

## Some Recent Developments in Inference for Geostatistical Functional Data

Algunos desarrollos recientes en inferencia para datos funcionales geoestadísticos

PIOTR KOKOSZKA<sup>1,a</sup>, MATTHEW REIMHERR<sup>2,b</sup>

<sup>1</sup>DEPARTMENT OF STATISTICS, COLORADO STATE UNIVERSITY, FORT COLLINS, USA

<sup>2</sup>DEPARTMENT OF STATISTICS, PENN STATE UNIVERSITY, STATE COLLEGE, USA

---

### Abstract

We review recent developments related to inference for functions defined at spatial locations. We also consider time series of functions defined at irregularly distributed spatial points or on a grid. We focus on kriging, estimation of the functional mean and principal components, and significance testing, giving special attention to testing spatio-temporal separability in the context of functional data. We also highlight some ideas related to extreme value theory for spatially indexed functional time series.

**Key words:** Functional data; Spatial statistics.

### Resumen

Revisamos desarrollos recientes relacionados con la inferencia de funciones definidas en locaciones espaciales. También consideramos series de tiempo funcionales definidas en puntos espaciales irregularmente distribuidos en una cuadrícula. Nos centramos en el kriging, la estimación de la media funcional y de los componentes principales, y en la prueba de significancia, dando especial atención a pruebas de separabilidad de espacio-tiempo en el contexto de datos funcionales. También destacamos algunas ideas relacionadas con la teoría de valores extremos para series de tiempo funcionales indexadas en el espacio.

**Palabras clave:** datos funcionales; estadística espacial.

---

<sup>a</sup>PhD. E-mail: [Piotr.Kokoszka@colostate.edu](mailto:Piotr.Kokoszka@colostate.edu)

<sup>b</sup>PhD. E-mail: [mreimherr@psu.edu](mailto:mreimherr@psu.edu)

## 1. Introduction

The objective of this paper is to review some recent developments in inference for geostatistical functional data. We thus consider data which are a collection of curves at spatial locations  $\{\mathbf{s}_k\}$ . Each  $t \mapsto X(\mathbf{s}_k; t)$  is a curve. For example, the  $\{\mathbf{s}_k\}$  can be locations of meteorological stations, and  $X(\mathbf{s}_k, t)$  can be some measurement, for example temperature or wind speed at time  $t$ . A sample of geostatistical spatial data has the form

$$\{X(\mathbf{s}_k, t), \mathbf{s}_k \in \mathbf{S}, t \in \mathcal{T}, k = 1, 2, \dots, N\}.$$

The spatial domain  $\mathbf{S}$  is typically a subset of the two-dimensional plane or sphere. The points  $\mathbf{s}_k$  at which observations are available are generally scattered over  $\mathbf{S}$  in some irregular manner. In some applications, e.g. satellite pictures, these points form a grid, but the methodology we discuss in this paper has been motivated by and generally applies to irregularly spaced locations  $\mathbf{s}_k$ . Some more general data structures consisting of time series of functions observed at spatial locations will also be considered.

Just as in the case of scalar geostatistical data, the model for functional geostatistical data is the random field  $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$ , with the difference that now each  $X(\mathbf{s})$  is a random function. We assume that each  $X(\mathbf{s})$  is an element of the Hilbert space  $L^2 = L^2(\mathcal{T})$ , with the interval  $\mathcal{T} = [0, 1]$  chosen merely for convenience. We assume that each  $X(\mathbf{s})$  is square integrable, i.e.

$$E \|X(\mathbf{s})\|^2 = \int EX^2(\mathbf{s}; t) dt < \infty.$$

If we assume that  $\mathbf{S}$  is the whole Euclidean space, we can define strict stationarity by the condition

$$\{X(\mathbf{s}_1 + \mathbf{h}), X(\mathbf{s}_2 + \mathbf{h}), \dots, X(\mathbf{s}_m + \mathbf{h})\} \stackrel{d}{=} \{X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_m)\}. \quad (1)$$

If this is the case, then  $E \|X(\mathbf{s})\|^2$  does not depend on  $\mathbf{s}$ . Square integrability and strict stationarity imply that the mean function

$$\mu(t) = EX(\mathbf{s}; t)$$

and the covariances

$$C(\mathbf{h}; t, u) = \text{Cov}(X(\mathbf{s}; t), X(\mathbf{s} + \mathbf{h}; u))$$

exist and do not depend on  $\mathbf{s}$ . When working with functional geostatistical data, it is important to keep in mind that the function  $\mathbf{h} \mapsto C(\mathbf{h}; t, u)$  is nonnegative definite only if  $t = u$ . In that case,  $C(\cdot; t) := C(\cdot; t, t)$  is the spatial covariance function of the scalar field  $\{X(\mathbf{s}; t)\}$ . It is often assumed that the covariances are isotropic, i.e.  $C(\mathbf{h}; t, u) = C(h; t, u)$  depends only on  $h = \|\mathbf{h}\|$ . *Unless stated otherwise, in the remainder of this paper, we assume that the field  $\{X(\mathbf{s})\}$  is square integrable, strictly stationary, and isotropic.*

We also consider a more complex data structure; at spatial locations we observe time series of functions. The data have the form  $X_n(\mathbf{s}; t)$ , where  $n$  is year,  $\mathbf{s}$  is a spatial location, and  $t$  is time within a year. For example  $X_n(\mathbf{s}; t)$  can be the curve of daily temperatures, the unit of  $t$  is day, in year  $n$ , and at location  $\mathbf{s}$ . Other examples are: total daily precipitation, wind strength, wind direction, solar radiation. Many of these daily time series are either maxima, minima or averages of data obtained in higher temporal resolutions. Historical records can be treated as geostatistical data, as meteorological stations are not uniformly distributed. However, computer weather prediction models produce complete records at dense grids that extend many decades into the future.

The paper is organized as follows. In Section 2, we review the subject of kriging in the context of functional data. Kriging is a fundamental tool in the field of spatial statistics, and so is a suitable starting point. We then turn, in Section 3, to the subject of estimation of the the mean function and the functional principal components from a sample of spatially distributed curves. Most methods of Sections 2 and 3 are implemented in the R package `geofd`. Examples of code are given in Chapter 9 of Kokoszka & Reimherr (2017). Section 4 is devoted to several significance tests in the context of spatial functional data. Testing of separability is presented in a separate section, Section 5, due to its importance in many inferential procedures for spatio-temporal data. We conclude with fairly recent research on extreme events defined in terms of spatially indexed functional time series. This work, whose many aspects are still under development, is outlined in Section 6.

It is not our objective to present a complete review of the subject, but rather to explain those recent developments which have attracted our interest. We strive not to merely provide references, but to explain the central points in detail that will allow the reader to gain sufficient understanding to decide if a specific development is relevant to her work or not.

We assume that the reader is familiar with the fundamental concepts of spatial statistics, explained, e.g., in initial chapters of Wackernagel (2003), Schabenberger & Gotway (2005) or Sherman (2011). Chapter 3 of Gelfand, Diggle, Fuentes & Guttorp (2010) also provides a sufficient introduction. Some familiarity with the fundamental concepts of functional data analysis would also be helpful; Chapters 1 and 3 of Kokoszka & Reimherr (2017) provide enough background.

## 2. Kriging

We observe functions  $X(\mathbf{s}_k)$  at spatial locations  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$ . Suppose  $\mathbf{s}$  is a different location, and we want to predict the function  $X(\mathbf{s})$  using the functions  $X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_N)$ . In the functional setting, the available data can be represented as a matrix of scalars

$$[X(\mathbf{s}_k; t_j), k = 1, 2, \dots, N, j = 1, 2, \dots, J],$$

assuming all functions are observed at the same points  $t_j$ . In some situations, the points  $t_j$  may depend on the location  $\mathbf{s}_k$ , so the scalar data points are  $X(\mathbf{s}_k, t_{kj})$ .

There are many approaches to computing a predictor  $\widehat{X}(\mathbf{s})$  of the unobserved curve  $X(\mathbf{s})$ . We start with a method which is the most direct extension of the approach used for scalar data.

Taking a functional point of view, we treat each curve  $X(\mathbf{s}_k)$  as an indivisible data object. We want to determine the weights  $w_k$  for the predictor of the form

$$\widehat{X}(\mathbf{s}) = \mu + \sum_{k=1}^N w_k (X(\mathbf{s}_k) - \mu). \quad (2)$$

Since the  $X(\mathbf{s}_k)$  are now functions, we calculate the weights by minimizing the expected  $L^2$  distance between  $\widehat{X}(\mathbf{s})$  and  $X(\mathbf{s})$ . We thus want to find weights  $w_1, w_2, \dots, w_N$  which minimize

$$E \left\| \widehat{X}(\mathbf{s}) - X(\mathbf{s}) \right\|^2 = E \int \left( \widehat{X}(\mathbf{s}; t) - X(\mathbf{s}; t) \right)^2 dt.$$

Observe that

$$\begin{aligned} \left\| \widehat{X}(\mathbf{s}) - X(\mathbf{s}) \right\|^2 &= \left\| \mu + \sum_{k=1}^N w_k (X(\mathbf{s}_k) - \mu) - X(\mathbf{s}) \right\|^2 \\ &= \left\langle \sum_{k=1}^N w_k (X(\mathbf{s}_k) - \mu) - (X(\mathbf{s}) - \mu), \sum_{\ell=1}^N w_\ell (X(\mathbf{s}_\ell) - \mu) - (X(\mathbf{s}) - \mu) \right\rangle \\ &= \sum_{k, \ell=1}^N w_k w_\ell \langle X(\mathbf{s}_k) - \mu, X(\mathbf{s}_\ell) - \mu \rangle \\ &\quad - 2 \sum_{k=1}^N w_k \langle X(\mathbf{s}_k) - \mu, X(\mathbf{s}) - \mu \rangle + \langle X(\mathbf{s}) - \mu, X(\mathbf{s}) - \mu \rangle. \end{aligned}$$

Define the functional spatial covariances as

$$C(\mathbf{s}, \mathbf{s}') = E [\langle X(\mathbf{s}) - \mu, X(\mathbf{s}') - \mu \rangle]. \quad (3)$$

Notice that for any two spatial points,  $C(\mathbf{s}, \mathbf{s}')$  is a real number, as opposed to an operator. Using this notation, we see that

$$E \left\| \widehat{X}(\mathbf{s}) - X(\mathbf{s}) \right\|^2 = \sum_{k, \ell=1}^N w_k w_\ell C(\mathbf{s}_k, \mathbf{s}_\ell) - 2 \sum_{k=1}^N w_k C(\mathbf{s}_k, \mathbf{s}) + C(\mathbf{s}, \mathbf{s}).$$

Finding the kriging weights is thus based on the same equations as for scalar data, but using the functional covariances (3). Therefore, the weights  $w_k$  are found by solving the system of equations

$$\sum_{\ell=1}^N C(\mathbf{s}_k, \mathbf{s}_\ell) w_\ell = C(\mathbf{s}_k, \mathbf{s}), \quad k = 1, 2, \dots, N. \quad (4)$$

To implement this approach in practice, we must estimate the covariances  $C(\mathbf{s}_k, \mathbf{s}_\ell)$  and  $C(\mathbf{s}_k, \mathbf{s})$ . We now explain how this can be done. Using the assumptions of stationarity and isotropy, we see that

$$\begin{aligned} C(\mathbf{s}, \mathbf{s}') &= E \int (X(\mathbf{s}; t) - \mu(t))(X(\mathbf{s}'; t) - \mu(t)) dt \\ &= \int C(\|\mathbf{s} - \mathbf{s}'\|; t) dt, \end{aligned}$$

where

$$C(h; t) = \text{Cov}(X(\mathbf{s} + \mathbf{h}; t), X(\mathbf{s}; t)), \quad h = \|\mathbf{h}\|,$$

is the spatial covariance function of the scalar field  $\{X(\cdot; t)\}$  obtained by fixing  $t$ . For each  $t$ , the covariance function  $h \mapsto C(h; t)$  can be estimated by any method that is applicable to covariance estimation for scalar spatial data; this is typically done by estimating the semivariogram.

The kriging predictor is often expressed in a slightly different way. We want to find a predictor  $X^*(\mathbf{s})$  of the form  $X^*(\mathbf{s}) = \sum_{k=1}^N \lambda_k X(\mathbf{s}_k)$ , subject to the condition  $\sum_{k=1}^N \lambda_k = 1$ . By contrast, equations (4) do not imply that  $\sum_{k=1}^N w_k = 1$ . In many applications, the weights  $w_j$  are very close to the weights  $\lambda_j$ , and the predicted function  $\widehat{X}(\mathbf{s})$  and  $X^*(\mathbf{s})$  are also very close.

In addition to these simple methods, (Delicado, Giraldo, Comas & Mateu 2010) give references to several other methods of kriging functional data. Another relatively simple kriging method they discuss, which gives predictions as good as more complex methods can be summarized as follows. The predictor has the form

$$\widehat{X}(\mathbf{s}; t) = \sum_{k=1}^N w_k(t) X(\mathbf{s}_k; t),$$

where the unknown functions  $w_k(\cdot)$  satisfy  $\sum_{k=1}^N w_k(t) = 1$ , for each  $t$ . The observed functions  $X(\mathbf{s}_k)$  and the unknown weight functions  $w_k(\cdot)$  are expanded using some basis system:

$$X(\mathbf{s}_k; t) = \sum_{m=1}^M c_{km} B_m(t), \quad w_k(t) = \sum_{m=1}^M b_{km} B_m(t).$$

In particular, one can use the estimated FPC's  $\hat{v}_m$  as the  $B_m$ . The coefficient matrix  $\mathbf{B} = [b_{km}, 1 \leq k \leq N, 1 \leq m \leq M]$  is also found by minimizing  $E\|\widehat{X}(\mathbf{s}) - X(\mathbf{s})\|^2$ . This approach is similar to the method of co-kriging used for multivariate data, e.g. Wackernagel (2003).

We have assumed so far that  $EX(\mathbf{s}; t) = \mu(t)$  does not depend on the location  $\mathbf{s}$ . Caballero, Giraldo & Mateu (2013) and Menafoglio, Secchi & Rosa (2013) show how to perform kriging under a more flexible assumption that  $EX(\mathbf{s}; t) = \sum_{\ell=1}^L \beta_\ell(t) f_\ell(\mathbf{s})$ .

### 3. Estimation of the mean function and of the FPC's

To estimate the mean function, we use the weighted average

$$\hat{\mu}(t) = \sum_{k=1}^N w_k X(\mathbf{s}_k; t), \quad (5)$$

and the problem becomes how to compute the optimal weights  $w_k$ . As in Section 2, in (5) we treat the curves  $X(\mathbf{s}_k)$  as indivisible statistical objects, and so we apply one weight to the whole curve. One could consider weights  $w_k(t)$ , defined as functions, but the simple approach we describe in this section is easy to implement, and produces estimates which significantly improve on the simple average. Figure 1 illustrates the difference between the simple average ( $w_k = 1/N$ ) and the weighted average we are about to derive, using annual temperature curves in Canada. There are plenty of locations in the southern part of Canada, especially in the densely populated South-East of the country, and very few in the North. If we postulate the model  $X(\mathbf{s}_k; t) = \mu(t) + \varepsilon(\mathbf{s}_k; t)$ , then  $\mu$  should represent the typical annual temperature profile for the whole country. The simple average will mostly reflect temperature curves in the region where there are many stations, i.e. mostly in the South. To obtain a more informative kriged mean function, stations far in the North must receive larger weights. Some stations in the south will receive even slightly negative weights.

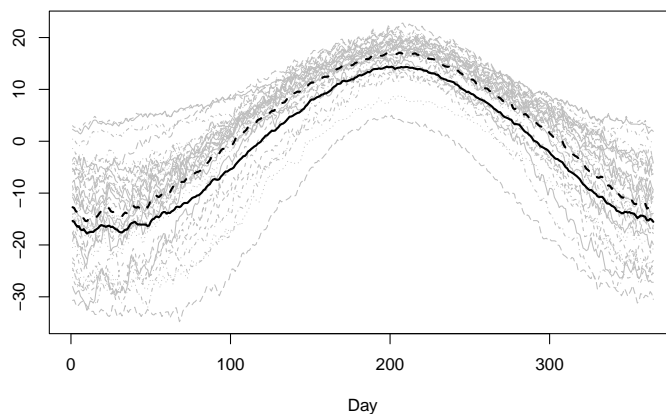


FIGURE 1: Gray lines: annual temperature curves at 35 locations in Canada (averaged over several decades). Dashed line: simple average of these curves. Continuous black line: mean function estimated by functional kriging.

We define the optimal weights as those that minimize

$$E \left\| \sum_{k=1}^N w_k X(\mathbf{s}_k) - \mu \right\|^2 = \int \left\{ \sum_{k=1}^N w_k X(\mathbf{s}_k; t) - \mu(t) \right\}^2 dt, \quad (6)$$

subject to the condition  $\sum_{k=1}^n w_k = 1$  which ensures that  $E\hat{\mu} = \mu$ . The method of Lagrange multipliers leads to the following system of  $N + 1$  linear equations:

$$\sum_{k=1}^N w_k = 1, \quad \sum_{k=1}^N w_k C(\mathbf{s}_k, \mathbf{s}_n) - r = 0, \quad n = 1, 2, \dots, N, \quad (7)$$

in which  $w_1, w_2, \dots, w_N, r$  are  $N+1$  unknowns. The functional covariances  $C(\mathbf{s}_k, \mathbf{s}_n)$  can be estimated as explained in Section 2.

A fundamental tool of functional data analysis is the Karhunen–Loève expansion. Every function  $X$  in  $L^2$  can be represented as  $X(t) = \sum_{j=1}^{\infty} \xi_j v_j(t)$ , where the  $v_j$  are orthonormal functions known as the functional principal components (FPC's). The main feature of this expansion is that the function  $X^{(p)}(t) = \sum_{j=1}^p \xi_j v_j(t)$  is a very good approximation of the function  $X(\cdot)$  even for small  $p$ ; in many applications  $p = 3$  is sufficient. A small number of *scores*  $\xi_1, \xi_2, \dots, \xi_p$  thus encodes an infinitely dimensional function. Details are explained in Chapter 11 of Kokoszka & Reimherr (2017). The functions  $v_j$  are the eigenfunctions of the covariance function  $c(t, s) = \text{Cov}(X(t), X(s))$ . Therefore, if the spatially indexed functions  $X(\mathbf{s})$  have the same distribution, then each of them admits the expansion

$$X(\mathbf{s}; t) = \mu(t) + \sum_{i=1}^{\infty} \xi_i(\mathbf{s}) v_i(t). \quad (8)$$

We now explain how to estimate the  $v_j$  in (8) under the assumptions of stationarity and isotropy stated in Section 1. We also assume that  $\mu(t) = 0$ ; in practice we would subtract the estimated mean function. We describe one of the two methods proposed by Gromenko, Kokoszka, Zhu & Sojka (2012). Both methods have very similar finite sample performance, and have significantly smaller MSE's compared to the standard method which does not take spatial dependence into account.

It is assumed that the  $v_j$  admit the expansion

$$v_j(t) = \sum_{\alpha=1}^K x_{\alpha}^{(j)} B_{\alpha}(t), \quad (9)$$

where the  $B_{\alpha}$  are elements of an *orthonormal* basis, for example Fourier basis (splines will not work because they are not orthogonal). The expansion (9) is already an estimation step. We would take  $K$  so large that each observed function  $X(\mathbf{s}_k)$  is well approximated by the sum  $X^{(K)}(\mathbf{s}_k; t) = \sum_{\alpha=1}^K b_{\alpha}(\mathbf{s}_k) B_{\alpha}(t)$ . The exact expansion is

$$X(\mathbf{s}; t) = \sum_{\alpha=1}^{\infty} b_{\alpha}(\mathbf{s}) B_{\alpha}(t), \quad b_{\alpha}(\mathbf{s}) = \langle X(\mathbf{s}), B_{\alpha} \rangle.$$

The  $b_{\alpha}(\mathbf{s}_k)$  form an observable field.

The  $v_j$  are determined (up to a sign) by the conditions  $C(v_j) = \lambda_j v_j$  and  $\|v_j\| = 1$ , where  $C(\cdot) = E[\langle X(\mathbf{s}), \cdot \rangle X(\mathbf{s})]$  is the covariance operator. Using the orthonormality of the  $B_j$ , we obtain

$$\begin{aligned}
C(B_j) &= E \left[ \left\langle \sum_{\alpha=1}^{\infty} b_{\alpha}(\mathbf{s}) B_{\alpha}, B_j \right\rangle \sum_{i=1}^{\infty} b_i(\mathbf{s}) B_i \right] \\
&= E \left[ b_j(\mathbf{s}) \sum_{i=1}^{\infty} b_i(\mathbf{s}) B_i \right] \\
&= \sum_{i=1}^{\infty} E[b_i(\mathbf{s}) b_j(\mathbf{s})] B_i.
\end{aligned} \tag{10}$$

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

Fix  $i$  and  $j$ , and define the scalar field  $z$  by  $z(\mathbf{s}) = b_i(\mathbf{s})b_j(\mathbf{s})$ . 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 (10) is then

$$\widehat{C}(B_j) = \sum_{i=1}^K \hat{r}_{ij} B_i. \tag{11}$$

Relation (11) defines the estimator  $\widehat{C}$  which acts on the span of  $B_j, 1 \leq j \leq K$ . Its eigenfunctions are of the form  $x = \sum_{1 \leq \alpha \leq K} x_{\alpha} B_{\alpha}$ . 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

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

we can write the above as a matrix equation

$$\widehat{\mathbf{R}}\mathbf{x} = \lambda\mathbf{x}. \tag{12}$$

Denote the solutions to (12) by

$$\hat{\mathbf{x}}^{(j)} = [\hat{x}_1^{(j)}, \hat{x}_2^{(j)}, \dots, \hat{x}_k^{(j)}]^T, \quad \hat{\lambda}_j, \quad 1 \leq j \leq K. \tag{13}$$

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

$$\hat{v}_j = \sum_{\alpha=1}^K \hat{x}_{\alpha}^{(j)} B_{\alpha} \tag{14}$$

are also orthonormal (because the  $B_j$  are orthonormal). The  $\hat{v}_j$  given by (14) are the estimators of the FPC's, and the  $\hat{\lambda}_j$  in (13) of the corresponding eigenvalues.

A natural question to ask is what happens if one applies standard software to the observed curves  $X(\mathbf{s}_k), 1 \leq k \leq K$ , in order to estimate the FPC's,  $v_j$ ,



in (8). The usual method, implemented in the function `pca` of the `fda` package (and in many other packages) is to use the the eigenfunctions of the sample covariance operator. It is suitable for a sample of iid curves or for curves forming a regularly spaced functional time series, but not for geostatistical functional data. We illustrate with an example. A comprehensive theory and more examples were developed by Hörmann & Kokoszka (2013). This material is also presented in Chapter 18 of Horváth & Kokoszka (2012).

For mean zero functions, the sample covariance operator  $\widehat{C}$  is defined by

$$\widehat{C}_N(x) = N^{-1} \sum_{k=1}^N \langle X_k, x \rangle X_k.$$

The usual estimated FPC's are defined by  $\widehat{C}_N(\hat{v}_j) = \hat{\lambda}_j \hat{v}_j$ . Consider a functional random field

$$X(\mathbf{s}; t) = \sum_{j=1}^{\infty} \xi_j(\mathbf{s}) e_j(t), \quad \mathbf{s} \in \mathbb{R}^d, \quad t \in [0, 1], \quad (15)$$

where  $\{e_j, j \geq 1\}$  is a complete orthonormal system and the  $\xi_j(\mathbf{s})$  are mean zero random variables with  $E[\xi_j(\mathbf{s})\xi_j(\mathbf{s} + \mathbf{h})] = \lambda_j \rho_j(h)$ ,  $h = \|\mathbf{h}\|$ , where  $\sum_{j=1}^{\infty} \lambda_j < \infty$  and each  $\rho_j(\cdot)$  is a positive correlation function. Recall that  $C(\cdot) = E[\langle X(\mathbf{s}), \cdot \rangle X(\mathbf{s})]$ . Direct verification shows that  $C(x) = \sum_{j=1}^{\infty} \lambda_j \langle e_j, x \rangle e_j$ , so the  $\lambda_j$  are the eigenvalues of  $C$ , and the  $e_j$  the corresponding eigenfunctions, i.e.  $v_j = e_j$ . Now consider a sequence  $\mathbf{s}_n \rightarrow \mathbf{0}$ . Because of the positive dependence,  $X(\mathbf{s}_n)$  is close to  $X(\mathbf{0})$ , so  $\widehat{C}_N$ , as an arithmetic average, is close to the random operator  $X^* = \langle X(\mathbf{0}), \cdot \rangle X(\mathbf{0})$ . Observe that  $X^*(X(\mathbf{0})) = \|X(\mathbf{0})\|^2 X(\mathbf{0})$ . Thus  $X(\mathbf{0})$  is an eigenfunction of  $X^*$ . Since it is random, it cannot be close to any of the  $v_j$ , even asymptotically.

The above is only an intuitive illustration of the point that if the spatial locations at which functions are observed have dense clusters, or if the the dependence is strong, then the eigenfunctions of  $\widehat{C}_N$  do not consistently estimate the FPC's  $v_j$  in (8).

## 4. Independence, Trend and Change Point Tests

In this section, we discuss several significance tests for geostatistical functional data. These tests are analogs of simpler tests for iid curves. The spacial dependence between the curves requires more complex forms of the test statistics. In addition, these statistics typically require complex numerical procedures to compute them. We therefore merely describe what has been done, and provide suitable references.

We begin with a test of independence of two functional random fields. It was proposed by Gromenko et al. (2012), and is described in detail in Chapter 17 of Horváth & Kokoszka (2012). The data are observed at  $N$  spatial locations:  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$ . At location  $\mathbf{s}_k$ , we observe two curves:  $X(\mathbf{s}_k)$  and  $Y(\mathbf{s}_k)$ . We assume that the sample  $\{X(\mathbf{s}_k)\}$  is a realization of a random field  $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$ ,

and the sample  $\{Y(\mathbf{s}_k)\}$  is a realization of a random field  $\{Y(\mathbf{s}), s \in \mathbf{S}\}$ . We want to test the null hypothesis:

$H_0$ : for each  $s \in \mathbf{S}$ , the random functions  $X(\mathbf{s})$  and  $Y(\mathbf{s})$  are independent

against the alternative that  $H_0$  does not hold. The test statistic detects departures from  $H_0$  that manifest themselves in the lack of the correlation between the projections  $\langle x, X(\mathbf{s}) \rangle$  and  $\langle y, Y(\mathbf{s}) \rangle$ , for any  $x, y \in L^2$ . In practice, the lack of correlation can be tested only for  $x$  and  $y$  from sufficiently large finite dimensional subspaces, those spanned by the first  $p$  FPC's of  $X(\mathbf{s})$ , and the first  $q$  FPC's of  $Y(\mathbf{s})$ . The idea of the test, thus requires that the pairs  $(X(\mathbf{s}), Y(\mathbf{s}))$  have the same distribution for every  $\mathbf{s} \in \mathbf{S}$ . The construction of the test assumes that both fields,  $\{X(\mathbf{s}), s \in \mathbf{S}\}$  and  $\{Y(\mathbf{s}), s \in \mathbf{S}\}$  are stationary. The test statistic  $\widehat{S}_N$  is constructed from the inner products  $\langle X(\mathbf{s}_k), \hat{v}_j \rangle, 1 \leq j \leq p$  and  $\langle Y(\mathbf{s}_k), \hat{u}_j \rangle, 1 \leq j \leq q$ , where the  $\hat{v}_j$  are estimated FPC's of the  $X$ -sample and the  $\hat{u}_j$  of the  $Y$ -sample. The  $\hat{v}_j$  and the  $\hat{u}_j$  be computed as described in Section 3. Under  $H_0$ , the test statistic  $\widehat{S}_N$  converges, as  $N \rightarrow \infty$ , to a chi-square distribution with  $pq$  degrees of freedom. Once the  $\hat{v}_j$  and the  $\hat{u}_j$  are computed, the tests statistic can be computed without difficulty using package `fda`. Since the critical values are those of the chi-square distribution, the test is easy to apply.

The problem of testing the equality of two mean functions, each defined for functions observed over a different spatial region, is considered in Gromenko & Kokoszka (2012).

Our next test is motivated by an interesting and extensively studied problem of space physics. The account presented here is based on Gromenko & Kokoszka (2013) and Gromenko, Kokoszka & Sojka (2017). We first describe the space physics problem, and then explain the idea of the test.

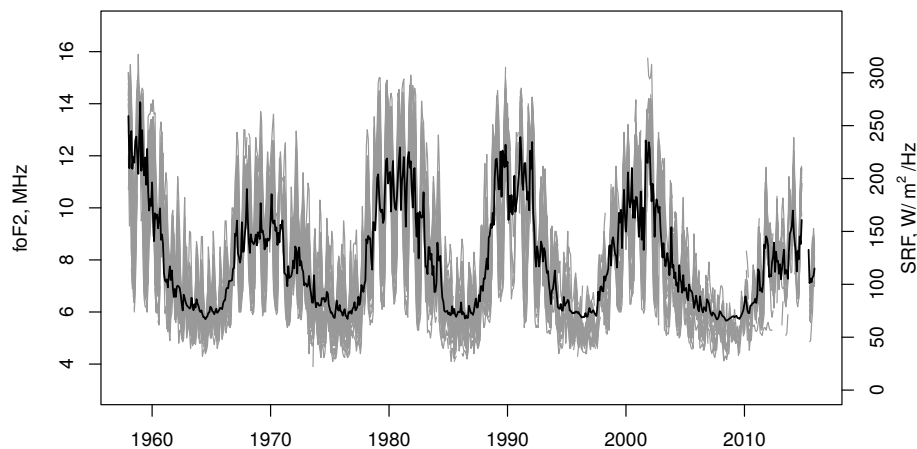


FIGURE 2: Gray lines represent ionosonde measurements obtained at observatories located in mid-latitude northern hemisphere, with the scale on the left-hand side. The black line represents the observed solar radio flux with the scale on the right-hand side.

Increased concentration of greenhouse gases in the upper atmosphere is associated with global warming in the lower troposphere (the atmosphere roughly below 10 km). Roble & Dickinson (1989) suggested that the increasing amounts of these radiatively active gases, mostly  $\text{CO}_2$  and  $\text{CH}_4$ , would lead to a global cooling in the ionosphere (atmosphere roughly 300 km above the Earth's surface). Rishbeth (1990) pointed out that this would result in a thermal contraction of the ionosphere. The height of the ionosphere can be approximately computed using data from a radar-type instrument called the ionosonde. Relevant measurements have been made for many decades by globally distributed ionosondes. In principle, these observations could be used to quantitatively test the hypothesis of Roble & Dickinson (1989). The difficulty in testing the contraction hypothesis comes from several sources. The height of the ionosphere depends on magnetic coordinates, the season, long term changes in the strength and direction of the internal magnetic field, and, most importantly, on the solar cycle; more solar radiation leads to greater ionization. This is illustrated in Figure 2. Another difficulty stems from the fact that ionosonde records are not complete. Most observation stations do not operate continuously for many decades. They start and end operation at different times, some of them are out of service for many years, or even decades. In the mid-latitude northern hemisphere, there are 81 ionosonde stations, but at any given time, data from no more than 40 are available, as shown in Figure 3. This means that the estimation methods described in Section 3 cannot be used, as they require complete records to compute various integrals. More complex methods that work for incomplete records are needed.

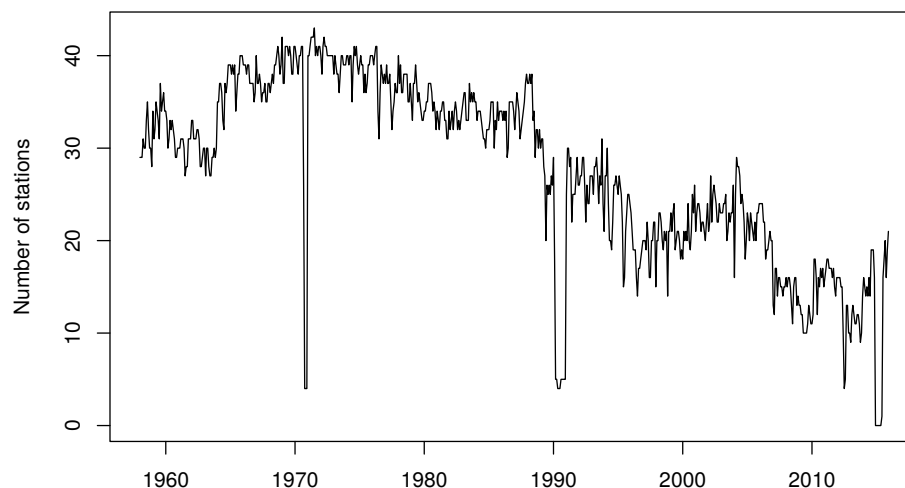


FIGURE 3: Number of available stations in the mid-latitude northern hemisphere.

Let  $Y(\mathbf{s}_k; \tau_i)$  be the original record at location  $\mathbf{s}_k$ , measured from 1958 to 2015, possibly with long gaps. The set of all locations is  $\{\mathbf{s}_k, 1 \leq k \leq K\}$ , and the set of time points at which measurements may be available is  $\{\tau_i, 1 \leq i \leq T\}$ ; in Gromenko, Kokoszka & Sojka (2017) these are months from January 1958 to December 2015. The following spatio-temporal model is postulated:

$$Y(\mathbf{s}; \tau) = \mu(\mathbf{s}; \tau) + \varepsilon(\mathbf{s}; \tau) + \theta(\mathbf{s}; \tau), \quad (16)$$

where  $\mathbf{s}$  is a generic location in a region of interest, and  $\tau$  is continuous time. A simple form of the mean function relevant to the space physics problem is

$$\mu(\mathbf{s}; \tau) = \beta_1 + \beta_2 \tau + \beta_3 \text{SRF}(\tau) + \beta_4 M(\mathbf{s}; \tau),$$

where  $\text{SRF}(\tau)$  is the solar radio flux, cf. Fig. 2, and  $M(\mathbf{s}; \tau)$  is a suitable function computed from the coordinates of the internal magnetic field. The interest lies in the estimation of the mean function, and testing if it contains a linear trend, i.e. testing  $H_0 : \beta_2 = 0$ . The function  $\mu(\cdot, \cdot)$  is treated as an unknown deterministic functional parameter. The second term,  $\varepsilon(\mathbf{s}; \tau)$ , describes the spatio-temporal variability away from the mean function. Stochastic modeling of this term is needed to develop inferential procedures. The term  $\theta(\mathbf{s}; \tau)$  represents a random error, which can be associated with measurement error. The details of the estimation and testing procedures are too complex to describe here. The conclusion is that  $\beta_2$  is significantly negative, confirming the hypothesis of global ionospheric contraction. The software to perform the test is available and can be used to test for the presence of global trends in other data of this type, for example in near-surface temperatures.

We conclude this section with a discussion of change-point detection for the mean function. The presentation is based on Gromenko, Kokoszka & Reimherr (2017). General theory for change points for scalar data is presented in Csörgő & Horváth (1997), several extensions to functional data are presented in Horváth & Kokoszka (2012). In the spatio-temporal setting we consider here, the data are assumed to follow the model

$$X_n(\mathbf{s}, t) = \mu_n(\mathbf{s}; t) + \varepsilon_n(\mathbf{s}; t).$$

In most applications,  $n$  denotes year,  $\mathbf{s}$ , as before, spatial location, and  $t$  time within a year. For a fixed  $\mathbf{s}$ ,  $\{X_n(\mathbf{s}), n = 1, 2, \dots\}$  is a time series of functions, one function per year. While we assume that the curves are observed densely in time, they are only observed at a finite number of spatial locations  $\{\mathbf{s}_k : k = 1, \dots, K\}$ . The goal is to determine if the mean functions,  $\mu_n$ , are the same across  $n$  or if there is a change in the mean. In other words, we aim to evaluate the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_N \quad \text{against} \quad H_A : \mu_1 = \dots = \mu_{n^*} \neq \mu_{n^*+1} = \dots = \mu_N.$$

The alternative of a sudden change at a single year  $n^*$  is to be viewed as a mathematical approximation. As with all testing problems, the null hypothesis of no change is tested, and the alternative specifies the violations of the null hypothesis which will be detected with the highest power. The tests described below, will also detect different violations of  $H_0$ , including gradual changes taking place over a few years.

To carry out the tests, we assume that the covariance structure of the  $\varepsilon_n$  is separable, a test of which will be discussed in section 5. In particular, we assume that

$$\text{Cov}(\varepsilon_n(\mathbf{s}; t), \varepsilon_n(\mathbf{s}'; t')) = v(t, t')u(\mathbf{s}, \mathbf{s}').$$

Notice that the decomposition above is unique only up to a constant, we thus assume that  $\int v(t, t) dt = 1$ . Gromenko, Kokoszka & Reimherr (2017) consider three different test statistics based on CUSUM's (cumulative sums), each of which uses the expansion (8) applied to the annual curves in different ways. They do not require a specific procedure for estimating  $v(t, t')$  and  $\mathbf{U} = \{u(\mathbf{s}_k, \mathbf{s}_{k'})\}$ , instead formulating their tests assuming they have consistent estimates. Examples of how to estimate these quantities are given in their appendix, as well as in Aston, Pigoli & Tavakoli (2016) and Constantinou, Kokoszka & Reimherr (2017). The first test involves using a temporal FPCA, normalizing by the corresponding eigenvalues, and then pooling across space. The tests statistics is

$$\hat{\Lambda}_1 = \frac{1}{N^2} \sum_{k=1}^K \hat{w}_k \sum_{i=1}^p \hat{\lambda}_i^{-1} \sum_{r=1}^r \left\langle \sum_{n=1}^N X_n(\mathbf{s}_k) - \frac{r}{N} \sum_{n=1}^N X_n(\mathbf{s}_k), \hat{v}_i \right\rangle^2.$$

The pooling weights,  $\hat{w}_k$ , can be chosen however the user prefers, but in practice, using weights as described in Section 3 works well. It has now become well recognized in the FDA community that dividing by estimated eigenvalues can cause stability problems with certain data. Thus a second test statistic was also proposed which omits normalizing by the  $\hat{\lambda}_i$ :

$$\hat{\Lambda}_2 = \frac{1}{N^2} \sum_{k=1}^K \hat{w}_k \sum_{i=1}^p \sum_{r=1}^r \left\langle \sum_{n=1}^r X_n(\mathbf{s}_k) - \frac{r}{N} \sum_{n=1}^N X_n(\mathbf{s}_k), \hat{v}_i \right\rangle^2.$$

For  $p$  large,  $\hat{\Lambda}_2$  is essentially an approximation of the third and final test statistic

$$\hat{\Lambda}_2^\infty = \frac{1}{N^2} \sum_{k=1}^K \hat{w}_k \sum_{i=1}^p \hat{\lambda}_i^{-1} \sum_{r=1}^r \left\| \sum_{n=1}^r X_n(\mathbf{s}_k) - \frac{r}{N} \sum_{n=1}^N X_n(\mathbf{s}_k) \right\|^2.$$

Each statistic has a slightly different limiting distribution. In particular, under  $H_0$

$$\begin{aligned} \hat{\Lambda}_1 &\xrightarrow{d} \sum_{k=1}^K w(k) \sum_{i=1}^d \int B_{ik}^2(t) dt \\ \hat{\Lambda}_2 &\xrightarrow{d} \sum_{k=1}^K w(k) \sum_{i=1}^d \lambda_i \int B_{ik}^2(t) dt \\ \hat{\Lambda}_2^\infty &\xrightarrow{d} \sum_{k=1}^K w(k) \sum_{i=1}^\infty \lambda_i \int B_{ik}^2(t) dt, \end{aligned}$$

where  $B_{ik}(t)$  are Brownian bridges which satisfy

$$\text{Cov}(B_{ik}(t), B_{i'k'}(t')) = 1_{i=i'} \min\{t, t'\}(1 - \min\{t, t'\})\sigma(\mathbf{s}_k, \mathbf{s}_{k'}).$$

These asymptotics can be used for choosing rejection regions with proper Type 1 error. To do so, one can use Monte Carlo, though (Gromenko, Kokoszka &

Reimherr 2017) also provide a normal approximation which works well when considering a relatively large number of spatial points. These tests can be implemented using the `sctp` package in R, which is available from <http://www.personal.psu.edu/~mlr36/codes.html> and includes an example code for implementing the methodology.

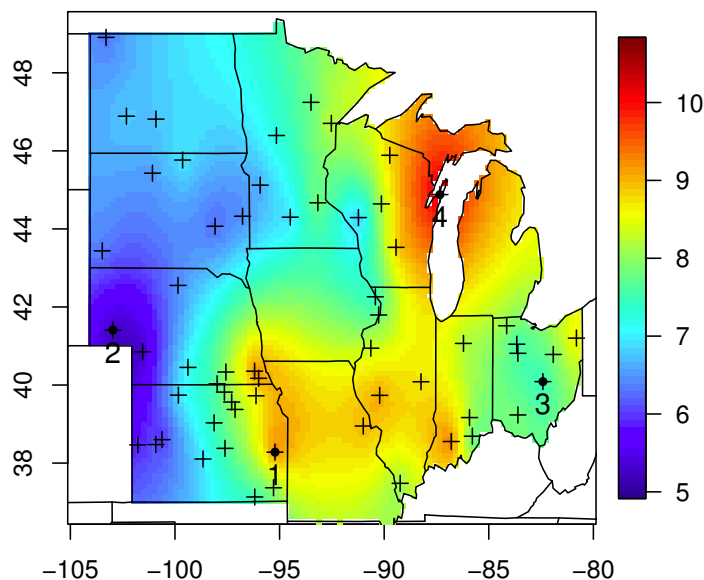


FIGURE 4: The spatial field showing the  $L^2$  distance between the mean log-precipitation before and after 1966. There is an increase in precipitation throughout the year in the area around location 4, decrease in the first half of the year in the area around location 1. Locations close to 2 and 3 do not show a large change nor a consistent pattern.

If a change point is detected, its location can be identified as the year when a test statistics attains maximum. Gromenko, Kokoszka & Reimherr (2017) applied the above test procedures (with suitable finite sample calibration) to test if the precipitation pattern over the midwest US states has changed. All tests led to the same conclusion: one change point in second half of the sixties. The pattern of the change is shown in Fig. 4. The biggest changes are in the area around Michigan Lake and south-west of the lake. To visualize the spatial distribution of change, the authors used the spatial field

$$\hat{\phi}(\mathbf{s}) = \|\hat{\mu}_1(\mathbf{s}) - \hat{\mu}_2(\mathbf{s})\|,$$

where

$$\hat{\mu}_1(\mathbf{s}; t) = \hat{r}^{-1} \sum_{n=1}^{\hat{r}} X_n(\mathbf{s}; t), \quad \hat{\mu}_2(\mathbf{s}; t) = (N - \hat{r})^{-1} \sum_{n=\hat{r}+1}^N X_n(\mathbf{s}; t).$$

They performed kriging with the exponential covariance model to obtain the heat map shown in Fig. 4. The application of the significance tests confirms that the

heat map shows a statistically significant change over a region, not a variation in the magnitude of change which may be due to chance.

## 5. Separability Tests

The second order structure of a random field of functions is described by the covariance function  $\sigma(\mathbf{s}, \mathbf{s}'; t, t') = \text{Cov}(X(\mathbf{s}, t), X(\mathbf{s}', t'))$ . Theoretical and computational aspects of most procedures can be significantly simplified if one can assume that

$$\sigma(\mathbf{s}, \mathbf{s}'; t, t') = u(\mathbf{s}, \mathbf{s}')v(t, t'), \quad (17)$$

that is, if the spatio-temporal covariance function factors into the product of a purely spatial and purely temporal covariance functions. The above condition is, for example, required to develop the change point methodology described in Section 4. If (17) holds, we say that the functional random field is *separable*. In particular, the spatial dependence structure is the same for any time  $t$ . Tests whose null hypothesis is (17) have been proposed by (Liu, Ray & Hooker 2017), (Aston et al. 2016) and (Constantinou et al. 2017). We discuss the latter two approaches here.

We begin by discussing two different procedures for estimating  $v(t, t')$  and  $\mathbf{U} = \{u(\mathbf{s}_k, \mathbf{s}_{k'})\}$ . Aston et al. (2016) used partial trace operators to estimate these quantities, while (Constantinou et al. 2017) combined bases expansions with a “flip-flop” approach, which basically iterates between estimating  $v$  and  $\mathbf{U}$ . Partial trace operators are used to eliminate either the temporal or spatial component of the covariance, in particular, define the estimated covariance function as

$$\hat{\sigma}(\mathbf{s}_k, \mathbf{s}'_k, t, t') = \frac{1}{N} \sum_{n=1}^N X_n(\mathbf{s}_k, t) X_n(\mathbf{s}'_k, t').$$

(In the above formula, and in the remainder of this section, we assume that the mean function is zero.) Then the spatial covariance is estimated using

$$\hat{\mathbf{U}}_{kk'} = \int \hat{\sigma}(\mathbf{s}_k, \mathbf{s}'_k, t, t) dt,$$

and the temporal covariance is estimated using

$$\hat{v}(t, t') = \frac{\sum_{k=1}^K \hat{\sigma}(\mathbf{s}_k, \mathbf{s}_k, t, t')}{\sum_k \int \hat{\sigma}(\mathbf{s}_k, \mathbf{s}_k, t, t) dt}.$$

The normalization above ensures that the trace of  $v(t, t')$  is normalized to be 1, though other normalizations are possible. One thing to note is that the full covariance  $\hat{\sigma}$  does not actually need to be estimated, as one can move to estimating the  $\hat{\mathbf{U}}$  and  $\hat{v}$  directly.

The second approach to estimating  $v$  and  $\mathbf{U}$  comes from (Constantinou et al. 2017) and begins with basis expansions. In particular, one approximates each temporal curve using  $J$  basis functions:

$$X_n(\mathbf{s}_k, t) \approx \sum_{j=1}^J \xi_{jn}(\mathbf{s}_k) e_j(t).$$

The basis  $e_j(t)$  can either be deterministic or data driven, such as FPCA. Testing separability of  $X_n$  is then translated into testing separability of  $\xi_{jn}(\mathbf{s}_k)$ , so that the null becomes

$$H_0 : \text{Cov}(\xi_{jn}(\mathbf{s}_k), \xi_{j'n}(\mathbf{s}_{k'})) = \mathbf{V}_{jj'} \mathbf{U}_{kk'}.$$

However, this is now a multivariate problem and we can utilize multivariate methods for estimating  $\mathbf{V}$  and  $\mathbf{U}$ , in particular, the maximum likelihood estimators, assuming the data are Gaussian, satisfy

$$\hat{\mathbf{V}} = \frac{1}{N} \sum_{n=1}^N \mathbf{\Xi}_n \hat{\mathbf{U}}^{-1} \mathbf{\Xi}_n^\top \quad \hat{\mathbf{U}} = \frac{1}{N} \sum_{n=1}^N \mathbf{\Xi}_n^\top \hat{\mathbf{V}}^{-1} \mathbf{\Xi}_n,$$

where  $\mathbf{\Xi}_n = \{\xi_{jn}(\mathbf{s}_k)\}$  is a  $J \times K$  matrix. To compute these estimators, one chooses an initial value, for example  $\hat{\mathbf{U}}_0 = \mathbf{I}$ , and then iterates between the two estimates, hence the name ‘‘flip-flop method’’.

With these estimators in hand, one can now carry out a number of different tests. To construct these tests first define

$$\mathbf{\Sigma} := E \text{vec}(\mathbf{\Xi}_n) \text{vec}(\mathbf{\Xi}_n)^\top,$$

where  $\text{vec}$  is the vectorization operator, which stacks the columns of a matrix into a vector. Under  $H_0$  one can show that

$$\mathbf{\Sigma} = \mathbf{V} \otimes \mathbf{U},$$

where  $\otimes$  is the Kronecker product. Constantinou et al. (2017) consider three different approaches

$$\begin{aligned} T_W &= N(\hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\mathbf{\Sigma}})^\top \widehat{\mathbf{W}}^+ (\hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\mathbf{\Sigma}}), \\ T_L &= N(J \log \det(\hat{\mathbf{U}}) + K \log \det(\hat{\mathbf{V}}) - \log \det(\hat{\mathbf{\Sigma}})), \\ T_F &= N \|\hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\mathbf{\Sigma}}\|_F^2. \end{aligned}$$

The first test uses the Moore–Penrose generalized inverse of the estimated covariance matrix of  $\hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\mathbf{\Sigma}}$ ,  $\widehat{\mathbf{W}}^+$ , to normalize the difference and create a Wald type test statistic. The form for  $\widehat{\mathbf{W}}$  is complex, but can be found in Constantinou et al. (2017). The generalized inverse is necessary since the symmetry of the matrices and the non-uniqueness of  $\mathbf{U}$  and  $\mathbf{V}$  create a number of linear constraints. This turns out to be a fairly unstable test statistic in terms of Type 1 error, and works well only for very small  $J$  and  $K$ . The second test is derived from the likelihood ratio and had previously been explored in the multivariate literature in ?. This test is also quite unstable, but only when using a chi-squared distribution as an asymptotic approximation. Mitchell, Genton & Gumpertz (2006) provided a Monte-Carlo method which provides a much more stable (in terms of type 1



error) test. The last test forgoes normalizing the test statistic and instead uses an asymptotic distribution given by a weighted sum of chi-squares. This test is very stable and, in the settings considered by Constantinou et al. (2017) exhibited excellent power. Lastly, we note that Aston et al. (2016) provided a testing procedure that focussed on testing the separability in the directions dictated by the eigenfunctions. They utilized a bootstrap approach instead of asymptotic distributions, and their individual tests can be pooled to construct a global test for evaluating separability overall.

## 6. Spatio-Temporal Extremes

In this section, we summarize the work of (French, Kokoszka, Stoev & Hall 2019) which deals with the computation of probabilities of heat waves. The raw data are spatially-indexed time series of daily temperature measurements,  $X(\mathbf{s}_k, j)$ , denoting the temperature at a spatial location  $\mathbf{s}_k$ ,  $k = 1, \dots, K$ , on day  $j$ . As argued above, due to the natural annual climate cycle, for each site, we partition  $\{X(\mathbf{s}_k, j)\}$  into years, and view the resulting 365-dimensional vectors as samples from a functional time series:

$$X_n(\mathbf{s}_k; \cdot) = \{X_n(\mathbf{s}_k; t_i), i \in \{1, 2, \dots, 365\}\}. \quad (18)$$

Here,  $t \mapsto X_n(\mathbf{s}_k; t)$  is the temperature curve at site  $\mathbf{s}_k$  for year  $n$ , viewed as a function of time  $t$  in days. In contrast to the setting of previous sections, the data used by French et al. (2019) are not historical records, but data generated by a computer climate model. These artificial data are of much higher quality than historical records; there are no gaps, and the daily records are available at 16,100 locations forming a grid covering much of North America. It is, at this point, not clear how to extend the methodology of (French et al. 2019) to historical records. The advantage of using computer model data is that they are predicted future temperatures (French et al. 2019 work with the period 2041-2070), which are more relevant to the prediction of future heat waves. On the other hand, these data do depend on a model, and the poor quality, geostatistical historical records are the real data.

French et al. (2019) propose many functionals that can quantify a heat wave, but here we focus on one specific approach that explains the general idea. A heat wave is characterized by its spatial and temporal extents and by its intensity. The intensity is typically quantified by a threshold. Public health concerns call for a fixed threshold, like 105°F. However, in climate studies of large spatial regions, with many climatic zones, such a fixed threshold is not appropriate. Also the variability of temperatures depends greatly on the geographical location, with coastal locations exhibiting much smaller variability than locations far away from large bodies of water. It is therefore reasonable to work with standardized temperatures

$$Z_n(\mathbf{s}_k, t_i) = \frac{X_n(\mathbf{s}_k, t_i) - \bar{X}(\mathbf{s}_k, t_i)}{\text{SD}(\mathbf{s}_k, t_i)}, \quad (19)$$

where

$$\bar{X}(\mathbf{s}_k, t_i) = \frac{1}{N} \sum_{n=1}^N X_n(\mathbf{s}_k, t_i) \quad \text{and} \quad \text{SD}^2(\mathbf{s}_k, t_i) = \frac{1}{N-1} \sum_{n=1}^N (X_n(\mathbf{s}_k, t_i) - \bar{X}(\mathbf{s}_k, t_i))^2. \quad (20)$$

If the  $Z_n(\mathbf{s}_k, t_i)$  exceed a fixed threshold  $z$ , e.g.  $z = 2$ , for a number of neighboring locations and over a period of time, then we have observed a heat wave (the  $Z_n(\mathbf{s}_k, t_i)$  are practically normal). The severity of a heat wave increases with the size of the region, the duration and the threshold  $z$  that is exceeded. Suppose the  $X_n(\mathbf{s}_k, t_i)$  are maximum daily temperatures, and set

$$Z_n^*(\mathbf{s}_k, t_j) = \frac{1}{\ell} \sum_{t_j - \ell < t_i \leq t_j} Z_n(\mathbf{s}_k, t_i).$$

This is the average maximum temperature over the  $\ell$  days preceding day  $t_j$ . Next, define

$$Z_n^*(t_j) = \min_{1 \leq k \leq K} Z_n^*(\mathbf{s}_k, t_j).$$

If  $Z_n^*(t_j) > z$ , then the average maximum temperature over  $\ell$  days over  $K$  (neighboring) locations exceeds,  $z$ ; this corresponds to a heat wave defined by this specific functional. We are interested in the probability of a heat wave in any given year. We assume that this probability does not depend on year  $n$ . We thus want to compute, for some relevant  $z > 0$ ,

$$p(z) = P(\exists j : Z_n(t_j) > z) = P\left(\max_{1 \leq j \leq J} Z_n^*(t_j) > z\right) = P(M_J > z),$$

where  $J = 365$ , and

$$M_J \stackrel{d}{=} M_{J,n} := \max_{1 \leq j \leq J} Z_n^*(t_j).$$

The concatenated sequence  $Z^*(t_j)$  is stationary and weakly dependent, so (see e.g. Beirlant, Goegebeur, Segers & Teugels 2006, Chapter 10), there are sequences  $a_J$  and  $b_J$  such that

$$\lim_{J \rightarrow \infty} P\left(\frac{M_J - b_J}{a_J} \leq z\right) \rightarrow H(z),$$

where  $H$  is a univariate Generalized Extreme Value distribution function. The function  $H$  depends on three parameters, which can be estimated, together with the constants  $a_J$  and  $b_J$ , using now standard **R** implementations.

Figure 5 shows examples of regions corresponding to 50, 150 and 450 neighboring locations. Figure 6 shows a map of the probability of a heat wave for  $d = 50$  for three durations  $\ell$ , with (a) corresponding to  $\ell = 3$ , (b) to  $\ell = 10$ , and (c) to  $\ell = 30$ . When  $\ell = 3$ , there is a surprisingly high probability of localized heat waves over the Labrador Peninsula. Such short heat spells may occur with probability approaching 50%, that is on average every second year. While our EVT approximation may break down for such high probabilities, it is nevertheless obvious that that part of Canada will see heat spells much more frequently than in the past. Generally, we see that the area around the Hudson Bay will experience an

increased frequency of hot spells lasting a few days. There is a noticeable drop in the probability of such a heat wave around the Rocky Mountain range. The probability is also very low along the Eastern seaboard of the United States. Increasing the duration to  $\ell = 10$  days, dramatically reduces the probability of a heat wave of the corresponding magnitude. The reader will note the different probability scale. Many parts of Canada once again show an increased probability of a heat wave of this magnitude, as well as parts of Iowa and Illinois, certain regions in Texas, and, most visibly, the Pacific Ocean off the Southern California coast. Increasing the duration to approximately 1 month ( $\ell = 30$ ), causes the probability of a heat wave to drop even further; generally, throughout North America, heat waves of this magnitude will occur with probability of less than 1%, i.e. once per one hundred years, on average. Over the Canadian plains and the Canadian Rockies, this probability increases only slightly to about 1.5%. There are two patches, in Arizona and Southern Texas, with probabilities elevated to 2-3%.

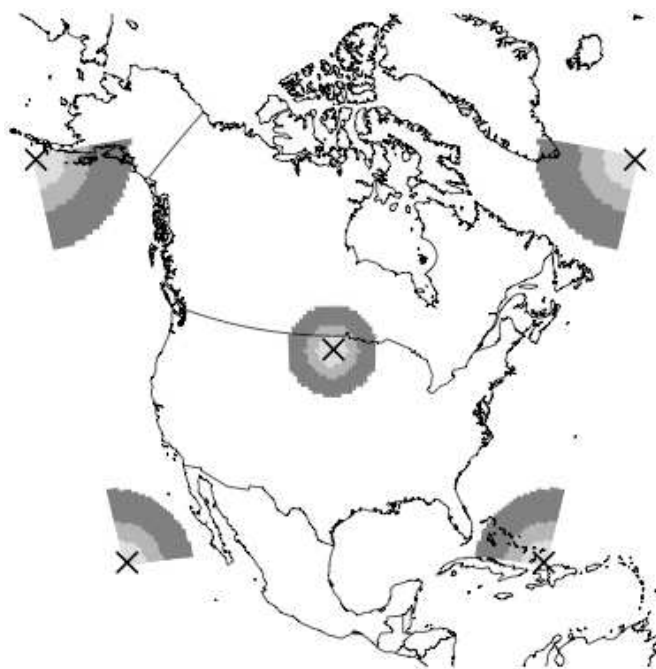


FIGURE 5: A map of the neighborhood structures for different locations using 50, 150, and 450 nearest neighbors. Each  $\times$  marks a neighborhood centroid and the sequences of grey shading mark the extents of the increasing neighborhood sizes.

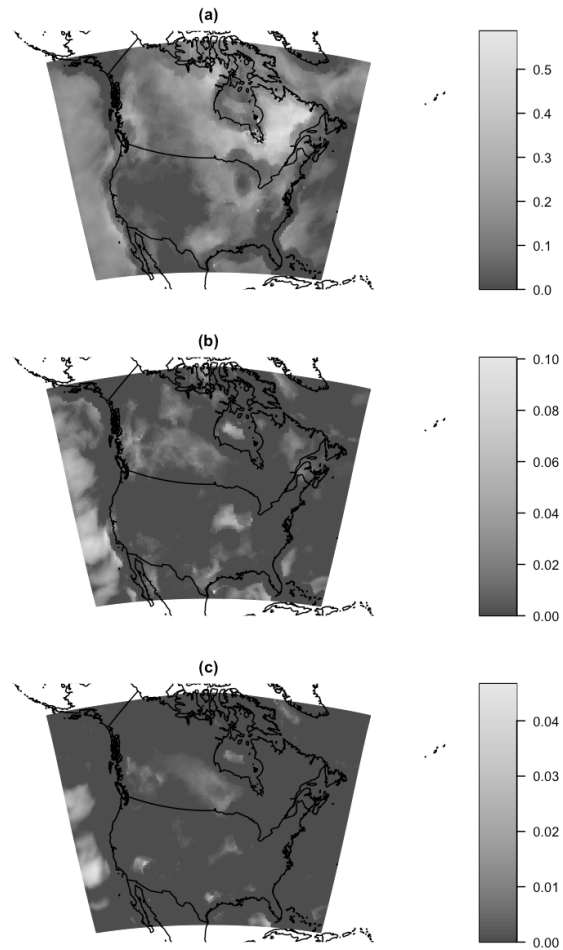


FIGURE 6: Probability of a heat wave with amplitude more than 2 standard deviations above the mean for spatial extent  $d = 50$  and durations of (a)  $\ell = 3$ , (b)  $\ell = 10$ , and (c)  $\ell = 30$ .

## 7. Conclusions

This article has focused on aspects of the statistical analysis of functions and time series of functions distributed over space that have attracted our interest. No claim is made that this survey is complete. There have been developments in the analysis of functional data arising in brain and other medical studies that have important spacial aspects, but do not readily fit into the more traditional and well understood paradigms of spatial statistics. One can expect that as new data sets consisting of functions that exhibit some form of spatial dependence arise, new tools may need to be developed to extract useful information.

## Acknowledgements

The research presented in this paper has been partially supported by the United States National Science Foundation grants DMS-0931948 *Spatio-temporal Statistical Framework for the Evaluation of Global Change in the Ionosphere* and DMS-1462067 *FRG: Collaborative Research: Extreme Value Theory for Spatially Indexed Functional Data*.

[Received: January 2017 — Accepted: November 2018]

## References

- Aston, J., Pigoli, D. & Tavakoli, S. (2016), ‘Tests for separability in nonparametric covariance operators of random surfaces’, *The Annals of Statistics* **6**, 1906–1948.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2006), *Statistics of Extremes: Theory and Applications*, John Wiley & Sons.
- Caballero, W., Giraldo, R. & Mateu, J. (2013), ‘A universal kriging approach for spatial functional data’, *Stochastic Environmental Research and Risk Assessment* **27**, 1553–1563.
- Constantinou, P., Kokoszka, P. & Reimherr, M. (2017), ‘Testing separability of space-time functional processes’, *Biometrika* **104**, 425–437.
- Csörgő, M. & Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, Wiley.
- Delicado, P., Giraldo, R., Comas, C. & Mateu, J. (2010), ‘Statistics for spatial functional data: some recent contributions’, *Environmetrics* **21**, 224–239.
- French, J., Kokoszka, P., Stoev, S. & Hall, L. (2019), ‘Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data’, *Computational Statistics and Data Analysis* **131**, 176–193.
- Gelfand, A. E., Diggle, P. J., Fuentes, M. & Guttorp, P., eds (2010), *Handbook of Spatial Statistics*, CRC Press.
- Gromenko, O. & Kokoszka, P. (2012), ‘Testing the equality of mean functions of spatially distributed curves’, *Journal of the Royal Statistical Society (C)* **61**, 715–731.
- Gromenko, O. & Kokoszka, P. (2013), ‘Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination’, *Computational Statistics and Data Analysis* **59**, 82–94.
- Gromenko, O., Kokoszka, P. & Reimherr, M. (2017), ‘Detection of change in the spatiotemporal mean function’, *Journal of the Royal Statistical Society (B)* **79**, 29–50.

- Gromenko, O., Kokoszka, P. & Sojka, J. (2017), ‘Evaluation of the global cooling trend in the ionosphere using functional regression models with incomplete curves’, *The Annals of Applied Statistics* **11**.
- Gromenko, O., Kokoszka, P., Zhu, L. & Sojka, J. (2012), ‘Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends’, *The Annals of Applied Statistics* **6**, 669–696.
- Hörmann, S. & Kokoszka, P. (2013), ‘Consistency of the mean and the principal components of spatially distributed functional data’, *Bernoulli* **19**, 1535–1558.
- Horváth, L. & Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer.
- Kokoszka, P. & Reimherr, M. (2017), *Introduction to Functional Data Analysis*, CRC Press.
- Liu, C., Ray, S. & Hooker, G. (2017), ‘Functional principal components analysis of spatially correlated data’, *Statistics and Computing* **27**, 1639–1654.
- Lu, N. & Zimmerman, D. (2005), ‘The likelihood ratio test for a separable covariance matrix’, *Statistics & Probability Letters* **73**, 449–457.
- Menafoglio, A., Secchi, P. & Rosa, M. D. (2013), ‘A universal kriging predictor for spatially dependent functional data of a hilbert space’, *Electronic Journal of Statistics* **7**, 2209–2240.
- Mitchell, M. W., Genton, M. G. & Gumpertz, M. L. (2006), ‘A likelihood ratio test for separability of covariances’, *Journal of Multivariate Analysis* **97**, 1025–1043.
- Rishbeth, H. (1990), ‘A greenhouse effect in the ionosphere?’, *Planetary and Space Science* **38**, 945–948.
- Roble, R. G. & Dickinson, R. E. (1989), ‘How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and thermosphere?’, **16**, 1441–1444.
- Schabenberger, O. & Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC.
- Sherman, M. (2011), *Spatial Statistics and Spatio–Temporal Data: Covariance Functions and Directional Properties*, Wiley.
- Wackernagel, H. (2003), *Multivariate Geostatistics*, Springer.