



---

**Título artículo / Títol article:** Hierarchical clustering of spatially correlated functional data

**Autores / Autors** Ramón Giraldo, Pedro Delicado, Jorge Mateu

**Revista:** *Statistica Neerlandica* (2012), volume 66, issue 4

**Versión / Versió:** Versió pre-print

**Cita bibliográfica / Cita bibliogràfica (ISO 690):** GIRALDO, Ramón; DELICADO, Pedro; MATEU, Jorge. Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica*, 2012, vol. 66, no 4, p. 403-421.

**url Repositori UJI:** <http://hdl.handle.net/10234/76907>

---

# Hierarchical clustering of spatially correlated functional data

Ramón Giraldo, Pedro Delicado and Jorge Mateu

**Reporte Interno de Investigación No. 12**

**Departamento de Estadística**

**Facultad de Ciencias**

**Universidad Nacional de Colombia**

**Bogotá, COLOMBIA**

## Abstract

Classification problems of functional data arise naturally in many applications. Several approaches have been considered for solving the problem of finding groups based on functional data. In this paper we are interested in detecting groups when the functional data are spatially correlated. Our methodology allows to find spatially homogeneous groups of sites when the observations at each sampling location consist of samples of random functions. In univariable and multivariable geostatistics various methods of incorporating spatial information into the clustering analysis have been considered. Here we extend these methods to the functional context in order to fulfill the task of clustering spatially correlated curves. In our approach we initially use basis functions to smooth the observed data and then we weight the dissimilarity matrix among curves by either the trace-variogram or the multivariable variogram calculated with the coefficients of the basis functions. As an illustration the methodology is applied to a real data set corresponding to average daily temperatures measured at 35 Canadian weather stations.

**Keywords:** Basis functions; Dissimilarity; Functional clustering; Geostatistics; Multi-variable variogram; Trace-variogram.

---

Research supported in part by the Spanish Ministry of Education and Science and FEDER through grants MTM2007-62923 and MTM2006-09920. The first author thanks to FUNDACION CAROLINA for financial support through a PhD scholarship.

Authors' Address: R. Giraldo is Associate Professor, Department of Statistics, Universidad Nacional de Colombia (e-mail: rgiraldoh@unal.edu.co). P. Delicado is Associate Professor, Department of Statistics and Operation Research, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. J. Mateu is Professor of Statistics, Department of Mathematics, Universitat Jaume I, E-12071 Castellón, Spain .

## 1. Introduction

Cluster analysis (unsupervised classification) provides an analytical technique for finding meaningful subgroups of individuals or objects. Specifically, the objective is to classify a sample of entities (individuals or objects) into a small number of mutually exclusive groups based on the similarities among the entities. In cluster analysis, unlike discriminant analysis, the groups are not predefined. Instead, the technique is used to identify the groups. There are two major methods of clustering: hierarchical clustering and  $K$ -means clustering. The first one involves the construction of a hierarchy of a treelike structure using agglomerative or divisive procedures. In contrast,  $K$ -means methods assign objects into clusters once the number of clusters to be formed is specified (Hair, Anderson, Tatham, and Black 1995).

Classification methods have been adapted to the case of geographically referenced samples in order to delineate relatively contiguous zones with similar attribute values (Bourgault, Journel, Rhoades, Corwin, and Lesch 1997). Oliver and Webster (1989), the first paper in this context, propose the use of the variogram (in a univariate case) or the variogram of the first principal component (in the multivariate case) to weight the dissimilarities between samples. In the same sense Bourgault, Marcotte, and Legendre (1992) propose using the multivariate variogram as a weighting function to increase dissimilarities among distant samples. In that paper the authors classify 147 rock samples according to six chemical variables and conclude that including spatial information into the analysis allows to obtain better results in agreement with known geology. An overview about classification for geographically referenced data is given in Atkinson and Lewis (2006), where an example for remote sensing is shown.

Functional data refer to data which consist of observed functions or curves evaluated at a finite subset of some interval. Functional versions for many branches of statistics have been given. An overview of statistical methods for analyzing functional data is shown in Ramsay and Silverman (2005), and recent developments in this field are given in special issues of several journals (González-Manteiga and Vieu 2007; Valderrama 2007). Cluster analysis has also been considered from a functional point of view. Approaches based on hierarchical,  $K$ -means and more formal model-based procedures have been studied in this setting. In the analysis of functional data, the vectors observed are usually converted into curves via a smoothing method. In clustering functional data, it is useful to smooth the observations and cluster smoothed curves rather than the observed data (Hitchcock, Booth, and Casella 2007). In the case of hierarchical clustering, standard methods can be invoked using the coefficients of basis functions as input (Hitchcock et al. 2007). Henderson (2006) shows an application of hierarchical clustering for functional data in a water quality study. Clarkson, Fraley, Gu, and Ramsay (2005) illustrate the use of this

technique classifying Canadian weather stations based on precipitation and temperature curves. Abraham, Cornillon, Matzner-Llober, and Mollinary (2003) and Tarpey and Kinateder (2003) consider  $K$ -means algorithms for functional data based on B-splines and Fourier basis, respectively. García-Escudero and Gordaliza (2005) and Cuesta-Albertos and Fraiman (2007) propose the use of trimmed  $K$ -means as a robust alternative to  $K$ -means. Other model-based approaches for clustering functional data are given in James and Sugar (2003), Lee (2004) and Chiou and Lee (2007), respectively.

The above mentioned statistical techniques for clustering functional data are focused on independent functions. However, in several disciplines of applied sciences there exists an increasing interest for modeling correlated functional data: it is the case when samples of functions are observed over a discrete set of time points (*temporally correlated functional data*) or when these functions are observed in different sites of a region (*spatially correlated functional data*). In this paper we combine hierarchical clustering methods for both geographically referenced data and functional data in order to give a solution to the problem of classifying spatially correlated curves. Our methodology allows to find groups of curves which are spatially homogeneous. We propose to weight the dissimilarities between curves by the trace-variogram (Giraldo, Delicado, and Mateu 2007) and by the multivariate variogram (Bourgault and Marcotte 1991) of the coefficients of basis functions used to smooth the observed data.

The paper is organized as follows. Section presents an overview on hierarchical clustering for both functional data and georeferenced data. We show how these methodologies can be joined together to carry out classification of spatially correlated curves. An application of the proposed method is given in Section . The paper ends with a brief discussion and suggestions for further research.

## 2. Hierarchical clustering of spatially correlated functional data

When considering the hierarchical clustering approach, a series of partitions takes place running from a single cluster containing all objects to  $n$  clusters each containing a single object. The simplest method to apply hierarchical clustering to real data is calculating a distance (dissimilarity) matrix between individuals and applying some agglomerative (or divisive) criterium (such as single, complete or average linkage) for finding groups of individuals (see for instance Everitt, Landau, and Leese 2001). The agglomerative techniques proceed by series of fusions of the  $n$  objects into groups, while the divisive methods separate the  $n$  objects successively into finer groupings. The results may be represented by a two dimensional diagram known as dendrogram which illustrates the fusions or divisions made at each successive stage of analysis. In this section we describe how this procedure

is extended to functional data, and how it can be adapted to a geostatistical setting. Then we show how these methods can be combined together for carrying out hierarchical classification of spatially correlated functional data.

Let us begin by considering first the case of clustering functional data. Hitchcock, Casella, and Booth (2006) show that in this context it is advantageous to use not the observed data, but rather a pre-smoothed version of the data. Suppose we have a sample of curves  $X_1(t), \dots, X_n(t)$  defined for  $t \in [a, b] \subset \mathbf{R}$ . It is assumed that these curves belong to the separable Hilbert space  $H$  of square integrable functions defined on  $[a, b]$ . We assume that the functions are expanded in terms of some basis functions by

$$X_i(t) = \sum_{l=1}^K a_{il} B_l(t) = \mathbf{a}_i^T \mathbf{B}(t), \quad i = 1, \dots, n. \quad (1)$$

A functional hierarchical cluster analysis is then developed as in the classical setting but considering instead the distance between curves  $X_i(t)$  and  $X_j(t)$  through the  $L_2$  norm

$$d_{ij} = \sqrt{\int_{[a,b]} (X_i(t) - X_j(t))^2 dt}.$$

Using the expansion (1) we have

$$\begin{aligned} d_{ij} &= \sqrt{\int_{[a,b]} (\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{B}(t) \mathbf{B}(t)^T (\mathbf{a}_i - \mathbf{a}_j) dt} \\ &= \sqrt{(\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{W} (\mathbf{a}_i - \mathbf{a}_j)}, \end{aligned} \quad (2)$$

where

$$\mathbf{W} = \int_{[a,b]} \mathbf{B}(t) \mathbf{B}(t)^T dt,$$

and  $\mathbf{a}_i$  and  $\mathbf{a}_j$  are vectors of basis coefficients for the  $i$ th and  $j$ th individuals. For any orthonormal basis such as the Fourier basis, the Gram matrix  $\mathbf{W}$  is the identity matrix. For other basis functions such as B-Splines,  $\mathbf{W}$  must be calculated by numerical integration. Once calculated the dissimilarity matrix, a standard agglomerative or divisive hierarchical clustering procedure can be applied.

In the case of geostatistical setting, clustering allows to find groups of contiguous sites with similar variable values. An overview of classification methods (supervised and unsupervised) in geostatistics is given in Atkinson and Lewis (2006).

Let  $\{\mathbf{Z}(x) = (Z_1(x), \dots, Z_m(x)) : x \in D\}$  be a  $m$  multivariable spatial process defined over a domain  $D \subset \mathbf{R}^d$ . The first method for clustering in geostatistics was proposed

by Oliver and Webster (1989). Their solution in the case of  $m = 1$  was to weight the dissimilarities  $d_{ij}$  between samples by

$$d_{ij}^w = d_{ij}\gamma(h), \quad (3)$$

where  $\gamma(h)$  corresponds to the variogram calculated for the distance between sites  $i, j$ . For  $m > 1$ , these authors propose using the variogram of the first principal component or a sum of variograms of the first few principal components. On the other hand Bourgault et al. (1992) generalize this idea by using

$$d_{ij}^w = d_{ij}\Gamma(h), \quad (4)$$

where  $\Gamma(h)$  is the multivariate variogram defined by

$$\Gamma(h) = \frac{1}{2}\mathbb{E}(\mathbf{Z}(x) - \mathbf{Z}(x+h))^T \mathbf{M}(\mathbf{Z}(x) - \mathbf{Z}(x+h)), \quad (5)$$

with  $\mathbf{M}$  a symmetric positive definite matrix used as a metric. In the particular case of  $\mathbf{M} = \mathbf{I}$ , the multivariate variogram defined in equation (5) is given by

$$\begin{aligned} \Gamma(h) &= \sum_{l=1}^m \frac{1}{2}\mathbb{E}(Z_l(x) - Z_l(x+h))^2 \\ &= \sum_{l=1}^m \gamma_l(h), \end{aligned} \quad (6)$$

where  $\gamma_l(h)$  is the variogram for the  $l$ -th variable. Another alternative for the matrix  $\mathbf{M}$  in (5) is the inverse of the variance-covariance matrix  $\Sigma^{-1}$ . In that case the multivariate variogram in equation (5) is a weighted sum of direct (single) variograms and cross-variograms.

Our aim is extending the approaches of Oliver and Webster (1992) and Bourgault et al. (1992) to the context of spatially correlated functional data. Let  $\{X_s(t) : s \in D \subset \mathbf{R}^d, t \in [a, b] \subset \mathbf{R}\}$  be a stationary isotropic functional random process. Let  $X_1(t), \dots, X_n(t)$  be a realization of this random process observed at  $n$  sites with coordinates  $x_1, \dots, x_n$ , respectively. We assume that these curves belong to the separable Hilbert space  $H$  of square integrable functions defined on  $[a, b]$ . We propose two alternatives for performing cluster analysis with this type of information. The first one consists in replacing the variogram in equation (3) by the trace-variogram function defined by Giraldo et al. (2007)

$$\gamma(h) = \frac{1}{2}\mathbb{E} \left[ \int_{[a,b]} (X_i(t) - X_j(t))^2 dt \right], \text{ with } h = \|x_i - x_j\|. \quad (7)$$

The function given in equation (7) can be estimated by using the classical method-of-moments by means of

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_{[a,b]} (X_i(t) - X_j(t))^2 dt, \quad (8)$$

where  $N(h) = \{(x_i, x_j) : \|x_i - x_j\| = h\}$ , and  $|N(h)|$  is the number of distinct elements in  $N(h)$ . For irregularly spaced data there are generally not enough observations separated by exactly  $h$ . Then  $N(h)$  is modified to  $\{(x_i, x_j) : \|x_i - x_j\| \in (h - \varepsilon, h + \varepsilon)\}$ , with  $\varepsilon > 0$  being a small value. Once we have estimated the trace-variogram for a sequence of  $K$  values  $h_k$ , we fit a parametric model  $\gamma_\alpha(h)$  (any of the classical and widely used models such as spherical, Gaussian, exponential or Matérn could well be used) to the points  $(h_k, \hat{\gamma}(h_k))$ ,  $k = 1, \dots, K$ , as if they were obtained under the classical geostatistical setting. Usually, this type of fitting is done by ordinary least squares (OLS) or weighted least squares (WLS) (see, for instance, Cressie, 1993). Note that the fitted parametric trace-variogram  $\gamma_{\hat{\alpha}}(h)$  is always a valid variogram because its properties are those of a parametric variogram fitted from a univariate geostatistical data set.

From equation (2) it is clear that the distance between two curves can be calculated from the distance between the coefficients of the basis functions. Consequently a second alternative for clustering spatially correlated functional data is estimating variograms and cross-variograms of coefficients of basis functions used for smoothing the observed data and applying the approach by Bourgault et al. (1992). Assuming that the curve for each sampling site  $i$ ,  $i = 1, \dots, n$ , is expanded by using (1), the coefficients in matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{pmatrix}_{(n \times K)}$$

form a realization of a  $K$  multivariable random field  $\{\mathbf{A}(x) = (A_1(x), \dots, A_K(x)) : x \in D \subset \mathbf{R}^d\}$  with  $\mathbb{E}(A_i(x)) = v_i$ , and variograms and cross-variograms matrix

$$\mathbf{\Upsilon}(h) = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \cdots & \gamma_{1K}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \cdots & \gamma_{2K}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{K1}(h) & \gamma_{K2}(h) & \cdots & \gamma_{KK}(h) \end{pmatrix}_{(K \times K)}, \quad (9)$$

where  $\gamma_{lq}(h) = \frac{1}{2} \mathbb{E} (A_l(x_i) - A_q(x_j))^2$ ,  $l, q = 1, \dots, K$ ,  $h = \|x_i - x_j\|$ . We propose using multivariable geostatistics (Wackernagel 1995) and specifically a linear model of coregionalization (LMC) for estimating the matrix (9). Based on this estimation we can calculate



the multivariate variogram of the coefficients of basis functions taking the trace of matrix (9), that is, calculating

$$\begin{aligned}\Gamma(h) &= \sum_{l=1}^K \frac{1}{2} \mathbb{E} (A_l(x) - A_l(x+h))^2 \\ &= \sum_{l=1}^K \gamma_u(h).\end{aligned}\tag{10}$$

Note that equation (10) coincides with equation (6) replacing the multivariable random field  $\{\mathbf{Z}(x)\}$  by  $\{\mathbf{A}(x)\}$ . Thus applying the Bourgault et al. (1992) method, the spatial weighted distance between two sites is given by equation (4) with  $d_{ij}$  calculated with equation (2) and  $\Gamma(h)$  obtained from equation (10).

### 3. Application: Temperature curves in Canada

The methodology is illustrated through an analysis of a well-known meteorological data set in FDA consisting of daily average (over the years 1960 to 1994) temperature measurements recorded at 35 weather stations of Canada (Figure 1, right panel). The data were obtained from Ramsay and Silverman's home page (<http://www.functionaldata.org/>). The geographical coordinates of weather stations (Figure 1, left panel) were obtained from the Meteorological Service of Canada (<http://www.climate.weatheroffice.ec.gc.ca/climateData/>). Following Ramsay and Silverman (2005) and Giraldo (2009) a Fourier basis with 65 functions is used to smooth the discrete temperature values observed at each weather station. The very large spatial area covered by this data set makes it difficult to apply statistical spatial tools based on the assumption of stationarity. For instance, distances (in order of thousands of kilometers) in latitude imply bigger weather differences than similar distances in longitude. Taking into account that we assume stationarity and isotropy in the models we introduce in this paper, we remove the spatial trend in a first step of the analysis. The smoothed data were detrended by using a functional regression model (Ramsay and Silverman 2005) with functional response (smoothed temperature curves) and two scalar covariates (longitude and latitude coordinates in decimal degrees), that is, we considered the functional linear model

$$X_i(t) = \hat{\alpha}(t) + \hat{\beta}_1(t)\text{Longitude}_i + \hat{\beta}_2(t)\text{Latitude}_i + e_i(t).\tag{11}$$

Then, a Fourier basis with 65 functions was used for smoothing the functional residuals given in equation (11). Based on the smoothed residual functions and by using the estimator in equation (8), the trace-variogram was calculated for several spatial lags. A spherical

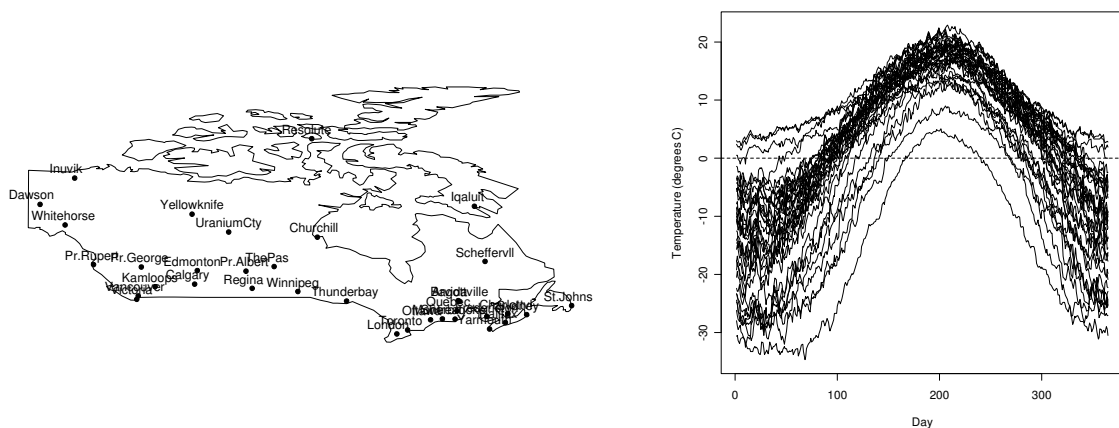


Figure 1: Averages (over 30 years) of mean daily temperature curves (*right panel*) recorded at 35 Canadian weather stations (*left panel*).

model (Cressie 1993) was fitted to the estimated trace-variogram by using the ordinary least square technique. The spherical model has been widely used in the literature, and was chosen in terms of its flexibility, simplicity and interpretability of its parameters. The estimated spherical model for the trace-variogram, was as follows: the nugget was equal to 0, the partial sill equal to  $\sigma^2 = 7769$ , and the range was  $\phi = 2184$ . So the estimated trace-variogram was  $\hat{\gamma}(h) = 7769(1.5h/2184 - 0.5(h/2184)^3)$  for  $h \leq 2184$ , and  $\hat{\gamma}(h) = 7769$  for  $h > 2184$ . Interpreting the range of the fitted model as in a classic geostatistical setting, we can say that sites separated up to 2184 kilometers are still correlated.

A LMC was fitted to the multivariable random field consisting of coefficients of the Fourier basis functions used for smoothing the residuals in equation (11). All single (direct) variograms and cross-variograms were modeled as a linear combination of nugget and exponential models. Based on the fitted LMC, we estimated the multivariate variogram by means of equation (5).

We performed three hierarchical functional cluster analysis. The first one was based on the matrix of Euclidean distances among the coefficients of the Fourier basis functions used for smoothing the temperature data (using equation 2), and the remaining ones by weighting this dissimilarity matrix by the trace-variogram and by the multivariate variogram obtained with the coefficients of the Fourier basis used for smoothing the residuals in equation (11). We used complete linkage as the agglomeration method. Complete linkage was also used by Clarkson et al. (2005) for clustering the same data set as considered here. Two threshold values were considered to obtain the clusters in each case. The first one (the largest threshold value) was obtained when applying the cluster quality measures of Davies-Bouldin, Calinsky-Harabasz, Hubert-Levine and Silhouette, respectively (Davies and Bouldin 1979; Milligan and Cooper 1985). In all cases these measures suggested two

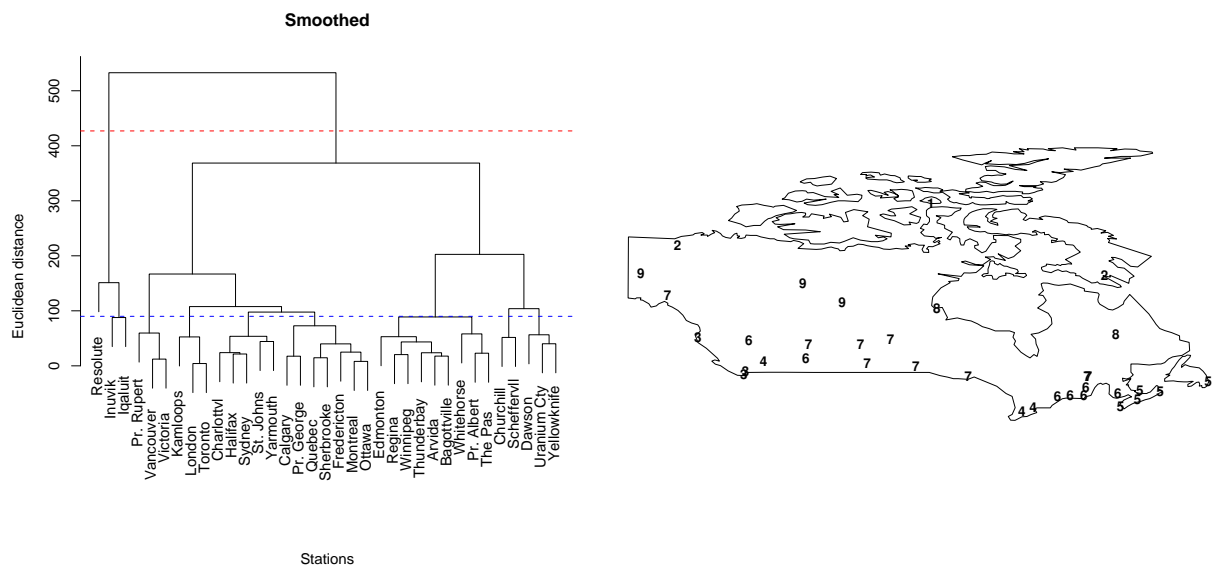


Figure 2: Left panel: Hierarchical clustering of temperature curves recorded at 35 weather stations of Canada. Right panel: Spatial distribution of stations according to the division into 9 groups.

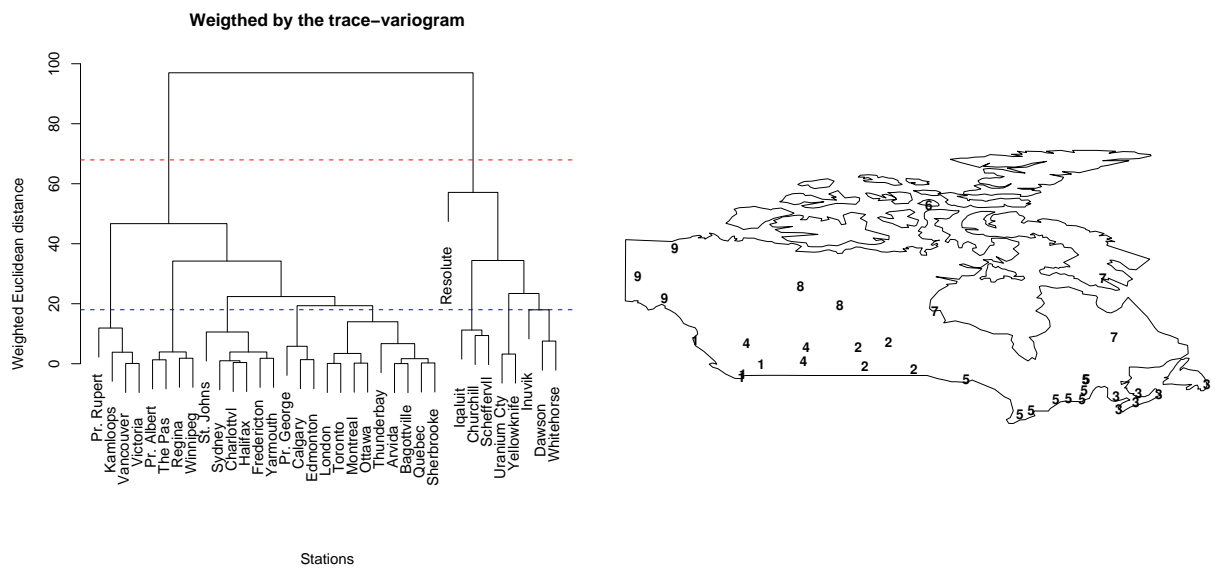


Figure 3: Left panel: Hierarchical clustering by weighting the distance matrix by the trace-variogram. Right panel: Spatial distribution of stations according to the division into 9 groups.

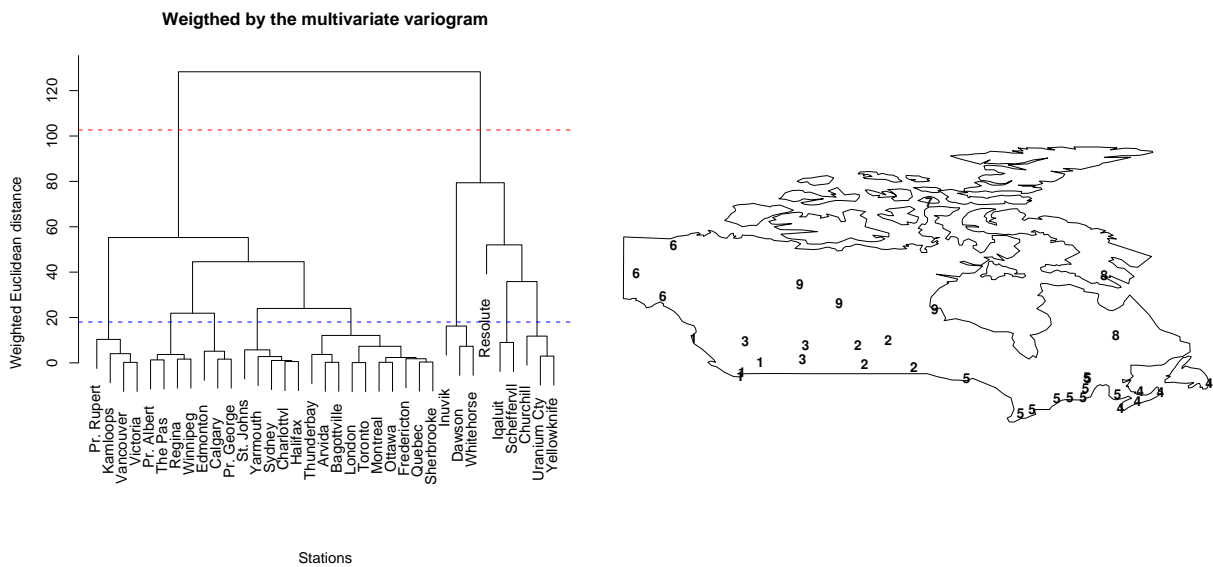


Figure 4: Left panel: Hierarchical clustering obtained by weighting the distance matrix by the multivariate variogram. Right panel: Spatial distribution of stations according to the division into 9 groups.

clusters. Cutting the dendrograms into two groups we could not identify clearly the effect of spatial correlation into the classification. For this reason in each case we chose visually a second threshold (the lowest threshold value) in order to obtain a more detailed partition (9 groups in all cases).

The dendrograms obtained are shown in Figures 2, 3 and 4. When we cut the dendrograms at the largest threshold, we can observe that in the classical functional cluster analysis (left panel in Figure 2) the stations Resolute, Inuvik and Iqaluit (the far northern stations) are very separated from the remaining ones. If we consider spatial weighting (left panel in Figures 3 and 4) these stations are clustered with other relatively closer stations such as Schererville (911 kilometers to Iqaluit) or Dawson (705 kilometers to Inuvik). These results reflect the effect of including the spatial weighting into the analysis.

To check if the clustering results make sense from a spatial point of view, we show the spatial distribution of clusters (right panels of Figures 2, 3 and 4), when we cut the dendrograms at the lowest threshold. We can observe that in general the clusters consist of stations that are located in the same region. However some exceptions to this rule are obtained when the classification is obtained by classical functional cluster (right panel in Figure 2). For instance if we compare the map in Figure 1 with the labels in the right panel of Figure 2 we note that Whitehorse (label 7 in Figure 2) in western Canada is assigned to the same cluster of other very separated stations (more than 4500 kilometers) such as

Thunder Bay in norther Ontario or Arvida and Bagotville in northeastern Quebec. In the same sense, Kamloops (label 4 in Figure 2) in the Pacific region (south central British Columbia) is assigned to the same cluster of stations separated more than 4000 kilometers as London and Toronto in southwestern Ontario. In summary, the comparison of the three analysis allows to conclude that the use of a spatial weighting is reflected in more spatially homogeneous groups. Clarkson et al. (2005) defined 7 groups of stations. These authors conclude that some cluster assignment makes sense (e.g., Pr Rupert, Vancouver and Victoria in the same cluster), as well as clusters which are difficult to interpret (e.g, Dawson in Western Canada in the same cluster of Regina or Winnipeg). In our analysis the use of a spatial weighting allows to give a solution to this problem.

We can also observe from Figures 3 and 4 that there are few differences between the clusters obtained with the two methods based on spatial weighting. When we weight the dissimilarity matrix by the trace-variogram, the station Churchill (see Figure 1) with label 7 in right panel of Figure 3 is assigned to the same cluster of Uranium City and Yellowknife (see Figures 1 and 3), whereas this station is assigned to the same cluster of Schefferville and Iqaluit when we use the multivariate variogram as weighting function (see Figures 1 and 4). Taking into account that Churchill is closer to Uranium City (around 1400 kilometers) than to Schefferville (2753 kilometers) or Iqaluit (4059 kilometers) it could be more reasonable to assign this station to its cluster in Figure 3. However a discussion from a meteorological point of view is obviously important in this case in order to make an appropriate assignment.

Another way of detrending the mean in FDA is by using a functional analysis of variance (FANOVA). In Chapter 13 of Ramsay and Silverman (2005), a FANOVA is performed with the Canadian temperature data set where a factor is defined by dividing Canada into Atlantic, Continental, Pacific and Artic meteorological zones and assigning each station to some of these zones. We also applied our proposal but considering the residuals of the FANOVA instead of those obtained with the functional regression model given in equation (11). The results in this case are completely analogous to those shown above, that is, we achieve to define groups of stations with similar temperature curves and located in zones relatively contiguos.

#### 4. Conclusions and further research

A methodology for clustering spatially correlated functional data has been proposed. Our proposal combines both geostatistical cluster analysis and functional cluster analysis in order to give a solution to the problem of detecting spatially homogeneous clusters of curves. Our approach has been applied to a climatological data set. The results show a

good performance of the proposed methodology, indicating from a descriptive point of view that this one can be used as a classification method when we need to take into account spatial correlation in cluster analysis of spatially referenced samples of curves.

We have only considered hierarchical clustering analysis. Alternative clustering techniques for spatially correlated functional data should also be considered. In addition, the use of other basis functions such as B-splines could also be considered.

# Bibliography

- ABRAHAM, C., CORNILLON, P., MATZNER-LLOBER, E., and MOLLINARY, N. (2003), “Unsupervised curve-clustering using B-splines,” *Scandinavian Journal of Statistics*, **50**, 581–595.
- ATKINSON, P. and LEWIS, P. (2006), “Geostatistical classification for remote sensing: an introduction,” *Computers & Geosciences*, **26**, 361–371.
- BOURGAULT, G., JOURNEL, A., RHOADES, D., CORWIN, D., and LESCH, S. (1997), “Geostatistical analysis of a soil salinity data set,” in *Advances in Agronomy*, Leiden: Academic Press, vol. 58, pp. 1–14.
- BOURGAULT, G. and MARCOTTE, D. (1991), “The multivariable variogram and its application to the linear coregionalization model,” *Mathematical Geology*, **23**, 899–928.
- BOURGAULT, G., MARCOTTE, D., and LEGENDRE, P. (1992), “The multivariable covariogram as a spatial weighting function in classification methods,” *Mathematical Geology*, **24**, 463–478.
- CHIOU, J. and LI, P. (2007), “Functional clustering and identifying substructures of longitudinal data,” *Journal Royal Statistical Society, Series B*, **69**, 679–699.
- CLARKSON, D., FRALEY, C., GU, C., and RAMSAY, J. (2005), *S+ Functional Data Analysis User Guide*, New York: Springer.
- CRESSIE, N. (1993), *Statistic for Spatial Data*, New York: John Wiley & Sons.
- CUESTA-ALBERTOS, J. and FRAIMAN, R. (2007), “Impartial trimmed k-means for functional data,” *Computational Statistics & Data Analysis*, **51**, 4814–4831.
- DAVIES, D. and BOULDIN, D. (1979), “A cluster separation measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, 224–227.
- EVERITT, B., LANDAU, S., and LEESE, M. (2001), *Cluster Analysis*, New York: Oxford University Press, Inc., 4th ed.

- GARCÍA-ESCUADERO, L. and GORDALIZA, A. (2005), “A proposal for robust curve clustering,” *Journal of Classification*, **22**, 185–201.
- GIRALDO, R. (2009), “Geostatistical Analysis of Functional Data,” Ph.D. thesis, Universitat Politècnica de Catalunya.
- GIRALDO, R., DELICADO, P., and MATEU, J. (2007), “Geostatistics for functional data: An ordinary kriging approach,” Tech. Rep. <http://hdl.handle.net/2117/1099>, Universitat Politècnica de Catalunya.
- GONZÁLEZ-MANTEIGA, W. and VIEU, P. (2007), “Statistics for functional data,” *Computational Statistics & Data Analysis*, **51**, 4788–4792.
- HAIR, J., ANDERSON, R., TATHAM, R., and BLACK, W. (1995), *Multivariate Data Analysis*, New Jersey: Prentice Hall.
- HENDERSON, B. (2006), “Exploring between site in water quality trends: a functional data analysis approach,” *Environmetrics*, **17**, 65–80.
- HITCHCOCK, D., BOOTH, J., and CASELLA, G. (2007), “The effect of pre-smoothing functional data on clustering analysis,” *Journal of Statistical Computation and Simulation*, **77**, 1043–1055.
- HITCHCOCK, D., CASELLA, G., and BOOTH, J. (2006), “Improved estimation of dissimilarities by presmoothing functional data,” *Journal of the American Statistical Association*, **101**, 211–222.
- JAMES, G. and SUGAR, A. (2003), “Clustering for sparsely sampled functional data,” *Journal of the American Statistical Association*, **98**, 397–408.
- LE, N. and ZIDEK, J. (1992), “Interpolation with uncertain spatial covariances: A bayesian alternative of kriging,” *Journal of Multivariate Analysis*, **43**, 357–374.
- LEE, H. (2004), “Functional Data Analysis: Classification and Regression,” Ph.D. thesis, Texas A&M University.
- MILLIGAN, G. and COOPER, M. (1985), “An examination of procedures of determining the number of clusters in a data set,” *Psychometrika*, **50**, 159–179.
- OLIVER, M. and WEBSTER, R. (1989), “A geostatistical basis for spatial weighting in multivariate classification,” *Mathematical Geology*, **21**, 15–35.
- RAMSAY, J. and SILVERMAN, B. (2005), *Functional Data Analysis. Second edition*, New York: Springer.



- TARPEY, T. and KINATEDER, K. (2003), “Clustering functional data,” *Journal of Classification*, **20**, 93–114.
- VALDERRAMA, M. (2007), “An overview to modelling functional data,” *Computational Statistics*, **22**, 331–334.
- WACKERNAGEL, H. (1995), *Multivariable Geostatistics: An Introduction with Applications*, Berlin: Springer-Verlag.