Journal of Classification 27:54-88 (2010)
DOI: 10.1007/s00357-010-9043-y

A Fuzzy Clustering Model for Multivariate Spatial Time Series

Renato Coppi

Sapienza University of Rome

Pierpaolo D'Urso

Sapienza University of Rome

Paolo Giordani

Sapienza University of Rome

Abstract: Clustering of multivariate spatial-time series should consider: 1) the spatial nature of the objects to be clustered; 2) the characteristics of the feature space, namely the space of multivariate time trajectories; 3) the uncertainty associated to the assignment of a spatial unit to a given cluster on the basis of the above complex features. The last aspect is dealt with by using the Fuzzy *C*-Means objective function, based on appropriate measures of dissimilarity between time trajectories, by distinguishing the cross-sectional and longitudinal aspects of the trajectories. In order to take into account the spatial nature of the statistical units, a spatial penalization term is added to the above function, depending on a suitable spatial proximity/ contiguity matrix. A tuning coefficient takes care of the balance between, on one side, discriminating according to the pattern of the time trajectories and, on the other side, ensuring an approximate spatial homogeneity of the clusters. A technique for determining an optimal value of this coefficient is proposed, based on an appropriate spatial autocorrelation measure. Finally, the proposed models are applied to the classification of the Italian provinces, on the basis of the observed dynamics of some socio-economical indicators.

Keywords: Array of space time data; Fuzzy clustering; Spatial autocorrelation function; Spatial penalization term; Dissimilarity measures between multivariate time trajectories.

This research was partially supported by the grant PRIN 2005 of the Italian Ministry of Education, University and Research ("Models and methods to handle information and uncertainty in knowledge acquisition processes"), which is gratefully acknowledged. The authors also thank the Editor and three anonymous reviewers for their insightful comments, which improved the quality of the paper.

Corresponding Author's Address: Renato Coppi, Department of Statistics, Probability and Applied Statistics, Sapienza University of Rome, P.le A. Moro, 5 - 00185 Rome, Italy, e-mail: renato.coppi@uniroma1.it

Published online 9 April 2010

1. Introduction

This work aims at proposing a new clustering model for classifying spatial units, based on a set of quantitative features observed at several time occasions. In accomplishing this task, the following aspects of the problem should be taken into account: 1) the spatial nature of the objects to be clustered; 2) the characteristics of the feature space, i.e. the space of multivariate time trajectories; 3) the uncertainty associated to the assignment of spatial units to specific clusters, which is increased by the complexity of the feature space. Concerning the first aspect, the main problem is how to exploit the spatial information in the clustering procedure, integrating this information with the one provided by the observed time trajectories pertaining to each spatial unit. One component of the spatial information is usually expressed in terms of a "spatial proximity" matrix (often represented by a "contiguity" matrix). Another component is linked to the spatial autocorrelation of the feature vectors, which should be incorporated in the procedure. The second aspect is related to the treatment of the specific information contained in the observed multivariate time series. In the classification framework it is essential to define appropriate measures of dissimilarity between time trajectories. These are in fact utilized to assess the internal homogeneity of the clusters and the heterogeneity of different clusters. As to the third aspect, the basic requirement is that the clustering technique should allow for a certain degree of flexibility in assigning the spatial units to the clusters, taking into account the complex nature of the information to be utilized to this purpose.

In the present paper, the last problem is dealt with by means of a fuzzy clustering technique, namely the Fuzzy *C*-Means model suitably modified in order to consider the above mentioned spatial information (as we will describe in the sequel). As a consequence, the uncertainty related to the assignment of a unit to a cluster will be expressed in terms of the respective degree of membership, adequately tuned according to an appropriately established fuzziness coefficient. Since the adoption of the Fuzzy *C*-Means criterion requires the introduction of dissimilarity measures on the feature space, constituted by the multivariate time trajectories, these measures are suitably defined. In this respect two types of dissimilarities are described according to whether we want to emphasize the "cross-sectional" or "longitudinal" aspects of the trajectories.

The problem of embodying the spatial information in the clustering procedure is tackled by the introduction of a "regularization/ penalization" term in the Fuzzy *C*-Means objective function. This consists of adding to the traditional part of the function a term which tends to increase the degrees of membership of the units, which are more "close" to a given spatial unit, in the same clusters in which the given spatial unit has higher degrees of membership. This regularization term is tuned by means of a multiplicative coefficient ("spatial penalty" coefficient) which weights its importance within the clustering criterion. The proposed clustering procedure provides a method for determining this coefficient, which exploits the information related to the spatial autocorrelation of the trajectories.

The paper is organized as follows. In Section 2 the approach based on the Fuzzy *C*-Means criterion with the spatial penalization term is described, with reference to the traditional feature space constituted by multivariate vectors. In Section 3 the above clustering model is extended to the treatment of multivariate time trajectories, by distinguishing the cross-sectional and longitudinal cases. In Section 4 the choice of the spatial penalty coefficient is discussed and a method for its determination, based on the spatial autocorrelation of the trajectories, is illustrated. In Section 5 the proposed clustering models are applied to the classification of the Italian provinces, on the basis of the observed dynamics of some socio-economical indicators. Some concluding remarks are made in Section 6.

2. Fuzzy *C*-Means Clustering Models for Spatial Units: The "Spatial Penalty" Approach

Several works devoted to the development of clustering models for spatial units have been proposed in the literature. In general, the peculiarity of these methodologies consists of their capability to suitably deal with the distinguishing characteristics of spatial data, that is, spatial dependence and spatial heterogeneity. A primary distinction among clustering models for spatial data can be made with respect to the objects to be classified:

1) Geographical areas (usually defined by means of administrative boundaries);
2) Pixels (image segmentation).

In class 1), the clustering models aim at determining groups of geographical areas such that the within group dispersion is minimized with the additional assumption that the configuration of the obtained clusters should satisfy particular spatial constraints (e.g., that the obtained clusters are formed by spatially contiguous areas). The empirical evidence suggests that spatial data are often characterized by positive spatial autocorrelation: neighbouring sites tend to have similar features. If such a spatial autocorrelation affects the observed data, this should be explicitly dealt with in the clustering model (instead of arbitrarily ignoring it) so that the resulting clusters may detect it. See, for instance, Gordon (1996). In class 2), the clustering models basically aim at assigning the pixels (i.e. the observation units) in an image to different classes according to their features. The standard clustering models do not take into account the information given by the spatial distribution of the pixels, but only the one given by the observed features. To overcome this problem, clustering algorithms have been adapted by suitably taking into account spatial information.

The clustering models with spatial information are thus distinguished with respect to the standard models in the sense that the features under investigation are both what we may call the "non-spatial" and the "spatial" ones. In particular, when fuzzy clustering techniques are adopted, the risk due to the possible presence of noise is reduced because the spatial information permits to increase or decrease the membership degree of a certain observation unit in a specific cluster according to the membership degrees of its neighbours.

With no claim of completeness, the following works may be assigned to class 1): Lefkovitch (1980), Ferligoj and Batagelj (1982, 1983, 1992), Murtagh (1985), Gordon (1996), Molenaar and Cheng (2000), Costanzo (2001), Di Nola, Loia, and Staiano (2002), Ng and Han (2002), Pilevar and Sukumar (2005), Ayala, Epifanio, Simó, and Zapater (2006), Hu and Sung (2006), Duan, Xu, Guo, Lee, and Yan (2007), Lawson, Simeon, Kulldorff, Biggeri, and Magnani (2007). Among the literature belonging to class 2), it is fruitful to mention Ambroise and Govaert (1998), Tolias and Panas (1998a,b), Pham and Prince (1999), Liew, Leung, and Lau (2000, 2003), Pham (2001), Liew and Yan (2003), Allende and Galbiati (2004), Cinque, Foresti, and Lombardi (2004), Kontos and Megalooikonomou (2005), Marçal and Castro (2005), Tran, Wehrens, and Buydens (2005), Chuang, Tzeng, Chen, Wu, and Chen (2006), Permuter, Francos, and Jermyn (2006), Krooshof, Tran, Postma, Melssen, and Buydens (2006), Xia, Feng, Wang, Zhao, and Zhang (2007). Notice, however, that the assignment of a few papers to either of the classes is somewhat questionable.

In the sequel, we focus our attention on the popular *C*-Means clustering model, suitably generalized to the case of spatial units. Two versions of the *C*-means clustering model are available in the literature: a) in the *Hard C*-Means (H*C*M) clustering model (MacQueen 1967), the observation units are assigned to exactly one of the *C* clusters; b) in the *Fuzzy C*-Means (F*C*M) clustering model (Bezdek 1981), the observation units are assigned to each and every cluster according to the so-called degree of membership.

Let $\mathbf{X}$ be the $(I \times J)$ data matrix concerning a set of I observation units on which the values taken by J variables are recorded. In its standard form, the F*C*M clustering model assigns I observation units to C clusters by minimizing the following loss function:

$$J(\mathbf{U},\mathbf{H},\mathbf{X}) = \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} d^{2}(\mathbf{x}_i,\mathbf{h}_c) = \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} \|\mathbf{x}_i - \mathbf{h}_c\|^{2}, \quad (1)$$

with respect to **U** and **H**. In (1) $\mathbf{x}_i$ denotes the generic i-th row of **X** pertaining to the i-th observation unit and $\mathbf{h}_c$ the generic c-th row of **H**, the centroids matrix of order $(C \times J)$, characterizing the c-th cluster. In F*C*M, the loss function in (1) is minimized subject to appropriate constraints on the $(I \times C)$ membership degrees matrix **U** with generic element u_{ic}, raised to the parameter $m > 1$, which controls the extent of membership sharing between fuzzy clusters. In particular, we impose that

$$u_{ic} \geq 0,\ i = 1, I,\ c = 1, C;\ \sum_{c=1}^{C} u_{ic} = 1,\ i = 1, I. \quad (2)$$

For the generic observation unit i and cluster c, if the degree of membership is close to 1, it follows that the observation unit strongly belongs to the cluster at hand. Instead, the H*C*M clustering model consists of minimizing (1) with m=1 under the following constraints

$$u_{ic} \in \{0,1\},\ i = 1, I,\ c = 1, C;\ \sum_{c=1}^{C} u_{ic} = 1,\ i = 1, I. \quad (3)$$

In the recent years, the F*C*M version has received a great deal of attention because of its flexibility in handling the real world complexity and uncertainty. To corroborate the previous statement, it can be said, as remarked by Hwang, De Sarbo, and Takane (2007), that the fuzzy clustering approach offers other major advantages over traditional clustering methods. First, the fuzzy clustering algorithm is computationally more efficient because dramatic changes in the value of cluster membership are less likely to occur in estimation procedures (McBratney and Moore 1985). Second, fuzzy clustering has been shown to be less afflicted by local optima problems (Heiser and Groenen 1997). Finally, the memberships for any given set of respondents indicate whether there is a second-best cluster almost as good as the best cluster—a result which traditional clustering methods cannot uncover (Everitt, Landau, and Leese 2001). Thus, in the sequel, we shall consider the F*C*M to deal with spatial units.

In order to suitably incorporate the spatial information in the clustering procedure, an important issue to be analyzed is connected with the available information regarding the spatial units. To this purpose a square matrix of order I, say **P**, is introduced. In case of image analysis, **P** is usually constructed as follows (first lag contiguity matrix). It is a symmetric matrix with zero diagonal elements and with off-diagonal elements given by ($i, i' = 1, I; i \neq i'$):

$$p_{ii'} = \begin{cases} 1 & \text{if pixel } i \text{ is contiguous to pixel } i', \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Notice that the diagonal elements of **P** are conventionally set equal to zero in order to allow the algebraic manipulation of **P** within the objective function and exploit the essential information provided by the contiguity between *different* spatial units. Thus, every row (column) of **P** contains 8 elements equal to 1 and *I*-8 zero elements. Obviously, in case of image analysis, the pixels have equal sizes and borderlines. When the spatial units are geographical areas, these are often constructed according to administrative boundaries. Thus, they usually have different sizes and shapes. It follows that several possible matrices describing the spatial information can be built. However, the contiguity matrix in (4) is almost always adopted. Thus, for each geographical area, one can set $p_{ii'} = 1$, if $i' \in N_i$ (N_i denotes the set of geographical areas contiguous to area *i*) and $p_{ii'} = 0$, otherwise (and $p_{ii} = 0$). Several alternatives can be also considered. For instance, it may be useful to assume that the score $p_{ii'}$ concerning spatial units *i* and i' increases according to the increase of the length of their borderline or according to monotone functions of the inverse of their distance. Notice that the specification of **P** when dealing with geographical areas has been extensively discussed in Gordon (1999).

A way for extending the F*C*M clustering model to spatial data consists of adding a suitable spatial penalty term in the minimization problem (1). In fact, (1) is replaced by

$$J(\mathbf{U},\mathbf{H},\mathbf{X}) + J_S(\mathbf{U},\mathbf{P},\mathbf{X}), \tag{5}$$

where $J_S(\mathbf{U},\mathbf{P},\mathbf{X})$ represents the spatial penalty term. The objective function in (5) is minimized with respect to **U** and **H** under the constraints given by (2). The component $J_S(\mathbf{U},\mathbf{P},\mathbf{X})$ can be constructed according to a twofold objective. On one side, it should act in such a way that the membership degree of an observation unit in a given cluster is negatively correlated with the membership degrees of the neighbouring observation units in the other clusters. On the other side, it should also take into account the existing spatial autocorrelation registered on the data set at hand.

Notice that the use of a spatial penalty term does not generally constrain the obtained clusters to be *solely* formed by neighbouring spatial units. Instead it *possibly* leads to clusters formed by neighbouring spatial units. More specifically, suppose that matrix **P** is a contiguity matrix as defined in (4). The spatial penalty term does not imply that contiguous areas *strictly* belong to the same cluster (in the sense that the respective membership degrees are close to 1 or, at least, higher than 0.5). The spatial penalty term *tries* to suitably determine clusters with contiguous areas. Nonetheless, when a few spatial units are characterized by features remarkably different from those of the neighbouring spatial units, the clustering procedure does not prohibit assigning these particular spatial units to different clusters. In this respect, it appears to be very flexible because it takes into account the spatial constraints *to a certain extent*, limiting the risk of altering the obtained clusters. In the literature, a reasonable choice for the spatial penalty term in (5) has been developed by Pham (2001). Such a proposal has been introduced for solving the image segmentation problem. However, it appears to be applicable also to the case of geographical areas. In fact, we have

$$J_S(\mathbf{U},\mathbf{P},\mathbf{X}) = J_S(\mathbf{U},\mathbf{P}) = \frac{\beta}{2}\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'} u_{i'c'}^m , \qquad (6)$$

with $\beta \geq 0$ and $C_c = \{1,\ldots,c-1,c+1,\ldots,C\}$. Notice that multiplying β by one half is useful for mathematical reasons when determining the optimal values of the membership degrees as it will be clear in the next section. The spatial penalty term in (6) is based on the following assumption. When a spatial unit i belongs to cluster c with a high membership degree, then (6) forces the neighbouring spatial units (w.r.t. spatial unit i) to have high membership degrees in cluster c, *as much as possible*. In other words, it is expected that a spatial unit with high (low) membership degree in cluster c, will have neighbouring areas with low (high) membership degrees in the remaining clusters, that is in clusters $k \in C_c$. It follows that (6) attempts to determine a spatially smoothed membership degrees matrix under the empirical evidence that neighbouring areas are often characterized by approximately similar features. Nonetheless, it may also occur that neighbouring geographical areas are described by pretty diverse profiles. In this respect, the parameter β plays the role of increasing or decreasing the emphasis of the spatial penalty term in the minimization of (5). It should be clear that the choice of the spatial parameter β is a complex and relevant issue. In fact, it is desirable that the value of β, to be chosen in advance, would be relatively high in case of concordance between the features of neighbouring geographical areas and relatively low in the opposite case. It is fruitful to observe that, when $\beta = 0$, the spatial smoothness does not act in the clustering problem and (6) reduces to the classical FCM clustering model given in (1) and (2). Extremely high values of β may lead to abnormal results of the clustering procedure in the sense that the spatial units are assigned to the clusters taking into account only the spatial constraints, ignoring, in practice, the observed features. In his work, Pham (2001) recognizes the problem due to the control of the trade-off between the spatial smoothing and the cluster interpretation. In order to solve it, a cross-validation procedure is suggested. See, for further details, Pham (2001). However, such a procedure appears to be of very limited use in case of geographical areas characterized by heterogeneous sizes, shapes and numbers of neighbouring areas. A possible alternative will be proposed in Section 4. It is a heuristic procedure which involves a measure of the spatial autocorrelation registered in the data at hand. In this respect, the choice of the optimal value of β is influenced by the observed data **X**, through the measure of spatial autocorrelation computed within each and every cluster, for any fuzzy partition considered in the procedure. It follows that our proposal leads to a spatial penalty term $J_s(\mathbf{U},\mathbf{P},\mathbf{X})$ as in (5), which suitably takes into account not only the membership degrees matrix **U** and the spatial information contained in **P** but also the data matrix **X**. Such a procedure will be adopted with reference to the clustering problem for geographical areas in case of three-way information, which, contrary to what concerns the two-way case, has received negligible attention in the literature. In the next Section, we will introduce a class of F*C*M models for clustering multivariate time trajectories observed on a set of spatial units. To this purpose we will suitably apply the spatial penalty term proposed by Pham (2001). Then, in Section 4, we will illustrate an appropriate method for choosing the value of the mixing parameter β, taking into account the spatial autocorrelation behaviour of the time trajectories.

3. Fuzzy *C*-means Clustering Models for Multivariate Spatial Time Trajectories

3.1 Spatial Time Data Array and Multivariate Spatial Time Trajectories

By considering an exploratory approach, we analyze the *three-way data array* of type "same objects × same quantitative variables × occasions", in which the *objects* are *spatial units* (geographical or territorial areas, pixels, etc.) and the *occasions* are *times*. This type of three-way data array is called *spatial-time data array*. For instance, suppose that a set of countries is analyzed with respect to some economic indicators during several years. This information can be stored in a spatial-time data array in which the objects are countries, the variables are economic indicators and the occasions are years.

A *spatial-time data array* can be algebraically formalized as follows: $\mathbf{X}\equiv\left\{x_{ijt} : i=1,I; j=1,J;\ t=1,T\right\}$, where i (i=1,I) indicates the spatial unit, j (j=1,J) the variable, and t (t=1,T) the time. Then, the generic element of **X**, x_{ijt}, represents the j-th variable observed on the i-th spatial unit at time t. We can denote $\mathbf{X}$ also in the following way: $\mathbf{X} \equiv \left\{ \mathbf{x}_{i\,t} : \mathrm{i}=1,I; t=1,T \right\}$, where $\mathbf{x}_{i\,t} = (x_{i\,1\,t},...,x_{i\,j\,t},...,x_{i\,J\,t})'$. The spatial-time data array $\mathbf{X}$ can be represented by a bi-dimensional matrix (stacked matrix) by combining two of the three indices i, j, t on the rows and assigning the remaining index to the columns. The various types of stacked matrices are distinguished by their generic "elements":

$$\mathbf{X}_i \equiv \left\{ x_{i\,j\,t} :\ j=1,J;\ t=1,T \right\},$$
$$\mathbf{X}_j \equiv \left\{ x_{i\,j\,t} :\ i=1,I; t=1,T \right\}, \text{ and}$$
$$\mathbf{X}_t \equiv \left\{ x_{i\,j\,t} :\ i=1,I; j=1,J \right\}.$$

Furthermore, the spatial-time data array $\mathbf{X}$ can be geometrically represented on a suitable vectorial space. By representing the elements of one of the three possible classification modes as vectors of a vectorial space, defined with regard to the other ones, we have three possible types of spaces (D'Urso 2000; Coppi and D'Urso 2001): *space of the "spatial units"* ($\Re^{J+1}$), *space of the "variables"* ($\Re^{I+1}$) and *space of the "time occasions"* ($\Re^{I\times J}$).

In this paper, we analyze only the case in which the spatial-time data array $\mathbf{X}$ is represented in the space of the *"spatial units"* $\Re^{J+1}$ (the first J dimensions correspond to the J variables and the last dimension is referred to the time). In this space, each spatial unit i is represented, for each time t, by vector $\mathbf{y}_{i\,t} = (x_{i\,1\,t},...,x_{i\,j\,t},...,x_{i\,J\,t},t)'$, $i=1,I;\quad t=1,T$. For fixed t, matrix $\mathbf{X}_t$ is represented by scatter $S_I(t) \equiv \left\{ \mathbf{y}_{i\,t} : i=1,I \right\}$. Then $\left\{ S_I(t) \equiv \left\{ \mathbf{y}_{i\,t} : i=1,I \right\} : t=1,T \right\}$ represents the set of scatters located on T hyperplanes parallel to the co-ordinate sub-space $\Re^J$. For fixed i, matrix $\mathbf{X}_i$ is represented by scatter $S_T(i) \equiv \left\{ \mathbf{y}_{i\,t} : t=1,T \right\}$ that describes the *trajectory* of the i-th spatial unit during the T time occasions (*spatial-time trajectory*). Then, the set of scatters

$$\left\{ S_T(i) \quad \left\{ \mathbf{y}_{i\,t} : t=1,T \right\} : i=1,I \right\}$$

represents the set of the time trajectories of the I spatial units (*multivariate spatial-time trajectories*). Each spatial-time trajectory crosses the T hyperplanes. In this connection, Figure 1 provides an example of the information contained in a spatial-time data array with I=4, J=2, T=4. In particular, on the upper left side, the map of the I=4 spatial units is displayed, whereas, on the upper right side, the "box" representing the three-way array is given. Finally, on the lower side of Figure 1, the set of scatters $\left\{ S_I(t) \equiv \left\{ \mathbf{y}_{i\,t} : i=1,I \right\} : t=1,T \right\}$ is depicted.

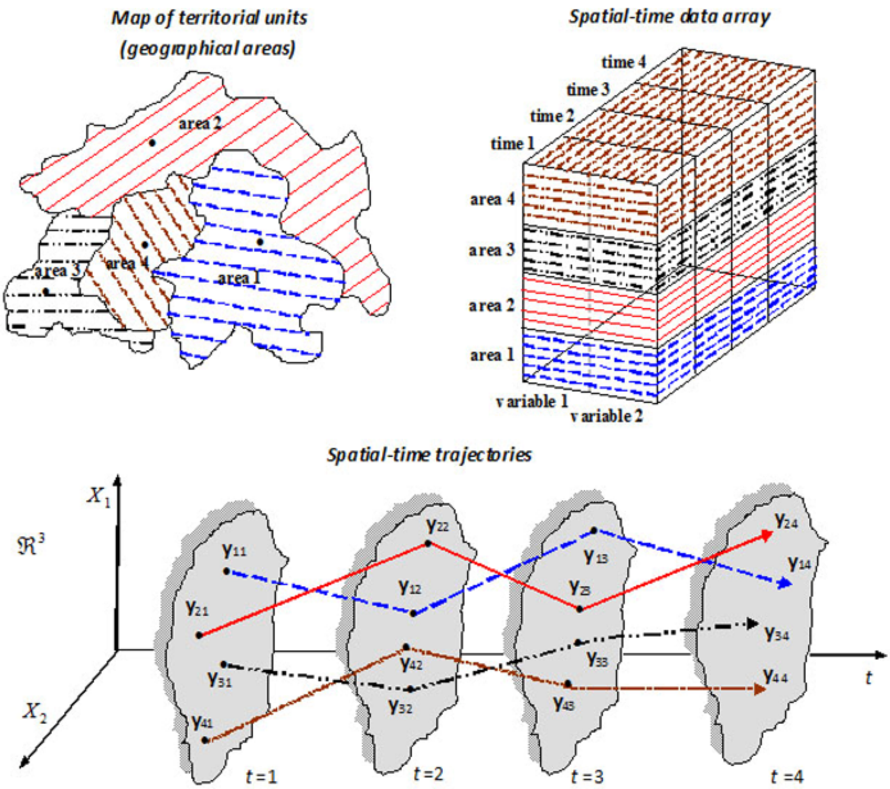


Figure1. Example of map of spatial units, spatial-time array and spatial-time trajectory

3.2 Fuzzy Clustering Models

By considering two types of dissimilarity measures for multivariate trajectories (see, e.g., Coppi and D'Urso 2001, 2006; D'Urso 2005) - i.e. the *cross sectional dissimilarity* that compares the instantaneous (positional) features of the trajectories and the *longitudinal dissimilarity* that captures the differences concerning the evolutive features (i.e. the "variational" patterns) of the trajectories measured by means of their velocities - and following a *fuzzy approach* (D'Urso 2005; Coppi and D'Urso 2006), we propose two types of fuzzy clustering models for classifying spatial units on the basis of multivariate time-varying empirical information, i.e. models for clustering spatial-time trajectories. Notice that, in the clustering process, in the definition of spatial-time trajectory we do not take into account the co-ordinate t, because this does not affect the clustering procedure. Thus, $\mathbf{X}_i \equiv \{ \mathbf{x}_{it} : t=1,T \}$ represents the i-th time trajectory, where $\mathbf{x}_{it} = (x_{i1t},...,x_{ijt},...,x_{iJt})'$, i=1,I; t=1,T.

In a fuzzy framework, in order to cluster I spatial units by considering the instantaneous characteristics of their multivariate time trajectories, i.e. the positions of the multivariate trajectories at each time occasion t (t=1, T) , we propose the following C*ross-Sectional* F*uzzy C-Means clustering model for Spatial-Time Trajectories* (CS-FCM-STT):

$$\begin{cases} \min : J^{CS}(\mathbf{U},\mathbf{H}^{CS},\mathbf{w}^{CS};\mathbf{P},\mathbf{X},C,m,\beta) = \sum_{i=1}^{I}\sum_{c=1}^{C}u_{ic}^{m}\sum_{t=1}^{T}\left(w_t^{CS}\ \left\|\mathbf{x}_{it}-\mathbf{h}_{ct}^{CS}\right\|\right)^2 \\ \qquad\qquad + \frac{\beta}{2}\sum_{i=1}^{I}\sum_{c=1}^{C}u_{ic}^{m}\sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'}u_{i'c'}^{m} & (7) \\ \sum_{c=1}^{C}u_{ic}=1,\ u_{ic}\geq 0; \quad (i=1,I;\, c=1,C) & (8) \\ \sum_{t=1}^{T}w_t^{CS}=1,\ w_t^{CS}\geq 0 \quad (t=1,\mathrm{T}) & (9) \end{cases}$$

where $\mathbf{U}\equiv\left\{u_{ic}: i=1,I; c=1,C\right\}$ is the membership degrees matrix, u_{ic} denotes the membership degree of the i-th spatial-time trajectory in the c-th cluster; $\mathbf{H}_t^{CS}\equiv\left\{\mathbf{h}_{ct}^{CS} : c=1,C; t=1,T\right\}$ is the centroids matrix, $\mathbf{h}_{ct}^{CS}$ is the vector of the c-th centroid at time t from the cross sectional point of view; $\mathbf{w}^{CS}$ $(w_1^{CS},...,w_t^{CS},...,w_T^{CS})$, w_t^{CS} is the t-th instantaneous weight associated to the instantaneous Euclidean distance $\left\|\mathbf{x}_{it}-\mathbf{h}_{ct}^{CS}\right\|$; $\mathbf{P}\equiv\left\{p_{ii'} : i,i'=1,I\right\}$ is a square matrix (I×I) with non-negative elements and p_{ii}=0; m>1 is a weighting exponent that controls the fuzziness of the clustering; $\beta\geq 0$ is the penalty coefficient, which tunes the contribution of the spatial penalization term (see Section 2). Notice that, for the "optimal" choice of m and C, we can consider suitable cluster-validity criteria (see Remark 2 in the sequel); for the appropriate selection of β , see Section 4.

The objective function in (7) is constituted by 2 terms:

- $\sum_{i=1}^{I}\sum_{c=1}^{C}u_{ic}^{m}\sum_{t=1}^{T}\left(w_t^{CS}\ \left\|\mathbf{x}_{it}-\mathbf{h}_{ct}^{CS}\right\|\right)^2$: this term, called *instantaneous within cluster dispersion term*, represents the objective function of the Cross-Sectional Fuzzy C-Means clustering model suggested by D'Urso (2004, 2005). It is a measure of the within cluster (cross-sectional) dissimilarities of the multivariate trajectories w.r.t. the centroids, appropriately weighted by the degrees of membership (taking into account the fuzziness coefficient m). Therefore, by minimizing this term we maximize the internal cohesion of the clusters, conditional on allowing for a certain degree of flexibility as indicated by m. Notice that the dissimilarity measure:

$$\sum_{t=1}^{T} d_w^{CS}\left(\mathbf{x}_{it}, \mathbf{h}_{ct}^{CS}\right)^2 = \sum_{t=1}^{T}\left(w_t^{CS} \left\|\mathbf{x}_{it} - \mathbf{h}_{ct}^{CS}\right\|\right)^2$$

is a sum of squared weighted Euclidean distances between object i and centroid c *at each time point*. It can be denominated "cross-sectional dissimilarity" between the observed time trajectory i and the centroid time trajectory c. We have that weight w_t^{CS} is *intrinsically* associated to the distance $d_w^{CS}\left(\mathbf{x}_{it}, \mathbf{h}_{ct}^{CS}\right)^2$ at time t, whileas the overall dissimilarity is just a sum of the squares of these weighted distances. This allows us to appropriately tune the influence of the various times when computing the dissimilarity between trajectories. In the following it will be clear that the w_t^{CS}'s will constitute specific parameters to be estimated within the clustering model.

- $\frac{\beta}{2}\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'} u_{i'c'}^m$: this represents the *spatial penalty term* (*spatial regularization term*). As it was explained in Section 2, the aim of this term is the following: for each spatial unit i and each cluster c, the sum of the membership degrees of the contiguous/neighboring spatial units (as indicated in matrix **P**) in all the clusters except cluster c (summarized in $C_c = \{1,...,c-1,c+1,...,C\}$) is constrained to be as small as possible. We can observe that the parameter β tunes the trade-off between internal cohesion based on the feature vectors $\mathbf{x}_i$ and the spatial homogeneity of the clusters, i.e.
  - $\beta >> 0 \Rightarrow$ contiguous/neighboring spatial units tend to belong to the same cluster;
  - $\beta = 0 \Rightarrow$ the spatial regularization is not taken into account.

The optimal iterative solution of (7)-(9) is:

$$u_{ic} = \frac{\left[\sum_{t=1}^{T}\left(w_t^{CS}\left\|\mathbf{x}_{it} - {}_1\mathbf{h}_{ct}^{CS}\right\|\right)^2 + \beta\sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'}u_{i'c'}^m\right]^{-1/(m-1)}}{\sum_{c''=1}^{C}\left[\sum_{t=1}^{T}\left(w_t^{CS}\left\|\mathbf{x}_{it} - {}_1\mathbf{h}_{c''t}^{CS}\right\|\right)^2 + \beta\sum_{i'=1}^{I}\sum_{c''\in C_c} p_{ii'}u_{i'c''}^m\right]^{-1/(m-1)}}, \quad (10)$$

$$w_t^{CS} = \frac{1}{\sum_{t'=1}^{T}\frac{\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \left\|\mathbf{x}_{it} - \mathbf{h}_{ct}^{CS}\right\|^2}{\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \left\|\mathbf{x}_{it} - \mathbf{h}_{ct'}^{CS}\right\|^2}}, \quad (11)$$

$$\mathbf{h}_{ct}^{CS} = \frac{\sum_{i=1}^{I} u_{ic}^{m} \mathbf{x}_{it}}{\sum_{i=1}^{I} u_{ic}^{m}}. \tag{12}$$

The previous iterative solution can be obtained by solving the constrained optimization (minimization) problem (7)-(9). The procedure is as follows. By fixing w_t^{CS} and $\mathbf{h}_{ct}^{CS}$, we set equal to zero the partial derivatives with respect to u_{ic} and λ (Lagrange multiplier) of the Lagrangian function:

$$\begin{aligned} L(\mathbf{u}_i, \lambda) = & \sum_{c=1}^{C} u_{ic}^{m} \sum_{t=1}^{T} \left(w_t^{CS} \left\| \mathbf{x}_{it} - \mathbf{h}_{ct}^{cs} \right\| \right)^2 \\ & + \frac{\beta}{2} \sum_{c=1}^{C} u_{ic}^{m} \sum_{i'=1}^{I} \sum_{c' \in C_c} p_{ii'} u_{i'c'}^{m} - \lambda \left(\sum_{c=1}^{C} u_{ic} - 1 \right) \end{aligned}. \tag{13}$$

We get:

$$\begin{cases} \dfrac{\partial L(\mathbf{u}_i, \lambda)}{\partial u_{i''c''}} = 0 \\ \dfrac{\partial L(\mathbf{u}_i, \lambda)}{\partial \lambda} = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} m\, u_{i''c''}^{m-1} \left(\sum_{t=1}^{T} \left(w_t^{CS} \left\| \mathbf{x}_{i''t} - \mathbf{h}_{c''t}^{cs} \right\| \right)^2 + \beta \sum_{i'=1}^{I} \sum_{c' \in C_c} p_{i''i'} u_{i'c'}^{m} \right) - \lambda = 0 & (14) \\ \sum_{c=1}^{C} u_{ic} - 1 = 0. & (15) \end{cases}$$

Notice that in (14) the coefficient (1/2) pertaining to β vanishes because the derivative results in a term corresponding to the product of u_{ic} and its neighbours, plus a term corresponding to the inverse product of the neighbours and u_{ic}. From (14) and by taking into account (15) we get (10). Then, by fixing u_{ic} and $\mathbf{h}_{ct}^{CS}$, we can compute w_t^{CS} in a similar manner. In fact, by setting equal to zero the partial derivatives with respect to w_t^{CS} and ξ (Lagrange multiplier) of the Lagrangian function:

$$\begin{aligned} L(\mathbf{w}^{CS}, \xi) = & \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^{m} \sum_{t=1}^{T} \left(w_t^{CS} \left\| \mathbf{x}_{it} - \mathbf{h}_{ct}^{CS} \right\| \right)^2 \\ & + \frac{\beta}{2} \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^{m} \sum_{i'=1}^{I} \sum_{c' \in C_c} p_{ii'} u_{i'c'}^{m} - \xi \left(\sum_{t=1}^{T} w_t^{CS} - 1 \right) \end{aligned}. \tag{16}$$

We get:

$$\begin{cases} \dfrac{\partial\, L_m(\mathbf{w}^{CS},\xi)}{\partial\, w_{t'}^{CS}} = 0 \\ \dfrac{\partial\, L_m(\mathbf{w}^{CS},\xi)}{\partial\, \xi} = 0 \end{cases} \Leftrightarrow \begin{cases} 2\, w_{t'}^{CS} \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} \sum_{t=1}^{T}\left(w_{t'}^{CS}\, \left\| \mathbf{x}_{i\,t'} - \mathbf{h}_{ct'}^{CS} \right\| \right)^2 \\ \quad -\xi = 0 & (17) \\ \sum_{t=1}^{T} w_t^{CS} - 1 = 0. & (18) \end{cases}$$

From (17) and using (18) we obtain (11). Finally, we can obtain (12) by solving an unconstrained minimization problem. The iterative algorithm for computing the optimal iterative solution is given in Table 1.

Remark 1 (Weights w_t^{CS})

Looking at the solution equation (11), we observe that the time weights have a statistical meaning. In fact, they appear to mirror the heterogeneity of the total intracluster "deviances" (i.e. $\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} \left\| \mathbf{x}_{i\,t} - \mathbf{h}_{ct}^{CS} \right\|^2$) across the different times. In particular, weight w_t^{CS} increases as long as the total intracluster "deviance" at time *t* decreases (compared with the remaining time occasions). Thus, the optimization procedure tends to give more emphasis to the time occasions capable to increase the within cluster similarity among the trajectories.

Remark 2 (Particular Cases)

From model (7)-(9), we can derive the following particular cases:

- *Cross-Sectional Fuzzy C-Means clustering model for Time Trajectories* (D'Urso 2004, 2005): by assuming that the *I* objects are not spatial or territorial units (β=0);
- *Spatial Fuzzy C-Means clustering model* (Pham 2001): by putting in (7) *T*=0 (see more detailed comparative comments in Sec. 6);
- *"Traditional" Fuzzy C-Means clustering model* (Bezdek 1981): by assuming that the *I* objects are not spatial or territorial units (β=0) and by putting in (7) *T*=0.

Remark 3 (Preprocessing)

In several occasions, the scores of the variables present artificial differences. Obviously, this may cause biased results. We thus suggest to

Table 1: Iterative algorithm for determining the optimal solution (CS-F*C*M-STT)

Step 0	Fix *m*, *C*. Randomly generate a suitable membership degrees matrix $\mathbf{U}^{(a)}$ with $a = 0$;
Step 1	Update the centroids $\mathbf{h}_{ct}^{CS(a+1)}$ according to (12);
Step 2	Update the time weights $w_t^{CS(a+1)}$ according to (11);
Step 3	Update the degrees of membership $u_{ic}^{(a+1)}$ according to (10);
Step 4	If $\left\Vert \mathbf{U}^{(a+1)} - \mathbf{U}^{(a)} \right\Vert < \tau$, where $\tau > 0$ is fixed in advance, then the algorithm has converged; otherwise set $a = a + 1$ and go to Step 1.

preprocess the data by normalization. More specifically, let x_{ijt}, $i = 1, I$; $j = 1, J; t = 1, T$, be the generic element of the observed three-way data matrix. We then get the generic normalized datum as:

$$\frac{x_{ijt}}{\overline{x}_{\cdot j \cdot}},\ i = 1, I; j = 1, J; t = 1, T;\ \text{where}\ \overline{x}_{\cdot j \cdot} = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T} x_{ijt}}{IT}.$$

Notice, however, that alternative pre-processing procedures can also be adopted (for more details, see, e.g, Coppi and D'Urso (2006)).

Remark 4 (Local Optima)
It should be underlined that the iterative algorithm given in Table 1 converges to, at least, a local optimum. To limit the risk of hitting local optima, more than one random start is recommended. However, it is recognized that fuzzy *C*-means clustering algorithms present a minor tendency of hittting local optima with respect to their traditional counterparts (see e.g., Bezdek, Keller, Krisnapuram, and Pal 1999). Moreover, empirical studies have shown that the fuzzy clustering algorithm is an efficient starting point for traditional clustering (Heiser and Groenen 1997).

For classifying a set of *I* spatial-time trajectories in a fuzzy manner, by taking into account the inter-temporal variations of the trajectories, we propose the following *Longitudinal Fuzzy C-Means Clustering model for Spatial-Time Trajectories* (L-F*C*M-STT):

$$\begin{cases} \min : J^L(\mathbf{U},\mathbf{H}^L,\mathbf{w}^L;\mathbf{P},\mathbf{X},C,m,\beta) = \sum_{i=1}^{I}\sum_{c=1}^{C} u_{i\,c}^m \sum_{t=2}^{T}\left(w_t^L \left\|\mathbf{v}_{i\,t}-\mathbf{h}_{ct}^L\right\|\right)^2 \\ \qquad\qquad + \frac{\beta}{2}\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'}u_{i'c'}^m & (19) \\ \sum_{c=1}^{C} u_{i\,c} = 1,\ u_{i\,c}\geq 0; \quad (i=1,I;\ c=1,C) & (20) \\ \sum_{t=2}^{T} w_t^L = 1,\ w_t^L \geq 0 \quad (t=2,T) & (21) \end{cases}$$

where $\mathbf{H}^L \equiv \left\{ \mathbf{h}_{ct}^L :\ c=1,C;t=1,T \right\}$ is the centroids velocity matrix, $\mathbf{h}_{ct}^L$ is the velocity vector of the trajectory of the c-th centroid in $[t-1,\ t]$ from the longitudinal point of view and $\mathbf{v}_{i\,t} = (\mathbf{x}_{i\,t} - \mathbf{x}_{i\,t-1})$ is the velocity vector of the trajectory of the i-th spatial unit; $\mathbf{w}^L \equiv (w_1^L,...,w_t^L,...,w_T^L)'$, w_t^L is a weight in $[t-1, t]$ (“*velocity” weight)*; $\left\|\mathbf{v}_{i\,t}-\mathbf{h}_{ct}^L\right\|$ is the Euclidean distance between the velocities of the time trajectory of the i-th spatial unit and the c-th centroid in $[t-1,\ t]$. In this case, $\sum_{i=1}^{I}\sum_{c=1}^{C} u_{i\,c}^m \sum_{t=1}^{T}\left(w_t^L \left\|\mathbf{v}_{i\,t}-\mathbf{h}_{ct}^L\right\|\right)^2$ represents the *longitudinal within cluster dispersion term*.

In the longitudinal case, the optimal iterative solution of model (19)-(21) is:

$$u_{ic} = \frac{\left[\sum_{t=2}^{T}\left(w_t^L\left\|\mathbf{v}_{it}-\mathbf{h}_{ct}^L\right\|\right)^2 + \beta\sum_{i'=1}^{I}\sum_{c'\in C_c} p_{ii'}u_{i'c'}^m\right]^{-1/(m-1)}}{\sum_{c''=1}^{C}\left[\sum_{t=2}^{T}\left(w_t^L\left\|\mathbf{v}_{it}-\mathbf{h}_{c''t}^L\right\|\right)^2 + \beta\sum_{i'=1}^{I}\sum_{c''\in C_c} p_{ii'}u_{i'c''}^m\right]^{-1/(m-1)}}, \tag{22}$$

$$w_t^L = \frac{1}{\sum_{t'=2}^{T}\frac{\sum_{i=1}^{I}\sum_{c=1}^{C}u_{ic}^m\left\|\mathbf{v}_{it}-\mathbf{h}_{ct}^L\right\|^2}{\sum_{i=1}^{I}\sum_{c=1}^{C}u_{ic}^m\left\|\mathbf{v}_{it}-\mathbf{h}_{ct'}^L\right\|^2}}, \tag{23}$$

$$\mathbf{h}_{ct}^L = \frac{\sum_{i=1}^{I}u_{ic}^m\mathbf{v}_{it}}{\sum_{i=1}^{I}u_{ic}^m}. \tag{24}$$

The analytic proof for obtaining the previous iterative solution (22)-(24) is similar to the one illustrated in the cross-sectional case. Also the iterative algorithm is the same.

Remark 5 (Weights w_t^L)

As noticed in Remark 1, time weights w_t^L (now referring to pairs of successive time occasions) are connected with the heterogeneity of the total intracluster "deviances". In the longitudinal case this is obviously related to the trajectories of "velocities". Larger weights are then given to the time intervals which increase the within cluster similarity among the above trajectories.

We can take into account, simultaneously, the information provided by the cross sectional and longitudinal models by considering a consensus method or a mixed clustering model that summarizes suitably the instantaneous and longitudinal features of the spatial-time trajectories in the clustering process. In this case, the benefits of the cross sectional and longitudinal models are inherited by the mixed model.

4. A Method for Selecting the Penalty Coefficient

As we already observed in Section 2, the selection of the optimal value of β is a complex issue. A possible way to solve it is represented by the following heuristic procedure. Notice that we assume that the values of C and m have already been chosen. For every specified value of β, the obtained clusters are constructed in such a way that the within cluster dispersion is minimized. However, it would be also desirable that all clusters are characterized by the maximal within cluster spatial autocorrelation. To this purpose, for fixed values of C and m, it is advisable to run the clustering algorithm for increasing values of β (e.g. from 0 to β_{Max}, with $\beta_{Max}>0$ chosen in advance and with increasing steps equal to $\beta_{Inc}>0$) and to choose the optimal value of β in such a way that the within cluster spatial autocorrelation is maximized.

4.1. Cross-Sectional Case

In order to evaluate the spatial autocorrelation, we define a sort of "compromise" matrix $\mathbf{X}_{Comp}$ over the T time occasions using the optimal weights $w_t^{CS}, t=1,T$. Thus, we get

$$\mathbf{X}_{Comp} = \sum_{t=1}^{T} w_t^{CS} \mathbf{X}_t \, , \tag{25}$$

which has order $(I \times J)$. In other words, $\mathbf{X}_{Comp}$ is a weighted mean of the data matrices collected during all the time occasions with weights given by w_t^{CS}'s.

In the univariate case (J=1), matrix $\mathbf{X}_{Comp}$ reduces to vector $\mathbf{x}_{Comp}$ and for the generic c-th cluster, $c=1,C$, a measure of the (univariate) spatial autocorrelation can be computed as

$$\rho_c = \frac{\left(\mathbf{x}_{Comp} - \overline{\mathbf{x}}_{Comp}\right)' \mathbf{U}_c^{1/2} \mathbf{P} \mathbf{U}_c^{1/2} \left(\mathbf{x}_{Comp} - \overline{\mathbf{x}}_{Comp}\right)}{\left(\mathbf{x}_{Comp} - \overline{\mathbf{x}}_{Comp}\right)' \mathbf{U}_c^{1/2} diag\left(\mathbf{P}'\mathbf{P}\right) \mathbf{U}_c^{1/2} \left(\mathbf{x}_{Comp} - \overline{\mathbf{x}}_{Comp}\right)}, \qquad (26)$$

where $\overline{\mathbf{x}}_{Comp}$ denotes the I-vector with elements equal to the average over I of the values $\overline{\mathbf{x}}_{Comp}$ and $\mathbf{U}_c$ is the square diagonal matrix of order I with generic diagonal element $u_{c\,ii}$=u_{ic}, $i=1,I$. Finally, $diag(\cdot)$ is the operator that creates a diagonal matrix whose elements in the main diagonal are the same as those of the square matrix in the argument. If $\mathbf{P}$ is a contiguity matrix as defined in (4), every diagonal element of $diag\left(\mathbf{P}'\mathbf{P}\right)$ contains the number of neighboring areas for the associated spatial unit. In order to take into account the structure of the c-th cluster, it is worth observing that in (26) the scores of $\mathbf{P}$ are scaled by means of the membership degrees of the spatial units in the cluster involved. It is important to observe that (26) is very similar to the measure proposed by Moran (1950). The main difference concerns matrix $\mathbf{U}_c$, which tunes the contribution of the neighbours. Also notice that measure (26) resembles the one proposed in Smouse and Peakall (1999) of which it represents a possible extension.

In order to develop a measure of spatial autocorrelation for the multivariate case ($J > 1$), it is convenient to rewrite (26) as

$$\rho_c = \frac{\left(\mathbf{Q}\mathbf{x}_{Comp}\right)' \mathbf{U}_c^{1/2} \mathbf{P} \mathbf{U}_c^{1/2} \left(\mathbf{Q}\mathbf{x}_{Comp}\right)}{\left(\mathbf{Q}\mathbf{x}_{Comp}\right)' \mathbf{U}_c^{1/2} diag\left(\mathbf{P}'\mathbf{P}\right) \mathbf{U}_c^{1/2} \left(\mathbf{Q}\mathbf{x}_{Comp}\right)}. \qquad (27)$$

in which $\mathbf{Q} = \mathbf{I}_I - \frac{\mathbf{1}_I \mathbf{1}_I'}{I}$ is the centering operator, where $\mathbf{I}_I$ is an identity matrix of order I and $\mathbf{1}_I$ is a column-vector of order I with unit elements. Taking into account (27), the spatial autocorrelation measure for the c-th cluster can be suitably extended as

$$\rho_c = \frac{tr\left(\left(\mathbf{Q}\mathbf{X}_{Comp}\right)' \mathbf{U}_c^{1/2} \mathbf{P} \mathbf{U}_c^{1/2} \left(\mathbf{Q}\mathbf{X}_{Comp}\right)\right)}{tr\left(\left(\mathbf{Q}\mathbf{X}_{Comp}\right)' \mathbf{U}_c^{1/2} diag\left(\mathbf{P}'\mathbf{P}\right) \mathbf{U}_c^{1/2} \left(\mathbf{Q}\mathbf{X}_{Comp}\right)\right)}, \qquad (28)$$

In order to determine an overall spatial autocorrelation measure for the obtained partition, we can compute the weighted mean of the measures in (26) or (28) with weights equal to the sum over the *I* spatial units of the membership degrees in the *C* clusters. Specifically, we obtain:

$$\rho_{Overall} = \frac{\sum_{c=1}^{C} \rho_c s_c}{\sum_{c=1}^{C} s_c} = \frac{\sum_{c=1}^{C} \rho_c s_c}{I}, \tag{29}$$

with $s_c = \sum_{i=1}^{I} u_{ic}$. After running the clustering models for increasing values of β, we choose that value β_{Opt} such that $\rho_{Overall}$, as defined in (29), is maximal. Summing up, the following algorithm can be adopted.

4.2. Longitudinal Case

In the longitudinal case, the three-way data matrix to be managed contains the velocities pertaining to the trajectories of the spatial units in $[t-1,t]$, $t=2,T$. In fact, the data corresponding to every interval of time $[t-1,t]$ are displayed in the (two-way) velocity matrices:

$$\mathbf{V}_t = \{v_{ijt} = x_{ijt} - x_{ijt-1}; i=1,I; j=1,J\},\ t=2,T.$$

The compromise matrix can then be defined as the weighted mean of matrices $\mathbf{V}_t$'s using weights w_t^L's. Therefore, we have:

$$\mathbf{V}_{Comp} = \sum_{t=2}^{T} w_t^L \mathbf{V}_t. \tag{30}$$

The spatial autocorrelation measure for the *c*-th cluster can then be derived using $\mathbf{V}_{Comp}$ instead of $\mathbf{X}_{Comp}$, as it was in (28), which is replaced by

$$\rho_c = \frac{tr\left(\left(\mathbf{QV}_{Comp}\right)' \mathbf{U}_c^{1/2} \mathbf{P} \mathbf{U}_c^{1/2} \left(\mathbf{QV}_{Comp}\right)\right)}{tr\left(\left(\mathbf{QV}_{Comp}\right)' \mathbf{U}_c^{1/2} diag\left(\mathbf{P}'\mathbf{P}\right) \mathbf{U}_c^{1/2} \left(\mathbf{QV}_{Comp}\right)\right)}. \tag{31}$$

Once again, a measure of the overall spatial autocorrelation can be determined according to (29) using, of course, (31). A similar algorithm to the one described in Table 2 can be adopted.

Table 2: Iterative algorithm for determining the optimal value of β (CS-F*C*M-STT)

Fix m, C, β_{Max} and β_{Inc}. Set $\beta = 0$, $\beta_{Opt} = 0$ and $\rho_{Opt} = 0$;

While $\beta \leq \beta_{Max}$; do:

Run the CS-F*C*M-STT model and compute $\rho_{Overall}(\beta)$ according to (29);

If $\rho_{Overall}(\beta) > \rho_{Opt}$; $\rho_{Opt} := \rho_{Overall}(\beta)$ and $\beta_{Opt} := \beta$;

$\beta := \beta + \beta_{Inc}$;

End.

5. Application

In order to show how our models work in practice, the results of an application to Italian socio-demographical data are presented and discussed. Specifically, the data under investigation refer to the values of J=7 variables observed on I=103 Italian provinces during T=5 time occasions (years 1995-1999). The J=7 variables on which our analysis is carried out are: Number of Foreign Residents per 1.000 inhabitants (hereinafter NFR), Rate of Natural Increase (RNI), Divorce Rate (DR), Number of Suicides per 100.000 inhabitants (NS), Average Expense per inhabitant for Theatrical and Musical Performances (ETMP), Number of Movie Halls open per 100.000 inhabitants (NMH), Rate of Unemployment (RU). Note that the data considered have been previously normalized as described in Remark 3. We need preprocessing them because the scores of the indicators at hand present artificial differences. Before performing the analysis, we expect that the data under examination present a high spatial autocorrelation. In fact, the features of the Italian provinces are usually highly affected by their geographical positions. Therefore, the use of the FCM clustering models with spatial penalty term seems to be a reasonable choice in view of obtaining clusters embodying spatial information.

5.1. Cross-Sectional Case

In order to determine groups of provinces characterized by homogeneous features together with a high spatial autocorrelation we first perform the CS-F*C*M-STT clustering model using the contiguity matrix as defined in (4). We decide to set C=3 and m=1.5. The choice of C=3 can be explained by observing that the Italian socio-demographical structure is usually distinguished with respect to three main areas: Northern Italy, Central Italy and Southern Italy. We expect that the clustering models will approximately find these three clusters of geographical areas on the basis of the variables at hand. Furthermore, we set m=1.5 because some prior investigations with the preprocessed data indicated that this allows us both to well determine homogeneous spatial units belonging to the given clusters and, at the same time, to detect those spatial units for which the memberships in the given clusters are rather fuzzy. Those spatial units will have relatively low membership degrees in all the clusters. If the maximal membership degree will be lower than 0.50, we conclude that the features of the corresponding spatial unit do not permit to exactly assign it to any cluster. Instead, when the maximal membership degree will be slightly higher than 0.50, the spatial unit at hand is assigned to a given cluster (hard clustering), but the extent to which it belongs to the cluster involved is pretty low and, implicitly, the associated fuzziness is high.

Finally, in order to select β in such a way that the within cluster spatial autocorrelation measure is maximized, we run the procedure described in Section 4.1 setting β_{Max}=7 and β_{Inc}=0.05. The values of the spatial autocorrelation measure according to (29), obtained by using increasing values of β from 0 to 7, are displayed in Figure 2. Notice that for each value of β, we run the clustering algorithm using 100 random starts in order to limit the risk of hitting local optima.

By inspecting Figure 2, we can observe that the spatial auto-correlation measure first increases (for $\beta \leq 0.65$) and takes the maximal value (0.44) when $\beta_{Opt} = 0.65$. In fact, for values of β higher than 0.65 we find that $\rho_{Overall}$ strictly decreases until β=3.80 (with value 0.36). Then, for values of β higher than 3.80, a decreasing trend is visible, which however seems to converge on value 0.33.

It is interesting to stress that, for low values of β, the obtained clusters are not formed by spatial units for which the overall spatial autocorrelation measure is high. Specifically, the resulting clusters are only partially composed by neighboring spatial units. Of course, it follows that $\rho_{Overall}$ takes low values because the more neighboring spatial units do not belong to the same clusters, the more the elements of $\mathbf{U}_c^{1/2}\mathbf{P}\mathbf{U}_c^{1/2}$ are close to 0 and, therefore, the numerator of (28) (on which (29) is based) takes small values. By contrast, for high values of β, the clusters tend to be solely formed by neighboring spatial units. In this case, the features of the spatial units play a limited role in the clustering process because the resulting clusters are mainly determined according to the spatial information. This implies that the values of $\rho_{Overall}$ are low because the within cluster spatial autocorrelation is low (the numerator of (28) is small because every cluster is formed by spatial units with heterogeneous

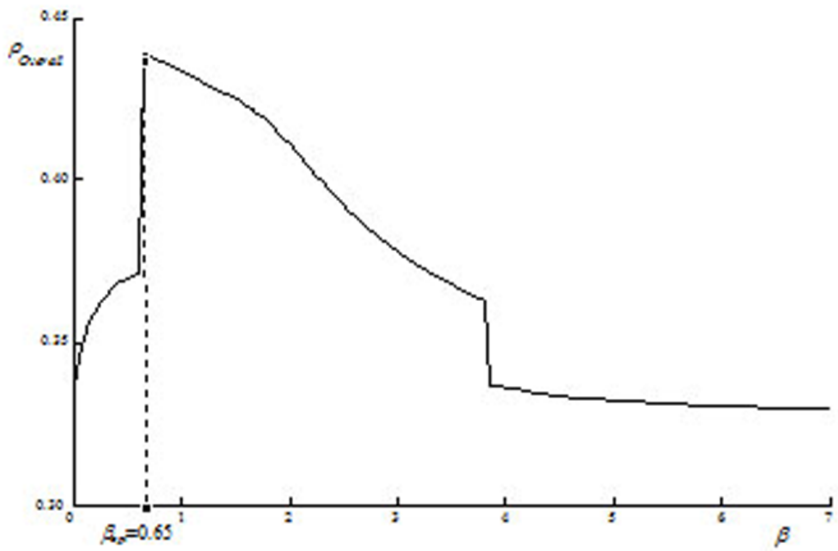


Figure 2. Value of the spatial autocorrelation measure according to (29) based on (28)

observed features). Finally, the clusters obtained in correspondence of β_{Opt} are constructed in such a way that the neighboring spatial units belonging to a given cluster are those with the more similar features, leading to the highest within cluster spatial autocorrelation. Therefore, in our application, we perform the CS-F*C*M-STT clustering model setting $\beta = 0.65$ (with C=3 and m=1.5). The obtained centroids (after applying the inverse pre-processing procedure) are displayed in Figure 3 and the membership degrees of the provinces in the clusters are summarized in Figure 4. Figure 4 is helpful in order to determine the memberships of the various provinces in the clusters and the associated (maximal) degrees. In particular, the colours (dark grey, grey or light grey) of the provinces highlight the assigned cluster in the hard clustering sense (membership degree higher than 0.50). In order to show the membership degrees of the provinces in the clusters, we report the associated (maximal) membership degrees. Moreover, some spatial units are not assigned to any cluster (all the membership degrees are lower than 0.50). These are depicted by white surfaces.

By inspecting Figure 4, we can observe that Cluster 3 (light grey coloured) is totally formed by provinces located in Southern Italy. Therefore, such a result seems to corroborate the existence of different patterns between Northern and Southern Italy. For what concerns the features under investigation, Figure 3 shows that the provinces belonging to Cluster 3 cohabit with an ever-during high rate of unemployment (19% on average). The population is getting older (the RNI values decrease

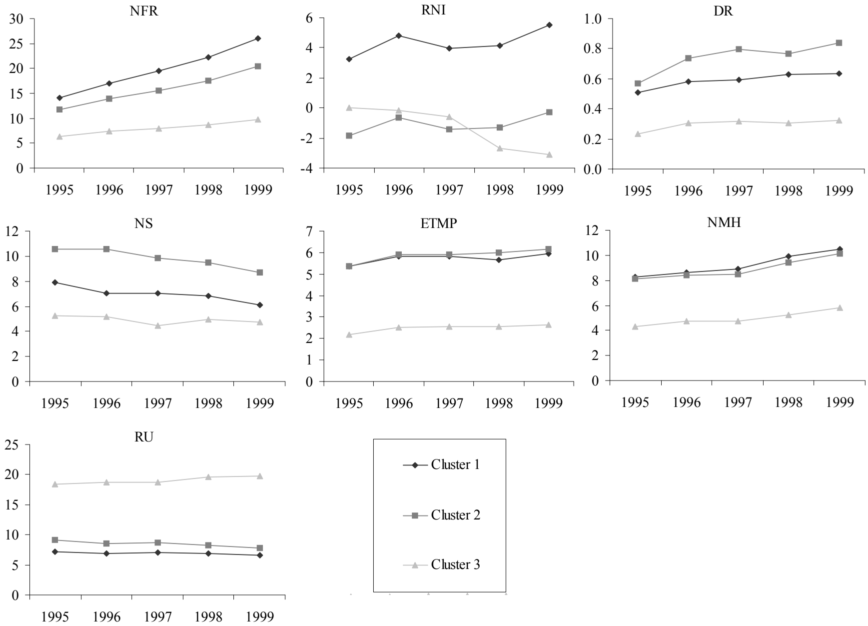
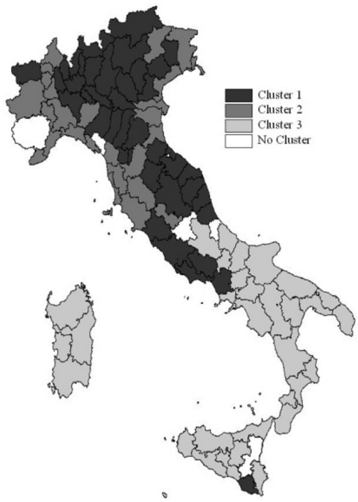


Figure 3. Centroids (after applying the inverse preprocessing procedure) resulting from CS-F*C*M-STT with $\beta = 0.65$, $C = 3$ and $m = 1.5$.

during the reference time starting from score 0.00 in 1995) and the number of foreign residents (NFR) is uniformly lower if compared with those of the other provinces. With respect to the other socio-demographical variables, we can see that the divorce rate (DR) is very low as well as the number of suicides (NS). The latter finding is an established sociological result: there exists an empirical trade-off between the number of suicides and the rate of unemployment. In other words, when the well-being increases, glum conditions more frequently arise. Finally, the ETMP and NMH scores are the lowest ones (if compared with those pertaining to the centroids of the other two clusters), thereby emphasizing a lower cultural level of Southern Italy with respect to the Northern area.

Clusters 1 (dark grey coloured) and 2 (grey coloured) are mainly composed by provinces located, respectively, in Northern and Central Italy, except for Caserta (membership degree = 0.72) and Ragusa (0.78) belonging to Cluster 1. Going into detail, Cluster 1 is mostly formed by the provinces belonging to some regions located in Northern-Eastern and Central areas of Italy (Lombardy, Veneto, Trentino Alto Adige, Marche, Emilia-Romagna and Tuscany). Notice that these provinces form a few



Prov.	Clus.	Memb. degree	Prov.	Clus.	Memb. degree	Prov.	Clus.	Memb. degree
Torino	2	0.95	Modena	1	0.96	Foggia	3	0.99
Vercelli	2	0.87	Bologna	1	0.70	Bari	3	0.54
Novara	1	0.84	Ferrara	2	0.62	Taranto	3	0.98
Cuneo	(2)	0.50	Ravenna	2	0.74	Brindisi	3	0.88
Asti	2	0.89	Forlì	1	0.77	Lecce	3	0.97
Alessandria	2	0.90	Pesaro-Urbino	1	0.99	Potenza	3	0.78
Aosta	1	0.71	Ancona	1	0.87	Matera	3	0.95
Imperia	2	0.91	Macerata	1	0.99	Cosenza	3	0.99
Savona	2	0.92	Ascoli Piceno	1	0.94	Catanzaro	3	0.99
Genova	2	0.61	Massa-Carrara	(2)	0.49	Reggio Calabria	3	0.98
La Spezia	2	0.60	Lucca	2	0.71	Trapani	3	0.85
Varese	1	0.90	Pistoia	1	0.85	Palermo	3	0.95
Como	1	0.99	Firenze	2	0.64	Messina	3	0.94
Sondrio	1	0.54	Livorno	2	0.97	Agrigento	3	0.93
Milano	1	0.83	Pisa	2	0.89	Caltanissetta	3	0.85
Bergamo	1	0.98	Arezzo	1	0.95	Enna	3	0.70
Brescia	1	0.97	Siena	2	0.55	Catania	(1)	0.49
Pavia	1	0.51	Grosseto	2	0.89	Ragusa	1	0.78
Cremona	1	0.91	Perugia	1	0.95	Siracusa	3	0.76
Mantova	1	0.89	Terni	2	0.69	Sassari	3	0.93
Bolzano	1	0.97	Viterbo	1	0.73	Nuoro	3	0.96
Trento	1	0.98	Rieti	(1)	0.46	Cagliari	3	0.92
Verona	1	0.79	Roma	1	0.74	Pordenone	1	0.72
Vicenza	1	0.97	Latina	1	0.93	Isernia	3	0.79
Belluno	2	0.56	Frosinone	1	0.71	Oristano	3	0.91
Treviso	1	0.93	Caserta	1	0.72	Biella	2	0.93
Venezia	2	0.73	Benevento	3	0.96	Lecco	1	0.99
Padova	1	0.97	Napoli	3	0.82	Lodi	1	0.94
Rovigo	2	0.64	Avellino	3	0.97	Rimini	1	0.98
Udine	2	0.91	Salerno	3	0.65	Prato	1	0.93
Gorizia	2	0.88	L'Aquila	3	0.51	Crotone	3	0.69
Trieste	2	0.59	Teramo	1	0.93	Vibo Valentia	3	0.90
Piacenza	2	0.74	Pescara	(1)	0.45	Verbano-C.-O.	2	0.76
Parma	1	0.70	Chieti	3	0.58			
Reggio Emilia	1	0.86	Campobasso	3	0.87			

Figure 4. Membership degrees resulting from CS-F*C*M-STT with $\beta = 0.65$, $C = 3$ and $m = 1.5$.

contiguous areas. In fact, together with the above-mentioned Southern provinces (Caserta and Ragusa), the only exception is Aosta. Therefore, Clusters 1 and 3 appear to be strongly affected by the spatial information. This holds to a smaller degree with respect to Cluster 2. In fact, all the provinces assigned to such a cluster are located in Northern and Central Italy. Nonetheless, some of them can be found in the Western part of Italy (regions Piedmont and Liguria) and some others in the Eastern part of Italy (Friuli Venezia Giulia). Moreover, some provinces of Emilia-Romagna and Tuscany (those not belonging to Cluster 1) also form Cluster 2.

A deeper insight into Clusters 1 and 2 can be attained by examining Figure 3. The analysis of the centroids allows us to state that the cultural life is pretty much the same as well as for the RU values, even if to a lesser extent. The figures concerning the DR and NS scores show that a more negative condition can be found in the provinces assigned to Cluster 2. This is also confirmed by the RNI values, which are uniformly decreasing and always lower than 0 during years 1995-1999. The NFR scores for Cluster 2 are increasing but always lower than those pertaining to Cluster 1. It is useful to notice that the RNI scores for Cluster 1 are always positive. This can be probably explained taking into account the continuously increasing number of foreign residents, which entails a high number of births (due to the younger age of the immigrants).

Finally, only five provinces are not assigned to any clusters. These are Rieti, Pescara and Catania (the maximal membership degrees are, respectively, 0.46, 0.45 and 0.49 in Cluster 1), Massa-Carrara (its membership degree in Cluster 2 is 0.49) and Cuneo (with membership degree in Cluster 2 slightly lower than 0.50).

The obtained time weights are as follows: $w_1^{CS} = 0.21$, $w_2^{CS} = 0.20$, $w_3^{CS} = 0.21$, $w_4^{CS} = 0.24$ and $w_5^{CS} = 0.14$. Thus, the first four time occasions approximately play a similar role in the clustering process, whereas the information concerning year 1999 appears to be less important (lowest time weight).

Let us now compare the above results with those obtained performing CS-F*C*M-STT without spatial penalty term, that is setting $\beta = 0$. In this case, the here-proposed clustering model coincides with the one introduced by D'Urso (2004). We now aim at evaluating whether a geographical classification of the Italian provinces can be still detected. We thus investigate to which extent the clusters can be interpreted in terms of the spatial location of the provinces. For comparative purposes, once again we set $C = 3$ and $m = 1.5$ and 100 random starts are considered. The centroids and the membership degrees are given in Figures 5 and 6, respectively.

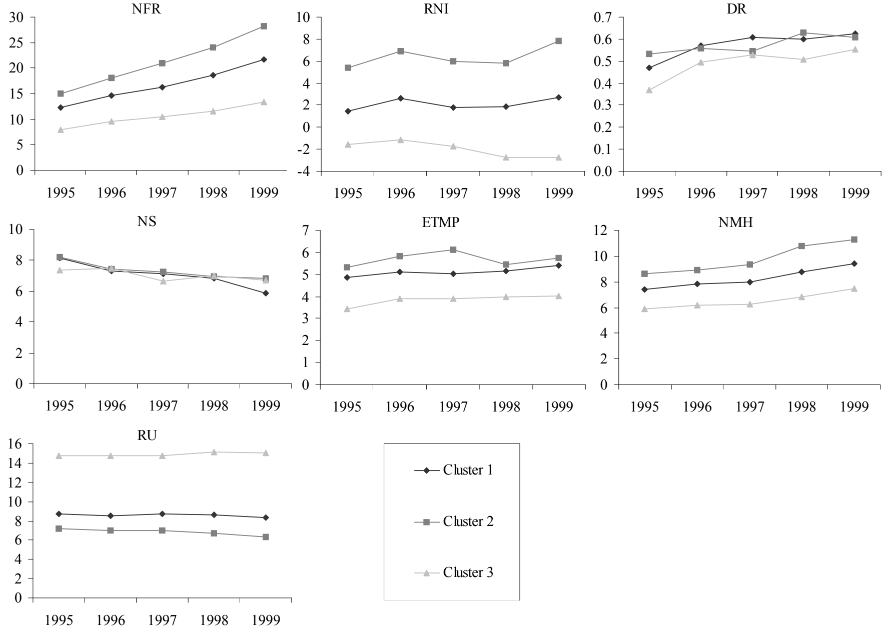
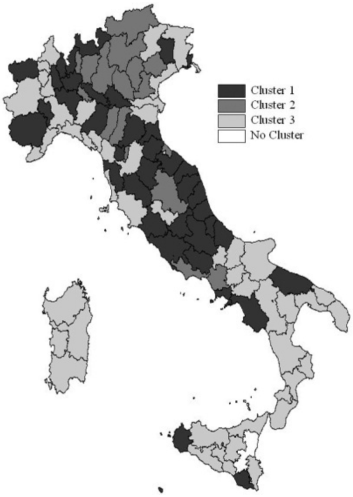


Figure 5. Centroids (after applying the inverse preprocessing procedure) resulting from from CS-F*C*M-STT with $\beta = 0$, $C = 3$ and $m = 1.5$.

The optimal time weights are $w_1^{CS} = 0.22$, $w_2^{CS} = 0.21$, $w_3^{CS} = 0.22$, $w_4^{CS} = 0.22$ and $w_5^{CS} = 0.13$. Therefore, as for the spatial case, the relevance of the information concerning year 1999 is lower than those of the first four years under examination. More remarkable differences can be found by observing the resulting centroids and membership degrees, even if the two solutions are comparable to some limited extent. In fact, once again, Cluster 3 (light grey coloured) is formed by most of the provinces located in Southern Italy. However, in addition, it also contains several provinces from both Northern and Central Italy. By observing Figure 5, it is easy to see that, similarly to the centroid of Cluster 3 (in Figure 3), such a centroid is characterized by the lowest values for NFR, RNI, ETMP and NMH and by the highest values for RU during all the time occasions if compared with the centroids of the other two clusters. Also the DR values are lower than those pertaining to the other centroids, but such differences are less remarkable as compared to those resulting from Figure 3. Instead, the NS values of the centroids are almost equal and, thus, do not help in distinguishing the obtained clusters.

By inspecting Figures 4 and 6, we can emphasize some links between Clusters 1 (dark grey coloured) determined by applying the clus-



Prov.	Clus.	Memb. degree	Prov.	Clus.	Memb. degree	Prov.	Clus.	Memb. degree
Torino	3	0.89	Modena	2	0.95	Foggia	3	0.99
Vercelli	3	0.97	Bologna	1	0.89	Bari	1	0.92
Novara	1	0.99	Ferrara	3	0.95	Taranto	3	0.98
Cuneo	1	1.00	Ravenna	1	0.68	Brindisi	3	0.73
Asti	1	0.68	Forli	1	0.97	Lecce	3	0.66
Alessandria	3	0.89	Pesaro-Urbino	1	0.78	Potenza	3	0.64
Aosta	1	0.88	Ancona	1	1.00	Matera	3	1.00
Imperia	3	0.68	Macerata	1	0.90	Cosenza	3	0.96
Savona	3	0.95	Ascoli Piceno	1	1.00	Catanzaro	3	0.89
Genova	3	0.88	Massa-Carrara	3	0.81	Reggio Calabria	3	0.95
La Spezia	3	0.97	Lucca	3	0.73	Trapani	1	0.63
Varese	1	0.99	Pistoia	1	0.99	Palermo	3	0.69
Como	1	0.89	Firenze	3	0.58	Messina	3	0.99
Sondrio	1	0.79	Livorno	3	0.99	Agrigento	3	0.98
Milano	1	0.81	Pisa	1	0.89	Caltanissetta	3	0.54
Bergamo	2	1.00	Arezzo	1	0.93	Enna	3	0.91
Brescia	2	0.99	Siena	1	0.95	Catania	(2)	0.46
Pavia	1	0.95	Grosseto	3	0.97	Ragusa	1	0.58
Cremona	1	0.99	Perugia	2	0.75	Siracusa	3	0.99
Mantova	1	0.94	Terni	3	0.97	Sassari	3	0.79
Bolzano	2	0.99	Viterbo	1	0.79	Nuoro	3	0.99
Trento	2	0.99	Rieti	1	0.85	Cagliari	3	0.89
Verona	2	0.62	Roma	1	0.78	Pordenone	1	0.90
Vicenza	2	0.99	Latina	2	0.92	Isernia	3	0.99
Belluno	3	0.81	Frosinone	1	0.89	Oristano	3	0.71
Treviso	2	0.99	Caserta	2	0.70	Biella	3	0.93
Venezia	3	0.93	Benevento	3	0.99	Lecco	2	0.98
Padova	2	0.52	Napoli	1	0.60	Lodi	2	0.99
Rovigo	3	0.99	Avellino	3	0.86	Rimini	2	0.95
Udine	3	0.82	Salerno	1	0.91	Prato	2	0.98
Gorizia	1	0.62	L'Aquila	1	0.67	Crotone	3	0.89
Trieste	3	0.77	Teramo	1	0.64	Vibo Valentia	3	0.97
Piacenza	3	0.84	Pescara	1	0.98	Verbano-C.-O.	3	0.89
Parma	1	0.96	Chieti	1	0.95			
Reggio Emilia	2	0.92	Campobasso	3	0.99			

Figure 6. Membership degrees resulting from CS-F*C*M-STT with $\beta = 0$, $C = 3$ and $m = 1.5$.

tering models with or without spatial penalty terms. In fact, 24 provinces are assigned to Cluster 1 in both cases. These provinces are principally located in the Northern side of Italy but a few Central (among them Pescara (membership degree = 0.98), Chieti (0.95) and Rome (0.78)) and Southern provinces (among them Bari (0.92), Salerno (0.91) and Naples (0.60)) are visible. Unfortunately, by comparing the centroids of Cluster 1 in Figures 3 and 5, some conflicting results can be seen. This especially occurs with respect to the NFR and RNI values, which are no longer the highest ones. Moreover, the RU values are not the lowest ones, as occurs for the solution with $\beta = 0.65$. These differences can be explained by observing that a lot of wealthy provinces are now assigned to a different cluster, i.e. Cluster 2. Some of these are Bolzano (0.99) and Trento (0.99) from Trentino Alto Adige, Vicenza (0.99) and Treviso (0.99) from Veneto, Bergamo (1.00), Brescia (0.99), Lodi (0.99) and Lecco (0.98) from Lombardy and Rimini (0.95), Modena (0.95) and Reggio Emilia (0.92) from Emilia Romagna.

The biggest differences between the solutions of the spatial and non spatial models can be found with respect to Cluster 2 (grey coloured). In fact, there does not exist any province belonging to Cluster 2 in both cases $\beta = 0$ and $\beta = 0.65$. We can approximately state that, in the non spatial case, Cluster 2 is composed by most of the provinces located in Trentino Alto Adige, Veneto, Lombardy (those previously mentioned along with many others), Emilia Romagna (only the above-mentioned three provinces) and in regions from Central Italy. The strong presence of provinces located in Northern Italy explains the values of the features under examination, as it can be depicted from Figure 5. Specifically, some of them have an excellent socio-economical profile: the RU values are the lowest ones whereas the ETMP and NMH are the highest ones, thus denoting a high cultural level of life. Moreover, the presence of foreign residents is the highest one as well as the rate of natural increase.

Finally, it should be noticed that in the non spatial case only one province (Catania) is not well assigned to any cluster. Also in the spatial case, Catania did not clearly belong to any cluster.

All in all, we can conclude that the clustering model with spatial penalty term has allowed us to determine a more reasonable partition of the Italian provinces in the sense that the provinces from Southern Italy are almost all assigned to exactly one cluster (Cluster 3). This seems to be consistent with the empirical evidence that the socio-demographical features (and, indeed, those under investigation) of the provinces located in Southern Italy are pretty different from those of the remaining parts of Italy. To a certain extent, in the spatial case, the two clusters (Clusters 1 and 2) composed by provinces from Northern and Central Italy seem to distinguish the highly productive areas (Cluster 1) from the remaining ones

(Cluster 2). On the contrary, the above spatial typology is not clearly recovered when the spatial information is ignored ($\beta = 0$).

5.2. Longitudinal case

The application of the CS-F*C*M-STT model with spatial penalty term showed that three geographically structured clusters can be found. These were constructed according to the levels of the variables during the time occasions under investigation. The clusters were thus obtained considering the T=5 'pictures' of the Italian provinces with respect to the examined socio-demographical variables in years 1995-1999. In this section, once more, we aim at clustering the Italian provinces. However, in this case, the information to be used concerns the variations (of the examined socio-demographical variables) during all the pairs of two consecutive years. Therefore, the spatial units are no longer studied in terms of the observed levels of the variables, but are examined with respect to the observed changes. We thus perform the L-F*C*M-STT model.

In the sequel we limit ourselves to briefly describing the relevant steps of the analysis and summarizing its results. First of all, after various runs of the procedures, we decided to set C=2 and m=2.2. Then, by applying the heuristic procedure proposed in Section 4.2, we obtained the optimal value β=3.70. Given this setting, the main features of the achieved classification are as follows.

1) The most recent time intervals (1996-97 and 1997-98) play a more relevant role in the clustering task.
2) A clear-cut distinction between provinces belonging to Northern and Central Italy (Cluster 1) on one side and provinces of Southern Italy (Cluster 2) on the other side, is obtained.
3) Most provinces present a very high degree of membership in one of the two clusters.
4) The analysis of the velocity centroids shows that in Cluster 1 the dynamics of NFR, RNI and DR is always positive. Instead, in Cluster 2, an increasing pattern is recored for NS and RU.
5) Ignoring the spatial information (by setting $\beta = 0$) and still using C=2 and m=2.2, the results of the clustering procedure do not show a remarkable change as compared to those obtained with β=3.70. However, in the non-spatial case, the partition presents a very high level of fuzziness (contrary to what happens when we embody the spatial information in the procedure).
6) With reference to the comments in point 5, it should be underlined that, in contrast with the instantaneous approach, the longitudinal one looks at the variations between successive time occasions. Of course, the longer the observed time series, the higher is the possibility of finding

differences between velocity trajectories. In the present application, the small number of considered years limits this possibility and, consequently, tends to lower the discriminative power of the trajectories. This is reflected by the low membership degrees provided by the clustering procedure with $\beta = 0$. Although the clustering structure is basically recovered, the associated fuzziness constitutes a sign of the uncertainty due to the relatively small differences between the velocity trajectories of the centroids of the two clusters. The partition coefficient:

$$PC = \frac{\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^2}{I}$$

by Bezdek (1974) is equal to 0.54. As it is well-known, PC takes values from C^{-1} in case of maximal fuzziness ($u_{ic} = C^{-1}, \forall i, c$) to 1 in case of crisp partition. As $C^{-1} = 0.50$, we can conclude that the obtained partition is very fuzzy. The addition of spatial information has, in this case, the effect of dramatically decreasing the fuzziness of the classification while providing a clear-cut geographical partition of the Country. In fact, note that $PC = 0.82$.

6. Concluding Remarks

The classification of univariate or multivariate time series has received much attention in the recent literature (see, e.g., Maharaj 2000). Most of these works are based on the characteristics of the models fitted to the observed time series (autoregressive models, mixture models, spectral analysis). Thus, the obtained clusters are identified by means of the parametric features of the models selected for representing the time patterns of the observational units. In this respect, the approach adopted in the present paper is completely different. In fact, it relies on the observed patterns rather than on their representation through statistical models. A key contribution in this connection is provided by the introduction of appropriate measures of dissimilarity between multivariate time trajectories. In this connection, it could be interesting to compare the performance of our proposal, carried out in a fuzzy context, with those of existing model-based techniques for clustering multivariate time series. This may offer a deeper insight into the strengths and weaknesses of the two approaches. However, the distinguishing feature of the proposed clustering procedure consists of its capability to deal with the spatial nature of the units to be clustered. Indeed, the classification task is accomplished by searching for a reasonable trade-off between cluster homogeneity based on the time pattern of the observed variables and, on

the other side, cluster homogeneity based on spatial contiguity. The utilization of a fuzzy setting for this problem provides more flexibility to its solution, allowing for different degrees of membership in the various classes of the obtained typology. This characteristic also allows us to identify the spatial units which do not fit very well in the above typology (showing almost equal degrees of membership in the different clusters). The objective functions J_1 and J_2, respectively in (7) and (19), express the above mentioned trade-off through a mixture of two component functions: one measuring cluster homogeneity in the feature space (i.e. the space of multivariate trajectories) and the other one measuring spatial homogeneity of the clusters. The mixing parameter β plays a crucial role in determining the balance between these two components. Therefore, it is necessary to find an appropriate value for β. The procedure suggested in this paper is based on establishing a statistical link between the patterns recorded in the feature space (summarized by the average values across the time occasions) and the spatial proximities of the units. This link is expressed by the index of spatial intracluster autocorrelation introduced in (28). This index depends not only on the contiguity of the spatial units (represented by **P**), but also on the degrees of membership and the time weights, as calculated from the solution of problems (7) or (19). Therefore, if we look at the solution which presents the maximum spatial intraclass autocorrelation (i.e. maximum spatial homogeneity in terms of both spatial contiguity and similarity of the time trajectories within the clusters), we have to choose the value of β to which there is associated the solution which maximizes the autocorrelation index.
In synthetic terms we can state that our method aims at maximizing:

[1] intracluster features' homogeneity subject to simultaneously maximizing:
[2] intracluster spatial neighborhood relationship;
[3] intracluster spatial autocorrelation.

Task [3] is accomplished by optimizing the mixing coefficient between the objective functions representing tasks [1] and [2], respectively. The way to solve the latter problem takes into account the charateristics of the time series to be clustered. In this connection, our proposal is specific for multivariate time trajectoeries as observed on spatial units, and differs from Pham (2001) model with respect to the following aspects:

1) Pham's model appears to be suitable mainly in the context of image segmentation;
2) the way to cope with the determination of the mixing coefficient β is integrated within the optimization tasks to be achieved when minimizing the objective function (7) (see task [3] above). Instead, in the Pham's model the problem of the mixing coefficient is formulated in a supplementary way in terms of a cross-validation procedure;

3) our model provides an organic integration between the technical features of fuzzy clustering of multivariate time trajectories and the specific features related to spatial information (neighborhood relation, spatial autocorrelation).

Summing up, the illustrated clustering procedure appears to constitute a coherent exploratory tool for classifying time trajectories referring to spatial units. However, though the spatially penalized Fuzzy C-Means objective function seems to work well in practice and moreover it can be framed within the theory of the FCM models with penalization terms (see, e.g., Coppi and D'Urso 2006), by no means it can be considered the only possible way to account for spatial proximity in the clustering task. In this respect, the extension of the spatial FCM model by Liew et al. (2000, 2003) appears to be a promising perspective of research. Both the Liew et al. (2000, 2003) and Pham (2001) models are generalizations of standard FCM for coping with the image segmentation problem. In both cases, the idea is to consider the spatial information in the clustering algorithm without imposing the clusters to be composed *exclusively* by contiguous pixels. Nonetheless, several differences can be highlighted. For instance, Liew et al. (2000, 2003) propose a new dissimilarity index to be embedded into the FCM objective function. Such a dissimilarity index is constructed by considering every window of nine pixels, that is the center pixel and its eight neighbouring pixels. In case of homogeneous features for the center pixel and its neighbours, the center pixel is smoothed by its neighbouring pixels when computing the membership degrees and the centroids. Otherwise, the influence of the neighbouring pixels on the centre pixel vanishes. It follows that the spatial FCM by Liew et al. (2000, 2003) is locally adaptive (with respect to every pixel) and tends to allow sharp boundaries between pixels. This is a relevant difference with respect to the spatial FCM model by Pham (2001), which can be seen as non-adaptive in the sense that the emphasis of the spatial information depending on β is evaluated with respect to the entire data set. Therefore, the extent to which the spatial information plays a relevant role in the clustering process does not vary among pixels. Among other things, another difference between the two clustering models regards, in Liew et al. (2000, 2003), the replacement of the classical square distance of the FCM as given in (1) by the above-mentioned new dissimilarity index, whereas, in Pham (2001), the spatial information is taken into account by the penalty term. Unfortunately, the extension of the Liew et al. (2000, 2003) model to the case of geographical units and, moreover, to spatial time series seems not to be straightforward. For instance, the spatial units and their neighbours will have different size, shape and border. A crucial point is whether it makes sense to take into account these characteristics in extending the dissimilarity index. It is not obvious how to do that, nor how to handle spatial units that have not neighbours (for instance, islands). The extension of such a model along with the investigation of alternative approaches to this problem will form the objective of our future research in this domain.

References

ALLENDE, H. and GALBIATI, J. (2004), "A Non-Parametric Filter for Digital Image Restoration, Using Cluster Analysis", *Pattern Recognition Letters*, *25*, 841–847.

AMBROISE, C. and GOVAERT, G. (1998), "Convergence of an EM-Type Algorithm for Spatial Clustering", *Pattern Recognition Letters*, *19*, 919–927.

AYALA, G., EPIFANIO, I., SIMÓ, A., and ZAPATER, V. (2006), "Clustering of Spatial Point Patterns", *Computational Statistics and Data Analysis*, *50*, 1016–1032.

BEZDEK, J.C. (1974), "Cluster Validity with Fuzzy Sets", *Journal of Cybernetics*, *3*, 58–72.

BEZDEK, J.C. (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press.

BEZDEK, J.C., KELLER, J., KRISNAPURAM, R., and PAL, N.R. (1999), *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, The Handbooks of Fuzzy Sets, 4, New York: Kluwer.

CHUANG, K.-S., TZENG, H.-L., CHEN, S., WU, J., and CHEN, T.J. (2006), "Fuzzy *C*-means Clustering with Spatial Information for Image Segmentation", *Computerized Medical Imaging and Graphics*, *30*, 9–15.

CINQUE, L., FORESTI, G., and LOMBARDI, L. (2004), "A Clustering Fuzzy Approach for Image Segmentation", *Pattern Recognition*, *37*, 1797–1807.

COPPI, R. and D'URSO, P. (2001), "The Geometric Approach to the Comparison of Multivariate Time Trajectories", in *Advances in Data Science and Classification*, eds. S. Borra, R. Rocci, M. Vichi, and M. Schader, Heidelberg: Springer-Verlag, 93–100.

COPPI, R. and D'URSO, P. (2006), "Fuzzy Unsupervised Classification of Multivariate Time Trajectories with the Shannon Entropy Regularization", *Computational Statistics and Data Analysis*, *50*, 1452–1477.

COSTANZO, G.D. (2001), "A Constrained *k*-means Clustering Algorithm for Classifying Spatial Units", *Statistical Methods and Applications*, *10*, 237–256.

DI NOLA, A., LOIA, V., and STAIANO, A. (2002), "An Evolutionary Approach to Spatial Fuzzy *C*-means Clustering", *Fuzzy Optimization and Decision Making*, *1*, 195–219.

DUAN, L., XU, L., GUO, F., LEE, J., and YAN, B. (2007), "A Local-Density Based Spatial Clustering Algorithm with Noise", *Information Systems*, *32*, 978–986.

D'URSO, P. (2000), "Dissimilarity Measures for Time Trajectories", *Journal of the Italian Statistical Society*, *1–3*, 1–31.

D'URSO, P. (2004), "Fuzzy *C*-means Clustering Models for Multivariate Time-Varying Data: Different Approaches", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *12*, 287–326.

D'URSO, P. (2005), "Fuzzy Clustering for Data Time Array with Inlier and Outlier Time Trajectories", *IEEE Transactions on Fuzzy Systems*, *13*, 583–604.

EVERITT, B.S., LANDAU, S. ,and LEESE, M. (2001), *Cluster Analysis* (4th ed.*),* London: Arnold Press.

FERLIGOJ A. and BATAGELJ, V. (1982), "Clustering with Relational Constraint", *Psychometrika*, *47*, 413–426.

FERLIGOJ A. and BATAGELJ, V. (1983), "Some Types of Clustering with Relational Constraint", *Psychometrika*, *48*, 541–552.

FERLIGOJ A. and BATAGELJ, V. (1992), "Direct Multicriteria Clustering Algorithm", *Journal of Classification*, *9*, 43–61.

GORDON, A.D. (1996), "A Survey of Constrained Classification", *Computational Statistics and Data Analysis*, *21*, 17–29.

GORDON, A.D. (1999), *Classification,* New York: Chapman & Hall/CRC.

HEISER, W.J. and GROENEN, P.J.F. (1997), "Cluster Differences Scaling with a Within-Clusters Loss Component and a Fuzzy Successive Approximation Strategy to Avoid Local Minima", *Psychometrika*, *62*, 63–83.

HU, T. and SUNG, S.Y. (2006), "A Hybrid EM Approach to Spatial Clustering", *Computational Statistics and Data Analysis*, *50*, 1188–1205.

HWANG, H., DE SARBO, W.S., and TAKANE Y. (2007), "Fuzzy Clusterwise Generalized Structured Component Analysis", *Psychometrika*, *72*, 181–198.

KONTOS, D. and MEGALOOIKONOMOU, V. (2005), "Fast and Effective Characterization for Classification and Similarity Searches of 2D and3D Spatial Region Data", *Pattern Recognition*, *38*, 1831–1846.

KROOSHOF, P.W.T., TRAN, T.N., POSTMA, G.J., MELSSEN, W.J., and BUYDENS, L.M.C. (2006) ,"Effects of Including Spatial Information in Clustering Multivariate Image Cata", *Trends in Analytical Chemistry*, *25*, 1067–1080.

LAWSON, A.B., SIMEON, S., KULLDORFF, M., BIGGERI, A., and MAGNANI, C. (2007), "Line and Point Cluster Models for Spatial Health Data", *Computational Statistics and Data Analysis*, *51*, 6027–6043.

LEFKOVITCH, L.P. (1980), "Conditional Clustering", *Biometrics*, *36*, 43–58.

LIEW, A.W.C., LEUNG, S.H., and LAU, W.H. (2000), "Fuzzy Image Clustering Incorporating Spatial Continuity", *IEE Proceedings of Visual Image Signal Process*, *147*, 185–192.

LIEW, A.W.C., LEUNG, S.H., and LAU, W.H. (2003), "Segmentation of Color Lip Images by Spatial Fuzzy Clustering", *IEEE Transactions on Fuzzy Systems*, *11*, 542–549.

LIEW, A.W.C. and YAN, H. (2003), "An Adaptive Spatial Fuzzy Clustering Algorithm for 3-D MR image Segmentation", *IEEE Transactions on Medical Imaging*, *22*, 1063–1075.

MACQUEEN, J.B. (1967), "Some Methods for Classifilcation and Analysis of Multivariate Observations", in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *2*, 281–297.

MARÇAL, A.R.S. and CASTRO, L. (2005) ,"Hierarchical Clustering of Multispectral Images Using Combined Spectral and Spatial Criteria", *IEEE Geoscience and Remote Sensing Letters*, *2*, 59–63.

MAHARAJ, E.A. (2000), "Clusters of Time Series", *Journal of Classification*, *17*, 298–314.

MCBRATNEY, A.B. and MOORE, A.W. (1985), "Application of Fuzzy Sets to Climatic Classification", *Agricultural and Forest Meteorology*, *35*, 165–185.

MOLENAAR, M. and Cheng, T. (2000), "Fuzzy Spatial Objects and Their Dynamics", *ISPRS Journal of Photogrammetry and Remote Sensing*, *55*, 164–175.

MORAN, P.A.P. (1950), "A Test for the Serial Independence of Residuals", *Biometrika 37*, 178–181.

MURTAGH, F. (1985), "A Survey of Algorithms for Contiguity-Constrained Clustering and Related Problems", *Computer Journal*, *28*, 82–88.

NG, R.T. and HAN, J. (2002), "CLARANS: A Method for Clustering Objects for Spatial Data Mining", *IEEE Transactions on Knowledge and Data Engineering*, *14*, 1003–1016.

PERMUTER, H., FRANCOS, J., and JERMYN, I. (2006), "A Study of Gaussian Mixture Models of Color and Texture Features for Image Classification and Segmentation", *Pattern Recognition*, *39*, 695–166.

PHAM, D.L. (2001), "Spatial Models for Fuzzy Clustering", *Computer Vision and Image Understandin, 84*, 285–297.

PHAM, D.L. and PRINCE, J.L. (1999), "Adaptive Fuzzy Segmentation of Magnetic Resonance Images", *IEEE Transactions on Medical Imaging 18*, 737–752.
PILEVAR, A.H. and SUKUMAR, M. (2005), "GCHL: A Grid-Clustering Algorithm for High-dimensional Very Large Spatial Data Bases", *Pattern Recognition Letters*, *26*, 999–1010.
SMOUSE, P.E. and PEAKALL, R. (1999), "Spatial Autocorrelation Analysis of Individual Multiallele and Multilocus Genetic Structure", *Heredity*, *82*, 561–573.
TOLIAS, Y.A. and PANAS, S.M. (1998), "On Applying Spatial Constraints in Fuzzy Image Clustering Using a Fuzzy Rule-based System", *IEEE Signal Processing Letters*, *5*, 245–247.
TOLIAS, Y.A. and PANAS, S.M. (1998), "Image Segmentation by a Fuzzy Clustering Algorithm Using Adaptive Spatially Constrained Membership Functions", *IEEE Transactions on Systems, Man, and Cybernetics A*, *28*, 359–369.
TRAN, T.N., WEHRENS, R., and BUYDENS, M.C. (2005), "Clustering Multispectral Images: A Tutorial", *Chemometrics and Intelligent Laboratory Systems*, *77*, 3–17.
XIA, Y., FENG, D., WANG, T., ZHAO, R,. and ZHANG, Y. (2007), "Image Segmentation by Clustering of Spatial Patterns", *Pattern Recognition Letters*, *28*, 1548–1555.