Schäfer F (2020b) GPvecchia: fast Gaussian-process inference using Vecchia approximations. R package version 0.1.3
- Katzfuss M, Guinness J, Lawrence E (2022) Scaled Vecchia approximation for fast computer-model emulation. *SIAM/ASA J Uncertain Quantif* 10(2):537–554
- Kay JE, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster JM, Bates SC, Danabasoglu G, Edwards J, Holland M, Kushner P, Lamarque JF, Lawrence D, Lindsay K, Middleton A, Munoz E, Neale R, Oleson K, Polvani L, Vertenstein M (2015) The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bull Am Meteorol Soc* 96(8):1333–1349
- Knapp A (2012) Global Bayesian nonstationary spatial modeling for very large datasets. Bachelor Thesis
- Lindgren F, Rue H, Lindström J (2011) An explicit link between gaussian fields and gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Ser B Stat Methodol* 73(4):423–498
- Ma C (2012) Stationary and isotropic vector random fields on spheres. *Math Geosci* 44(6):765–778
- Ma C (2015) Isotropic covariance matrix functions on all spheres. *Math Geosci* 47(6):699–717
- Menegatto V (2020) Positive definite functions on products of metric spaces via generalized stieltjes functions. *Proc Am Math Soc* 148(11):4781–4795
- Paciorek C, Schervish M (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17(5):483–506
- Schäfer F, Katzfuss M, Owahdi H (2021) Sparse Cholesky factorization by Kullback–Leibler minimization. *SIAM J Sci Comput* 43(3):A2019–A2046
- Stein ML (2005) Nonstationary spatial covariance functions. Technical Report No. 21, University of Chicago
- Stein ML et al (2007) Spatial variation of total column ozone on a global scale. *Ann Appl Stat* 1(1):191–210
- Vecchia A (1988) Estimation and model identification for continuous spatial processes. *J R Stat Soc Ser B* 50(2):297–312
- Vihola M (2012) Robust adaptive metropolis algorithm with coerced acceptance rate. *Stat Comput* 22(5):997–1008
- Yaglom A (1987) *Correlation theory of stationary and related random functions*, vol 1. Springer, New York

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.