

# Towards a complete picture of stationary covariance functions on spheres cross time

Philip White<sup>1,\*</sup> and Emilio Porcu<sup>2</sup>

<sup>1</sup>*Brigham Young University, Department of Statistics, Provo, UT, USA*  
e-mail: [pwhite@stat.byu.edu](mailto:pwhite@stat.byu.edu)

<sup>2</sup>*School of Mathematics and Statistics, Newcastle University*  
*Newcastle upon Tyne, NE1 7RU. U.K.*  
e-mail: [emilio.porcu@newcastle.ac.uk](mailto:emilio.porcu@newcastle.ac.uk)

**Abstract:** With the advent of wide-spread global and continental-scale spatiotemporal datasets, increased attention has been given to covariance functions on spheres over time. This paper provides results for stationary covariance functions of random fields defined over  $d$ -dimensional spheres cross time. Specifically, we provide a bridge between the characterization in Berg and Porcu (2017) for covariance functions on spheres cross time and Gneiting’s lemma (Gneiting, 2002) that deals with planar surfaces.

We then prove that there is a valid class of covariance functions similar in form to the Gneiting class of space-time covariance functions (Gneiting, 2002) that replaces the squared Euclidean distance with the great circle distance. Notably, the provided class is shown to be positive definite on every  $d$ -dimensional sphere cross time, while the Gneiting class is positive definite over  $\mathbb{R}^d \times \mathbb{R}$  for fixed  $d$  only.

In this context, we illustrate the value of our adapted Gneiting class by comparing examples from this class to currently established nonseparable covariance classes using out-of-sample predictive criteria. These comparisons are carried out on two climate reanalysis datasets from the National Centers for Environmental Prediction and National Center for Atmospheric Research. For these datasets, we show that examples from our covariance class have better predictive performance than competing models.

**Keywords and phrases:** Bayesian statistics, covariance functions, global data, great circle distance, spatiotemporal statistics, sphere.

Received August 2018.

## 1. Introduction

In recent years, there has been a sharp increase in the prevalence of global or continental-scale spatiotemporal data due to satellite imaging, climate re-analyses, and wide-spread monitoring networks. Although Earth is not exactly spherical (it flattens at the pole), it is commonly believed that the Earth can be well approximated by a sphere (Gneiting, 2013; Castruccio and Stein, 2013). With the goal of modeling data over large spatial scales, while accounting for

---

\*Corresponding Author.

the geometry of the Earth, there has recently been fervent research on modeling and inference for random fields on spheres as well as on spheres cross time. Recent examples provide a comprehensive overview of these topics, including Gneiting (2013), Jeong and Jun (2015), Porcu, Bevilacqua and Genton (2016), Berg and Porcu (2017), and Porcu, Alegría and Furrer (2018).

Under Gaussianity, the covariance function is core to spatiotemporal modeling, inference, and prediction. Covariance functions are positive definite and showing that a candidate function is positive definite over spheres cross time often requires mathematical tools from harmonic analysis. Following the works of Schoenberg (1942) and Gneiting (2013) on spheres, the mathematical characterization of covariance functions on spheres cross time has been given by Berg and Porcu (2017). In addition, Porcu, Bevilacqua and Genton (2016) provide examples of covariance functions for practitioners. As a special case of these covariance classes, some have adapted these classes for temporal models on circles (one-dimensional spheres) to account for seasonal patterns in temporal autocorrelation (see Shirota and Gelfand, 2017; White and Porcu, 2019). Generalizations in the area of mathematical analysis include Guella, Menegatto and Peron (2016a,b, 2017) and Barbosa and Menegatto (2017).

For a random field on  $\mathbb{R}^d \times \mathbb{R}$  with stationary covariance function  $C : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ , Gneiting (2002) showed the following characterization: if  $C$  is continuous, bounded, symmetric, and integrable (over  $\mathbb{R}^d$ ), then  $C$  is a covariance function if and only if the function  $C_\omega : \mathbb{R} \rightarrow \mathbb{R}$ , defined by

$$C_\omega(u) = \int_{\mathbb{R}^d} e^{i\mathbf{h}^\top \omega} C(\mathbf{h}, u) d\mathbf{h}, \quad u \in \mathbb{R}, \quad (1)$$

is a covariance function for almost every  $\omega \in \mathbb{R}^d$ . Here,  $i$  is the unit imaginary number and  $\top$  is the transpose operator. This characterization has been the crux of many important results in spatiotemporal covariance modeling. Examples include the Gneiting class (Gneiting, 2002), Schlather's generalized class (Schlather, 2010), component-wise anisotropic covariances (Porcu, Gregori and Mateu, 2006), multivariate geostatistical modeling (Apanasovich and Genton, 2010), quasi-arithmetic construction (Porcu, Mateu and Christakos, 2010) and nonstationary models (Porcu, Mateu and Bevilacqua, 2007). For a given positive integer  $d$ , (1) proves the validity of the Gneiting class of covariance functions:

$$C(\mathbf{h}; u) = \frac{\sigma^2}{\psi(u^2)^{d/2}} \varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(u^2)}\right), \quad \mathbf{h} \in \mathbb{R}^d, u \in \mathbb{R}, \quad (2)$$

where  $\|\cdot\|$  is the Euclidean norm. The function  $\varphi : [0, \infty) \rightarrow \mathbb{R}_+$  is completely monotonic; that is,  $\varphi$  is infinitely differentiable on  $(0, \infty)$ , satisfying  $(-1)^n \varphi^{(n)}(t) \geq 0$ ,  $n \in \mathbb{N}$ . Here,  $\varphi^{(n)}$  denotes  $n$ th derivative and we use  $\varphi^{(0)}$  for  $\varphi$ , where  $\varphi(0)$  is required to be finite. The function  $\psi : [0, \infty) \rightarrow \mathbb{R}_+$  is strictly positive with a completely monotonic derivative. Here and throughout,  $\sigma^2$  is used to represent the spatiotemporal variance; that is, a scaling factor of a spatiotemporal correlation function. We also note that the function  $C$  in (2) is positive definite in  $\mathbb{R}^d \times \mathbb{R}$  for a given positive integer  $d$ , but it is not positive definite on every  $d$ -dimensional Euclidean space cross time.

Our paper focuses on two aspects of covariance modeling on  $d$ -dimensional spheres cross time. We first focus on Criterion (1) and its analogue on spheres over time. Our result provides an additional equivalence condition to those provided in Berg and Porcu (2017). We then provide an adaptation of the Gneiting class (2) to spherical domains, and show that it is positive definite over all  $d$ -dimensional spheres (including the Hilbert sphere) cross time. Further, our proof is based on direct construction, allows us to avoid Fourier inversion, and does not require a convergence argument that was originally used by Gneiting (2002). Porcu, Bevilacqua and Genton (2016) considered a variant of this problem, modifying the Gneiting class based on temporal rescaling of the spatial component. This idea was also suggested by Gneiting (2002). In addition to a new Gneiting class for spheres over time, we adapt Heine's class of covariance functions (Heine, 1955), originally proposed over two-dimensional Euclidean spaces, to  $d$ -dimensional spheres cross time.

For estimation and prediction with the new covariance class, (5), we take a Bayesian approach using nearest neighbor Gaussian processes (NNGP) (Datta et al., 2016a,b). Bayesian models allow for simple and rigorous uncertainty quantification through a single probabilistic framework that does not rely on asymptotic assumptions. Because Gaussian process (GP) models for large datasets, as we have with globally sampled spatiotemporal data, are often computationally intractable, we use the NNGP as a surrogate. Modeling with the NNGP enables scalable model fitting, inference, and prediction for real-data examples. Our discussion here adds to application areas for NNGPs as they have not been used for global data in the literature.

For our data examples, we use daily near-surface temperature and cloud coverage from the first week of 2017 (Kalnay et al., 1996). We only use the first week to keep computation times short. To be clear, we do not claim that covariance functions from our new covariance classes are preferable for all datasets. Indeed, for some datasets that we tested, but that we do not present here, the new covariance functions in this manuscript showed little or no predictive advantage. However, we highlight these datasets because they show that our new Gneiting class yields practical predictive benefits in some cases.

We start by giving background (Section 2) for the theoretical results given in Section 3. In Section 3, we provide the analogue of (1) for covariance functions over spheres cross time. Additionally, we adapt the Gneiting class (2) to spheres cross time and show that, using a subclass of completely monotonic functions, the adapted Gneiting class can be used on all  $d$ -dimensional spheres cross time. Then, we provide an adaptation of the Heine covariance function, originally proposed in  $\mathbb{R} \times \mathbb{R}$ , to spheres cross time. Proofs of the theoretical results are technical and are deferred to the Appendix A. We also provide a supplementary result in Appendix A related to our main result in Section 3. We then turn our attention to modeling data using covariance functions from our adapted Gneiting class in Section 4. In Section 5, we draw upon our modeling discussion for simulation studies and real data analyses. In our simulation studies, we explore parameter identifiability for examples from our adapted Gneiting covariance class and highlight some limitations. In our data examples, covariance functions

from our adapted Gneiting covariance class have better out-of-sample predictive performance than covariance models currently in the literature, using mean absolute error, mean squared error, and continuous ranked probability scores as model comparison criteria. Finally, we provide concluding remarks in Section 6.

## 2. Background

Let  $d$  be a positive integer. Here, we consider stationary Gaussian random fields on  $d$ -dimensional unit spheres  $\mathbb{S}^d$  cross time (in  $\mathbb{R}$ ), where  $\mathbb{S}^d$  is defined to be  $\{\mathbf{s} \in \mathbb{R}^{d+1} : \|\mathbf{s}\| = 1\}$ . We use the unit sphere without loss of generality. These random fields are denoted  $\{Y(\mathbf{s}, t), \mathbf{s} \in \mathbb{S}^d, t \in \mathbb{R}\}$ . We assume Gaussianity in modeling (Section 4) which implies that finite dimensional distributions are completely specified by the mean and covariance function of the random field.

As a metric on  $\mathbb{S}^d$ , we use the great circle distance  $\theta : \mathbb{S}^d \times \mathbb{S}^d \rightarrow [0, \pi]$ , defined as the mapping

$$(\mathbf{s}_1, \mathbf{s}_2) \mapsto \arccos(\mathbf{s}_1^\top \mathbf{s}_2), \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{S}^d.$$

We then consider covariance functions based on  $\theta(\mathbf{s}_1, \mathbf{s}_2)$  and time difference  $u = |t_1 - t_2|$ ,

$$\text{cov}(Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2)) = C(\theta(\mathbf{s}_1, \mathbf{s}_2), |t_1 - t_2|), \quad (\mathbf{s}_i, t_i) \in \mathbb{S}^d \times \mathbb{R}, \quad (3)$$

where we take  $\theta$  as an abbreviation for  $\theta(\mathbf{s}_1, \mathbf{s}_2)$ . Porcu, Alegria and Furrer (2018) refer to such covariance functions as spatially geodesically isotropic and temporally symmetric, and Berg and Porcu (2017) provide a mathematical characterization for these functions.

Covariance functions are positive definite, meaning that for any collection  $\{(\mathbf{s}_k, t_k)\}_{k=1}^N \subset \mathbb{S}^d \times \mathbb{R}$  and constants  $\{a_k\}_{k=1}^N, \sum_{k,h} a_k a_h C(\theta(\mathbf{s}_k, \mathbf{s}_h), |t_k - t_h|) \geq 0$ . It is worth noting that classes of positive definite functions on  $\mathbb{S}^d \times \mathbb{R}$  are nested, meaning that positive definiteness on  $\mathbb{S}^d \times \mathbb{R}$  implies positive definiteness on  $\mathbb{S}^{d'} \times \mathbb{R}$  for  $d' < d$ , but the converse is not necessarily true.

Porcu, Bevilacqua and Genton (2016) proposed the *inverted* Gneiting class and define it as

$$C(\theta; u) = \frac{\sigma^2}{\psi_{[0,\pi]}(\theta)^{1/2}} \varphi\left(\frac{u^2}{\psi_{[0,\pi]}(\theta)}\right), \quad \theta \in [0, \pi], u \in \mathbb{R}, \quad (4)$$

with  $\varphi$  and  $\psi$  as defined in (2), and where  $\psi_{[0,\pi]}$  denotes the restriction of  $\psi$  to the interval  $[0, \pi]$ . In contrast to (2) which scales Euclidean distance by a function of the temporal lag, (4) rescales the temporal lag by a function of the great circle distance. This was also mentioned in Gneiting (2002). It might be more intuitive to rescale space with time, as was done in (2), proposing a structure like

$$C(\theta; u) = \frac{\sigma^2}{\psi(u^2)} \varphi\left(\frac{\theta}{\psi(u^2)}\right), \quad \theta \in [0, \pi], u \in \mathbb{R}, \quad (5)$$

where, in this case, we do not need to restrict any of the functions  $\varphi$  and  $\psi$  to the interval  $[0, \pi]$ . Also, one might note that the function  $\psi$  is not raised to the power  $d/2$  as in (2). Showing this construction is valid is nontrivial and receives an exposition in Section 3.2.

One choice of  $\varphi$ , used to construct covariance functions in (2) and (4) is the Matérn class  $\varphi(t) = \mathcal{M}_{\alpha, \nu}(t)$ ,  $t \geq 0$ ,  $\alpha, \nu > 0$ , defined as

$$\mathcal{M}_{\alpha, \nu}(t) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{t}{\alpha}\right)^\nu \mathcal{K}_\nu\left(\frac{t}{\alpha}\right), \quad (6)$$

where  $\mathcal{K}_\nu$  is the MacDonald function (Gradshteyn and Ryzhik, 2007). One appeal of this class is the parameter  $\nu$  that governs the smoothness at the origin (Stein, 1999). Unfortunately,  $\mathcal{M}_{\alpha, \nu}(\theta)$  is not positive definite on  $d$ -dimensional spheres, unless  $\nu \in (0, 1/2]$  (Gneiting, 2013), which makes this function less appealing to model spatial processes that are sufficiently smooth.

### 3. Theoretical results

#### 3.1. The generalized Gneiting lemma on spheres cross time

We begin our discussion with a criterion for covariance functions defined over  $d$ -dimensional spheres cross time. Let  $\mathcal{G}_k^\lambda$  be the  $k$ th normalized Gegenbauer polynomial of order  $\lambda > 0$  (Dai and Xu, 2013). Gegenbauer polynomials form an orthonormal basis for the space of square-integrable functions, denoted as  $\mathcal{L}^2([0, \pi], \sin \theta^{d-1} d\theta)$ .

**Theorem 1.** *Let  $d$  be a positive integer. Let  $C : [0, \pi] \times \mathbb{R} \rightarrow \mathbb{R}$  be continuous and bounded with the  $k$ th related Gegenbauer transform, defined as*

$$b_{k,d}(u) = \int_0^\pi C(\theta, u) \mathcal{G}_k^{(d-1)/2}(\cos \theta) \sin \theta^{d-1} d\theta, \quad u \in \mathbb{R}, k = 0, 1, \dots, \quad (7)$$

with  $b_{k,d} : \mathbb{R} \rightarrow \mathbb{R}$ , satisfying  $\sum_{k=0}^\infty \int_{\mathbb{R}} |b_{k,d}(u)| du < \infty$ . Then, the following assertions are equivalent:

1.  $C(\theta, u)$  is the covariance function of a random field on  $\mathbb{S}^d \times \mathbb{R}$ ;
2. the function  $C_\tau : [0, \pi] \rightarrow \mathbb{R}$ , defined as

$$C_\tau(\theta) = \int_{-\infty}^{+\infty} e^{-iu\tau} C(\theta, u) du, \quad (8)$$

- is the covariance function of a random field on  $\mathbb{S}^d$  for almost every  $\tau \in \mathbb{R}$ ;
3. for all  $k = 0, 1, 2, \dots$ , the functions  $b_{k,d} : \mathbb{R} \rightarrow \mathbb{R}$ , defined through (7), are continuous, positive definite on  $\mathbb{R}$ , and  $\sum_k b_{k,d}(0) < \infty$ .

Some comments are in order. Equivalence of 1 and 3 was shown by Berg and Porcu (2017). The result completes the picture that had been started by Gneiting (2013), Berg and Porcu (2017) and Porcu, Bevilacqua and Genton

(2016). Equivalence of 1 and 2 provides the analogue of Gneiting's criterion in (1) for spheres cross time. Thus, Theorem 1 gives insight into relationships between covariance functions on spheres and covariance functions on Euclidean spaces. In fact, application of Theorem 1 provides a useful criterion (see Appendix A) relating spatiotemporal covariances on  $\mathbb{R} \times \mathbb{R}$  with covariance functions on  $\mathbb{S}^3 \times \mathbb{R}$ .

The proof of Theorem 1 is technical, and we defer it to Appendix A to avoid mathematical obfuscation.

### 3.2. New classes of covariance functions on spheres cross time

We now detail our findings with two new classes of covariance functions on spheres over time. To do this, we need to introduce a new class of special functions. A function  $\varphi : [0, \infty) \rightarrow \mathbb{R}$  is called a Stieltjes function if

$$\varphi(t) = \int_{[0, \infty)} \frac{\mu(d\xi)}{t + \xi}, \quad t \geq 0, \quad (9)$$

where  $\mu$  is a positive and bounded measure. We require throughout  $\varphi(0) = 1$ , which implies that  $\int \xi^{-1} \mu(d\xi) = 1$ . Let us call  $\mathcal{S}$  the set of Stieltjes functions. It has been proved that  $\mathcal{S}$  is a convex cone (Berg, 2008), with the inclusion relation  $\mathcal{S} \subset \mathcal{C}$ , where  $\mathcal{C}$  is the set of completely monotone functions. The relation (9) shows that the function  $t \mapsto (1+t)^{-1}$ ,  $t \geq 0$ , is a Stieltjes function. Using the fact that  $\varphi \in \mathcal{S}$  if and only if  $1/\varphi$  is a completely Bernstein function (for a definition, see Porcu and Schilling, 2011) we can get a wealth of examples, as the book by Schilling, Song and Vondracek (2012) provides an entire catalogue of completely Bernstein functions. We finally note that completely Bernstein functions are infinitely differentiable over  $(0, \infty)$  and have a completely monotonic derivative.

We are now able to state the following result.

**Theorem 2.** *Let  $C : [0, \pi] \times \mathbb{R}$  be the mapping defined through (5), where  $\varphi$  is a Stieltjes function on the positive real line, with  $\varphi(0) = 1$ , and  $\psi$  is strictly positive with a completely monotone derivative. Then,  $C$  is a covariance function on  $\mathbb{S}^d \times \mathbb{R}$  for all positive integers  $d$ .*

Again, the proof is deferred to Appendix A. This result completes the adaptation of the Gneiting class (Gneiting, 2002) to spheres cross time. Theorem 2 allows  $C$  in (5) to be positive definite on every  $d$ -dimensional sphere under the condition that the function  $\varphi$  is a Stieltjes function. As already mentioned, the class is rich, and there is a whole catalogue available from the book by Schilling, Song and Vondracek (2012). In addition, our proof in Appendix A does not require any Fourier inversion techniques, nor convergence arguments as those used in Gneiting (2002).

We also note that the Matérn function  $\mathcal{M}_{\alpha, \nu}$  cannot be used for the purposes of Theorem 2. This is not surprising, as arguments in Gneiting (2013) show that the Matérn covariance function in (6) can only be used in Theorem 2 for  $0 < \nu \leq 1/2$ . If one is interested in smoother realizations over spheres, then a common method involves using the Euclidean distance on spheres (Gneiting,

2013; Porcu, Alegria and Furrer, 2018), also called chordal distance, in (2). In this case, any choice for  $\nu > 0$  preserves positive definiteness. At the same time, using chordal distance has a collection of drawbacks that have inspired constructive criticism in Banerjee (2005), Gneiting (2013), Porcu, Alegria and Furrer (2018) and Alegria and Porcu (2017), to cite a few. We explore both possibilities and compare them in terms of predictive performance in Section 5.

To introduce another class of covariance functions, we define the complementary error function  $\operatorname{erfc}$  as

$$\operatorname{erfc}(t) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-\xi^2) d\xi, \quad t \geq 0,$$

and  $\operatorname{erfc}(t) = 2 - \operatorname{erfc}(-t)$  when  $t$  is negative. We can show the following result.

**Theorem 3.** *Let  $\psi_{[0,\pi]}$  be the restriction to  $[0, \pi]$  of a positive function with a completely monotonic derivative. Then,*

$$C(\theta, u) = e^{-|u|} \operatorname{erfc} \left( \sqrt{\psi_{[0,\pi]}(\theta)} - \frac{|u|}{2\sqrt{\psi_{[0,\pi]}(\theta)}} \right) + e^{|u|} \operatorname{erfc} \left( \sqrt{\psi_{[0,\pi]}(\theta)} + \frac{|u|}{2\sqrt{\psi_{[0,\pi]}(\theta)}} \right), \quad (10)$$

$\theta \in [0, \pi]$ ,  $u \in \mathbb{R}$ , is a covariance function on  $\mathbb{S}^d \times \mathbb{R}$  for all  $d = 1, 2, \dots$

The class presented in (10) is related to a covariance class on  $\mathbb{R} \times \mathbb{R}$  considered by Heine (1955). Again, the proof is provided in Appendix A.

## 4. Modeling nonseparable spatiotemporal data

### 4.1. Hierarchical process modeling for spatiotemporal data

We illustrate the utility of one of our covariance classes (Theorem 2) using hierarchical NNGP models in a Bayesian setting. In spatial and spatiotemporal analyses, Bayesian models are often preferred for hierarchical modeling because they allow for simple and rigorous uncertainty quantification through a single probabilistic framework that does not rely on asymptotic assumptions (see, e.g., Gelman et al., 2014; Banerjee, Carlin and Gelfand, 2014; Cressie and Wikle, 2015).

Spatiotemporal random effects for point-referenced data are often specified through a functional prior using a Gaussian process (GP). Gaussian processes are natural choices for modeling data that vary in space and time. However, likelihood computations for hierarchical GP models require inverting a square matrix with dimension equal to the size of the data, making GP models intractable in “big-data” settings. Many have addressed this computational bottleneck using either low-rank or sparse matrix methods (see Heaton et al., 2018, for a review and comparison of some of these methods).

Low-rank methods depend on selecting representative points, often called knots, that are used to approximate the original process (see, e.g., Higdon, 2002; Banerjee et al., 2008; Cressie and Johannesson, 2008; Stein, 2008). These models tend to oversmooth and often have poor predictive performance (see Stein, 2014; Heaton et al., 2018).

In contrast to low-rank methods, inducing sparsity in either the covariance matrix or the precision matrix can reduce the computational burden. Covariance tapering creates sparsity in the covariance matrix by using compactly supported covariance functions (see, e.g., Furrer, Genton and Nychka, 2006; Kaufman, Schervish and Nychka, 2008). These methods are generally effective for parameter estimation and interpolation; however, the allowable class of covariance functions is limited. On the other hand, inducing sparsity in the precision matrix has been leveraged to approximate GPs using Markov random fields (Lindgren, Rue and Lindstroem, 2011) or using conditional likelihoods (Vecchia, 1988; Stein, Chi and Welty, 2004). These approaches were extended to process modeling by Gramacy and Apley (2015) and Datta et al. (2016a). For discussion and further extension of these approaches, see Katzfuss and Guinness (2017). Unlike local approximate Gaussian processes (Gramacy and Apley, 2015), the NNGP is itself a GP (Datta et al., 2016a) and has good predictive performance relative to other “fast” GP methods (See Heaton et al., 2018).

To specify an NNGP, we begin with a parent GP over  $\mathbb{R}^d \times \mathbb{R}$  or  $\mathbb{S}^d \times \mathbb{R}$ . Nearest neighbor Gaussian process models induce sparsity in the precision matrix of the parent Gaussian model by assuming conditional independence given neighborhood sets constructed from directed acyclic graphs, yielding huge computational benefits (Datta et al., 2016a,b). Modeling, model fitting, and prediction details for NNGP models are given in Appendix B.

#### 4.2. Examples of covariance functions

Here, we turn our attention to six nonseparable covariance functions used in simulation studies in Section 5.1 in our data analyses (Sections 5.2 and 5.3). For all examples, we parameterize the models with variance  $\sigma^2$  and use  $c_s$  and  $c_t$  to represent the strictly positive spatial and temporal scale parameters, respectively.

Explicitly, we consider two special cases from the Gneiting class (2) with  $\varphi(t) = \mathcal{M}_{c_s, \nu}(t)$ ,  $t \geq 0$ , for  $\mathcal{M}_{c_s, \nu}$ , defined in (6), obtained when  $\nu = 1/2$  and  $\nu = 3/2$ ,

$$C(\mathbf{h}, u) = \frac{\sigma^2}{\left(1 + \left(\frac{\|\mathbf{u}\|}{c_t}\right)^\alpha\right)^{\delta + \beta d/2}} \mathcal{M}_{c_s, \nu} \left( \frac{\|\mathbf{h}\|}{\left(1 + \left(\frac{\|\mathbf{u}\|}{c_t}\right)^\alpha\right)^{\beta/2}} \right), \quad (11)$$

where  $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$ . These choices correspond to  $\mathcal{M}_{c_s, 1/2}(t) = \exp(-t/c_s)$  and  $\mathcal{M}_{c_s, 3/2}(t) = \exp(-t/c_s)(1 + t/c_s)$ ,  $t \geq 0$ . Here, we work on the sphere, thus  $\|\cdot\|$  refers to chordal distance. The parameters restriction is  $\delta > 0$ ,  $\beta \in (0, 1]$  and  $\alpha \in (0, 2]$ .



For the remaining examples,  $(\theta, u) \in [0, \pi] \times \mathbb{R}$ . Next, we consider a pair of similar covariance models from the inverted Gneiting class (Porcu, Bevilacqua and Genton, 2016) in (4):

$$C(\theta, u) = \frac{\sigma^2}{\left(1 + \left(\frac{\theta}{c_s}\right)^\alpha\right)^{\delta + \beta/2}} \exp\left(-\frac{|u|^{2\gamma}}{c_t^{2\gamma} \left(1 + \left(\frac{\theta}{c_s}\right)^\alpha\right)^{\beta\gamma}}\right), \quad (12)$$

where  $\delta > 0$ , and where  $\beta$ ,  $\alpha$  and  $\gamma$  belong to the interval  $(0, 1]$ . The second we consider uses the generalized Cauchy covariance function (Gneiting and Schlather, 2004) for the temporal margin, that is  $\psi(u) = (1 + (|u|/c_s)^\gamma)^{-\lambda}$ :

$$C(\theta, u) = \frac{\sigma^2}{\left(1 + \left(\frac{\theta}{c_s}\right)^\alpha\right)^{\delta + \beta/2}} \left(1 + \frac{|u|^{2\gamma}}{c_t^{2\gamma} \left(1 + \left(\frac{\theta}{c_s}\right)^\alpha\right)^{\beta\gamma}}\right)^{-\lambda}, \quad (13)$$

with  $\delta, \lambda > 0$ , and where  $\beta$ ,  $\alpha$  and  $\gamma$  belong to the interval  $(0, 1]$ .

As a first example from our new adapted Gneiting class on spheres cross time (see Theorem 2), we chose the Stieltjes function

$$\varphi(t) = \kappa \frac{1 - e^{-2\sqrt{t+1}}}{\sqrt{t+1}}, \quad t \geq 0,$$

with  $\kappa := 1/(1 - e^{-2})$  a normalization constant. We then pick the function  $\psi(t) = (1 + t^\alpha)^\delta$ , that is a Bernstein function for  $\alpha, \delta \in (0, 1]$ . Thus, we have

$$C(\theta, u) = \frac{\sigma^2 \kappa}{\left(1 + \left(\frac{|u|}{c_t}\right)^\alpha\right)^\delta} \frac{1 - e^{-2 \left(1 + \frac{\theta}{c_s \left(1 + \left(\frac{|u|}{c_t}\right)^\alpha\right)^\delta}\right)^{1/2}}}{\left(1 + \frac{\theta}{c_s \left(1 + \left(\frac{|u|}{c_t}\right)^\alpha\right)^\delta}\right)^{1/2}}. \quad (14)$$

For the second, we again propose a generalized Cauchy covariance function for the spatial margin, obtaining

$$C(\theta, u) = \frac{\sigma^2}{\left(1 + \left(\frac{|u|}{c_t}\right)^\alpha\right)^{\delta + \beta/2}} \left(1 + \frac{\theta^\gamma}{c_s^\gamma \left(1 + \left(\frac{|u|}{c_t}\right)^\alpha\right)^{\beta\gamma}}\right)^{-\lambda}, \quad (15)$$

where  $\delta > 0$ ,  $\beta, \gamma \in (0, 1]$ ,  $\alpha \in (0, 2]$  and  $\lambda > 0$ .

For all models in our simulation study and our data analyses, we include an independent Gaussian error term with variance  $\tau^2$  in the model. The variance  $\tau^2$  is often called a nugget and accounts for potential discontinuities at the origin of the covariance function. In other words, the nugget accounts for sources of uncertainty that are not explained or captured by our spatiotemporal model.

### 4.3. Model comparison

To compare models that differ in terms of covariance specification, we propose the following criteria for comparing predictions to hold-out data  $y_i$ : 90% predictive interval coverage, predictive mean square or absolute error (defined as  $(\mathbf{E}(Y_i | \mathbf{Y}_{obs}) - y_i)^2$  or  $|\mathbf{E}(Y_i | \mathbf{Y}_{obs}) - y_i|$ , respectively), where  $\mathbf{Y}_{obs}$  denotes observations. Besides these common criteria, we also use a strictly proper scoring rule (Gneiting and Raftery, 2007), the continuous ranked probability score (CRPS), defined as

$$\text{CRPS}(F_i, y_i) = \int_{-\infty}^{\infty} (F_i(x) - \mathbf{1}(x \geq y_i))^2 dx = \mathbf{E} | Y_i - y_i | - \frac{1}{2} \mathbf{E} | Y_i - Y'_i |, \quad (16)$$

where  $Y_i$  and  $Y'_i$  follow the predictive distribution  $F_i$  (see Brown, 1974; Matheson and Winkler, 1976, for early discussion on CRPS).. An empirical estimate of the continuous ranked probability score, using  $M$  posterior predictive samples  $Y_{i,1}, \dots, Y_{i,M}$  from  $Y_i | \mathbf{Y}_{obs}$ , is

$$\text{CRPS}(\hat{F}_i, y_i) = \frac{1}{M} \sum_{j=1}^M | Y_{i,j} - y_i | - \frac{1}{2M^2} \sum_{j=1}^M \sum_{k=1}^M | Y_{i,j} - Y_{i,k} |. \quad (17)$$

We average continuous ranked probability scores over all hold-out data to obtain a single value for comparison.

## 5. Simulation and data examples

Using only covariance functions (14) and (15), we provide a brief simulation study in Section 5.1 to explore the identifiability of covariance model parameters using an NNGP model. In Sections 5.2 and 5.3, we illustrate practical predictive advantages of the new Gneiting class using spherical distance (Theorem 2) using two climate reanalysis datasets from the National Centers for Environmental Prediction and National Center for Atmospheric Research (Kalnay et al., 1996). For both analyses, we use the first week of the 2017 dataset.

### 5.1. Simulation study

Here, we present simulation studies to explore the empirical identifiability of covariance model parameters. To do this, we simulate many datasets that differ in their covariance specification, using either (14) and (15). For each covariance function, we fix parameters and generate 1,000 datasets from the following generative model:

$$\begin{aligned} Y(\mathbf{s}, t) &= w(\mathbf{s}, t) + \epsilon(\mathbf{s}, t), \\ w(\mathbf{s}, t) &\sim \text{GP}(0, C((\mathbf{s}, t), (\mathbf{s}', t'))), \\ \epsilon(\mathbf{s}, t) &\sim \text{GP}(0, \tau^2 \delta_{\mathbf{s}}^s \delta_t^t), \end{aligned} \quad (18)$$

where observations lie on an evenly spaced grid of latitudes ranging from  $-90$  to  $-60$  ( $5^\circ$  spacing) and longitudes between  $-180$  and  $0$  ( $5^\circ$  spacing). This grid is repeated from times 1 to 10, giving  $N = 4330$ . While we use a full GP specification for simulation and the dataset is not particularly large when fitting a single model, we fit these data using a hierarchical NNGP model with  $m = 25$  neighbors because this mirrors modeling approach in Section 4 and because we fit 1,000 simulated datasets per simulation (6,000 in total). Neighbors are selected using simple rectangular neighborhood sets using great-circle (spherical) distance to define nearness (see Datta et al., 2016b).

When simulating from (18) using (14) as the covariance function, we use with  $\sigma^2\kappa = 4$ ,  $c_s = 0.2$ ,  $c_t = 2$ ,  $\alpha = 1/2$ ,  $\delta = 1/2$ , and  $\tau^2 = 1$  for the noise term  $\epsilon(\mathbf{s}, t)$ . For  $\tau^2$  and  $\sigma^2$ , we use inverse-gamma priors with 0.1 as both the shape and scale (corresponding to the rate of a gamma distribution) parameters. We assume  $c_s \sim \text{Unif}(0, \pi)$ ,  $c_t \sim \text{Unif}(0, 10)$ ,  $\alpha \sim \text{Unif}(0, 1]$ , and  $\delta \sim \text{Unif}(0, 1]$ , *a priori*.

When using (15) in the generative model (18), we set  $\sigma^2 = 4$ ,  $c_s = 0.2$ ,  $c_t = 2$ ,  $\alpha = 1$ ,  $\beta = 1/2$ ,  $\delta = 3/4$ ,  $\lambda = 1$ ,  $\gamma = 1/2$ , and  $\tau^2 = 1$  for the noise term  $\epsilon(\mathbf{s}, t)$ . For simplicity, we constrain  $\delta + \beta/2 = 1$ ,  $\gamma = 1/2$ , and  $\lambda = 1$ . As before, we use inverse-gamma priors with 0.1 as both the shape and scale parameters for  $\tau^2$  and  $\sigma^2$ . As before, we assume  $c_s \sim \text{Unif}(0, \pi)$ ,  $c_t \sim \text{Unif}(0, 10)$ ,  $\alpha \sim \text{Unif}(0, 2]$ ,  $\beta \sim \text{Unif}(0, 1]$ , and  $\delta \sim \text{Unif}(0, 1]$ , *a priori*.

We explore the effect of fixing some model parameters to examine how model identifiability is affected. Specifically, we consider fixing combinations of  $c_s$ ,  $c_t$ , and  $\alpha$  in (14) and (15) to the true value. In this setting, we re-fit the models described above, keeping all other specifications the same as described. We present the 90% empirical coverage rates for all parameters in Table 1. In this table, dashes indicate that parameters are fixed. For (14), we see improved coverage rates (i.e., closer to 90%) for  $\sigma^2$  and  $\alpha$  when range parameters  $c_s$  and  $c_t$  are fixed; however,  $\delta$  shows significant under coverage when range parameters are fixed. The results for (15) are similar. When range parameters are fixed, coverage rates for  $\sigma^2$  and  $\beta$  are closer to 90%. For this covariance model,  $\alpha$  has slightly worse coverage when range parameters are fixed.

TABLE 1  
90% empirical coverage rates. Here, we use dashes when parameters are fixed on the parameters used in simulating the data.

Model	$\sigma^2$	$\tau^2$	$c_s$	$c_t$	$\alpha$	$\delta$ or $\beta$
(14)	0.72	0.87	0.71	0.84	0.74	0.81
(14)	0.79	0.87	—	—	0.82	0.62
(14)	0.80	0.89	—	—	—	0.55
(15)	0.73	0.88	0.67	0.89	0.75	0.99
(15)	0.82	0.89	—	—	0.69	0.92
(15)	0.82	0.90	—	—	—	0.92

These simulation studies highlight some of the limitations in estimating parameters of covariance functions from Theorem 2. For both (14) and (15), we note that there is limited parameter identifiability, particularly for spatial range

parameter  $c_s$  and spatiotemporal variance  $\sigma^2$ . Although we do not provide identifiability proofs, this result seems similar to the limited identifiability of the Matérn class that is identifiable up to  $\sigma^2 c_s^{-2\nu}$  (see Zhang, 2004). Apparently, this issue becomes even more complex under the space-time setting, and the lack of theoretical results for space-time asymptotics and equivalence of Gaussian measures make this problem very difficult. In this simulation, we find improved parameter identifiability when we fix spatial range parameters. We emphasize that the primary goal of our analyses is comparing predictive performance. However, if the unbiased estimation of covariance parameters is the primary goal, then a multi-stage fitting process can improve estimation (Mardia and Marshall, 1984).

## 5.2. Surface air temperature reanalysis data

For this section, we utilize the 2017 National Centers for Environmental Prediction/National Center for Atmospheric Research daily average 0.995 sigma level (near-surface) temperature reanalysis data (Kalnay et al., 1996). Air temperature at 0.995 sigma level is defined to be the temperature taken at an air pressure  $0.995 \times$  surface air pressure.

The foundations of global temperature change are well established (see, e.g., Folland et al., 2001; Hansen et al., 2006, 2010). Furthermore, air temperature changes have, along with other changes in climate, a wide and deep impact on global biological systems (see Parmesan and Yohe, 2003; Thomas et al., 2004; Held and Soden, 2006). For these reasons, many climate models are dedicated to understanding past and predicting future temperature changes (see Simmons et al., 2004, for some discussion and comparisons about the various analyses of surface air temperature).

The daily near-surface temperature reanalysis data represent daily temperature averages over a global grid with  $2.5^\circ$  spacing for latitude and longitude. We thin the data to  $5^\circ$  spacing for latitude and longitude to carry out a model comparison on the hold-out data. In this dataset, we have observations at 2522 unique spatial locations, giving 17654 total observations. The averages of near-surface temperature over the first week of January are plotted in Figure 1. Additionally, we give the density estimate of near-surface temperature for each day in Figure 1. Figure 1 shows that the overall temperature distribution is similar across days and demonstrates a clear spatial structure. Because our covariance model allows space to be scaled by time while using the spherical distance, we expect that our model may be able to capture the strong spatial structure in this data more effectively than the models with which we compare them.

With 17654 data, carrying out fully Bayesian inference using a full GP model is computationally burdensome; thus, we utilize a NNGP model. For these models, we use simple neighborhood selection presented in Datta et al. (2016b) using  $m = 25$  neighbors, using the five nearest neighbors at the five most recent times, including the current time. We utilize two covariance models from each of the

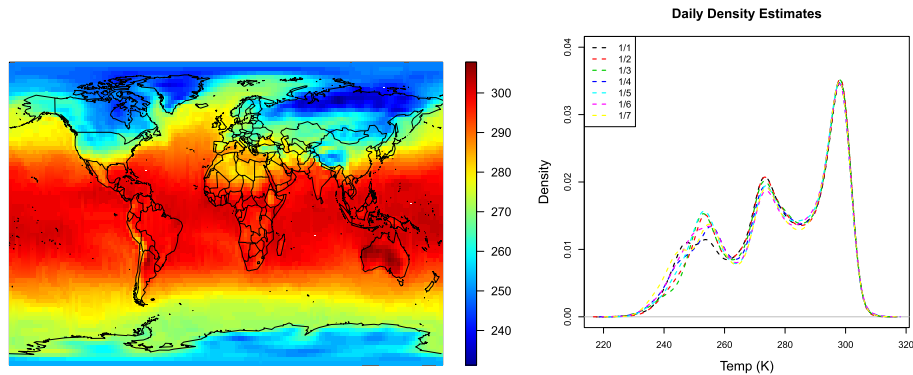


FIG 1. (Left) Near-surface temperature heat map in degrees Kelvin (K) (Right) Kernel density estimates of global near-surface temperature for each day.

Gneiting class, the inverted-Gneiting class using spherical distance, and our new Gneiting class (Theorem 2). These models are presented in (11) to (15). For all models, we use inverse-gamma prior distributions for  $\tau^2$  and  $\sigma^2$  with 0.1 for both the shape and scale parameters (corresponding to the rate parameter of a gamma distribution). We use prior distributions of  $c_s \sim \text{Unif}(0, \pi)$ ,  $c_t \sim \text{Unif}(0, 10)$ ,  $\alpha \sim \text{Unif}(0, 2]$ , and  $\beta \sim \text{Unif}(0, 1]$  for correlation function parameters. Because many covariance models have limited parameter identifiability, we constrain  $\delta + \beta d/2 = 1$  for (11), and  $\delta + \beta/2 = 1$  for (12), (13), and (15).

We compare these six models in terms of predictive performance on a randomly selected subset of the hold-out data. In total, we use 1000 locations over the week, giving 7000 hold-out observations. These hold-out locations are plotted in Figure 5 in Appendix C. We compare these models in terms of predictive root mean squared error, mean absolute error, continuous ranked probability score, and 90% prediction interval coverage, as discussed in Section 4.3. For predictions, neighbors are chosen to make prospective predictions using 25 neighbors (See Appendix B for details on modeling and prediction).

The results presented are based on 25,000 posterior draws after a burn-in of 5,000 iterations using a Gibbs sampler presented in Appendix B. These posterior samples are used for prediction and posterior inference. The results of the model comparison are given in Table 2. For this data, the covariance models from our class (14) and (15) had the best out-of-sample predictive performance, and the model (14) was the very best. For comparison, the Gneiting class using chordal distance had continuous ranked probability scores 7% and 16% higher than the best model for  $\nu = 1/2$  and  $\nu = 3/2$ , respectively. Both models from the inverted Gneiting class with spherical distance were 13% worse in terms of continuous ranked probability scores.

TABLE 2

Predictive performance of competing covariance models for the cloud cover dataset. For brevity in the table, let *PRMSE* and *PMAE* denote predictive mean squared error and mean absolute error, respectively. Relative CRPS is scaled such that the lowest is one. Bolded entries are used to indicate best model performances, i.e. lowest *PRMSE*, *PMAE*, and *CRPS* and 90% interval coverage closest to 90%.

Equation	PRMSE	PMAE	90% Coverage	CRPS	Relative CRPS
(11) and $\nu = 1/2$	6.44	4.58	0.91	3.50	1.07
(11) and $\nu = 3/2$	7.15	5.08	0.91	3.81	1.16
(12)	6.79	4.85	<b>0.90</b>	3.70	1.13
(13)	6.78	4.84	<b>0.90</b>	3.69	1.13
(14)	<b>6.02</b>	<b>4.26</b>	<b>0.90</b>	<b>3.28</b>	<b>1.00</b>
(15)	<b>6.04</b>	<b>4.40</b>	<b>0.90</b>	<b>3.35</b>	<b>1.02</b>

For the best model, (14), we provide posterior summaries in Table 3. Additionally, we display the correlation contour as function of spherical distance  $\theta$  and time  $t$  for the posterior mean in Figure 2. Correlation is very persistent as a function of time; thus, decreases in autocorrelation are almost completely attributable to changes in spatial location.

TABLE 3

Posterior summaries for the near-surface air temperature dataset for parameters for the model fit using (14). Percentiles (2.5% and 97.5%) represent posterior percentiles.

	Mean	Standard Deviation	2.5%	97.5%
$\tau^2$	21.140	0.553	20.088	22.250
$\sigma^2$	105.897	4.253	97.953	114.498
$c_s$	0.994	0.024	0.952	1.025
$c_t$	2.783	3.243	0.026	9.624
$\alpha$	0.017	0.020	0.001	0.067
$\delta$	0.090	1.54e-3	0.088	0.092

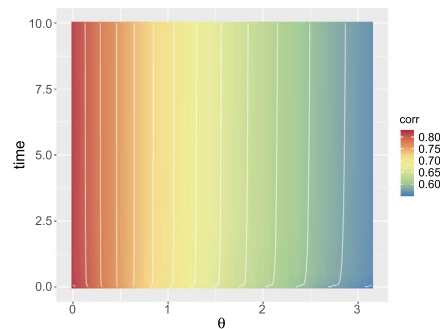


FIG 2. Posterior mean correlation contour plot for the near-surface air temperature.

We can also compute posterior summaries for  $\sigma^2/(\sigma^2 + \tau^2)$ , interpreted as the proportion of total variance attributable to the spatiotemporal random effect. Here, the posterior mean of  $\sigma^2/(\sigma^2 + \tau^2)$  is 0.834. In other words, our selected

covariance model accounts for 83.4% of the total variance. In Table 3 and Figure 2,  $c_s$  and  $c_t$  suggest that the surface temperature process exhibits persistent correlation over both space and time. Low values of  $\alpha$  in (14) add temporal smoothness.

### 5.3. Total cloud coverage reanalysis data

For this section, we utilize the 2017 National Centers for Environmental Prediction/National Center for Atmospheric Research daily average total cloud coverage reanalysis data (Kalnay et al., 1996). Total cloud coverage is defined as the fraction of the sky covered by any visible clouds, regardless of type. Total cloud coverage takes values between 0 and 100, representing a percentage of cloud coverage. Values of total cloud coverage close to 0 indicate clear skies, values from 40 to 70 percent represent broken cloud cover, and overcast skies correspond with 70 to 100 percent.

The degree of cloudiness impacts how much solar energy radiates to the Earth (see, e.g., Svensmark and Friis-Christensen, 1997). Total cloud coverage, like changes in global surface temperature, has been impacted by global climate changes (see, e.g., Melillo et al., 1993; Wylie et al., 2005), and changes in total cloud coverage are linked with many biological changes (see Pounds, Fogden and Campbell, 1999). Thus, tracking, predicting, and anticipating changes in cloudiness have important implications for understanding global climate changes and their effects on ecosystems.

The daily total cloud coverage reanalysis data represent daily averages and are given on a global grid with  $1.9^\circ$  spacing for latitude and  $1.875^\circ$  spacing for longitude. The spatial averages of cloud coverage over the first week of January are plotted in Figure 3. This map shows clear spatial variability that suggests that a spatial model is appropriate. We provide density estimates of total cloud coverage for each day of the week in Figure 3. These density estimates show that cloud coverage is similar across days.

Again, we thin the data to  $3.8^\circ$  spacing for latitude and  $3.75^\circ$  for longitude to carry out model comparison on hold-out data. In total, we have 4512 unique spatial locations, giving 31584 total observations. With 31584 data, carrying out fully Bayesian inference using a full Gaussian process model is intractable; thus, we utilize a nearest neighbor Gaussian process model. For these models, we use the same neighborhood formulations and fit the same covariance models with the same prior distribution to these data as we did in Section 5.2.

To obtain a test set to compare the six competing models, we randomly select 1000 locations and predict at these locations over the time-span of our data, giving a test set of size 7000. These hold-out locations are plotted in Figure 6 in Appendix C. As before, we compare these models in terms of predictive mean squared error, mean absolute error, continuous ranked probability scores, and 90% prediction interval coverage, as discussed in Section 4.3. The results of the model comparison are given in Table 4.

Again, an example from our new class was best and models in terms of prediction for the total cloud coverage dataset. All competing models were at

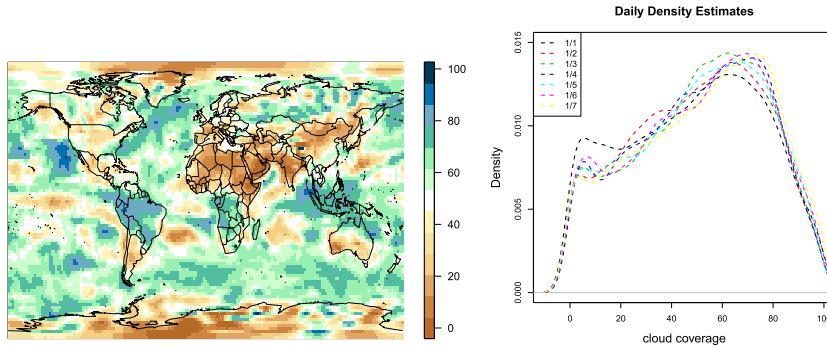


FIG 3. (Left) Total cloud coverage heat map in percentage, taking values 0 to 100 (Right) Kernel density estimates of total cloud coverage for each day

TABLE 4

Predictive performance of competing covariance models for the total cloud coverage dataset. For brevity in the table, let PRMSE and PMAE denote predictive mean squared error and mean absolute error, respectively. Relative CRPS is scaled such that the lowest is one. Bolded entries are used to indicate best model performance, i.e. lowest PRMSE, PMAE, and CRPS and 90% interval coverage closest to 90%.

Equation	PRMSE	PMAE	90% Coverage	CRPS	Relative CRPS
(11) and $\nu = 1/2$	13.92	11.08	0.95	7.85	1.32
(11) and $\nu = 3/2$	13.35	10.51	0.94	7.53	1.26
(12)	19.89	8.27	<b>0.90</b>	8.27	1.39
(13)	19.31	8.18	<b>0.90</b>	8.15	1.37
(14)	13.24	10.41	0.95	7.51	1.26
(15)	<b>10.68</b>	<b>7.66</b>	0.95	<b>5.96</b>	<b>1.00</b>

TABLE 5

Posterior summaries for the total cloud coverage dataset for parameters for the model fit using (15). Percentiles (2.5% and 97.5%) represent posterior percentiles.

	Mean	Std. Err.	2.5%	97.5%
$\tau^2$	22.280	3.601	16.764	31.598
$\sigma^2$	595.93	7.677	580.318	610.792
$c_s$	0.102	0.004	0.094	0.110
$c_t$	6.762	1.794	3.416	9.803
$\alpha$	0.350	0.071	0.232	0.516
$\beta$	0.952	0.049	0.817	0.999

least 26% worse in terms of continuous ranked probability score compared to the best model (15).

For the best predictive model in (15), we provide posterior summaries in Table 5. Additionally, we display the correlation contour as function of spherical distance  $\theta$  and time  $t$  for the posterior mean in Figure 4. The scale of the plots in Figure 4 are not the same as Figure 2. Correlation falls off sharply as a function of great circle distance. In this way, the total cloud coverage dataset differs greatly from the near-surface temperature dataset which demonstrated very persistent autocorrelation over space and time.



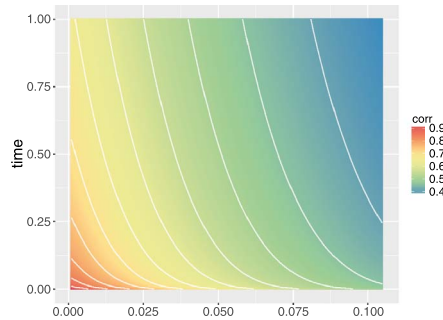


FIG 4. Posterior mean correlation contour plot for total cloud coverage data.

For the total cloud coverage data, spatiotemporal variance  $\sigma^2$  accounts for 96.40% of the total variance  $\sigma^2 + \tau^2$  (see Table 5). In Table 5, the parameter  $c_t$  suggests that the surface temperature process exhibits persistent correlation over time; however, as discussed, the parameter  $c_s$  indicates rapid decay as a function of space. The separability parameter  $\beta \in [0, 1]$  is close to one, meaning that the covariance process is nonseparable.

## 6. Discussion and conclusion

In this paper, we generalize the Gneiting criteria for nonseparable covariance functions (Gneiting, 2002) in Theorem 1 and present new classes of nonseparable covariance models for spatiotemporally data in Theorems 2 and 3. In a simulation study, we explored the identifiability of covariance parameters for two covariance functions from Theorem 2 and noted that these covariance functions have limited parameter identifiability. However, some of these challenges are remedied by fixing the spatial range parameter. We then illustrate the utility of our new Gneiting-like class using spherical distance through two climate reanalysis datasets from the National Centers for Environmental Prediction and National Center for Atmospheric Research (Kalnay et al., 1996). In these two data examples, covariance models from Theorem 2 outperform similar stationary nonseparable covariance models from Gneiting (2002) and Porcu, Bevilacqua and Genton (2016) using continuous ranked probability scores, root mean squared error, and mean absolute error. As discussed, we do not suggest that covariance functions from our new covariance classes are preferable for all datasets. However, these results highlight the benefit of allowing spatial distance to be scaled by the difference in time and the importance of using the spherical distance relative to Euclidean distance or chordal distance for these datasets.

The result in Theorem 1 presents a key for extending results obtained in Euclidean spaces to spheres cross time. Due to the lack of literature for multivariate cross-covariance models on spheres over time, with the notable exception of Alegria et al. (2017), we recommend this as a valuable area of expansion. In addition, the development of nonstationary covariance models for spheres cross time is an important direction for future research.

## Appendix A: Proofs for Theorems

In this Appendix, we provide proofs for all theorems presented in the manuscript. In addition, we provide a supplemental result (Theorem 4).

### Proof of Theorem 1

We start by proving the implication (1)  $\rightarrow$  (2). Let  $C \in \Psi_{d,T}$ . Then, according to Theorem 3.3 in Berg and Porcu (2017),  $C$  admits the expansion

$$C(\theta, u) = \sum_{k=0}^{\infty} b_{k,d}(u) \mathcal{G}_k^{(d-1)/2}(\cos \theta), \quad (\theta, u) \in [0, \pi] \times \mathbb{R}, \quad (19)$$

with  $b_{k,d}$  being positive definite on  $\mathbb{R}$  for all  $k = 0, 1, \dots$ , and with the constraint that  $\sum_{k=0}^{\infty} b_{k,d}(0) < \infty$ . Let us start by noting that the assumption  $\sum_k \int |b_{k,d}(u)| du < \infty$  implies  $C(\theta, \cdot) \in L_1(\mathbb{R})$  for all  $\theta \in [0, \pi]$ . In fact,

$$\begin{aligned} \int_{-\infty}^{\infty} |C(\theta, u)| du &= \int_{-\infty}^{\infty} \left| \sum_{k=0}^{\infty} b_{k,d}(u) \mathcal{G}_k^{(d-1)/2}(\cos \theta) \right| du \\ &\leq \sum_{k=0}^{\infty} \int_{-\infty}^{\infty} |b_{k,d}(u)| du < \infty, \end{aligned}$$

where the last step is justified by the fact that, for normalized Gegenbauer polynomials,  $|\mathcal{G}_k^\lambda(u)| \leq \mathcal{G}_k^\lambda(1)$ ,  $\lambda > 0$ .

Let  $C_\tau$  be the function defined through (8). Since  $C(\theta, \cdot) \in L_1(\mathbb{R})$  for all  $\theta$ , and using Lebesgue's theorem, we have

$$\begin{aligned} C_\tau(\theta) &= \int_{-\infty}^{+\infty} e^{-iu\tau} C(\theta, u) du \\ &= \int_{-\infty}^{+\infty} e^{-iu\tau} \sum_{k=0}^{\infty} b_{k,d}(u) \mathcal{G}_k^{(d-1)/2}(\cos \theta) du \\ &= \sum_{k=0}^{\infty} \widehat{b}_{k,d}(\tau) \mathcal{G}_k^{(d-1)/2}(\cos \theta), \quad \theta \in [0, \pi], \end{aligned}$$

where  $\widehat{b}_{k,d}(\tau) = \int_{-\infty}^{+\infty} e^{-iu\tau} b_{k,d}(u) du$ . Clearly, for any  $k = 0, 1, \dots$  we have that  $\widehat{b}_{k,d}$  is nonnegative and additionally  $\widehat{b}_{k,d} \in L_1(\mathbb{R})$ . To complete the proof, we invoke the theorem of Schoenberg (1942) and thus need to show that  $\sum_k \widehat{b}_{k,d}(\tau) < \infty$  for all  $\tau \in \mathbb{R}$ . Again invoking Lebesgue's theorem we have

$$\begin{aligned} \sum_{k=0}^{\infty} \widehat{b}_{k,d}(\tau) &= \sum_{k=0}^{\infty} \int_{-\infty}^{+\infty} e^{-iu\tau} b_{k,d}(u) du \\ &= \int_{-\infty}^{+\infty} e^{-iu\tau} \sum_{k=0}^{\infty} b_{k,d}(u) du \end{aligned}$$

$$= \int_{-\infty}^{+\infty} e^{-iu\tau} B_d(u) du,$$

with  $B_d(u) = \sum_{k=0}^{\infty} b_{k,d}(u)$ . Using the fact that  $\sum_{k=0}^{\infty} b_{k,d}(0) < \infty$  and that  $b_{k,d}(0) \geq |b_{k,d}(u)|$  for all  $u \in \mathbb{R}$  (because  $b_{k,d}$  are positive definite for all  $k$ ), we get

$$\infty > \sum_{k=0}^{\infty} b_{k,d}(0) \geq \sum_{k=0}^{\infty} |b_{k,d}(u)| \geq \sum_{k=0}^{\infty} b_{k,d}(u) = B_d(u), \quad u \in \mathbb{R},$$

showing that  $B_d$  is bounded and continuous. Further,  $B_d$  is positive definite on  $\mathbb{R}$  because positive definite functions are a convex cone being closed under pointwise convergence. To complete the result, we need to prove that  $B_d \in L_1(\mathbb{R})$ . This comes from the fact that

$$B_d(u) = C(0, u) \in L_1(\mathbb{R}).$$

The proof is completed.

To prove (2)  $\rightarrow$  (1), we let  $C_\tau$  as defined through (8) and suppose that  $C_\tau \in \Psi_d$  a.e.  $\tau \in \mathbb{R}$ . By Schoenberg (1942) theorem, we have that

$$\tilde{b}_{k,d}(\tau) := \kappa \int_0^\pi C_\tau(\theta) \mathcal{G}_k^{(d-1)/2}(\cos \theta) \sin \theta^{d-1} d\theta, \quad \tau \in \mathbb{R}, \quad (20)$$

is nonnegative, where  $\kappa > 0$  (Berg and Porcu, 2017). Using again Schoenberg (1942) theorem, we can write  $C_\tau$  as

$$C_\tau(\theta) = \sum_{k=0}^{\infty} \tilde{b}_{k,d}(\tau) \mathcal{G}_k(\cos \theta), \quad \theta \in [0, \pi].$$

Since  $C_\tau \in \Psi_d$  a.e.  $\tau$ , this in turn implies  $\infty > C_\tau(0) = \sum_k \tilde{b}_{k,d}(\tau)$  for all  $\tau \in \mathbb{R}$ . We now define

$$\tilde{B}_d(\tau) := \sum_{k=0}^{\infty} \tilde{B}_{k,d}(\tau), \quad \tau \in \mathbb{R}.$$

Apparently  $\tilde{B}_d$  is nonnegative. Let us now show that  $\tilde{B}_d \in L_1(\mathbb{R})$ . To do so, we note that

$$\begin{aligned} \int_{-\infty}^{+\infty} |\tilde{B}_d(u)| du &= \int_{-\infty}^{+\infty} \left| \sum_{k=0}^{\infty} \tilde{b}_{k,d}(u) \right| du \\ &\leq \sum_{k=0}^{\infty} \int_{-\infty}^{+\infty} |\tilde{b}_{k,d}(u)| du < \infty, \end{aligned}$$

so that  $\tilde{B}_d \in L_1(\mathbb{R})$  as asserted. This in turn implies that  $b_{k,d} \in L_1(\mathbb{R})$  for all  $k = 0, 1, \dots$ . Thus, we can define a function  $C : [0, \pi] \times \mathbb{R} \rightarrow \mathbb{R}$  through

$$C(\theta, u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu\tau} C_\tau(\theta) d\tau = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu\tau} \sum_{k=0}^{\infty} \tilde{b}_{k,d}(\tau) \mathcal{G}_k(\cos \theta)$$

$$= \sum_{k=0}^{\infty} b_{k,d}(u) \mathcal{G}_k(\cos \theta), \quad \theta \in [0, \pi], u \in \mathbb{R},$$

and where  $b_{k,d}(\cdot) = 1/(2\pi) \int e^{i\cdot\tau} \tilde{b}_{k,d}(\tau) d\tau$  is positive definite on  $\mathbb{R}$  for all  $n \in \mathbb{N}$ . Thus, the proof is completed by invoking Theorem 3.3 in Berg and Porcu (2017) and by verifying that

$$\sum_{k=0}^{\infty} b_{k,d}(0) = \sum_{k=0}^{\infty} \int_{-\infty}^{+\infty} \tilde{b}_{k,d}(\tau) d\tau < \infty.$$

We now prove the implication (2)  $\longrightarrow$  (3). Since  $C_\tau \in \Psi_d$  for almost every  $\tau \in \mathbb{R}$ , we have that (20) holds. This implies that

$$b_{k,d}(u) = \frac{1}{\pi} \int_{-\infty}^{+\infty} e^{iu\tau} \tilde{b}_{k,d}(\tau) d\tau, \quad k = 0, 1, \dots,$$

is positive definite. Summability of the sequence  $\{b_{k,d}(u)\}_{k=0}^{\infty}$  at  $u = 0$  follows easily from previous arguments.

To prove the implication (3)  $\longrightarrow$  (2), using (7) we have

$$\begin{aligned} b_{k,d}(u) &= \int_0^\pi C(\theta, u) \mathcal{G}_k^{(d-1)/2}(\cos \theta) \sin \theta^{d-1} d\theta \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{iu\tau} \int_0^\pi C_\tau(\theta) \mathcal{G}_k(\theta) \sin \theta^{d-1} d\theta d\tau, \end{aligned}$$

which shows that  $C_\tau \in \Psi_d$  for all  $\tau$  because the positive definiteness of  $b_{k,d}(\cdot)$  implies, by Lemma 4.3 in Berg and Porcu (2017), that the inner integral here,  $\int_0^\pi C_\tau(\theta) \mathcal{G}_k(\theta) \sin \theta^{d-1} d\theta$ , is nonnegative.

To conclude, the implication (3)  $\longrightarrow$  (1) has been shown by Berg and Porcu (2017). □

**Proof of Theorem 2**

We start by noting the beautiful formula

$$\int_0^\infty e^{-rx} e^{-r\xi} dr = \frac{1}{x + \xi}, \quad \xi > 0, x \geq 0.$$

We now consider the function

$$H_\xi(\theta, u) = \int_0^\infty e^{-r\theta} e^{-r\xi\psi(u^2)} dr = \frac{1}{\psi(u^2)} \left( \xi + \frac{\theta}{\psi(u^2)} \right)^{-1},$$

where  $(\theta, u) \in [0, \pi] \times \mathbb{R}$ . This shows that  $H_\xi$  is positive definite on every  $d$ -dimensional sphere cross time ( $\mathbb{R}$ ) because the mapping  $\theta \mapsto \exp(-r\theta)$  is positive definite on every  $d$ -dimensional sphere  $\mathbb{S}^d$  (Gneiting, 2013, Theorem 7) and

because  $u \mapsto \exp(-\xi r \psi(u^2))$  is positive definite on the real line. Since  $\varphi \in \mathcal{S}$ , we have, using (9), that

$$C(\theta, u) = \frac{1}{\psi(u^2)} \varphi\left(\frac{\theta}{\psi(u^2)}\right) = \int_0^\infty H_\xi(\theta, u) \mu(d\xi),$$

and this proves the assertion.  $\square$

### **Proof of Theorem 3**

We make use of the arguments in the proof of Theorem 2, in concert with formula (15) on page 15 of Bateman (1954):

$$\begin{aligned} & \frac{\pi}{2} \int_0^\infty \cos(t\omega) \exp(-(1+\omega^2)x) \frac{d\omega}{1+\omega^2} \\ &= e^{-u} \operatorname{erfc}\left(\sqrt{x} - \frac{t}{2\sqrt{x}}\right) + e^u \operatorname{erfc}\left(\sqrt{x} + \frac{t}{2\sqrt{x}}\right), \end{aligned}$$

for  $x, t \geq 0$ . We now replace  $x$  with  $\psi_{[0,\pi]}(\theta)$  and  $t$  with  $|u|$ . Since the composition of the negative exponential with a positive functions having completely monotonic derivative provides a completely monotonic function, we can invoke Theorem 7 in Gneiting (2013) to infer that the mixture above provides, in view of analogous arguments to the proof of Theorem 2, a positive definite function on  $\mathbb{S}^d \times \mathbb{R}$  for all  $d$ . The proof is completed.  $\square$

### **Supplementary Theorem for Section 3.1**

We now show how Theorem 1 can be useful to understand connections and analogues between the class of positive definite functions  $\mathbb{R} \times \mathbb{R}$  and positive definite functions on the circle  $\mathbb{S}^1$  cross  $\mathbb{R}$ .

**Theorem 4.** *Let  $\varphi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be a covariance function that is symmetric in both arguments. Let  $\varphi_\tau : \mathbb{R} \rightarrow \mathbb{R}$  be the function defined by*

$$\varphi_\tau(x) = \int_{-\infty}^{+\infty} e^{iu\tau} \varphi(x, u) du, \quad x \in \mathbb{R}. \quad (21)$$

*and suppose that such an integral is well defined. Let  $\varphi_\tau(x) = 0$  whenever  $|x| \geq \pi$ . Call  $C(\theta, u) = \varphi_{[0,\pi]}(\theta, u)$ ,  $\theta \in [0, \pi]$ ,  $u \in \mathbb{R}$ , where the restriction to  $[0, \pi]$  is with respect to the first argument. Then,  $C(\theta, u)$  is a covariance function on  $\mathbb{S}^3 \times \mathbb{R}$ .*

*Proof.* Since  $\varphi$  is positive definite in  $\mathbb{R} \times \mathbb{R}$ , by Lemma 1 in Gneiting (2002) we get that  $\varphi_\tau$  is positive definite in  $\mathbb{R}$  a.e.  $\tau \in \mathbb{R}$ . Additionally,  $\varphi_\tau(x) = 0$  whenever  $|x| \geq \pi$ . Call  $\psi_\tau$  the restriction of  $\varphi_\tau$  to  $[0, \pi]$ . By Corollary 3 in Gneiting (2013) we have that the coefficients  $\tilde{b}_{k,1}(\tau)$ ,  $\tau \in \mathbb{R}$  in the Schoenberg expansion of  $\psi_\tau$ , as defined in (20) are nonnegative and strictly decreasing in  $k$  for any fixed  $\tau \in \mathbb{R}$ . This implies that  $\varphi_\tau(\theta)$  is positive definite in  $\mathbb{S}^3$  for almost every  $\tau \in \mathbb{R}$ . Application of Theorem 1, Assertion 2, shows that  $C(\theta, u)$  is positive definite in  $\mathbb{S}^3 \times \mathbb{R}$ . The proof is completed.  $\square$

## Appendix B: Modeling details for the nearest neighbor Gaussian process

Suppose we begin with a parent Gaussian process over  $\mathbb{R}^d \times \mathbb{R}$  or  $\mathbb{S}^{d-1} \times \mathbb{R}$ . Nearest neighbor Gaussian processes induce sparsity in the precision matrix of the parent Gaussian process by assuming conditional independence given neighborhood sets (Datta et al., 2016a,b). Let  $\mathcal{S} = \{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), \dots, (\mathbf{s}_k, t_k)\}$  of  $k$  distinct location-time pairs denote the reference set, where we allow time to act as a natural ordering and impose an ordering on the locations observed at identical times. Then, we define neighborhood sets  $N_{\mathcal{S}} = \{N(\mathbf{s}_i, t_i); i = 1, \dots, k\}$  over the reference set with  $N(\mathbf{s}_i, t_i)$  consisting of the  $m$  nearest neighbors of  $(\mathbf{s}_i, t_i)$ , selected from  $\{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), \dots, (\mathbf{s}_{i-1}, t_{i-1})\}$ . If  $i \leq m + 1$ ,  $N(\mathbf{s}_i, t_i) = \{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), \dots, (\mathbf{s}_{i-1}, t_{i-1})\}$ . For the Gibbs sampler, we need to define an inverse of the neighborhood set, which we call  $U(\mathbf{s}_i, t_i)$ . The set  $U(\mathbf{s}_i, t_i)$  consists of all sites that include  $(\mathbf{s}_i, t_i)$  in their neighborhood sets.

Along with  $\mathcal{S}$ ,  $N_{\mathcal{S}}$  defines a Gaussian directed acyclic graph  $\mathbf{w}_{\mathcal{S}}$  with a joint distribution

$$p(\mathbf{w}_{\mathcal{S}}) = \prod_{i=1}^k p(w(\mathbf{s}_i, t_i) \mid \mathbf{w}_{N(\mathbf{s}_i, t_i)}) = \prod_{i=1}^k \mathcal{N}(\mathbf{w}(\mathbf{s}_i, t_i) \mid \mathbf{B}_{(\mathbf{s}_i, t_i)} \mathbf{w}_{N(\mathbf{s}_i, t_i)}, \mathbf{F}_{(\mathbf{s}_i, t_i)}), \quad (22)$$

where  $\mathcal{N}$  is a normal distribution,

$$\begin{aligned} \mathbf{B}_{(\mathbf{s}_i, t_i)} &= C_{(\mathbf{s}_i, t_i), N(\mathbf{s}_i, t_i)} C_{N(\mathbf{s}_i, t_i)}^{-1}, \\ \mathbf{F}_{(\mathbf{s}_i, t_i)} &= \sigma^2 - \mathbf{B}_{(\mathbf{s}_i, t_i)} C_{N(\mathbf{s}_i, t_i), (\mathbf{s}_i, t_i)}, \end{aligned}$$

where  $C_{(\mathbf{s}_i, t_i), N(\mathbf{s}_i, t_i)}$  is a vector of covariances between  $(\mathbf{s}_i, t_i)$  and its neighbors, we define  $C_{N(\mathbf{s}_i, t_i)}$  to be the covariance matrix for the neighbors of  $(\mathbf{s}_i, t_i)$ , and  $\mathbf{w}_{N(\mathbf{s}_i, t_i)}$  is the subset of  $\mathbf{w}_{\mathcal{S}}$  corresponding to neighbors  $N(\mathbf{s}_i, t_i)$  (Datta et al., 2016a). Datta et al. (2016a) extend this Gaussian directed acyclic graph to a GP. This Gaussian process formulation only requires storage of  $k$   $m \times m$  distance matrices and requires many fewer floating point operations than full Gaussian process models (see Datta et al., 2016a). Like any other GP model, the NNGP can be utilized hierarchically for spatiotemporal random effects. In this article, we use NNGP as an alternative to the full Gaussian process specification.

We envision our model taking the following form:

$$\begin{aligned} Y(\mathbf{s}, t) &= \mathbf{x}(\mathbf{s}, t)^\top \boldsymbol{\beta} + w(\mathbf{s}, t) + \epsilon(\mathbf{s}, t), \\ w(\mathbf{s}, t) &\sim NNGP(0, C((\mathbf{s}, t), (\mathbf{s}', t'))), \\ \epsilon(\mathbf{s}, t) &\sim GP(0, \tau^2 \delta_{\mathbf{s}}^{\mathbf{s}} \delta_t^t), \end{aligned} \quad (23)$$

where  $Y(\mathbf{s}, t)$  is a spatiotemporal process measured (with error),  $\mathbf{x}(\mathbf{s}, t)$  are  $p$  spatiotemporal covariates, and  $\delta_a^b$  is the Kronecker delta function. We define  $C((\mathbf{s}, t), (\mathbf{s}', t'))$  using a covariance model discussed in Section 2 or Section 3. We recommend using inverse gamma (IG) prior distributions for  $\tau^2$  from the pure error term and  $\sigma^2$  from the covariance function because this selection

gives closed form full conditional distributions. If the outcomes and covariates are centered, then an intercept is unnecessary. If covariates are not available, then  $\mathbf{x}(\mathbf{s}, t)^\top \boldsymbol{\beta}$  is replaced with  $\mu$ . Here, for more compact notation, we index space-time location pairs with  $i$  as  $(\mathbf{s}_i, t_i)$  and refer to corresponding outcomes, covariates, and spatiotemporal random effects as  $y_i$ ,  $\mathbf{x}_i$ , and  $w_i$ , respectively.

The prior mean and variance for regression coefficients  $\boldsymbol{\beta}$  are  $m_\beta$  and  $V_\beta^{-1}$ , respectively. Additionally, let  $a_V$  and  $a_\tau$  be shape parameters for the inverse gamma prior distributions for  $\sigma^2$  and  $\tau^2$ . Similarly, let  $b_V$  and  $b_\tau$  be scale parameters (corresponding to rate parameter of the gamma distribution) for the inverse gamma prior distributions for  $\sigma^2$  and  $\tau^2$ .

The full conditional distributions for the Gibbs sampler, which we denote  $\cdot | \dots$ , are

$$\begin{aligned}\boldsymbol{\beta} | \dots &\sim \mathcal{N}_p(V_\beta^* m_\beta^*, V_\beta^*) \\ \tau^2 | \dots &\sim IG(a_\tau^*, b_\tau^*) \\ \sigma^2 | \dots &\sim IG(a_V^*, b_V^*) \\ \mathbf{w}_i | \dots &\sim \mathcal{N}_1(V_{w_i}^* m_{w_i}^*, V_{w_i}^*).\end{aligned}$$

To express  $V_{w_i}^*$  and  $m_{w_i}^*$ , we must define some additional terms. First, we let  $B_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)}$  be the scalar in  $B_{(\mathbf{s}', t')}$  corresponding to  $(\mathbf{s}_i, t_i)$ . Second, we define

$$a_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)} = w(\mathbf{s}', t') - \sum_{(\mathbf{s}_j, t_j) \in N(\mathbf{s}', t'), (\mathbf{s}_j, t_j) \neq (\mathbf{s}_i, t_i)} B_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)} w(\mathbf{s}_i, t_i).$$

For more details, see Datta et al. (2016a).

The parameters of the full conditional distributions are as follows:

$$\begin{aligned}V_\beta^* &= \left( \mathbf{X}^\top \mathbf{X} / \tau^2 + V_\beta^{-1} \right)^{-1} \\ m_\beta^* &= V_\beta^{-1} m_\beta + \sum_i \mathbf{x}_i (y_i - w_i) / \tau^2 \\ V_{w_i}^* &= \left( 1/\tau^2 + F_{(\mathbf{s}_i, t_i)}^{-1} + \sum_{(\mathbf{s}', t') : (\mathbf{s}', t') \in U(\mathbf{s}_i, t_i)} B_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)}^2 / F_{(\mathbf{s}', t')} \right)^{-1} \\ m_{w_i}^* &= (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) / \tau^2 + B_{(\mathbf{s}_i, t_i)} \mathbf{w}_{N(\mathbf{s}_i, t_i)} / F_{(\mathbf{s}_i, t_i)} + \\ &\quad \sum_{(\mathbf{s}', t') : (\mathbf{s}', t') \in U(\mathbf{s}_i, t_i)} B_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)} a_{(\mathbf{s}', t'), (\mathbf{s}_i, t_i)} / F_{(\mathbf{s}', t')} \\ a_V^* &= a_V + n/2 \\ b_V^* &= b_V + \sigma^2 \sum_i (w_i - B_{(\mathbf{s}_i, t_i)} \mathbf{w}_{N(\mathbf{s}_i, t_i)})^\top (w_i - B_{(\mathbf{s}_i, t_i)} \mathbf{w}_{N((\mathbf{s}_i, t_i))}) / F_{(\mathbf{s}_i, t_i)} \\ a_\tau^* &= a_\tau + \frac{n}{2} \\ b_\tau^* &= b_\tau + \frac{1}{2} \sum_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - w_i)^2.\end{aligned}$$

Prediction at an arbitrary location and time requires selection of  $m$  nearest neighbors from the reference set  $\mathcal{S}$  for that location-time pair. We discuss two ways of selecting  $m$  neighbors. In theory, any location-time pair from the reference set can be selected as a neighbor for any prediction. If we allow predictions to depend on data occurring after the prediction time, then we call this a *retrospective* prediction since such a prediction could only be made retrospectively. On the other hand, a *prospective* prediction limits neighbor selection to elements of  $\mathcal{S}$  that occur at the same time or prior to the time of prediction. Datta et al. (2016b) selects neighbors for prospective predictions, and we do the same in our analyses.

Predicted spatiotemporal random effects at location-time pairs follow a conditional normal distribution, where conditioning is limited to its neighbors. For any space-time pair  $(\mathbf{s}, t)$ , the conditional distribution of the random effect is

$$\mathbf{w}(\mathbf{s}) | \mathbf{w}_{N(\mathbf{s})} \sim \mathcal{N} \left( C_{\mathbf{s}, N(\mathbf{s})} C_{N(\mathbf{s})}^{-1} \mathbf{w}_{N(\mathbf{s})}, C(\mathbf{s}, \mathbf{s}) - C_{\mathbf{s}, N(\mathbf{s})} C_{N(\mathbf{s})}^{-1} C_{\mathbf{s}, N(\mathbf{s})}^{\top} \right). \quad (24)$$

Then, the posterior prediction  $Y(\mathbf{s}, t) | \mathbf{Y}$  is  $\mathbf{x}(\mathbf{s}, t)^{\top} \boldsymbol{\beta} + w(\mathbf{s}, t) + \epsilon_i(\mathbf{s}, t)$ , where, in practice, posterior samples of  $\boldsymbol{\beta}$ ,  $w(\mathbf{s}, t)$ , and  $\tau^2$  are used to sample from the posterior predictive distribution. Predictions at hold-out location-time pairs can be used to compare competing models.

### Appendix C: Locations of hold-out data from Section 5

Here, we provide the locations of hold-out data used for model validation. The locations for hold-out air temperature are given in Figure 5. The locations for hold-out cloud coverage are in Figure 6.

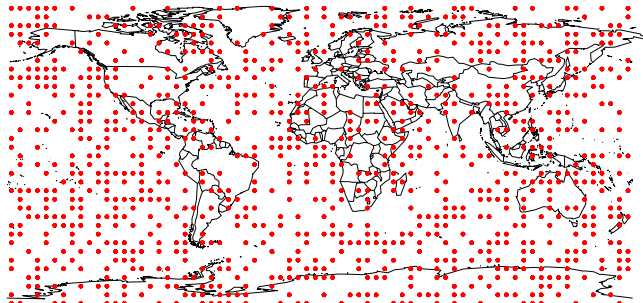


FIG 5. Hold-out locations used for predictive performance for the near-surface air temperature.



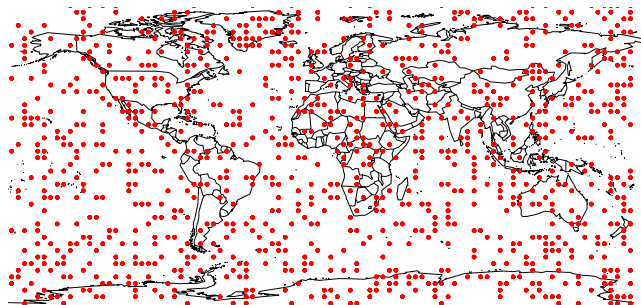


FIG 6. Hold-out locations used for predictive performance for the total cloud coverage dataset.

## References

- ALEGRIA, A. and PORCU, E. (2017). The Dimple Problem related to Space-Time Modeling under the Lagrangian Framework. *Journal of Multivariate Analysis* **162** 110-121. [MR3719338](#)
- ALEGRIA, A., PORCU, E., FURRER, R. and MATEU, J. (2017). Covariance Functions for Multivariate Gaussian Fields Evolving Temporally over Planet Earth. *Technical Report, University Federico Santa Maria*.
- APANASOVICH, T. and GENTON, M. (2010). Cross-Covariance Functions for Multivariate Random Fields Based on Latent Dimensions. *Biometrika* **97** 15-30. [MR2594414](#)
- BANERJEE, S. (2005). On Geodetic Distance Computations in Spatial Modeling. *Biometrics* **61** 617-625. [MR2140936](#)
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Crc Press. [MR3362184](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian Predictive Process Models for Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 825-848. [MR2523906](#)
- BARBOSA, V. S. and MENEGATTO, V. A. (2017). Strict Positive Definiteness on Products of Compact Two-Point Homogeneous Spaces. *Integral Transforms and Special Functions* **28** 56-73. [MR3574321](#)
- BATEMAN, H. (1954). *Tables of Integral Transforms, Volume I*. McGraw-Hill Book Company, New York. [MR0065685](#)
- BERG, C. (2008). Stieltjes-Pick-Bernstein-Schoenberg and Their Connection to Complete Monotonicity. In *Quantitative Methods for Current Environmental Issues* 15-45.
- BERG, C. and PORCU, E. (2017). From Schoenberg Coefficients to Schoenberg Functions. *Constructive Approximation* **45** 217-241. [MR3619442](#)
- BROWN, T. A. (1974). Admissible Scoring Systems for Continuous Distributions Technical Report No. P-5235, The Rand Corporation, Santa Monica, California.

- CASTRUCCIO, S. and STEIN, M. L. (2013). Global Space-Time Models for Climate Ensembles. *Ann. Appl. Statist.* **7** 1593–1611. [MR3127960](#)
- CRESSIE, N. and JOHANNESSON, G. (2008). Fixed Rank Kriging for Very Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 209–226. [MR2412639](#)
- CRESSIE, N. and WIKLE, C. K. (2015). *Statistics for Spatio-Temporal Data*. John Wiley & Sons. [MR2848400](#)
- DAI, F. and XU, Y. (2013). *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer. [MR3060033](#)
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016a). Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association* **111** 800–812. [MR3538706](#)
- DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. and SCHAAP, M. (2016b). Nonseparable Dynamic Nearest Neighbor Gaussian Process Models for Large Spatio-Temporal Data with an Application to Particulate Matter Analysis. *The Annals of Applied Statistics* **10** 1286–1316. [MR3553225](#)
- FOLLAND, C. K., RAYNER, N. A., BROWN, S., SMITH, T., SHEN, S., PARKER, D., MACADAM, I., JONES, P., JONES, R. N. and NICHOLLS, N. (2001). Global Temperature Change and its Uncertainties since 1861. *Geophysical Research Letters* **28** 2621–2624.
- FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics* **15** 502–523. [MR2291261](#)
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis* **2**. CRC press Boca Raton, FL. [MR3235677](#)
- GNEITING, T. (2002). Nonseparable, Stationary Covariance Functions for Space-Time Data. *Journal of the American Statistical Association* **97** 590–600. [MR1941475](#)
- GNEITING, T. (2013). Strictly and Non-Strictly Positive Definite Functions on Spheres. *Bernoulli* **19** 1327–1349. [MR3102554](#)
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102** 359–378. [MR2345548](#)
- GNEITING, T. and SCHLATHER, M. (2004). Stochastic Models That Separate Fractal Dimension and the Hurst Effect. *SIAM Rev.* **46** 269–282. [MR2114455](#)
- GRADSHTEYN, I. S. and RYZHIK, I. M. (2007). *Tables of Integrals, Series, and Products*, seventh ed. Academic Press, Amsterdam. [MR2360010](#)
- GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian Process Approximation for Large Computer Experiments. *Journal of Computational and Graphical Statistics* **24** 561–578. [MR3357395](#)
- GUELLA, J. C., MENEGATTO, V. A. and PERON, A. P. (2016a). An Extension of a Theorem of Schoenberg to a Product of Spheres. *Banach Journal of Mathematical Analysis* **10** 671–685. [MR3543906](#)
- GUELLA, J. C., MENEGATTO, V. A. and PERON, A. P. (2016b). Strictly

- Positive Definite Kernels on a Product of Spheres II. *SIGMA* **12**. [MR3566202](#)
- GUELLA, J. C., MENEGATTO, V. A. and PERON, A. P. (2017). Strictly Positive Definite Kernels on a Product of Circles. *Positivity* **21** 329–342. [MR3613000](#)
- HANSEN, J., SATO, M., RUEDY, R., LO, K., LEA, D. W. and MEDINA-ELIZADE, M. (2006). Global Temperature Change. *Proceedings of the National Academy of Sciences* **103** 14288–14293.
- HANSEN, J., RUEDY, R., SATO, M. and LO, K. (2010). Global Surface Temperature Change. *Reviews of Geophysics* **48**.
- HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D., KATZFUSS, M. et al. (2018). A Case Study Competition among Methods for Analyzing Large Spatial Data. *Journal of Agricultural, Biological and Environmental Statistics* 1–28.
- HEINE, V. (1955). Models for Two-Dimensional Stationary Stochastic Processes. *Biometrika* **42** 170–178. [MR0071673](#)
- HELD, I. M. and SODEN, B. J. (2006). Robust Responses of the Hydrological Cycle to Global Warming. *Journal of Climate* **19** 5686–5699.
- HIGDON, D. (2002). Space and Space-Time Modeling using Process Convolutions. In *Quantitative Methods for Current Environmental Issues* 37–56. Springer. [MR2059819](#)
- JEONG, J. and JUN, M. (2015). A Class of Matern-like Covariance Functions for Smooth Processes on a Sphere. *Spatial Statistics* **11** 1–18. [MR3311853](#)
- KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G. and WOOLLEN, J. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American meteorological Society* **77** 437–471.
- KATZFUSS, M. and GUINNESS, J. (2017). A General Framework for Vecchia Approximations of Gaussian Processes. *arXiv preprint arXiv:1708.06302*.
- KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance Tapering for Likelihood-Based Estimation in Large Spatial Data Sets. *Journal of the American Statistical Association* **103** 1545–1555. [MR2504203](#)
- LINDGREN, F., RUE, H. and LINDSTROEM, J. (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 423–498. [MR2853727](#)
- MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika* **71** 135–146. [MR0738334](#)
- MATHESON, J. E. and WINKLER, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science* **22** 1087–1096.
- MELILLO, J. M., MCGUIRE, A. D., KICKLIGHTER, D. W., MOORE, B., VOROSMARTY, C. J. and SCHLOSS, A. L. (1993). Global Climate Change and Terrestrial Net Primary Production. *Nature* **363** 234.
- PARMESAN, C. and YOHE, G. (2003). A Globally Coherent Fingerprint of Climate Change Impacts across Natural Systems. *Nature* **421** 37.
- PORCU, E., ALEGRÍA, A. and FURRER, R. (2018). Modeling Temporally Evolv-

- ing and Spatially Globally Dependent Data. *International Statistical Review* **To Appear**. [MR3852415](#)
- PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-Temporal Covariance and Cross-Covariance Functions of the Great Circle Distance on a Sphere. *Journal of the American Statistical Association* **111** 888–898. [MR3538713](#)
- PORCU, E., GREGORI, P. and MATEU, J. (2006). Nonseparable Stationary Anisotropic Space–Time Covariance Functions. *Stochastic Environmental Research and Risk Assessment* **21** 113–122. [MR2307631](#)
- PORCU, E., MATEU, J. and BEVILACQUA, M. (2007). Covariance Functions which are Stationary or Nonstationary in Space and Stationary in Time. *Statistica Neerlandica* **61** 358–382. [MR2355065](#)
- PORCU, E., MATEU, J. and CHRISTAKOS, G. (2010). Quasi-Arithmetic Means of Covariance Functions with Potential Applications to Space-Time Data. *J. Multivariate Anal.* **100** 1830–1844. [MR2535390](#)
- PORCU, E. and SCHILLING, R. (2011). From Schoenberg to Pick-Nevanlinna: Towards a Complete Picture of the Variogram Class. *Bernoulli* **17** 441–455. [MR2797998](#)
- POUNDS, J. A., FOGDEN, M. P. and CAMPBELL, J. H. (1999). Biological Response to Climate Change on a Tropical Mountain. *Nature* **398** 611.
- SCHILLING, R., SONG, R. and VONDRACEK, Z. (2012). *Bernstein Functions. Theory and Applications*. De Gruyter. [MR2978140](#)
- SCHLATHER, M. (2010). Some Covariance Models Based on Normal Scale Mixtures. *Bernoulli* **16** 780–797. [MR2730648](#)
- SCHOENBERG, I. J. (1942). Positive Definite Functions on Spheres. *Duke Mathematical Journal* **9** 96–108. [MR0005922](#)
- SHIROTA, S. and GELFAND, A. E. (2017). Space and Circular Time Log Gaussian Cox Processes with Application to Crime Event Data. *The Annals of Applied Statistics* **11** 481–503. [MR3693535](#)
- SIMMONS, A., JONES, P., DA COSTA BECHTOLD, V., BELJAARS, A., KÅLLBERG, P., SAARINEN, S., UPPALA, S., VITERBO, P. and WEDI, N. (2004). Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research: Atmospheres* **109**.
- STEIN, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York. [MR1697409](#)
- STEIN, M. L. (2008). A Modeling Approach for Large Spatial Datasets. *Journal of the Korean Statistical Society* **37** 3–10. [MR2420389](#)
- STEIN, M. L. (2014). Limitations on Low Rank Approximations for Covariance Matrices of Spatial Data. *Spatial Statistics* **8** 1–19. [MR3326818](#)
- STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating Likelihoods for Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 275–296. [MR2062376](#)
- SVENSMARK, H. and FRIIS-CHRISTENSEN, E. (1997). Variation of Cosmic Ray Flux and Global Cloud Coverage—a Missing Link in Solar-Climate Relationships. *Journal of Atmospheric and Solar-Terrestrial Physics* **59** 1225–1232.

- THOMAS, C. D., CAMERON, A., GREEN, R. E., BAKKENES, M., BEAUMONT, L. J., COLLINGHAM, Y. C., ERASMUS, B. F., DE SIQUEIRA, M. F., GRAINGER, A. and HANNAH, L. (2004). Extinction Risk from Climate Change. *Nature* **427** 145.
- VECCHIA, A. V. (1988). Estimation and Model Identification for Continuous Spatial Processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 297–312. [MR0964183](#)
- WHITE, P. and PORCU, E. (2019). Nonseparable Covariance Models on Circles Cross Time: A Study of Mexico City Ozone. *Environmetrics* e2558.
- WYLIE, D., JACKSON, D. L., MENZEL, W. P. and BATES, J. J. (2005). Trends in Global Cloud Cover in Two Decades of HIRS Observations. *Journal of Climate* **18** 3021–3031.
- ZHANG, H. (2004). Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics. *Journal of the American Statistical Association* **99** 250–261. [MR2054303](#)