Title

A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials

Permalink

https://escholarship.org/uc/item/4zs6107g

Journal

The Annals of Applied Statistics, 12(1)

ISSN

1932-6157

Authors

Fan, Minjie
Paul, Debashis
Lee, Thomas CM
et al.

Publication Date

2018-03-01

DOI

10.1214/17-aoas1104

Peer reviewed

A MULTI-RESOLUTION MODEL FOR NON-GAUSSIAN RANDOM FIELDS ON A SPHERE WITH APPLICATION TO IONOSPHERIC ELECTROSTATIC POTENTIALS

Minjie Fan[*], **Debashis Paul**[*], **Thomas C. M. Lee**[*], **Tomoko Matsuo**[†]

[*]Department of Statistics, University of California, Davis, ONE Shields Avenue, Davis, California 95616, USA

[†]Department of Aerospace, Engineering Sciences, University of Colorado, Boulder, 429 UCB, Boulder, Colorado 80309, USA

Abstract

Gaussian random fields have been one of the most popular tools for analyzing spatial data. However, many geophysical and environmental processes often display non-Gaussian characteristics. In this paper, we propose a new class of spatial models for non-Gaussian random fields on a sphere based on a multi-resolution analysis. Using a special wavelet frame, named *spherical needlets*, as building blocks, the proposed model is constructed in the form of a sparse random effects model. The spatial localization of needlets, together with carefully chosen random coefficients, ensure the model to be non-Gaussian and isotropic. The model can also be expanded to include a spatially varying variance profile. The special formulation of the model enables us to develop efficient estimation and prediction procedures, in which an adaptive MCMC algorithm is used. We investigate the accuracy of parameter estimation of the proposed model, and compare its predictive performance with that of two Gaussian models by extensive numerical experiments. Practical utility of the proposed model is demonstrated through an application of the methodology to a data set of high-latitude ionospheric electrostatic potentials, generated from the LFM-MIX model of the magnetosphere-ionosphere system.

Keywords

Non-Gaussian random field; multi-resolution analysis; isotropic process on a sphere; MCMC; ionospheric electrostatic potential; LFM-MIX model

1. Introduction.

Gaussian random fields (GRF) have provided a very successful modeling framework for analyzing spatial data. Two key ingredients of the success of GRF modeling are: (i) the

behavior of the underlying stochastic process is entirely characterized by the mean and covariance functions, thereby facilitating deep theoretical investigations; and (ii) computations for both estimation and prediction primarily involve matrix algebra. However, there are many geophysical and environmental processes that exhibit a significant degree of non-Gaussianity, such as turbulent fields [Barndorff-Nielsen (1979), Berg et al. (2016)], meteorological variables including relative vorticity of wind and oceanic currents, wind velocity, air temperature and humidity [Perron and Sura (2013)], precipitation [De Oliveira, Kedem and Short (1997), Wallin and Bolin (2015)] and ionospheric electric fields [Cousins and Shepherd (2012)].

1.1. High-latitude ionospheric electric field.

The process that motivates the methodological developments in this paper is the high-latitude ionospheric electric field, originating from solar wind-magnetosphere-ionosphere interactions. Energy and momentum deposition associated with highly variable electric fields lead to global disturbance in the Earth's partially ionized upper atmosphere. This impacts the drag force on low-Earth-orbit satellites and debris, deteriorating our ability of tracking these objects to mitigate potential collisions; and affects radio signal propagation, hindering robust performance of modern technological systems including telecommunication, navigation and positioning. In spite of their scientific importance and societal relevance, current general circulation models of the upper atmosphere are incapable of adequately reproducing these phenomena. One of the obstinate issues originates from systematic underestimation of energy and momentum sources resulting from an inadequate representation of the variability of the high-latitude ionospheric electric fields [Codrescu, Fuller-Rowell and Foster (1995), Matsuo, Richmond and Hensel (2003)].

The electric fields exhibit considerable variability across a range of scales in both space and time. Cousins and Shepherd (2012) showed in their analysis of the data obtained from the *Super Dual Auroral Radar Network (SuperDARN)*, an international network of ground-based high-frequency (HF) radars, that the small-scale spatial and temporal variability of the electric fields displays heavier-tailed behavior than a normal distribution. While a lot of efforts have been made to model the large-scale variability as a GRF through decomposing the process into spherical harmonics (SH) or empirical orthogonal functions (EOF) [see Cousins, Matsuo and Richmond (2013a) and references therein], little attempt has been made so far to characterize the non-Gaussian small-scale counterpart. Although SH and EOF basis functions can well represent large-scale features, they are not effective in capturing small-scale details. Moreover, the small-scale variability can form a significant part of the total variability. Through analyzing the SuperDARN data, Cousins, Matsuo and Richmond (2013b) found that together with the mean, the first three EOFs, which are characterized by global spatial scales and long time scales, can only explain approximately 50% of the observed squared electric field.

Most existing numerical simulations of the upper atmosphere general circulation only account for the large-scale electric fields. Because the amount of Joule heating that results from collisions between neutrals and ions drifting under the effects of the electric and magnetic fields is proportional to the square of the electric field, neglecting the small-scale

electric field variability in general circulation models can lead to significant underestimation of energy and momentum inputs into the upper atmosphere. While simple GRF models have been used to parameterize the effects of the large-scale electric field variability on general circulation models [Codrescu et al. (2000), Matsuo and Richmond (2008)], there is no viable stochastic parameterization scheme for generating the non-Gaussian small-scale electric fields in a manner consistent with observations. The focus of this paper is therefore to model the small-scale electric field variability for the purpose of incorporating such a model as an adaptive random field generator in general circulation models to account for missing energy and momentum sources. This is a crucial step toward accurately characterizing the variability of the electric fields by random field modeling, and representing the energy budget of solar wind-magnetosphere-ionosphere interactions in numerical models, for facilitating not only further scientific understanding but also practical applications.

1.2. LFM-MIX model output.

We consider a simulated data set of high-latitude ionospheric electrostatic potentials generated from the LFM-MIX model, a state-of-the-art coupled magnetosphere-ionosphere model [Lyon, Fedder and Mobarry (2004)] capable of running at multiple resolutions [Wiltberger et al. (2017)]. Note that the aforementioned electric fields are electrostatic fields given as the negative gradient of the electrostatic potentials. The domain of the LFM-MIX model relevant to magnetosphere-ionosphere coupling is the high-latitude regions (i.e., latitude 45° or −45°) of the Northern and Southern Hemispheres. LFM-MIX global simulation results were compared to real observations and empirical models of high-latitude ionospheric electrodynamics, and good agreement was reported [Zhang et al. (2011), Wiltberger et al. (2017), Kleiber et al. (2016)]. Kleiber et al. (2013) and Heaton et al. (2015) constructed spatio-temporal statistical emulators of the LFM-MIX model, and using the emulators, they estimated the input parameters of the LFM-MIX model through matching the model output with real observations.

Global simulations using the LFM-MIX model especially at higher resolutions require a considerable amount of high-performance computing resources, which by itself provides another motivation for the stochastic parameterization of the small-scale electric field variability for further numerical investigations using upper atmosphere general circulation models. Herein, we focus on the Quad resolution output of the LFM-MIX model in the high-latitude region of the Northern Hemisphere since the model output in the Southern Hemisphere can be analyzed similarly. Through an exploratory analysis of the Quad resolution model output, we found that the large-scale features (i.e., the mean and the first four EOFs) of the electrostatic potentials in the Northern Hemisphere show patterns similar to those derived from real observations in Matsuo, Richmond and Nychka (2002), Cousins, Matsuo and Richmond (2013b) [see Section S.2.1 of the supplementary material Fan et al. (2018)]. Figure 1 shows the small-scale component of the high-latitude ionospheric electrostatic potentials in the Northern Hemisphere after subtracting the estimated large-scale component (see Section 5.1). The non-Gaussianity of the small-scale component is illustrated by the Q–Q plots of the replications over time at different locations. Thus, the small-scale component of the high-latitude ionospheric electrostatic potentials is naturally a non-Gaussian random field on a sphere.

1.3.   Challenges and contributions.

One popular approach for dealing with non-Gaussian processes is to transform the data by some nonlinear function such that they can be modeled by a GRF; see Cressie (1993), De Oliveira, Kedem and Short (1997), Xu and Genton (2017) for examples. The success of this approach relies heavily on finding a suitable transformation for the data. As the covariance structure of the latent GRF becomes more complicated, the transformation could make the derived non-Gaussian process more difficult to interpret. Among various approaches to modeling non-Gaussian processes directly without any transformation, Palacios and Steel (2006) proposed a class of non-Gaussian spatial models based on scale mixing of a Gaussian process; Røislien and Omre (2006) developed a *t*-distributed random field model with heavy-tailed marginal probability density functions, which is a generalization of the familiar multivariate *t*-distribution; and Wallin and Bolin (2015) derived non-Gaussian geostatistical models with a Matérn covariance structure from stochastic partial differential equations driven by non-Gaussian noise.

Modeling random fields on a sphere instead of a Euclidean space poses an additional technical challenge. Processes on a sphere are usually constructed by restricting processes on $\mathbb{R}^3$ to the sphere, but this may cause physically unrealistic distortions, especially for the covariance structure at long distances [Gneiting (2013)]. There have been efforts to develop valid covariance functions on a sphere directly [see Jun and Stein (2008), Guinness and Fuentes (2016) for examples], but having a covariance function only does not suffice to fully define a non-Gaussian process. Stein (2007) and Cressie and Johannesson (2008), among others, used a fixed rank approach in which GRFs on a sphere are approximated by a linear combination of basis functions with normally distributed coefficients. The stochastic properties of the resulting process are determined by the distribution of the coefficients and the choice of the basis functions.

In this paper, we propose a new class of multi-resolution spatial models for non-Gaussian random fields on a sphere, which are constructed to mimic the characteristics of the aforementioned high-latitude ionospheric electrostatic potentials. The construction of the model starts with a sparse random effects model, represented in terms of a special multi-resolution wavelet frame named spherical needlets. The spatial localization of needlets, together with carefully chosen random coefficients in this representation, ensure that the resulting process is non-Gaussian and isotropic. The model is then expanded to include a spatially varying variance profile. Motivated by the specific application, we consider a special case where the variance depends on the latitude only. Analytical properties of needlets and the special formulation of the model enable us to develop efficient estimation and prediction procedures.

We apply the proposed model to the simulated high-latitude ionospheric electrostatic potentials generated from the LFM-MIX model. The results reveal that the amount of Joule heating is significantly increased by taking into account and modeling the small-scale variability of the electrostatic potentials using the proposed model. This indicates the necessity of including the small-scale variability in the calculation of the Joule heating. In comparison, modeling the small-scale variability using the popular Gaussian Matérn model

fails to produce significantly more energy than the large-scale component only. In current general circulation models, the Joule heating rate is underestimated, and hence often multiplied by an arbitrary factor such that the latitudinal distribution of the modeled upper atmosphere temperature matches well with observed climatology [Matsuo and Richmond (2008)]. However, such an ad hoc treatment of the energy source has a number of shortcomings. In particular, since the electric fields also drive winds through the ion-drag force, the energy and momentum resulting from the high-latitude electric fields cannot be arbitrarily altered and need to be consistently treated. We also notice that the proposed model has a higher chance of producing extremely high energy than its Gaussian version due to the non-Gaussianity of the model. In fact, the upper atmosphere responds to extreme energy and momentum sources differently from moderately elevated ones. Thus, the proposed model's ability to produce extreme energy and momentum sources is promising to improve the accuracy of numerically modeling the magnetosphere-ionosphere coupling.

The remainder of the paper is organized as follows. In Section 2, we first give a brief introduction to spherical needlets, and using needlets as building blocks, we then construct a sparse random effects model for non-Gaussian random fields on the unit sphere. The model is also expanded to include a spatially varying variance profile. In Section 3, we describe the computational details of model fitting, prediction and unconditional simulation. Their good performance is demonstrated in Section 4 by extensive numerical experiments. In Section 5, we apply the proposed model to the high-latitude ionospheric electrostatic potentials. Some relevant issues and future directions are discussed in Section 6.

2. Model construction for non-Gaussian random fields.

In this section, we construct a class of spatial models for non-Gaussian random fields on the unit sphere through a sparse random effects model that uses a multi-resolution representation in the form of a spherical needlet frame.

2.1. Spherical needlets.

We first give a brief introduction to spherical needlets and some of their important properties. More details can be found in Marinucci and Peccati (2011), Narcowich, Petrushev and Ward (2006), Fan (2015).

Let $\mathbf{s}$ denote a point on the unit sphere $\mathbb{S}^2 = \left\{ \mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1 \right\}$, and (θ, φ) represent the same point in spherical coordinates, where θ and φ are the colatitude and longitude, respectively. Complex-valued spherical harmonics (SH), denoted by $\{ Y_{lm}(\theta, \varphi), l = 0, 1, \ldots, m = -l, \ldots, l \}$, are the spherical analogue of the Fourier basis on the unit circle. They form an orthonormal basis for the Hilbert space $L^2(\mathbb{S}^2)$, the space of square integrable functions on the unit sphere. The index l determines the frequency level of SH functions.

Needlets are constructed in terms of SH functions based on two key ideas, namely, (a) a discretization of the unit sphere; and (b) a Littlewood–Paley decomposition. The discretization of $\mathbb{S}^2$ is achieved by an exact *quadrature formula*: for every $l \in \mathbb{N}$, there exist a

finite subset $\mathcal{X}_l = \left\{ \xi_{lk} \right\}_{k=1}^{n_l} \subset \mathbb{S}^2$ (*quadrature points*) and positive weights $\left\{ \lambda_{lk} \right\}_{k=1}^{n_l}$ (*quadrature weights*) such that, for any polynomial f of degree at most l,

$$\int_{\mathbb{S}^2} f(\mathbf{s}) \, d\mathbf{s} = \sum_{k=1}^{n_l} \lambda_{lk} f(\xi_{lk}). \tag{2.1}$$

The Littlewood–Paley decomposition is defined through a function b on $\mathbb{R}^+$ satisfying: (i) $b(\cdot) > 0$ on (B^{-1}, B) for some $B > 1$, and equal to zero on $(B^{-1}, B)^c$; (ii) $\sum_{j=0}^{\infty} b^2(y/B^j) = 1$ for all $y \ge 1$ (iii) $b(\,\cdot\,) \in C^M(\mathbb{R}^+)$ for some $M \in \mathbb{N} \cup \{\infty\}$. Based on these specifications, a class of spherical needlets is defined as follows. For $\mathbf{s} \in \mathbb{S}^2$.

$$
\begin{aligned}
\psi_{jk}(\mathbf{s}) &= \sqrt{\lambda_{jk}} \sum_{l = \lceil B^{j-1} \rceil}^{\lfloor B^{j+1} \rfloor} b\!\left(\frac{l}{B^j}\right) \sum_{m=-l}^{l} Y_{lm}(\xi_{jk}) \bar{Y}_{lm}(\mathbf{s}) \\
&= \sqrt{\lambda_{jk}} \sum_{l = \lceil B^{j-1} \rceil}^{\lfloor B^{j+1} \rfloor} b\!\left(\frac{l}{B^j}\right)\!\left(\frac{2l+1}{4\pi}\right) P_l\!\left(\langle \xi_{jk}, \, \mathbf{s} \rangle\right),
\end{aligned}
\tag{2.2}
$$

where $j \in \mathbb{N} \cup \{0\}$ encodes the *scale* or *frequency* of needlets, (with a slight abuse of notation) $\left\{ \xi_{jk} \right\}_{k=1}^{p_j}$ are a set of quadrature points, and $\left\{ \lambda_{jk} \right\}_{k=1}^{p_j}$ are the corresponding quadrature weights, where $p_j = n_{Cj}$ with $C_j = 2\lfloor B^{j+1} \rfloor$ and n_l is as in (2.1). The point ζ_{jk} determines the *location* of the needlet ψ_{jk}, for each $k = 1, \ldots, p_j$. Besides, $\langle \cdot, \cdot \rangle$ denotes the dot product on $\mathbb{R}^3$, and P_l represents the lth Legendre polynomial. Note that the last equality holds due to the *Addition Theorem for SH functions* [Atkinson and Han (2012), Theorem 2.9], and it also shows that needlets are real-valued functions.

Needlets possess several attractive properties. They are localized in both the spatial and frequency domains, with quasi-exponentially increasing concentration around the quadrature point ζ_{jk} as the frequency level j increases. Together with Y_{00}, a constant function on the unit sphere, the collection of needlets also form a *Parseval tight frame*, that is, for any $f \in L^2(\mathbb{S}^2)$,

$$\int_{\mathbb{S}^2} |f(\mathbf{s})|^2 d\mathbf{s} = \left| \langle f, Y_{00} \rangle_{L^2} \right|^2 + \sum_{j=0}^{\infty} \sum_{k=1}^{p_j} \left| \langle f, \psi_{jk} \rangle_{L^2} \right|^2,$$

where $\langle \,\cdot\,, \,\cdot\, \rangle_{L^2}$ denotes the inner product on $L^2(\mathbb{S}^2)$ and

$$\langle f, \psi_{jk} \rangle_{L^2} = \int_{\mathbb{S}^2} f(\mathbf{s}) \psi_{jk}(\mathbf{s}) d\mathbf{s}$$

is called the *needlet coefficient* of *f* corresponding to the index pair (*j*, *k*), denoted by β_{jk}. The tight frame property, together with the localization in both the spatial and frequency domains, ensure that needlets can be used to perform a *multiresolution analysis* of functions in $L^2(\mathbb{S}^2)$.

In our construction, we use the symmetric spherical *t*-designs on $\mathbb{S}^2$ [Womersley (2015)] as the quadrature points. The number of quadrature points $n_l = l^2/2 + l/2 + \mathcal{O}(1)$, and the quadrature weights are all equal to $4\pi/n_l$. Moreover, the function b is chosen based on the second specification in Marinucci and Peccati (2011, Chapter 10.2.2) with $b(\cdot) \in C^\infty$ [see Section S.1.1 of the supplementary material Fan et al. (2018) for details], and *B* is specified as 2 according to Narcowich, Petrushev and Ward (2006).

2.2. Sparse random effects model.

In this subsection, we construct a sparse random effects model for scalar random fields on the unit sphere using needlets as building blocks. Let $\left\{X(\mathbf{s}): \mathbf{s} \in \mathbb{S}^2\right\}$ be a zero-mean scalar random field on the unit sphere, which can be represented in terms of needlets, that is,

$$X(\mathbf{s}) = \sum_{j=J_0}^{J} \sum_{k=1}^{p_j} c_{jk} \psi_{jk}(\mathbf{s}),$$

(2.3)

where $J \geq J_0 \geq 0$ determine the frequency range of the information contained in the process. In general, the representation is not unique due to the fact that $\{\psi_{jk}\}_{j,k}$ form a tight frame, which is an overcomplete system. Thus, we need to impose certain probabilistic assumptions on the random coefficients c_{jk}'s such that meaningful inference on model parameters is enabled with pooling of information. Another consideration is to make the process *X* non-Gaussian. This can be achieved through assuming that c_{jk}'s have a non-Gaussian joint probability distribution. In order to incorporate both scale-dependent variations and non-Gaussianity, we make the following key structural assumptions:

- For each index pair (*j*, *k*), there are parameters $\nu_{jk} > 2$ and $\sigma_{jk} > 0$ such that $c_{jk} \sim \sigma_{jk} t(\nu_{jk})$, where $t(\nu)$ denotes the *t*-distribution with ν degrees of freedom. Besides, c_{jk}'s are independent.

The independence assumption of the coefficients c_{jk}'s is motivated by the fact that, for a two-weakly isotropic process with mild regularity conditions, the correlation among pairs of needlet coefficients decay rapidly as the corresponding quadrature points and frequency levels become more separated [Baldi et al. (2009), Marinucci and Peccati (2011)].

The *t*-distribution assumption and the spatial localization of needlets ensure non-Gaussian stochastic properties of the process *X* because for any given location **s**, $X(\mathbf{s})$ is approximately a weighted sum of a small number of c_{jk}'s, which are independent and non-Gaussian. The same does not hold if needlets are replaced by basis functions with global support, such as SH functions, since in that case, $X(\mathbf{s})$ becomes a weighted sum of a large number of c_{jk}'s and the central limit theorem kicks in, resulting in approximately Gaussian

behavior. This specification also has a distinct computational advantage, which will be discussed in Section 3.2.

We have the following theorem characterizing the covariance structure of the process X in a special case.

Theorem 1. If we consider the special case where $\sigma_{jk} = \sigma_j$ and $\nu_{jk} = \nu_j$ with $\nu_j > 2$ for all $k = 1, \ldots, p_j$, the resulting process X is two-weakly isotropic with an oscillating covariance function

$$C(\mathbf{s}, \mathbf{t}) = \mathrm{Cov}(X(\mathbf{s}), X(\mathbf{t})) = \sum_{j = J_0}^{J} \frac{\nu_j \sigma_j^2}{\nu_j - 2} \sum_l b^2\left(\frac{l}{B^j}\right)\left(\frac{2l + 1}{4\pi}\right) P_l(\langle \mathbf{s}, \mathbf{t}\rangle), \qquad (2.4)$$

where $\mathbf{s}, \mathbf{t} \in \mathbb{S}^2$. The covariance function decays quasi-exponentially with respect to the great-circle distance between the two locations $\mathbf{s}$ and $\mathbf{t}$

$$|C(\mathbf{s}, \mathbf{t})| \leq \sum_{j = J_0}^{J} \frac{\nu_j \sigma_j^2}{\nu_j - 2} \frac{c_M B^{2j}}{\left[1 + B^j \arccos(\langle \mathbf{s}, \mathbf{t}\rangle)\right]^M}, \qquad (2.5)$$

where c_M is some constant depending on M. Obviously, when $b(\cdot) \in C^\infty$, (2.5) holds for all $M \in \mathbb{N}$.

The proof is deferred to Appendix A. The isotropy of the process is mainly due to the assumption that σ_{jk} and ν_{jk} do not vary with respect to k. Extensions to anisotropic models are possible by allowing σ_{jk} and ν_{jk} to vary spatially, for example, $\sigma_{jk} = f(\zeta_{jk})\sigma_j$ and $\nu_{jk} = g(\zeta_{jk})\nu_j$ for certain functions f and g. For simplicity, in the rest of the paper, we focus on the special case where $\sigma_{jk} = \sigma_j$ and $\nu_{jk} = \nu_j = \nu$.

There are some connections of the derived covariance function (2.4) with existing literature. Lindgren, Rue and Lindström (2011) discussed an approach for constructing scalar GRFs on $\mathbb{R}^d$ and $\mathbb{S}^2$ with oscillating covariance functions through stochastic partial differential equations. These covariance functions have an oscillating structure similar to our proposal, but they do not necessarily decay quasi-exponentially with respect to the great-circle distance. Moreover, there is no closed-form expression for the covariance functions when they are defined on $\mathbb{S}^2$. Schoenberg (1942) showed that $\{C(\theta) : 0 \leq \theta \leq \pi\}$ is a valid isotropic continuous covariance function on $\mathbb{S}^2$ if and only if it has the following representation in terms of Legendre polynomials:

$$C(\theta) = \sum_{l = 0}^{\infty} a_l P_l(\cos \theta),$$

where θ is the great-circle distance between two locations, $a_l \geq 0$ for all $l \geq 0$ and $\sum_{l = 0}^{\infty} a_l < \infty$. This general form was used by Terdik et al. (2015) and Guinness and Fuentes (2016) through specifying specific a_l's to construct isotropic covariance functions on $\mathbb{S}^2$.

We now consider an observation model based on the above formulation. Suppose that we have observations on X at n different locations, $\mathbf{s}_1,\ldots,\mathbf{s}_n$, and these observations, denoted by $Z(\mathbf{s}_i)$, are corrupted by observational errors. Then

$$Z(\mathbf{s}_i) = X(\mathbf{s}_i) + e_i, \ i = 1, \ldots, n, \tag{2.6}$$

where e_i's are the observational errors modeled as i.i.d. $\mathcal{N}(0, \tau^2)$. In practice, we can specify the parameter J as some value $J_{\max}$ based on data resolution. The parameter σ_j usually decays as the frequency level j increases. For example, we may assume that $\sigma_j = f(j)\sigma$, where $f(j) = B^{-aj/2}$ with $a > 2$ [see Section S.1.2 of the supplementary material Fan et al. (2018) for justification]. The hyperparameter a controls the decay rate of the magnitude of the coefficients c_{jk}'s from low to high frequency levels. In general, σ_j can be estimated without imposing any particular structure.

In Figure 2, we give a few examples of the correlation function of the process X with $\sigma_j = 2^{-aj/2}$ for various a and ranges of j. The curve oscillates more dramatically as a increases, and decays more rapidly as J_0 becomes larger. The parameter ν has no impact on the correlation structure, but it measures the degree of non-Gaussianity of the process. In all the cases, the curve decays quasi-exponentially fast as we have shown in Theorem 1.

To demonstrate the effect of the non-Gaussianity assumption for the coefficients c_{jk}'s, we simulate the process X when c_{jk} is distributed as $\sigma_j t(\nu)$ and normally distributed with the same variance, respectively. We specify $\sigma_j = 2^{-aj/2}$ with $a = 3$, $J_0 = 2$, $J = 4$ and $\nu = 2.5$. Figure 9 in Section S.3 of the supplementary material [Fan et al. (2018)] shows the projection of the simulations onto frequency levels $j = 3,4$, that is, $X^{(j)} := \sum_{k=1}^{p_j} c_{jk} \psi_{jk}$. We can see that the non-Gaussian coefficients yield more positive and negative extreme values than the Gaussian ones. The Q–Q plots of 10,000 i.i.d. simulations of the process X at a specific location, displayed in Figure 10 in Section S.3 of the supplementary material [Fan et al. (2018)], confirm its non-Gaussianity when c_{jk}'s are t-distributed. Checking the non-Gaussianity at one location is sufficient due to the isotropy of the process.

2.3. Axially symmetric model with a variance profile.

Spatial data on a global scale usually exhibit nonstationary behavior. In this subsection, we therefore expand the sparse random effects model to include a spatially varying variance profile. Let $\left\{ g(\mathbf{s}), \ \mathbf{s} \in \mathbb{S}^2 \right\}$ be a function that characterizes the spatially varying variance structure. The function g is modeled through a log-linear basis representation

$$g(\mathbf{s}) = \exp(\mathbf{b}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\eta}), \tag{2.7}$$

where $\mathbf{b}(\mathbf{s}) = (1, b_1(\mathbf{s}),\ldots,b_r(\mathbf{s}))^{\mathrm{T}}$ is a vector of basis functions (including the intercept) evaluated at location $\mathbf{s}$, and $\boldsymbol{\eta} = \left(\eta_0, \boldsymbol{\eta}_{-0}^{\mathrm{T}}\right)^{\mathrm{T}}$ is the corresponding coefficient vector. Through an exploratory analysis of the LFM-MIX model output, we notice that the variance of the high-latitude ionospheric electrostatic potentials shows a strong dependence on the latitude, while only a moderate dependence on the longitude. Motivated by this, for simplicity and

interpretability, we ignore the longitudinal dependence, and assume that $g(\mathbf{s})$ is a function of the co-latitude θ only, that is,

$$g(\mathbf{s}) \equiv g(\theta) = \exp\left(\mathbf{b}^{\mathrm{T}}(\theta)\boldsymbol{\eta}\right). \tag{2.8}$$

Moreover, the variance is assumed to vary smoothly over $\mathbb{S}^2$ and, therefore, the basis functions are specified as cubic B-splines due to their numerical stability, with the first B-spline replaced by the intercept. Then we have the following non-Gaussian model with a variance profile:

$$X(\mathbf{s}) = g(\theta) \sum_{j=J_0}^{J} \sum_{k=1}^{p_j} c_{jk}\psi_{jk}(\mathbf{s}). \tag{2.9}$$

Since the function g depends on the latitude only, this model is axially symmetric [Jones (1963)]. We refer to the model given by (2.8) and (2.9) as the *axially symmetric non-Gaussian needlet model*, abbreviated as *AXING-need*. As a generalization of the model, one may use known functions of spatial covariates instead of $\mathbf{b}(\theta)$ in the log-linear basis representation (2.8) if they are available.

We now consider the simple observation model (2.6) with X defined by (2.9). Denote the vector of observations by $\mathbf{Z} = (Z_1,\ldots,Z_n)^{\mathrm{T}}$, where $Z_i = Z(\mathbf{s}_i)$, and the vector of observational errors by $\mathbf{e} = (e_1,\ldots,e_n)^{\mathrm{T}}$. The AXING-need model can be written in matrix-vector form

$$\mathbf{Z} = \mathbf{GAc} + \mathbf{e}, \tag{2.10}$$

where $\mathbf{G} := \mathrm{diag}\{g(\mathbf{s}_1),\ldots,g(\mathbf{s}_n)\}$, $\mathbf{A}$ is the n-by-$\sum_j p_j$ design matrix with the ith row corresponding to the value of needlets at $\mathbf{s}_i$, and columns ordered according to the index pair (j,k), and $\mathbf{c}$ is the coefficient vector consisting of c_{jk}'s.

3. Model fitting and prediction.

In this section, we describe the computational details of model fitting, prediction and unconditional simulation. The latter two are used to test the performance of the proposed AXING-need model, and compare it with Gaussian models. Computing maximum likelihood estimates (MLE) in the AXING-need model is intractable due to the absence of a closed form of the likelihood for the parameters. Evaluation of the likelihood requires high-dimensional integrals over the distribution of the random coefficients c_{jk}'s. Thus, we adopt a Bayesian approach, and implement a Markov Chain Monte Carlo (MCMC) algorithm that samples from the posterior distribution of the parameters. Before going into the details, we impose the following structural restrictions for the ease and stability of the computation: (i) the parameter ν is fixed and known [Wolfe, Godsill and Ng (2004)] and (ii) the parameters η_0 and σ_{J_0} are nonidentifiable. We choose to fix η_0 at 0 and treat σ_{J_0} as a free parameter, so that the latter can be sampled in a Gibbs step. According to our numerical experiments, this is more efficient than sampling η_0. We denote the vector of the remaining parameters by

$$\boldsymbol{\theta} = \left(\sigma_{J_0}^2, \ldots, \sigma_J^2, \tau^2, \boldsymbol{\eta}_{-0}^{\mathrm{T}}\right).$$

In the following derivation of the algorithm, generic notations are used: $[U]$ denotes the distribution (or density) of a random variable U and $[U|V]$ denotes the conditional distribution (or density) of U given V.

3.1. Prior specification.

We assume that the parameters are *a priori* independent, that is, $[\boldsymbol{\theta}] = \left[\sigma^2_{J_0}\right] \cdots \left[\sigma^2_J\right]\left[\tau^2\right]\left[\boldsymbol{\eta}_{-0}\right]$.

The prior distributions of σ^2_j and τ^2 are specified as the noninformative Jeffreys' priors $\left[\sigma^2_j\right] = 1/\sigma^2_j$ and $[\tau^2] = 1/\tau^2$. The prior distribution of $\boldsymbol{\eta}_{-0}$ is assumed to be $\mathcal{N}\left(\mathbf{0}, \tau^2_{\boldsymbol{\eta}}\mathbf{I}_r\right)$, where the hyperparameter $\tau^2_{\boldsymbol{\eta}}$ is chosen to be sufficiently large such that the prior distribution is nearly noninformative.

3.2. Adaptive MCMC.

Implementation of an MCMC sampling scheme for the AXING-need model seems challenging due to the dimensionality and distribution of the random coefficients c_{jk}'s. One important observation that significantly reduces the computational burden is that the t-distribution (for c_{jk}'s) belongs to the class of scale mixtures of Gaussians (SMOG) [Andrews and Mallows (1974), West (1987)]. Thus, the coefficient c_{jk} can be expressed as $\sqrt{V_{jk}}G_{jk}$, where $V_{jk} \sim \mathcal{IG}\left(v/2, v\sigma^2_j/2\right)$ with $\mathcal{IG}(\alpha, \beta)$ denoting an inverse gamma distribution, and $G_{jk} \sim \mathcal{N}(0, 1)$ independent of V_{jk}. Since c_{jk}'s are independent, the same holds for V_{jk}'s and G_{jk}'s. This representation leads to an efficient MCMC algorithm in which a Gibbs sampler is used.

Let $\mathbf{V}$ denote the vector stacked by V_{jk}'s, and $\boldsymbol{\sigma}^2$ denote the vector consisting of $\sigma^2_{J_0}, ..., \sigma^2_J$.

We use a Gibbs sampler to sample from $[\mathbf{c},\mathbf{V},\boldsymbol{\theta}|\mathbf{Z}]$ so that the full conditional distributions of $\mathbf{c}$, $\mathbf{V}$, $\boldsymbol{\sigma}^2$ and τ^2 all have closed forms. In particular, the full conditional distribution of $\mathbf{c}$ (i.e., $[\mathbf{c}|\mathbf{Z},\mathbf{V},\boldsymbol{\theta}]$) is multivariate Gaussian. Sampling from it requires $\mathcal{O}\left(p^3\right)$ operations with $p = \sum_j p_j$, which is computationally intractable for large p. Nonetheless, our numerical experiments indicate that the subblocks $[\mathbf{c}_j|\mathbf{Z},\mathbf{V},\boldsymbol{\theta}]$, $j = J_0,...,J$ are weakly correlated, where $\mathbf{c}_j = (c_{j1}, ...,c_{jp_j})^\mathrm{T}$. This can be attributed to the fact that needlets are spatially localized and bandlimited. Thus, the sampling step for $\mathbf{c}$ is achieved through successive draws from the conditional subblocks $[\mathbf{c}_j|\mathbf{Z},\mathbf{V},\boldsymbol{\theta},\mathbf{c}_{-j}]$, $j = J_0,...,J$, where $\mathbf{c}_{-j}$ denotes the vector obtained by dropping $\mathbf{c}_j$ from $\mathbf{c}$. Partitioning the vector $\mathbf{c}$ into subblocks avoids the Cholesky decomposition of a large-scale matrix, while the weak correlation among pairs of the subblocks mitigates the potential problem of slow convergence for the Gibbs sampler. For sufficiently large j, we may further speed up the computation by sampling from $[c_{jk}|\mathbf{Z},\mathbf{V},\boldsymbol{\theta},\mathbf{c}_{-jk}]$, $k = 1,...,p_j$. The full conditional distribution of $\boldsymbol{\eta}_{-0}$ is not available in closed form. Thus, we sample from $[\boldsymbol{\eta}_{-0}|\mathbf{Z}, \mathbf{c},\mathbf{V}, \boldsymbol{\sigma}^2, \tau^2]$ using an adaptive Metropolis step [Andrieu and Thoms (2008), Algorithm 4], and incorporate it into the Gibbs sampler.

Let $\mathbf{A}_j$ denote the columns of $\mathbf{A}$ corresponding to j, and $\mathbf{V}_j = (V_{j1},...,V_{jp_j})^\mathrm{T}$. Then the aforementioned adaptive Metropolis-within-Gibbs sampler is summarized as follows:

Algorithm 1.

1. *Sample* $\mathbf{c}_j$ *from* $\left[\mathbf{c}_j | \mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}, \mathbf{c}_{-j}\right] = \mathcal{N}\left(\widehat{\boldsymbol{\mu}}_j, \widehat{\boldsymbol{\Sigma}}_j\right)$, *where*

$$\widehat{\boldsymbol{\Sigma}}_j = \tau^2 \left(\mathbf{A}_j^{\mathrm{T}} \mathbf{G}^2 \mathbf{A}_j + \tau^2 \, \mathrm{diag}\!\left(\mathbf{V}_j\right)^{-1} \right)^{-1},$$

and

$$\widehat{\boldsymbol{\mu}}_j = \frac{1}{\tau^2} \widehat{\boldsymbol{\Sigma}}_j \mathbf{A}_j^{\mathrm{T}} \mathbf{G}\!\left(\mathbf{Z} - \mathbf{G} \mathbf{A}_{-j} \mathbf{c}_{-j} \right).$$

2. *Sample* $\mathbf{V}$ *from* $[\mathbf{V}|\mathbf{Z},\mathbf{c},\boldsymbol{\theta}]$, *where* $V_{jk}|\mathbf{Z},\mathbf{c},\boldsymbol{\theta}$ *are independent and distributed as*

$$\mathcal{IG}\!\left(\frac{\nu + 1}{2}, \frac{c_{jk}^2 + \nu\sigma_j^2}{2} \right).$$

3. *Sample* $\boldsymbol{\sigma}^2$ *from* $[\boldsymbol{\sigma}^2|\mathbf{Z},\mathbf{c},\mathbf{V},\tau^2,\boldsymbol{\eta}]$, *where* $\sigma_j^2|\mathbf{Z}, \mathbf{c}, \mathbf{V}, \tau^2, \boldsymbol{\eta}$ *are independent and distributed as*

$$\mathcal{G}\!\left(\frac{\nu p_j}{2}, \frac{\nu}{2} \sum_{k=1}^{p_j} \frac{1}{V_{jk}} \right),$$

where $\mathcal{G}(\alpha, \beta)$ *denotes a gamma distribution.*

4. *Sample* τ^2 *from*

$$\left[\tau^2 \middle| \mathbf{Z}, \mathbf{c}, \mathbf{V}, \sigma^2, \boldsymbol{\eta} \right] = \mathcal{IG}\!\left(\frac{n}{2}, \frac{(\mathbf{Z} - \mathbf{G}\mathbf{A}\mathbf{c})^{\mathrm{T}}(\mathbf{Z} - \mathbf{G}\mathbf{A}\mathbf{c})}{2} \right).$$

5. 5. *Sample* $\boldsymbol{\eta}_{-0}$ *using the adaptive Metropolis step from*

$$\left[\boldsymbol{\eta}_{-0} \middle| \mathbf{Z}, \mathbf{c}, \mathbf{V}, \sigma^2, \tau^2 \right] \propto \exp\!\left\{ -\frac{1}{2\tau^2}(\mathbf{Z} - \mathbf{G}\mathbf{A}\mathbf{c})^{\mathrm{T}}(\mathbf{Z} - \mathbf{G}\mathbf{A}\mathbf{c}) \right\} \exp\!\left\{ -\frac{1}{2\tau_{\boldsymbol{\eta}}^2} \boldsymbol{\eta}_{-0}^{\mathrm{T}} \boldsymbol{\eta}_{-0} \right\}.$$

The proposal distribution is chosen as

$$Q\!\left(\boldsymbol{\eta}_{-0}^* \middle| \boldsymbol{\eta}_{-0}\right) \sim \mathcal{N}\!\left(\boldsymbol{\eta}_{-0}, \gamma \boldsymbol{\Sigma}\right),$$

where γ is a parameter adaptively tuned with the goal of achieving the optimal acceptance rate [Gelman, Roberts and Gilks (1996)], and $\boldsymbol{\Sigma}$ is adaptively updated to approximate the covariance matrix of the full conditional distribution of $\boldsymbol{\eta}_{-0}$.

Choosing a good initial value of the parameter vector $\boldsymbol{\theta}$ is crucial for successful convergence of the algorithm. We provide a simple but effective method to specify the initial value through fitting the model with $c_{jk} \sim \mathcal{N}\left(0, v\sigma_j^2/(v-2)\right)$ for all j, k. The MLE under this misspecified model, which assumes a Gaussian process for the observations, is easily computable, and close to the true value according to our numerical experiments. Thus, it is reasonable to specify the initial value of $\boldsymbol{\theta}$ in terms of the MLE under the misspecified model. By default, the initial values of $\mathbf{c}$ and $\mathbf{V}$ are specified as a zero vector and an all-ones vector, respectively.

3.3. Spatial prediction.

One main focus of spatial statistics is to predict values at unobserved locations based on observed values. In this subsection, we present a spatial prediction method for the proposed AXING-need model. Let $\mathbf{Z}^* = (Z(\mathbf{t}_1),\ldots,Z(\mathbf{t}_{n_P}))^{\mathrm{T}}$, where $\mathbf{t}_1,\ldots,\mathbf{t}_{n_P}$ are n_P unobserved locations. Then we have $\mathbf{Z}^* = \mathbf{G}^*\mathbf{A}^*\mathbf{c} + \mathbf{e}^*$, where $\mathbf{G}^*$, $\mathbf{A}^*$ and $\mathbf{e}^*$ are $\mathbf{G}$, $\mathbf{A}$ and $\mathbf{e}$ evaluated at the unobserved locations, respectively. From a fully Bayesian perspective, the posterior predictive distribution is

$$[\mathbf{Z}^*|\mathbf{Z} = \mathbf{z}] = \int_{\boldsymbol{\theta},\,\mathbf{c}} [\mathbf{Z}^*|\boldsymbol{\theta}, \mathbf{c}, \mathbf{Z} = \mathbf{z}]\,[\boldsymbol{\theta}, \mathbf{c}|\mathbf{Z} = \mathbf{z}]d\boldsymbol{\theta}\,d\mathbf{c}, \tag{3.1}$$

where $\mathbf{z}$ is the actual observed value of $\mathbf{Z}$, and

$$\mathbf{Z}^*\Big|\boldsymbol{\theta},\, \mathbf{c},\, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}\left(\mathbf{G}^*\mathbf{A}^*\mathbf{c}, \tau^2\mathbf{I}_{n_P}\right). \tag{3.2}$$

The posterior predictive distribution does not have a closed form, but it can be sampled using the MCMC samples from $[\boldsymbol{\theta},\mathbf{c}|\mathbf{Z} = \mathbf{z}]$. Let $\{\boldsymbol{\theta}^{(l)}, \boldsymbol{\theta}^{(l)}, l = 1,\ldots,L\}$ denote the MCMC samples. For each l, we draw a sample $\mathbf{Z}^{*(\mathbf{l})}$ from $[\mathbf{Z}^*|\boldsymbol{\theta}^{(l)},\mathbf{c}^{(l)},\mathbf{Z} = \mathbf{z}]$ in (3.2). The empirical distribution of $\{\mathbf{Z}^{*(l)}, l = 1,\ldots,L\}$ is approximately the posterior predictive distribution. Thus, the posterior predictive mean, standard deviation, quantiles and posterior predictive intervals can be computed based on these samples, respectively. For example, the posterior predictive mean is estimated by $L^{-1}\sum_{l=1}^{L}\mathbf{Z}^{*(l)}$ and the $(1 - a)100\%$ posterior predictive interval is estimated by finding the $(100a/2)$th and $(100(1 - a/2))$th percentiles of $\{\mathbf{Z}^{*(l)}, l = 1,\ldots,L\}$.

3.4. Unconditional simulation.

Utility of spatial models is not restricted to spatial prediction. For example, Genton and Kleiber (2015, Section 7.4) emphasized the importance of using spatial models for simulation of random fields. It is also well known that a misspecified spatial model does not necessarily mean significantly inferior predictive performance [Stein (1988)]. Thus, simulation of random fields from a fitted spatial model provides an alternative way to test the performance of the model. Specifically, it is used to check whether the proposed AXING-need model can successfully capture the characteristics of data. The following algorithm describes the details of simulating a random field from a fitted AXING-need model (2.10) at locations $\mathbf{s}_1, \ldots, \mathbf{s}_{n_S}$.

Algorithm 2.

1. To incorporate the uncertainty of parameter estimates into the simulation, we randomly select one of the MCMC samples $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\}$, *denoted by* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\sigma}_{J_0}^2, \ldots, \widehat{\sigma}_J^2, \widehat{\tau}^2, \widehat{\boldsymbol{\eta}}_{-0} \right)$, *as an estimate of the parameters. The matrix* $\mathbf{G}$ *is evaluated with* $\boldsymbol{\eta}_{-0} = \widehat{\boldsymbol{\eta}}_{-0}$ *at the locations.* $\mathbf{s}_1, \ldots, \mathbf{s}_{n_S}$, *and the design matrix* $\mathbf{A}$ *is also evaluated at these locations*

2. *Generate a coefficient vector* $\mathbf{c}$, *where* $c_{jk} \sim \widehat{\sigma}_j t(\nu)$ *for all j, k, and they are independent.*

3. *Then* $\mathbf{X} = \mathbf{GAc}$ *is an unconditional simulation of random fields from the fitted AXING-need model, where the term of observational errors is omitted.*

4. Numerical experiments.

4.1. Accuracy of parameter estimation.

In this subsection, we investigate the accuracy of parameter estimation by a Monte Carlo simulation study. The data are generated from the proposed AXING-need model (2.10), and the sampling locations are on a mildly perturbed HEALPix grid [Górski et al. (2005)] with 768 grid points. To reduce the computational burden of the simulation study, we assume that there are only two levels of needlets, that is, $J_0 = 2$, $J = 3$. As a function of co-latitude, log g is represented as a linear combination of cubic B-splines with one interior knot $\pi/2$, where the first B-spline is replaced by the intercept. We consider two different settings of the coefficient vector $\boldsymbol{\eta}_{-0}$, and specify $\eta_0 = 0$, $\nu = 4$, $\sigma_2 = 1.25$, $\sigma_3 = 0.4419$, $\tau = 0.1$.

For each simulation run, the parameters $(\boldsymbol{\eta}_{-0}, \sigma_2, \sigma_3, \tau)$ are estimated by the posterior sample means using the adaptive MCMC algorithm in Section 3.2, in which $\tau_\eta = 10$. The chain was run for 400,000 iterations, with the first 200,000 samples discarded as a burn-in period, and every 200th sample is collected to compute the posterior sample means. Figure 3 displays the boxplots of the parameter estimates based on 100 simulation runs. We can see that all the parameter estimates have small biases and low variability. Moreover, the pointwise median curve of the function g matches well with the true one. These demonstrate the effectiveness of the estimation procedure using the adaptive MCMC algorithm.

4.2. Predictive performance comparison.

In this subsection, we compare the predictive performance of the proposed AXING-need model with that of two Gaussian models, which are *Gaussian needlet (Gau-need)* and *Gaussian Matérn (Gau–Matérn)* models, the latter being widely used in spatial statistics. The purpose of this comparison is to understand the effects of two factors on spatial predictive performance: Gaussian versus non-Gaussian, and a Matérn covariance structure versus an oscillating one constructed through a needlet representation of the process. In the Gau-need model, the coefficient vector $\mathbf{c}$ in (2.10) is assumed to be multivariate Gaussian, that is,

$$\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}),$$

where Λ is a diagonal matrix, and $\mathrm{Var}\left(c_{jk}\right) = \sigma_j^2$ for all j, k. The Gau–Matérn model assumes a nonstationary Matérn covariance function with latitude-dependent variance

$$C(\mathbf{s},\ \mathbf{t}) = g(\mathbf{s})\, g(\mathbf{t})\, M(\|\mathbf{s} - \mathbf{t}\|; \kappa, a),$$

where $\mathbf{s},\ \mathbf{t} \in \mathbb{S}^2$, $\|\mathbf{s} - \mathbf{t}\|$ denotes the chordal distance between the two locations $\mathbf{s}$ and $\mathbf{t}$, and g is given by (2.8). Besides, $M(r; \kappa, a)$ is the Matérn correlation function

$$M(r; \kappa, a) = \frac{2^{1-\kappa}}{\Gamma(\kappa)}(ar)^\kappa K_\kappa(ar),$$

where K_κ is the modified Bessel function of the second kind, Γ is the gamma function, $\kappa > 0$ is the smoothness parameter, and $a > 0$ is the spatial scale parameter, whereby 1/a controls the range of correlation.

The data are generated from the AXING-need model, where the parameter ν is specified as 2.5,3,4 to illustrate the impact of the degree of non-Gaussianity on prediction accuracy. The specification of the remaining model parameters and sampling locations are the same as the first setting in Section 4.1. For each cross-validation replication, we randomly select 500 locations outside a fixed longitudinal region with width 30° as the training set to estimate the parameters. The remaining locations are held out as the test set to evaluate the predictive performance. For the AXING-need model, the parameters are estimated by the adaptive MCMC algorithm described in Section 3.2, in which $\tau_\eta = 10$. The chain was run for 1,000,000 iterations, with the first 500,000 samples discarded as a burn-in period, and every 500th sample is collected as the posterior samples. The posterior predictive mean and quantiles (see Section 3.3) are used for prediction. For the Gaussian models, the parameters are estimated by the maximum likelihood method, and the prediction is performed by the usual kriging. The first panel of Figure 4 shows the training and test set separation for one of the cross-validation replications. Held-out locations in and out of the longitudinal region are used to test the predictive performance in terms of long-range and short-range predictions, respectively. We assess the prediction accuracy using five scoring rules: the mean absolute error (MAE), the mean squared prediction error (MSPE), the continuous ranked probability score (CRPS) [Gneiting and Raftery (2007)], and the quantile scores at the levels of 5% and 95% [Gneiting and Ranjan (2011)], averaged over all the predicted locations. The CRPS for the AXING-need model is computed based on the MCMC samples [see Section S.1.3 of the supplementary material Fan et al. (2018) for details]. The quantile score for a quantile forecast q at the level $\alpha \in (0,1)$ is defined as $\mathrm{QS}_\alpha(q, y) = (\mathbb{1}\{y < q\} - \alpha)(q - y)$, where y is the observed value. We repeat the cross-validation procedure 20 times, and summarize the prediction accuracy of the models in terms of the scoring rules in the remaining panels of Figure 4. We can see that when $\nu = 3,4$, the AXING-need and Gau-need models give similar results, and both are better than the Gau–Matérn model. When $\nu = 2.5$, the Gau-need model significantly deviates from the non-Gaussian characteristics of the data, and thus performs worse than the AXING-need model. Again, the Gau–Matérn model has the worst predictive performance among the three. The difference in predictive performance between the Gau-

need and Gau–Matérn models becomes more pronounced for the long-range prediction than the short-range one. This can be explained by the following facts: the Gau–Matérn model can reasonably approximate the true covariance structure at short distances, but it fails to capture the oscillating covariance structure at long distances due to its constraint of nonnegative covariance.

Gneiting, Balabdaoui and Raftery (2007) proposed the concept of *sharpness* to evaluate the predictive performance in probabilistic prediction, which provides a predictive distribution instead of a point prediction for the value of a random field at an unobserved location. Sharpness is measured by the concentration of a predictive distribution around the true value, and the more concentrated the predictive distribution is, the sharper the prediction. We adopt the mean length of predictive intervals ("mean" refers to averaging over all the predicted locations) and the corresponding coverage probability to assess the sharpness of the predictions produced by the models. The posterior predictive interval (see Section 3.3) and the usual symmetric predictive interval are used for the AXING-need model and the Gaussian models, respectively. Table 1 displays the mean lengths of 50% and 90% predictive intervals and the corresponding coverage probabilities, averaged over 20 cross-validation replications. In general, the AXING-need model yields the shortest predictive intervals with satisfactory coverage probabilities.

5. Application to high-latitude ionospheric electrostatic potentials.

In this section, we demonstrate the effectiveness of the proposed AXING-need model and the associated statistical methodology through applying them to the LFMMIX model output: the high-latitude ionospheric electrostatic potentials, which are introduced in Section 1. A detailed description of the LFM-MIX model can be found in Wiltberger et al. (2017). The LFM-MIX model is capable of running at multiple resolutions including Single, Double and Quad. Herein, we use Quad resolution simulations since they contain more small-scale details than the other two. At the Quad resolution, the sampling locations are on a regular 1×1 degree geomagnetic latitude-longitude grid with latitudes ranging from $45°$ to $90°$. The electrostatic potentials are generated every two minutes during the time period from March 20, 2008, to April 16, 2008 (inclusive), and thus there are 20,160 time points in total. We denote the observations by $Z(s_i, t_r)$, $i = 1,\ldots,N$, $r = 1,\ldots,T$, where s_i and t_r represent location and time, respectively.

5.1. Data preprocessing.

To extract the small-scale component of the electrostatic potentials, we subtract a crude estimate of the large-scale component from the observations. The large-scale component is estimated from the data by the empirical orthogonal function (EOF) method, which was applied to the LFM-MIX model output in Kleiber et al. (2013). Let $\mathbf{Z} = (Z(s_i, t_r))_{1 \leq r \leq T, 1 \leq i \leq N}$ denote the $T \times N$ matrix of observations with each row corresponding to a time point. We center $\mathbf{Z}$ by subtracting the vector of column averages from each row. The singular value decomposition (SVD) is then applied to $\mathbf{Z}$, that is, $\mathbf{Z} = \mathbf{UDV}^{\mathrm{T}}$, where $\mathbf{U} = (u_{rk})_{1 \leq r \leq T, 1 \leq k \leq T}$ is a $T \times T$ orthogonal matrix, $\mathbf{D} = \mathrm{diag}(d_1,\ldots, d_T)$ with decreasing singular values d_k's, and $\mathbf{V} = (v_{ik})_{1 \leq i \leq N, 1 \leq k \leq T}$ is an $N \times T$ matrix with orthonormal columns, that is, EOFs.

Element-wise, we can express the observations as $Z(\mathbf{s}_i, t_r) = \sum_{k=1}^{T} d_k u_{rk} v_{ik}$ The cumulative sum of the squared singular values indicates that the first four EOFs can explain approximately 95% of the total variability in the data. Section S.2.1 of the supplementary material [Fan et al. (2018)] gives a scientific explanation of the EOFs through comparing them with those derived from real observed data. Thus, the large-scale component can be estimated by $\sum_{k=1}^{K} d_k u_{rk} v_{ik}$ with $K = 4$, and the residuals are denoted by

$$r(\mathbf{s}_i, t_r) = Z(\mathbf{s}_i, t_r) - \sum_{k=1}^{K} d_k u_{rk} v_{ik}.$$

The sampling locations of the electrostatic potentials are restricted to the highlatitude region of the Northern Hemisphere, while the AXING-need model is constructed for data on a global scale due to the grids on which spherical needlets are placed. There are several ways of handling this problem. First, we can still fit the same model to the data, which are in the high-latitude region, but the needlets outside the region would not have any observations contained in their effective support, and hence the corresponding coefficients could not be effectively estimated from the data [Chu, Clyde and Liang (2009), Section 3.1]. Second, we may simplify the problem as discussed in Heaton et al. (2015) through transforming the data domain to a disk in $\mathbb{R}^2$ by the polar projection, but our model needs to be modified accordingly (it is still under the same framework as a random effects model), especially using an appropriate set of basis functions constructed on the disk instead of the sphere. Herein, we follow the idea of Weimer (1995) and Ruohoniemi and Baker (1998), in which the data are stretched from the high-latitude region to the entire sphere. This is achieved by multiplying the co-latitude θ with a stretching factor α to obtain a new co-latitude $\theta' = \alpha\theta$, where in our case $\alpha = 180/45 = 4$. When applied to random fields, one drawback of this approach is that the correlation structure may be distorted by the stretching, especially near the low-latitude boundary of the data. However, the magnitude of the electrostatic potentials in the lower latitude region ($\theta > 30°$) is close to zero, and the correlation contours in the higher latitude region ($\theta < 30°$) have elongated shapes that are wider in the east-west direction [Cousins, Matsuo and Richmond (2013b)]. These findings suggest that the distortion is not as serious as it seems.

We notice that there are still certain large-scale features remaining in the residuals, which cannot be directly modeled by the AXING-need model since it only captures small-scale details at needlet frequency levels higher than or equal to J_0. Thus, we consider the following augmented model (the term of observational errors is omitted here):

$$r(\mathbf{s}_i, t_r) = g(\mathbf{s}_i', t_r)\left(\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{lm}(t_r) Y_{lm}(\mathbf{s}_i') + \sum_{j=J_0}^{J}\sum_{k=1}^{p_j} c_{jk}(t_r)\psi_{jk}(\mathbf{s}_i')\right),$$

where the first term within the parentheses involving SH functions is included to represent the remaining large-scale features, $\mathbf{s}_i'$ is the point obtained by stretching $\mathbf{s}_i$, and the random coefficients a_{lm} and c_{jk} all depend on the time t_r. To ensure that the SH functions and needlets cover the full frequency range with the minimum overlap, we specify $L = 3$ and $J_0 = 2$ given $B = 2$. We also specify $J = 4$. This indicates that there are in total three levels of

needlets in the model, and the finest details it can capture are at the frequency level $l = 31$ in terms of SH functions. For simplicity, instead of simultaneously estimating the parameters of the SH and needlet terms, we remove a crude estimate of the SH term from the residuals. First, we further assume that the function g does not vary over time, and the residuals are i.i.d. replications over time. These two assumptions are made primarily to obtain a crude estimate of g by the moment estimator, denoted by $\hat{g}(\mathbf{s}_i')$, based on the replications over time. The residuals are then standardized by dividing by $\hat{g}(\mathbf{s}_i')$. We filter out the SH term through regressing the standardized residuals on SH functions up to $L = 3$. Last, we multiply the residuals after regression with $\hat{g}(\mathbf{s}_i')$, and treat them as the small-scale component of the electrostatic potentials. Note that a formal statistical analysis on the augmented model is left for future work.

5.2. Model fitting and prediction results.

We fit the AXING-need, Gau-need and Gau–Matérn models to the small-scale component of the electrostatic potential at the first time point (see the top left panel of Figure 1) since we are mainly interested in modeling the small-scale spatial variability. For future work, we may extend our model to spatio-temporal settings under the same framework with additional temporal structures on the coefficients c_{jk}'s. Without loss of generality, the Earth is treated as a unit sphere. The function g is assumed to be a function of co-latitude and the basis functions $b_i(\theta')$'s ($\theta' = 4\theta$ due to the stretching) are chosen as natural cubic B-splines with two interior knots and one boundary knot [see Section S.2.2 of the supplementary material Fan et al. (2018) for details]. To reduce the computational burden, we randomly select 4000 locations from the original grid to fit the models. We choose $\nu = 3$ among 2.5, 3 and 4 since it yields the best predictive performance (in terms of the MAE, MSPE and CRPS) when predicting at the remaining locations. Moreover, too large ν ($\nu > 4$) makes the model too close to Gaussian, and ν has to be larger than 2 to ensure the existence of the first two moments. For the AXING-need model, the parameters are estimated by the adaptive MCMC algorithm in Section 3.2, in which $\tau_\eta = 10$. The chain was run for 600,000 iterations, with the first 400,000 samples discarded as a burn-in period, and every 200th sample is collected as the posterior samples. The MCMC diagnostics can be found in Section S.2.3 of the supplementary material [Fan et al. (2018)]. For the Gaussian models, the parameters are estimated by the maximum likelihood method. Table 2 lists the parameter estimates and the corresponding standard errors for the models. The standard errors are estimated by the posterior sample standard deviations and the parametric bootstrap with 200 bootstrap samples for the AXING-need model and the Gaussian models, respectively. The parametric bootstrap method was used in spatial settings in Xu and Genton (2017), Fan et al. (2017).

Apart from model fitting, we further compare the predictive performance of the models. To reduce the computational burden, we only compare the Gau-need and Gau–Matérn models since the parameter estimation and spatial prediction for the AXING-need model involves more time-consuming MCMC iterations. The scientific implications and benefits of using a non-Gaussian model compared with Gaussian models will be illustrated in Section 5.3. Section S.2.6 of the supplementary material [Fan et al. (2018)] gives the training and test set separation in a cross-validation procedure, specified according to the spatial coverage of the

SuperDARN data. We predict on the test set using the parameter estimates in Table 2. The scoring rules MAE, MSPE and CPRS are used to assess the prediction accuracy, averaged over all the predicted locations. We repeat the cross-validation procedure 100 times, and calculate the mean and standard deviation (in parentheses) of the scores, which are displayed in Table 3. There is no significant difference in predictive performance between the two models, with the Gau-need model performing slightly better than the Gau–Matérn model in terms of the MSPE and CRPS. Table 3 also shows the mean lengths of 50% and 90% predictive intervals and the corresponding coverage probabilities, averaged over the 100 cross-validation replications. The Gau-need model yields longer predictive intervals, but they have more satisfactory coverage probabilities.

5.3.   Scientific implications.

As mentioned in Section 1, one important scientific goal of modeling the electrostatic potentials is to generate their simulations for estimating high-latitude energy inputs into the upper atmosphere in general circulation models. In Figure 5, we compare the simulations of the fitted AXING-need, Gau-need and Gau–Matérn models, for which the parameter estimates are given in Table 2. These simulations are generated such that their patterns are comparable across the models [see Section S.2.5 of the supplementary material Fan et al. (2018)]. We can see that the magnitude of the simulations of the Gau–Matérn model is much smaller than that of the AXING-need and Gau-need models. The most activity (i.e., the red and blue patches in the online version of this figure) of the simulations of all the models occurs near 75° latitude, which is consistent with the location of the general auroral zone [Hunsucker and Hargreaves (2007), Chapter 6.2.1]. Compared with the Gaussian models, the red and blue patches of the simulations of the AXING-need model have the sharpest and clearest boundaries. This can be attributed to its non-Gaussianity, which makes only a small number of the coefficients c_{jk}'s nonzero and the rest of them almost zero. Additional simulations of the fitted AXING-need model are shown in Section S.2.4 of the supplementary material [Fan et al. (2018)].

Finally, we quantify the amount of Joule heating produced by the small-scale component of the electrostatic potentials. The estimation of the Joule heating rates (i.e., the amount of Joule heating per second) based on the electrostatic potentials is given in Appendices B and C. For each fitted model, we compute the overall integrated Joule heating rate of 1000 simulated electrostatic potentials in the high-latitude region of the Northern Hemisphere, and plot the corresponding histogram in Figure 6. The overall integrated Joule heating rates for the AXING-need and Gau-need models are significantly larger than that for the large-scale component only, while the latter is almost the same as the overall integrated Joule heating rates for the Gau–Matérn model. Note that the large-scale component is what has been subtracted from the data in the preprocessing step. Moreover, the overall integrated Joule heating rates for the AXING-need model have a heavier right tail than those for the Gau-need model. Specifically, the 95th and 99th percentiles of the overall integrated Joule heating rates for the AXING-need model are 9.6842 and 11.6769, while those for the Gau-need model are 9.2981 and 9.9659. Consequently, the AXING-need model can produce significantly larger extreme values of the Joule heating rate than the Gau-need model. All of these suggest that the proposed AXING-need model can provide a potentially viable remedy

to systematic biases of general circulation models resulting from the underestimation of high-latitude energy inputs.

6. Discussion.

We have introduced a new class of multi-resolution spatial models for non-Gaussian random fields on a sphere. They are constructed in the form of a sparse random effects model using spherical needlets as building blocks. The spatial localization of needlets, together with carefully chosen random coefficients, ensure the model to be non-Gaussian and isotropic. We have shown that the proposed model has an oscillating and quasi-exponentially decaying covariance function. The model has also been expanded to include a spatially varying variance profile. The special formulation of the model enables us to develop efficient estimation and prediction procedures under a hierarchical Bayesian framework. We have investigated the accuracy of parameter estimation of the proposed model, and compared its predictive performance with that of two Gaussian models by extensive numerical experiments. The effectiveness of the proposed model is also demonstrated through an application of the methodology to the high-latitude ionospheric electrostatic potentials generated from the LFM-MIX model.

The computational speed of each iteration of the adaptive MCMC algorithm is primarily determined by the sampling step for the coefficient vector $\mathbf{c}$, even though we have already partitioned it into subblocks $\mathbf{c}_j$'s to speed up the computation. Recall that $\mathbf{c}_j$ is sampled from $\mathcal{N}(\widehat{\boldsymbol{\mu}}_j, \widehat{\boldsymbol{\Sigma}}_j)$. When the number of observations is much smaller than that of needlets, we can reduce the computational burden significantly through applying the Sherman–Morrison–Woodbury formula to $\widehat{\boldsymbol{\Sigma}}_j$. This coincides with the algorithm proposed in Bhattacharya, Chakraborty and Mallick (2016).

Many terrestrial processes arising in geophysical and environmental sciences are vector fields tangential to the surface of the Earth. Along the lines of Fan et al. (2017), extensions to models for tangential vector fields on a sphere are possible through applying spherical differential operators to the proposed model. Then the tangential vector fields are represented in terms of *vectorial needlets*. The proposed model can also be extended to models for anisotropic random fields, as mentioned in Section 2.2, by allowing σ_{jk} and ν_{jk} to vary spatially, and dynamic spatiotemporal random fields through imposing a time series model on the coefficients c_{jk}'s [Cressie, Shi and Kang (2010)].

Acknowledgments.

Appendix A:: PROOF OF THEOREM 1

Since c_{jk}'s are independent, we obtain

$$\mathrm{Cov}(X(\mathbf{s}), X(\mathbf{t})) = \sum_{j=J_0}^{J} \sum_{k=1}^{p_j} \mathrm{Var}\!\left(c_{jk}\right) \psi_{jk}(\mathbf{s}) \psi_{jk}(\mathbf{t})$$

$$= \sum_{j=J_0}^{J} \frac{v_j \sigma_j^2}{v_j - 2} \sum_{k=1}^{p_j} \psi_{jk}(\mathbf{s}) \psi_{jk}(\mathbf{t}).$$

Using (2.2), it follows that

$$\sum_{k=1}^{p_j} \psi_{jk}(\mathbf{s}) \psi_{jk}(\mathbf{t}) = \sum_{l=\lceil B^{j-1}\rceil}^{\lfloor B^{j+1}\rfloor} \sum_{l'=\lceil B^{j-1}\rceil}^{\lfloor B^{j+1}\rfloor} b\!\left(\frac{l}{B^j}\right) b\!\left(\frac{l'}{B^j}\right)\!\left(\frac{2l+1}{4\pi}\right)\!\left(\frac{2l'+1}{4\pi}\right) \times \sum_{k=1}^{p_j} \lambda_{jk} P_l\!\left(\left\langle \zeta_{jk}, \mathbf{s}\right\rangle\right) P_{l'}\!\left(\left\langle \zeta_{jk}, \mathbf{t}\right\rangle\right)$$

$$= \sum_{l=\lceil B^{j-1}\rceil}^{\lfloor B^{j+1}\rfloor} \sum_{l'=\lceil B^{j-1}\rceil}^{\lfloor B^{j+1}\rfloor} b\!\left(\frac{l}{B^j}\right) b\!\left(\frac{l'}{B^j}\right)\!\left(\frac{2l+1}{4\pi}\right)\!\left(\frac{2l'+1}{4\pi}\right) \times \int_{\mathbb{S}^2} P_l(\langle \mathbf{x}, \mathbf{s}\rangle) P_{l'}(\langle \mathbf{x}, \mathbf{t}\rangle)\, d\mathbf{x},$$

where the last step is due to the quadrature formula (2.1). Now, by the fact that P_l is real-valued and the *addition theorem for SH functions*, we have

$$\left(\frac{2l+1}{4\pi}\right)\!\left(\frac{2l'+1}{4\pi}\right) \int_{\mathbb{S}^2} P_l(\langle \mathbf{x}, \mathbf{s}\rangle) P_{l'}(\langle \mathbf{x}, \mathbf{t}\rangle)\, d\mathbf{x}$$

$$= \int_{\mathbb{S}^2} \sum_{m=-l}^{l} \sum_{m'=-l'}^{l'} \bar{Y}_{lm}(\mathbf{x}) Y_{lm}(\mathbf{s}) Y_{l'm'}(\mathbf{x}) \bar{Y}_{l'm'}(\mathbf{t})\, d\mathbf{x}$$

$$= \sum_{m=-l}^{l} \sum_{m'=-l}^{l'} Y_{lm}(\mathbf{s}) \bar{Y}_{l'm'}(\mathbf{t}) \int_{\mathbb{S}^2} Y_{l'm'}(\mathbf{x}) \bar{Y}_{lm}(\mathbf{x})\, d\mathbf{x}$$

$$= \sum_{m=-l}^{l} \sum_{m'=-l'}^{l'} Y_{lm}(\mathbf{s}) \bar{Y}_{l'm'}(\mathbf{t}) \delta_{ll'} \delta_{mm'}$$

$$= \delta_{ll'} \sum_{m=-l}^{l} Y_{lm}(\mathbf{s}) \bar{Y}_{lm}(\mathbf{t})$$

$$= \delta_{ll'} \left(\frac{2l+1}{4\pi}\right) P_l(\langle \mathbf{s}, \mathbf{t}\rangle),$$

where $\delta_{ab} = 1$ if $a = b$, and 0 otherwise. Note that the third equality follows from the orthonormality of SH functions, and the last one is due to another use of the *addition theorem*. Substituting this into the above equation, we obtain

$$\sum_{k=1}^{p_j} \psi_{jk}(\mathbf{s}) \psi_{jk}(\mathbf{t}) = \sum_{l=\lceil B^{j-1}\rceil}^{\lfloor B^{j+1}\rfloor} b^2\!\left(\frac{l}{B^j}\right)\!\left(\frac{2l+1}{4\pi}\right) P_l(\langle \mathbf{s}, \mathbf{t}\rangle).$$

Thus,

$$\text{Cov}(X(\mathbf{s}), X(\mathbf{t})) = \sum_{j=J_0}^{J} \frac{v_j \sigma_j^2}{v_j - 2} \sum_{l=\left\lceil B^{j-1} \right\rceil}^{\left\lfloor B^{j+1} \right\rfloor} b^2\left(\frac{l}{B^j}\right)\left(\frac{2l+1}{4\pi}\right) P_l\left(\langle \mathbf{s}, \mathbf{t} \rangle\right).$$

By the bound used in the proof of Baldi et al. (2009, Lemma 3),

$$\left| \sum_{l=\left\lceil B^{j-1} \right\rceil}^{\left\lfloor B^{j+1} \right\rfloor} b^2\left(\frac{l}{B^j}\right)\left(\frac{2l+1}{4\pi}\right) P_l\left(\langle \mathbf{s}, \mathbf{t} \rangle\right) \right| \leq \frac{c_M B^{2j}}{\left[1 + B^j \arccos(\langle \mathbf{s}, \mathbf{t} \rangle)\right]^M},$$

where where c_M is some constant depending on M. Note that a direct proof of the above inequality can be derived from Narcowich, Petrushev and Ward (2006, Theorem 3.5) or Dai and Xu (2013, Theorem 2.6.7).

Thus,

$$|\text{Cov}(X(\mathbf{s}), X(\mathbf{t}))| \leq \sum_{j=J_0}^{J} \frac{v_j \sigma_j^2}{v_j - 2} \frac{c_M B^{2j}}{\left[1 + B^j \arccos(\langle \mathbf{s}, \mathbf{t} \rangle)\right]^M}.$$

Appendix B:: DERIVING ELECTRIC FIELD FROM ELECTROSTATIC POTENTIAL

Denote the electric field by $\mathbf{E}$ and the electrostatic potential by Φ_E. Then

$$\mathbf{E} = -\nabla\Phi_E \approx -\frac{1}{R}\frac{\partial\Phi_E}{\partial\theta}\widehat{\boldsymbol{\theta}} - \frac{1}{R}\frac{1}{\sin\theta}\frac{\partial\Phi_E}{\partial\phi}\widehat{\boldsymbol{\phi}},$$

where ∇ is the gradient operator defined on $\mathbb{R}^3$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\phi}}$ are two of the spherical unit vectors with $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\phi}}$ pointing southward and eastward, respectively, and $R \approx 6.5 \times 10^6 m$ is the radius of the ionosphere. Note that the radial component of the electric field is ignored in this approximation since its magnitude is relatively small compared with that of the tangential component. Recall that the electrostatic potential in the high-latitude region of the Northern Hemisphere can be expressed as

$$\Phi_E(\theta, \phi) = g(\theta')\sum_{j,k} c_{jk}\psi_{jk}(\theta', \phi) = g(4\theta)\sum_{j,k} c_{jk}\psi_{jk}(4\theta, \phi),$$

where $\theta' = 4\theta \in [0, \pi]$ due to the stretching of the data to the entire sphere. Then we have

$$\frac{\partial \Phi_{\mathrm{E}}}{\partial \theta} = 4\left(\frac{\partial g}{\partial \theta'}\Big|_{\theta' \,=\, 4\theta} \sum_{j,\,k} c_{jk} \psi_{jk}(4\theta, \phi) + g(4\theta) \sum_{j,\,k} c_{jk} \frac{\partial \psi_{jk}}{\partial \theta'}\Big|_{\theta' \,=\, 4\theta}\right),$$

and

$$\frac{1}{\sin \theta} \frac{\partial \Phi_{\mathrm{E}}}{\partial \phi} = \frac{\sin \theta'}{\sin \theta} g(4\theta) \sum_{j,\,k} c_{jk} \frac{1}{\sin \theta'} \frac{\partial \psi_{jk}}{\partial \phi}.$$

Recall that

$$\psi_{jk}(\theta', \phi) = \sqrt{\lambda_{jk}} \sum_{l} b\left(\frac{l}{B^j}\right) \frac{2l+1}{4\pi} \times P_l\left(x_{jk} \sin \theta' \cos \phi \,+\, y_{jk} \sin \theta' \sin \phi \,+\, z_{jk} \cos \theta'\right).$$

Then

$$\frac{\partial \psi_{jk}}{\partial \theta'} = \sqrt{\lambda_{jk}}\left(x_{jk} \cos \theta' \cos \phi \,+\, y_{jk} \cos \theta' \sin \phi \,-\, z_{jk} \sin \theta'\right) \times \sum_{l} b\left(\frac{l}{B^j}\right) \frac{2l+1}{4\pi} \frac{dP_l(u)}{du}\Big|_{u \,=\, u'},$$

and

$$\frac{1}{\sin \theta'} \frac{\partial \psi_{jk}}{\partial \phi} = \sqrt{\lambda_{jk}}\left(-x_{jk} \sin \phi \,+\, y_{jk} \cos \phi\right) \sum_{l} b\left(\frac{l}{B^j}\right) \frac{2l+1}{4\pi} \frac{dP_l(u)}{du}\Big|_{u \,=\, u'},$$

where $u' = x_{jk} \sin \theta' \cos \phi + y_{jk} \sin \theta' \sin \phi + z_{jk} \cos \theta'$ Note that $dP_l(u)/du$ can be efficiently computed by a recursive formula.

APPENDIX C:: COMPUTING IONOSPHERIC JOULE HEATING RATE

According to Palmroth et al. (2005), the ionospheric Joule heating rate can be estimated by

$$\hat{P}_{\mathrm{JH}}(\theta, \phi) = \Sigma_{\mathrm{P}}(\theta, \phi)\left|\mathbf{E}(\theta, \phi)\right|^2 \approx \Sigma_{\mathrm{P}}(\theta, \phi)\left(E_\theta(\theta, \phi)^2 + E_\phi(\theta, \phi)^2\right),$$

Where Σ_{P} is the height-integrated Pedersen conductivity and E_θ and E_φ are the are the zonal and meridional components of the electric field given in Appendix B, that is,

$$E_\theta = -\frac{1}{R} \frac{\partial \Phi_{\mathrm{E}}}{\partial \theta},$$

and

$$E_\phi = -\frac{1}{R} \frac{1}{\sin \theta} \frac{\partial \Phi_{\mathrm{E}}}{\partial \phi}.$$

The Joule heating rate integrated over a subset S of the ionosphere is defined as

$$P_{\text{IJH}}(\mathcal{S}) = \int_{\mathcal{S}} P_{\text{JH}}(\theta, \phi)\, dS,$$

where dS is the area element of the ionosphere. In particular, when $\mathcal{S}$ is the high-latitude region of the Northern Hemisphere (co-latitude $\theta \leq \pi/4$),

$$P_{\text{IJH}} = \int_{\theta \leq \pi/4} P_{\text{JH}}(\theta, \phi)\, dS.$$

REFERENCES

Andrews DF and Mallows CL (1974). Scale mixtures of normal distributions. J. Roy. Statist. Soc. Ser. B 36 99–102. MR0359122

Andrieu C and Thoms J (2008). A tutorial on adaptive MCMC. Stat. Comput 18 343–373.

Atkinson K and Han W (2012). Spherical Harmonics and Approximations on the Unit Sphere: An Introduction Lecture Notes in Math. 2044 Springer, Heidelberg MR2934227

Baldi P, Kerkyacharian G, Marinucci D and Picard D (2009). Asymptotics for spherical needlets. Ann. Statist 37 1150–1171.

Barndorff-Nielsen O (1979). Models for non-Gaussian variation, with applications to turbulence. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci 368 501–520. MR0554672

Berg J, Natarajan A, Mann J and Patton EG (2016). Gaussian vs non-Gaussian turbulence: Impact on wind turbine loads. Wind Energy 19 1975–1989.

Bhattacharya A, Chakraborty A and Mallick BK (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. Biometrika 103 985–991. MR3620452 [PubMed: 28435166]

Chu J-H, Clyde MA and Liang F (2009). Bayesian function estimation using continuous wavelet dictionaries. Statist. Sinica 19 1419–1438.

Codrescu MV, Fuller-Rowell TJ and Foster JC (1995). On the importance of E-field variability for Joule heating in the high-latitude thermosphere. Geophys. Res. Lett 22 2393–2396.

Codrescu MV, Fuller-Rowell TJ, Foster JC, Holt JM and Cariglia SJ (2000). Electric field variability associated with the Millstone Hill electric field model. J. Geophys. Res 105 5265–5273.

Cousins EDP, Matsuo T and Richmond AD (2013a). SuperDARN assimilative mapping. J. Geophys. Res 118 7954–7962.

Cousins EDP, Matsuo T and Richmond AD (2013b). Mesoscale and large-scale variability in high-latitude ionospheric convection: Dominant modes and spatial/temporal coherence. J. Geophys. Res 118 7895–7904.

Cousins EDP and Shepherd SG (2012). Statistical characteristics of small-scale spatial and temporal electric field variability in the high-latitude ionosphere. J. Geophys. Res 117 A03317.

Cressie N (1993). Statistics for Spatial Data. Wiley, New York, NY.

Cressie N and Johannesson G (2008). Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B. Stat. Methodol 70 209–226.

Cressie N, Shi T and Kang EL (2010). Fixed rank filtering for spatio-temporal data. J. Comput. Graph. Statist 19 724–745. With supplementary material available online. MR2732500

Dai F and Xu Y (2013). Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer, New York, NY.

De Oliveira V, Kedem B and Short DA (1997). Bayesian prediction of transformed Gaussian random fields. J. Amer. Statist. Assoc 92 1422–1433.

Fan M (2015). A note on spherical needlets. Preprint Available at arXiv:1508.05406.

Fan M, Paul D, Lee CMT and Matsuo T (2017). Modeling tangential vector fields on a sphere. J. Amer. Statist. Assoc To appear.

Fan M, Paul D, Lee CMT and Matsuo T (2018). Supplement to "A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials." DOI: 10.1214/17-AOAS1104SUPP.

Gelman A, Roberts GO and Gilks WR (1996). Efficient Metropolis jumping rules. Bayesian Stat 5 599–608.

Genton MG and Kleiber W (2015). Cross-covariance functions for multivariate geostatistics. Statist. Sci 30 147–163. MR3353096

Gneiting T (2013). Strictly and non-strictly positive definite functions on spheres. Bernoulli 19 1327–1349.

Gneiting T, Balabdaoui F and Raftery AE (2007). Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B. Stat. Methodol 69 243–268.

Gneiting T and Raftery AE (2007). Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc 102 359–378.

Gneiting T and Ranjan R (2011). Comparing density forecasts using threshold-and quantileweighted scoring rules. J. Bus. Econom. Statist 29 411–422.

Górski KM, Hivon E, Banday AJ, Wandelt BD, Hansen FK, Reinecke M and Bartelmann M (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. Astrophys. J 622 759.

Guinness J and Fuentes M (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. J. Multivariate Anal 143 143–152.

Heaton MJ, Kleiber W, Sain SR and Wiltberger M (2015). Emulating and calibrating the multiple-fidelity Lyon–Fedder–Mobarry magnetosphere-ionosphere coupled computer model. J. R. Stat. Soc. Ser. C. Appl. Stat 64 93–113. MR3293920

Hunsucker RD and Hargreaves JK (2007). The High-Latitude Ionosphere and Its Effects on Radio Propagation. Cambridge Univ. Press, Cambridge.

Jones RH (1963). Stochastic processes on a sphere. Ann. Math. Stat 34 213–218. MR0170378

Jun M and Stein ML (2008). Nonstationary covariance models for global data. Ann. Appl. Stat 2 1271–1289.

Kleiber W, Sain SR, Heaton MJ, Wiltberger M, Reese CS, Bingham D (2013). Parameter tuning for a multi-fidelity dynamical model of the magnetosphere. Ann. Appl. Stat 7 1286–1310.

Kleiber W, Hendershott B, Sain SR and Wiltberger M (2016). Feature-based validation of the Lyon–Fedder–Mobarry magnetohydrodynamical model. J. Geophys. Res 121 11921200.

Lindgren F, Rue H and Lindström J (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B. Stat. Methodol 73 423–498.

Lyon JG, Fedder JA and Mobarry CM (2004). The Lyon–Fedder–Mobarry (LFM) global MHD magnetospheric simulation code. J. Atmos. Sol.-Terr. Phys 66 1333–1350.

Marinucci D and Peccati G (2011). Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications. Cambridge Univ. Press, Cambridge.

Matsuo T and Richmond AD (2008). Effects of high-latitude ionospheric electric field variability on global thermospheric Joule heating and mechanical energy transfer rate. J. Geophys. Res 113 A07309.

Matsuo T, Richmond AD and Hensel K (2003). High-latitude ionospheric electric field variability and electric potential derived from DE-2 plasma drift measurements: Dependence on IMF and dipole tilt. J. Geophys. Res 108 SIA 1–1–SIA 1–15.

Matsuo T, Richmond AD and Nychka DW (2002). Modes of high-latitude electric field variability derived from DE-2 measurements: Empirical Orthogonal Function (EOF) analysis. Geophys. Res. Lett 29 11–1–11–4.

Narcowich FJ, Petrushev P and Ward JD (2006). Localized tight frames on spheres. SIAM J. Math. Anal 38 574–594.

Palacios MB and Steel MFJ (2006). Non-Gaussian Bayesian geostatistical modeling. J. Amer. Statist. Assoc 101 604–618. MR2281244

Palmroth M, Janhunen P, Pulkkinen TI, Aksnes A, Lu G, Østgaard N, Watermann J, Reeves GD and Germany GA (2005). Assessment of ionospheric Joule heating by GUMICS-4 MHD simulation, AMIE, and satellite-based statistics: Towards a synthesis. Ann. Geophysicae 23 2051–2068.

Perron M and Sura P (2013). Climatology of non-Gaussian atmospheric statistics. J. Climate 26 1063–1083.

Røislien J and Omre H (2006). T-distributed random fields: A parametric model for heavytailed well-log data. Math. Geol 38 821–849.

Ruohoniemi JM and Baker KB (1998). Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations. J. Geophys. Res 103 20797–20811.

Schoenberg IJ (1942). Positive definite functions on spheres. Duke Math. J 9 96–108.

Stein ML (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. Ann. Statist 16 55–63.

Stein ML (2007). Spatial variation of total column ozone on a global scale. Ann. Appl. Stat 1 191–210. MR2393847

Terdik G (2015). Angular spectra for non-Gaussian isotropic fields. Braz. J. Probab. Stat 29 833–865.

Wallin J and Bolin D (2015). Geostatistical modelling using non-Gaussian Matérn fields. Scand. J. Stat 42 872–890.

Weimer DR (1995). Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients. J. Geophys. Res 100 19595–19607.

West M (1987). On scale mixtures of normal distributions. Biometrika 74 646–648. MR0909372

Wiltberger M, Rigler EJ, Merkin V and Lyon JG (2017). Structure of high latitude currents in magnetosphere-ionosphere models. Space Sci. Rev 206 575–598.

Wolfe PJ, Godsill SJ and Ng W-J (2004). Bayesian variable selection and regularization for time-frequency surface estimation. J. R. Stat. Soc. Ser. B. Stat. Methodol 66 575–589. MR2088291

Womersley RS (2015). Efficient spherical designs with good geometric properties. Available at http://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/.

Xu G and Genton MG (2017). Tukey g-and-h random fields. J. Amer. Statist. Assoc 112 1236–1249.

Zhang B, Lotko W, Wiltberger MJ, Brambles OJ and Damiano PA (2011). A statistical study of magnetosphere–ionosphere coupling in the Lyon–Fedder–Mobarry global MHD model. J. Atmos. Sol.-Terr. Phys 73 686–702.
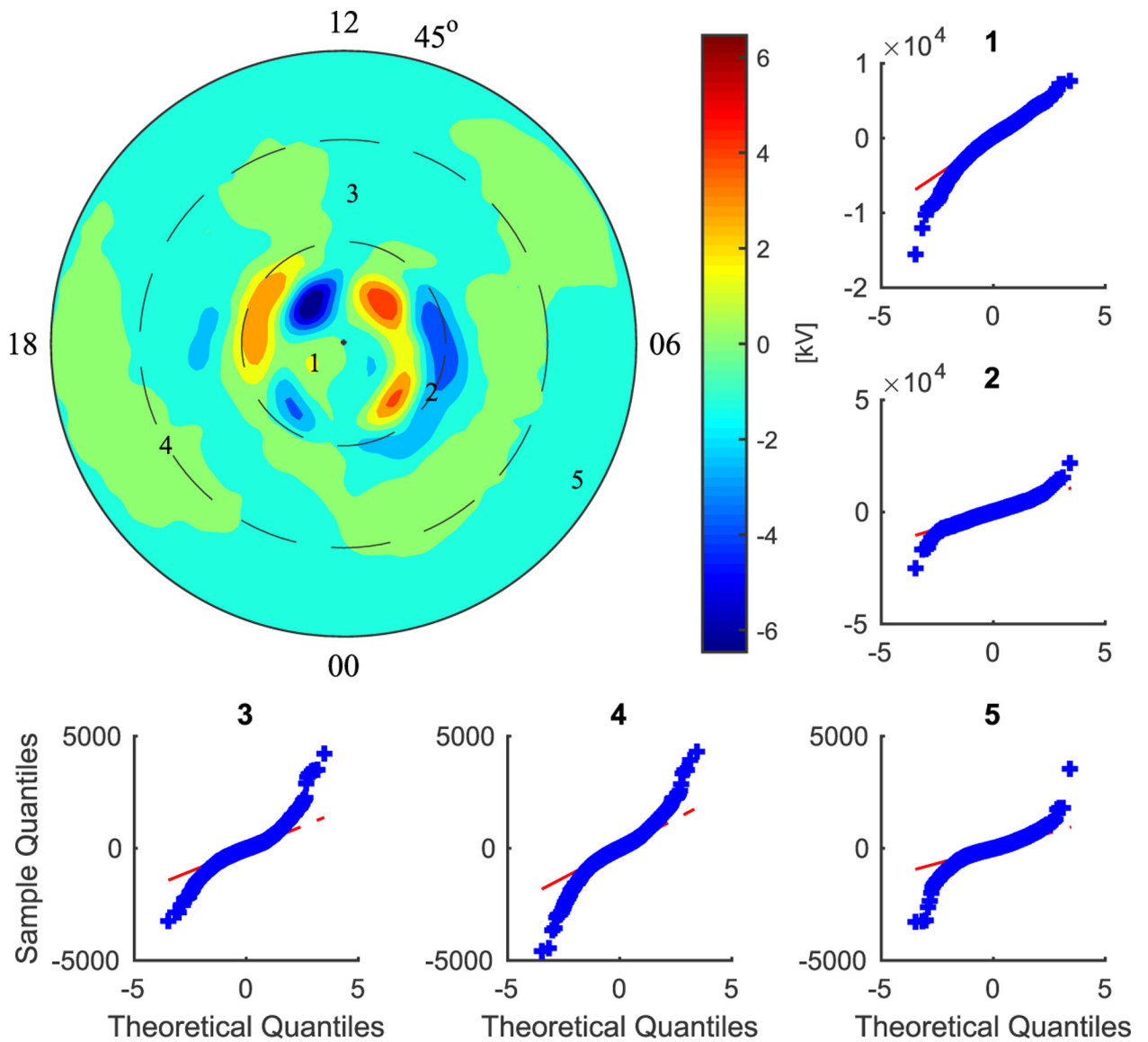
Fig. 1.
Small-scale component of the high-latitude ionospheric electrostatic potentials in the Northern Hemisphere generated from the LFM-MIX model at the Quad resolution. Top left panel: the small-scale component at the first time point; Other panels: *the non-Gaussianity of the small-scale component is illustrated by the Q–Q plots of the replications over time at different locations.*
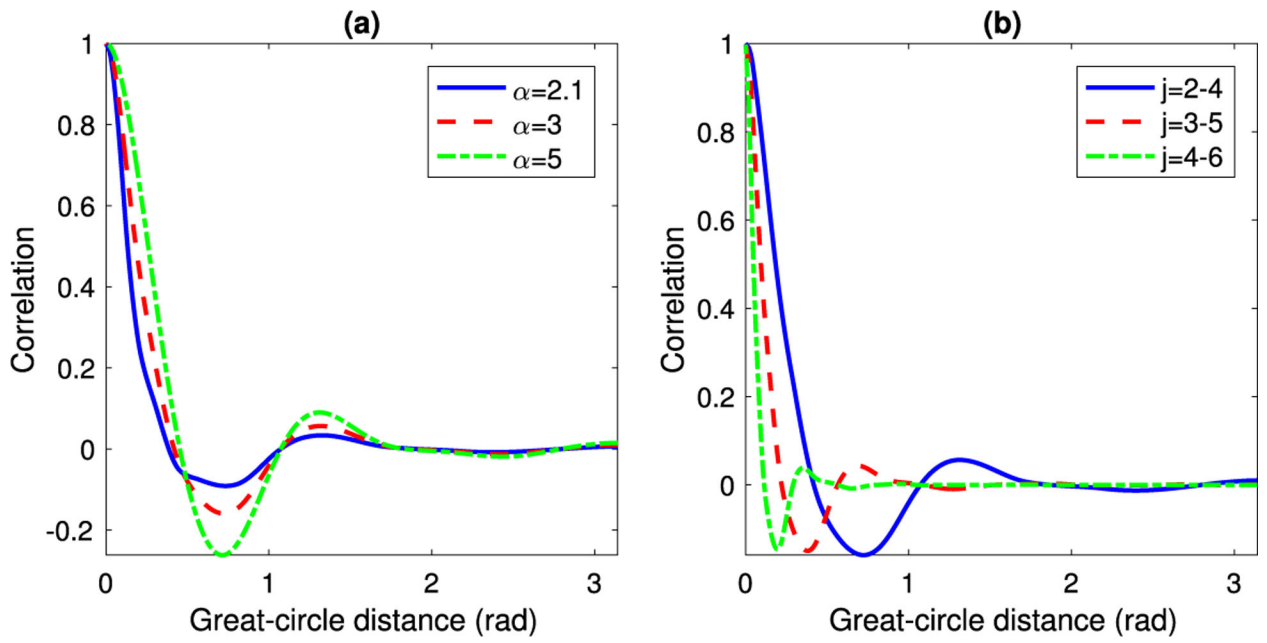
Fig. 2.
(a) *Correlation functions for various α with j ranging from* 2 *to* 4; (b) *correlation functions for various ranges of j with α = 3.*
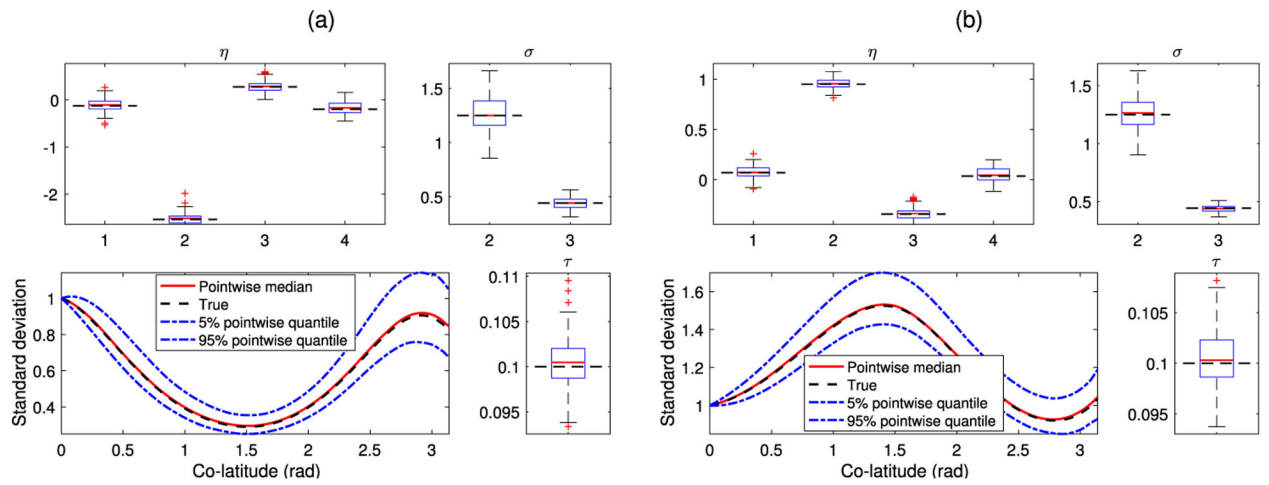
Fig. 3.

Results of the Monte Carlo simulation study under two different settings of η_{-0}. in panels (a) and (b) *The dashed horizontal line accompanying the boxplots shows the true value of the parameters.* Top left block: *boxplots of estimated η_{-0};* Bottom left block: *estimated and true curves of g* (*i.e., the standard deviation profile up to a constant*); Top right block: *boxplots of estimated σ_2 and σ3;* Bottom right block: *boxplot of estimated τ.*
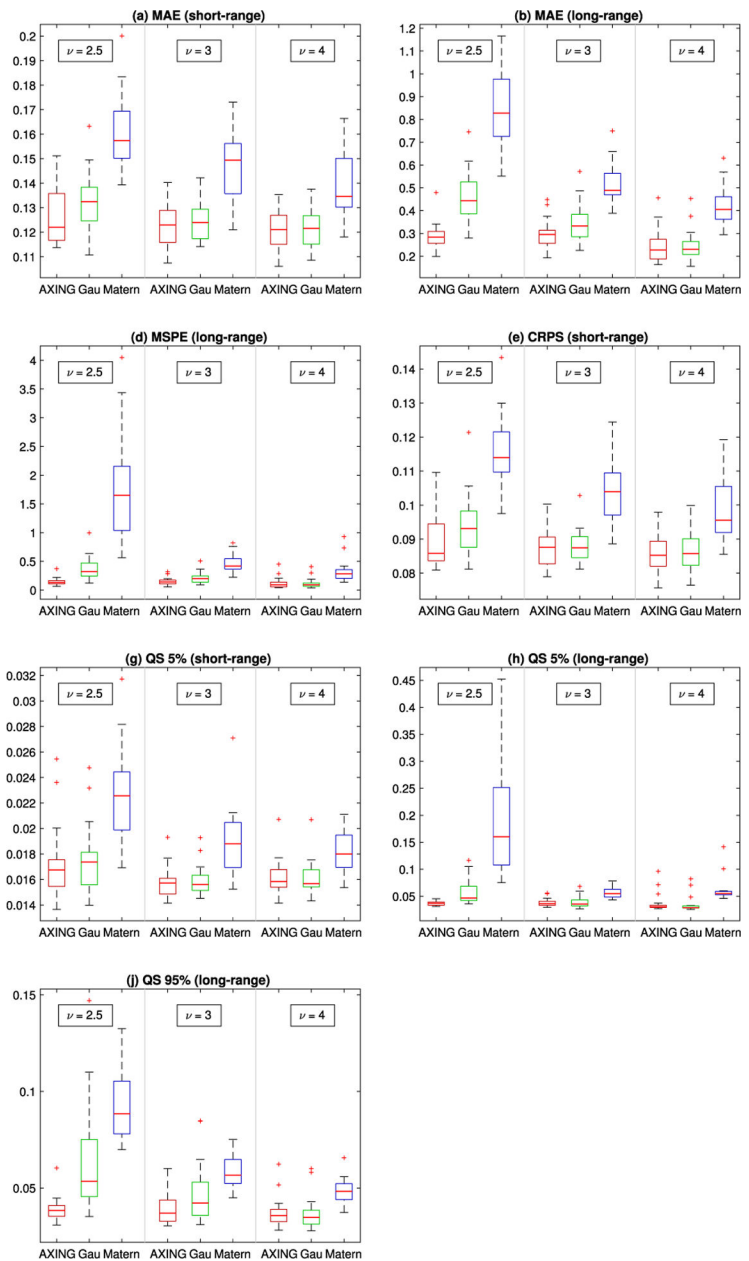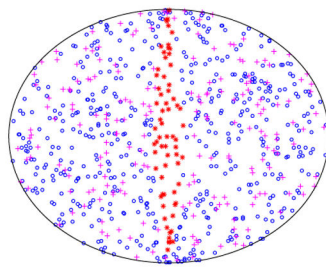
Fig. 4.

First panel: *Training and test set separation for one of the cross-validation replications. The sphere has been projected to an ellipse by the Hammer projection. "*o*" observed locations; "*+*" short-range predicted locations; "***" long-range predicted locations;* Other panels: *Boxplots of five scoring rules for the AXING-need, Gau-need and Gau–Matérn models with 20 cross-validation replications.* (a)–(b) *MAE*; (c)–(d) *MSPE*; (e)–(f) *CRPS*; (g)–(h) *QS 5%*; (i)–(j) *QS 95%.*
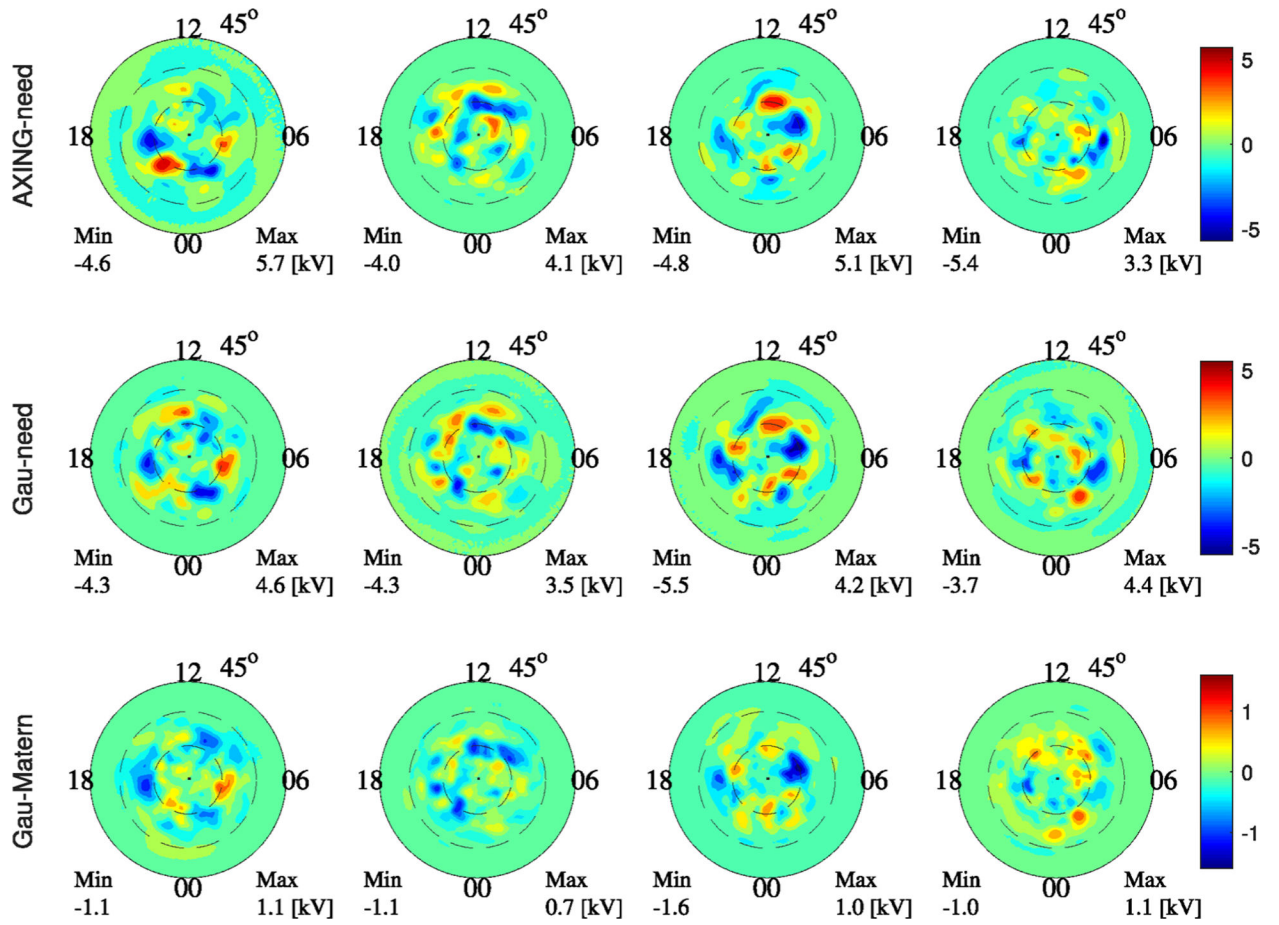
Fig. 5.
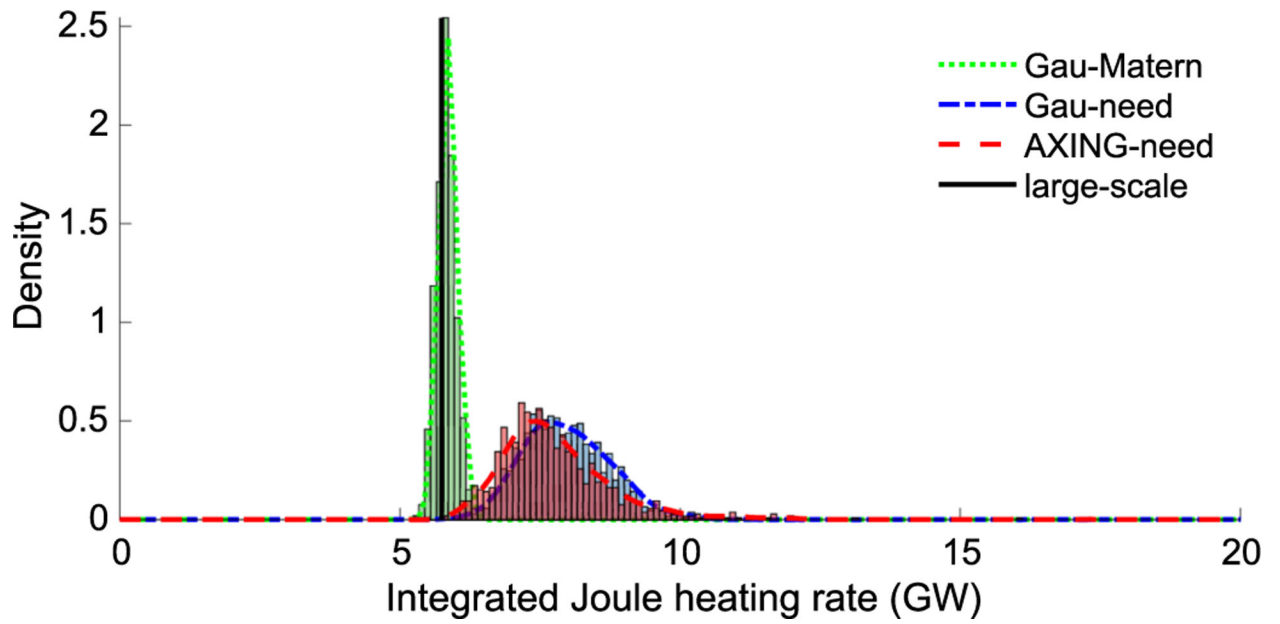*Comparison of simulations of the AXING-need, Gau-need and Gau–Matérn models.*

Fig. 6.
Histograms of the overall integrated Joule heating rates of 1000 electrostatic potentials in the high-latitude region of the Northern Hemisphere simulated from the AXING-need, Gan-need and Gau–Matérn models.

Table 1

Coverage probabilities (CP) and mean lengths (mLen) of 50% and 90% predictive intervals, averaged over 20 cross-validation replications

		Short-range prediction				Long-range prediction			
v	Model	CP (50%)	mLen	CP (90%)	mLen	CP (50%)	mLen	CP (90%)	mLen
2.5	AXING-need	49.4%	0.21	88.8%	0.50	53.2%	0.55	91.8%	1.34
	Gau-need	48.3%	0.21	88.2%	0.51	34.9%	0.57	79.7%	1.39
	Gau-Matérn	52.6%	0.28	90.6%	0.69	43.4%	1.06	81.7%	2.59
3	AXING-need	49.0%	0.20	90.0%	0.50	50.5%	0.52	88.6%	1.26
	Gau-need	48.2%	0.20	89.2%	0.49	44.0%	0.50	84.7%	1.22
	Gau-Matérn	51.3%	0.25	90.6%	0.61	44.4%	0.87	93.2%	2.12
4	AXING-need	50.7%	0.20	89.6%	0.50	63.5%	0.50	93.2%	1.23
	Gau-need	50.4%	0.20	89.4%	0.49	59.5%	0.48	92.1%	1.16
	Gau-Matérn	52.1%	0.24	90.6%	0.60	54.3%	0.78	93.5%	1.90

Table 2

Parameter estimates for the AXING-need ($\nu = 3$), Gau-need and Gau–Matérn models applied to the small-scale component of the high-latitude ionospheric electrostatic potential at the first time point. The corresponding standard errors are shown in parentheses. Note that η_0 is fixed at 0 for the AXING-need and Gau-need models

Model	AXING-need	Gau-need	Gau-Matérn
η_0	0 (–)	0 (–)	–2.037 (0.16)
η_1	0.949 (0.036)	0.959 (0.011)	0.659 (0.10)
η_2	1.579 (0.086)	1.619 (0.012)	1.372 (0.29)
η_3	–1.209 (0.046)	–1.190 (0.012)	–1.078 (0.070)
σ_2	0.148 (0.023)	0.252 (0.014)	-
σ_3	0.0363 (0.0035)	0.0665 (0.0032)	-
σ_4	4.09e-3 (2.62e-4)	6.47e-3 (2.11e-04)	-
κ	-	-	2.857 (0.14)
$1/a$	-	-	0.102 (0.0072)
τ	0.0280 (3.71e-04)	0.0282 (3.18e-04)	7.76e-03 (1.16e-04)

Table 3

Predictive performance of the Gau-need and Gau–Matérn models for the small-scale component of the high-latitude ionospheric electrostatic potential at the first time point. All the scores are averaged over 100 cross-validation replications, where the standard deviations are shown in parentheses

Model	MAE (kV)	MSPE (kV^2)	CRPS	CP (50%)	mLen	CP (90%)	mLen
Gau-need	0.138 (0.024)	0.146 (0.074)	0.102 (0.019)	68.2%	0.23	92.5%	0.57
Gau-Matérn	0.136 (0.031)	0.158 (0.092)	0.113 (0.029)	43.6%	0.08	68.8%	0.21