On the performance of two clustering methods for spatial functional data

Elvira Romano1 · Jorge Mateu2 · Ramon Giraldo3

Received: 3 July 2013 / Accepted: 18 August 2015 / Published online: 30 August 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The performance of two clustering strategies for spatially correlated functional data based on the same measure of spatial dependence is examined and compared. In particular, the role of the spatial dependence computed by the trace-variogram function is analyzed. The main features of both procedures is shown through a simulation study based on a variety of practical scenarios easily encountered in the analysis of spatial functional data. An application on real data based on salinity curves is also presented.

Keywords Clustering · Geostatistics · Salinity · Spatial Functional Data · Trace-variogram function

1 Introduction

The analysis of data varying over a continuum has contributed to the development of a new branch of statistics named functional data analysis (FDA) (Ramsay and Silverman 2005; Ferraty and Vieu 2006). Data object of analysis are (usually smooth) curves or
surfaces over some domain, hence the term functional data. A distinctive feature of this kind of analysis is that it makes use of the information of the slopes and curvatures of the curves, as reflected in their derivatives.

A number of statistical methods has been extended to their functional counterpart. Some recent references can be found in Gonzalez-Manteiga and Vieu (2007), Valderrama (2007), Ferraty (2010) and Ferraty and Romain (2011). However, the latest challenges for FDA are driven by the analysis of spatially dependent curves. This is the case when samples of functions are observed at different sites of a region (the so-called spatially correlated functional data) or when these functions are observed over a discrete set of time points (temporally correlated functional data). If the functional data are spatially dependent, basic FDA methods fail since the assumptions of dependence are ignored. Contributions in this framework deal with a wide variety of scientific topics (Delicado et al. 2010).

In this paper we focus on clustering. Clustering is a method in which a set of objects is divided into several groups (clusters) in such a way that members within the same cluster tend to be similar, and distinct from members of other clusters (see Bock 1974; Celeux et al. 1989; Gordon 1999; Hartigan 1975). We can find a large number of clustering methods for (non-spatial) functional data under various data assumptions, under different clustering structures and homogeneity concepts, and with or without some optimization criterion. A good review of these clustering approaches can be found in Jacques and Preda (2014). As described in Jacques and Preda (2014), we can distinguish three methodologies: dimension reduction before clustering, nonparametric methods using specific distances or dissimilarities between curves, and model-based clustering methods. In all these cases the main difficulties are related to the infinite dimensional space to which the data belongs to. In this sense the distinction among the three mentioned methods consists mainly in how they handle the infinite dimensional problem.

None of these methods consider the spatial correlation of a curve with its neighboring curves. However, in a wide variety of environmental and ecological problems the assumption of spatial independence is not appropriate and thus clustering results may be misleading. The use of a simple curve clustering procedure does not take into account the influence of the spatial correlation. In this sense, it is generally true that missing the spatial correlation structure in a functional clustering context produces unexpected results, providing many times clusters of curves that are clearly not similar in shape and/or behavior. Thus, apart from considering individual characteristics of each individual curve, we should consider the curve spatial location, and search for those curves that are homogeneous not only with respect to the shape, say, but also with respect to their spatial location (considering this as another individual characteristic).

This has led to the development of several clustering methods incorporating spatial dependence information between curves: model-based approaches (Jiang and Serban 2012), clusterwise procedures (Romano et al. 2010a; Romano and Verde 2011), hierarchical (Giraldo et al. 2012) and dynamical (Romano et al. 2010b) methods.

The purpose of these clustering methods is to find subgroups of homogeneous curves by including the spatial component within the clustering process in several different ways. Model-based (Jiang and Serban 2012) and clusterwise methods (Romano et al. 2010a; Romano and Verde 2011) consider the spatial relation among functions by
introducing a functional model in the clustering process. In Jiang and Serban (2012),
a nonparametric clustering model with spatially correlated errors is introduced. It is
designed for summarizing the time dependent cluster effects of a large number of
time-varying random functions which are spatially interdependent. In Romano et al.
(2010a) and Romano and Verde (2011) two different functional models optimize some
criteria of spatial variability among the curves in each cluster. In particular, Romano
et al. (2010a) aim at achieving a spatio-functional kriging model for each cluster
by minimizing the spatial variability measure among the geo-referenced curves. Then
Romano and Verde (2011), by extending Romano et al. (2010a) to the multivariate case,
attempt to discover spatial functional linear regression models with two functional
predictors, an interaction term and with spatially correlated residuals.

The hierarchical (Giraldo et al. 2012) and dynamical (Romano et al. 2010b) methods
both use a measure of spatial association to emphasize and delineate the average spa-
tial dependence among curves. This measure is given by the so-called trace-variogram
function (defined for the first time in Giraldo et al. (2011)). Both methods account for
the functional dynamics of the curves (in terms of the time) and their spatial rela-
tions. The hierarchical method takes the functional dynamics into account through the
Euclidean distance among the curves, while in the dynamic method the dynamics are
considered through the squared Euclidean distance among the curves computed using
the variogram function for each cluster. However, the spatial dependence plays a quite
different role in the clustering process. In the hierarchical case, the spatial dependence
is considered by the trace-variogram function computed within the distance measure
between two functional objects. In the dynamic case, such dependence is considered
by the local trace-variogram function computed for each cluster.

In this paper we focus our attention on these two last explorative methods that per-
form (spatial) classification of curves without any model assumption in the clustering
structure. Our main aim is to compare the two methods on aspects related to theory
and application, and give directions on which method has to be used under particular
scenarios. We choose these two methods since from a theoretical point of view our aim
is to evaluate the role of the trace-variogram function and its effects on the clustering
outcomes. To deal with this issue, we first describe the two methods, then we ana-
lyze, compare and discuss the performance of the two techniques through a simulation
plan designed to evaluate different levels of spatial functional correlation. Finally an
application on salinity data is discussed. The article is organized as follows. Section 2
introduces the methodological framework to deal with spatial functional data. Section
3 provides a description of the two considered clustering methods. In Sect. 4 we ana-
lyze, compare and underline the main features of both procedures through an intensive
simulation study based on a variety of practical scenarios easily encountered in the
analysis of spatial functional data. An application to salinity data is presented in Sect.
5. The paper ends with some conclusions and a discussion.

2 Measures of spatial variability for geostatistical functional data

Geostatistical functional data include a broad family of spatially dependent functional
data. For a good account of this type of data, the reader is referred to Giraldo et al.
(2012). They may be defined as curves spatially located in a region and generated by the following spatial functional process (Giraldo et al. 2011)

\[
\{ \chi_s : s \in D \subseteq \mathbb{R}^d \}
\]

(1)

where \( s \) is a generic data location in a fixed \( d \)-dimensional (usually \( d = 2 \)) Euclidean space \( D \subseteq \mathbb{R}^d \) with positive volume. We assume to observe a set of functions at \( n \) locations, \( (\chi_{s_1}(t), \ldots, \chi_{s_i}(t), \ldots, \chi_{s_n}(t)) \) for \( t \in T = [a, b] \subseteq \mathbb{R} \) and \( s_i \in D \), for \( i = 1, \ldots, n \) defining the set of functional observations. Each function is assumed to belong to a Hilbert space

\[
L_2(T) = \left\{ f : T \rightarrow \mathbb{R}, \text{ such that } \int_T f(t)^2 \, dt < \infty \right\}
\]

with the inner product \( \langle f, g \rangle = \int_T f(t)g(t) \, dt \).

Moreover, for a fixed site \( s_i \), the observed data is assumed to follow the model

\[
\chi_{s_i}(t) = \mu_{s_i}(t) + \epsilon_{s_i}(t), \quad i = 1, \ldots, n
\]

(2)

where \( \epsilon_{s_i}(t) \) are zero-mean residuals and \( \mu_{s_i}(\cdot) \) is a mean function which summarizes the main structure of \( \chi_{s_i} \). For each \( t \), we assume that the process is a second-order stationary functional random process, that formally means that the expected value \( \mathbb{E}(\chi_s(\cdot)) \) and the variance \( \mathbb{V}(\chi_s(\cdot)) \) do not depend on the spatial location, that is,

- \( \mathbb{E}(\chi_{s}(t)) = \mu(t), \forall t \in T, s \in D. \)
- \( \mathbb{V}(\chi_{s}(t)) = \sigma^2(t), \forall t \in T, s \in D. \)

In addition, we have that

- \( \mathbb{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = Cov(h, t), \) where \( h = ||s_i - s_j||, \forall t \in T, \) and \( \forall s_i, s_j \in D. \)
- \( \frac{1}{2} \mathbb{V}(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h, t) = \gamma_{s_i,s_j}(t), \) where \( h = ||s_i - s_j||, \forall t \in T, \) and \( \forall s_i, s_j \in D. \)

Note that under the second-order stationary assumption, \( \frac{1}{2} \mathbb{V}(\chi_{s_i}(t) - \chi_{s_j}(t)) \) with \( h = ||s_i - s_j|| \) could have been written as \( \frac{1}{2} \mathbb{V}(\chi_s(t) - \chi_{s'}(t)), \) where \( \chi_{s'}(t) \) is the curve located at a distance \( h \) from \( \chi_s(t) \). However, we do prefer for clarity the more general formulation, as with this we can directly compare with the local functions used by the dynamic procedure. Since we are assuming that the mean function is constant over \( D \), the function \( \gamma(h, t) \), called the variogram of \( \chi_s(t) \), can be expressed by

\[
\gamma(h, t) = \gamma_{s_i,s_j}(t) = \frac{1}{2} \mathbb{V}(\chi_{s_i}(t) - \chi_{s_j}(t)) = \frac{1}{2} \mathbb{E} [(\chi_{s_i}(t) - \chi_{s_j}(t))^2].
\]

(3)

By considering the integral on \( T \) of this expression, using Fubini’s theorem and following Delicado et al. (2010), a measure of spatial variability is given by

\[
\gamma(h) = \frac{1}{2} \mathbb{E} \left[ \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 \, dt \right], \quad \text{for } s_i, s_j \in D \text{ with } h = ||s_i - s_j||
\]
which is the so-called trace-variogram (Giraldo et al. 2011). It can be estimated as

\[
\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{s_i, s_j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt, \tag{4}
\]

where \( N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\} \), 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 \( \{(s_i, s_j) : \|s_i - s_j\| \in (h - \varepsilon, h + \varepsilon)\} \), with \( \varepsilon > 0 \) being a small value.

The trace-variogram extends to the functional context a classical approach considered in multivariate geostatistics consisting of minimizing the trace of the mean-squared prediction error matrix. We thus adopt, with the trace-variogram function, an extension of this minimization criterion to the functional context, by replacing the summation by an integral.

Following Delicado et al. (2010) and Giraldo et al. (2011), we consider that the functions can be expanded in terms of some basis functions

\[
\chi_{si}(t) = \sum_{l=1}^{Z} a_{il} B_l(t) = a_i^T B(t), \quad i = 1, \ldots, n \tag{5}
\]

The coefficients of the curves can be consequently organized in a matrix form as follows

\[
A = \begin{pmatrix}
a_{11} & a_{12} & \ldots & a_{1Z} \\
a_{21} & a_{22} & \ldots & a_{2Z} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \ldots & a_{nZ}
\end{pmatrix}_{n \times Z}
\]

and thus, the empirical trace-variogram function can be expressed by

\[
\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i, j \in N(h)} \left[ (a_i - a_j)^T W (a_i - a_j) \right] \forall i, j \text{ with } \|s_i - s_j\| = h
\]

where \( a_i, a_j \) are vectors of the basis coefficients for the \( \chi_{s_i} \) and \( \chi_{s_j} \) curves, and \( W = \int_T B(t) B(t)^T dt \) is a Gram matrix. Note that \( W \) is the identity matrix for any orthonormal basis, while for any other basis such as a B-spline basis function, \( W \) is computed by numerical integration.

Once the trace-variogram has been estimated for a sequence of \( P \) values \( h_p \), with \( p = 1, \ldots, P \), a parametric model \( \gamma_\theta(h) \) can be fitted (any of the classical and widely used models such as spherical, Gaussian, exponential or Matérn could well be used) to the points \( (h_p, \hat{\gamma}(h_p)), \quad p = 1, \ldots, P \) by ordinary least squares (OLS) or weighted least squares (WLS) (Cressie 1993). We note that the fitted parametric trace-variogram \( \hat{\gamma}_\theta(h) \) is always a valid variogram because its properties are those of a parametric variogram fitted from a univariate geostatistical data set.
The trace-variogram, as defined before, is used to describe the spatial variability among functional data across an entire spatial domain. In this case, all possible location pairs are considered. However, this kind of spatial variability could be influenced by unusual or rapid-changing behaviors. The concept of (local) spatial variability in relation to a specific location, proposed by Romano et al. (2010b), overcomes this problem.

Formally speaking, and coherently with the above definition, given a curve $\chi_{s_i}(t)$, at a specific spatial location $s_i$, the centered variogram can be expressed by

$$\gamma^{s_i}(h) = \frac{1}{2} \mathbb{E} \left[ \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \right]$$

(6)

for each $s_j \neq s_i \in D$ for which $\|s_i - s_j\| = h$. Differently from the classical variogram function for functional data, the centered variogram is defined as the variability function of a set of curves with respect to a fixed location $s_i$.

And similarly to the variogram function, the centered variogram of the curve $\chi_{s_i}(t)$, as a function of the lag $h$, can be estimated through the method-of-moments as

$$\hat{\gamma}^{s_i}(h) = \frac{1}{2 |N^{s_i}(h)|} \sum_{s_j \in N^{s_i}(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt$$

(7)

where $N^{s_i}(h) \subset N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h, s_i \text{ fixed}\}$ is defined as $N^{s_i}(h) = \{(s_i, s_j) : \|s_i - s_j\| = h, s_i \text{ fixed}\}$ where $s_j$ with $j = 1, ..., n$ is a set of varying locations, and such that $N(h) = \bigcup_{s_i} N^{s_i}(h)$ and $|N(h)| = \sum_i |N^{s_i}(h)|$.

Through straightforward algebraic operations, it is possible to show that the variogram function is a weighted average of the centered variograms

$$\hat{\gamma}(h) = \frac{1}{2 |N(h)|} \sum_{i=1}^n \left( \frac{1}{2 |N^{s_i}(h)|} \sum_{s_j \in N^{s_i}(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \right) 2 |N^{s_i}(h)|$$

(8)

and thus

$$\hat{\gamma}(h) = \frac{1}{2 |N(h)|} \sum_{i=1}^n \hat{\gamma}^{s_i}(h) 2 |N^{s_i}(h)| .$$

(9)

3 Variogram-based clustering methods for spatially dependent functional data

We now present and compare two methods (hierarchical and dynamical procedures) representing two main explorative branches of clustering techniques. To our knowledge, these are the only two explorative methods focusing on spatial classification of curves based on the same variogram function without any model assumption in the clustering structure. The former involves a hierarchy structure of the data while the
latter a partition. The presentation of the methods in this section draws heavily from Giraldo et al. (2012) and Romano et al. (2010b).

3.1 Hierarchical clustering

The hierarchical method we refer to for spatial functional data comes from Giraldo et al. (2012). It can be considered a natural extension to the functional framework of the approaches proposed for geostatistical data, where the $L_2$ norm between two curves $\chi_{s_i}, \chi_{s_j}$ is replaced by a weighted norm among the geo-referenced functions. In particular two alternatives are proposed, respectively for the univariate and multivariate contexts. When taking into account the existing spatial correlation, these methods allow to find groups of curves which are spatially homogeneous, result that cannot be obtained using classical (non-spatial) classification methods, as the weighting function increases dissimilarities among distant samples.

The first alternative is based on a weighted dissimilarity measure among the geo-referenced curves expressed by

$$d_w(\chi_{s_i}(t), \chi_{s_j}(t)) = d(\chi_{s_i}(t), \chi_{s_j}(t)) \gamma_{s_i s_j}(h)$$

where $d(\chi_{s_i}(t), \chi_{s_j}(t)) = \sqrt{\int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt}$ is the distance between the curves without considering the spatial component, and $\gamma_{s_i s_j}(h)$ corresponds to the trace-variogram function calculated for the distance between sites $s_i$ and $s_j$. Once the trace-variogram has been estimated, a parametric model is fitted following classical geostatistical estimation procedures (Giraldo et al. 2012).

Taking into account that the distance between two curves can also be calculated from the distance between the coefficients of the basis functions (Giraldo et al. 2011, 2012), a second alternative consists in estimating variograms and cross-variograms of the coefficients of the basis functions used to smooth the observed data, and then following the approach in Borgault et al. (1992). In this case assuming that the curve for each sampling site $s_i$, with $i = 1, \ldots, n$, is expanded by using (5), the coefficients in the matrix $A$ form a realization of a $Z$ multivariate random field $\{A(s) = (A_1(s), \ldots, A_Z(s)) : s \in D \subset \mathbb{R}^d\}$ with a variogram and cross-variogram matrix of the form

$$\Upsilon(h) = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \cdots & \gamma_{1Z}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \cdots & \gamma_{2Z}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{Z1}(h) & \gamma_{Z2}(h) & \cdots & \gamma_{ZZ}(h) \end{pmatrix}_{Z \times Z},$$

where $\gamma_{lq}(h) = \frac{1}{2} \mathbb{E} \left[ A_l(s_i) - A_q(s_j) \right]^2$ with $l, q = 1, \ldots, Z$ and $h = \|s_i - s_j\|$. Then multivariable geostatistics and specifically a linear model of coregionalization (LMC) to estimate the matrix (11) can be used.

As in any hierarchical clustering method, we need a distance (or a dissimilarity measure) between individuals or curves, and a distance between groups. We have defined
the distance between curves in (10). For the case of groups, we used the complete linkage based on maximum distances between groups as the agglomeration method. Complete-linkage clustering is one of several methods of agglomerative hierarchical clustering. In a first step, each curve belongs to a single-event cluster. The clusters are then combined into larger clusters until all elements are in the same cluster according to a link function. The link in the case of complete linkage contains all element pairs, and the distance between clusters equals the distance between those two elements that are farthest away from each other. The shortest of these links that remains at any step causes the merging of the two involved clusters. Complete linkage was also used by Clarkson et al. (2005). Note also that the second alternative of estimating variograms and cross-variograms from the coefficients of the curves is quite difficult in practice, as fitting a coregionalization linear model is restrictive when the number \( Z \) of basis functions increases. However, we have been able to use this method in our application with \( Z = 15 \), a number obtained by cross-validation.

### 3.2 Dynamical clustering

The dynamical clustering algorithm (DCA) was first proposed by Diday (1971) and Celeux et al. (1989), and defines a method which finds simultaneously the right partition of given data into a fixed number of clusters, and a set of representative syntheses obtained through the optimization of a fitting criterion.

Let \( \{ \chi_{s_1}(t), \ldots, \chi_{s_n}(t) \} \) with \( t \in T \) and \( s_i \in D \) with \( i = 1, \ldots, n \) be the sample of spatially located functional data. The dynamic method we present follows Romano et al. (2010b) and aims at partitioning functional data by minimizing the spatial association. The method optimizes a goodness-of-fit criterion between the centered trace-variogram function \( \gamma_{s_i}^k(h) \) and a theoretical variogram function \( \gamma_k^*(h) \) for each cluster \( C_k \) \((k = 1, \ldots, K)\) as follows

\[
\Delta(h) = \sum_{k=1}^{K} \sum_{\chi_{s_i}(t) \in C_k} (\gamma_{s_i}^k(h) - \gamma_k^*(h))^2
\]

where \( \gamma_{s_i}^k \) is the centered trace-variogram, which describes the spatial dependence between a curve \( \chi_{s_i}(t) \) at the site \( s_i \) and all the other curves \( \chi_{s_j}(t) \) at different spatial lags \( h \). Note that the index \( k \) indicates that we are only considering the cluster \( C_k \). Indeed \( \gamma_{s_i}^k \) describes the spatial variability between a curve in a site and all the other curves in the cluster \( C_k \). Thus, this algorithm could be seen as a DCA on the centered trace-variogram, instead of a clustering on the curves. In particular, the prototype here is the trace-variogram. This setting allows to evaluate the membership of a curve \( \chi_{s_i}(t) \) with respect to the spatial variability structure of an area.

The detection of the partition and, simultaneously, of the prototype set is based on the minimization of the criterion which is performed through an iterative algorithm which converges to a locally optimal solution. The number of clusters \( K \) is an input parameter. As such, when performing dynamic clustering, the number of clusters is not known nor fixed, and we need diagnostics to determine the number of clusters in
each case. In particular starting from a random initialization, the algorithm alternates representation and allocation steps until it reaches the convergence to a stationary value of the criterion $\Delta(h)$.

In the representation step, the theoretical function $\gamma^*_k(h)$ of the set of curves $\chi_{s_i}(t) \in C_k$ is estimated for each cluster $C_k$. This involves the computation of the empirical trace-variogram and model fitting by the ordinary least square method. In the allocation step, the function $\gamma^{s_i}_k(h)$ is computed for each curve $\chi_{s_i}(t)$. Then a curve $\chi_{s_i}(t)$ is allocated to a cluster $C_k$ by evaluating its matching with the spatial variability structure of the clusters according to the following rule

$$\sum_{h<h^* \in [m; M]} (\gamma^{s_i}_k(h) - \gamma^*_k(h))^2 \rho^{s_i}_k(h) < \sum_{h<h^* \in [m; M]} (\gamma^{s_i}_k(h) - \gamma^*_k'(h))^2 \rho^{s_i}_k'(h) \quad \forall k \neq k'$$

where

a. $\rho^{s_i}_k(h) = \frac{|N^{s_i}_k(h)|}{|N_k|}$ and $\rho^{s_i}_k'(h) = \frac{|N^{s_i}_k(h)|}{|N_{k'}|}$ are the weights computed for a fixed $s_i$ in terms of the number of location pairs $|N^{s_i}_k(h)|$ and $|N^{s_i}_k(h)|$ that are separated by a distance $h$ in the clusters $k$ and $k'$, and $|N_k|$ and $|N_{k'}|$ stand for the total number of locations pairs in the clusters $k$ and $k'$.

b. $m = \min_k h^*_k$, $M = \max_k h^*_k$ where $h^*_k$ is the spatial distance at which the variogram $\gamma^*_k$ for each cluster $k$ reaches its sill.

According to the above allocation criterion, only one level $h^*$ is chosen such that for $h > h^*$ there is no spatial correlation. This rule facilitates the spatial aggregation process leading to a tendency to form regions of spatially correlated curves. In particular, $h^*$ is set within the range $[m, M]$. The consistency between the representation of the clusters and the allocation criterion guarantees the convergence of the criterion to a minimum value (Celeux et al. 1989). This procedure also guarantees that $\gamma^*_k(h)$ minimizes the spatial variability of each cluster.

### 3.3 A comparison

Here we present a short comparison of the methods underlining the main characteristics together with their advantages and disadvantages. In particular the comparison has been mainly focused on two factors: (a) how each method treats the spatial component, and (b) the aims and their consequent performances.

Both clustering approaches view the realization of a spatial functional process as a set of spatially interdependent curves with cluster membership primarily characterized by the spatial variability of the curves. Both methods are strongly related according to the spatial structure they incorporate into the functional analysis. The hierarchical approach considers the estimates of the covariances between geo-referenced curves in the clustering process by generalizing an approach proposed in the spatial field. In particular, the trace-variogram function is considered as a weight in the distance
measure. In the dynamical case, this spatial structure is reflected in defining a center of the clusters at a specific location obtained by optimizing a goodness-of-fit criterion.

A characteristic of both methods is that it is assumed that the spatial functional process is second-order stationary, meaning explicitly that the second-order structure given by the covariance depends only on the distance, as defined in (4). It is true that this affects the expected value, considered to be constant. This is a classical assumption in geostatistics which should not be in contrast with the existence of clusters within the data. In a sense, a stochastic process is stationary (and isotropic) if its statistical properties do not change under translation and rotation. And we do have lots of examples of clustered stochastic processes that are stationary. The events within a cluster can be translated and the process keeps the statistical property. So we can have both concepts for the same process. We could have opted in the paper to assume a less restrictive assumption, which is that of second-order mean-reweighted stationarity, under which everything states the same as in second-order stationarity except that the mean may not be constant. However, this assumption is not completely well studied so far, and it is not that common in the literature of geostatistics. We thus preferred to keep second-order stationarity, as this is compatible with having clusters in the data. Also note that this assumption is also compatible with certain local characteristics, case that the dynamic method tends to detect.

In the hierarchical clustering method the trace-variogram is computed for all the curves as a step prior to clustering. However, in the dynamical method the trace-variogram is computed for all the curves in the cluster, and then it is optimized iteratively in the clustering process. Thus, from a computational point of view the hierarchical approach is faster than the dynamical one since the trace-variogram function is computed only once, contrary to what happens in the dynamical case in which for each cluster a trace-variogram is computed.

The resulting nature of the clustering results reflects the different ways of including the spatial component. In particular, while the dynamic method identifies homogeneous local-areas (named local clusters) by means of different trace-variogram functions (interpretable in terms of the observed phenomenon), the hierarchical method highlights global spatial functional clusters (named global clusters) obtained by using the same trace-variogram function. In many applied contexts, stationary processes exhibit stochastic variations, and thus the use of the same trace-variogram function for taking into account the (global) spatial dependence structure of functional data is not a realistic assumption since complex spatial structure can not be efficiently described by a single spatial variability. In this sense, the clusters obtained by the hierarchical method are representative of homogenous spatial areas, but not strictly representative for heterogeneous spatial areas.

According to the aims, both methods are satisfactory since independent of the output of the two procedures (a partition and a hierarchy) they both detect clusters of spatially dependent functional data. In the dynamic method, the trace-variogram functions and their locations are assumed to be spatially uniform representatives of the obtained clusters. In the hierarchical method, the spatial diversity is directly obtained according to the hierarchy that gives an order of spatial diversity among the functional data.
In contrast to the classical clustering methods for functional data where the spatial location is not considered, they both have their own advantages. On one hand, the dynamic approach built upon the trace-variogram, allows to model the spatial relation among the functional data by a set of representative trace-variogram functions. On the other hand, the hierarchical approach built upon a weighted spatial similarity, aims to identify groups of sites which display similar spatio-temporal characteristics by a unique trace-variogram function. Thus dynamic clustering is probably the best choice when the spatial dependence among curves is affected by local covariates, and can not be described and summarized by the same trace-variogram function. On the opposite case, the hierarchical approach can be chosen when the spatial dependence can be described by a single trace-variogram function as the unique characteristic of a spatial structure.

4 A simulation study

To analyze and compare the performances of both clustering procedures, a simulation study has been performed. Since our main aim is to evaluate the strength of the two methodologies in capturing the clustering schemes while taking care of the spatial structure, we have considered a simulation plan that takes into account a number of spatial functional relationships among any two functions observed at any two different sites. For completeness, a comparison of the two methods with a classical K-means algorithm on functional data (without considering the spatial component) (Tarpey and Kinateder 2003; Tarpey 2007) has been discussed.

Consider given a set of curves \( \chi_{s_1}(t), \chi_{s_2}(t), \ldots, \chi_{s_n}(t) \) located at \( s_1, s_2, \ldots, s_n \) in \( \mathbb{R}^2 \), and \( t \in T \). Assume that a generic curve \( \chi_s(t) \) is generated by the general model

\[
\chi_s(t) = \mu_s(t) + \epsilon_s(t), \quad t \in T
\]

with mean \( \mu_s(t) \) and \( \epsilon_s(t) \) a Gaussian random field with zero mean and a spatial functional covariance function expressed by \( \text{Cov}(h, u) = \text{Cov}\{\chi_{s_i}(t_1), \chi_{s_j}(t_2)\} \), for any couple of locations \( s_i, s_j \) separated by a spatial distance \( h = \|s_i - s_j\| \), and a temporal distance \( u = t_1 - t_2 \). We considered three main scenarios defined according to the nature of the covariance function. These were purely spatial, separable spatio-functional, and non-separable spatio-functional structures, following the setup in Sun and Genton (2012). In addition, for each of these scenarios we considered several structures of the mean function \( \mu_s(t) \).

For each considered scenario, we generated 588 curves in three clusters of data \( C_1, C_2, C_3 \), containing each one 196 curves. These 196 curves were located over \( s_1, \ldots, s_{196} \) locations on a grid of size 14 × 14 in the unit square. Each curve was formed by 50 equally spaced time points in [0, 1]. Figure 1 shows a representation of the set of 196 locations for each of the three considered clusters (schematized with color blue, green and red). For each cluster, we considered three main levels of spatial variability: a strong spatial variability inside each cluster with large differences, a strong spatial variability inside the cluster, but mild variability among clusters, and
Fig. 1 Grid of $196 \times 3$ simulated locations over which the curves are generated. The colors represent the three considered clusters, $C_1$, $C_2$, $C_3$ (color figure online)

one level of spatial variability for each cluster varying from very strong, medium to low variability.

This has been done in order to have a strong spatial correlation within clusters but also some (weak) spatial correlation between clusters. In the separable and non-separable scenarios in addition to the levels of spatial variability we add three levels of functional variability: (a) medium and equal for each cluster; (b) very low and equal for each cluster, and (c) ranging from low to strong for each cluster.

Note that the considered block-structured setting defining the locations of the curves is such defined to ease interpretation and visual inspection of the differences amongst the curves. We simulated other kinds of spatial settings and our methods were not sensitive to these other structures; this is also underlined in the real data application.

Let us have a closer look at the three considered covariance function structures. We first used a stationary purely spatial covariance function which takes the form

$$C_{ov}(h) = (1 - \nu) \exp (-c |h|) + \nu I \{h = 0\}$$

where $c > 0$ controls the spatial correlation intensity, and $\nu \in (0, 1]$ is the nugget effect. For convenience, we fixed $\nu = 0.04$ for all the cases. The choice of the values of the parameter $c$ was done considering that the smaller the value of $c$, the stronger the spatial dependence. We thus considered the following values: (a) $c_1 = 0.01$, $c_2 = 0.5$, $c_3 = 0.9$, corresponding to a strong variability inside each cluster with a clear difference; (b) $c_1 = 0.1$, $c_2 = 0.3$, $c_3 = 0.6$, corresponding to strong but not
very different, and (c) $c_1 = 0.1, c_2 = 1, c_3 = 2$, corresponding to very strong, medium and low spatial variability.

As a second covariance structure, we considered a separable spatio-temporal covariance function of the form

$$Cov_{SEP} (h, u) = Cov \{ \chi_{s_i} (t_1), \chi_{s_j} (t_2) \} = Cov_S (h) Cov_T (u)$$

(16)

where $Cov_S (h)$ is given in (15) and $Cov_T (u)$ is a stationary, purely functional covariance function of the Cauchy type

$$Cov_T (u) = \left( 1 + a |u|^{2 \alpha} \right)^{-1}$$

(17)

with a time span $u = |t_1 - t_2|$. Here $a > 0$ is the scale parameter in time, that here is fixed to $a = 1$ for convenience, and $\alpha \in [0, 1]$ is the parameter that controls the strength of the functional variability. For each level of spatial correlation, the following values were considered: (a) $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$, (b) $\alpha_1 = \alpha_2 = \alpha_3 = 0.01$, and (c) $\alpha_1 = 0.1 \alpha_2 = 0.3 \alpha_3 = 0.4$.

Finally, as a third covariance structure, we considered a symmetric but non-separable spatio-temporal covariance function of the form

$$Cov_{NSEP} (h, u) = \frac{1 - v}{1 + au^{2 \alpha}} \left[ \exp \left\{ - \frac{c \|h\|}{(1 + a |u|^{2 \alpha})^{\beta}} \right\} + \frac{v}{1 - v} I \{ h = 0 \} \right]$$

(18)

where the parameter $0 \leq \beta \leq 1$ controls the degree of non-separability. We fixed $\beta = 0.9$ to have the maximum possible non-separability degree, while for the other parameters $c$ and $\alpha$ we considered the same combination of values as in the second scenario.

The combination of all the considered scenarios provided a set of 42 simulated datasets, each one formed by 588 spatially correlated curves. Figures 2 and 3 give an idea of some simulated functions following the above spatial and temporal structures. Out of the 42, 21 datasets corresponded to the case where the mean function $\mu_s(t)$ was zero in all cases. The other 21 datasets were analyzed under a presence of a trend, with varying values of the mean function depending on the cluster membership ($\mu_{C_1} = 5$, $\mu_{C_2} = 10$, $\mu_{C_3} = 15$).

In particular, we simulated the following cases: (a) 6 datasets for the purely spatial covariance function, the first 3 (denoted by IS$_v$ with $v = 1, 2, 3$) with a zero-constant mean function, and the other 3 (denoted by IMS$_v$ with $v = 1, 2, 3$) with a variable mean function (trend-presence case); (b) 18 datasets under the spatio-temporal separable covariance function, the first 9 (denoted by I$_v$ with $v = 1, \ldots, 9$) with a zero-constant mean function, and the other 9 (denoted by IM$_v$ with $v = 1, \ldots, 9$) with a variable mean function (trend-presence case); and (c) 18 datasets under the spatio-temporal non-separable covariance function, the first 9 (denoted by I$_v^{ns}$ with $v = 1, \ldots, 9$) with a zero-constant mean function, and the other 9 (denoted by PIM$_v^{ns}$ with $v = 1, \ldots, 9$) with a variable mean function (trend-presence case). All these cases, and the corresponding parameter setting used are shown in Table 1.
As an illustrative example, Fig. 2 shows two particular simulated datasets, one under the spatio-temporal separable covariance function \( I_s^1 \), and the other under the non-separable function \( I_{ns}^0 \), and under the homogeneous case, with a zero-constant mean function. Two cases under the presence of a trend, corresponding to a varying mean function, are shown in Fig. 3. It is clearly noted the effect of the varying mean function on the curves for each cluster. Since a large number of basis functions causes over-fitting, and with a too-small number we lose important aspects of the functions (Ramsay and Silverman 2005), the functional observations were smoothed using B-splines basis with a number of basis functions evaluated over a number of knots selected
Table 1 The datasets and their parameters for each simulated scenario: 6 datasets IS\(_v\) with \(v = 1, 2, 3\), IMS\(_v\) with \(v = 1, 2, 3\) for the purely spatial covariance function, 18 datasets \(I^s_v\) with \(v = 1, \ldots, 9\), IM\(_v^s\) with \(v = 1, \ldots, 9\) for the separable covariance function, and 18 datasets \(I^{ns}_v\) with \(v = 1, \ldots, 9\), IM\(_v^{ns}\) with \(v = 1, \ldots, 9\) for the non-separable covariance function

| Dataset id | \(\alpha_1\) | \(\alpha_2\) | \(\alpha_3\) | \(c_1\) | \(c_2\) | \(c_3\) | \(\mu_{c_1}\) | \(\mu_{c_2}\) | \(\mu_{c_3}\) |
|------------|--------------|--------------|--------------|-------|-------|-------|----------|----------|----------|
| IS\(_1\)   | 0.01         | 0.5          | 0.9          | 0     | 0     | 0     | 0        | 0        | 0        |
| IS\(_2\)   | 0.1          | 0.3          | 0.6          | 0     | 0     | 0     | 0        | 0        | 0        |
| IS\(_3\)   | 0.1          | 1            | 2            | 0     | 0     | 0     | 0        | 0        | 0        |
| \(I^s_1\)  | 0.2          | 0.2          | 0.2          | 0.01  | 0.5   | 0.9   | 0        | 0        | 0        |
| \(I^s_2\)  | 0.2          | 0.2          | 0.2          | 0.1   | 0.3   | 0.6   | 0        | 0        | 0        |
| \(I^s_3\)  | 0.2          | 0.2          | 0.2          | 0.1   | 1     | 2     | 0        | 0        | 0        |
| \(I^s_4\)  | 0.01         | 0.01         | 0.01         | 0.01  | 0.5   | 0.9   | 0        | 0        | 0        |
| \(I^s_5\)  | 0.01         | 0.01         | 0.01         | 0.1   | 0.3   | 0.6   | 0        | 0        | 0        |
| \(I^s_6\)  | 0.01         | 0.01         | 0.01         | 0.1   | 1     | 2     | 0        | 0        | 0        |
| \(I^s_7\)  | 0.01         | 0.01         | 0.01         | 0.01  | 0.3   | 0.6   | 0        | 0        | 0        |
| \(I^s_8\)  | 0.01         | 0.3          | 0.4          | 0.1   | 0.3   | 0.6   | 0        | 0        | 0        |
| \(I^s_9\)  | 0.01         | 0.3          | 0.4          | 0.1   | 0.3   | 0.6   | 0        | 0        | 0        |
| IMS\(_1\)  | 0.01         | 0.5          | 0.9          | 5     | 10    | 15    |          |          |          |
| IMS\(_2\)  | 0.1          | 0.3          | 0.6          | 5     | 10    | 15    |          |          |          |
| IMS\(_3\)  | 0.1          | 1            | 2            | 5     | 10    | 15    |          |          |          |
| IM\(_1^{ns}\) | 0.2        | 0.2          | 0.2          | 0.01  | 0.5   | 0.9   | 5        | 10       | 15       |
| IM\(_2^{ns}\) | 0.2        | 0.2          | 0.2          | 0.1   | 0.3   | 0.6   | 5        | 10       | 15       |
| IM\(_3^{ns}\) | 0.2        | 0.2          | 0.2          | 0.1   | 1     | 2     | 5        | 10       | 15       |
| IM\(_4^{ns}\) | 0.01       | 0.01         | 0.01         | 0.01  | 0.5   | 0.9   | 5        | 10       | 15       |
| IM\(_5^{ns}\) | 0.01       | 0.01         | 0.01         | 0.1   | 0.3   | 0.6   | 5        | 10       | 15       |
| IM\(_6^{ns}\) | 0.01       | 0.01         | 0.01         | 0.1   | 1     | 2     | 5        | 10       | 15       |
| IM\(_7^{ns}\) | 0.01       | 0.3          | 0.4          | 0.01  | 0.5   | 0.9   | 5        | 10       | 15       |
| IM\(_8^{ns}\) | 0.1         | 0.3          | 0.4          | 0.1   | 0.3   | 0.6   | 5        | 10       | 15       |
| IM\(_9^{ns}\) | 0.1         | 0.3          | 0.4          | 0.1   | 0.3   | 0.6   | 5        | 10       | 15       |

The presence of \(M\) in the name of the scenario indicates that we are under a trend-presence case.

by cross-validation (Giraldo et al. 2011). An important choice in the clustering process consists of evaluating the best theoretical variogram model that fits the data. This is a common procedure in geostatistical analysis, and here we evaluated three different well-known parametric models, exponential, Gaussian, and spherical structures. We chose for both clustering methods the exponential one, as this model showed the best fit in 73% of the cases.

In order to evaluate the ability of the compared methods to detect the clusters, we used the Rand Index (Hubert and Arabie 1985; Rand 1971) as a measure of concordance between the true clustering structure and the output produced by the clustering algorithms. Obviously, as noted before, the hierarchical clustering method provides a single unique solution, the hierarchy, which is in opposition to the dynamic method that depends on the initialization step as the K-means method does. We compensate...
Table 2  Averaged rand index for the dynamic (DC) and hierarchical (HC) clustering methods, for the scenarios considered

| Dataset id | RI<sub>DC</sub> | RI<sub>HC</sub> | Dataset id | RI<sub>DC</sub> | RI<sub>HC</sub> |
|------------|---------------|---------------|------------|---------------|---------------|
| IS<sub>1</sub> | 0.772         | 0.997         | IS<sub>M</sub> | 0.811        | 0.911         |
| IS<sub>2</sub> | 0.527         | 0.538         | IS<sub>M</sub> | 0.710        | 0.760         |
| IS<sub>3</sub> | 0.835         | 0.800         | IS<sub>M</sub> | 0.661        | 0.795         |
| I<sub>S</sub><sub>1</sub> | 0.986         | 0.927         | IM<sub>S</sub><sub>1</sub> | 0.906        | 0.931         |
| I<sub>S</sub><sub>2</sub> | 0.997         | 0.695         | IM<sub>S</sub><sub>2</sub> | 0.699        | 0.963         |
| I<sub>S</sub><sub>3</sub> | 0.685         | 0.942         | IM<sub>S</sub><sub>3</sub> | 0.773        | 0.801         |
| I<sub>S</sub><sub>4</sub> | 0.738         | 0.901         | IM<sub>S</sub><sub>4</sub> | 0.773        | 0.801         |
| I<sub>S</sub><sub>5</sub> | 0.997         | 0.777         | IM<sub>S</sub><sub>5</sub> | 0.887        | 0.801         |
| I<sub>S</sub><sub>6</sub> | 0.589         | 0.950         | I<sub>M</sub><sub>S</sub><sub>6</sub> | 0.599        | 0.722         |
| I<sub>S</sub><sub>7</sub> | 0.611         | 0.901         | I<sub>M</sub><sub>S</sub><sub>7</sub> | 0.711        | 0.801         |
| I<sub>S</sub><sub>8</sub> | 0.946         | 0.950         | I<sub>M</sub><sub>S</sub><sub>8</sub> | 0.846        | 0.850         |
| I<sub>S</sub><sub>9</sub> | 0.646         | 0.977         | I<sub>M</sub><sub>S</sub><sub>9</sub> | 0.716        | 0.812         |
| I<sub>Ns</sub><sub>1</sub> | 0.519         | 0.997         | I<sub>M</sub><sub>Ns</sub><sub>1</sub> | 0.589        | 0.817         |
| I<sub>Ns</sub><sub>2</sub> | 0.730         | 0.997         | I<sub>M</sub><sub>Ns</sub><sub>2</sub> | 0.810        | 0.807         |
| I<sub>Ns</sub><sub>3</sub> | 0.843         | 0.977         | I<sub>M</sub><sub>Ns</sub><sub>3</sub> | 0.903        | 0.917         |
| I<sub>Ns</sub><sub>4</sub> | 0.778         | 0.997         | I<sub>M</sub><sub>Ns</sub><sub>4</sub> | 0.688        | 0.891         |
| I<sub>Ns</sub><sub>5</sub> | 0.790         | 0.977         | I<sub>M</sub><sub>Ns</sub><sub>5</sub> | 0.690        | 0.728         |
| I<sub>Ns</sub><sub>6</sub> | 0.750         | 0.997         | I<sub>M</sub><sub>Ns</sub><sub>6</sub> | 0.850        |               |
| I<sub>Ns</sub><sub>7</sub> | 0.741         | 0.997         | I<sub>M</sub><sub>Ns</sub><sub>7</sub> | 0.823        | 0.922         |
| I<sub>Ns</sub><sub>8</sub> | 0.765         | 0.911         | I<sub>M</sub><sub>Ns</sub><sub>8</sub> | 0.801        | 0.899         |
| I<sub>Ns</sub><sub>9</sub> | 0.522         |               | I<sub>M</sub><sub>Ns</sub><sub>9</sub> | 0.592        |               |

this difference by considering the Rand Index under 100 repetitions of the algorithm in order to have an average value of the Rand Index (in addition, in the dynamic case, we get independence of the initial partition). This index, whose value is in the range [0, 1], measures the degree of matching between the true partition of the data, which arises from the simulation scenario, and the partition given as the output of the proposed clustering method. The value 0 indicates that the two partitions do not agree on any pair of items, while 1 means that the partitions are exactly the same. Table 2 reports the results for both algorithms showing the average of the Rand Index under 100 repetitions. We should note that in any case, the standard deviation was larger than 0.035.

For each simulated dataset, and under the hierarchical procedure, we used the Calinsky-Harabasz criteria (Davies and Bouldin 1979) to define the optimum number of groups. According to this, the cutting of the tree into three clusters was the best option for most of the analyzed datasets. Only in some few cases (IS<sub>3</sub>, I<sub>S</sub><sub>4</sub>, I<sub>S</sub><sub>6</sub>, I<sub>S</sub><sub>9</sub>, I<sub>M</sub><sub>ns</sub><sub>5</sub>, I<sub>M</sub><sub>ns</sub><sub>6</sub>, I<sub>M</sub><sub>ns</sub><sub>9</sub>) for which we do not report the value of the rand index, this was not the best option. In most of these cases, the within spatial cluster variability has different intensity (from very strong to low) and the weight in the distance (10) is not able to diversify the spatial structure of each cluster. In general, the hierarchical method performs better when the functional characteristics of the spatial clusters are different.
across space, and when there are low differences among the spatial variability of the clusters.

The dynamic method detected most of the simulated clusters. However, as shown by \( \text{RI}_{\text{DC}} \) (see Table 2) there were some scenarios (IS\(_2\), I\(_7\), I\(_9\), IM\(_1\), IM\(_3\), IM\(_5\)) in which the dynamic method reported low values of the rank index, indicating that it missed the original clusters. These cases are those in which the clusters show similar spatial variability. Thus in the step of aggregation of the curves into the clusters, the dynamic method does not discriminate in which clusters the functions must be allocated since they have similar structures.

These results highlight the main difference among the two methods: for the hierarchical method the spatial relation among functional data, expressed by the trace-variogram function, is an equal and global weight for all the functional data spatially located in the clustering process; on the contrary, in the dynamic one the centered trace-variogram represents a local variability with respect to which all curves are evaluated in the allocation step and leads to define local weights for different clusters.

The good performance of both methods is highlighted by the big proportion (81 and 79 %, respectively) of situations in which they are able to capture the different spatial structures considered in the simulated scenarios.

For completeness, we also compared the performance of both clustering methods with that of the K-means method for functional data. In particular, we applied the K-means method to the simulated datasets, noting that this method misses the spatial information coming from the locations of the curves. The RI for the K-means method in all the scenarios considered in Table 2 was always between 0.5 and 0.75. This is due to the strength of the method to discover functional clusters with similar shape, however these results are partially inexact if we look for the spatial locations of the curves. By observing the classification obtained by this method (see Fig. 4), we note that in 40 % of the cases the curves are not well classified in terms of their spatial relation.

The classification results for some selected cases (IS\(_3\), I\(_9\), and IM\(_9\)), and for illustrative purposes, are shown in Fig. 4. Comparing this figure with Fig. 1, both methods (hierarchical and dynamic) are doing a good job in detecting the true partitions, contrary to the K-means method which is not able to capture the spatial structure. In particular, the results show that both clustering methods are able to recognize the right partition in terms of functional and spatial relationships under a variety of scenarios. We can only observe a slight misclassification for the third scenario. Both methods (hierarchical and dynamic) are thus superior in detecting clusters when compared with the more classical and non-spatial K-means clustering method. In particular different aspects should be considered for an appropriate choice of the method. Both methods account for the functional dynamics of the curves (in terms of the time) and their spatial relations. The temporal dynamic is handled similarly by both methods. In particular, in the hierarchical method, the functional dynamic is taken into account through the Euclidean distance among the curves, while in the dynamic method it is considered through the squared Euclidean distances among the curves, computed by the variogram function for each cluster. However, the spatial dependence plays a very different role in the clustering process. In the hierarchical case, it is considered by the trace-variogram function computed among all the functional objects, weighting the
Fig. 4 Classification results for some selected scenarios IS₃, IS₅, and IS₆. Results obtained by the DC method (first row), by the HC method (second row), and by the K-means method for functional data (third row).

distance measure between two functional objects. In the dynamic case, it is considered by the local trace-variogram function computed for each cluster.

A further characteristic can be underlined by observing the performances of both methods on some simulated datasets with different spatial variability in the clusters. For example, if we look at the results of the methods for the simulated datasets IM₆ and IS₃, cases where the dataset is characterized by clusters with different spatial variability structure, we note that the dynamic method is more accurate in discovering and modeling this variability, as shown by RI. In this sense, the hierarchical method seems preferable for discovering spatial functional clusters adapting the global character of the spatial variation, and that is why we name the resulting clusters as “global”. On the contrary, the dynamic method seems preferable for verifying the existence of spatial functional clusters that consider the effect of local changes in the spatial functional variability. For this reason we define them as “local” clusters.

5 Application: clustering of salinity curves

In this section we perform functional cluster analysis with a real dataset by means of hierarchical and dynamical spatial functional clustering. A comparison between
the results obtained with both methods is provided. The information corresponds to salinity data measured at 21 monitoring stations of the lagoonal-estuarine system comprised by the Ciénaga Grande de Santa Marta (CGSM) and Complex of Pajarales (CP) (see Fig. 5). Biweekly data from October 1988 to March 1991 were recorded (left panel, Fig. 6). Several works taken place in CGSM and CP have shown that the salinity is one of the variables that governs the ecosystem changes. For this reason it is important to know its spatial distribution. Giraldo et al. (1995) performed a cluster analysis over this same dataset but the spatial dependence was not taken into account. Although the results found in that work were useful, it was difficult to explain a few of them from a biological point, due to not having considered the geographic proximity as a relevant factor. Here we retake that analysis but applying spatial clustering for functional data.
To perform both clustering methods, the data was initially smoothed out by using (5) with $B_l(t), l = 1, \ldots, Z$ B-splines basis, and $Z = 15$ (right panel, Fig. 6). The number of basis functions was obtained by cross-validation. These basis functions were evaluated over 12 knots uniformly distributed, and spanning the whole range of the temporal domain. Note that the shape of the basis functions is determined by the position of the knots. Scaling or translating the knot vector does not alter the basis functions.

**5.1 Hierarchical clustering of salinity curves**

We performed a functional hierarchical clustering analysis under two alternative perspectives. One is based on weighting the distance matrix by the trace-variogram function (10), and the other uses the multivariate variogram for the fitted basis coefficients $a_i's$. We thus need to estimate the trace-variogram function, and a linear model of coregionalization (between the spatial variables formed with the basis coefficients) to calculate the matrix in (11). These estimations should be done using stationary functions. However, the salinity data is not stationary as the mean increases from east to west, that is, the stations located in Complex of Pajarales have a higher salinity level than stations located in the Ciénaga Grande de Santa Marta (see left panel in Fig. 6). Only the station LBA in the northeast side of Ciénaga Grande de Santa Marta has a level as high as the stations located in Complex of Pajarales (Giraldo et al. 1995). For this reason, and with the purpose of detrending the data, we initially performed a functional regression analysis with the smoothed salinity curves (right panel in Fig. 6) as the functional response, and the geographical coordinates as scalar covariates. Specifically we estimated, using ordinary least squares, the model

$$
\chi_i(t) = \alpha(t) + \beta_1(t) \text{Longitude}_i + \beta_2(t) \text{Latitude}_i + \epsilon_i(t),
$$

where $\chi_i(t), i = 1, \ldots, 21$ are the salinity curves, and $\alpha(t), \beta_1(t),$ and $\beta_2(t)$ are the functional parameters of interest. The estimated functional parameters are shown in Fig. 7. We note that the parameters are significantly different from zero (only $\beta_1(t)$ is significantly equal to zero in some time periods). The estimation of the mean $\hat{\alpha}(t)$ shows that the level of salinity increased during the time period of the study. According to the sign of $\hat{\beta}_1$, the salinity decreased from west to east, that is, when the longitude increased (right panel in Fig. 5) the salinity level decreased (in those time periods where $\hat{\beta}_1$ has a negative sign and is significantly different from zero). This is coherent with the fact that the level of salinity is higher in CP stations than in CGSM stations (Fig. 6). The sign of $\hat{\beta}_2$ indicates that the salinity level increased from south to north (Fig. 5). This trend is particularly due to the stations located in the south of the system (RFU and BRF, Fig. 5), as they have a low level of salinity because they are directly influenced by the Fundacion river.

Once estimated the functional regression model, we used the residual curves $e_i(t), i = 1, \ldots, 21$ to estimate the spatial correlation structure. These curves were also smoothed by using a B-splines basis with 15 functions (Fig. 8). We followed two procedures to estimate the spatial correlation structure. We first estimated the trace-
Fig. 7 Estimated parameters (from left to right $\alpha(t)$, $\beta_1(t)$, and $\beta_2(t)$) for the functional regression model between the salinity curves and the geographical coordinates of stations where data were recorded. Dark lines correspond to estimations, and dashed lines to confidence intervals.

In addition to the two specified methods for spatial functional cluster analysis, and for comparison purposes, we also performed a classical (non spatial) functional cluster analysis. To carry out the hierarchical spatially correlated clustering analysis with the salinity data, we calculated the weighting dissimilarity matrix in (10) with $\gamma_{ij}$ estimated from both the trace-variogram function and the multivariate variogram. The dendrograms obtained are shown in Fig. 9. Two threshold values were considered. The largest threshold value was obtained by applying the cluster quality measures of Davies-Bouldin, Calinski-Harabasz, Hubert-Levine and Silhouette, respectively (Davies and Bouldin 1979). In all cases these measures suggested two clusters. The lowest threshold value was chosen visually, and with a purely empirical criterion, in order to get five clusters in all cases. This allows to detect more specifically the effect of taking into account the spatial dependence in the analysis.

The most relevant result when we compare the three dendrograms in Fig. 9, and when considering the highest threshold value, is that with the classical functional cluster analysis the station LBA (located in the northwest side of CGSM) is assigned...
Fig. 8 Left panel residual curves of the functional regression model between salinity curves and the geographical coordinates of the monitoring stations of Ciénaga Grande de Santa Marta and Complex of Pajarales. Residuals were smoothed by a B-splines basis with 15 functions. Right panel Trace-variogram fitted to the residual curves.

Fig. 9 Dendograms and clusters of monitoring stations of Ciénaga Grande de Santa Marta and Complex of Pajarales obtained according to three methods: classical functional clustering (left panel), spatially correlated functional clustering weighting by the trace-variogram function (central panel), and spatially correlated functional clustering weighting by the multivariate variogram (right panel). Two threshold values are considered in each case. The largest (dark line) was obtained by using the cluster quality measures. The lowest (dashed line) is given empirically such that we obtain five clusters in each case.
to the same group of some of the stations in CP (which have high salinity values). However, with the methods based on spatially correlated functional clustering the station LBA is assigned to other clusters formed by closer stations of CGSM (with exception of station VCH in the dendrogram of the right panel). Although the salinity level in the LBA station is as high as in the stations in CP, from a biological point of view this one should be assigned to a different cluster because it is spatially far from the stations in CP (the maximum distance between stations is 34 km, and LBA is separated from all the stations in CP more than 22 km) and its high salinity values are due to different environmental conditions. High salinity values in LBA are due to its proximity to the Caribbean sea, whereas the high values in the west side of CP are caused by interruptions of natural flows of water (due to natural events and specially to human intervention (see Botero and Mancera (1996)). In this sense the application of methods based on spatially functional clustering are a better option. Another important result that we can observe in Fig. 9, considering now the lowest threshold value, is that the clusters obtained by the methods based on the spatial dependence are very similar (central and right panels in Fig. 9). The only difference is given by the cluster to which the station VCH is assigned. A detailed overview of the dissimilarity matrix shows that although the salinity levels in stations VCH, RSE and RFU are very different, the $L_2$ norm between these stations (the distance between the smoothed curves without considering the spatial component) is quite low, and consequently they are assigned to the same cluster. The difference is due to the fact that the trace-variogram detects a higher spatial dependence than the multivariate variogram. The ecological knowledge of the ecosystem indicates that the most appropriate result is the one obtained by the spatially correlated functional clustering method when the dissimilarity matrix is weighted by the trace-variogram function.

### 5.2 Dynamic clustering of salinity curves

The application of the dynamic method to the salinity curves needs of two input parameters: the number of clusters, and the parametric model fitted to the variogram function. In relation to the number of clusters, we used several values for $K$, and selected $K = 2$ as that value that provided the minimum value of the optimized criterion $\Delta(\cdot)$. We also evaluated three different variogram models, exponential, spherical and Gaussian, and according to the results in Table 3, we chose the exponential one.

The dynamic method detects two main spatial regions (see Fig. 10), which consider the two more different areas in terms of spatial distribution and functional shape of the analyzed phenomena. As in the hierarchical method, the station LBA is assigned

| Trace-variogram model | $\Delta(\cdot)$ |
|-----------------------|----------------|
| Exponential           | $2.8e+4$       |
| Gaussian              | $3.3e+4$       |
| Spherical             | $3.5e+4$       |
Fig. 10  Detected clusters of monitoring stations of Ciénaga Grande de Santa Marta and Complex of Pajarales obtained according to the dynamic method. The two clusters are identified by the two colors, black and red (color figure online)

Fig. 11  Theoretical prototype variogram models for the two detected clusters

to the same group of closer stations of CGSM, contrary to the classical clustering algorithm. In addition, the two different spatio-functional variability structures can be summarized by the variogram prototypes of each cluster (see Fig. 11). Here it is possible to note that the variability in the second detected cluster rises at a lower rate compared to that in the first detected cluster. The variogram prototypes differ in the slope of the curves, in the maximum value of variability, and in the maximum value of the lag distance $h$. This last aspect involves that the maximum distance between two curves of a cluster can be different. In particular, the first variogram prototype reaches the highest level of variance, $\gamma(h) = 15,200$, for a lag distance of $h = 3.9 \times 10^4$, while for the second cluster the highest level of variance is $\gamma(h) = 12,900$ at a lag distance $h = 3.25 \times 10^4$. For both the variograms there is no nugget effect. These results highlight that in the first cluster there is a higher spatial dependence.

6 Conclusions and discussion

We have presented two clustering methods for functional data that handle spatial correlation. This is the key factor of the presented methods, and what makes the
difference with more classical clustering methods for curves. The paper has focused on the analysis, performance and comparison between the hierarchical and dynamical clustering methods under a large variety of practical scenarios. In addition, we have analyzed a dataset on salinity curves. Based on the simulations and real data results, we note that in general, apart from the different clustering outputs of the two methods (partition and hierarchy), the best choice between the hierarchical and the dynamic procedures depends on the particular aim of the analysis, and on the spatial functional data structure. In particular, the dynamic method seems preferable for detecting local spatial functional clusters, whereas the hierarchical method appears preferable for global spatial clusters. Both methods are superior in detecting clusters when compared with the more classical and non-spatial K-means clustering method.

On the weak part, the hierarchical method depends much upon the selected threshold value, but this is a long standing problem with these type of clustering methods. In addition, this method also assumes a well-fitted theoretical model to the trace-variogram or to the multivariate variogram. This is an issue that has to be reinforced when selecting a particular model. The dynamic method depends much on the optimization criterion used, and caution has to be taken into account on not being trapped in local maxima or minima. This can be quite tricky in practical terms.

Both methods are based on the assumption that data, as usually assumed in geostatistics, are stationary. However, alternative ways of including spatial dependence among functional data when the process is not stationary is an interesting topic for future research. We could extend the methods compared in this paper by considering directional variograms for treating non-stationary spatial functional process. In this setting, it is important to have computationally efficient procedures to estimate different degrees of spatial variability.

The salinity dataset will be made available in a new version of the library geofda of the free software R (Giraldo et al. 2012). This library is now available in its first version from http://www.docentes.unal.edu.co/rgiraldoh/docs/. We shall soon release a second version. The methods here implemented are in Matlab and R and the code is available upon contacting the authors.

Acknowledgments We are grateful to the AE and the referees whose comments and suggestions have clearly improved an earlier version of the manuscript. Work partially funded by Grant MTM2013-43917-P from the Spanish Ministry of Science and Education and Grant P1-1B2012-52.

References

Bock, H.-H.: Automatische Klassifikation (Clusteranalyse). Vandenhoeck & Ruprecht, Gottingen (1974)

Borgault, G., Marcotte, D., Legendre, P.: The multivariable covariogram as a spatial weighting function in classification methods. Math. Geol. 24, 463–478 (1992)

Botero, L., Mancera, E.: Síntesis de los cambios de origen antropico ocurridos en los últimos 40 años en la cienaga grande de Santa Marta (Colombia). Revista de la Academia Colombiana de Ciencias 20, 465–474 (1996)

Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., Ralambondrainy, H.: Classification automatique des données: environnement statistique et informatique. Dunod (1989)

Clarkson, D., Fraley, C., Gu, C., Ramsay, J.: S+ Functional Data Analysis User Guide. Springer, New York (2005)

Cressie, N.A.C.: Statistics for Spatial Data. John Wiley & Sons, New York (1993)
Cressie, N.A.C., Collins, L.B.: Analysis of spatial point patterns using bundles of product density isarithms. J. Agric. Biol. Environ. Stat. 6, 118–135 (2001)
Davies, D., Bouldin, D.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1, 224–227 (1979)
Delicado, P., Giraldo, R., Comas, C., Mateu, J.: Statistics for spatial functional data: some recent contributions. Environmetrics 21, 224–239 (2010)
Diday, E.: La méthode des Neuroes dynamiques. Revue de Statistique Appliquée 19, 19–34 (1971)
Ferraty, F.: Statistical methods and problems in infinite-dimensional spaces. J. Multivar. Anal. 101, 305–490 (2010)
Ferraty, F., Romain, Y.: The Oxford Handbook of Functional Data Analysis. Oxford University Press, New York (2011)
Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York (2006)
Giraldo, R., Martinez, J., Hurtado, H., Zea, S., Madera, R.: Análisis de clasificación de series temporales: El caso de la salinidad en la Ciénaga Grande de Santa Marta. Anales del Instituto de Investigaciones Marinas de Punta de Bete 24, 123–134 (1995)
Giraldo, R., Delicado, P., Mateu, J.: Ordinary kriging for function-valued spatial data. Environ. Ecol. Stat. 18, 411–426 (2011)
Giraldo, R., Delicado, P., Mateu, J.: Hierarchical clustering of spatially correlated functional data. Stat. Neerlandica 66, 403–421 (2012)
Giraldo, R., Mateu, J., Delicado, P.: Geofd: an R package for function-valued geostatistical prediction. Revista Colombiana de Estadística 35, 383–405 (2012)
Gonzalez-Manteiga, W., Vieu, P.: Statistics for functional data. Comput. Stat. Data Anal. 51, 4788–4792 (2007)
Gordon, A.D.: Classification, 2nd edn. Chapman & Hall, Boca Raton (1999)
Hartigan, J.A.: Clustering Algorithms. John Wiley, New York (1975)
Hubert, L., Arabie, P.: Comparing partitions. J. Classif. 2, 193–218 (1985)
Jiang, H., Serban, N.: Clustering random curves under spatial interdependence: classification of service accessibility. Technometrics 54, 108–119 (2012)
Jacques, J., Preda, C.: Functional data clustering: a survey. Adv. Data Anal. Classif. 8(3), 231–255 (2014)
Ramsay, J.E., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005)
Rand, W.M.: Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. 66, 846–850 (1971)
Romano, E., Balzanella, A., Verde, R.: Clustering Spatio-functional data: a model-based approach. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin-Heidelberg (2010)
Romano, E., Balzanella, A., Verde, R.: A regionalization method for spatial functional data based on variogram models: an application on environmental data. In: Torelli, N., Pesarin, F., Bar-Hen, A. (eds.) Advances in Theoretical and Applied Statistics. Springer, New York (2010)
Romano, E., Verde, R.: Clustering geostatistical functional data. In: Di Ciaccio, A., Coli, M., Angulo, J.M. (eds.) Advanced Statistical Methods for the analysis of large data-sets, Series Studies in Theoretical and Applied Statistics. Springer, Berlin (2011)
Sun, Y., Genton, M.G.: Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. Environmetrics 23, 54–64 (2012)
Tarpey, T., Kinater, K.: Clustering functional data. J. Classif. 20, 93114 (2003)
Tarpey, T.: Linear transformations and the k-means clustering algorithm: applications to clustering curves. Am. Stat. 61, 34–40 (2007)
Valderrama, M.: An overview to modelling functional data. Comput. Stat. 22, 331–334 (2007)