Modeling temporally evolving and spatially globally dependent data

Porcu, Emilio ; Alegria, Alfredo ; Furrer, Reinhard

Abstract: The last decades have seen an unprecedented increase in the availability of data sets that are inherently global and temporally evolving, from remotely sensed networks to climate model ensembles. This paper provides an overview of statistical modeling techniques for space–time processes, where space is the sphere representing our planet. In particular, we make a distintion between (a) second order-based approaches and (b) practical approaches to modeling temporally evolving global processes. The former approaches are based on the specification of a class of space–time covariance functions, with space being the two-dimensional sphere. The latter are based on explicit description of the dynamics of the space–time process, that is, by specifying its evolution as a function of its past history with added spatially dependent noise. We focus primarily on approach (a), for which the literature has been sparse. We provide new models of space–time covariance functions for random fields defined on spheres cross time. Practical approaches (b) are also discussed, with special emphasis on models built directly on the sphere, without projecting spherical coordinates onto the plane. We present a case study focused on the analysis of air pollution from the 2015 wildfires in Equatorial Asia, an event that was classified as the year's worst environmental disaster. The paper finishes with a list of the main theoretical and applied research problems in the area, where we expect the statistical community to engage over the next decade.

DOI: https://doi.org/10.1111/insr.12266

International Statistical Review

International Statistical Review (2018), 0, 0, 1–34 doi:10.1111/insr.12266

Modeling Temporally Evolving and Spatially Globally Dependent Data

Emilio Porcu[1], Alfredo Alegria[2] and Reinhard Furrer[3]

[1]*School of Mathematics, Statistics and Physics, University of Newcastle, Newcastle Upon Tyne, England, and Department of Mathematics, University Federico Santa Maria, Valparaiso 2360102, Chile*
E-mail: georgepolya01@gmail.com
[2]*School of Mathematics, Statistics and Physics, University of Newcastle, Newcastle Upon Tyne, England*
E-mail: alfredo.alegria.jimenez@gmail.com
[3]*Department of Mathematics and Department of Computational Science, University of Zurich, 8057 Zurich, Switzerland*
E-mail: reinhard.furrer@math.uzh.ch

Summary

The last decades have seen an unprecedented increase in the availability of data sets that are inherently global and temporally evolving, from remotely sensed networks to climate model ensembles. This paper provides an overview of statistical modeling techniques for space–time processes, where space is the sphere representing our planet. In particular, we make a distintion between (a) second order-based approaches and (b) practical approaches to modeling temporally evolving global processes. The former approaches are based on the specification of a class of space–time covariance functions, with space being the two-dimensional sphere. The latter are based on explicit description of the dynamics of the space–time process, that is, by specifying its evolution as a function of its past history with added spatially dependent noise.
We focus primarily on approach (a), for which the literature has been sparse. We provide new models of space–time covariance functions for random fields defined on spheres cross time. Practical approaches (b) are also discussed, with special emphasis on models built directly on the sphere, without projecting spherical coordinates onto the plane.
We present a case study focused on the analysis of air pollution from the 2015 wildfires in Equatorial Asia, an event that was classified as the year's worst environmental disaster. The paper finishes with a list of the main theoretical and applied research problems in the area, where we expect the statistical community to engage over the next decade.

Key words: Covariance functions; great circle; massive data sets; spheres.

1 Introduction

The strong evidence of a changing climate over the last century (IPCC, 2013) has prompted the scientific community to seek comprehensive strategies for monitoring the state of the climate system over the entire planet. The surge in satellite observations, the increase in computational and storage availability, as well as an increase in the horizontal resolution of global climate models, the deployment of new global and automated networks (the ARGO floats and the AERONET, see Holben *et al*., 1998) and the recent development of smartphone-based data,

which potentially allow every user on the planet to contribute scientific data (Citizen Science), have generated an increase in the size of the data of orders of magnitude. Such an increase in the volume, variate and velocity of globally indexed data serves as a strong catalyst for the statistical community to develop models that are inherently global and time evolving.

The research questions underpinning global space–time models span from optimal interpolation (kriging) for global coverage over both space and time (for variables such as temperature and precipitation but also, more recently, ozone, carbon dioxide and aerosol optical depth), to interpolation in the input space of Earth System Models by generating fast approximations (emulators) that can be used to perform fast and affordable sensitivity analysis. While the aforementioned topics are of high scientific interest, the development of appropriate statistical methodologies for global and temporal data has been limited, with advances in two seemingly very different directions.

The construction of models on the sphere cross temporal horizon calls for rigorous development of a theory that would allow for valid processes with the proper distance over the curved surface of the planet. Under the assumption of Gaussianity, the second-order structure can be explicitly specified and the properties of the process can be studied directly from its functional form. The theory for this approach has been actively developed over the last decade, but thus far has been limited to the large-scale structure for the covariance function, such as isotropy or stationarity across longitudes. We refer to this approach as the *second order-based approach*.

Alternative definitions of the space–time process rely on either the decoupling of the spatial and temporal parts through the specification of dynamics in time, or its representation as a solution of a stochastic partial differential equation (SPDE). These techniques are designed primarily for inferential purposes, and they are particularly suitable for modern, massive data sets. This increased suitability for inference, however, comes at the expense of convenient expressions that allow for an understanding of the underlying theoretical properties of the process. We refer to this as the *practical approach*.

We start by describing the second order-based approaches, which rely on the first- and second-order moments of the underlying random field on the sphere cross time. In particular, the focus becomes the space–time covariance, where space is the two-dimensional sphere. For stochastic processes on a sphere, the reader is referred to Jones (1963), Marinucci and Peccati (2011) and the thorough reviews in Gneiting (2013) and Jeong *et al.* (2017). For space–time stochastic processes on the sphere, we refer the reader to the more recent approaches in Porcu *et al.* (2016a), Berg and Porcu (2017) and Jeong and Jun (2015). Generalisations to multivariate space–time processes have been considered in Alegria *et al.* (2017). The increasing interest in modeling stochastic processes over spheres or spheres cross time with an explicit covariance function is also reflected in work in areas as diverse as mathematical analysis (Schoenberg, 1942; Gangolli, 1967; Hannan, 1970; Menegatto, 1994; 1995; Chen *et al.*, 2003; Menegatto *et al.*, 2006; Beatson *et al.*, 2014; Guella *et al.*, 2017; 2016b; Barbosa & Menegatto, 2017; Guella *et al.*, 2016a), probability theory (Baldi & Marinucci, 2006; Lang & Schwab, 2013; Hansen *et al.*, 2015; Clarke *et al.*, 2018), spatial point processes (Møller *et al.*, 2018), spatial geostatistics (Christakos & Papanicolaou, 2000; Gneiting, 2002; Hitczenko & Stein, 2012; Huang *et al.*, 2012; Gerber *et al.*, 2017), space–time geostatistics (Christakos, 1991; 2000; Christakos *et al.*, 2000; Porcu *et al.*, 2016a; Berg & Porcu, 2017) and mathematical physics (Istas, 2005; Leonenko & Sakhno, 2012; Malyarenko, 2013).

The natural metric to be used on the sphere is the geodesic or great circle distance (details are explained in subsequent sections). However, if this metric is used in space–time covariance models defined on Euclidean spaces cross time, these are generally not valid on the sphere cross time. This fact is inherited from merely spatial covariance functions. For instance, the Matérn covariance function (Stein, 1999) is positive definite on the sphere only under a severe restriction on the smoothing parameter (Gneiting, 2013). A technical argument that explains

why space–time covariances in Euclidean spaces are not valid on the sphere if coupled with the geodesic distance is given in (Berg & Porcu, 2017, theorem 3.3), who show that covariance functions on spheres cross time have a very specific Fourier representation, being different to that of space–time covariances in Euclidean distances (see theorem 4.3.1 in Gneiting *et al*., 2007).

The resulting need for new space–time covariance function theory has motivated a rich literature based on positive definite functions based on great circle distance.

Techniques based on covariance functions are certainly accurate both in terms of likelihood inference and kriging predictions; yet they imply a high computational cost when dealing with large data sets over the space–time domain. The main computational hurdle is the calculation of the determinant and of the quadratic form based on the inverse of the covariance matrix. The so-called *big N problem* in this case requires a compromise between statistical and computational efficiencies. The reader is referred to Bevilacqua *et al.* (2010) and Bevilacqua *et al.* (2012), Stein (2005b), Furrer *et al.* (2006) and Furrer *et al.* (2016), amongst others, for statistical approaches based on covariance functions, which aim to mitigate such a computational burden. A notable approach to the problem of prediction for very large data sets can be found in Cressie and Johannesson (2008).

As for practical approaches, when analysing modern, massive data sets arising from remotely sensed data, climate models or reanalysis data products, a model that fully specifies the covariance function would require a prohibitive amount of information to be stored in the covariance matrix, as well as a prohibitive number of flops and iterations for maximising the likelihood or exploring the Markov chain when performing Bayesian inference.

In the context of data in Euclidean spaces, one of the most popular and natural alternatives is to explicitly describe the dynamics of the process by specifying the evolution as a function of its past history with added spatially dependent noise. This approach has received strong support from reference textbooks on space–time modeling (see Cressie & Wikle, 2011), and recent studies have extended this methodology to the context of space–time data. Richardson *et al*. (2017, 2016) and Tebaldi and Sansó (2009), amongst others, recommend the use of an explicit description of the dynamics of the process by specifying its evolution as a function of the spatial distribution of the process. Dynamic space–time models have a long history in Euclidean spaces (Cressie & Wikle, 2011, with the references therein), but the literature on the sphere is more sparse, with the exception of Castruccio and Stein (2013) and related work. This paper will review the recent literature on this approach, with a particular emphasis on scalable methods for large data sets. Dynamics on large regions of Earths surface have been studied by Kang *et al.* (2010), who consider aerosol data from multi-angle imaging spectroradiometers along the Americas, the Atlantic Ocean and the western part of Europe and North Africa. Other relevant applications are proposed in Oleson *et al.* (2013); Nguyen *et al.* (2014); Banerjee *et al.* (2008).

In recent years, a powerful modeling approach has emerged based on the consideration of a space–time global process as a solution to a SPDE defined over the sphere and in time. Earlier work proposed in Jones and Zhang (1997) was based on a diffusion-injection equation, which is just one of a multitude of SPDE-based models commonly used to describe physical processes (Christakos, 2000). Later studies focused on specifying an SPDE over the sphere or, more generally, on Riemannian geometries, and more recently under the Integrated Nested Laplace Approximation environment (Rue & Tjelmeland, 2002; Lindgren *et al.*, 2011; Lindgren & Rue, 2015; Cameletti *et al.*, 2012). The key intuition is to use the SPDE approach as a link to approximate a continuous stochastic process with a Gaussian Markov random field, a discretised version of the first. This has remarkable computational advantages (see Lindgren *et al*., 2011) and also allows one to build flexible models by providing flexible functional expressions for the differential operator. Some notable approaches to modeling global data under this

framework can be found in Bolin and Lindgren (2011). Although substantial effort has been made to introduce the Markov field architecture coupled with the Integrated Nested Laplace Approximation routine for spatial data, to our knowledge, only few studies have focused on the sphere cross time, which will be surveyed in the succeeding text.

The paper is structured as follows. Section 2 provides the necessary background material on random fields over spheres or spheres cross time, including their covariance functions. Section 3 details the *second order-based approach*, with construction principles for characterising space–time covariance functions on the sphere cross time domain. New covariance functions are also introduced with formal proofs given in the Appendix. Section 4 is devoted to *practical approaches*, with emphasis on dynamical methods as well as methods based on SPDE and Gaussian Markov random fields. Section 5 presents a case study focused on the analysis of air pollution from the 2015 wildfires in Equatorial Asia, an event that was classified as the year's worst environmental disaster. A massive global space–time data set of air quality from NASA's MERRA-2 reanalysis, consisting of more than 12 million data points, is provided. Approaches from both categories (a) and (b) are compared and their relative merits are discussed. The paper ends with a discussion in Section 6 and with a list of research problems in the area, where we expect the statistical community to engage over the next decade. Technical proofs are deferred to the Appendix, where we also give some necessary background material, as well as a list of other new space–time covariance functions that can be used on spheres cross time.

2 BACKGROUND

2.1 Spatial Fields on Spheres, Coordinates and Metrics

We consider the unit sphere $\mathbb{S}^2$, defined as $\mathbb{S}^2 = \{\boldsymbol{s} \in \mathbb{R}^3, \|\boldsymbol{s}\| = 1\}$, where $\|\cdot\|$ denotes Euclidean distance. Every point $\boldsymbol{s}$ on the sphere $\mathbb{S}^2$ has spherical coordinates $\boldsymbol{s} = (\phi, \vartheta)$, with $\phi \in [0, \pi]$ and $\vartheta \in [0, 2\pi)$ being respectively the polar and the azimuthal angles (equivalently, latitude and longitude). The extension to a sphere with an arbitrary radius is straightforward. For planet Earth, the radius is approximately 6 371 km, although Earth is not exactly a sphere.

The natural distance on the sphere is the geodesic or great circle distance, defined as the mapping $d_{\mathrm{GC}} : \mathbb{S}^2 \times \mathbb{S}^2 \to [0, \pi]$ so that

$$d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \arccos\left(\langle \boldsymbol{s}_1, \boldsymbol{s}_2 \rangle\right) = \arccos\left(\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos\vartheta\right),$$

with $\boldsymbol{s}_i = (\phi_i, \vartheta_i)$, $i = 1, 2$, and $\langle \cdot, \cdot \rangle$ denoting the classical dot product on the sphere, and where $\vartheta = |\vartheta_1 - \vartheta_2|$. Thus, the geodesic distance describes an arc between any pair of points located on the spherical shell. Throughout, we shall use $d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and d_{GC} interchangeably to denote the geodesic distance, whenever no confusion can arise.

An approximation of the true distance between any two points on the sphere is the chordal distance $d_{\mathrm{CH}}(\boldsymbol{s}_1, \boldsymbol{s}_2)$, given by

$$d_{\mathrm{CH}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = 2 \sin\left(\frac{d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2)}{2}\right), \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}^2.$$

The chordal distance defines a segment below the arc joining two points on the spherical shell.

We consider zero mean Gaussian fields $\{X(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{S}^2\}$ with finite second-order moments. Thus, the finite dimensional distributions are completely specified by the covariance function $C_{\mathcal{S}} : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$, defined by

$$C_{\mathcal{S}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{C}\mathrm{ov}\left(X(\boldsymbol{s}_1), X(\boldsymbol{s}_2)\right), \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}^2.$$

Covariance functions are positive definite: for any N dimensional collection of points $\{\boldsymbol{s}_i\}_{i=1}^N \subset \mathbb{S}^2$ and constants $c_1, \ldots, c_N \in \mathbb{R}$, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i C_{\mathcal{S}}\left(\boldsymbol{s}_i, \boldsymbol{s}_j\right) c_j \geq 0, \tag{2.1}$$

see Bingham (1973). We also define the variogram $\gamma_{\mathcal{S}}$ of X on $\mathbb{S}^2$ as half the variance of the increments of X at the given points on the sphere, namely,

$$2\gamma_{\mathcal{S}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{V}\mathrm{ar}\left(X(\boldsymbol{s}_2) - X(\boldsymbol{s}_1)\right), \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}^2.$$

For a discussion about variograms on spheres, the reader is referred to Huang *et al.* (2012) and Gneiting (2013) and the references therein.

The simplest process in the Euclidean framework is the isotropic process, that is, a process that does not depend on a particular direction, but just on the distance between points. We say that $C_{\mathcal{S}}$ is geodesically isotropic if $C_{\mathcal{S}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \psi_{\mathcal{S}}(d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2))$, for some $\psi_{\mathcal{S}} : [0, \pi] \to \mathbb{R}$. $\psi_{\mathcal{S}}$ is called the geodesically isotropic part of $C_{\mathcal{S}}$ (Daley & Porcu, 2013). Henceforth, we shall refer to both $C_{\mathcal{S}}$ and $\psi_{\mathcal{S}}$ as covariance functions, in order to simplify exposition. For a characterisation of geodesic isotropy, the reader is referred to Schoenberg (1942) and the essay in Gneiting (2013). The definition of a geodesically isotropic variogram is analogous.

While a geodesically isotropic process on $\mathbb{S}^2$ is the natural counterpart to an isotropic process in Euclidean space, it is not necessarily an appropriate process for globally referenced data. Although it may be argued that on a sufficiently small scale atmospheric phenomena lack any structured flow, this does not apply in general for synoptic or mesoscale processes such as prevailing winds, which follow regular patterns dictated by atmospheric circulation.

As a first approximation for large-scale atmospheric phenomena, a process may be assumed to be non-stationary for different latitudes, but stationary for the same longitude. Indeed, it is expected that physical quantities such as surface temperature will display an interannual variability that is lower in the tropics than at mid-latitude. Therefore, we define the covariance $C_{\mathcal{S}}$ to be *axially symmetric* if

$$C_{\mathcal{S}}\left(\boldsymbol{s}_1, \boldsymbol{s}_2\right) = \mathcal{C}_{\mathcal{S}}(\phi_1, \phi_2, \vartheta_1 - \vartheta_2), \qquad (\phi_i, \vartheta_i) \in [0, \pi] \times [0, 2\pi), i = 1, 2. \tag{2.2}$$

Additionally, an axially symmetric Gaussian field $X(\boldsymbol{s})$ is called longitudinally reversible if

$$\mathcal{C}_{\mathcal{S}}\left(\phi_1, \phi_2, \vartheta\right) = \mathcal{C}_{\mathcal{S}}\left(\phi_1, \phi_2, -\vartheta\right), \qquad \phi_i \in [0, \pi], \vartheta \in [-2\pi, 2\pi), i = 1, 2. \tag{2.3}$$

An alternative notion of isotropy can be introduced if we assume that the sphere $\mathbb{S}^2$ is embedded in the three-dimensional Euclidean space $\mathbb{R}^3$ and that the Gaussian process X is defined on $\mathbb{R}^3$. The covariance $C_{\mathcal{S}}$ can then be defined by restricting X on $\mathbb{S}^2$, which gives rise to the name of Euclidean isotropy or radial symmetry, because it depends exclusively on the chordal distance (d_{CH}) between the points. Following Yadrenko (1983) and Yaglom (1987), for any $C_{\mathcal{S}}$ being isotropic in the Euclidean sense in $\mathbb{R}^3$, the function $C_{\mathcal{S}}(d_{\mathrm{CH}})$ is a valid covariance function on $\mathbb{S}^2$. This principle has been used to create space–time and multivariate covariance functions, and we come back to this in detail in Section 3.5.

2.2 Space–Time Random Fields and Covariance Functions

We now describe zero mean Gaussian fields $\{Z(\boldsymbol{s}, t), \boldsymbol{s} \in \mathbb{S}^2, t \in \mathbb{R}\}$ evolving temporally over the sphere $\mathbb{S}^2$. Henceforth, we assume that Z has a finite second-order moment.

Given the Gaussianity assumption, we focus on the covariance function $C : \left(\mathbb{S}^2 \times \mathbb{R}\right)^2 \to \mathbb{R}$, defined as

$$C\left((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)\right) = \mathbb{C}\mathrm{ov}\left(Z(\boldsymbol{s}_1, t_1), Z(\boldsymbol{s}_2, t_2)\right), \qquad (\boldsymbol{s}_i, t_i) \in \mathbb{S}^2 \times \mathbb{R}, i = 1, 2. \tag{2.4}$$

The definition of positive definiteness is analogue to that given in Equation 2.1.

The covariance C is called separable (Gneiting *et al.*, 2007) if there exists two mappings $C_{\mathcal{S}} : (\mathbb{S}^2)^2 \to \mathbb{R}$ and $C_{\mathcal{T}} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$C\left((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)\right) = C_{\mathcal{S}}(\boldsymbol{s}_1, \boldsymbol{s}_2) C_{\mathcal{T}}(t_1, t_2), \qquad (\boldsymbol{s}_i, t_i) \in \mathbb{S}^2 \times \mathbb{R}, \quad i = 1, 2. \tag{2.5}$$

Separability can be desirable from a computational standpoint: for a given set of colocated observations (i.e. when for every instant t, the same spatial sites have available observations), the related covariance matrix factorises into the Kronecker product of two smaller covariance matrices. However, it has been deemed physically unrealistic, as the degree of spatial correlation is the same at points near or far from the origin in time (Gneiting *et al.*, 2007); a constructive criticism is offered in Stein (2005a). There is a rich literature on non-separable covariance functions defined on Euclidean spaces; see the reviews in Gneiting *et al.* (2007), Mateu *et al.* (2008) and Banerjee *et al.* (2014), with the references therein.

2.3 Temporally Evolving Geodesic Isotropies and Axial Symmetries

A common assumption for space–time covariance functions over spheres cross time is that of geodesic isotropy coupled with temporal symmetry: there exists a mapping $\psi : [0, \pi] \times \mathbb{R} \to \mathbb{R}$ such that

$$C\left((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)\right) = \psi\left(d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2), t_1 - t_2\right), \qquad (\boldsymbol{s}_i, t_i) \in \mathbb{S}^2 \times \mathbb{R}, \quad i = 1, 2. \tag{2.6}$$

Following Berg and Porcu (2017), we call $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$ the class of continuous functions ψ such that $\psi(0, 0) = \sigma^2 < \infty$ and the identity (2.6) holds. The functions ψ are called the geodesically isotropic and temporally symmetric parts of the space–time covariance functions C. Also, we refer equivalently to C or ψ as covariance functions, in order to simplify the exposition. In Appendix A, we introduce the more general class $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$ and show many relevant facts about it.

Equation 2.6 can be generalised to the case of temporal non-stationarity, and the reader is referred to Estrade *et al.* (2016) for a mathematical approach to this problem.

As in the purely spatial case, isotropic models are seldom used in practical applications as they are deemed overly simplistic. Yet they can serve as building blocks for more sophisticated models that can account for local anisotropies and non-stationarities.

We now couple spatial axial symmetry with temporal stationarity, so that, for the covariance C in Equation 2.4, there exists a continuous mapping $\mathcal{C} : [0, \pi]^2 \times [-2\pi, 2\pi] \times \mathbb{R} \to \mathbb{R}$ such that

$$C\left((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)\right) = \mathcal{C}(\phi_1, \phi_2, \vartheta_1 - \vartheta_2, t_1 - t_2), \qquad (\phi_i, \vartheta_i, t_i) \in [0, \pi] \times [0, 2\pi] \times \mathbb{R},$$

$i = 1, 2$. Additionally, we call a temporally stationary-spatially axially symmetric random field $Z(\boldsymbol{s}, t)$ longitudinally reversible if

$$\mathcal{C}(\phi_1, \phi_2, \vartheta, u) = \mathcal{C}(\phi_1, \phi_2, -\vartheta, u), \qquad \phi_i \in [0, \pi], \vartheta \in [-2\pi, 2\pi), u \in \mathbb{R}, i = 1, 2. \tag{2.7}$$

The use of statistical models based on axially symmetric and longitudinal reversible stochastic processes on the sphere is advocated in Stein (2007) for the analysis of total column ozone.

We show throughout the paper that the construction in Equation 2.7 is especially important for implementing dynamical models as described in Section 4.1, where a solely spatial version $\mathcal{C}_{\mathcal{S}}(\phi_1, \phi_2, \vartheta)$ is used. Note that an axially symmetric covariance function $\mathcal{C}$ might be separable, in which case

$$\mathcal{C}(\phi_1, \phi_2, \vartheta, u) = \mathcal{C}_{\mathcal{S}}(\phi_1, \phi_2, \vartheta)\mathcal{C}_{\mathcal{T}}(u), \qquad (\phi_i, \vartheta, u) \in [0, \pi] \times [-2\pi, 2\pi) \times \mathbb{R},$$

$i = 1, 2$, with $\mathcal{C}_{\mathcal{S}}$ and $\mathcal{C}_{\mathcal{T}}$ being marginal covariances in their respective spaces.

3 Second-Order Approaches

In this section, we provide a list of techniques that are used in the literature to implement space–time covariance functions on the two-dimensional sphere cross time.

3.1 Spectral Representations and Related Stochastic Expansions

Spectral representations in Euclidean spaces have been known since the work of Schoenberg (1938) and extended to space–time in Cressie and Huang (1999) and then in Gneiting (2002). The analogue of spectral representations over spheres was then provided by Schoenberg (1942). The case of the sphere cross time has been unknown until the recent work of Berg and Porcu (2017): finding a spectral representation for geodesically isotropic space–time covariance functions is equivalent to providing a characterisation of the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$. Namely, Berg and Porcu (2017) establish that a continuous mapping ψ is a member of the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$ if and only if

$$\psi(d_{\mathrm{GC}}, u) = \sum_{k=0}^{\infty} C_{k,\mathcal{T}}(u) P_k(\cos d_{\mathrm{GC}}), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}, \tag{3.1}$$

where $\{C_{k,\mathcal{T}}(\cdot)\}_{k=0}^{\infty}$ is a sequence of temporal covariance functions with the additional requirement that $\sum_{k=0}^{\infty} C_{k,\mathcal{T}}(0) < \infty$ in order to guarantee the variance $\sigma^2 = \psi(0, 0)$ to be finite. Here, $P_k(x)$ denotes the k-th Legendre polynomial, $x \in [-1, 1]$ (see Dai & Xu, 2013, for more details). We refer to (3.1) as the spectral representation of the space–time covariance $\psi(d_{\mathrm{GC}}, u)$.

Berg and Porcu (2017) showed a general representation for the case of the n-dimensional sphere $\mathbb{S}^n$ cross time (see Appendix A for details). This fact is not merely a mathematical artefact but also the key for modeling strategies, as shown in Porcu *et al.* (2016a).

Some comments are in order. Clearly, $\psi_{\mathcal{S}}(d_{\mathrm{GC}}) = \psi(d_{\mathrm{GC}}, 0)$ is the geodesically isotropic part of a spatial covariance defined over the sphere, a characterisation of which can be found in the notable work by Schoenberg (1942); see also the recent review by Gneiting (2013). Furthermore, representation (2.6) implies that ψ is separable if and only if the elements $C_{k,\mathcal{T}}$ of the sequence $\{C_{k,\mathcal{T}}(\cdot)\}_{k=0}^{\infty}$ in Equation 3.1 are of the form

$$C_{k,\mathcal{T}}(u) = b_k \, \rho_{\mathcal{T}}(u), \qquad u \in \mathbb{R}, \quad k \in \{0, 1, \ldots\},$$

with $\{b_k\}_{k=0}^{\infty}$ being a uniquely determined probability sequence, and $\rho_{\mathcal{T}}$ a temporal covariance function. We follow Daley and Porcu (2013) and call $\{b_k\}_{k=0}^{\infty}$ a 2-Schoenberg sequence to emphasise that the coefficients b_k depend on the dimensions of the two-dimensional sphere where ψ is defined.

It can be proved that the covariance functions from the series expansion in (3.1) have a corresponding Gaussian process with an associated series expansion as well. Indeed, direct inspection together with the addition theorem for spherical harmonics (Marinucci & Peccati,

2011) shows that (3.1) corresponds to a Gaussian process on $\mathbb{S}^2 \times \mathbb{R}$, defined by the stochastic expansion of Jones (1963):

$$Z(\boldsymbol{s}, t) = \sum_{k=0}^{\infty} \sum_{\ell=-k}^{k} A_{k,\ell}(t) \mathcal{Y}_{k,\ell}(\boldsymbol{s}), \qquad (\boldsymbol{s}, t) \in \mathbb{S}^2 \times \mathbb{R}, \tag{3.2}$$

where $\mathcal{Y}_{k,\ell}$ are the deterministic spherical harmonics on $\mathbb{S}^2$ (Dai & Xu, 2013), and each of the zero mean Gaussian processes $\{A_{k,\ell}(t)\}_{k=0,\ell=-k}^{\infty,\ k}$ satisfies

$$\mathbb{E}\left(A_{k,\ell}(t) A_{k',\ell'}(t')\right) = \delta_{k=k'} \delta_{\ell=\ell'} C_{k,\mathcal{T}}(t - t'), \qquad t, t' \in \mathbb{R}, \tag{3.3}$$

with δ denoting the classical Kronecker delta and $\{C_{k,\mathcal{T}}\}_{k=0}^{\infty}$ a sequence of temporal covariances summable at zero.

Unfortunately, the representation does not allow for many interesting closed form examples. Hence, Porcu *et al.* (2016a) propose a spectral representation valid on any n-dimensional sphere cross $\mathbb{R}$, namely,

$$\psi(d_{\mathrm{GC}}, u) = \sum_{k=0}^{\infty} C_{\mathcal{T}}(u)^k \cos(d_{\mathrm{GC}})^k, \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}, \tag{3.4}$$

where $C_{\mathcal{T}} : \mathbb{R} \to [0, 1]$ is a temporal correlation function. Some examples from Porcu *et al.* (2016a) are reported in Table 1. Details are explained in Appendix A.

Clarke *et al.* (2018) take advantage of stochastic representation (3.2) to study the regularity properties of a Gaussian process Z on $\mathbb{S}^2 \times \mathbb{R}$, in terms of dynamical fractal dimensions and Hölder exponents. This is achieved by taking a double Karhunen–Loève expansion, for example, by expanding each term $A_{k,\ell}(t)$ in terms of basis functions (in particular, Hermite polynomials). Such a double Karhunen–Loève representation is also the key for fast simulation. Clarke *et al.* (2018) propose truncating the order of the double expansions and evaluating the error bound in the L_2 sense.

Note how in each of the parametric families outlined in Table 1, the spatial margin $\psi_{\mathcal{S}}(d_{\mathrm{GC}}) = \psi(d_{\mathrm{GC}}, 0)$ is an analytic function. In particular, we have that the spatial margin is either non-differentiable or infinitely differentiable at the origin.

Table 1. *Parametric families of covariance functions on* $\mathbb{S}^2 \times \mathbb{R}$ *obtained through the representation in Equation 3.4.*

Family	Analytic Expression	Parameter Range
Negative Binomial	$\psi(r, u) = \left(\frac{1-\varepsilon}{1-\varepsilon g(u)\cos(r)}\right)^{\tau}$	$\varepsilon \in (0,1), \tau > 0$
Multiquadric	$\psi(r, u) = \left(\frac{(1-\varepsilon)^2}{1+\varepsilon^2-2\varepsilon g(u)\cos(r)}\right)^{\tau}$	$\varepsilon \in (0,1), \tau > 0$
Sine Series	$\psi(r, u) = \frac{1}{2} \mathrm{e}^{g(u)\cos(r)-1}\left(1 + g(u)\cos(r)\right)$	
Sine Power	$\psi(r, u) = 1 - 2^{-\alpha}\left(1 - g(u)\cos(r)\right)^{\alpha/2}$	$\alpha \in (0, 2]$
Adapted Multiquadric	$\psi(r, u) = \left(\frac{(1+g^2(u))(1-\varepsilon)}{1+g^2(u)-2\varepsilon g(u)\cos(r)}\right)^{\tau}$	$\varepsilon \in (0,1), \tau > 0$
Poisson	$\psi(r, u) = \exp\left(\lambda(\cos(r) g(u) - 1)\right)$	$\lambda > 0$

The second column reports the analytic expression, where g is any correlation function on the real line. An additional condition is required for the Sine Power family (refer to Porcu *et al.*, 2016a for details). All of the members ψ in the second column are rescaled so that $\psi(0, 0) = 1$. We use the abuse of notation r for the great circle distance d_{GC}.

Spectral representations for stochastic processes over spheres, with the additional feature of axial symmetry, have been proposed in Jones (1963), Narcowich (1995), Hitczenko and Stein (2012) and Castruccio and Stein (2013). We are not aware of extensions of this representation to space–time, but the work in Jones (1963) suggests that axial symmetry over the sphere coupled with temporal symmetry should be obtained by relaxing assumption (3.3) for the processes $A_{k,\ell}$ in expansion (3.2).

An alternative approach to modeling axially symmetric processes over spheres cross time is proposed in Castruccio and Guinness (2017), which is based on the stochastic representations in Jones (1963).

3.2 Scale Mixture Representations

Scale mixtures provide a powerful and elegant way to build members of the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$. We sketch the general principle here and then report some examples. Let F be a positive measure on the positive real line. Let $\psi_{\mathcal{S}}(d_{\mathrm{GC}}; \xi)$ be a geodesically isotropic spatial covariance for any $\xi \in \mathbb{R}_+$. Also, let $C_{\mathcal{T}}(u; \xi)$ be a covariance for any ξ. Then, arguments in Porcu and Zastavnyi (2011) and Gneiting *et al.* (2007) show that the function

$$\psi(d_{\mathrm{GC}}, u) = \int_{\mathbb{R}_+} \psi_{\mathcal{S}}(d_{\mathrm{GC}}; \xi) C_{\mathcal{T}}(u; \xi) F(\mathrm{d}\xi), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}$$

is a non-separable geodesically isotropic and temporally symmetric space–time covariance function.

The scale mixture approach offers a nice interpretation in terms of the construction of the associated process. Let Ξ be a random variable with probability distribution F, which has a finite first moment. Let X and Y be a purely spatial and a purely temporal Gaussian process, respectively. Also, suppose that the random variable Ξ and the two processes are mutually independent. Let

$$Z(\boldsymbol{s}, t \mid \Xi = \xi) = X(\boldsymbol{s}\xi) Y(t\xi), \qquad (\boldsymbol{s}, t) \in \mathbb{S}^2 \times \mathbb{R},$$

where ξ is a realisation from Ξ and let $Z(\boldsymbol{s}, t) = \mathbb{E} Z(\boldsymbol{s}, t \mid \Xi)$, with expectation $\mathbb{E}$ taken with respect to F. Then, the covariance of Z admits a scale mixture representation.

One might be tempted to believe that all classes of non-separable covariance functions based on scale mixtures and defined over $\mathbb{R}^3 \times \mathbb{R}$ can be adapted to the case $\mathbb{S}^2 \times \mathbb{R}$. This is, in general, not true. For instance, Gaussian scale mixtures as proposed in Schlather (2010) cannot be implemented on the sphere because the function $C_{\mathcal{S}}(d_{\mathrm{GC}}; \xi) = \exp(-(d_{\mathrm{GC}}/\xi)^{\alpha})$ is positive definite on the sphere only when $\alpha \in (0, 1]$ (see Gneiting, 2013). Some caution needs be taken when considering the scale mixture-based covariances defined on Euclidean spaces by Porcu *et al.* (2007); Porcu and Mateu (2007) and Fonseca and Steel (2011), amongst others.

To illustrate some valid examples, further notation is needed. A function $f : [0, \infty) \to (0, \infty)$ is called completely monotonic if it is infinitely differentiable on the positive real line and if its j-th order derivatives $f^{(j)}$ satisfy $(-1)^j f^{(j)}(t) \geq 0$, $t > 0$, $j = 0, 1, 2, \ldots$. A scale mixture argument in Porcu *et al.* (2016a) shows that the function

$$\psi(d_{\mathrm{GC}}, u) = \frac{\sigma^2}{\gamma_{\mathcal{T}}(u)^3} f\left(d_{\mathrm{GC}} \gamma_{\mathcal{T}}(u)\right), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R} \tag{3.5}$$

is a non-separable covariance function on $\mathbb{S}^2 \times \mathbb{R}$ provided f is completely monotonic, with $f(0) = 1$, and $\gamma_{\mathcal{T}}$ is a strictly positive variogram, with $\gamma_{\mathcal{T}}(0) = 1$. Here, σ^2 denotes the

variance. We consider a special case of Equation 3.5, namely,

$$\psi(d_{\mathrm{GC}}, u) = \frac{\sigma^2}{\left(1 + \frac{|u|}{b_T}\right)^3} \exp\left(-\frac{d_{\mathrm{GC}}\left(1 + \frac{|u|}{b_T}\right)}{b_S}\right), \tag{3.6}$$

which corresponds to $f(t) = \exp(-t)$, $t \geq 0$, with $\gamma_T(u) = (1 + |u|)$, $u \in \mathbb{R}$. Here, σ^2, b_S and b_T are positive parameters associated with the variance, spatial scale and temporal scale of the field, respectively. Figure 1 shows a realisation using $\sigma^2 = 1$, $b_S = 0.3$ and $b_T = 0.5$.

Another scale mixture approach allows for an adaptation of the Gneiting class (Gneiting, 2002; Zastavnyi & Porcu, 2011). Again, Porcu *et al.* (2016a) show how to derive a space–time covariance of the type

$$\psi(d_{\mathrm{GC}}, u) = \frac{\sigma^2}{g_{[0,\pi]}(d_{\mathrm{GC}})^{1/2}} f\left(\frac{u}{g_{[0,\pi]}(d_{\mathrm{GC}})}\right), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}. \tag{3.7}$$

Here, f is completely monotonic, with $f(0) = 1$, and $g_{[0,\pi]}$ is the restriction to the interval $[0, \pi]$ of a function, $g : [0, \infty) \to \mathbb{R}_+$, having a completely monotonic derivative (see, Porcu and Schilling, 2011, for a description with classes of functions having this property).

We now outline a new result within the scale mixture-based approach. Specifically, we consider quasi-arithmetic means, as defined in Porcu *et al.* (2009): these allow one to obtain space–time covariances with given margins in space and time. The construction is defined as

$$\psi(d_{\mathrm{GC}}, u) = \mathcal{Q}_f\left(\psi_{\mathcal{S}}(d_{\mathrm{GC}}), C_{\mathcal{T}}(u)\right), \qquad (d_{\mathrm{GC}}, u) \in \mathbb{S}^2 \times \mathbb{R}, \tag{3.8}$$

where $\mathcal{Q}_f(x, y) = f\left(1/2 f^{-1}(x) + 1/2 f^{-1}(y)\right)$ for a function f that is completely monotonic on the positive real line and has a proper inverse, f^{-1}, which is always well defined because completely monotonic functions are strictly decreasing. There is a wide list of such functions available, and the reader is referred to Schilling *et al.* (2012) and to Porcu and Schilling (2011). Quasi-arithmetic means include all the other means (e.g. the geometric, harmonic and the Gini means) as special cases, and the reader is referred to Porcu *et al.* (2009) for a complete description of this framework.

Using the same arguments as in Porcu *et al.* (2009), Theorem 6.3 in Appendix C shows the conditions for which ψ in Equation 3.8 is a geodesically isotropic space–time covariance function.

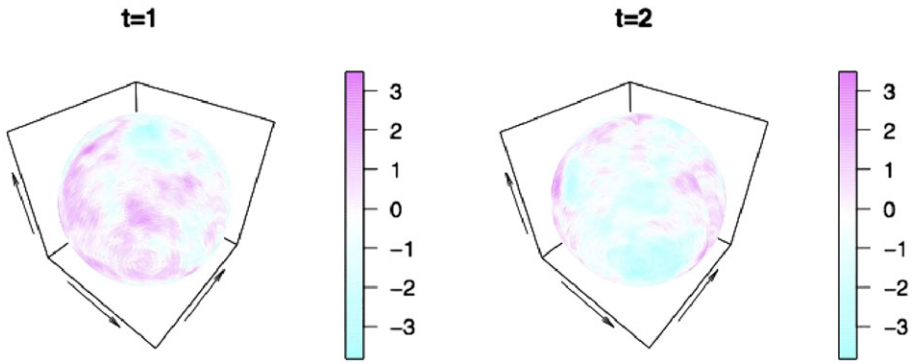

Figure 1. *Simulated space–time data from covariance* (3.6) *for over 17 000 spatial sites on* $\mathbb{S}^2$ *and two temporal instants. [Colour figure can be viewed at wileyonlinelibrary.com]*

Table 2. *Wendland correlation $\varphi_{\mu,k}(\cdot)$ and Matérn correlation $\mathcal{M}_\nu(\cdot)$ with increasing smoothness parameters k and ν.*

k	$\varphi_{\mu,k}(r)$	ν	$\mathcal{M}_\nu(r)$	$SP(k)$
0	$(1-r)_+^{\mu}$	0.5	e^{-r}	0
1	$(1-r)_+^{\mu+1}(1+r(\mu+1))$	1.5	$e^{-r}(1+r)$	1
2	$(1-r)_+^{\mu+2}\left(1+r(\mu+2)+r^2(\mu^2+4\mu+3)\frac{1}{3}\right)$	2.5	$e^{-r}\left(1+r+\frac{r^2}{3}\right)$	2
3	$(1-r)_+^{\mu+3}\left(1+r(\mu+3)+r^2(2\mu^2+12\mu+15)\frac{1}{5}\right.$ $\left.+r^3(\mu^3+9\mu^2+23\mu+15)\frac{1}{15}\right)$	3.5	$e^{-r}\left(1+\frac{r}{2}+r^2\frac{6}{15}+\frac{r^3}{15}\right)$	3

$SP(k)$ means that the sample paths of the associated Gaussian field are k times differentiable.

3.3 Space–Time Compact Supports

Let ψ be a member of the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$. We say that ψ is dynamically compactly supported on the sphere if there exists a function $h : \mathbb{R} \to (0, c]$, with $c \leq \pi$, such that for every fixed temporal lag u_0, the function $\psi_{\mathcal{S}}(d_{\mathrm{GC}}; u_0) = \psi(d_{\mathrm{GC}}, u_0)$ is a covariance function on $\mathbb{S}^2$ and is compactly supported on the interval $[0, h(u_0))$.

Let $(x)_+$ denote the positive part of a real number x. For $\mu > 0$ and $k = 0, 1, 2, \ldots$, we define Wendland functions $\varphi_{\mu,k}$ (Wendland, 1995) by

$$\varphi_{\mu,k}(x) = (1-x)_+^{\mu+k}\,\mathbb{P}_k(x), \qquad x \geq 0, \tag{3.9}$$

where $\mathbb{P}_k$ is a polynomial of order k. Special cases are depicted in the second column of Table 2.

We provide new results in the remainder of this subsection. Using scale mixture arguments, we can show that

$$\psi(d_{\mathrm{GC}}, u) = \frac{\sigma^2}{c^\alpha} h(|u|)^\alpha \varphi_{\mu,0}\left(\frac{d_{\mathrm{GC}}}{h(|u|)}\right) = \frac{\sigma^2}{c^\alpha} h(|u|)^\alpha \left(1 - \frac{d_{\mathrm{GC}}}{h(|u|)}\right)_+^{\mu}, \tag{3.10}$$

$(d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}$, for $\alpha \geq 3$ and $\mu \geq 4$, is a covariance function on $\mathbb{S}^2 \times \mathbb{R}$ provided h is positive, decreasing and convex on the positive real line, with $h(0) = c$, $0 < c \leq \pi$, and $\lim_{t\to\infty} h(t) = 0$. The additional technical restriction on μ and α is explained in Theorem 6.4 in Appendix C, where a formal proof is given. An example can better clarify things. A potential candidate that can be used as dynamical support in (3.10) is the function $h(t) = c(1+t)^{-1/\alpha}$, $t \geq 0$, $\alpha \geq 3$ and $0 < c \leq \pi$. Then, the previous construction becomes

$$\psi(d_{\mathrm{GC}}, u) = \frac{\sigma^2}{(1+|u|)}\left(1 - \frac{d_{\mathrm{GC}}}{c(1+|u|)^{-1/\alpha}}\right)_+^{\mu}, \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}, \tag{3.11}$$

which shows that the spatial margin $\psi(d_{\mathrm{GC}}, 0)$ has compact support c. If $c = \pi$, then $\psi(d_{\mathrm{GC}}, 0)$ becomes globally supported. Note how the compact support becomes smaller as the temporal lag increases, which is a very intuitive property.

The following result characterises a class of dynamically supported Wendland functions on the sphere cross time, being a generalisation of (3.10) to arbitrary $k \in \mathbb{N}$.

Theorem 3.1. *Let $0 < c \leq \pi$. Let $h : [0, \infty) \to (0, c]$ be positive, decreasing and convex on the positive real line, with $\lim_{t\to\infty} h(t) = 0$. Let k be a positive integer and $\alpha \geq 2k + 2$. Let $\varphi_{\mu,k}$ be the Wendland function defined in Equation 3.9. Then,*

$$\psi(d_{GC}, u) = \sigma^2 h(|u|)^\alpha \varphi_{\mu,k}\left(\frac{d_{GC}}{h(|u|)}\right), \qquad (d_{GC}, u) \in [0, \pi] \times \mathbb{R} \tag{3.12}$$

is a covariance function on the sphere cross time, provided $\mu \geq k + 4$.

A more general statement involves compactly supported space–time covariance functions on the circle, where no parametric forms are imposed on the involved functions.

Theorem 3.2. *Let $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a covariance function that is symmetric in both the first and second arguments, such that $\varphi(x, u) = 0$ whenever $|x| \geq \pi$, for all $u \in \mathbb{R}$. Call ψ the restriction of φ to the interval $[0, \pi]$ with respect to the first argument. Then, ψ is a geodesically isotropic and temporally symmetric covariance function on the circle $\mathbb{S}^1$ cross time $\mathbb{R}$.*

The earlier result is unfortunately insufficient for building compactly supported models over spheres cross time. Theorem 3.1 offers a special case with dynamically supported Wendland functions. To obtain a general assertion on the basis of Theorem 3.2, we give some hints in the research problems at the end of the manuscript.

3.4 Covariance Models under the Lagrangian framework

Environmental, atmospheric and geophysical processes are often influenced by prevailing winds or ocean currents (Gneiting *et al.*, 2007). In this type of situation, the general idea of a Lagrangian reference frame applies. Alegria and Porcu (2017) considered a simple Lagrangian framework that allows for transport effects over spheres. Namely, they consider random orthogonal (3×3) matrices $\mathcal{R}$ with determinants identically equal to one, such that $\mathcal{R}^{-1} = \mathcal{R}^{\top}$, with $\top$ denoting the transpose operator. Standard theory on random orthogonal rotations shows that

$$\mathcal{R} = QDQ^{-1},$$

with Q denoting a matrix containing the eigenvectors of $\mathcal{R}$, and D a diagonal matrix containing the associated eigenvalues. Also, we have that each eigenvalue λ_k, $k = 1, 2, 3$, can be uniquely written as $\lambda_k = \exp(\imath\kappa_k)$, with $\imath$ being the unit imaginary number and κ_k real, for $k = 1, 2, 3$. Then, following Gantmacher (1960), one can define the t-th real power of $\mathcal{R}$, as

$$\mathcal{R}^t = Q\left(\operatorname{diag}\left(\exp(\imath\kappa_k t)\right)\right) Q^{-1}.$$

Let X be a Gaussian process on $\mathbb{S}^2$ with geodesically isotropic covariance $C_{\mathcal{S}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \psi_{\mathcal{S}}(d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2))$. Define

$$Z(\boldsymbol{s}, t) = X\left(\mathcal{R}^t \boldsymbol{s}\right), \qquad \boldsymbol{s} \in \mathbb{S}^2, t \in \mathbb{R}. \tag{3.13}$$

Then, the space–time covariance function with transport effect can be expressed as

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2, u) = \mathbb{E}\left(C_{\mathcal{S}}(d_{\mathrm{GC}}(\mathcal{R}^u \boldsymbol{s}_1, \boldsymbol{s}_2))\right), \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}^2, \quad u \in \mathbb{R}, \tag{3.14}$$

where the expectation is taken with respect to the random rotation $\mathcal{R}$. The fact that the resulting covariance is still geodesically isotropic in the spatial component is non-trivial. It is shown formally, for some specific choice of the random rotation $\mathcal{R}$, in Alegria and Porcu (2017), at least for the case of the sphere $\mathbb{S}^2$. Some comments are in order. The resulting field Z in Equation 3.13 is not Gaussian (it is Gaussian conditionally on $\mathcal{R}$ only). Also, obtaining closed forms for the associated covariance is generally difficult. Alegria and Porcu (2017) provide some special cases.

For the specification of the random rotation $\mathcal{R}$ in the Lagrangian covariance (3.14), various choices can be physically motivated and justified. The reader is referred to Gneiting *et al.* (2007) for a thorough discussion. For instance, the random rotation might represent a prevailing wind as in Gupta and Waymire (1987), it might be the westerly wind considered by Haslett

and Raftery (1989), or it might be updated dynamically according to the current state of the atmosphere. Of course, the model in Equation 3.14 represents only a first step into Lagrangian modeling over spheres. The concept of transport effects should be put into a broader context, for example, by following the discussion on pages 319–320 of Cressie and Wikle (2011), which shows that the Lagrangian model offered here is a very special case, obtained when the spatial field is moving at a constant velocity. Innovative approaches to space–time data under transport effects can be found in Wikle *et al.* (2001) and Cressie *et al.* (2010).

The work by Kent *et al.* (2011) has brought attention to the so-called dimple problem in Euclidean spaces: for some classes of space–time covariance functions, a dimple is present if $Z(\boldsymbol{s}_{\texttt{here}}, t_{\texttt{now}})$ is more correlated with $Z(\boldsymbol{s}_{\texttt{there}}, t_{\texttt{tomorrow}})$ than with $Z(\boldsymbol{s}_{\texttt{there}}, t_{\texttt{now}})$. The authors establish conditions for the presence of the dimple in the Gneiting covariance (Gneiting, 2002) and argue that the dimple is a counterintuitive property for modeling space–time data because it contradicts a natural monotonicity assumption of the covariance. Additional works related to the dimple effect are Cuevas *et al.* (2017). The formulation of the dimple problem on the sphere cross time is due to Alegria and Porcu (2017). Further, the authors show conditions for which there might be a dimple in the transport effect covariance in Equation 3.14.

3.5 Covariance Models based on Chordal Distances

On the basis of the results in Yadrenko (1983) and Yaglom (1987), Jun and Stein (2007, 2008) exploit the fact that, for any covariance function $C_{\mathcal{S}}$ on $\mathbb{R}^3$, the function $C_{\mathcal{S}}(d_{\mathrm{CH}})$ is a covariance function on $\mathbb{S}^2$. The Matérn function (e.g. Stein, 1999) is defined as

$$\mathcal{M}_\nu(x) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} x^\nu \mathcal{K}_\nu(x), \qquad x \geq 0, \tag{3.15}$$

where $\nu > 0$ governs the mean square differentiability of the associated process Z on $\mathbb{S}^2$, which is k times mean differentiable if and only if $\nu > k$. Here, $\mathcal{K}_\nu$ is a modified Bessel function.

The Matérn model coupled with chordal distance, that is, $\mathcal{M}_\nu(d_{\mathrm{CH}})$, is valid on $\mathbb{S}^2$ for any positive ν. Unfortunately, arguments in Gneiting (2013) show that $\mathcal{M}_\nu(d_{\mathrm{GC}})$ is a valid model on $\mathbb{S}^2$ only for $0 < \nu \leq 1/2$, making its use impractical. Space–time models based on geodesic distance inherit this limitation. The same argument of the Matérn covariance is used in Guinness and Fuentes (2016) and Jeong and Jun (2015) to assert that the chordal distance might be preferable with respect to the great circle distance. For instance, a Matérn-Gneiting (Gneiting, 2002; Porcu & Zastavnyi, 2011) type covariance function based on chordal distance might be easily implemented. For a positive valued temporal variogram $\gamma_{\mathcal{T}}$ with $\gamma_{\mathcal{T}}(0) = 1$, the function

$$C(d_{\mathrm{CH}}, u) = \frac{\sigma^2}{\gamma_{\mathcal{T}}(u)^{3/2}} \mathcal{M}_\nu\left(\frac{d_{\mathrm{CH}}}{\gamma_{\mathcal{T}}(u)}\right), \qquad (d_{\mathrm{CH}}, u) \in [0, 2] \times \mathbb{R} \tag{3.16}$$

is a covariance function on $\mathbb{S}^2 \times \mathbb{R}$ for any positive ν. The spatial margin $C(d_{\mathrm{CH}}, 0)$ is of Matérn type and keeps all the desirable features in terms of differentiability at the origin.

The use of chordal distance has been extensively criticised in the literature. For instance, because the chordal distance underestimates the true distance between the points on the sphere, Porcu *et al.* (2016a) argue that this fact has a non-negligible impact on the estimation of the spatial scale. Moreover, Gneiting (2013) argues that the chordal distance is counter to spherical geometry for larger values of the great circle distance and thus may result in physically unrealistic distortions. Further, covariance functions based on chordal distance inherit the limitations of isotropic models in Euclidean spaces in modeling covariances with negative values. For instance, a covariance based on chordal distance on $\mathbb{S}^2$ does not allow for values lower than

$-0.21\sigma^2$, with σ^2 being the variance as before. Instead, properties of Legendre polynomials (see Szegö, 1939) imply that correlations based on geodesic distance can attain any value between -1 and $+1$. Another argument in favour of the great circle distance is that the differentiability of a given covariance function depending on the great circle distance can be modeled by imposing a given rate of decay of the associated 2-Schoenberg coefficients $\{b_k\}_{k=0}^{\infty}$, as shown in Møller *et al.* (2018) or even by modeling the rate of decay of the associated 2-Schoenberg functions in (3.1), as shown by Clarke *et al.* (2018).

A good alternative to the model in Equation 3.16 is the space–time Wendland model based on the great circle distance, as defined through Equation 3.10. Table 2 highlights a striking connection between the two approaches, showing that Wendland functions allow for a parameterisation of the differentiability at the origin in the same way that Matérn does. The compact support of the Wendland functions can imply some problems in terms of loss of accuracy of kriging predictors, as shown in Stein (1999). However, recent encouraging results (Bevilacqua *et al.*, 2018) show that such a loss is negligible under infill asymptotics.

3.6 Non-stationary and Anisotropic Space–Time Covariance Functions

The literature regarding the construction of non-stationary space–time covariance functions is sparse. The works of Jun and Stein (2007, 2008) are notable exceptions. The methods proposed by the authors are based on the coupling of the chordal distance with certain classes of differential operators. Estrade *et al.* (2016) have generalised the Berg–Porcu (Berg & Porcu, 2017) representation of geodesically isotropic-temporally symmetric covariance functions on $\mathbb{S}^d \times \mathbb{R}$. In particular, two generalisations are obtained: an extension of the Berg–Porcu class to the case of temporally non-stationary covariances and a new class that allows for local anisotropy. Anisotropy is also considered in the tour de force by Hitczenko and Stein (2012), on the basis of chordal distances. Anisotropic components can be induced through the use of Wigner matrices, as explained in Marinucci and Peccati (2011).

More recently, Alegria and Porcu (2016) have considered geodesically isotropic space–time covariance functions that allow the separation of the linear from the cyclical component in the temporal lag. The authors show that such an approach offers considerable gains in terms of predictive performance, particularly in the presence of temporal cyclic components or in the presence of non-stationarities that are normally removed when detrending the data. A recent discussion about non-stationary approaches can be found in Fuglstad *et al.* (2015) who work under the SPDE framework.

4 Practical Approaches

4.1 Dynamical Approaches

When analysing massive data sets arising from, for instance, remotely sensed networks or satellite constellations, climate models or reanalysis data product, a model that fully specifies the covariance function becomes impractical. Indeed, it would require a prohibitive amount of information to be stored in the covariance matrix, as well as a prohibitive number of flops and iterations for maximising the likelihood or exploring the posterior distribution. Hence, a compromise between inferential feasibility and model flexibility must be achieved. These limitations arise independently on the space where the random field is defined. For the case of the sphere, the drawbacks are magnified by the considerably more sparse literature on the construction of valid global space–time processes, as shown in the previous section, as well as by the extremely large size of global data.

In Euclidean spaces, a very popular approach to drastically reducing the complexity is to separate the spatial and temporal components and to describe the dynamics of the process by specifying its evolution as a function of the past. Variability is then achieved by assuming a random spatial innovation. For modeling global climate fields, this approach has been further simplified by modeling the temporal dynamics through covariates only (Furrer *et al.*, 2007; Geinitz *et al.*, 2015).

The dynamical approach has received strong support from reference textbooks in space–time modeling (see, e.g. Cressie & Wikle, 2011), and, in recent years, there has been some further development of this methodology in the context of global space–time data (Castruccio & Stein, 2013). In this regard, the expectation–maximisation algorithm has been efficiently implemented by Fassó *et al.* (2016) and successfully used at continental level for multivariate space–time data (Finazzi & Fassó, 2014). Thanks to the Matérn covariance implemented on the sphere, this code works also at the global level.

Let us assume, with no loss of generality, that the process has zero mean and that it is observed at some locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N \in \mathbb{S}^2$, for equally spaced time points t. We denote the process at time t by $\boldsymbol{Z}_t = (Z(\boldsymbol{s}_1, t), \ldots, Z(\boldsymbol{s}_N, t))^\top$, and we specify its dynamics through the recursive equation

$$\boldsymbol{Z}_t = \mathcal{E}_t(\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_{t-2}, \ldots, \boldsymbol{Z}_{t-p}) + \boldsymbol{\varepsilon}_t, \tag{4.1}$$

where $\mathcal{E}_t$ incorporates the evolution of the past trajectory of the process up to time $t - p$, and $\boldsymbol{\varepsilon}_t = (\varepsilon(\boldsymbol{s}_1, t), \ldots, \varepsilon(\boldsymbol{s}_n, t))^\top \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ is an innovation vector with purely spatial global covariance matrix $\boldsymbol{\Sigma}$.

The considerable benefit of such an approach is that the space–time structure of the model is specified by the dynamical evolution $\mathcal{E}_t$ and the spatial innovation $\boldsymbol{\varepsilon}_t$. Hence, the temporal and spatial part of the model are decoupled, and this allows for a considerably more convenient inference. Recent work on satellite data has proposed coupling the dynamical approach dimension reduction techniques and fixed rank kriging (see Cressie & Johannesson, 2008, Nguyen *et al.*, 2014) in particular to further reduce the parameter dimensionality and to achieve a fit for very large data sets (fixed rank filtering, see Kang *et al.*, 2010; Cressie *et al.*, 2010). While these approaches have been very effective for interpolating large data sets, they do not explicitly account for the spherical geometry but rather project the data onto the Euclidean space. Here, we consider dynamical models for space–time global data with an explicit definition of the covariance function in $\mathbb{S}^2$.

A particularly appealing class of models with (4.1) is the vector autoregressive model VAR(p), defined as

$$\boldsymbol{Z}_t = \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{Z}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{4.2}$$

where $\boldsymbol{\Phi}_j$ are $N \times N$ matrices that encode the temporal dependence at lag $t - j$. This model results in a space–time precision matrix that is block banded and hence would greatly reduce the storage burden arising from massive data sets as shown in Section 5.

4.1.1 The innovation structure

While models such as (4.2) allow for substantial computational savings, in typical climate model applications, the number of locations is larger than 10 000. Thus, even the likelihood of the innovation $\boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ is impossible to evaluate due to memory issues on a laptop. Therefore, additional structure on the model or spatial design of the data must be leveraged in

order to perform feasible inference on the full data set. Castruccio and Stein (2013) consider the Fourier transform cross longitude of an axially symmetric process with a finite number of longitude bands, M, that is,

$$\begin{aligned} &\boldsymbol{\varepsilon}_t(\phi,\vartheta) = \sum_{k=1}^{M} e^{\imath\vartheta k}\mathfrak{f}(k;\phi)\tilde{\varepsilon}_t(k;\phi), \\ &\operatorname{corr}\left(\widetilde{\boldsymbol{\varepsilon}}_t(k;\phi),\widetilde{\boldsymbol{\varepsilon}}_t(k';\phi')\right) = \delta_{k=k'}\,\rho(k;\phi,\phi'), \qquad \phi,\phi' \in [0,\pi], \vartheta \in [0,2\pi), \end{aligned} \tag{4.3}$$

with $\tilde{\varepsilon}_t(k;\phi)$ being the Fourier process for wavenumber k and latitude ϕ. Here, $\mathfrak{f}(k;\phi)$ is the spectrum at latitude ϕ and wavenumber k, and, for any pair of latitudes (ϕ,ϕ'), the function $\rho(k;\phi,\phi')$ defines a spectral correlation (it is also called *coherence*). Jun and Stein (2008) showed that there is a considerable computational benefit if the data are on an $N = N_\phi \times N_\vartheta$ (latitude $\times$ longitude) regular grid over the sphere and if the process is axially symmetric. Indeed, they showed how the resulting covariance matrix is block circulant and, most importantly, block diagonal in the spectral domain, thus requiring only $O(N_\phi \times N_\vartheta^2)$ entries to store it instead of $O(N_\phi^2 \times N_\vartheta^2)$, and $O(N_\phi \times N_\vartheta^3)$ flops instead of $O(N_\phi^3 \times N_\vartheta^3)$. Castruccio and Stein (2013) showed that such structure can be used to perform inference on massive data sets from computer model ensembles (in the range of 10^7 to 10^9 data points) by first estimating the spectrum $\mathfrak{f}(k;\phi)$ for each of the N_ϕ latitudinal bands in parallel, and then conditionally estimating the structure of the spectral correlation $\rho(k;\phi,\phi')$. This approach has been extended to analyse three-dimensional temperature profiles in a regular grid for a data set larger than one billion data points, allowing for an extension of axially symmetric models in three dimensions (Castruccio & Genton, 2016).

The spectral approach is not just a mere strategy to simplify inference but also the key to generalising axially symmetric processes to exhibit non-stationarities. This is done by imposing additional structure in the process transformed in the spectral domain, while still guaranteeing positive definiteness of the corresponding covariance functions. In particular, it is still possible to retain the computational convenience of the gridded geometry while assuming non-stationary models cross longitudes. Castruccio and Genton (2014) explored this idea by assuming that the $\widetilde{\varepsilon}_t(k;\phi)$ in (4.3) are correlated across frequencies, with a fully non-parametric dependence structure to be estimated using time replicates, and they showed how the axially symmetric assumption is badly violated for temperature data.

In order to allow for ocean transitions, Castruccio and Guinness (2017) impose the spectrum $\mathfrak{f}(\cdot;\phi)$ in (4.3) to depend on longitude as well and call it *evolutionary spectrum* (Priestley, 1965). Given two spectra, $\mathfrak{f}_i(k;\phi)$, $i = 1,2$ and a mapping $b_{\text{land}} : [0,\pi] \times [0,2\pi) \to [0,1]$, an evolutionary spectrum is attained through the convex combination

$$\mathfrak{f}(k;\phi,\vartheta) = \mathfrak{f}_1(k;\phi) b_{\,\text{land}}(\phi,\vartheta) + \mathfrak{f}_2(k;\phi)\left\{1 - b_{\text{land}}(\phi,\vartheta)\right\}, \tag{4.4}$$

so that b_{land} plays the role of modulating the relative contribution of the land regime. Castruccio and Guinness (2017) showed how this approach is able to capture the majority of the non-stationarity occurring over a single latitudinal band. Recently, Jeong *et al.* (2017) proposed an extension of this approach incorporating mountain ranges in the evolutionary spectrum in the context of wind fields. Even if the data are not on a regular grid, Horrell and Stein (2015) showed that it is still possible to interpolate satellite data on a gridded structure using interpolated likelihoods to leverage spectral methods. They propose first performing kriging on the original observations, interpolating them over a grid and evaluating the likelihood of these pseudo-observations.

The spectral approach to global data allows one to achieve a fit for data sets of remarkable size to generalise axially symmetric models to capture longitudinal non-stationarities. Such a great improvement, however, comes at the cost of a loss of interpretability of the notion of distance. Indeed, the wavenumber k differs in physical length for different latitudes. Thus, interpreting the dependence structure across latitudes is problematic, especially near the poles, where the physical distance amongst points is very small. Additionally, a process specified with a latitudinally varying spectrum is not, in general, mean square continuous at the poles. Conditions for regularity at these two singular points have been discussed in Castruccio and Guinness (2017).

4.1.2 The temporal structure

For sufficiently aggregated data, (4.2) can be further simplified by assuming $\mathbf{\Phi}_j = \text{diag}(\phi_{ii;j})$, so that inference can be performed in parallel for each location. Simple diagnostics have shown that this structure is adequate for data spanning from multi-decadal to monthly data, while submonthly data would likely require a more sophisticated neighbouring-dependent structure. A key feature of some geophysical processes is their dependence along thousands of miles. These teleconnections would require a more complex structure for $\mathbf{\Phi}_j$ that would link far away locations.

4.2 The Stochastic Partial Differential Equation Approach

We start with a brief description of the SPDE approach as in Lindgren *et al.* (2011) and Bolin and Lindgren (2011). We skip all the mathematical details and will stick to the main idea. Clearly, we shall detail our exposition by working on the sphere. The main idea in Lindgren *et al.* (2011) is to evade direct specification of the Matérn function $\mathcal{M}_\nu$ as defined in Equation 3.15 in order to be able to work on any manifold, which of course includes the sphere $\mathbb{S}^2$. To quote Lindgren *et al.* (2011) verbatim, *our main objective is to construct Matérn fields on the sphere, which is important for the analysis of global spatial and spatiotemporal models.* Also, the authors note that they want to avoid using the Matérn covariance adapted with chordal distance in order to avoid *the interpretational disadvantage of using chordal distances to determine the correlation between points.*

The solution is to consider the SPDE having a Gaussian field with a Matérn covariance function as a solution: for a merely spatial process X on $\mathbb{S}^2$, the authors study the SPDE defined through

$$(\kappa^2 - \Delta)^\alpha X(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{S}^2, \tag{4.5}$$

where $\kappa > 0$, Δ is the Laplace–Beltrami operator, and $\mathcal{W}$ is Gaussian white noise on the sphere. Here, the positive exponent α depends on the parameter ν in (3.15), as well as on the dimension of the sphere.

In order to provide a computationally convenient approximation of (4.5), Lindgren *et al.* (2011) find a very ingenious computationally efficient Hilbert space approximation, namely, the weak solution to (4.5) is found in some approximation space spanned by some basis functions. The computational efficiency is then attained by imposing local basis functions, that is, compactly supported basis functions. This all boils down to approximating the field X with a Gaussian Markov field, $\boldsymbol{x}$, with precision matrix $\boldsymbol{Q}$. This idea is then generalised in Bolin and Lindgren (2011) through nested SPDE models.

The SPDE approach proposed in Lindgren *et al.* (2011) is ingeniously coupled with hierarchical models by Cameletti *et al.* (2012) to provide a space–time model. Our exposition is

adapted to the domain $\mathbb{S}^2 \times \mathbb{R}$. The authors propose a model of the type

$$Z(\boldsymbol{s},t) = \boldsymbol{\Upsilon}_t^\top \boldsymbol{\beta} + W_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}),$$
$$W_t(\boldsymbol{s}) = aW_{t-1}(\boldsymbol{s}) + \mathcal{W}_t(\boldsymbol{s}), \qquad (\boldsymbol{s},t) \in \mathbb{S}^2 \times \mathbb{R}.$$

Here, $\boldsymbol{\Upsilon}_t$ is the vector of covariates and $\boldsymbol{\beta}$ is a parameter vector. The process $\varepsilon_t(\boldsymbol{s})$ expresses the measurement error at time t and location $\boldsymbol{s}$ on the sphere. The latent process $W_t(\boldsymbol{s})$ is autoregressive over time, and $\mathcal{W}_t(\boldsymbol{s})$ is a purely spatial process with Matérn covariance function as in Equation 3.15. The bridge to the SPDE approach thus comes from the following assumptions: $\mathcal{W}$ is approximated through a Gaussian Markov structure with a given precision matrix. The same approximation, but with another precision matrix, is then assigned to the process $W_1(\boldsymbol{s})$. Using the dynamic approach, Cameletti *et al.* (2012) show that the whole latent process W has a Gaussian Markov structure with a sparse precision matrix that can be calculated explicitly.

A direct space–time formulation of the SPDE approach is also suggested in Lindgren *et al.* (2011).

5 Data Example

The 2015 El Niño Southern Oscillation was registered as the most intense over the last two decades (Wolter & Timlin, 2011). In September–October 2015, strong El Niño Southern Oscillation conditions, coupled with the Indian Ocean Dipole, suppressed precipitation and resulted in a dry, highly flammable landscape in Equatorial Asia. The extent of the haze from fires in the region was the largest recorded since 1997. The increased particulate matter concentration over the densely populated area resulted in tens of millions of people being exposed to very unhealthy and even hazardous air quality (as defined by the Pollutant Standard Index National Environment Agency, 2016) and one of the worst environmental disasters on record. The assessment of exposure and mortality from this event is critical for the implementation of future mitigation strategies. The estimation of these numbers with the associated uncertainty, including the official estimates from scientific studies performed by local governments, has received widespread media attention (Shannon *et al.*, 2016). While it is possible to focus on regional data to provide local exposure estimates, such simulations are very hard to perform and have been attempted only by Crippa *et al.* (2016), while all other studies have been focused on more affordable and readily available data on a global scale.

Here, we focus on this event using the MERRA-2 reanalysis (Molod *et al.*, 2015) data of daily (aggregated from 3-hourly) Aerosol Optical Depth (AOD); see Figure 2. Values around 0.01 mean clear sky conditions, 0.4 very hazy conditions and AOD measurements around 1 indicate extremely toxic conditions. We have also removed the trend by subtracting the monthly mean (location-wise) from every point. We use this data set to compare different statistical models fitted globally but focused on the region of interest. The practical approaches discussed in Section 4 allow for full estimation over the 2 months (for a total of more than 12 million points), but that is not possible with second-order approaches. Therefore, we limit our analysis to the first 6 days of October. While a detailed study would require an additional comparison in terms of exposure maps from population estimates, we avoid this for the sake of simplicity and brevity.

5.1 The Second-Order Approach

We fit the data set with a second-order approach as detailed in Section 3. Despite subsampling in time, the data set is still too large for a full analysis. Hence, further subsampling in space was

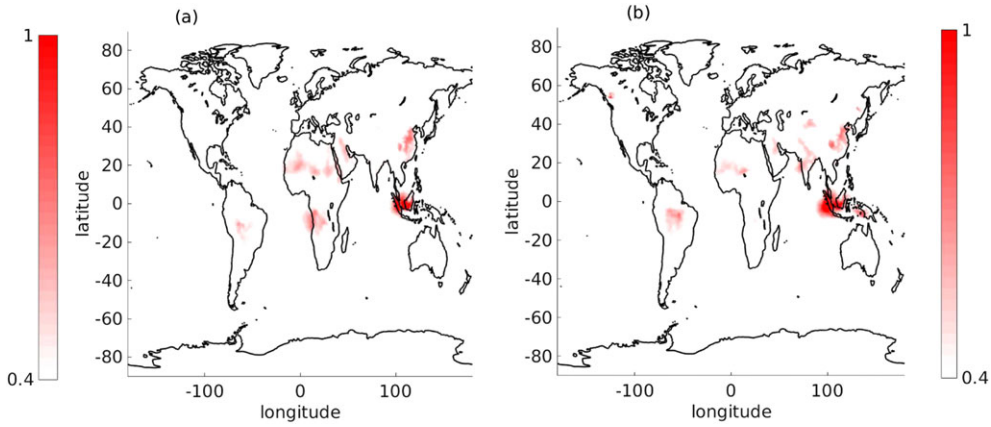

Figure 2. *Averages for Aerosol Optical Depth for the months of September (a) and October (b). [Colour figure can be viewed at wileyonlinelibrary.com]*

performed and we continued to work with a grid of 10°in latitude and longitude. We remove a small portion of observations, which correspond to latitudes greater than 85°and lower than −85°to avoid numerical problems due to the concentration of points on the poles.

We have considered four models:

Model 1. A modified Gneiting covariance, as detailed by Equation 3.6.
Model 2. A dynamically compactly supported covariance, as in Equation 3.11.

To explain the choices for models 3 and 4, we consider the Gneiting (2002) function

$$K(r,u) = \frac{\sigma^2}{\left(1+\frac{r}{b_S}\right)^3} \exp\left(-\frac{|u|}{b_T\left(1+\frac{r}{b_S}\right)^{1/2}}\right), \qquad r \geq 0. \tag{5.1}$$

Model 3. A Gneiting type covariance coupled with the great circle distance, that is, $\psi(d_{\mathrm{GC}},u) = K(d_{\mathrm{GC}},u)$, with K as in (5.1).
Model 4. A Gneiting type covariance K as in Equation 5.1, coupled with the chordal distance, d_{CH}.

The proposed models have the three parameters $(\sigma^2, b_S, b_T)^\top$, indicating the variance, spatial and temporal range, respectively. Inference was performed using a pairwise composite likelihood (CL) approach, which provides approximate but asymptotically unbiased estimates for very large data sets. We use the CL method for observations whose spatial distance is less than 6 378 km (equivalent to 1 radians on a unit sphere).

Table 3 shows CL estimates and the Log-CL values at the maximum; the models have similar performance in terms of CL. Figure 3 illustrates the empirical spatial (semi-) variogram in terms of the great circle distance for different temporal lags in comparison to the theoretical models. The models are indeed able to capture this large-scale feature of the data by fitting well the empirical variograms.

We now compare the models in terms of their predictive performance. We use the kriging predictor and a drop-one prediction strategy. We consider the following indicators: Mean Squared Prediction Error (MSPE), Log-Score (LSCORE) and Continuous Ranked Probability Score (CRPS). Table 4 contains the indicators for each model. Small values indicate better predictions.

Table 3. *Composite likelihood estimates and Log-CL value at the maximum for both second-order models.*

Model	$\hat{\sigma}^2$	$\hat{b}_S$	$\hat{b}_T$	Log-CL
1	8.564×10^{-3}	816.38	6.320	2223640
2	8.564×10^{-3}	3 845.93	6.255	2223574
3	8.562×10^{-3}	1 651.90	2.202	2223928
4	8.563×10^{-3}	765.36	5.601	2223828

The units for the spatial and temporal scales are kilometres and days, respectively. CL, composite likelihood.

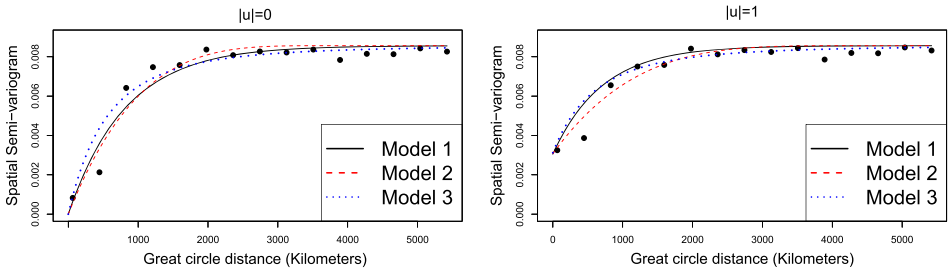

Figure 3. *Empirical spatial (semi-) variograms versus theoretical covariances according to models 1, 2 and 3 for different temporal lags. [Colour figure can be viewed at wileyonlinelibrary.com]*

Table 4. *Predictive scores for models 1–4.*

Model	MSPE	LSCORE	CRPS
1	5.069×10^{-3}	−1.052	0.107
2	5.307×10^{-3}	−0.935	0.106
3	4.442×10^{-3}	−1.288	0.117
4	4.527×10^{-3}	−1.281	0.118

MSPE, Mean Squared Prediction Error; LSCORE, Log-Score; CRPS, Continuous Ranked Probability Score.

According to MSPE and LSCORE, the predictive performance of models 3 and 4 is better than that of models 1 and 2. The use of the great circle distance (model 3) results in a 2% improvement in terms of MSPE compared with the use of chordal distance (model 4). Model 3 also outperforms model 4 in terms of LSCORE and Continuous Ranked Probability Score.

5.2 Dynamical Approach

We now fit the same data set with a dynamical model. We choose (4.2) with $p = 1$, that is, a VAR(1) process with a diagonal autoregressive structure. Providing a location-specific temporal structure is likely to be too flexible for data subsampled for only 6 days, but further model selection approaches to reduce model complexity were deemed out of the scope of this work. Because the data are gridded, we choose a spectral model (4.3) with a latitudinally varying spectrum and a coherence for the same wavenumber but independence otherwise. While the model was subsampled in time for a comparison with the second-order approach, the inference

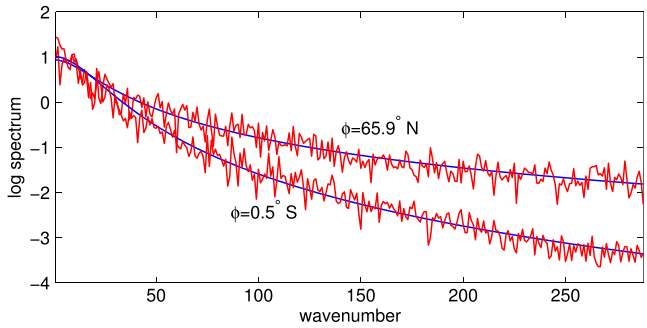

Figure 4. *Fitted parametric (blue) and non-parametric (i.e. periodogram and red) log spectrum for two different latitudinal bands, one at the equator and one at high northern latitudes. The periodogram was obtained by averaging across all longitudes and times. [Colour figure can be viewed at wileyonlinelibrary.com]*

Table 5. *Comparison between different models in terms of number of parameters (excluding the temporal ones), computational time (minutes), normalised loglikelihood and Bayesian Information Criterion for* $K = 6$ *days from October 1–6, 2015.*

	Model 1	Model 2	Model (4.3)
# param	3	3	161
Time (minutes)	238	236	4.2
Δ-loglik$/(N_\phi N_\vartheta K)$	−3.24	−3.24	0
BIC$\times 10^4$	−4.50	−4.55	−19.0

was scalable and the fit of the entire data set did not require significant additional computational time.

Even if the likelihood evaluation can sidestep the big N problem by storing the results in the spectral domain via a Whittle likelihood (Whittle, 1953), it is not possible to obtain a maximum likelihood for the entire model. Because the model's structure composes of hundreds of parameters, optimisation over the entire space would be impossible. Therefore, the estimation is performed step-wise. First, the temporal dependence is estimated. Then, we estimate the longitudinal dependence followed by the latitudinal dependence. Hence, the parameter estimates are to be regarded as local approximations.

The spectral model allows for a different spectral shape at different latitudes, and it is flexible enough to capture very different behaviours, as shown in the diagnostics in Figure 4. Indeed, AOD residuals near the equator display a much smoother behaviour than at high latitudes, where the spectrum is more flat, but the model is able to fit both behaviours naturally.

5.3 Comparison for Equatorial Asia

While both models are fitted globally, the interest lies in their relative performance in the area of interest, that is, in Equatorial Asia. Table 5 shows a comparison of the aforementioned models in terms of the marginal likelihood in this area, defined as all points with latitudes between $-11.3°$ and $15.3°$ ($N_\phi = 53$) and longitudes between $93°$ and $137°$ ($N_\vartheta = 71$). The dynamically specified model almost uniformly outperforms models 1 and 2 in the second-order approach: it is faster, richer in complexity and yet achieves an overwhelmingly better fit and so it is clearly more suitable for a more detailed analysis in the area.

While these results are strongly in favour of the dynamically specified approach, it must be pointed out that this particular setting, that is, a regular grid in space and equal observations in time, is such that a spectral model is clearly the best choice. An application with more complicated geometries such as with satellite data would have required an adaptation of spectral methods along the lines of Horrell and Stein (2015), and likely resulted in a worse fit. While comparison with an irregular design would have added to the discussion, we avoid it for the sake of brevity.

6 Research Problems

This section is devoted to a list of research problems that we consider relevant for future research.

1. Non-stationary Space–Time Covariances. The models proposed in Section 3 are geodesically isotropic and temporally symmetric. Because real phenomena on the globe are notably non-stationary, it is necessary to use those models as building blocks for more sophisticated constructions. In particular, the development of non-stationary models is necessary and a promising direction for research is to extend the work of Guella *et al.* (2017, 2016a, 2016b) who define kernels over products of n-dimensional spheres. Another direction for research might be to take into account differences in local geometry of the sphere representing planet Earth. Thus, the natural solution is to work on processes evolving temporally over Riemannian manifolds. In this direction, the works of Menegatto *et al.* (2006) and the recent work by Barbosa and Menegatto (2017) might be very useful.

2. Models Based on Convolutions. Convolution arguments have been used for Gaussian processes defined over spheres (no time) in order to index the associated fractal dimensions. Hansen *et al.* (2015) show that Gaussian particles provide a flexible framework for modeling and simulating three-dimensional star-shaped random sets. The authors show that, if the kernel is a von Mises–Fisher density or uniform on a spherical cap, the correlation function of the associated random field admits a closed form expression. We are not aware of any convolution argument for space–time covariance functions. The work of Ziegel (2014) might be very useful in this direction.

3. Physical-based Constructions. Constructions based on dynamical models or moving averages have been proposed by Ailliot *et al.* (2011) and Schlather (2010). We are not aware of such extensions to the case $\mathbb{S}^2 \times \mathbb{R}$, but certainly it would be worth studying. Some other constructions based on physical characteristics of the space–time process might be appealing for modeling several real processes. In this direction, it would be desirable to study the approaches proposed in Christakos (2000) as well as Kolovos *et al.* (2004), Christakos (1991), Christopoulos and Tsantili (2016) and Christakos *et al.* (2000).

4. Multivariate Space–Time Models. Often, several variables are observed over the same spatial location and same temporal resolution. Thus, there is substantial need for space–time multivariate covariance models that are geodesically isotropic or axially symmetric in the spatial component. The literature in this subject is very sparse, with the notable exceptions of Jun (2011) and Alegria *et al.* (2017). The main difficulty is in adapting the construction principles that have been proposed for multivariate space–time covariances in Euclidean spaces. Most of these construction principles are based on Cramér's theorem (see 2015, with the references therein). Thus, the characterisation theorems proposed by Alegria *et al.* (2017) lay the foundations for substantial future work in this direction.

5. Matérn Analogues over Spheres. Find the counterpart of the Matérn covariance function on the sphere. This would allow for a covariance function that indexes the differentiability at the

origin of the associated Gaussian field. Notable attempts have been made by Guinness and Fuentes (2016), with partial success. Møller *et al.* (2018) suggest modeling the n-Schoenberg coefficients (see Appendix A for details) while imposing a given rate of decay, but this does not allow for explicit closed forms.

5. Compact Supports with Differentiable Temporal Margins. A potential drawback of the space–time construction as in Theorem 3.1 is that it only allows for temporal margins that are non-differentiable at the origin. An important step ahead would be to improve such a limitation. Following the arguments in the proof of Theorem 3.1, the crux would be in proving that, for some mapping $h : [0, \infty) \to \mathbb{R}$, the function

$$h(u^2)^{\alpha+d}\,{}_1F_2\left(1+k; \frac{\mu+2}{2}+k, \frac{\nu+3}{2}+k; -n^2h(u^2)^2/4\right), \qquad u \in \mathbb{R}, \quad n \in \mathbb{N}$$

is positive definite for all $n \in \mathbb{N}$.

6. Spectral Constructions à la Stein. Stein (2005a) proposes a class of spectral densities in $\mathbb{R}^d \times \mathbb{R}$ of the type

$$\mathfrak{f}(\boldsymbol{\omega}, \tau) = ((1 + \|\boldsymbol{\omega}\|)^{\alpha_1} + (1 + |\tau|)^{\alpha_2})^{-\alpha_3}, \qquad (\boldsymbol{\omega}, \tau) \in \mathbb{R}^d \times \mathbb{R}, \quad \alpha_i > 0,$$

for $i = 1, 2, 3$. Under the condition $\alpha_3 < \alpha_1/d + \alpha_2$, $\mathfrak{f}$ is in $L_1(\mathbb{R}^d \times \mathbb{R})$. The partial Fourier transforms allow for indexing the differentiability at the origin in a similar way as the Matérn covariance function. Further, the resulting covariance is smoother away from the origin than at the origin, which is important as reported in Theorem 1 in Stein (2005a). There is no analogue of $\mathfrak{f}$-based constructions for the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$. One should necessarily start from the Berg–Porcu (Berg & Porcu, 2017) characterisation and try to find a related approach that allows one to obtain such a construction on spheres cross time.

7. Local Anisotropies and Transport Effects. The geodesically isotropic models should be extended to allow for local anisotropies as in Paciorek and Schervish (2006). In this direction, a major step should be made in order to generalise the Lagrangian model in Equation 3.14 to the non-stationary case.

8. Strictly Positive Definite Functions. A function $C : (\mathbb{S}^2 \times \mathbb{R})^2 \to \mathbb{R}$ is called strictly positive definite when inequality

$$\sum_{i,j}^{N} c_i C\left((\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j)\right) c_j > 0$$

holds for any $\{c_k\}_{k=1}^N \subset \mathbb{R}$, unless $c_1 = \cdots = c_N = 0$, and $\{\boldsymbol{s}_k, t_k\}_{k=1}^N \subset \mathbb{S}^d \times \mathbb{R}$. There is substantial work on strict positive definiteness, and the reader is referred to Chen *et al.* (2003), Menegatto (1994), Menegatto (1995) and Menegatto *et al.* (2006). A characterisation of the subclass of $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$ (see Appendix A) with members ψ such that the corresponding covariances C are strictly positive definite is still elusive.

9. Walks Through Dimensions. The literature on walks through dimensions is related to operators that allow one, for a given positive definite function on the n-dimensional sphere, to obtain new classes of positive definite functions on n'-dimensional spheres, with $n \neq n'$. The application of such operators has consequences on the differentiability at the origin of the involved functions. Walks on spheres have been proposed by Beatson *et al.* (2014), Ziegel (2014) and

Massa *et al.* (2017). This last work extends the previous work to the case of complex spheres. It would be timely to obtain walks through dimensions for the members of the classes $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$, for n a positive integer.

10. Optimal Prediction on Spheres Cross Time. Much work needs to be done in order to assess asymptotic optimal kriging prediction over spheres cross time when the covariance is misspecified. The work of Arafat *et al.* (2016) bridges the gap between equivalence of Gaussian measures and asymptotic optimal prediction over spheres. A relevant direction of research would be to evaluate the screening effect on spheres cross time. After the work of Stein (1999), there is nothing related to spectral behaviours over spheres that allow for evaluation of the corresponding screening effect.

11. Fast Simulations on Spheres. Lang and Schwab (2013) and Clarke *et al.* (2018) propose fast simulations through truncation of Karhunen–Loève expansions. It would be very interesting to propose analogues of circulant embedding methods, where the difficulty is mainly due to the fact that it is not possible to set a regular grid on the sphere.

12. Chordal versus Great Circle Again. Moreno Bevilacqua (personal communication) points out that he has tried the following experiment. Simulate any set of points on the sphere cross time. Take any element ψ from the class $\mathcal{P}(\mathbb{S}^2, \mathbb{R})$. Replace the great circle distance with the chordal distance and calculate the corresponding matrix realisations. With all the experiments, he always found strictly positive eigenvalues. Thus, the question is suppose that $\psi \in \mathcal{P}(\mathbb{S}^2, \mathbb{R})$. Then, is it true that $\psi(d_{\mathrm{CH}})$ is positive definite on $\mathbb{S}^2 \times \mathbb{S}^2 \times \mathbb{R}$?

Appendix A: Mathematical Background

We need some notation in order to illustrate the following sections. This material largely follows the exposition in Berg and Porcu (2017).

Let n be a positive integer. We denote by $\mathbb{S}^n$ the n-dimensional unit sphere of $\mathbb{R}^{n+1}$, given as

$$\mathbb{S}^n = \{\boldsymbol{s} \in \mathbb{R}^{n+1} \mid \|\boldsymbol{s}\| = 1\},\ n \geq 1. \tag{6.1}$$

We also consider

$$\mathbb{S}^\infty = \{(s_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} | \sum_{k=1}^{\infty} s_k^2 = 1\},$$

which is the unit sphere in the Hilbert sequence space ℓ_2 of square summable real sequences. Following Berg and Porcu (2017), we thus define the class $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$ as the class of continuous functions $\psi : [0, \pi] \times \mathbb{R} \to \mathbb{R}$ such that

$$C((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)) = \psi(d_{\mathrm{GC}}(\boldsymbol{s}_1, \boldsymbol{s}_2), t_1 - t_2), \qquad (\boldsymbol{s}_i, t_i) \in \mathbb{S}^n \times \mathbb{R}, \quad i = 1, 2,$$

where d_{GC} has already been defined as the great circle distance. Berg and Porcu (2017) define $\mathcal{P}(\mathbb{S}^\infty, \mathbb{R})$ as $\bigcap_{n\geq 1} \mathcal{P}(\mathbb{S}^n, \mathbb{R})$. The inclusion relation

$$\mathcal{P}(\mathbb{S}^1, \mathbb{R}) \supset \mathcal{P}(\mathbb{S}^2, \mathbb{R}) \supset \cdots \supset \mathcal{P}(\mathbb{S}^\infty, \mathbb{R})$$

is strict, and the reader is referred to Berg and Porcu (2017) for details.

We recall the definition of Gegenbauer polynomials $C_k^{(\lambda)}$, given by the generating function (see Dai & Xu, 2013)

$$(1 - 2xz + r^2)^{-\lambda} = \sum_{k=0}^{\infty} C_k^{(\lambda)}(z) r^k, \quad |r| < 1, z \in \mathbb{C}. \tag{6.2}$$

Here, $\lambda > 0$. We have the classical orthogonality relation:

$$\int_{-1}^{1} (1 - x^2)^{\lambda - 1/2} C_k^{(\lambda)}(x) C_h^{(\lambda)}(x)\, \mathrm{d}x = \frac{\pi \Gamma(k + 2\lambda) 2^{1-2\lambda}}{\Gamma^2(\lambda)(k + \lambda)k!} \delta_{h=k} \tag{6.3}$$

with Γ denoting the Gamma function.

It is of fundamental importance that $|C_k^{(\lambda)}(x)| \leq C_k^{(\lambda)}(1)$, $x \in [-1, 1]$. The special value $\lambda = (n - 1)/2$ is relevant for the sphere $\mathbb{S}^n$ because of the relation to spherical harmonics, which is illustrated in Berg and Porcu (2017) as follows. A spherical harmonic of degree k for $\mathbb{S}^n$ is the restriction to $\mathbb{S}^n$ of a real-valued harmonic homogeneous polynomial in $\mathbb{R}^{n+1}$ of degree k. Together with the zero function, the spherical harmonics of degree k form a finite dimensional vector space denoted $\mathcal{H}_k(n)$. It is a subspace of the space $\mathcal{C}(\mathbb{S}^n)$ of continuous functions on $\mathbb{S}^n$. We have

$$N_k(n) = \dim \mathcal{H}_k(n) = \frac{(n)_{k-1}}{k!}(2n + n - 1),\ k \geq 1, \quad N_0(n) = 1, \tag{6.4}$$

(see Dai & Xu, 2013, p. 3). Here, $(n)_{k-1}$ denotes the Pochammer symbol.

The surface measure of the sphere is denoted as ω_n, and it has total mass

$$||\omega_n|| = \frac{2\pi^{(n+1)/2}}{\Gamma((n + 1)/2)}. \tag{6.5}$$

The spaces $\mathcal{H}_k(n)$ are mutually orthogonal subspaces of the Hilbert space $L^2(\mathbb{S}^n, \omega_n)$ that they generate. This means that any $F \in L^2(\mathbb{S}^n, \omega_n)$ has an orthogonal expansion

$$F = \sum_{k=0}^{\infty} S_k,\ S_k \in \mathcal{H}_k(n), \quad ||F||_2^2 = \sum_{k=0}^{\infty} ||S_k||_2^2, \tag{6.6}$$

where the first series converges in $L^2(\mathbb{S}^n, \omega_n)$ and the second series is Parseval's equation (Berg & Porcu, 2017). Here, S_k is the orthogonal projection of F onto $\mathcal{H}_k(n)$, given as

$$S_k(\xi) = \frac{N_k(n)}{||\omega_n||} \int_{\mathbb{S}^n} c_k(n, \langle s, \eta \rangle) F(\eta)\, \mathrm{d}\omega_n(\eta). \tag{6.7}$$

See the addition theorem for spherical harmonics (Schoenberg, 1942). For $\lambda = (n - 1)/2$, $c_k(n, x)$ is defined as the normalised Gegenbauer polynomial

$$c_k(n, x) = C_k^{((n-1)/2)}(x) / C_k^{((n-1)/2)}(1) = \frac{k!}{(n - 1)_k} C_k^{((n-1)/2)}(x), \tag{6.8}$$

which is 1 for $x = 1$.

Specialising the orthogonality relation (6.3) to $\lambda = (n-1)/2$ and using Equations (6.4) and (6.5), Berg and Porcu (2017) get for $n \in \mathbb{N}$

$$\int_{-1}^{1} (1-x^2)^{n/2-1} c_k(n,x) c_h(n,x)\, \mathrm{d}x = \frac{||\omega_n||}{||\omega_{n-1}|| N_k(n)} \delta_{h=k}. \tag{6.9}$$

The following result characterises completely the class $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$.

Theorem 6.1. (Berg & Porcu, 2017) *Let $n \in \mathbb{N}$ and let $\psi : [0,\pi] \times \mathbb{R} \to \mathbb{C}$ be a continuous function. Then, ψ belongs to the class $\mathcal{P}(\mathbb{S}^n, \mathbb{R})$ if and only if there exists a sequence $\{\varphi_{k,n}\}_{k=0}^{\infty}$ of positive definite functions on $\mathbb{R}$, with $\sum_k \varphi_{k,n}(0) < \infty$, such that*

$$\psi(d_{GC}, u) = \sum_{k=0}^{\infty} \varphi_{k,n}(u) c_k(n, \cos d_{GC}), \tag{6.10}$$

and the earlier expansion is uniformly convergent for $(d_{GC}, u) \in [0,\pi] \times \mathbb{R}$. We have

$$\varphi_{k,n}(u) = \frac{N_k(n)||\omega_{n-1}||}{||\omega_n||} \int_0^{\pi} \psi(x,u) c_k(n,x) \sin x^{n-1}\, \mathrm{d}x. \tag{6.11}$$

We also report the characterisation of the class $\mathcal{P}(\mathbb{S}^\infty, \mathbb{R})$ obtained by the same authors.

Theorem 6.2. (Berg & Porcu, 2017) *Let $\psi : [0,\pi] \times \mathbb{R} \to \mathbb{R}$ be a continuous function. Then, ψ belongs to $\mathcal{P}(\mathbb{S}^\infty, \mathbb{R})$ if and only if there exists a sequence $\{\varphi_k\}_{k=0}^{\infty}$ of positive definite functions on $\mathbb{R}$, with $\sum_k \varphi_k(0) < \infty$, such that*

$$\psi(d_{GC}, u) = \sum_{k=0}^{\infty} \varphi_k(u) \cos^k d_{GC}, \tag{6.12}$$

and the earlier expansion is uniformly convergent for $(d_{GC}, u) \in [0,\pi] \times G$.

Some comments are in order. Theorems 6.1 and 6.2 are fundamental to creating the examples illustrated in Table 1. Also, they are crucial to solving some of the open problems that we reported in Section 6.

Appendix B: List of Permissible Models on $\mathbb{S}^2 \times \mathbb{R}$

We report a list of examples in Table A1.

Table A1. *Parametric families of geodesically isotropic space–time covariance functions.*

Family	Expression for $\psi(r,u)$			Parameter Restriction (or Remarks)
Fonseca-Steel	$(1+(\gamma_S(r)+\gamma_T(u)))\,(1+\gamma_S(r))^{-\lambda_1}\,\frac{\mathcal{K}_{\lambda_1}\left(2\sqrt{(a+\gamma_S(r))^{\delta}}\right)}{\mathcal{K}_{\lambda_1}(2\sqrt{a^{\delta}})}$			$a_1,\delta,\lambda_i>0, i=0,1.$
Clayton PMC	$\left((1+r^{\alpha})^{\rho_1}+(1+\lvert u\rvert^{\beta})^{\rho_2}-1\right)^{-\rho_3}$			$\alpha,\beta,\rho_1,\rho_2\in(0,1];\ \rho_3>0.$
Gumbel PMC	$\exp\left(-(r^{\alpha_1}+u^{\alpha_2})^{\alpha_3}\right)$			$\alpha_i\in(0,1], i=1,2,3.$
PBG	Equation (3.5)			
	Families for f	Expression	Parameters	
	Dagum	$f(r)=1-\left(\frac{r^{\beta}}{1+r^{\beta}}\right)^{\tau}$	$\beta,\tau\in(0,1]$	
	Matérn	Equation (3.15)	$0<\nu\leq 1/2$	
	Gen. Cauchy	$f(r)=(1+r^{\alpha})^{-\beta/\alpha}$	$\alpha\in(0,1],\beta>0$	
	Pow. Exponential	$f(r)=\exp(-r^{\alpha})$	$\alpha\in(0,1]$	
Adaptive Gneiting	Equation (3.7)			f as in PBG family
	Families for $g_{[0,\pi]}$	Expression	Parameters	
	Dagum	$g_{[0,\pi]}(r)=1+\left(\frac{r^{\beta}}{1+r^{\beta}}\right)^{\tau}$	$\beta,\tau\in(0,1]$	
	Gen. Cauchy	$g_{[0,\pi]}(r)=(1+r^{\alpha})^{\beta/\alpha}$	$\alpha\in(0,1],\beta\leq\alpha$	
	Power	$g_{[0,\pi]}(r)=c+r^{\alpha}$	$\alpha\in(0,1],c>0$	
PBG2	Equation (3.10)			h positive, decreasing and convex with $\lim_{t\to\infty}h(t)=0.$
Dynamical Wendland	Equation (3.12)			h positive, decreasing and convex with $\lim_{t\to\infty}h(t)=0.$
	Families for $\varphi_{\mu,k}$	Expression	Parameters	
	Dynamical Wendland 2	$h(\lvert u\rvert)^{\alpha}\left(1-\frac{r}{h(\lvert u\rvert)}\right)_+^{6}\left(1+6\frac{r}{h(\lvert u\rvert)}\right)$	$\alpha\geq 4$	
	Dynamical Wendland 4	$h(\lvert u\rvert)^{\alpha}\left(1-\frac{r}{h(\lvert u\rvert)}\right)_+^{8}\left(1+8\frac{r}{h(\lvert u\rvert)}+\frac{63}{3}\left(\frac{r}{h(\lvert u\rvert)}\right)^{2}\right)$	$\alpha\geq 4$	

PMC is the acronym for Porcu–Mateu–Christakos (Porcu *et al.*, 2009). PBG is the acronym for Porcu–Bevilacqua–Genton (Porcu *et al.*, 2016a). We use the abuse of notation r for the great circle distance d_{GC}.

International Statistical Review (2018), **0**, 0, 1–34

Appendix C. Mathematical Proofs

C1. The Quasi-Arithmetic Class on Spheres

For ease of exposition, we slightly deviate from the notation of the paper and denote the composition of f to g by $f \circ g$. Let us recall the expression of quasi-arithmetic covariances as in Equation 3.8:

$$\psi(d_{\mathrm{GC}}, u) = f\left(\frac{1}{2} f^{-1} \circ \psi_{\mathcal{S}}(d_{\mathrm{GC}}) + \frac{1}{2} f^{-1} \circ C_{\mathcal{T}}(u)\right), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}.$$

By Bernstein's theorem (Feller, 1966, p. 439), a function $f : [0, \infty) \to \mathbb{R}$ is completely monotonic if and only if

$$f(t) = \int_{[0,\infty)} \mathrm{e}^{-\xi t} H(\mathrm{d}\xi), \qquad t \geq 0, \tag{C1}$$

where H is a positive and bounded measure. A Bernstein function is a continuous mapping on the positive real line, having a first derivative being completely monotonic. We are now able to give a formal assertion for the validity of the quasi-arithmetic construction.

Theorem 6.3. *Let $f : [0, \infty) \to \mathbb{R}_+$ be a completely monotonic function. Let $f_1 : [0, \infty) \to \mathbb{R}$ be a continuous function such that $f^{-1} \circ f_1$ is a Bernstein function. Let $C_{\mathcal{T}} : \mathbb{R} \to \mathbb{R}$ be a continuous, symmetric covariance function such that $f^{-1} \circ C_{\mathcal{T}}$ is a temporal variogram. Call $\psi_{\mathcal{S}} = f_{1,[0,\pi]}$ the restriction of f_1 to the interval $[0, \pi]$. Then,*

$$\psi(d_{GC}, u) = \mathcal{Q}_f\left(\psi_{\mathcal{S}}(d_{GC}), C_{\mathcal{T}}(u)\right), \qquad (d_{GC}, u) \in [0, \pi] \times \mathbb{R} \tag{C2}$$

is a geodesically isotropic space–time covariance function on $\mathbb{S}^n \times \mathbb{R}$, for all $n = 1, 2, 3, \ldots$.

Proof. Denote by g the composition $f^{-1} \circ f_1$ and call $g_{[0,\pi]}$ the restriction of g to $[0, \pi]$, obtained through $g_{[0,\pi]} = f^{-1} \circ f_{1,[0,\pi]}$. By assumption, g is a Bernstein function. Thus, arguments in Porcu and Schilling (2011), with the references therein, show that the function h, defined through

$$h(t; \xi) = \exp(-\xi g(t)), \qquad t, \xi \geq 0$$

is a completely monotonic function for any positive ξ. Thus,

$$h_{[0,\pi]}(d_{\mathrm{GC}}; \xi) = \exp\left(-\xi f^{-1} \circ f_{1,[0,\pi]}(d_{\mathrm{GC}})\right), \qquad d_{\mathrm{GC}} \in [0, \pi]$$

is the restriction of a completely monotonic function to the interval $[0, \pi]$. Invoking theorem 7 in Gneiting (2013), we obtain that $h_{[0,\pi]}$ is a geodesically isotropic covariance function on any n-dimensional sphere. Additionally, because $f^{-1} \circ C_{\mathcal{T}}$ is a temporal variogram, by Schoenberg's theorem (Schoenberg, 1938), we deduce that $k(u; \xi) = \exp(-\xi f^{-1} \circ C_{\mathcal{T}}(u))$, $u \in \mathbb{R}$, is a covariance function on the real line for every positive ξ. Thus, the scale mixture covariance

$$\begin{aligned}\psi(d_{\mathrm{GC}}, u) &= \int_{[0,\infty)} h_{[0,\pi]}(d_{\mathrm{GC}}; \xi) k(u; \xi) H(\mathrm{d}\xi) \\ &= \int_{[0,\infty)} \exp\left(-\xi g_{[0,\pi]}(d_{\mathrm{GC}}) - \xi f^{-1} \circ C_{\mathcal{T}}(u)\right) H(\mathrm{d}\xi) \\ &= \mathcal{Q}_f(\psi_{\mathcal{S}}(d_{\mathrm{GC}}), C_{\mathcal{T}}(u)), \qquad (d_{\mathrm{GC}}, u) \in [0, \pi] \times \mathbb{R}\end{aligned}$$

is a covariance function on $\mathbb{S}^n \times \mathbb{R}$ for all $n \in \mathbb{N}$. □

C2. Proofs for Section 3.3

Theorem 6.4. *Let h be a positive, decreasing and convex function on the positive real line, with $h(0) = c$, $0 < c \leq \pi$, and $\lim_{t\to\infty} h(t) = 0$. Let $\alpha \geq 3$ and $\mu \geq 4$. Then, Equation 3.10 defines a geodesically isotropic and temporally symmetric covariance function on $\mathbb{S}^3 \times \mathbb{R}$.*

Proof. The proof is based on scale mixture arguments in concert with the calculations in Porcu *et al.* (2016b). In particular, we have that the function in Equation 3.10 is the result of the scale mixture of the function $\psi_{\mathcal{S}}(d_{\mathrm{GC}}; \xi) = (1 - d_{\mathrm{GC}}/\xi)_+^n$ with the function $C_{\mathcal{T}}(u; \xi) = \xi^n \left(1 - \frac{\xi}{h(|u|)}\right)_+^{\gamma}$, $\xi > 0, t \geq 0, \gamma \geq 1, n \geq 2$. In particular, arguments in lemmas 3 and 4 in Gneiting (2013) show that $\psi_{\mathcal{S}}$ is a covariance function on $\mathbb{S}^3$ for every positive ξ. Under the required conditions on the function h, we have that $C_{\mathcal{T}}$ is a covariance function on $\mathbb{R}$ for every positive ξ. Thus, the scale mixture arguments in Porcu *et al.* (2016b), with $\mu = n + \gamma + 1$ and $\alpha = n + 1$ can now be applied, obtaining the result. □

A more sophisticated argument is required to show that the structure in Equation 3.12 is positive definite on the circle cross time. We start with the proof of Theorem 3.2 because its arguments will be partially used to prove Theorem 3.1.

Proof of Theorem 3.2. Denote by $\varphi_{[0,\pi]}$ the restriction of φ to the interval $[0, \pi]$ with respect to the first argument. Let $n \in \mathbb{N}$. Consider the sequence of functions $b_n(\cdot)$, defined through

$$\begin{aligned}b_n(u) &= \frac{2}{\pi} \int_0^{\pi} \cos(nz) \psi(z, u) \mathrm{d}z = \frac{2}{\pi} \int_0^{\pi} \cos(nz) \varphi_{[0,\pi]}(z, u) \mathrm{d}z \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \cos(nx) \varphi(x, u) \mathrm{d}x.\end{aligned} \tag{C3}$$

Because φ is positive definite on $\mathbb{R} \times \mathbb{R}$, arguments in lemma 1 in Gneiting (2002) show that $b_n(u)$ is a covariance function on $\mathbb{R}$ for all $n \in \mathbb{N}$. Additionally, we have

$$\sum_{n=0}^{\infty} b_n(0) = \sum_{n=0}^{\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} \cos(nx) \varphi(x, 0) \mathrm{d}x = \sum_{n=0}^{\infty} \hat{\varphi}_{\mathbb{N}}(n) < \infty,$$

where $\hat{\varphi}_{\mathbb{N}}$ denotes the Fourier transform of $\varphi(x, 0)$ restricted to natural numbers. Thus, we get that $\sum_n b_n(0) < \infty$ because the Fourier transform of a positive definite function is non-negative and integrable. We can thus invoke theorem 3.3 in Berg and Porcu (2017) to obtain that $\psi(d_{\mathrm{GC}}, u) = \varphi_{[0,\pi]}(d_{\mathrm{GC}}, u)$ is a covariance function on the circle $\mathbb{S}^1$ cross time. □

Proof of Theorem 3.1. We give a proof of the constructive type. Let $k \in \mathbb{N}$. Arguments in theorem 1 of Porcu *et al.* (2016b) show that the function

$$C(x,u) = \sigma^2 h(|u|)^{\alpha} \varphi_{\mu,k}\left(\frac{|x|}{h(|u|)}\right), \qquad (x,u) \in \mathbb{R} \times \mathbb{R} \tag{C4}$$

is positive definite on $\mathbb{R} \times \mathbb{R}$ provided $\alpha \geq 2k+3$ and $\mu \geq k+4$. Arguments in Porcu *et al.* (2017) show that

$$\begin{aligned} b_n(u) &= \int_{\infty}^{-\infty} \cos(nx) C(x,u) \mathrm{d}x \\ &\propto \sigma^2 h(|u|)^{\alpha+d}\, {}_1F_2\left(1+k; \frac{\mu+2}{2}+k, \frac{\nu+3}{2}+k; -n^2 h(|u|)^2/4\right), \quad u \in \mathbb{R}, \quad n \in \mathbb{N}. \end{aligned}$$

From the argument in Equation C3 in concert with Lemma 1 in Gneiting (2002), we have that $b_n(u)$ is positive definite on the positive real line for each n. Additionally, arguments in Porcu *et al.* (2017) show that, for $\mu \geq k+4$, $b_n(u)$ is strictly decreasing in n. Application of proposition 3.6 of Berg and Porcu (2017) shows that (3.12) is positive definite on $\mathbb{S}^3 \times \mathbb{R}$. The proof is completed.

Acknowledgements

We gratefully acknowledge the NASA scientists responsible for MERRA-2 products. The authors thank Anastassia Baxevani, Moreno Bevilacqua, David Bolin, George Christakos, Noel Cressie, Alessandro Fassó, Francesco Finazzi, Dionisios Hristopoulos, Giovanna Jona-Lasinio, Finn Lindgren, Gianluca Mastrantonio, Havard Rue, Bruno Sansó and Felipe Tagle for thorough discussions during the preparation of the manuscript. Emilio Porcu is supported by Fondecyt Regular number 1170290. Reinhard Furrer acknowledges the support of the Swiss National Science Foundation grant 175529.

References

Ailliot, P., Baxevani, A., Cuzol, A., Monbet, V. & Raillard, N. (2011). Space-time models for moving fields with an application to significant wave height fields. *Environmetrics*, **22**(3), 354–369.

Alegria, A. & Porcu, E. (2016). *Space-Time Geostatistical Models with both Linear and Seasonal Structures in the Temporal Components*. *Technical Report, University Federico Santa Maria*, Available at arXiv:1702.01400.

Alegria, A. & Porcu, E. (2017). The dimple problem related to space-time modeling under the Lagrangian framework. *J. Multivar. Anal.*, **162**, 110–121.

Alegria, A., Porcu, E., Furrer, R. & Mateu, J. (2017). *Covariance Functions for Multivariate Gaussian Fields Evolving Temporally over Planet Earth*. Technical Report, University Federico Santa Maria.

Alonso-Malaver, C., Porcu, E. & Giraldo, E. M. (2015). Multivariate and multiradial Schoenberg measures with their dimension walk. *J. Multivar. Anal.*, **133**, 251–265.

Arafat, A., Porcu, E., Bevilacqua, M. & Mateu, J. (2016). *Equivalence and orthogonality of Gaussian measures on spheres*. Technical Report, University Federico Santa Maria.

Baldi, P. & Marinucci, D. (2006). Some characterizations of the spherical harmonics coefficients for isotropic random fields. *Statistics and Probability Letters*, **77**, 490–496.

Banerjee, S., Gelfand, A. E., Finley, A. O. & Sang, H. (2008). Gaussian prediction process models for large spatial data sets. *J. R. Stat. Soc.: Series B*, **70**, 825–848.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. CRC Press.

Barbosa, V. S. & Menegatto, V. A. (2017). Strict positive definiteness on products of compact two–point homogeneous spaces. *Integral Transforms Spec. Funct.*, **28**(1), 56–73.
Beatson, R. K., zu Castell, W. & Xu, Y. (2014). Pólya criterion for (strict) positive definiteness on the sphere. *IMA J. Numer. Anal.*, **34**, 550–568.
Berg, C. & Porcu, E. (2017). From Schoenberg coefficients to Schoenberg functions. *Constr. Approx.*, **45**, 217–241.
Bevilacqua, M., Gaetan, C., Mateu, J. & Porcu, E. (2012). Estimating space and space-time covariance functions: A weighted composite likelihood approach. *J. Am. Stat. Assoc.*, **107**, 268–280.
Bevilacqua, M., Mateu, J., Porcu, E., Zhang, H. & Zini, A. (2010). Weighted composite likelihood-based tests for space-time separability of covariance functions. *Stat. Comput.*, **20**(3), 283–293.
Bevilacqua, M., Faouzi, T., Furrer, R. & Porcu, E. (2018). Estimation and prediction using generalized wendland covariance function under fixed domain asymptotics. *Ann. Stat.* To Appear.
Bingham, N. H. (1973). Positive definite functions on spheres. *Proc. Camb. Philos. Soc.*, **73**, 145–156.
Bolin, D. & Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Stat.*, **5**(1), 523–550.
Cameletti, M., Lindgren, F., Simpson, D. & Rue, H. (2012). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 1–23.
Castruccio, S. & Genton, M. G. (2014). Beyond axial symmetry: An improved class of models for global data. *Stat*, **3**(1), 48–55.
Castruccio, S. & Genton, M. G. (2016). Compressing an ensemble with statistical models: An algorithm for global 3d spatio-temporal temperature. *Technometrics*, **58**(3), 319–328.
Castruccio, S. & Guinness, J. (2017). An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. *J. R. Stat. Soc.: Series C*, **66**(2), 329–344.
Castruccio, S. & Stein, M. L. (2013). Global space-time models for climate ensembles. *Ann. Appl. Stat.*, **7**(3), 1593–1611.
Chen, D., Menegatto, V. A. & Sun, X. (2003). A necessary and sufficient condition for strictly positive definite functions on spheres. *Proceedings of the American Mathematical Society*, **131**, 2733–2740.
Christakos, G. (1991). On certain classes of spatiotemporal random fields with application to space-time data processing. *Syst. Man Cybern.*, **21**(4), 861–875.
Christakos, G. (2000). *Modern Spatiotemporal Geostatistics.* Oxford University Press.
Christakos, G., Hristopoulos, D. T. & Kolovos, A. (2000). Stochastic flowpath analysis of multiphase flow in random porous media. *SIAM J. Appl. Math.*, **60**(5), 1520–1542.
Christakos, G., Hristopulos, D. T. & Bogaert, P. (2000). On the physical geometry hypotheses at the basis of spatiotemporal analysis of hydrologic geostatistics. *Adv. Water Resour.*, **23**, 799–810.
Christakos, G. & Papanicolaou, V. (2000). Norm-dependent covariance permissibility of weakly homogeneous spatial random fields. *Stochastic Environmental Research and Risk Assessment*, **14**(6), 1–8.
Christopoulos, D. & Tsantili, I. (2016). Space-time models based on random fields with local interactions. *Internat. J. Modern Phys. B*, **29**, 26.
Clarke, J., Alegria, A. & Porcu, E. (2018). Regularity Properties and Simulations of Gaussian Random Fields on the Sphere cross Time. Technical Report, University Federico Santa Maria. *Electron. J. Stat.*, **12**(1), 399–426.
Cressie, N. & Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.*, **94**(448), 1330–1340.
Cressie, N. & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc., Series B*, **70**, 209–226.
Cressie, N. & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data.* New York: John Wiley & Sons.
Cressie, N., Shi, T. & Kang, E. (2010). Fixed rank filtering for spatio-temporal data. *J. Comput. Graph. Statist.*, **19**(3), 724–745.
Crippa, P., Castruccio, S., Archer-Nicholls, G. B., Lebron, M., Kuwata, A., Thota, S., Sumin, E., Butt, C., Wiedinmyer, W. & Spracklen, D. V. (2016). Population exposure to hazardous air quality due to the 2015 fires in equatorial Asia. *Scientific Reports*, **6**, Article number: 37074.
Cuevas, F., Porcu, E. & Bevilacqua, M. (2017). Contours and dimple for the gneiting class of space-time correlation functions. *Biometrika*, **104**(4), 995–1001.
Dai, F. & Xu, Y. (2013). *Approximation Theory and Harmonic Analysis on Spheres and Balls.* New York: Springer.
Daley, D. J. & Porcu, E. (2013). Dimension walks and schoenberg spectral measures. *Proceedings of the American Mathematical Society*, **141**, 1813–1824.
Estrade, A., Fariñas, A. & Porcu, E. (2016). *Characterization Theorems for Covariance Functions on the n-Dimensional Sphere Across Time*. Technical Report, University Federico Santa Maria.
Fassó, A., Finasi, F. & Ndongo, F. (2016). European population exposure to airborne pollutants based on a multivariate spatio-temporal model. *Journal of Agricultural, Biological, and Environmental Statistics*, **21**(3), 492–511.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications (Vol. II).* New York: John Wiley & Sons.
Finazzi, F. & Fassó, A. (2014). D-STEM: A software for the analysis and mapping of environmental space-time variables. *Journal of Statistical Software*, **62**, 1–29.
Fonseca, T. C. O. & Steel, M. F. J. (2011). A general class of nonseparable space-time covariance models. *Environmetrics*, **22**(2), 224–242.
Fuglstad, G. A., Simpson, D., Lindgren, F. & Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spat. Stat.*, **14**, 505–531.
Furrer, R., Bachoc, F. & Du, J. (2016). Asymptotic Properties of Multivariate Tapering for Estimation and Prediction. *J. Multivar. Anal.*, **149**, 177–191.
Furrer, R., Genton, M. G. & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.*, **15**, 502–523.
Furrer, R., Sain, S. R., Nychka, D. W. & Meehl, G. A. (2007). Multivariate bayesian analysis of atmosphere-ocean general circulation models. *Environ. Ecol. Stat.*, **14**(3), 249–266.
Gangolli, R. (1967). Positive definite kernels on homogeneous spaces and certain stochastic processes related to levy's brownian motion of several parameters. *Ann Inst H Poincare*, **3**, 121–226.
Gantmacher, F. (1960). *The Theory of Matrices.* Chelsea Pub Co.
Geinitz, S., Furrer, R. & Sain, S. R. (2015). Bayesian multilevel analysis of variance for relative comparison across sources of global climate model variability. *International Journal of Climatology*, **35**(3), 433–443.
Gerber, F., Mösinger, L. & Furrer, R. (2017). Extending R packages to support 64-bit compiled code: An illustration with spam64 and GIMMS NDVI $_{3g}$ data. *Computers & Geosciences*, **104**, 107–119.
Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.*, **97**, 590–600.
Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19**(4), 1327–1349.
Gneiting, T., Genton, M. G. & Guttorp, P. (2007). Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry. In *Statistical methods for spatio-temporal systems*, pp. 151–175.
Guella, J. C., Menegatto, V. A. & Peron, A. P. (2016a). Strictly positive definite kernels on a product of spheres ii. *SIGMA*, **12**(103).
Guella, J. C., Menegatto, V. A. & Peron, A. P. (2016b). An extension of a theorem of Schoenberg to a product of spheres. *Banach J. Math. Anal.*, **10**(4), 671–685.
Guella, J. C., Menegatto, V. A. & Peron, A. P. (2017). Strictly positive definite kernels on a product of circles. *Positivity*, **21**(1), 329–342.
Guinness, J. & Fuentes, M. (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. *J. Multivariate Anal.*, **143**, 143–152.
Gupta, V. K. & Waymire, E. (1987). On Taylor's hypothesis and dissipation in rainfall. *Journal of Geophysical Research*, **92**(3), 9657–9660.
Hannan, E. J. (1970). *Multiple Time Series.* New York: John Wiley & Sons.
Hansen, L. V., Thorarinsdottir, T. L., Ovcharov, E. & Gneiting, T. (2015). Gaussian random particles with flexible Hausdorff dimension. *Adv. in Appl. Probab.*, **47**(2), 307–327.
Haslett, J. & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Irelands wind-power resource. *Applied Statistics*, **38**, 1–50.
Hitczenko, M. & Stein, M. L. (2012). Some theory for anisotropic processes on the sphere. *Statistics Methodology*, **9**, 211–227.
Holben, B. N. *et al.* (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, **66**, 1–16.
Horrell, M. T. & Stein, M. L. (2015). A covariance parameter estimation method for polar-orbiting satellite data. *Statistica Sinica*, **25**(1), 41–59.
Huang, C., Zhang, H. & Robeson, S. (2012). A simplified representation of the covariance structure of axially symmetric processes on the sphere. *Statist. Probab. Lett.*, **82**, 1346–1351.
IPCC. (2013). Climate Change 2013: The Physical Science Basis. In *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Eds. Stocker, T. F. *et al.* New York: Cambridge.
Istas, J. (2005). Spherical and hyperbolic fractional Brownian motion. *Electronic Communications in Probability*, **10**, 254–262.
Jeong, J. & Jun, M. (2015). A class of Matérn-like covariance functions for smooth processes on a sphere. *Spat. Stat.*, **11**, 1–18.
Jeong, J., Jun, M. & Genton, M. G. (2017). Spherical process models for global spatial statistics. *Stat. Sci.*, **32**, 501–513.
Jones, R. & Zhang, Y. (1997). Models for Continuous Stationary Space-Time Processes. In *Modelling longitudinal and spatially correlated data*, lecture notes in statistics, pp. 289–298.

Jones, R. H. (1963). Stochastic processes on a sphere. *The Annals of Mathematical Statistics*, **34**, 213–218.
Jun, M. (2011). Non-stationary cross-covariance models for multivariate processes on a globe. *Scand. J. Stat.*, **38**, 726–747.
Jun, M. & Stein, M. L. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics*, **49**, 468–479.
Jun, M. & Stein, M. L. (2008). Nonstationary covariance models for global data. *Ann. Appl. Stat.*, **2**(4), 1271–1289.
Kang, E. L., Cressie, N. & Shi, T. (2010). Using temporal variability to improve spatial mapping with application to satellite data. *Canad. J. Statist.*, **38**(2), 271–289.
Kent, J. T., Mohammadzadeh, M. & Mosammam, A. M. (2011). The dimple in Gneiting's spatial-temporal covariance model. *Biometrika*, **98**(2), 489–494.
Kolovos, A., Christakos, G., Hristopulos, D. T. & Serre, M. L. (2004). Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Environmental Applications, Advances in Water Resources*, **27**(8), 815–830.
Lang, A. & Schwab, C. (2013). Isotropic random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Annals of Applied Probabilty*, **25**, 3047–3094.
Leonenko, N. & Sakhno, L. (2012). On spectral representation of tensor random fields on the sphere. *Stoch. Anal. Appl.*, **31**, 167–182.
Lindgren, F. & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.*, **63**(19), 1–25.
Lindgren, F., Rue, H. & Lindstroem, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc.: Series B*, **73**, 423–498.
Malyarenko, A. (2013). *Invariant Random Fields on Spaces with a Group Action.* Springer, New York.
Marinucci, D. & Peccati, G. (2011). *Random Fields on the Sphere, Representation, Limit Theorems and Cosmological Applications.* New York: Cambridge.
Massa, E., Peron, A. & Porcu, E. (2017). Positive definite functions on complex spheres, and their walks through dimensions. *SIGMA*, **13**, 088.
Mateu, J., Porcu, E. & Gregori, P. (2008). Recent advances to model anisotropic space-time data. *Stat. Methods Appl.*, **17**, 209–223.
Menegatto, V. A. (1994). Strictly positive definite kernels on the Hilbert sphere. *Appl. Anal.*, **55**, 91–101.
Menegatto, V. A. (1995). Strictly positive definite kernels on the circle. *Rocky Mountain J. Math.*, **25**, 1149–1163.
Menegatto, V. A., Oliveira, C. P. & Peron, A. P. (2006). Strictly positive definite kernels on subsets of the complex plane. *Comput. Math. Appl.*, **51**, 1233–1250.
Møller, J., Nielsen, M., Porcu, E. & Rubak, E. (2018). Determinantal point process models on the sphere. *Bernoulli*, **24**(2), 1171–1201.
Molod, A., Takacs, L., Suarez, M. & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356.
Narcowich, F. J. (1995). Generalized Hermite interpolation and positive definite kernels on a Riemannian manifold. *J. Math. Anal. Appl.*, **190**, 165–193.
National Environment Agency (2016). *Air pollution control*. Retrieved from http://www.nea.gov.sg/anti-pollution-radiation-protection/air-pollution-control/.
Nguyen, H., Katzfuss, M., Cressie, N. & Braverman, A. (2014). Spatio-temporal data fusion for very large remote sensing datasets. *Technometrics*, **56**(2), 174–185.
Oleson, J. J., Kumar, N. & Smith, B. J. (2013). Spatiotemporal modeling of irregularly spaced aerosol optical depth data. *Environ. Ecol. Stat.*, **20**(2), 297–314.
Paciorek, C. J. & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, **17**(5), 483–506.
Porcu, E. & Mateu, J. (2007). Mixture-based modeling for space-time data. *Environmetrics*, **18**, 285–302.
Porcu, E. & Schilling, R. (2011). From Schoenberg to Pick-Nevanlinna: Towards a complete picture of the variogram class. *Bernoulli*, **17**(1), 441–455.
Porcu, E. & Zastavnyi, V. P. (2011). Characterization theorems for some classes of covariance functions associated to vector valued random field. *J. Multivar. Anal.*, **102**, 1293–1301.
Porcu, E., Mateu, J. & Bevilacqua, M. (2007). Covariance functions which are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica*, **61**(3), 358–382.
Porcu, E., Mateu, J. & Christakos, G. (2009). Quasi-arithmetic means of covariance functions with potential applications to space-time data. *J. Multivar. Anal.*, **100**(8), 1830–1844.
Porcu, E., Bevilacqua, M. & Genton, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Am. Stat. Assoc.*, **11**, 888–898.

Porcu, E., Bevilacqua, M. & Genton, M. G. (2016). *Space-Time Covariance Functions with Dynamical Compact Supports*. Technical Report, University Federico Santa Maria.

Porcu, E., Zastavnyi, V. P. & Bevilacqua, M. (2017). Buhmann covariance functions, their compact supports, and their smoothness. *Dolomite Research Notes on Approximation*, **10**, 33–42.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Stat. Soc.: Series B*, **27**, 204–237.

Richardson, R., Kotta, A. & Sansó, B. (2016). Bayesian non-parametric modeling for integro-difference equations. *Stat. Comput.*, **1–15**.

Richardson, R., Kotta, A. & Sansó, B. (2017). Flexible integro-difference equation modeling for spatio-temporal data. *Comput. Statist. Data Anal.*, **109**, 182–198.

Rue, H. & Tjelmeland, H. (2002). Fitting Gaussian markov random fields to Gaussian fields. *Scand. J. Stat.*, **29**, 31–49.

Schilling, R., Song, R. & Vondracek, Z. (2012). *Bernstein Functions. Theory and Applications.* De Gruyter.

Schlather, M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli*, **16**(3), 780–797.

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Ann. Math.*, **25**(39), 811–841.

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

Shannon, N., Koplitz, L. J., Mickley, M. E., Marlier, J. J., Buonocore, P. S., Kim, T. L., Sulprizio, M. P., DeFries, R., Jacob, D. J., Schwartz, J., Pongsiri, M. & Myers, S. S. (2016). Public health impacts of the severe haze in equatorial Asia in September–October 2015: Demonstration of a new framework for informing fire management strategies to reduce downwind smoke exposure. *Environ. Res. Lett.*, **11**(9), Letter number: 094023.

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer.

Stein, M. L. (2005). Space-time covariance functions. *J. Am. Stat. Assoc.*, **100**(469), 310–321.

Stein, M. L. (2005). Statistical methods for regular monitoring data. *J. R. Stat. Soc.: Series B*, **67**, 667–687.

Stein, M. L. (2007). Spatial variation of total column ozone on a global scale. *Ann. Appl. Stat.*, **1**, 191–210.

Szegő, G. (1939). *Orthogonal Polynomials*, Vol. XXIII. COLLOQUIUM PUBLICATIONS, American Mathematical Society.

Tebaldi, C. & Sansó, B. (2009). Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *J. R. Stat. Soc.: Series A*, **172**, 83–106.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.*, **4**, 389–396.

Whittle, P. (1953). The analysis of multiple stationary time series. *J. R. Stat. Soc.: Series B*, **15**(1), 125–139.

Wikle, C. K., Milliff, R. F., Nychka, D. & Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *J. Am. Stat. Assoc.*, **96**(454), 382–397.

Wolter, K. & Timlin, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.*, **31**, 1074–1087.

Yadrenko, A. M. (1983). *Spectral Theory of Random Fields.* New York: Optimization Software.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*, Vol. I. New York: Springer.

Zastavnyi, V. P. & Porcu, E. (2011). Characterization theorems for the gneiting class of space–time covariance functions. *Bernoulli*, **17**(1), 456–465.

Ziegel, J. (2014). Convolution roots and differentiability of isotropic positive definite functions on spheres. *Proceedings of the American Mathematical Society*, **142**, 2053–2077.

[*Received June 2017, accepted February 2018*]