6-2012

Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends

Oleksandr Gromenko
Utah State University

Piotr Kokoszka
Utah State University

Lie Zhu
Utah State University

Jan J. Sojka
Utah State University

Recommended Citation
Gromenko, O., Kokoszka, P., Zhu, L., & Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. Annals of Applied Statistics, 6(2), 669-696.

The Annals of Applied Statistics
2012, Vol. 6, No. 2, 669–696
DOI: 10.1214/11-AOAS524

ESTIMATION AND TESTING FOR SPATIALLY INDEXED CURVES WITH APPLICATION TO IONOSPHERIC AND MAGNETIC FIELD TRENDS[1]

By Oleksandr Gromenko, Piotr Kokoszka, Lie Zhu and Jan Sojka

Utah State University

We develop methodology for the estimation of the functional mean and the functional principal components when the functions form a spatial process. The data consist of curves $X(\mathbf{s}_k;t), t\in[0,T]$, observed at spatial locations $\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_N$. We propose several methods, and evaluate them by means of a simulation study. Next, we develop a significance test for the correlation of two such functional spatial fields. After validating the finite sample performance of this test by means of a simulation study, we apply it to determine if there is correlation between long-term trends in the so-called critical ionospheric frequency and decadal changes in the direction of the internal magnetic field of the Earth. The test provides conclusive evidence for correlation, thus solving a long-standing space physics conjecture. This conclusion is not apparent if the spatial dependence of the curves is neglected.

1. Introduction. The contribution of this paper to statistics is two-fold: (1) we develop estimation methodology for the functional mean and the functional principal components (FPCs) when the functions form a spatial field; (2) we propose a significance test to determine if two families of curves observed at the same spatial locations are uncorrelated. The contribution to space physics consists in solving a controversy regarding the impact of long-term changes in the internal magnetic field of the Earth on long-term ionospheric trends. The required physics background is provided later in this section, and in Section 8.

The data is modeled as curves $X(\mathbf{s}_k;t), t\in[0,T]$, observed at spatial locations $\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_N$. Such functional data structures are quite common, but typically the spatial dependence and the spatial distribution of the points $\mathbf{s}_k$ are not taken into account. A fundamental question is how to estimate the

Received January 2011; revised August 2011.
[1]Supported in part by NSF Grants DMS-08-04165 and DMS-09-31948.
Key words and phrases. Ionospheric trends, functional data analysis, spatial statistics.

mean function of curves indexed by spatial locations. Clearly, curves located at close-by points look similar and must be given smaller weights than curves at points far apart. In addition to the mean function, FPCs play a fundamental role in functional data analysis. Good estimators of FPCs are needed to construct reliable testing and classification procedures, but such issues have been addressed only in the contexts of independent curves, with focus on sparsity and measurement error. The geophysical data that motivate this research are available at fine temporal grids and are measured with errors that are negligible relative to the objectives of the statistical analysis. A focus of recent geophysical research is on the detection and estimation of global and/or regional long-term trends (the global warming paradigm), so before a statistical analysis is undertaken, the data are typically smoothed to remove daily or even annual periodicity. The question we address is how to combine the temporal trajectories available at many spatial locations to obtain meaningful summary trends. We argue that one can do better than using simple averaging. The focus of this paper is thus on combining information from spatially dependent curves, which are smooth and available at all time points.

Many environmental and geophysical data sets fall into the framework considered in this paper. The data set that motivated this research consists of the curves of the ionospheric F2-layer critical frequency, foF2. Three such curves are shown in Figure 1. In principle, foF2 curves are available at over

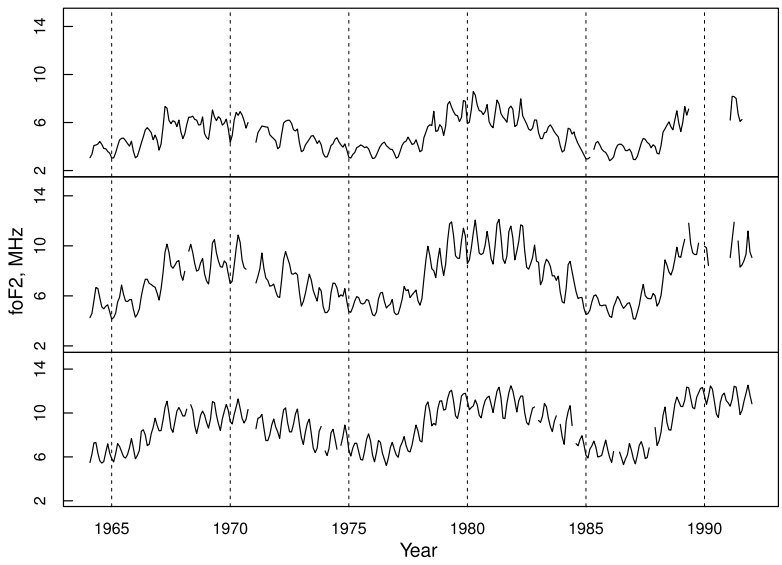

FIG. 1. *F2-layer critical frequency curves at three locations. Top to bottom (latitude in parentheses): Yakutsk (62.0), Yamagawa (31.2), Manila (14.7). The functions exhibit a latitudinal trend in amplitude.*

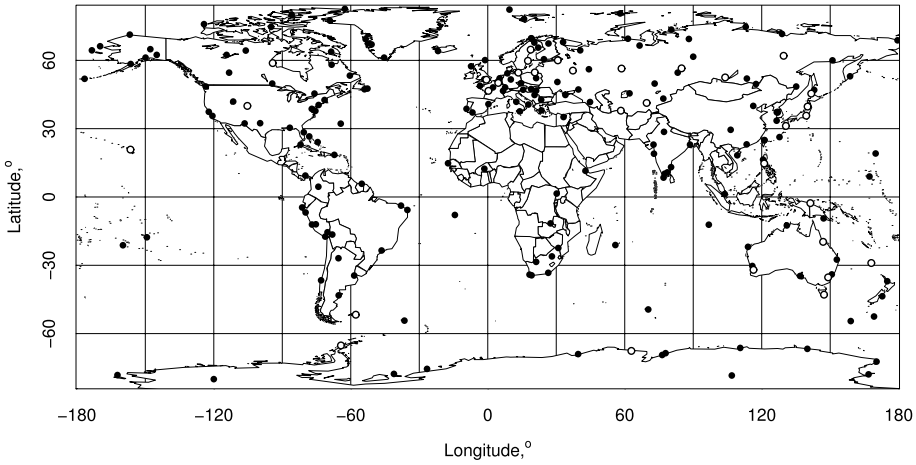

FIG. 2. *Locations of 218 ionosonde stations. Circles represent the 32 stations with the longest complete records.*

200 locations throughout the globe (see Figure 2), but sufficiently complete data are available at only 30–40 locations which are very unevenly spread; for example, there is a dense network of observatories over Europe and practically no data over the oceans. The study of this data set has been motivated by the hypothesis of Roble and Dickinson (1989), who suggested that the increasing amounts of (radiative) greenhouse gases should lead to global cooling in mesosphere and thermosphere, as opposed to the global warming in lower troposphere; cf. Figure 3. Rishbeth (1990) pointed out that such cooling would result in a thermal contraction and the global lowering of the ionospheric peak densities, which can be computed from the critical frequency foF2. The last twenty years have seen very extensive research in this area; see Laštovička et al. (2008) for a partial overview. One of the difficulties in determining a global trend is that the foF2 curves appear to exhibit trends in opposing directions over various regions. A possible explanation suggests that these trends are caused by long-term trends in the magnetic field of the Earth. There is, however, currently not agreement in the space physics community if this is indeed the case. In general, to make any trends believable, a suitable statistical modeling and a proper treatment of "errors and uncertainties" is called for [Ulich, Clilverd and Rishbeth (2003)]. This paper makes a contribution in this direction. Space physics data measured at terrestrial observatories always come in the form of temporal curves at fixed spatial locations. In Maslova et al. (2009), Maslova et al. (2010a) and Maslova et al. (2010b) the tools of functional data analysis were used to study such data, but the spatial dependence of the curves was not fully exploited.

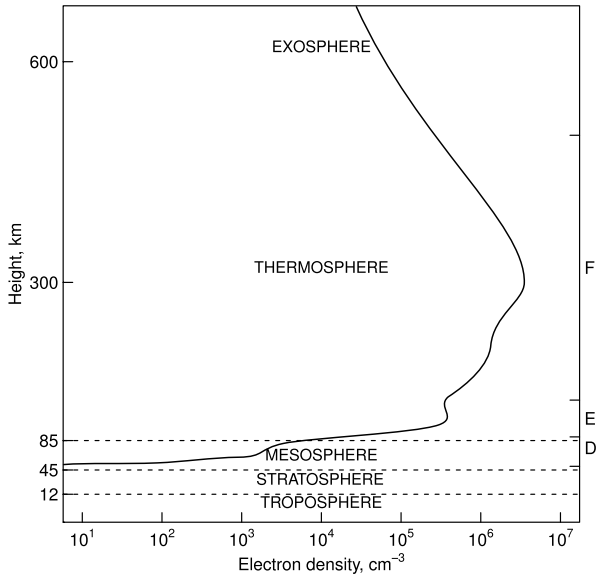

FIG. 3. *Typical profile of day time ionosphere. The curve shows electron density as a function of height. The right vertical axis indicates the D, E and F regions.*

Spatio-temporal modeling has received a great deal of attention of late; see Part V of Gelfand et al. (2010) and Chapters 3, 4 and 6 of Finkenstaedt, Held and Isham (2007) which discuss spatio-temporal models for geostatistical data. There has, however, not been much research specifically on spatially indexed functional data; Delicado et al. (2010) review recent contributions. For geostatistical functional data, several approaches to kriging have been proposed; see Yamanishi and Tanaka (2003), Nerini, Monestiez and Manté (2010), Giraldo, Delicado and Mateu (2011a) and Bel et al. (2011).

Throughout the paper, $\{X(\mathbf{s})\}$ denotes a random field defined on a spatial domain and taking values in the Hilbert space $L^2 = L^2([0,1])$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt, \qquad f, g \in L^2.$$

The value of the function $X(\mathbf{s}) \in L^2$ at time $t \in [0,1]$ is denoted by $X(\mathbf{s};t)$. We postulate the model

$$X(\mathbf{s};t) = \mu(t) + \sum_{i=1}^{\infty} \xi_i(\mathbf{s})e_i(t), \qquad \xi_i(\mathbf{s}) = \langle X(\mathbf{s}) - \mu, e_i\rangle, \tag{1.1}$$

where the e_i form a complete orthonormal system. Note that the mean function μ and the FPCs e_i do not depend on $\mathbf{s}$. A sufficient condition for this is that the distribution in L^2 of the function $X(\mathbf{s})$ does not depend on

the location $\mathbf{s}$. A stronger sufficient condition is the strict stationarity of the field $\{X(\mathbf{s})\}$.

For the applications we have in mind, it is enough to assume that the spatial domain is a subset of the plane or a two-dimensional sphere. On the plane, the distance between points is the usual Euclidean distance; on the sphere, we use the chordal distance defined as the Euclidean distance in the three-dimensional space. The reason for using the chordal distance is that any spatial covariance functions in $\mathbb{R}^3$ restricted to the unit sphere is then also a covariance function on the sphere. Denoting the latitude by L and the longitude by l, the chordal distance, $0 \le d_{k,\ell} \le 2$, between two points, $\mathbf{s}_k, \mathbf{s}_\ell$, on the unit sphere is given by

$$d_{k,\ell} = 2\left[\sin^2\left(\frac{L_k - L_\ell}{2}\right) + \cos L_k \cos L_\ell \sin^2\left(\frac{l_k - l_\ell}{2}\right)\right]^{1/2}. \tag{1.2}$$

For arbitrary (not necessarily spatially indexed) functions, $X_1, X_2, \ldots, X_N$, the sample mean is defined as $\bar{X}_N = N^{-1}\sum_{n=1}^N X_n$, and the sample covariance operator as

$$\widehat{C}(x) = N^{-1}\sum_{n=1}^{N}[\langle (X_n - \bar{X}_N), x\rangle (X_n - \bar{X}_N)], \qquad x \in L^2.$$

The sample FPCs are computed as the eigenfunctions of $\widehat{C}$. These are the estimates produced by several software packages, including the popular `R` package `fda`; see Ramsay, Hooker and Graves (2009). The consistency of the sample mean and the sample FPCs relies on the assumption that the functional observations form a simple random sample. If the functions $X_k = X(\mathbf{s}_k)$ are spatially distributed, the sample mean and the sample FPCs need not even be consistent; see Hörman and Kokoszka (2012). This happens if the spatial dependence is strong or if there are clusters of the points $\mathbf{s}_k$. We will demonstrate that better estimators are available and we will use them as part of the procedure for testing the independence of two functional fields $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$ and $\{Y(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$. The procedure is based on the observed pairs of functions $(X(\mathbf{s}_k), Y(\mathbf{s}_k)), 1 \le k \le N$. The test we propose is applied to ionosonde (X) and magnetic (Y) curves, and conclusively shows that the temporal evolution of these two families is strongly correlated.

The remainder of the paper is organized as follows. Sections 2 and 3 focus, respectively, on the estimation of the mean function and the FPCs in a spatial setting. Section 4 demonstrates by means of a simulation study that the methods we propose improve on the standard approach, and discusses their relative performance and computational cost. In Section 5 we develop a test for the correlation of two functional spatial fields. This test requires estimation of a covariance tensor. After addressing this issue in Section 6,

we study in Section 7 the finite sample properties of several implementations of the test. Finally, in Section 8 we apply the methodology developed in the previous section to test for the correlation between the ionospheric critical frequency and magnetic curves.

2. Estimation of the mean function. We propose three methods of estimating the mean function μ, which we call M1, M2, M3. As will become apparent in this section, several further variants, not discussed here, are conceivable. But the results of Section 4 show that while all these methods offer an improvement over the simple sample mean, their performance is comparable. We represent the observed functions as

$$X(\mathbf{s}_k;t) = \mu(t) + \varepsilon(\mathbf{s}_k;t), \qquad k = 1, 2, \ldots, N, \tag{2.1}$$

where ε is an unobservable field with $E\varepsilon(\mathbf{s};t) = 0$. All methods assume that the function valued field $\varepsilon(\cdot)$ is strictly stationary and isotropic, even though weaker, more technical assumptions could be made for the specific methods. Methods M1 and M2 are akin to the kriging technique advocated by Giraldo, Delicado and Mateu (2011a) in that they treat the curves as single entities and seek to minimize the integrated mean squared error. Method M3 is similar in spirit to the approach to functional kriging developed by Nerini, Monestiez and Manté (2010) and Giraldo, Delicado and Mateu (2011b) who use cokriging of basis coefficients.

Methods M1 and M2 estimate μ by the weighted average

$$\hat{\mu}_N = \sum_{n=1}^{N} w_n X(\mathbf{s}_n). \tag{2.2}$$

The optimal weights w_k are defined to minimize $E\|\sum_{n=1}^{N} w_n X(\mathbf{s}_n) - \mu\|^2$ subject to the condition $\sum_{n=1}^{N} w_n = 1$ ($\|x\|^2 = \int_0^1 x^2(t)\,dt$). Using the method of the Lagrange multiplier, we see that the unknowns $w_1, w_2, \ldots, w_N, r$ are solutions to the system of $N+1$ equations

$$\sum_{n=1}^{N} w_n = 1, \qquad \sum_{k=1}^{N} w_k C_{kn} - r = 0, \qquad n = 1, 2, \ldots, N, \tag{2.3}$$

where

$$C_{k\ell} = E[\langle \varepsilon(\mathbf{s}_k), \varepsilon(\mathbf{s}_\ell)\rangle]. \tag{2.4}$$

Set $\mathbf{w} = (w_1, \ldots, w_N)^T$. An easy way to solve the equations in (2.3) is to compute $\mathbf{v} = \mathbf{C}^{-1}\mathbf{1}$, where $\mathbf{C} = [C_{k\ell}, 1 \le k, \ell \le N]$, and then set $\mathbf{w} = a\mathbf{v}$, where a is a constant such that $\mathbf{1}^T\mathbf{w} = 1$.

Method M1. At each time point t_j, we fit a parametric spatial model to the scalar field $X(\mathbf{s};t_j)$. To focus attention, we provide formulas for the

exponential model

$$\operatorname{Cov}(X(\mathbf{s}_k;t_j),X(\mathbf{s}_\ell;t_j))=\sigma^2(t_j)\exp\left\{-\frac{d(\mathbf{s}_k,\mathbf{s}_\ell)}{\rho(t_j)}\right\}. \tag{2.5}$$

It is clear how they can be modified for other popular models.

Observe that under model (2.5),

$$\begin{aligned}C_{k\ell}&=E\int(X(\mathbf{s}_k;t)-\mu(t))(X(\mathbf{s}_\ell;t)-\mu(t))\,dt\\&=\int\operatorname{Cov}(X(\mathbf{s}_k;t_j),X(\mathbf{s}_\ell;t_j))\,dt\\&=\int\sigma^2(t)\exp\left\{-\frac{d(\mathbf{s}_k,\mathbf{s}_\ell)}{\rho(t)}\right\}dt.\end{aligned}$$

One way to estimate $C_{k\ell}$ is to set

$$\widehat{C}_{k\ell}=\int\hat{\sigma}^2(t)\exp\left\{-\frac{d(\mathbf{s}_k,\mathbf{s}_\ell)}{\hat{\rho}(t)}\right\}dt, \tag{2.6}$$

with the estimates $\hat{\sigma}^2(t_j)$ and $\hat{\rho}(t_j)$ obtained using some version of empirical variogram, (4.7) or (4.8) in this study.

Another way to proceed is to replace the $\hat{\rho}(t_j)$ by their average $\hat{\rho}=m^{-1}\times\sum_{j=1}^m\hat{\rho}(t_j)$, where m is the count of the t_j at which the variogram is estimated successfully. Then, the $C_{k\ell}$ are approximated by

$$\widehat{C}_{k\ell}=\left(\int\hat{\sigma}^2(t)\,dt\right)\exp\left\{-\frac{d(\mathbf{s}_k,\mathbf{s}_\ell)}{\hat{\rho}}\right\}.$$

As explained above, in order to compute the weights w_j in (2.3), it is enough to know the matrix $\mathbf{C}$ only up to a multiplicative constant. Thus, we may set

$$\widehat{C}_{k\ell}=\exp\left\{-\frac{d(\mathbf{s}_k,\mathbf{s}_\ell)}{\hat{\rho}}\right\}. \tag{2.7}$$

Once the matrix $\mathbf{C}$ has been estimated, we compute the weights w_j, and estimate the mean via (2.2).

If (2.6) is used, we refer to this method as M1a; if (2.7) is used, we call it M1b.

Method M1 relies on the estimation of the variograms at every point t_j. Method M2, described below, requires only one optimization, so it is much faster than M1.

Method M2. We define the *functional* variogram

$$\begin{aligned}2\gamma(\mathbf{s}_k,\mathbf{s}_\ell)&=E\|X(\mathbf{s}_k)-X(\mathbf{s}_\ell)\|^2\\&=2E\|X(\mathbf{s}_k)-\mu\|^2-2E[\langle X(\mathbf{s}_k)-\mu,X(\mathbf{s}_\ell)-\mu\rangle]\\&=2E\|X(\mathbf{s})-\mu\|^2-2C_{k\ell}.\end{aligned} \tag{2.8}$$

The variogram (2.8) can be estimated by its empirical counterparts, like (4.7) or (4.8), with the $|X(\mathbf{s}_k) - X(\mathbf{s}_\ell)|$ replaced by

$$\|X(\mathbf{s}_k) - X(\mathbf{s}_\ell)\| = \left\{ \int (X(\mathbf{s}_k; t) - X(\mathbf{s}_\ell; t))^2 \, dt \right\}^{1/2}.$$

Next, we fit a parametric model, for example, we postulate that

$$\gamma(\mathbf{s}_k, \mathbf{s}_\ell) = \sigma_f^2 \left(1 - \exp\left\{ -\frac{d(\mathbf{s}_k, \mathbf{s}_\ell)}{\rho_f} \right\} \right). \tag{2.9}$$

The subscript f is used to emphasize the *functional* variogram. Denoting by $\hat{\rho}_f$ the resulting NLS estimate, we estimate the C_{kl} by (2.7) with $\hat{\rho}$ replaced by $\hat{\rho}_f$.

Method M3. This method uses a basis expansion of the functional data, it does not use the weighted sum (2.2). Suppose $B_j, 1 \le j \le K$, are elements of a functional basis with K so large that for each k

$$X(\mathbf{s}_k) \approx \sum_{j \le K} \langle B_j, X(\mathbf{s}_k) \rangle B_j \tag{2.10}$$

to a good approximation. By (2.1), we obtain for every j

$$\langle B_j, X(\mathbf{s}_k) \rangle = \langle B_j, \mu \rangle + \langle B_j, \varepsilon(\mathbf{s}_k) \rangle, \qquad k = 1, 2, \ldots, N. \tag{2.11}$$

For every fixed j, the $\langle B_j, X(\mathbf{s}_k) \rangle$ form a stationary and isotropic scalar spatial field with a constant unknown mean $\langle B_j, \mu \rangle$. This mean can be estimated by postulating a covariance structure for each $\langle B_j, X(\mathbf{s}_k) \rangle$, for example,

$$\operatorname{Cov}(\langle B_j, X(\mathbf{s}_k) \rangle, \langle B_j, X(\mathbf{s}_\ell) \rangle) = \sigma_j^2 \exp\left\{ -\frac{d(\mathbf{s}_k, \mathbf{s}_\ell)}{\rho_j} \right\}.$$

The mean $\langle B_j, \mu \rangle$ is estimated by a weighted average of the $\langle B_j, X(\mathbf{s}_k) \rangle$ (the weights depend on j). Denote the resulting estimate by $\hat{\mu}_j$. The mean function μ is then estimated by $\hat{\mu}(t) = \sum_{j \le K} \hat{\mu}_j B_j(t)$.

3. Estimation of the principal components. Assume now that the mean function μ has been estimated, and this estimate is subtracted from the data. To simplify the formulas, in the following we thus assume that $EX(\mathbf{s}) = 0$.

We consider analogs of methods M2 and M3. Extending Method M1 is possible, but presents a computational challenge because a parametric spatial model would need to be estimated for every pair (t_i, t_j). For the ionosonde data studied in Section 8, there are 336 points t_j. Estimation on a single data set would be feasible, but not a simulation study based on thousands of replications.

In both approaches, which we term CM2 and CM3, the FPCs are estimated by expansions of the form

$$v_j(t) = \sum_{\alpha=1}^{K} x_\alpha^{(j)} B_\alpha(t), \tag{3.1}$$

where the B_α are elements of an *orthonormal* basis. We first describe an analog of method M3, which is conceptually and computationally simpler.

Method CM3. The starting point is the expansion

$$X(\mathbf{s};t)=\sum_{j=1}^{\infty}\xi_j(\mathbf{s})B_j(t),$$

where, by the orthonormality of the B_j, the $\xi_j(\mathbf{s})$ form an observable field $\xi_j(\mathbf{s}_k)=\langle B_j, X(\mathbf{s}_k)\rangle$. Using the orthonormality of the B_j again, we obtain

$$\begin{aligned} C(B_j) &= E\left[\left\langle\sum_{\alpha=1}^{\infty}\xi_\alpha(\mathbf{s})B_\alpha, B_j\right\rangle\sum_{i=1}^{\infty}\xi_i(\mathbf{s})B_i\right] \\ &= E\left[\xi_j(\mathbf{s})\sum_{i=1}^{\infty}\xi_i(\mathbf{s})B_i\right] \qquad (3.2)\\ &= \sum_{i=1}^{\infty}E[\xi_i(\mathbf{s})\xi_j(\mathbf{s})]B_i. \end{aligned}$$

Thus, to estimate C, we must estimate the means $E[\xi_i(\mathbf{s})\xi_j(\mathbf{s})]$.

Fix i and j, and define the scalar field z by $z(\mathbf{s})=\xi_i(\mathbf{s})\xi_j(\mathbf{s})$. We can postulate a parametric model for the covariance structure of the field $z(\cdot)$, and use an empirical variogram to estimate $\mu_z=Ez(\mathbf{s})$ as a weighted average of the $z(\mathbf{s}_k)$. Denote the resulting estimate by $\hat{r}_{ij}$.

The empirical version of (3.2) is then

$$\widehat{C}(B_j)=\sum_{i=1}^{K}\hat{r}_{ij}B_i. \qquad (3.3)$$

Relation (3.3) defines the estimator $\widehat{C}$ which acts on the span of B_j, $1\le j\le K$.

Its eigenfunctions are of the form $x=\sum_{1\le\alpha\le K}x_\alpha B_\alpha$. Observe that

$$\widehat{C}(x)=\sum_{\alpha}x_\alpha\sum_{i}\hat{r}_{i\alpha}B_i=\sum_{i}\left(\sum_{\alpha}\hat{r}_{i\alpha}x_\alpha\right)B_i.$$

On the other hand,

$$\lambda x=\sum_{i}\lambda x_i B_i.$$

Since the B_i form an orthonormal basis, we obtain

$$\sum_{\alpha}\hat{r}_{i\alpha}x_\alpha=\lambda x_i.$$

Setting

$$\mathbf{x}=[x_1,x_2,\ldots,x_K]^T,\qquad \widehat{\mathbf{R}}=[\hat{r}_{ij},1\le i,j\le K],$$

we can write the above as a matrix equation

$$\widehat{\mathbf{R}}\mathbf{x} = \lambda \mathbf{x}. \tag{3.4}$$

Denote the solutions to (3.4) by

$$\hat{\mathbf{x}}^{(j)} = [\hat{x}_1^{(j)}, \hat{x}_2^{(j)}, \ldots, \hat{x}_k^{(j)}]^T, \qquad \hat{\lambda}_j, \qquad 1 \le j \le K. \tag{3.5}$$

The $\hat{\mathbf{x}}^{(j)}$ satisfy $\sum_{\alpha=1}^{K} \hat{x}_\alpha^{(j)} \hat{x}_\alpha^{(i)} = \delta_{ij}$. Therefore, the $\hat{v}_j$ defined by

$$\hat{v}_j = \sum_{\alpha=1}^{K} \hat{x}_\alpha^{(j)} B_\alpha \tag{3.6}$$

are also orthonormal (because the B_j are orthonormal). The $\hat{v}_j$ given by (3.6) are the estimators of the FPCs, and the $\hat{\lambda}_j$ in (3.5) of the corresponding eigenvalues.

As in method M3, the value of K can be taken to the number of basis functions used to create the functional objects in `R`, so it can be a relatively large number, for example, $K = 49$. Even though the range of j in (3.5) and (3.6) runs up to K, only the first few estimated FPCs $\hat{v}_j$ would be used in further work.

Method CM2. Recall that under the assumption of zero mean function, the covariance operator is defined by $C(x) = E[\langle X(\mathbf{s}), x\rangle X(\mathbf{s})]$. It can be estimated by the simple average

$$\frac{1}{N}\sum_{n=1}^{N} \langle X(\mathbf{s}_n), \cdot\rangle X(\mathbf{s}_n) = \frac{1}{N}\sum_{n=1}^{N} C_k, \tag{3.7}$$

where C_k is the operator defined by

$$C_k(x) = \langle X(\mathbf{s}_k), x\rangle X(\mathbf{s}_k).$$

As for the mean, more precise estimates can be obtained by using the weighted average

$$\widehat{C} = \sum_{k=1}^{N} w_k C_k. \tag{3.8}$$

Before discussing the estimation of the weights w_k, we comment that the FPCs v_j and their eigenvalues λ_j can be estimated using (3.8) and the representation (3.1). As in method CM3, set $x = \sum_{1\le\alpha\le K} x_\alpha B_\alpha$, and observe that

$$\widehat{C}(x) = \sum_{j=1}^{K} \left(\sum_{\alpha=1}^{K} s_{j\alpha} x_\alpha \right) B_j,$$

where

$$s_{j\alpha} = \sum_{k=1}^{N} w_k \langle X_k, B_j\rangle \langle X_k, B_\alpha\rangle.$$

Thus, analogously to (3.4), we obtain a matrix equation $\mathbf{Sx} = \lambda\mathbf{x}$, from which the estimates of the v_j, λ_j can be found as in (3.5) and (3.6).

We now return to the estimation of the weights w_k in (3.8). One way to define the optimal weights is to require that they minimize the expected Hilbert–Schmidt norm of $\widehat{C} - C$. Recall that the Hilbert–Schmidt norm of an operator K is defined by

$$\|K\|_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} \|K(e_i)\|^2 = \sum_{i=1}^{\infty} \int |K(e_i)(t)|^2 \, dt,$$

where $\{e_i, i \geq 1\}$ is any orthonormal basis in L^2. Since $\|\cdot\|_{\mathcal{S}}$ is a norm in the the Hilbert space $\mathcal{S}$ of the Hilbert–Schmidt operators with the inner product

$$\langle K_1, K_2 \rangle_{\mathcal{S}} = \sum_{i=1}^{\infty} \langle K_1(e_i), K_2(e_i) \rangle,$$

we can repeat all algebraic manipulations needed to obtain the weight w_i in (2.2). The optimal weights in (3.8) thus satisfy

$$\sum_{n=1}^{N} w_n = 1, \qquad \sum_{k=1}^{N} w_k \kappa_{kn} - r = 0, \qquad n = 1, 2, \ldots, N, \tag{3.9}$$

where

$$\kappa_{k\ell} = E[\langle C_k - C, C_\ell - C \rangle_{\mathcal{S}}].$$

Finding the weights thus reduces to estimating the expected inner products $\kappa_{k\ell}$.

Since method M2 of Section 2 relies only on estimating inner product in the Hilbert space L^2, it can be extended to the Hilbert space $\mathcal{S}$. First observe that, analogously to (2.8),

$$E\|C_k - C_\ell\|_{\mathcal{S}}^2 = 2E\|C_k - C\|_{\mathcal{S}}^2 - 2\kappa_{k\ell}.$$

We can estimate the variogram

$$\gamma_C(d) = E\|\langle X(\mathbf{s}), \cdot \rangle X(\mathbf{s}) - \langle X(\mathbf{s}+\mathbf{d}), \cdot \rangle X(\mathbf{s}+\mathbf{d})\|_{\mathcal{S}}^2, \qquad d = \|\mathbf{d}\|$$

by fitting a parametric model. In formulas (4.7) and (4.8), the squared distances $(X(\mathbf{s}_k) - X(\mathbf{s}_\ell))^2$ must be replaced by the squared norms $\|C_k - C_\ell\|_{\mathcal{S}}^2$. These norms are equal to

$$\|C_k - C_\ell\|_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} \int \left(f_{ik} X_k(t) - f_{i\ell} X_\ell(t)\right)^2 dt,$$

where

$$f_{ik} = \int X_k(t) e_i(t) \, dt.$$

The inner products f_{ik} can be computed using the `R` package `fda`.

4. Finite sample performance of the estimators. In this section we report the results of a simulation study designed to compare the performance of the methods proposed in Sections 2 and 3 in a realistic setting motivated by the ionosonde data. It is difficult to design an exhaustive simulation study due to the number of possible combinations of the point distributions, dependence structures, shapes of mean functions and the FPCs and ways of implementing the methods (choice of spatial models, variogram estimation etc.). We do, however, think that our study provides useful information and guidance for practical application of the proposed methodology.

Data generating processes. We generate functional data at location $\mathbf{s}_k$ as

$$X(\mathbf{s}_k;t) = \mu(t) + \sum_{i=1}^{p} \xi_i(\mathbf{s}_k) e_i(t), \tag{4.1}$$

where the e_i are orthonormal functions; cf. model (1.1).

To evaluate the estimators of the mean, we use $p = 2$ and

$$e_1(t) = \sqrt{2}\sin(2\pi t \cdot 6), \qquad e_2(t) = \sqrt{2}\sin(2\pi t/2). \tag{4.2}$$

We use two mean functions

$$\mu(t) = a\sqrt{2}\sin(2\pi t \cdot 3), \qquad a = 2, \tag{4.3}$$

and

$$\mu(t) = a\sqrt{t}\sin(2\pi t \cdot 3), \qquad a = 1. \tag{4.4}$$

The mean function (4.3) resembles the mean shape for the ionosonde data. It is, however, a member of the Fourier basis, and can be isolated using only one basis function, what could possibly artificially enhance the performance of method M3. We therefore also consider the mean function (4.4). Combining the mean function (4.3) and the FPCs (4.2), we obtain functions which very closely resemble the shapes of the ionosonde curves. In the above formulas, time is rescaled so that $t \in [0,1]$.

To evaluate the estimators of the FPCs, we use $p = 3$ and

$$X(\mathbf{s}_k;t) = \xi_1(\mathbf{s}_k)\frac{e_1(t) + e_2(t)}{\sqrt{2}} + \xi_2(\mathbf{s}_k)e_3(t), \tag{4.5}$$

where $e_1(t) = \sqrt{2}\sin(2\pi t \cdot 7)$, $e_2(t) = \sqrt{2}\sin(2\pi t \cdot 2)$, $e_3(t) = \sqrt{2}\sin(3\pi t \cdot 3)$. Direct verification, which uses the independence of the fields ξ_1 and ξ_2, shows that the FPCs are $v_1 = 2^{-1/2}(e_1 + e_2)$ and $v_2 = 2_3$ (for the parameters of the ξ_i specified below).

To complete the description of the data generating processes, we must specify the dependence structure of the scalar spatial fields ξ_1 and ξ_2. A common assumption for the Karhunen–Loéve expansions used in statistical inference is that the score processes ξ_i are independent, and this is what we assume. We use the exponential and Gaussian models (8.4) with chordal

distances (1.2) between the locations described below. To make simulated data look similar to the real foF2 data, we chose $\sigma_1 = 1$, $\rho_1 = \pi/6$ for $\xi_1(\mathbf{s})$ field and $\sigma_2 = 0.1$, $\rho_2 = \pi/4$ for $\xi_2(\mathbf{s})$ field.

The locations $\mathbf{s}_k$ are selected to match the locations of the real ionosonde stations. For the sample size 218 we use all available locations, as shown in Figure 2. Size 32 corresponds to the ionosondes with the longest record history. We also consider a sample of size 100; the 100 stations were selected randomly out of the 218 stations.

Details of implementation. All methods require the specification of a parametric spatial model for the variogram. Even though for some methods the variograms are defined for L^2- or $\mathcal{S}$-valued objects, only a *scalar* model is required. In this simulation study we use the exponential and Gaussian models.

Methods M3, CM2 and CM3 require the specification of a basis $\{B_j\}$ and the number K of the basis functions. We use the Fourier basis and $K = 1 + 4[\sqrt{\#\{t_j\}}]$, where $\#\{t_j\}$ is the count of the points at which the curves are observed. For our real and simulated data $K = 1 + 4[\sqrt{336}] = 73$, a number that falls between the recommended values of 49 and 99 for the number of basis functions. Specifically, the basis functions B_j are

$$\{1, \sqrt{2}\sin(2\pi t \cdot i), \sqrt{2}\cos(2\pi t \cdot i); i = 1, 2, \ldots, 36\}. \tag{4.6}$$

All methods require the estimation of a parametric model on an empirical variogram. There are several versions of the empirical variogram for scalar fields. The classical estimator proposed by Matheron is given by

$$\hat{\gamma}(d) = \frac{1}{|N(d)|} \sum_{N(d)} (X(\mathbf{s}_k) - X(\mathbf{s}_l))^2, \tag{4.7}$$

where $N(d) = \{(\mathbf{s}_i, \mathbf{s}_j) : d_{\mathbf{s}_i,\mathbf{s}_j} = d; i, j = 1, \ldots, N\}$ and $|N(d)|$ is the number of distinct pairs in $N(d)$. A robust estimator proposed by Cressie and Hawkins is defined as

$$\hat{\gamma}(d) = \left(\frac{1}{|N(d)|} \sum_{N(d)} |X(\mathbf{s}_k) - X(\mathbf{s}_l)|^{1/2} \right)^4 \Bigg/ \left(0.457 + \frac{0.494}{|N(d)|} \right). \tag{4.8}$$

For details, we refer to Section 4.4 of Schabenberger and Gotway (2005), where other ways of variogram estimation are also discussed. In our study we use only estimators (4.7) and (4.8), and refer to them, respectively, as MT and CH.

Results of the simulation study. For comparison of different methods we introduce the quantity L which is the average of the integrated absolute differences between real and estimated mean functions or FPCs. For the mean function, L is defined by

$$L = \frac{1}{R} \sum_{r=1}^{R} \int |\hat{\mu}_r(t) - \mu(t)| \, dt, \tag{4.9}$$

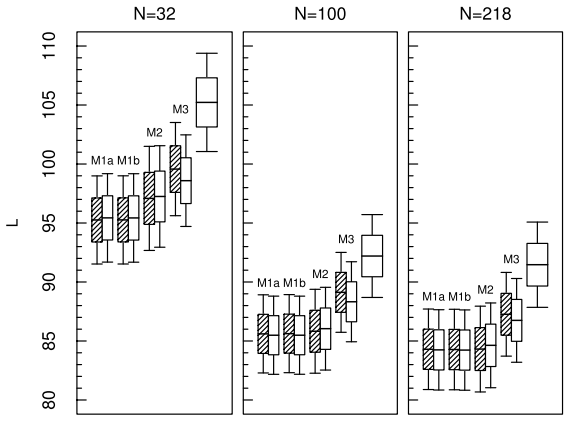

FIG. 4. *Errors in the estimation of the mean function for sample sizes:* 32, 100, 218. *The dashed boxes are estimates using the CH variogram, empty are for the MT variogram. The rightmost box for each N corresponds to the simple average. The bold line inside each box plot represents the average value of L (4.9). The upper and lower sides of rectangles show one standard deviation, and horizontal lines show two standard deviations. The rightmost boxes correspond to the standard method.*

where R is the number of replications; we use $R = 10^3$. For the FPCs the definition is fully analogous. We also compute the standard deviation for L, based on the normal approximation for R independent runs.

The results of the simulations for the mean function (4.4) are shown in Figure 4. The data generating processes have exponential covariance functions. If the ξ_i in (4.1) have Gaussian covariances, the results are not visually distinguishable. The errors values for mean (4.3) are slightly different, but the relative position of the box plots practically does not change. All methods M1, M2 and M3 are significantly better than the sample average. Method M2 strikes the best balance between the computational cost and the precision of estimation. Note that methods M1 and M2 were designed to minimizes the expected L^2 distance, and all three methods are compared using the L^1 distance, so this comparison does not a priori favor them. In the context of forecasting, using different loss functions to evaluate the forecasts than to design them can lead to spurious conclusions; see Gneiting (2011). In our context, if the L^2 distance is used to compare the methods, the ranking and conclusions are the same.

Errors in the estimation of the FPCs in model (4.5) are shown in Figure 5. The displayed errors are those for the ξ_i with exponential covariances and the CH variogram. The results for Gaussian covariances and the MT variogram are practically the same. The performance of methods CM2 and CM3 is comparable, and they are both much better than using the eigenfunctions of the empirical covariance operator (3.7), which is the standard method

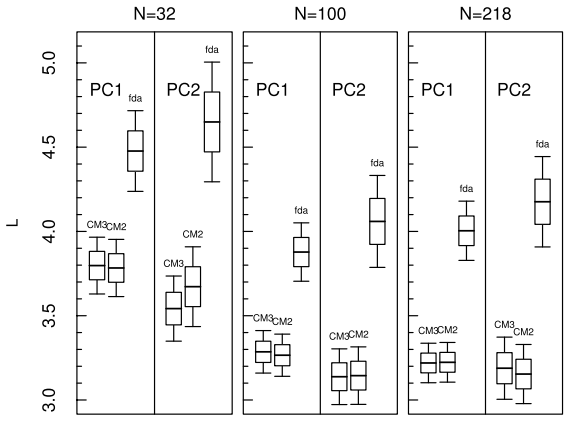

FIG. 5. *Errors in the estimation of the FPCs for sample sizes:* 32, 100, 218. *The bold line inside each box plot represents the average value of L. The upper and lower sides of rectangles show one standard deviation, and horizontal lines show two standard deviations. The rightmost boxes correspond to the standard method.*

implemented in the fda package. The computational complexity of methods CM2 and CM3 is the same.

Conclusions. For simulated data generated to resemble the ionosonde data, all methods introduced in Sections 2 and 3 have integrated absolute deviations (away from a true curve) statistically significantly smaller than the standard methods designed for i.i.d. curves. Methods M2 and CM2, based on weighted averages estimated using functional variograms, offer a computationally efficient and unified approach to the estimation of the mean function and of the FPCs in this spatial setting.

5. Testing for correlation of two spatial fields. Motivated by the problem of testing for correlation between foF2 and magnetic curves, described in detail in Section 8, we now propose a relevant statistical significance test.

There are N spatial locations: $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. At location $\mathbf{s}_k$, we have two curves:

$$X_k = X(\mathbf{s}_k) = X(\mathbf{s}_k; t), \qquad t \in [0,1],$$

and

$$Y_k = Y(\mathbf{s}_k) = Y(\mathbf{s}_k; t), \qquad t \in [0,1].$$

We want to test the null hypothesis that the collections of curves $\{X_k, 1 \le k \le N\}$ and $\{Y_k, 1 \le k \le N\}$ are uncorrelated in a sense defined below. The null distribution is derived assuming a stronger condition that these two families are independent. Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009) introduced measures of dependence based on the distance of characteristic functions which allow to test independence (rather than just

lack of correlation) of random variables X and Y given a sample of i.i.d. observations (X_k, Y_k). The extension of their theory to the case of spatially dependent observations (X_k, Y_k) is not obvious, so we consider only a test for linear dependence.

The idea of the test is as follows. To lighten the notation, assume that

$$EX_k(t) = 0 \quad \text{and} \quad EY_n(t) = 0.$$

The mean functions will be estimated and subtracted using one of the methods of Section 2. We approximate the curves X_n and Y_n by the expansions

$$X_n(t) \approx \sum_{i=1}^{p} \langle X_n, v_i \rangle v_i(t), \qquad Y_n(t) \approx \sum_{j=1}^{q} \langle Y_n, u_j \rangle u_j(t),$$

where the v_i and the u_j are the corresponding FPCs. At this point, the functions $v_i, 1 \le i \le p$ and $u_j, 1 \le j \le q$ are deterministic, so the independence of the curves X_n of the curves Y_n implies the independence of the vectors

$$[\langle X_n, v_1 \rangle, \langle X_n, v_2 \rangle, \ldots, \langle X_n, v_p \rangle]^T, \qquad 1 \le n \le N$$

and

$$[\langle Y_n, u_1 \rangle, \langle Y_n, u_2 \rangle, \ldots, \langle Y_n, u_q \rangle]^T, \qquad 1 \le n \le N.$$

Then, under H_0, the expected value of the sample covariances

$$A_N(i,j) = \frac{1}{N} \sum_{n=1}^{N} \langle X_n, v_i \rangle \langle Y_n, u_j \rangle \tag{5.1}$$

is zero. If their estimated versions are large as a group, that is, if some of the estimated $A_N(i,j)$ are too large, we reject the null hypothesis.

To construct a test statistic, we introduce the quantities

$$V_{k\ell}(i,i') = E[\langle v_i, X_k \rangle \langle v_{i'}, X_\ell \rangle], \qquad U_{k\ell}(j,j') = E[\langle u_j, Y_k \rangle \langle u'_j, Y_\ell \rangle].$$

Note that $V_{k\ell}(i,i') = 0$ and $U_{k\ell}(j,j') = 0$, if the observations in each sample are independent (and have mean zero). Thus, the $V_{k\ell}(i,i')$ and the $U_{k\ell}(j,j')$ are specific to dependent data, they do not occur in the currently available testing procedures developed for independent curves. Setting $X_{ik} = \langle v_i, X_k \rangle, Y_{jk} = \langle u_j, Y_k \rangle$, observe that if the X_{ik} are uncorrelated with the Y_{jk}, then

$$\begin{aligned} E[\sqrt{N} A_N(i,j) \sqrt{N} A_N(i',j')] &= \frac{1}{N} E\left[\sum_{k=1}^{N} X_{ik} Y_{jk} \sum_{\ell=1}^{N} X_{i'\ell} Y_{j'\ell} \right] \\ &= \frac{1}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} E[X_{ik} X_{i'\ell}] E[Y_{jk} Y_{j'\ell}] \\ &= \frac{1}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} V_{k\ell}(i,i') U_{k\ell}(j,j'). \end{aligned}$$

The normalized covariance tensor of the $\sqrt{N}A_N(i,j)$ thus has entries

$$\sigma_N(i,j;i',j') = \frac{1}{N}\sum_{k,\ell=1}^{N} V_{k\ell}(i,i')U_{k\ell}(j,j'). \tag{5.2}$$

The idea of the test is to approximate the distribution of the matrix

$$\mathbf{A}_N = [A_N(i,j), 1 \le i \le p, 1 \le j \le q]$$

via $\sqrt{N}\mathbf{A}_N \approx \mathbf{Z}$, where $\mathbf{Z}$ is a $p \times q$ Gaussian matrix whose elements have covariances $E[Z(i,j)Z(i',j')] = \sigma_N(i,j;i',j')$.

We now explain how to implement this idea. Denote by $\hat{\lambda}_i, \hat{\gamma}_j$ and $\hat{v}_i, \hat{u}_j$ the eigenvalues and the eigenfunctions estimated either by method CM2 or CM3. The covariances $A_N(i,j)$ are then estimated by

$$\hat{A}_N(i,j) = \frac{1}{N}\sum_{n=1}^{N}\langle X_n, \hat{v}_i\rangle\langle Y_n, \hat{u}_j\rangle.$$

If the observations within each sample are independent, an appropriate test statistic is

$$N\sum_{i=1}^{p}\sum_{j=1}^{q}\hat{\lambda}_i^{-1}\hat{\gamma}_j^{-1}\hat{A}_N^2(i,j).$$

Since $\lambda_i = E[\langle v_i, X\rangle^2]$, this is essentially the sum of all correlations, and it tends to a chi-squared distribution with pq degrees of freedom, as shown in Kokoszka et al. (2008). This is, however, not that case for dependent data. To explain, set

$$\mathbf{a}_N = \mathrm{vec}(\mathbf{A}_N),$$

that is, $\mathbf{a}_N$ is a column vector of length pq consisting of the columns of $\mathbf{A}_N$ stacked on top of each other, starting with the first column. Then $\sqrt{N}\mathbf{a}_N$ is approximated by a Gaussian vector $\mathbf{z}$ with covariance matrix $\mathbf{\Sigma}$ constructed from the entries (5.2). It follows that

$$\hat{S}_N = N\hat{\mathbf{a}}_N^T\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{a}}_N \approx \chi^2_{pq}, \tag{5.3}$$

where $\hat{\mathbf{a}}_N = \mathrm{vec}(\hat{\mathbf{A}}_N)$. The entries of the matrix $\hat{\mathbf{\Sigma}}$ are

$$\hat{\sigma}_N(i,j;i',j') = \frac{1}{N}\sum_{k,\ell=1}^{N}\hat{V}_{k\ell}(i,i')\hat{U}_{k\ell}(j,j'), \tag{5.4}$$

where $\hat{V}_{k\ell}(i,i')$ and $\hat{U}_{k\ell}(j,j')$ are estimators of $V_{k\ell}(i,i')$ and $U_{k\ell}(j,j')$, respectively. The test rejects H_0 if $\hat{S}_N > \chi^2_{pq}(1-\alpha)$, where $\chi^2_{pq}(1-\alpha)$ is the $100(1-\alpha)$th percentile of the chi-squared distribution with pq degrees of

freedom. One can use Monte Carlo versions of the above test, for example, the test is based on the approximation

$$\hat{T}_N := N\hat{\mathbf{a}}_N^T\hat{\mathbf{a}}_N \approx \mathbf{w}^T\hat{\mathbf{\Sigma}}\mathbf{w}, \tag{5.5}$$

where the components of $\mathbf{w}$ are i.i.d. standard normal.

The test procedure can be summarized as follows:

(1) Subtract the mean functions, estimated by one of the methods of Section 2, from both samples.

(2) Estimate the FPCs by method CM2 or CM3.

(3) Using a model for the covariance tensor (5.4) (see Section 6), compute the test statistic $\hat{S}_N$. (This tensor is not needed to compute $\hat{T}_N$, but it is needed to find its Monte Carlo distribution.)

(4) Find the P-value using either a Monte Carlo distribution or the χ^2 approximation.

We now turn to the important issue of modeling and estimation of the matrix $\mathbf{\Sigma}$.

6. Modeling and estimation of the covariance tensor. The estimation of the $V_{k\ell}(i,i')$ involves only the X_n, and the estimation of the $U_{k\ell}(j,j')$ only the Y_n, so we describe only the procedure for the $V_{k\ell}(i,i')$. We assume that the mean has been estimated and subtracted, so that we can define

$$C_h(x) = E[\langle X(\mathbf{s}), x\rangle X(\mathbf{s}+\mathbf{h})], \qquad h = \|\mathbf{h}\|. \tag{6.1}$$

The estimation of the $V_{k\ell}(i,i')$ relies on the identity

$$V_{k\ell}(i,i') = \langle C_h(v_i), v_{i'}\rangle, \qquad h = d(\mathbf{s}_k, \mathbf{s}_\ell).$$

To propose a practical approach to the estimation of $\mathbf{\Sigma}$, we consider an extension of the multivariate intrinsic model; see, for example, Chapter 22 of Wackernagel (2003). A most direct extension is to assume that

$$C_h = Cr(h), \tag{6.2}$$

where C is a covariance operator, that is, a symmetric positive definite operator with summable eigenvalues, and $r(h)$ is a correlation function of a scalar random field. Since $r(0) = 1$, we have $C = C_0$, so C in (6.2) must be the covariance operator of each $X(\mathbf{s})$. If we assume the intrinsic model (6.2), then

$$V_{k\ell}(i,j) = \langle r(h)C(v_i), v_j\rangle = \lambda_i\delta_{ij}r(d(\mathbf{s}_k,\mathbf{s}_l)). \tag{6.3}$$

To allow more modeling flexibility, we postulate that

$$V_{k\ell}(i,j) = \lambda_i\delta_{ij}r_i(d(\mathbf{s}_k,\mathbf{s}_l)). \tag{6.4}$$

Under (6.3) [equivalently, under (6.2)], each scalar field $\langle X(\mathbf{s}), v_i\rangle$ has the same correlation function, only their variances are different. Under (6.4), the fields $\langle X(\mathbf{s}), v_i\rangle$ can have different correlation functions. As will be seen below, model (6.4) also leads to a valid covariance matrix.

The correlations $r_i(d(\mathbf{s}_k, \mathbf{s}_l))$ and the variances λ_i can be estimated using a parametric model for the scalar field $\xi_i(\mathbf{s}) = \langle X(\mathbf{s}), v_i\rangle$. The resulting estimates $\hat{r}_i(d(\mathbf{s}_k, \mathbf{s}_l))$ and $\hat{\lambda}_i$ lead to the estimates $\hat{V}_{k\ell}(i,j)$ via (6.4). Analogous estimates of the functional field Y are $\hat{\gamma}_j(d(\mathbf{s}_k, \mathbf{s}_l)), \hat{\tau}_j$ and $\hat{U}_{k\ell}(i,j)$.

For ease of reference, we note that under model (6.4) and H_0, the covariance tensor,

$$\left[\frac{1}{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\hat{V}_{k\ell}(i,i')\hat{U}_{k\ell}(j,j'), 1 \le i, i' \le p, 1 \le j, j' \le q\right],$$

has the following matrix representation:

$$\hat{\mathbf{\Sigma}} = \operatorname{diag}\left(\sum_{k=1}^{N}\sum_{\ell=1}^{N}\hat{\mathbf{\Sigma}}_{\xi_1}(k,\ell)\hat{\mathbf{\Sigma}}_{\eta_1}(k,\ell), \ldots, \sum_{k=1}^{N}\sum_{\ell=1}^{N}\hat{\mathbf{\Sigma}}_{\xi_p}(k,\ell)\hat{\mathbf{\Sigma}}_{\eta_q}(k,\ell)\right), \tag{6.5}$$

where

$$\hat{\mathbf{\Sigma}}_{\xi_i}(k,\ell) = \frac{1}{\sqrt{N}}\hat{\lambda}_i\hat{r}_i(d(\mathbf{s}_k, \mathbf{s}_\ell))$$

and

$$\hat{\mathbf{\Sigma}}_{\eta_j}(k,\ell) = \frac{1}{\sqrt{N}}\hat{\gamma}_j\hat{\tau}_j(d(\mathbf{s}_k, \mathbf{s}_\ell)).$$

This form is used to construct the Monte Carlo tests discussed in Section 7.

The matrices $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ are positive definite; see Horváth and Kokoszka (2012) for the verification.

7. Size and power of the correlation test. As in Section 4, our objective is to evaluate the finite sample performance of the test introduced in Section 5 in a realistic setting geared toward the application presented in Section 8.

Data generating processes. We generate samples of zero mean Gaussian processes

$$X(\mathbf{s};t) = \sum_{i=1}^{p}\xi_i(\mathbf{s})v_i(t); \qquad Y(\mathbf{s};t) = \sum_{j=1}^{q}\eta_j(\mathbf{s})u_j(t). \tag{7.1}$$

The process X is designed to resemble in distribution appropriately transformed and centered foF2 curves; the process Y the centered magnetic curves. Following the derivation presented in Section 8, we use $p = 7$ and $q = 1$. The curves v_i and u_1 are the estimated FPCs of the real data. The

TABLE 1
Models and estimated covariance parameters for the transformed foF2 curves and the magnetic curves

Spatial field	Model	Parameters		
		c_0	σ^2	ρ
$\boldsymbol{\eta}$	Gaussian	–	5.99 ± 0.48	0.32 ± 0.04
$\boldsymbol{\xi}_1$	Gaussian	–	20.05 ± 2.20	0.12 ± 0.03
$\boldsymbol{\xi}_2$	–	–	3.30 ± 0.43	–
$\boldsymbol{\xi}_3$	Exponential	–	2.63 ± 0.52	0.16 ± 0.07
$\boldsymbol{\xi}_4$	Gaussian	–	2.66 ± 0.39	0.18 ± 0.05
$\boldsymbol{\xi}_5$	–	–	2.74 ± 0.32	–
$\boldsymbol{\xi}_6$	Gaussian	0.16 ± 0.02	0.85 ± 0.24	0.17 ± 0.06
$\boldsymbol{\xi}_7$	–	–	1.22 ± 0.18	–

scalar Gaussian spatial fields ξ_i and η_1 follow parametric models estimated for real data; details of the models are presented in Table 1. The ξ_i are independent. Under H_0, the ξ_i are independent of η_1. The dependence under H_A can be generated in many ways. We considered the following scenarios: ξ_1 and η_1 are dependent, ξ_i and η_1 are independent for $i \neq 1$, then ξ_2 and η_1 are dependent, ξ_i and η_1 are independent for $i \neq 2$, etc. To produce two dependent spatial fields ξ_i and η, we generated N i.i.d. pairs $\mathbf{x}_i = [x_{1i}, x_{2i}]^T$, $1 \leq i \leq N$, where

$$\mathbf{x}_i \sim N\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Then we merged all x_{1i} into vector $\mathbf{y}_1 = [x_{11}, \ldots, x_{1N}]^T$ and all x_{2i} into vector $\mathbf{y}_2 = [x_{21}, \ldots, x_{2N}]^T$. Performing the Cholesky rotation, we obtain correlated spatial vectors:

$$\boldsymbol{\xi}_i = \mathbf{V}\mathbf{y}_1, \qquad (\boldsymbol{\Sigma}_{\xi_i} = \mathbf{V}\mathbf{V}^T), \qquad \boldsymbol{\eta} = \mathbf{U}\mathbf{y}_2, \qquad (\boldsymbol{\Sigma}_{\eta} = \mathbf{U}\mathbf{U}^T).$$

We used sample sizes $N = 32$ and $N = 100$ corresponding to the locations determined as in Section 4.

Testing procedures. We studied the finite sample behavior of three methods, which we call S, SM and T. Method S rejects H_0 if the statistic $\hat{S}_N$ (5.3) exceeds a chi-square critical value. Method SM uses a Monte Carlo distribution of the statistic $\hat{S}_N$: after estimating all parameters from the data and assuming the Gaussian distribution of the fields ξ_i and η_1, we can replicate the values of the statistic $\hat{S}_N$ under H_0 using the covariance matrix (6.5). Method T uses the statistics $\hat{T}_N$ (5.5), and approximates its distribution by the Monte Carlo distribution of $\mathbf{w}^T\hat{\boldsymbol{\Sigma}}\mathbf{w}$, as explained in Section 5. For determining the critical values in methods SM and T, we used 10^7 Monte Carlo replications. The empirical size and power are based on 10^5 independent runs.

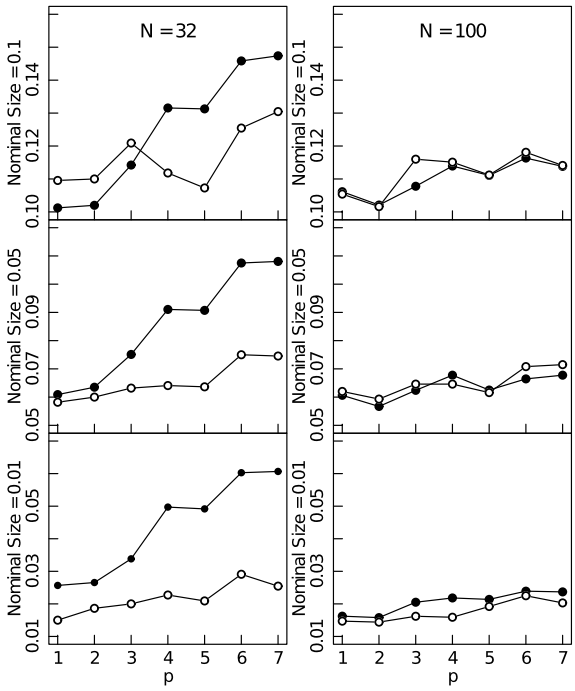

FIG. 6. *Size of the correlation test as a function of p. Solid disks represent method S (based on χ^2 distribution). Circles represent method SM (based on the Monte Carlo distribution).*

Conclusions. As Figure 6 shows, the empirical size is higher than the nominal size, and it tends to increase with the number p of principal components used to construct the test, especially for $N = 32$. The usual recommendation is to use p, which explains about 85% of the variance. For the foF2 data with $N = 32$, this corresponds to $p = 4$. Applied to real data in Section 8, all tests (S, SM and T) lead to extremely strong rejections, so the inflated empirical size is not a problem. Figure 6 also shows that the Monte Carlo approximation is useful for $N = 32$, this is the sample size we must use in Section 8. The size of test T is practically indistinguishable from that of test SM. Figure 7 shows the power of method SM; power curves for method T are practically the same, method S has higher power. The simulation study shows that a strong rejection when the test is applied to real data can be viewed as reliable evidence of dependence.

8. Application to critical ionospheric frequency and magnetic curves. In this section we apply the correlation test, which uses the estimation methodology of Sections 2 and 3, to foF2 and magnetic curves.

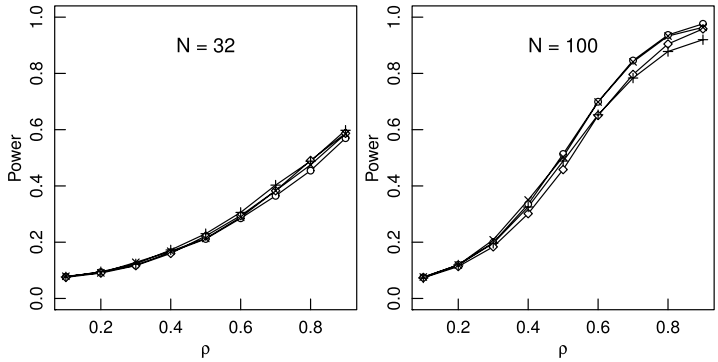

FIG. 7. *Power of the correlation test SM as a function of the population correlation ρ. Each line represents one of the four possible correlated spatial fields: $\boldsymbol{\xi}_1 - \boldsymbol{\eta}$, $\boldsymbol{\xi}_2 - \boldsymbol{\eta}$, $\boldsymbol{\xi}_3 - \boldsymbol{\eta}$, $\boldsymbol{\xi}_4 - \boldsymbol{\eta}$. The test was performed using $p = 4$, which explains about 85% of variance of the foF2 curves. Since all curves in the graphs are practically the same, we do not specify which curve represents a particular dependent pair $\boldsymbol{\xi}_i - \boldsymbol{\eta}$.*

Description of the data. The F2 layer of the ionosphere is the upper part of the F layer shown in Figure 3. The F2 layer electron critical frequency, foF2, is measured using an instrument called the ionosonde, a type of radar. The foF2 frequency is used to estimate the location of the peak electron density, so an foF2 trend corresponds to a trend in the average height of the ionosphere over a spatial location. The foF2 data have therefore been used to test the hypothesis of ionospheric global cooling discussed in the Introduction. Hourly values of foF2 are available from the SPIDR database `http://spidr.ngdc.noaa.gov/spidr/` for more than 200 ionosondes. We use monthly averages for 32 selected ionosondes, with sufficiently complete records, for the period 1964–1992. Their locations are shown in Figure 2. Three typical foF2 curves are shown in Figure 1. We omit the details of the procedure for obtaining curves like those shown in Figure 1, but we emphasize that it requires a great deal of work. In particular, the SPIDR data suffer from two problems. First, for some data, the amplitude is artificially magnified ten times, and needs to be converted into standard units (MHz). Second, in many cases, missing observations are not replaced by the standard notation 9999, but rather just skipped. Thus, if one wants to use equally-spaced time series, skipped data must be found and replaced by missing values. For filling in missing values, we perform linear interpolation. We developed a customized C++ code to handle these issues. We emphasize that one of the reasons why this global data set has not been analyzed so far is that useable data have been derived only over relatively small regions, for example, Western Europe, and more often for a single location.

As explained in Section 1, the foF2 data are used to test hypotheses on long-term ionospheric trends. We thus removed annual and higher frequency

variations using 16 month averaging with MODWT filter; see Chapter 5 of Percival and Walden (2000). This leads to 32 time series at different locations, each containing 336 equally-spaced temporal observations. The amplitude of the foF2 curves exhibits a nonlinear latitudinal trend; it decreases as the latitude increases; see Figure 1. To remove this trend, which may potentially bias the test, we assume that the foF2 signal, $F(\mathbf{s};t)$, at location $\mathbf{s}$ follows the model

$$F(\mathbf{s};t) = G(L(\mathbf{s}))X(\mathbf{s};t), \tag{8.1}$$

where $X(\mathbf{s};t)$ is a constant amplitude field, and $G(\cdot)$ is a scaling function which depends only on the *magnetic* latitude L (in radians). Since the trend in the amplitude of $F(\mathbf{s};t)$ is caused by the solar radiation which is nonlinearly proportional to the zenith angle, we postulate that the function $G(\cdot)$ has the form

$$G(L) = a + b\cos^c(L). \tag{8.2}$$

The parameters a, b, c are estimated as follows. Let $\mathbf{s}_0$ be the position of the ionosonde closest to the magnetic equator. For identifiability, we set $G(L(\mathbf{s}_0)) = 1$. For the remaining locations $\mathbf{s}_k$, we compute $\hat{G}(L(\mathbf{s}_k))$ as the average, over all 336 time points t_j of the ratio $F(\mathbf{s}_k;t_j)/F(\mathbf{s}_0;t_j)$. Figure 8 shows these ratios as a function of the magnetic and geographic latitude. The ratios in the magnetic latitude show much less spread, and this is another reason why we work with the magnetic latitude. The curve $G(L)$ (8.2) is fitted to the $\hat{G}(L(\mathbf{s}_k))$ in magnetic latitude by nonlinear least squares. The fitted values are $a = 0.5495$, $b = 0.4488$, $c = 4.2631$.

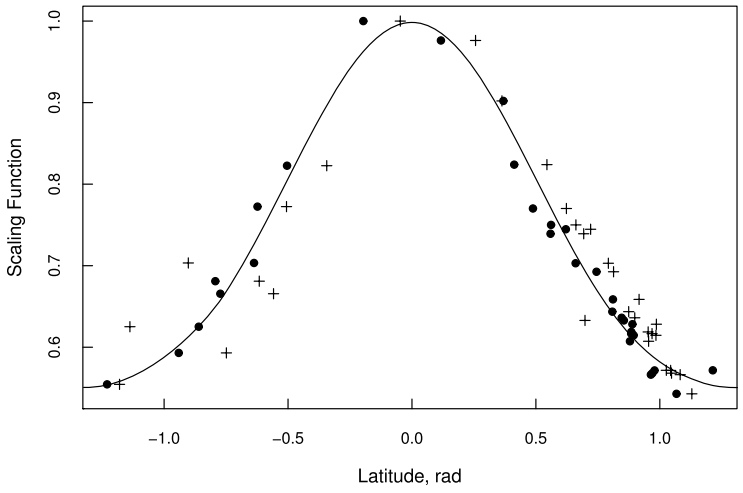

FIG. 8. *Dots represent the scaling function $G_L(\mathbf{s}_i)$ in the magnetic coordinate system and crosses are the same in the geographic coordinate system. Line is the best fit for G_L in the magnetic coordinate system.*

We now describe how we construct the curves that reflect the relevant long-term changes in the internal magnetic field of the earth. The height of the F2 layer (and so the foF2 frequency) can be affected by a vertical plasma drift which responds to the magnetic field. The vertical plasma drift is due to the wind effect, and is given by [we use the same notation as in Mikhailov and Marin (2001)]

$$W = (V_{nx} \cos D - V_{ny} \sin D) \sin I \cos I + V_{nz} \sin^2 I.$$

In the above formula, V_{nx}, V_{ny} and V_{nz} are, respectively, meridional (parallel to constant longitude lines), zonal (parallel to constant latitude lines) and vertical components of the thermospheric neutral wind; I and D are inclination and declination of the earth magnetic field. Detailed figures are provided in Chapter 13 of Kivelson and Russell (1997). Usually $V_{nz} \ll V_{nx}, V_{ny}$, and assuming that the difference between magnetic and geographic coordinates, D, is small (at least for low- and mid-latitude regions), we can simplify the above formula to $W = V_{nx} \sin I \cos I$. Thus, only the meridional thermospheric wind is significant. Measuring neutral wind components (V_{nx}, V_{ny}, V_{nz}) is difficult, and long-term wind records are not available. We therefore replace V_{nx} by its average. For the test of correlation, the specific value of this average plays no role, so we define the magnetic curves as

$$Y(\mathbf{s};t) = \sin I(\mathbf{s};t) \cos I(\mathbf{s};t). \tag{8.3}$$

The curves $I(\mathbf{s};t)$ are computed using the international geomagnetic reference field (IGRF); the software is available at http://www.ngdc.noaa.gov/IAGA/vmod/.

The test is applied to the curves $X(\mathbf{s}_k;t)$ defined by (8.1) and (8.2), and to the curves $Y(\mathbf{s}_k;t)$ defined by (8.3).

Application of the correlation test. We first estimate and subtract the mean functions of the fields $X(\mathbf{s}_k)$ and $Y(\mathbf{s}_k)$ using method M2 (the other spatial methods give practically the same estimates). The principal components v_i and u_i are estimated using method CM2 (method CM3 gives practically the same curves).

We apply the test, for all $1 \le p \le 7$ and $q = 1$. The first seven eigenvalues of the field X [computed per (3.4) or its analog for method CM2] explain about 95% of the variance. The first eigenvalue of the field Y explains about 99% of the variance. The eigenfunction u_1 is approximately equal to the linear function: $u_1(t) \sim t$. This means that at any location, after removing the average, the magnetic field either linearly increases or decreases, with slopes depending on the location; see Figure 9. To lighten the notation, we drop the "hats" from the estimated scores and denote the zero mean vector $[xi_i(\mathbf{s}_1), \ldots, \xi_i(\mathbf{s}_N)]^T$ by $\boldsymbol{\xi}_i$, and $[\eta_1(\mathbf{s}_1), \ldots, \eta_1(\mathbf{s}_N)]^T$ by $\boldsymbol{\eta}$. The

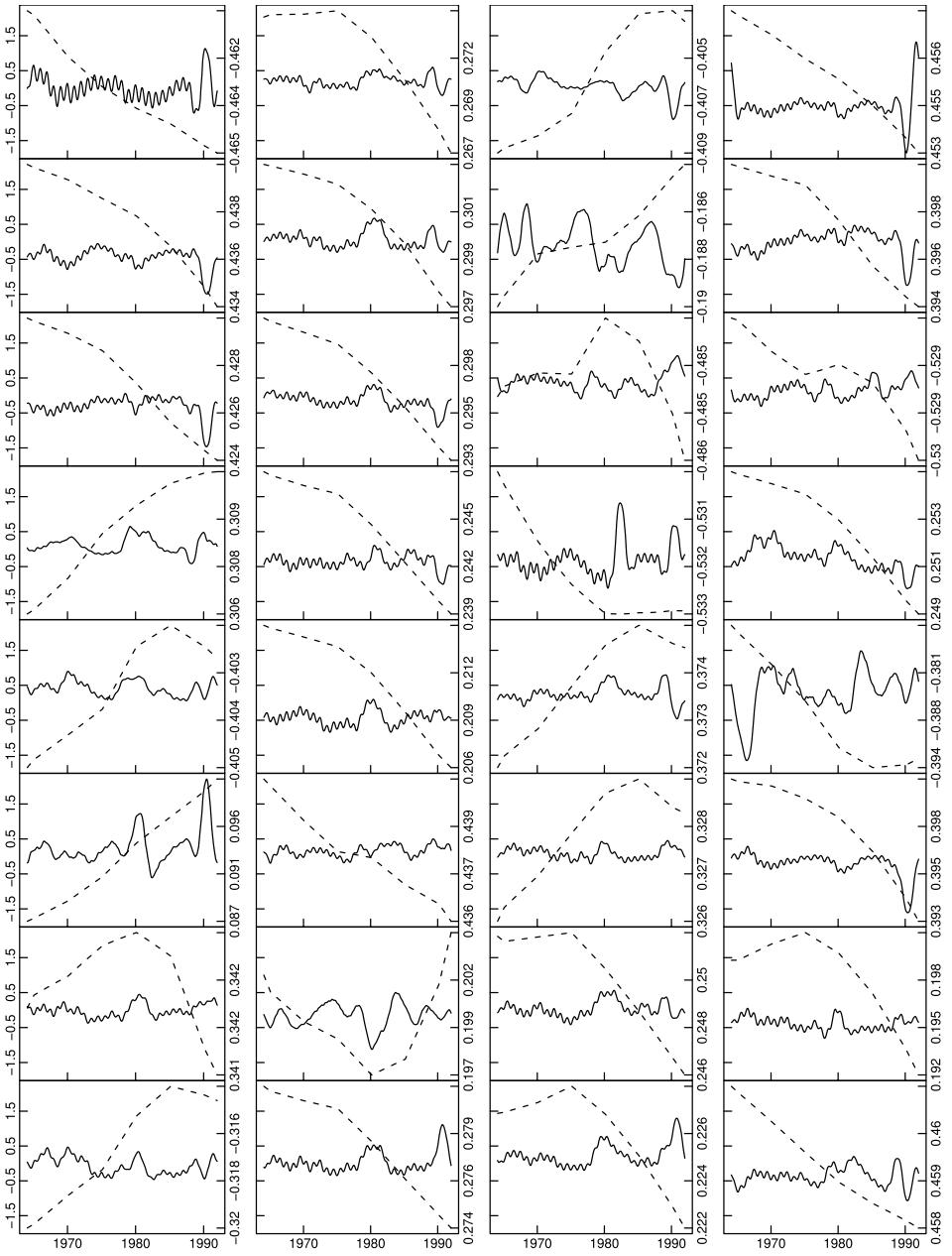

FIG. 9. *Transformed and centered foF2 curves (continuous) and centered magnetic curves (dashed) at 32 locations denoted with circles in Figure 2. The scales for the two families of curves are different. The foF2 curves have the same scale, it is shown on the right vertical axes in MHz. The scale of the magnetic curves changes, it is shown on the right vertical axes in each box (unitless).*

TABLE 2

P-values of the correlation tests applied to the transformed foF2 data. The first column shows the number of FPCs, the second column shows cumulative variances computed as the ratios of the eigenvalues estimated using method CM2. Testing procedures S, SM and T are defined in Section 7. The "simple" procedure neglects the spatial dependence of the curves

		Spatial			
p	CV, %	S	SM	T	Simple
1	47.88	$6.22 \cdot 10^{-5}$	$3.05 \cdot 10^{-4}$	$3.05 \cdot 10^{-4}$	0.035
2	62.59	$3.26 \cdot 10^{-6}$	$2.91 \cdot 10^{-4}$	$2.99 \cdot 10^{-4}$	0.095
3	73.67	$4.53 \cdot 10^{-8}$	$2.43 \cdot 10^{-4}$	$2.32 \cdot 10^{-4}$	0.043
4	84.40	$1.47 \cdot 10^{-26}$	$1.6 \cdot 10^{-7}$	$2.24 \cdot 10^{-5}$	0.039
5	88.70	$4.95 \cdot 10^{-26}$	$2.6 \cdot 10^{-7}$	$2.27 \cdot 10^{-5}$	0.046
6	92.21	$6.73 \cdot 10^{-27}$	$5.9 \cdot 10^{-7}$	$2.21 \cdot 10^{-5}$	0.060
7	94.57	$2.12 \cdot 10^{-32}$	$1.6 \cdot 10^{-7}$	$1.92 \cdot 10^{-5}$	0.030

covariances $\mathbf{\Sigma}_{\xi_i}$ and $\mathbf{\Sigma}_\eta$ are estimated using parametric spatial models determined by the inspection of the empirical variograms. In this application, it is sufficient to use two covariance models:

$$\begin{aligned} &\text{Gaussian: } c(\mathbf{s}_k, \mathbf{s}_\ell) = c_0 + \sigma^2 \exp\{-d^2(k,\ell)/\rho^2\}, \\ &\text{Exponential: } c(\mathbf{s}_k, \mathbf{s}_\ell) = c_0 + \sigma^2 \exp\{-d(k,\ell)/\rho\}. \end{aligned} \tag{8.4}$$

When the scores do not have a spatial structure, we use the sample variance (flat variogram). The estimated models and their parameters are listed in Table 1.

The P-values for different numbers of FPCs $1 \le p \le 7$ are summarized in Table 2. Independent of p and a specific implementation of the test, all P-values are very small, and so the rejection of the null hypothesis is conclusive; we conclude that there is a statistically significant correlation between the foF2 curves $X(\mathbf{s}_k)$ and the magnetic curves $Y(\mathbf{s}_k)$. We also applied a version of our test which neglects any spatial dependence, this is the test proposed by Kokoszka et al. (2008). The P-values hover around the 5% level, but still point toward rejection. The evidence is, however, much less clear cut. This may partially explain why this issue has been a matter of much debate in the space physics community. The correlation between the foF2 and magnetic curves is far from obvious. Figure 9 shows these pairs at all 32 locations. It is hard to conclude by eye that the direction of the magnetic field change impacts the foF2 curves.

Discussion. A very important role in our analysis is played by the transformation (8.2). Applying the test to the original foF2 curves, $F(\mathbf{s}_k; t)$, gives the P-values 0.209 ($p = 1$) and 0.011 ($p = 2$) for the spatial S test, and 0.707 ($p = 1$), 0.185 ($p = 2$), 0.139 ($p = 3$) for the "simple" test. As explained above,

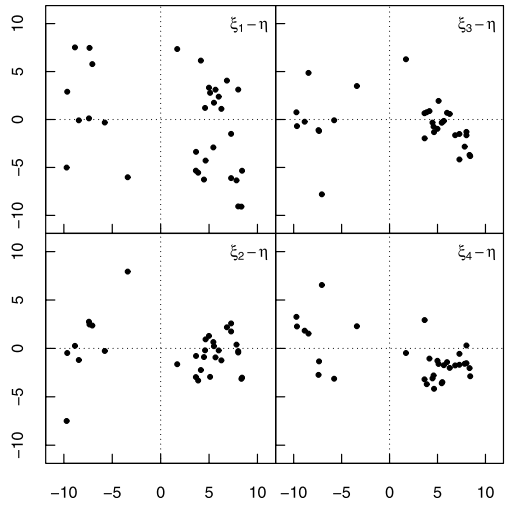

FIG. 10. *Scatter plots of the scores* $\boldsymbol{\xi}_i, i = 1, 2, 3, 4$ *of the foF2 curves, vertical axes, against the scores* $\boldsymbol{\eta}$ *of the magnetic curves, horizontal axes.*

the amplitude of the field $F(\mathbf{s}_k; t)$ evolves with the latitude. This invalidates the assumption of a mean function which is independent of the spatial location. Thus, even for the spatial test, the mean function confounds the first FPC. However, the spatial estimation of the mean function and of the FPCs "quickly corrects" for the violation of assumptions, and the null hypothesis is rejected for $p \geq 2$. When the spatial structure is neglected (and no latitudinal transformation is applied) no correlation between the foF2 curves and magnetic curves is found.

The rejection of the null hypothesis means that after adjusting the foF2 curves for the latitude and the global mean, their regional variability is correlated with the regional changes in the magnetic field. This conclusion agrees with recent space physics research [see Cnossen and Richmond (2008) and Laštovička (2009)], and can, to some extent, be visually confirmed, post-analysis, by the examination of the scatter plots shown in Figure 10. It implies that long-term magnetic trends must be considered as additional covariates in testing for long-term trends in the foF2 curves. The main covariate is the solar activity which drives the shape of the mean function, but, as explained in the Introduction, the impact of the concentration of the greenhouse gases is of particular interest; see Qian et al. (2009), among many other contributions.

A broader conclusion of the work presented in this paper is that methods of functional data analysis must be applied with care to curves obtained at spatial locations. Neglecting the spatial dependence can lead to incorrect conclusions and biased estimates. The same applies to space physics research. If trends or models are estimated separately at each spatial location,

one should not rely on results obtained by some form of a simple averaging. This is, however, the prevailing approach. Interestingly, the results related to global ionospheric trends are often on the borderline of statistical significance. Standard t-tests lead either to rejection or acceptance, depending on a specific method used (a similar phenomenon is observed in the last column of Table 2). It is hoped that the methodology developed in this paper will be useful in addressing such issues.

Acknowledgment. We are grateful to Levan Lomidze for his help in processing the ionosonde data.

REFERENCES

Bel, L., Bar-Hen, A., Petit, R. and Cheddadi, R. (2011). Spatio-temporal functional regression on paleoecological data. *J. Appl. Stat.* **38** 695–704. MR2773575

Cnossen, I. and Richmond, A. D. (2008). Modelling the effects of changes in the Earth's magnetic field from 1957 to 1997 on the ionospheric hmF2 and foF2 parameters. *Journal of Atmospheric and Solar-Terrestrial Physics* **70** 1512–1524.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics* **21** 224–239. MR2842240

Finkenstädt, B., Held, L. and Isham, V., eds. (2007). *Statistical Methods for Spatio-Temporal Systems. Monographs on Statistics and Applied Probability* **107**. Chapman & Hall/CRC, Boca Raton, FL. MR2307967

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P., eds. (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL. MR2761512

Giraldo, R., Delicado, P. and Mateu, J. (2011a). Ordinary kriging for function-valued spatial data. *Environ. Ecol. Stat.* **18** 411–426.

Giraldo, R., Delicado, P. and Mateu, J. (2011b). A generalization of cokriging and multivariable spatial prediction for functional data. Technical report, Univ. Politécnica de Catalunya, Barcelona.

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106** 746–762.

Hörmann, S. and Kokoszka, P. (2012). Consistency of the mean and the principal components of spatially distributed functional data. *Bernoulli*. To appear.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, Berlin.

Kivelson, M. G. and Russell, C. T., eds. (1997). *Introduction to Space Physics*. Cambridge Univ. Press, Cambridge.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* **36** 207–222. MR2431682

Laštovička, J. (2009). Global pattern of trends in the upper atmosphere and ionosphere: Recent progress. *Journal of Atmospheric and Solar-Terrestrial. Physics* **71** 1514–1528.

Laštovička, J., Akmaev, R. A., Beig, G., Bremer, J., Emmert, J. T., Jacobi, C., Jarvis, J. M., Nedoluha, G., Portnyagin, Y. I. and Ulich, T. (2008). Emerging pattern of global change in the upper atmosphere and ionosphere. *Annales Geophysicae* **26** 1255–1268.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2009). Removal of nonconstant daily variation by means of wavelet and functional data analysis. *Journal of Geophysical Research* **114** A03202.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2010a). Estimation of Sq variation by means of multiresolution and principal component analyses. *Journal of Atmospheric and Solar-Terrestial Physics* **72** 625–632.

Maslova, I., Kokoszka, P., Sojka, J. and Zhu, L. (2010b). Statistical significance testing for the association of magnetometer records at high-, mid- and low latitudes during substorm days. *Planetary and Space Science* **58** 437–445.

Mikhailov, A. V. and Marin, D. (2001). An interpretation of the foF2 and hmF2 long-term trends in the framework of the geomagnetic control concept. *Annales Geophysicae* **19** 733–748.

Nerini, D., Monestiez, P. and Manté, C. (2010). Cokriging for spatial functional data. *J. Multivariate Anal.* **101** 409–418. MR2564350

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **4**. Cambridge Univ. Press, Cambridge. MR1770693

Qian, L., Burns, A. G., Solomon, S. C. and Roble, R. G. (2009). The effect of carbon dioxide cooling on trends in the F2-layer ionosphere. *Journal of Atmospheric and Solar-Terrestrial Physics* **71** 1592–1601.

Ramsay, J., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

Rishbeth, H. (1990). A greenhouse effect in the ionosphere? *Planetary and Space Science* **38** 945–948.

Roble, R. G. and Dickinson, R. E. (1989). How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and thermosphere? *Geophysical Research Letters* **16** 1441–1444.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL. MR2134116

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127

Ulich, T., Cliilverd, M. A. and Rishbeth, H. (2003). Determining long-term change in the ionosphere. *Eos, Transactions American Geophysical Union* **84** 581–585.

Wackernagel, H. (2003). *Multivariate Geostatistics*, 3rd ed. Springer, New York.

Yamanishi, Y. and Tanaka, Y. (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *J. Japanese Soc. Comput. Statist.* **15** 307–317. MR2027947

O. Gromenko
P. Kokoszka
Department of Mathematics and Statistics
Utah State University
Logan, Utah 84322-3900
USA
E-mail: agromenko@gmail.com
Piotr.Kokoszka@usu.edu

L. Zhu
J. Sojka
Department of Physics and
Center for Atmospheric and Space Science
Utah State University
Logan, Utah 84322-4405
USA
E-mail: zhu@cc.usu.edu
sojka@cc.usu.edu