Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2012)
61, *Part* 5, *pp.* 715–731

Testing the equality of mean functions of ionospheric critical frequency curves

Oleksandr Gromenko
Utah State University, Logan, USA

and Piotr Kokoszka
Colorado State University, Fort Collins, USA

[Received April 2011. Final revision December 2011]

Summary. The paper develops a significance test for evaluating the equality of the mean functions in two samples of spatially indexed functional data. The problem is motivated by an important question in space physics research which is related to the hypothesis of ionospheric global cooling (as opposed to the conjectured global warming of near surface atmosphere). The critical electron frequency of the ionosphere's F2 region, foF2, can be used to test conjectures about the trends in the ionosphere empirically. We apply the proposed test to the foF2-records over eastern and western Europe to verify whether there is a conjectured difference between first-order behaviour of these records over regions with different magnetic inclinations. It is found that the difference between the means is statistically significant for night-time records. The implications of this result are discussed. Finite sample performance of the test proposed is validated via numerical simulations.

Keywords: Functional data analysis; Ionosphere; Spatial statistics; Two-sample problem

1. Introduction

Over the last two decades, functional data analysis has established itself as an important and dynamic area of statistics which has provided intuitive and computationally feasible approaches to many applied problems. Ramsay and Silverman (2005) offers an excellent and accessible introduction to the central ideas of the field. Ramsay *et al.* (2009) provide a concise introduction which focuses on computational issues. Many recent developments are studied in Ferraty and Vieu (2006), Bosq and Blanke (2007), Ferraty and Romain (2011), Shi and Choi (2011) and Horváth and Kokoszka (2012). Relatively little attention has, however, focused on inferential methods for spatially indexed curves, even though data of this type are quite common; a data object is a curve $X(\mathbf{s}_k)$ observed at location $\mathbf{s}_k$. In many cases, the curves are functions of time so $X(\mathbf{s}_k; t)$ is the value of the function $X(\mathbf{s}_k)$ at time t. Such a data structure is a special case of a spatiotemporal process which calls for statistical tools specifically designed for it. Many environmental and geophysical data sets are of this type. A well-known example is the Canadian weather data set which consists of temperature and precipitation curves at 35 locations; see Ramsay and Silverman (2005). Another example is the Australian rainfall data set, which was recently studied by Delaigle and Hall (2010), which consists of daily rainfall measurements from 1840 to 1990 at 191 Australian weather stations. Snow water curves measured at several dozen locations in Western US states over many decades provide climatic information which

Address for correspondence: Oleksandr Gromenko, Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA.
E-mail: agromenko@gmail.com

 0035–9254/12/61715

is of importance for urban and agricultural development planning in the western states with little summer time rainfall; see for example Carroll *et al.* (1995) and Carroll and Cressie (1996). Another important example is pollution curves: $X(\mathbf{s}_k; t)$ is the concentration of a pollutant at time t at location $\mathbf{s}_k$. Data of this type were studied by Kaiser *et al.* (2002). In many studies, $X(\mathbf{s}_k; t)$ is the count at time t of infectious disease cases, where $\mathbf{s}_k$ is a representative location, e.g. a 'middle point' of a county. Still another example arises in modelling brain activity based on continuous time records obtained from probes placed at different locations in the brain; see Aston and Kirch (2011) and references therein. Delicado *et al.* (2010) have reviewed other examples and recent contributions to the methodology for spatially indexed functional data.

The data which motivate the research that is presented in this paper are not well known in the statistical community but have played a central role in space physics research for many decades. Since the 1930s the ionosphere has been studied by an instrument called the ionosonde, which is a type of radar vertically emitting a frequency spectrum and recording the profile of the signal returned. The returned profile is called the ionogram; an example is given in Fig. 1. The ionogram contains implicit information about the physical properties of the ionosphere directly above the location of the ionosonde. The specific data set that we study, derived from ionograms, consists of hourly records of the so-called electron critical frequency in that the F2 ionosphere region, foF2, at 13 locations in Europe; we describe this data set in detail in Section 2. The F2 region is the main part of the ionosphere's F region in the range of heights 250–350 km above sea level; Fig. 2. For a brief overview of the structure of the ionosphere and its properties see, for example, chapter 1 in Kelly (2009).

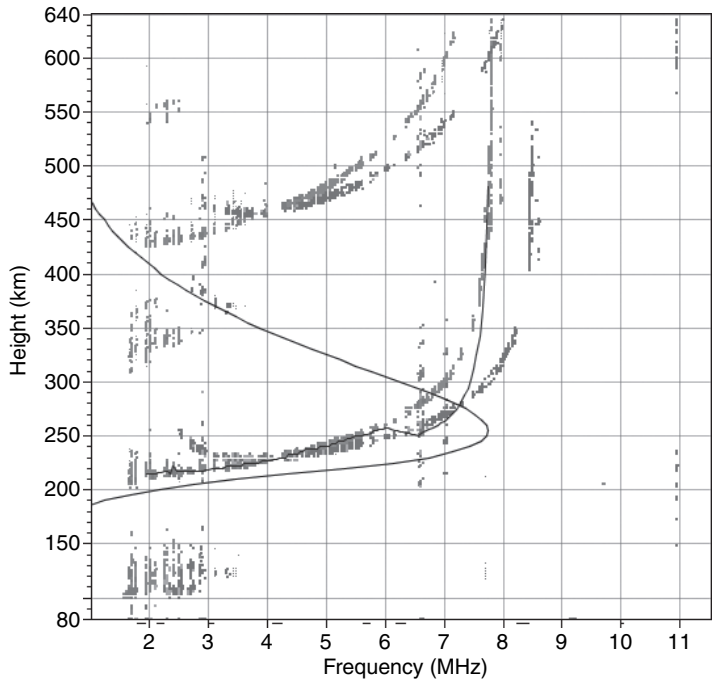

Fig. 1. Example of a digital ionogram recorded at Juliusruh ionosonde (•, returned frequencies and their virtual heights): the frequency at which the virtual height tends to ∞ is called the electron critical frequency; the 'bell-shaped' curve is the restored profile of the ionosphere (see Fig. 2) and the rightmost point of this curve is the electron critical frequency

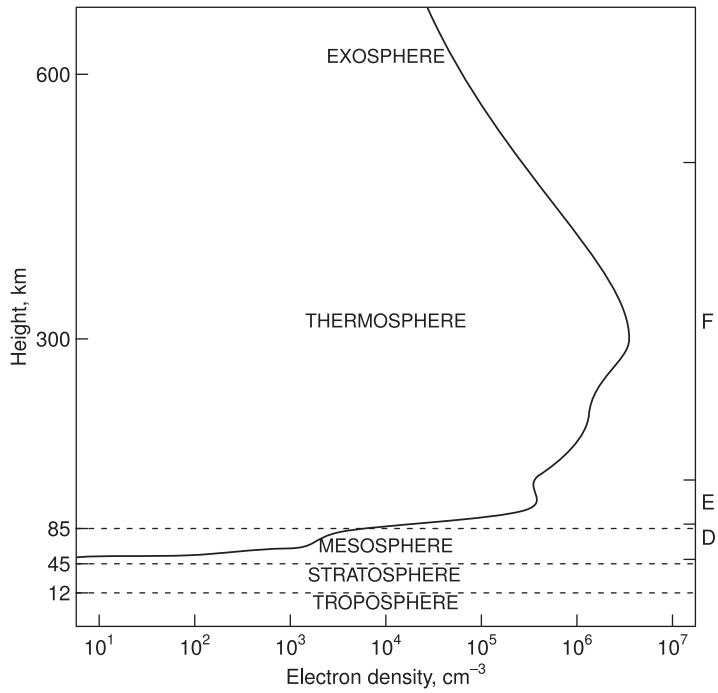

Fig. 2. Typical profile of the ionosphere: electron density as a function of height; the right-hand vertical axis indicates the D, E and F regions of the ionosphere

To describe the specific problem that is studied in this paper, we must provide some broader background. The increased concentration of greenhouse gases in the upper atmosphere has been associated with the global warming in the lower troposphere. Roble and Dickinson (1989) suggested that the increasing amounts of these radiatively active gases would lead to global cooling in the mesosphere and thermosphere. Shortly afterwards, Rishbeth (1990) pointed out that this would result in a thermal contraction and the global lowering of the ionospheric peak densities. The peak density height of the F2 region can be approximately computed by using the critical frequency foF2. Thus, if the hypothesis of Roble and Dickinson (1989) were true, cooling of the ionosphere would result in a systematic global change in foF2 which, in space physics research, is referred to as a global foF2 trend. It initially appeared that the evaluation of such a trend in the ionosphere might be easier than the evaluation of a global warming trend in the near surface atmosphere which exhibits a very strong local variability, making a definition of a global warming trend much less obvious. Consequently, during the last two decades, trends in foF2 (defined in several ways) have been extensively studied by many researchers; Lastovicka *et al.* (2006) and Lastovicka (2009) offer reviews of the relevant literature. In spite of this great interest and extensive research, the nature of the trends in the upper ionosphere is not yet fully understood, and there is no universal agreement on the existence of a global trend. A central issue is that trends over some regions appear to be upward, and over other regions to be downward. These trend estimates have been obtained by regression methods that are not fully justified and validated for the spatially distributed time series data, but they prompted explanations that are different from the global contraction hypothesis of Rishbeth (1990). In particular, Danilov and Mikhailov (1999) and Mikhailov and Marin (2000) argued that the trends in the upper ionosphere can be related to changes in geomagnetic activity. A different explanation is that the trends could

be due to long-term changes in the internal magnetic field of the Earth; see Foppiano *et al.* (1999). Recently Elias (2009) pointed out that it is more likely that observed trends in foF2 are a combined result of the influence of several factors. The precise understanding of the influence of different factors is needed to make a reliable conclusion about the origin and significance of the observed trends but, because of the extremely complex nature of the Earth's ionosphere, separation of the influence of different factors from a physical perspective is almost impossible. For this reason, a reliable statistical analysis is needed. In particular, it is important to determine whether the trends over two regions are indeed different, or if the apparent differences are spurious and stem from using regression methods that are not suitable for the foF2-data. The present paper makes a contribution in this direction by proposing a new methodology.

Dependence makes the study of spatially indexed functional data different from the more common analysis of functional objects. Most methodological and theoretical developments in functional data analysis have been motivated by data obtained from designed experiments in which functional observations on subjects can be treated as independent. In this paper, we carefully take into account the spatial dependence between the curves to develop a significance test for testing whether the mean curves over two disjoint regions are different. We apply the new test to establish whether there is a difference in the mean functions of foF2, due to the Earth's magnetic field. We analyse foF2-data recorded over eastern and western Europe. The inclination of the Earth's magnetic field is positive in the eastern part and negative in the western part of Europe. We use this property to separate Europe into two regions.

Whereas there is extensive literature on comparisons of spatial fields in different settings such as medical imaging and forecasts (see for example Hering and Genton (2011) and Gilleland *et al.* (2009) and references therein), in functional data analysis spatial dependence is usually ignored. A test for the equality of the mean functions in two independent samples of independent curves is proposed in Horváth and Kokoszka (2012). It is extended to time series data in Horváth *et al.* (2011). These are asymptotic tests whose test statistic is a quadratic form constructed from estimated variance operators (long-run variances for time series data). The test statistic is asymptotically χ^2 distributed under the null hypothesis of equal mean functions. An approach of this type is in principle possible for spatially correlated curves, but it gives very poor results in small samples. We therefore pursue a different approach. Very small sample sizes combined with spatial dependence are two main concerns that we must tackle.

The paper is organized as follows. In Section 2, we provide a detailed description of the foF2-data. Section 3 introduces the notation required, formalizes the problem and proposes a statistical model for the data. In Section 4, we introduce an iterative method for the estimation of mean functions and of the covariance matrix. With these preliminaries, we construct in Section 5 the test statistic and study its properties. This allows us to apply the test to the foF2-data in Section 6 and to arrive at well-supported conclusions. Appendix A contains an algorithmic description of one of the estimation procedures that are used in the paper.

2. Description of the ionosonde data

The foF2-data are collected by a global network of ionosonde stations which consist of over 200 observatories, but only a few dozen have operated for sufficiently long periods of time to provide data that are useful to study long-term ionospheric trends. The complete raw records are stored at the National Oceanic and Atmospheric Administration Web site `http://spidr.ngdc.noaa.gov/spidr/`. The quality of the raw data from most stations is poor. The main problem is that, even if the measurements are listed as equidistant, in reality they are not. Some absent data are flagged as missing, and some are not. The gaps range from several hours to

Table 1. Summary information for the ionosonde stations

Code	*Location*	*Latitude (deg)*	*Longitude (deg)*	*Magnetic inclination*
AZ136	Arkhangelsk	64.4	40.5	+
DB049	Dourbes	50.1	4.6	−
JR055	Juliusruh	54.6	13.4	+
KI167	Kiruna	67.8	20.4	+
KL154	Kaliningrad	54.7	20.6	+
LN047	Lannion	48.8	−3.4	−
LY164	Lycksele	64.7	18.8	+
MO155	Moscow	55.5	37.3	+
PQ052	Pruhonice	50.0	14.6	+
PT046	Poitiers	46.6	0.3	−
SL051	Slough	51.5	−0.6	−
UP158	Uppsala	59.8	17.6	+
US057	South Uist	57.4	−7.3	−

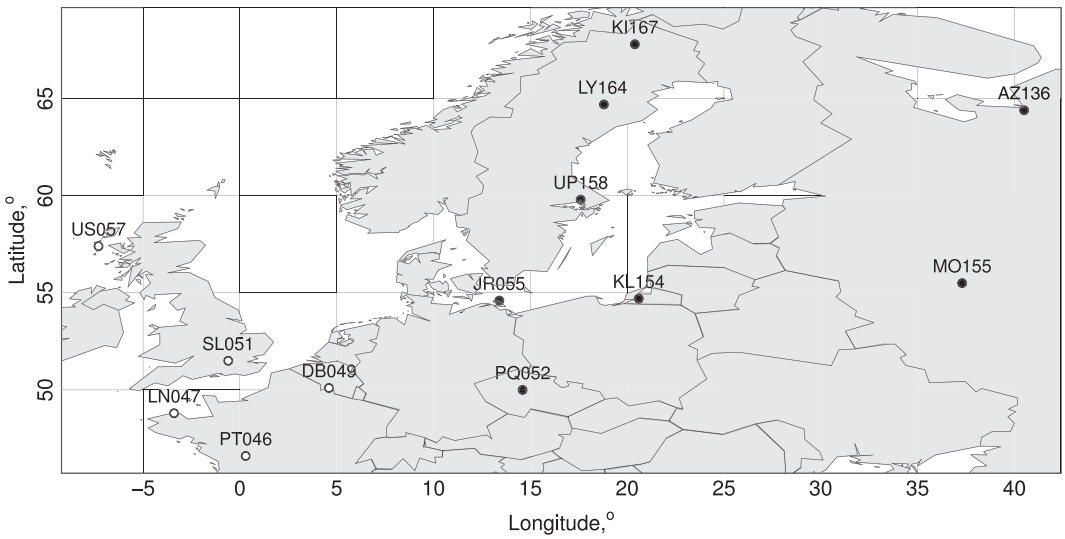

Fig. 3. Locations of 13 selected ionosonde stations with the corresponding five-letter codes: ○, stations with negative magnetic inclination; ●, stations with positive magnetic inclination

several months and must be identified and filled. The second problem is that in December the foF2-signal is artificially increased 10 times. Because of the amount of observations (over 70 000 in 8 years per curve (station)), cleaning must be done algorithmically. We developed a customized C++ code which cleans the foF2-data and performs interpolation; the code is available on request. We suppose that the absence of sufficiently long records of clean data at many locations has been a practical major obstacle in studying the global foF2-trend. We hope that our code will be useful to the space physics community.

We selected 13 ionosonde stations with hourly records of foF2 starting from January 1972 and ending in December 1980. All stations are in Europe. The main criterion used for ionosonde station selection was the lack of long periods of missing observations. The ionosonde stations selected are listed in Table 1. For ease of reference, we also provide a map of Europe with the

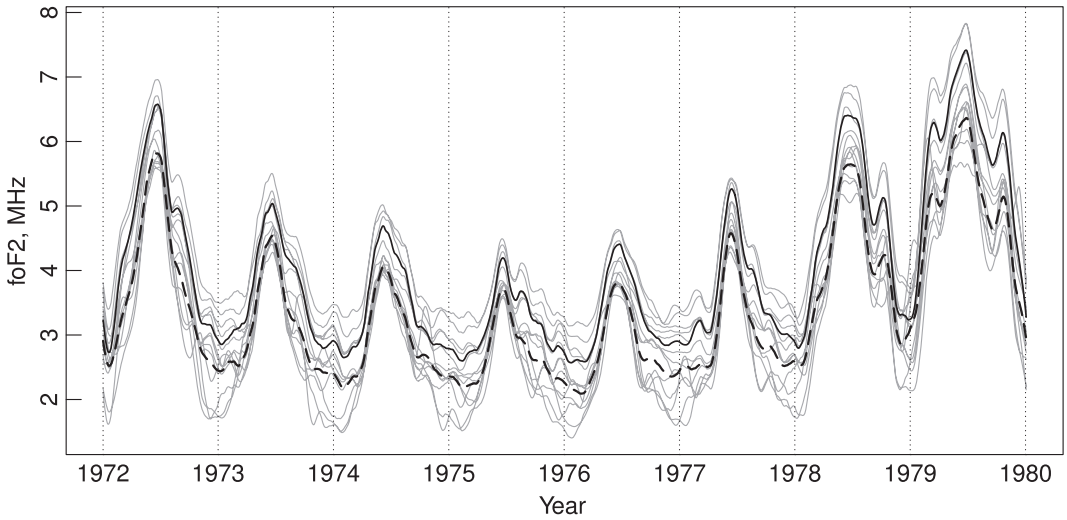

Fig. 4. Ionosonde data (——): mean function for stations with negative inclination (——) and mean function for stations with the positive inclination (– – –)

locations of the ionosonde stations; Fig. 3. The time interval was selected to match two criteria: maximum data availability and coverage of a solar minimum period. The second criterion is needed to reduce latitudinal variability of the foF2-signal due to the solar activity. For further reduction of the effects due to the solar activity, we calculated nightly sample averages of data recorded between 22 and 2 local (longitude) time at each day. All such averages are calculated on the basis of 3–5 consecutive temporal observations. This results in one data point per day. Finally, we used the maximum overlap discrete wavelet transform to smooth the noisy curves; see chapter 5 of Percival and Walden (2000). We used filter length roughly corresponding to averaging over 32 days. The data curves thus prepared and transformed are shown in Fig. 4, which also shows the estimated mean functions for the negative and positive inclination samples. We call such smooth nightly averages ionosonde data. The question that we answer in this paper is whether the two mean curves (for positive and negative inclination), whose estimates are shown in Fig. 4, are significantly different.

3. A functional spatiotemporal model for the ionosonde data

This section introduces the requisite Hilbert space framework within which the testing problem can be clearly stated. Following the usual approach that is adopted in functional data analysis, we assume that the smoothed foF2-curves belong to the Hilbert space of square integrable functions on the interval $[0, 1]$, which is the rescaled time interval from January 1st, 1972, to December 31st, 1980. This space is denoted $L^2 = L^2([0, 1])$ and is equipped with the inner product $\langle f, g\rangle = \int f(t)\, g(t)\, \mathrm{d}t$ and the norm $\|f\|^2 = \int f^2(t)\, \mathrm{d}t$. The curve at location $\mathbf{s}$ is denoted $X(\mathbf{s})$, and the value at time $t \in [0, 1]$ is denoted by $X(\mathbf{s}; t)$. We treat the function $X(\mathbf{s})$ as a random element of the space L^2.

We consider a spatial domain $\mathcal{D}$ which is separated into two disjoint regions $\mathcal{D}_1$ and $\mathcal{D}_2$; $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$. Let $\mathbf{s}^{(1)} \in \mathcal{D}_1$ and $\mathbf{s}^{(2)} \in \mathcal{D}_2$ denote the generic locations in these two regions, and $\mathbf{s}_k^{(1)}$ and $\mathbf{s}_l^{(2)}$ denote the spatial locations in regions $\mathcal{D}_1$ and $\mathcal{D}_2$ respectively, at which data curves are available. The number of observations in region $\mathcal{D}_j$ is N_j. We reserve indices $j, j' = 1, 2$ to specify the region.

We assume that the random functions in the two regions may differ only in the mean function, so the spatiotemporal model for the data is

$$X(\mathbf{s}^{(j)};t)=\mu_j(t)+\varepsilon(\mathbf{s}^{(j)};t), \qquad j=1,2, \tag{1}$$

where $\varepsilon(\cdot)$ is a strictly stationary zero-mean random field in L^2. To define the covariance function, we must assume that

$$E[\|\varepsilon(s)\|^2]=E\left[\int \varepsilon^2(\mathbf{s};t)\,\mathrm{d}t\right]<\infty.$$

It follows that the expectations $E[\langle\varepsilon(\mathbf{s}),\varepsilon(\mathbf{s}')\rangle]$ exist and depend only on the distance between $\mathbf{s}$ and $\mathbf{s}'$, and the distribution of $\varepsilon(\mathbf{s})$ does not depend on $\mathbf{s}$.

In this paper, the distance $d_{k,l}$ between two points $\mathbf{s}_k$ and $\mathbf{s}_l$ is a chordal distance defined as

$$d_{k,l}=2\left\{\sin^2\left(\frac{L_k-L_l}{2}\right)+\cos(L_k)\cos(L_l)\sin^2\left(\frac{l_k-l_l}{2}\right)\right\}^{1/2}, \tag{2}$$

where L denotes the latitude and l the longitude. 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.

In this framework, the testing problem can be formulated as

$$\begin{aligned} H_0&:\mu_1(t)=\mu_2(t) && (\|\mu_1-\mu_2\|^2=0);\\ H_{\mathrm{A}}&:\mu_1(t)\neq\mu_2(t) && (\|\mu_1-\mu_2\|^2>0). \end{aligned}$$

In the remainder of this section we tighten our assumptions in a way that makes a construction of a test possible. The functional error field admits the Karhunen–Loève expansion

$$\varepsilon(\mathbf{s};t)=\sum_{i=1}^{\infty}\xi_i(\mathbf{s})\,v_i(t),$$

in which the functional principal components (FPCs) v_i are unknown L^2-valued parameters. Later in the paper we truncate the infinite sum by taking into account the first p summands which capture the desired level of the variance. Let $\boldsymbol{\xi}_i$ be a column vector comprised of the scores $\xi_i(\mathbf{s}_k)$, $\boldsymbol{\xi}_i=(\xi_i(\mathbf{s}_1^{(1)}),\ldots,\xi_i(\mathbf{s}_{N_1}^{(1)}),\xi_i(\mathbf{s}_1^{(2)}),\ldots,\xi_i(\mathbf{s}_{N_2}^{(2)}))^{\mathrm{T}}$. Below we use two assumptions: the vector fields $\boldsymbol{\xi}_i$ are normal, $\boldsymbol{\xi}_i\sim N(\mathbf{0},\boldsymbol{\Gamma}_i)$, where $\boldsymbol{\Gamma}_i=\mathrm{var}(\boldsymbol{\xi}_i)$, and the fields $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_{i'}$ are independent if $i'\neq i$.

The assumption of normality can be verified empirically. The QQ-plots in Fig. 5 do not contradict it in any obvious way. The assumption of the independence of score functions is required to construct a computable test statistic and is needed for most inferential procedures, Delaigle and Hall (2010) discussed this point. The scores at the same location are always uncorrelated, i.e. $E[\xi_i(\mathbf{s})\,\xi_{i'}(\mathbf{s})]=0$, if $i'\neq i$. Our assumptions imply that $E[\xi_i(\mathbf{s})\,\xi_{i'}(\mathbf{s}')]=0$, for arbitrary locations $\mathbf{s}$ and $\mathbf{s}'$.

The above development shows that, to construct a useful test statistic, we must address the estimation of the mean function and of the FPCs in a spatial setting. We turn to these issues in Section 4.

4. Estimation of the mean and covariance functions

In this section, we propose an extension of the methodology for the estimation of the mean function and the FPCs that are introduced by Gromenko *et al.* (2012). There are several new elements: the modifications required to deal with two samples, with possibly different means, the

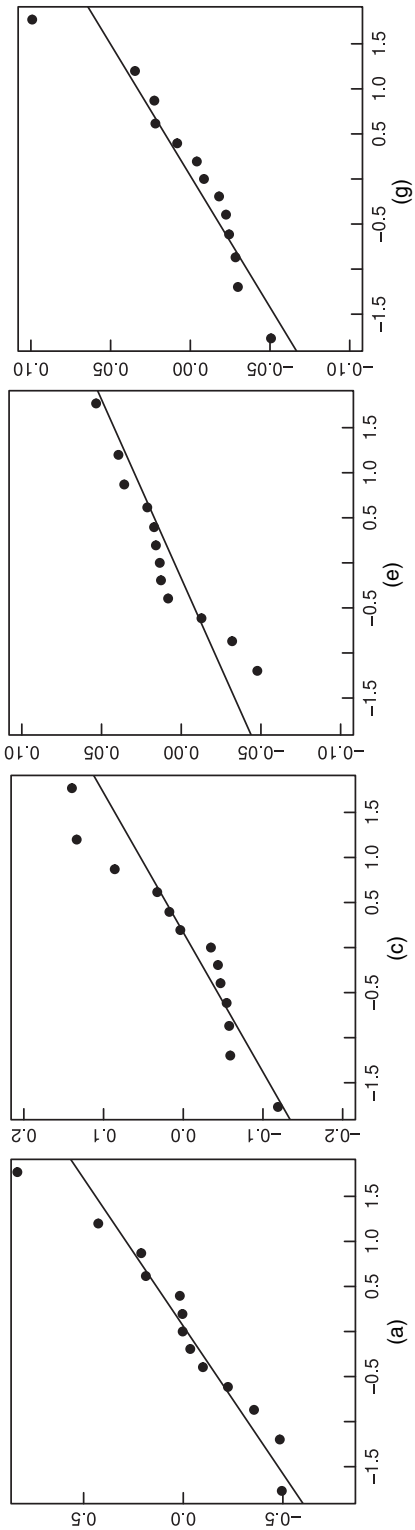

Fig. 5. Normal *QQ*-plots for the estimated scores (a) ξ_1, (b) ξ_2, (c) ξ_3, (d) ξ_4, (e) ξ_5, (f) ξ_6 and (g) ξ_7

introduction of an iterative estimation process and the choice of the bin size in variogram estimation.

Recall that $\mathbf{s}_k^{(j)}$ refers to a location in region $j = 1, 2$. The mean functions μ_j are estimated by weighted sums of observations:

$$\hat{\mu}_j(t) = \sum_{k=1}^{N_j} w_k^{(j)} X(\mathbf{s}_k^{(j)}; t), \qquad \sum_{k=1}^{N_j} w_k^{(j)} = 1, \qquad j = 1, 2. \tag{3}$$

To find the weights $w_k^{(j)}$, it is convenient to use matrix notation. We introduce the following vectors and matrices:

$$\mathbf{1}_1 = \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0}_{N_2} \end{pmatrix}, \qquad \mathbf{1}_2 = \begin{pmatrix} \mathbf{0}_{N_1} \\ \mathbf{1}_{N_2} \end{pmatrix}, \qquad \mathbf{I}_1 = \begin{pmatrix} I_{N_1 \times N_1} & \mathbf{0}_{N_1 \times N_2} \\ \mathbf{0}_{N_2 \times N_1} & \mathbf{0}_{N_2 \times N_2} \end{pmatrix}, \qquad \mathbf{I}_2 = \begin{pmatrix} \mathbf{0}_{N_1 \times N_1} & \mathbf{0}_{N_1 \times N_2} \\ \mathbf{0}_{N_2 \times N_1} & I_{N_2 \times N_2} \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{w}_1 &= (w_1^{(1)}, \ldots, w_{N_1}^{(1)}, 0, \ldots, 0), \\ \mathbf{w}_2 &= (0, \ldots, 0, w_1^{(2)}, \ldots, w_{N_2}^{(2)}), \\ \mathbf{X}(t) &= (X(\mathbf{s}_1^{(1)}; t), \ldots, X(\mathbf{s}_{N_2}^{(2)}; t))^{\mathrm{T}}. \end{aligned}$$

Then, the estimates are given by

$$\hat{\mu}_j(t) = \mathbf{w}_j \mathbf{X}(t), \qquad \mathbf{w}^{(j)} \mathbf{1}_j = 1, \qquad j = 1, 2.$$

To give closed form expressions for the weights which minimize the constrained least-squared errors, we define the covariance matrix $\mathbf{C}$, which has the block form

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}.$$

The elements of the submatrices $\mathbf{C}_{jj'}$ are given by

$$\begin{aligned} c_{jj'}(k, l) &= E[\langle X(\mathbf{s}_k^{(j)}) - \mu_j, X(\mathbf{s}_l^{(j')}) - \mu_{j'} \rangle] \\ &= \sum_{i=1}^{p} E[\xi_i(\mathbf{s}_k^{(j)}) \xi_i(\mathbf{s}_l^{(j')})], \qquad 1 \leqslant k \leqslant N_j, \quad 1 \leqslant l \leqslant N_{j'}. \end{aligned} \tag{4}$$

(The estimation of the $c_{jj'}(k, l)$ is discussed at the end of this section.) Using the method of Lagrange multipliers, we find the optimal weights:

$$\begin{aligned} \mathbf{w}_1 &= (\mathbf{1}_1^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{1}_1)^{-1} \mathbf{1}_1^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{I}_1, \\ \mathbf{w}_2 &= (\mathbf{1}_2^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{1}_2)^{-1} \mathbf{1}_2^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{I}_2. \end{aligned} \tag{5}$$

The covariance matrix $\mathbf{C}$ requires the estimation of the FPCs, $v_i(t)$, which, in turn, requires the estimation of the mean functions $\mu_j(t)$ and centring the data. Thus, the following iterative approach is used.

(a) Compute the simple mean (average) for each sample and subtract it from the curves in each sample.
(b) Estimate the FPCs and their number required to capture the desired level of variability by using method CM3 proposed in Gromenko *et al.* (2012) and calculate the scores. We review method CM3 in Appendix A.
(c) Estimate the covariance matrix $\mathbf{C}$ by using equation (4), find the weights by using equations (5) and estimate the mean functions by using the weighted sums (3).

(d) Subtract the mean functions from the curves in each sample and repeat steps (b)–(d) until a suitably defined convergence of the mean function estimates is reached. We discuss the convergence in Section 6.

To calculate the expectations $E[\xi_i(\mathbf{s}_k^{(j)})\xi_i(\mathbf{s}_l^{(j')})]$ appearing in the definition of the matrix $\mathbf{C}$, we use parametric modelling, which is common in spatial statistics:

$$E[\xi_i(\mathbf{s}_k^{(j)})\xi_i(\mathbf{s}_l^{(j')})] = \mathrm{cov}\{\xi_i(\mathbf{s}_k^{(j)}), \xi_i(\mathbf{s}_l^{(j')})\} = f_i(d_{k,l}, \sigma^2, \rho, \dots),$$

where $d_{k,l}$ is a chordal distance between $\mathbf{s}_k$ and $\mathbf{s}_l$ that is defined in Section 3. The selected parametric models for scores as well as estimated parameters are summarized in Section 6. They are estimated by using weighted least square fitting of a robust variogram estimator of Hawkins and Cressie (1984). We emphasize that there is no guarantee that the iterative algorithm converges for an arbitrary data set.

When the sample size is small, estimation of the empirical variogram can be a challenge. In fact, improper selection of the bin parameter could potentially cause significant overestimation of the sill and the range, which may affect the conclusion of the test. To achieve stability of the estimator of the empirical variogram the bin parameter should be selected in such a way that each lag interval contains at least 30 distinct distances; see, for example, section 2.4 in Cressie (1993). In the case of small sample size this recommendation is difficult to fulfil; thus, visual validation both for variogram estimation and its fitting is essential. In our study, we use two criteria for validation. The first criterion is that the empirical variogram should have a flat-shaped region for large lags. The second criterion is that the estimated range should be less than the characteristic size of the spatial domain. We also recommend the use of a simple estimator of the variance and FPCs to approximate the magnitude order of the sill. Usually, the maximum lag is taken to be half of the characteristic size of the spatial domain. Finally, we draw attention to the fact that overestimation of the sill can cause under-rejection of the null hypothesis which in practice is more preferable than over-rejection. We discuss bin size selection for the ionosonde data in Section 6.

5. Test procedure

To test the null hypothesis that was formulated in Section 3, we use the statistic

$$\hat{S} = \|\hat{\mu}_1 - \hat{\mu}_2\|^2 = \int \{\hat{\mu}_1(t) - \hat{\mu}_2(t)\}^2 \mathrm{d}t, \tag{6}$$

where the mean functions $\hat{\mu}_1$ and $\hat{\mu}_2$ are estimated by using the iterative procedure that was described in Section 4. A natural approach for the small sample is to use a Monte Carlo (MC) test based on equation (6), which is feasible under the assumptions stated in Section 3 (normality and the independence of score processes). We describe this approach first. Then we introduce tests based on a gamma approximation to the distribution of statistic (6) under hypothesis H_0.

Both the MC test and the gamma test use the expansion under H_0

$$\hat{S}_0 = \sum_{i=1}^{p} \{(\mathbf{w}_1 - \mathbf{w}_2)\hat{\boldsymbol{\xi}}_i\}^2 \tag{7}$$

in which the $\hat{\boldsymbol{\xi}}_i$ are the scores estimated as explained in Section 4. The weights $\mathbf{w}_j$ are also estimated but are treated as deterministic in both procedures. This is because the variability of the mean function estimates is quantified by the variability of the scores.

5.1. Monte Carlo approximation

The MC method relies on the following procedure.

(a) Estimate the mean functions μ_j, the weights $w_k^{(j)}$ and the FPCs v_i as described in Section 4 and Appendix A.
(b) Calculate the test statistics $\hat{S}$ by using equation (6).
(c) Generate scores $\tilde{\boldsymbol{\xi}}_i \sim N(\mathbf{0}, \boldsymbol{\Gamma}_i)$, where $\boldsymbol{\Gamma}_i$ is the estimated covariance matrix. (Recall that the variances $\boldsymbol{\Gamma}_i = E[\boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\mathrm{T}}]$ are estimated by using a parametric model.)
(d) Calculate the MC statistic $\tilde{S}$ by plugging the weights $w_k^{(j)}$ and the generated scores $\tilde{\boldsymbol{\xi}}_i$ into equation (7).
(e) Compare the MC statistic and the observed test statistics.
(f) Repeat steps (c)–(e) M times to reach a required precision. (We use $M = 10^6$, which provide precision of three orders of magnitude.)
(g) The ratio of the MC statistics which are larger than the test statistics $\hat{S}$ and the number of iterations M is the MC P-value.

The MC procedure is time consuming. We now describe a procedure based on a gamma approximation of the test statistics. This procedure is much faster and has comparable finite sample properties in synthetic data sets generated to resemble the ionosonde data.

5.2. Gamma approximation

The test statistics (7) can be written as

$$\hat{S}_0 = \sum_{k=1}^{p} \{(\mathbf{w}_1 - \mathbf{w}_2)\hat{\boldsymbol{\xi}}_k\}^2 = \sum_{k=1}^{p} \hat{\sigma}_k^2 z_k^2, \tag{8}$$

where the z_i are standard normal and $\hat{\sigma}_i^2 = (\mathbf{w}_1 - \mathbf{w}_2)\hat{\boldsymbol{\Gamma}}_i(\mathbf{w}_1 - \mathbf{w}_2)^{\mathrm{T}}$. It is not difficult to verify that $\hat{\sigma}_k^2 = O(N^{-1})$, and so the characteristic function of $\hat{S}$ admits the approximation

$$\prod_{k=1}^{p} (1 - 2i\sigma_k^2 t)^{-1/2} \approx \left(1 - 2i \sum_{k=1}^{p} \sigma_k^2 t\right)^{-1/2}. \tag{9}$$

The right-hand side of approximation (9) is the characteristic function of the gamma density

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \qquad x > 0,$$

with

$$\begin{aligned} \alpha &= \tfrac{1}{2}, \\ \beta &= 2 \sum_{k=1}^{p} \sigma_k^2. \end{aligned} \tag{10}$$

Instead of using expression (10), a better approximation is obtained by using exact values given by

$$\begin{aligned} \alpha &= E[\hat{S}_0]^2 / \mathrm{var}(\hat{S}_0), \\ \beta &= \mathrm{var}(\hat{S}_0) / E[\hat{S}_0]. \end{aligned} \tag{11}$$

This requires estimates of $E[\hat{S}]$ and $\mathrm{var}(\hat{S})$ that are valid under hypothesis H_0. Treating the weights $w_k^{(j)}$ as deterministic and using equation (8), we obtain

$$E[\hat{S}_0] = \sum_{i=1}^{p} \hat{\sigma}_i^2 = (\mathbf{w}_1 - \mathbf{w}_2)\hat{\mathbf{C}}(\mathbf{w}_1 - \mathbf{w}_2)^{\mathrm{T}} \tag{12}$$

and

$$E[\hat{S}_0^2] = 3\sum_{i=1}^{p}\hat{\sigma}_i^4 + 2\sum_{i>i'}^{p}\hat{\sigma}_i^2\hat{\sigma}_{i'}^2,$$

$$E[\hat{S}_0]^2 = \sum_{i=1}^{p}\hat{\sigma}_i^4 + 2\sum_{i>i'}^{p}\hat{\sigma}_i^2\hat{\sigma}_{i'}^2,$$

so that

$$\mathrm{var}(\hat{S}_0) = E[\hat{S}_0^2] - E[\hat{S}_0]^2 = 2\sum_{i=1}^{p}\hat{\sigma}_i^4. \tag{13}$$

For the ionosonde data the difference between the 'approximate' and the 'exact' values given by expressions (10) and (11) respectively is about 20%, which is the main cause of the difference between P-values in Table 3 in Section 6.

We can now summarize the testing procedure based on the gamma approximation.

(a) Estimate the mean functions μ_j, the weights $w_k^{(j)}$ and the FPCs v_i as described in Section 4.
(b) Compute the test statistics $\hat{S}$ by using equation (6).
(c) Estimate the variances $\mathbf{\Gamma}_i = E[\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}]$ by using a spatial parametric model for each $1 \leqslant i \leqslant p$.
(d) Estimate the parameters α and β by using either expression (10) or expression (11).
(e) Compute the P-value:

$$P\text{-value} = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_{\hat{S}}^{\infty} x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)\mathrm{d}x.$$

6. Application to the ionosonde data

In this section we describe the application of the methodology that we propose to the ionosonde data. We begin with the details of implementation.

We estimated the variograms by using the robust estimator of Hawkins and Cressie (1984) and fitted parametric spatial models by using weighted non-linear least squares. As explained at the end of Section 4, the bin parameter needs to be selected with great care to obtain a reasonable estimate of the spatial parametric model. The number of distinct distances is $C_2^{13} = 78$ and the maximum lag is about 20° or 0.3 rad; see Fig. 3. By letting each lag contain $N(h_i) = 20$ distinct distances we conclude that the bin parameter should be 5° or 0.06 rad. Because $N(h_i) < 30$ visual validation is needed in each step of the iterative algorithm.

Visual examination revealed that for some score processes $\boldsymbol{\xi}_i$ the exponential model with zero nugget offers the best fit. Under this model, the covariances are given by

$$\mathrm{cov}\{\xi_i(\mathbf{s}_k^{(j)}), \xi_i(\mathbf{s}_l^{(j')})\} = \sigma_i^2\exp(-d_{k,l}/\rho_i).$$

The estimated sills σ_i^2 and ranges ρ_i are summarized in Table 2. 'Simple' stands for the ordinary variance estimation.

Using the iterative approach that was introduced in Section 4, we estimate the mean functions for the two regions that were described in Section 2. The convergence is evaluated by means of the quantity

$$R_i = \int_0^1 |\mu^{(i+1)}(t) - \mu^{(i)}(t)|\,\mathrm{d}t,$$

where the index i denotes the number of iterations. For the ionosonde data, the convergence is reached after four iterations, as shown in Fig. 6.

Table 2. Estimates for the parametric covariance models fitted to the ionosonde data

Spatial field	*Model*	σ^2	ρ
ξ_1	Exponential	1.57×10^{-1}	0.056
ξ_2	Simple	1.72×10^{-2}	—
ξ_3	Exponential	6.56×10^{-3}	0.036
ξ_4	Simple	3.92×10^{-3}	—
ξ_5	Exponential	2.17×10^{-3}	0.028
ξ_6	Exponential	2.76×10^{-3}	0.025
ξ_7	Simple	1.49×10^{-3}	—

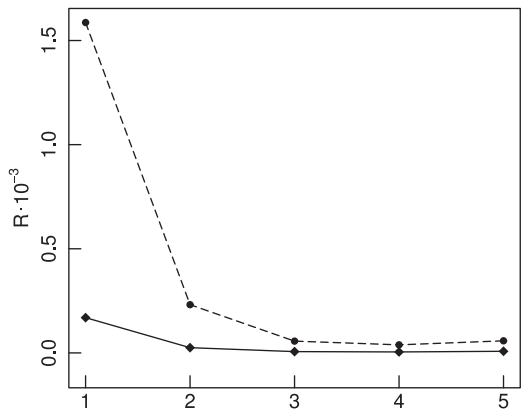

Fig. 6. Convergence of the iterative method for the estimation of the means R_i as a function of iteration i: – – –, western; ———, eastern

These estimates allow us to construct the following decomposition of variance:

$$\sum_{k=1}^{N_1} \|X(\mathbf{s}_k^{(1)}) - \hat{\mu}_1\|^2 + \sum_{l=1}^{N_2} \|X(\mathbf{s}_l^{(2)}) - \hat{\mu}_2\|^2 = \sum_{i=1}^{N_1+N_2} \left\{ \sum_{k=1}^{N_1} \hat{\xi}_i^2(\mathbf{s}_k^{(1)}) + \sum_{l=1}^{N_2} \hat{\xi}_i^2(\mathbf{s}_l^{(2)}) \right\}.$$

The sum on the right-hand side is replaced by the sum

$$V(p) := \sum_{i=1}^{p} \left\{ \sum_{k=1}^{N_1} \hat{\xi}_i^2(\mathbf{s}_k^{(1)}) + \sum_{l=1}^{N_2} \hat{\xi}_i^2(\mathbf{s}_l^{(2)}) \right\},$$

with p so large that $V(p)$ exceeds a specified percentage of $V(N_1 + N_2)$. This procedure is fairly standard for independent functional data. For spatially indexed functions, the estimated score processes $\hat{\boldsymbol{\xi}}_i$ take into account the spatial correlations.

In some cases, the results of a test for functional data depend on the selection of the cut-off value p, with the usual recommendation being to use the value of p which explains between 80% and 90% of the variance. For the ionosonde data, $p = 1$ already explains over 75% of the variance, and the results of the test, for both MC and gamma implementations, do not significantly depend on p. Table 3 shows the P-values for the MC method, gamma method with the exact parameters α and β calculated by using expression (11) ('Exact'), gamma method which uses

Table 3. *P*-values for various numbers of the FPCs p, $1 \leqslant p \leqslant 7$, for night and day data

Target cumulative variance (%)	*Results for spatial model*					*Results for simple model*		
	Final p	*Final cumulative variance (%)*	*MC*	*Exact*	*Approximate*	*Final p*	*Final cumulative variance (%)*	*Exact*
75	1	75.45	0.023	0.0244	0.0244	—	—	—
80	2	85.14	0.024	0.0250	0.0296	1	82.49	1.42×10^{-3}
90	4	91.81	0.024	0.0254	0.0334	2	91.53	1.23×10^{-3}
95	7	95.17	0.025	0.0257	0.0357	3	95.32	1.15×10^{-3}

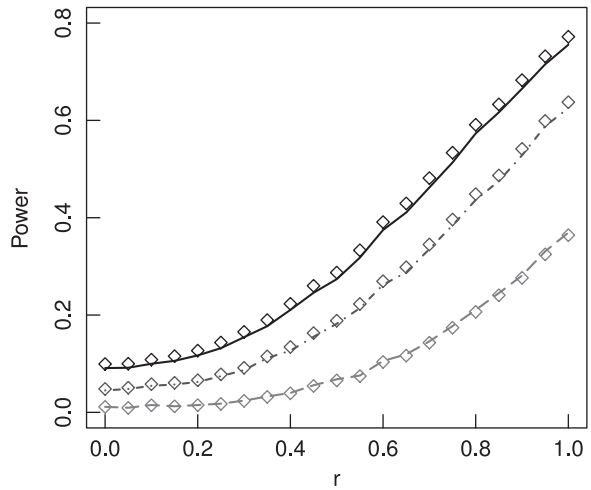

Fig. 7. Power function of the test: 'exact' gamma test for confidence levels α 10% (———), 5% (· – · –) and 1% (— —) (◇, MC test)

expression (10) ('Approximate') and the *P*-values obtained by ignoring the spatial dependence ('Simple'). In the last case we used the estimation procedure and the gamma method which neglect spatial dependence. We conclude that there is enough evidence to support the alternative hypothesis. For the ionosonde data the conclusion of our analysis is that there is statistically significant evidence to claim that the mean function over Europe changes owing to magnetic inclination. This conclusion agrees with the findings of Cnossen and Richmond (2008) who used a physical ionospheric model.

The remainder of this section is devoted to the statistical validation of the conclusions that were presented above. Owing to very small sample sizes, the only feasible way of assessing the final sample performance of the tests is through a simulation study. The results of any such study depend on the model for the data-generating process. We have taken great care to ensure that the stochastic structure of the simulated data resembles that of the real data. The key assumptions that are made to simulate the data are the normality and the independence of the score processes $\boldsymbol{\xi}_i$.

The data-generating process is given by

Table 4. Empirical size of the tests applied to data generated to resemble the ionosonde data

p	*Results (%) for spatial model*									*Results (%) for simple model*		
	MC			*Exact*			*Approximate*			*Exact*		
	10%	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*
1	9.96	4.89	1.02	9.57	4.66	0.93	9.57	4.66	0.93	20.17	11.12	1.66
2	9.94	5.05	1.01	9.34	4.70	1.01	8.89	4.20	0.76	19.51	10.91	1.86
3	10.03	4.97	1.00	9.29	4.76	0.98	8.60	4.02	0.64	19.51	11.14	1.99
4	9.96	4.96	1.02	9.25	4.77	1.02	8.51	3.91	0.62	19.46	11.09	2.01
5	10.04	5.02	1.01	9.23	4.69	1.02	8.34	3.80	0.61	19.41	11.07	2.08
6	9.97	4.86	0.92	9.23	4.65	0.97	8.26	3.67	0.54	19.31	11.03	2.02
7	9.88	4.95	0.93	9.12	4.63	0.96	8.15	3.60	0.53	19.17	10.99	2.04

$$\begin{aligned} X(\mathbf{s}^{(1)};t) &= m_1(t) + \sum_{i=1}^{p} \xi_i(\mathbf{s}^{(1)}) v_i(t), \\ X(\mathbf{s}^{(2)};t) &= m_2(t) + \sum_{i=1}^{p} \xi_i(\mathbf{s}^{(2)}) v_i(t). \end{aligned} \tag{14}$$

To evaluate the empirical size, we set $m_1(t) = m_2(t) = 0$. To evaluate the empirical power, we set $m_1(t) = \mu_1(t)$, the estimated mean function for the western region, and $m_2 = (1-r)\mu_1 + r\mu_2$, where μ_2 is the estimated mean function for the eastern region. The power is then a function of the 'separation' $0 \leqslant r \leqslant 1$. The FPCs v_i are those estimated from the real data. The scores are generated from the zero-mean normal distribution $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{\Gamma}_i)$, using the Cholesky decomposition. For the empirical size we select $1 \leqslant p \leqslant 7$ and for the power $p = 7$. The number of replications of process (14) is 10^5. The results for the empirical size and the power are summarized in Table 4 and Fig. 7 respectively. Starting with Table 4, we see that the MC test has more accurate empirical size, very close to the nominal size. The gamma test (exact and approximate) is slightly too conservative. The exact method without taking into account spatial correlation has inflated size. This agrees with the very small P-values that were reported for real data and shows that conclusions based on this method must be treated as tentative.

Acknowledgements

We thank two referees and the Associate Editor for their insightful comments which helped to improve the quality of the paper. This research was partially supported by National Science Foundation grant DMS-0931948.

Appendix A: Estimation of the functional principal components

In this appendix we summarize a procedure for the estimation of the FPCs when the curves are spatially correlated. The method that we outline is called MC3 in Gromenko *et al.* (2012). It is easy to implement and has good size and power in finite samples.

(a) Assuming that $E[X(\mathbf{s})] = 0$ postulate the expansion

$$X(\mathbf{s}) \approx \sum_{m=1}^{K} \eta_m(\mathbf{s}) B_m,$$

where B_m is an orthonormal basis and $\eta_m(\mathbf{s})$ form an observable field $\eta_m(\mathbf{s}_k) = \langle B_m, X(\mathbf{s}_k)\rangle$. The value of K can be taken to be the number of basis functions that are used to create the functional objects in R, e.g. $K = 49$.

(b) Fix n and m, and define the scalar field z by $z(\mathbf{s}_k) = \eta_n(\mathbf{s}_k)\,\eta_m(\mathbf{s}_k)$.

(c) Estimate $\mu_z = E[z(\mathbf{s})]$ as a weighted average of the $z(\mathbf{s}_k)$ for all n and m and denote the resulting estimate by $\hat{r}_{nm}$.

(d) Find the solution of the matrix equation

$$\hat{\mathbf{R}}\mathbf{x} = \lambda\mathbf{x}, \tag{15}$$

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_K)^{\mathrm{T}},$$
$$\hat{\mathbf{R}} = (\hat{r}_{nm},\ 1 \leqslant n, m \leqslant K).$$

We denote the solutions to equation (15) by

$$\hat{\mathbf{x}}^{(n)} = (\hat{x}_1^{(n)}, \hat{x}_2^{(n)}, \ldots, \hat{x}_k^{(n)})^{\mathrm{T}}, \qquad \hat{\lambda}_j, \qquad 1 \leqslant n \leqslant K. \tag{16}$$

(e) The FPCs v_n, $1 \leqslant n \leqslant K$, are estimated by

$$\hat{v}_n = \sum_{\alpha=1}^{K} \hat{x}_\alpha^{(n)} B_\alpha. \tag{17}$$

Because the B_n are orthonormal the estimated FPCs are also orthonormal. The $\hat{\lambda}_n$ in expression (16) are estimators of the corresponding eigenvalues.

References

Aston, J. A. D. and Kirch, C. (2011) Estimation of the distribution of change-points with application to fMRI data. *Research Reports*. University of Warwick, Coventry.

Bosq, D. and Blanke, D. (2007) *Inference and Prediction in Large Dimensions*. Hoboken: Wiley.

Carroll, S. S. and Cressie, N. (1996) A comparison of geostatistical methodologies used to estimate snow water equivalent. *Wat. Resour. Bull.*, **32**, 267–278.

Carroll, S. S., Day, G. N., Cressie, N. and Carroll, T. R. (1995) Spatial modeling of snow water equivalent using airborne and ground-based snow data. *Environmetrics*, **6**, 127–139.

Cnossen, I. and Richmond, A. (2008) Modelling the effects of changes in the Earth's magnetic field from 1957 to 1997 on the ionospheric hmf2 and fof2 parameters. *J. Atmos. Sol. Terrest. Phys.*, **70**, 1512–1524.

Cressie, N. A. C. (1993) *Statistics for Spatial Data*. Hoboken: Wiley.

Danilov, A. and Mikhailov, A. (1999) Long-term trends in the parameters of the f2-region: a new approach. *Geomagn. Aeron.*, **39**, 473–479.

Delaigle, A. and Hall, P. (2010) Defining probability density function for a distribution of random functions. *Ann. Statist.*, **38**, 1171–1193.

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

Elias, A. (2009) Trends in the F2 ionospheric layer due to long-term variations in the Earth's magnetic field. *J. Atmos. Sol. Terrest. Phys.*, **71**, 1602–1609.

Ferraty, F. and Romain, Y. (eds) (2011) *The Oxford Handbook of Functional Data Analysis*. Oxford: Oxford University Press.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.

Foppiano, A., Cid, L. and Jara, V. (1999) Ionospheric long-term trends for South American mid-latitudes. *J. Atmos. Sol. Terrest. Phys.*, **61**, 717–723.

Gilleland, E., Ahijevych, D., Brown, B., Casati, B. and Ebert, E. (2009) Intercomparison of spatial forecast verification methods. *Weath. Forecast.*, **24**, 1416–1430.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Ann. Appl. Statist.*, to be published.

Hawkins, D. M. and Cressie, N. (1984) Robust kriging—a proposal. *J. Int. Ass. Math. Geol.*, **16**, 3–18.

Hering, A. and Genton, M. (2011) Comparing spatial predictions. *Technometrics*, **53**, 414–425.

Horváth, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. New York: Springer. To be published.

Horváth, L., Kokoszka, P. and Reeder, R. (2011) Estimation of the mean of functional time series and a two sample problem. *Technical Report*. University of Utah, Logan.

Kaiser, M. S., Daniels, M. J., Furakawa, K. and Dixon, P. (2002) Analysis of particulate matter air pollution using Markov random field models of spatial dependence. *Environmetrics*, **13**, 615–628.
Kelly, M. C. (2009) *The Earth's Ionosphere*. New York: Academic Press.
Lastovicka, J. (2009) Global pattern of trends in the upper atmosphere and ionosphere: recent progress. *J. Atmos. Sol. Terrest. Phys.*, **71**, 1514–1528.
Lastovicka, J., Mikhailov, A., Ulich, T., Bremer, J., Elias, A., Ortiz de Adler, N., Jara, V., del Rio, R. A., Foppiano, A., Ovalle, E. and Danilov, A. (2006) Long-term trends in foF2: a comparison of various methods. *J. Atmos. Sol. Terrest. Phys.*, **68**, 1854–1870.
Mikhailov, A. and Marin, D. (2000) Geomagnetic control of the fof2 long-term trends. *Ann. Geophys.*, **18**, 653–665.
Percival, D. B. and Walden, A. T. (2000) *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.
Ramsay, J., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and MATLAB*. New York: Springer.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. New York: Springer.
Rishbeth, H. (1990) A greenhouse effect in the ionosphere? *Planet. Space Sci.*, **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? *Geophys. Res. Lett.*, **16**, 1441–1444.
Shi, J. Q. and Choi, T. (2011) *Gaussian Process Regression Analysis for Functional Data*. Boca Raton: CRC Press.