Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Spatial robust fuzzy clustering of COVID 19 time series based on B-splines

Pierpaolo D’Urso a,*, Livia De Giovanni b, Vincenzina Vitale a

a Department of Social and Economic Sciences, Sapienza University of Rome, P.za Aldo Moro, 5 00185 Rome, Italy
b Department of Political Sciences Politiche, LUISS university, Viale Romania, 32 00197 Rome, Italy

A R T I C L E   I N F O

Article history:
Received 1 March 2021
Received in revised form 24 April 2021
Accepted 4 May 2021
Available online xxxx

Keywords:
Fuzzy Partitioning Around Medoids
Robust clustering
Time series clustering
Spatial clustering
COVID-19

A B S T R A C T

The aim of the work is to identify a clustering structure for the 20 Italian regions according to the main variables related to COVID-19 pandemic. Data are observed over time, spanning from the last week of February 2020 to the first week of February 2021. Dealing with geographical units observed at several time occasions, the proposed fuzzy clustering model embedded both space and time information. Properly, an Exponential distance-based Fuzzy Partitioning Around Medoids algorithm with spatial penalty term has been proposed to classify the spline representation of the time trajectories. The results show that the heterogeneity among regions along with the spatial contiguity is essential to understand the spread of the pandemic and to design effective policies to mitigate the effects.

© 2021 Published by Elsevier B.V.

1. Introduction

Severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2) is the name given to the new 2019 coronavirus while COVID-19 is that given to the disease associated with the virus. This new strain of coronavirus, not previously identified in humans, has exponentially spread out from the capital of the Chinese province of Hubei, Wuhan, the epicentre of the contagion, worldwide causing a pandemic with serious effects on national health systems (Li et al., 2020).

* Corresponding author.
E-mail addresses: pierpaolo.durso@uniroma1.it (P. D’Urso), ldegiovanni@luiss.it (L. De Giovanni), vincenzina.vitale@uniroma1.it (V. Vitale).

https://doi.org/10.1016/j.spasta.2021.100518
2211-6753/© 2021 Published by Elsevier B.V.

Please cite this article as: P. D’Urso, L. De Giovann and V. Vitale, Spatial robust fuzzy clustering of COVID 19 time series based on B-splines. Spatial Statistics (2021) 100518, https://doi.org/10.1016/j.spasta.2021.100518.
In Italy the outbreak of COVID-19 started at the end of February in Lombardy and then hit Northern Italy. As a consequence the Italian government imposed a nationwide lockdown (on 9 March 2020) in order to mitigate the disease and avoid further spreading of the virus. Flows of people travelled from North to South when the national lockdown measures were first announced on 8 March 2020.

The lockdown was gradually suspended at the beginning of May 2020 as the number of new daily active cases and the number of deaths due to COVID-19 steadily declined, also considering its economic cost. People could gather provided they maintained social distancing. Nonetheless, it was not possible to move freely between regions until 3 June 2020. The count of the new daily infections remained stable until August 2020, when the number of new daily cases, on a national level, increased to over 1200.

In October, with an average of 10000 new daily cases and the exponential growth of the incidence rate, Italy faced its second pandemic wave involving all territories, from the North to the South. It is worth noting that the higher number of new daily infections in the second wave was also strictly related to the considerably higher number of tested people by day.

New restrictive measures were introduced by the Italian authorities that varied based on the classification of Italian provinces into three areas – red, orange and yellow – corresponding to three risk profiles on the basis of the value of 21 indicators. The 21 indicators are divided into three different categories: process indicators on monitoring capacity; process indicators on the ability to diagnose, investigate and manage contacts; result indicators relating to transmission stability and the resilience of health services. The red colour corresponds to higher risk. They contributed to control the pandemic spread in Italy and are still adopted to this day.

On 2 December 2020, Health Minister presented the guidelines of the Strategic Plan for anti-SARS-CoV-2/COVID-19 vaccination (Decree 2 January 2021) to Parliament, produced by the Ministry of Health, Extraordinary Commissioner for the Emergency, Higher Institute of Health, Agenas and Aifa. As foreseen by the Plan itself, on 8 February 2021, the document was published that updates the categories and the order of priority for the second phase of the vaccination campaign against Covid-19 based on the evolution of knowledge and information on vaccines.

The official data for the COVID-19 pandemic are collected, daily, by the Italian Civil Protection Department. The GitHub repository\(^1\) contains multiple counts, such as the new incident cases, the total cases, the currently positives, the number of hospitalised patients, the number of patients in Intensive Care Units, the deaths. They are available at regional level and, in such cases, at provincial level.

Nonetheless the new starting vaccination campaign, on 27 January 2021, the Italian Civil Protection Department recorded 2487036 total cases and 85418 deaths due to COVID-19.

We argue that the lack of uniformity in reporting and signalling the COVID-19 new infections and deaths, the latter due above all to the presence of comorbidities, affects the reliability of the estimates of the incidence and mortality/fatality rates; in particular for the latter, the under-estimation is also related to the certification of the “COVID-19” victims that could be done only among swabbed individuals.

According to the last bulletin provided by the Italian National Institute of Health (ISS) (Iss, 2021), the mean age of died patients positive to SARS-CoV-2 infection was 81 years, with women having an older age than men (median age 86 years for women versus median age 80 for men). Moreover, it is worth noting that the median age of patients dying for COVID-19 was 83 years while that of infected people was 48 years.

Taking into account gender, out of the 85418 patients died for SARS-CoV-2, 43.7% was female. Based on an opportunistic sample of only 6381 patients died in hospital, the average number of diseases was 3.6 and the 84.9% of the sample had 2 or more comorbidities.

In this study, we focused on the daily time-series of the cumulative cases over population (per 10,000 inhabitants), of the cumulative cases over monitored cases and of the cumulative deaths over population (per 10,000 inhabitants), spanning from 2020-02-24 to 2021-02-08.

\(^{1}\) http://www.protezionecivile.gov.it/attivita-rischi/rischio-sanitario/emergenze/coronavirus.
From a methodological point of view, one deals with spatial units observed at several time occasions. The corresponding spatial time data array $X$ is a two-way data array, therefore an array of the type: spatial objects $\times$ occasions.

Several space–time series clustering techniques have been proposed in the literature: some of them take into account the spatial dependence in the non-spatial time series clustering models by defining a suitable spatial dissimilarity measure (Izakian et al., 2013), while others are density-based (Ester et al., 1996; Wang et al., 2006; Birant and Kut, 2007; Ienco and Bordogna, 2016; Xie et al., 2016) or model-based (Basford and McLachlan, 1985; Viroli, 2011; Torabi, 2014, 2016; Disegna et al., 2017) clustering techniques.

The approach followed in this work belongs to the broad group of the spatially constrained time series clustering techniques (Hu and Sung, 2006; Coppi et al., 2010; Gao and Yu, 2016). More specifically, it is based on the fuzzy clustering algorithm for spatial-time data proposed by D’Urso et al. (2019) that, in turn, is an extension of the fuzzy C-Means algorithm for time series with spatial information developed in Coppi et al. (2010). Both fuzzy algorithms proposed in Coppi et al. (2010) and D’Urso et al. (2019) include a spatial penalization term in the objective function. The role of this term and of the related tuning parameter is to smooth the membership degrees of all units contiguous to the generic $i$th unit in all clusters to which $i$th unit does not belong.

The timeseries are transformed onto (finite dimensional) vectors of coefficients for dimension reduction purposes. This is obtained by projecting each time series onto a finite dimensional functional basis. Afterwards, robust clustering is applied to the resulting basis coefficients. Projection onto a suitable functional basis is considered, in particular the cubic B-spline basis.

The need for robust clustering methods is naturally justified by the possible and frequent presence of outlier time series. Notice that outliers are likely to occur in large, and sometimes fully automatized, data sets. Part of the variability of data can be removed by the regularization or smoothing procedure considered, but other sources of variability could be due to particularities of the data themselves. To achieve robustness, in the paper the Exponential distance is considered (Wu and Yang, 2002; D’Urso and De Giovanni, 2014).

In this study, the Exponential distance-based Fuzzy C-Medoids clustering algorithm based on B-splines with spatial penalty term (BS-Exp-FCMd-S) is applied to clustering of time series related to COVID-19 pandemic.

Summing up the proposed clustering method used for analysing the Covid time series presents different benefits inherited by its structural features. In particular:

**Clustering procedure:** adopting the PAM (Partitioning Around Medoids) approach, the cluster prototypes (i.e., medoids) are units actually observed and not “virtual” units like the “centroids” derived with a fuzzy c-means (Bezdek, 1981). Overall, having non-fictitious representative units available makes interpreting the obtained clusters easier (Kaufman and Rousseeuw, 2005a). In addition, PAM procedure provides a “timid robustification” of the c-means clustering (García-Scudero and Gordaliza, 1999; García-Scudero et al., 2010).

**Robustness:** we consider in the clustering process a suitable transformation (exponential transformation) of the distance measure (exponential distance) capable to smooth the effect of the outliers and to tune their influence properly. In particular, the exponential distance gives less importance to outliers and more importance to those data points laying close to the bulk of the data set (for more details see Wu and Yang, 2002; D’Urso et al., 2015).

**Fuzziness:** fuzzy clustering appears more attractive than the traditional clustering methods when it is difficult to identify a clear boundary among clusters (McBratney and Moore, 1985; Wedel and Kamakura, 2000). In addition, the memberships indicate whether there is a second-best cluster almost as good as the best cluster, a scenario which standard clustering methods cannot uncover (Everitt et al., 2011). Furthermore, fuzzy clustering is attractive because it is easily compatible with distribution free methods (Hwang et al., 2007) and it is computationally efficient (McBratney and Moore, 1985; Heiser and Groenen, 1997). For more details, see D’Urso (2015).
**Temporal information**: the time series are “re-parametrized” by means of the B-splines; in this manner, the temporal information of the time series is preserved considering a reduced number of values (the estimates of the splines coefficients) instead of the numerous actual observations of the time series. In this way we are able to preserve time information in the clustering process in a more parsimonious way from a computational point of view. Furthermore, our method inherits all the theoretical benefits of the B-splines in the time series clustering process (de Boor, 2001).

**Spatial information**: our clustering method is capable to take into account the spatial information by means of a spatial penalty term defined on the basis of the following assumption: “…when a spatial unit belongs to a cluster with a high membership degree, then the penalty term forces the neighbouring spatial units to have high membership degrees in the cluster, as much as possible. In other words, it is expected that a spatial unit with high (low) membership degree in a cluster, will have neighbouring areas with low (high) membership degrees in the remaining clusters. It follows that the spatial penalty term attempts to determine a spatially smoothed membership degrees under the empirical evidence that neighbouring spatial units are often characterized by approximately similar features. Nonetheless, it may also occur that neighbouring spatial units are described by pretty diverse profiles. In this respect, there is a parameter which plays the role of increasing or decreasing the emphasis of the spatial penalty term in the clustering process” (Coppi et al., 2010). For this purpose, the parameter is chosen on the basis of an appropriate spatial autocorrelation measure.

The paper is structured as follows. Section 2 describes data formalization and processing, Section 3 the proposed BS-Exp-FCMd-S clustering model while Section 4 the application of the model to COVID-19 data. The last section addresses some conclusions.

### 2. Data formalization and B-spline modelling

A two-way data array of type “same objects \( \times \) same quantitative variable \( \times \) occasions”, in which the objects are spatial units (geographical areas, pixels, etc.) and the occasions are times, is called spatial-time data array.

In this study, the Italian regions are analysed with respect to variables related to COVID-19 pandemic, collected daily over one year. Therefore, the corresponding objects of the spatial-time data array are represented by the Italian regions, the corresponding variable by a COVID-19 variable and the occasions by the days under consideration.

In a formal way, a spatial-time data matrix can be algebraically formalized as (D’Urso, 2000, 2004, 2005):

\[
X \equiv \{ x_i(t) : i = 1, \ldots, I; t = 1, \ldots, T \}
\]  \hspace{1cm} (1)

where \( i \) indicates the generic spatial unit and \( t \) the generic time; \( x_i(t) \) is the value of the variable observed for the \( i \)th unit at time \( t \). The time data matrix \( X \) can be conveniently represented as the set of vectors:

\[
x_i \equiv \{ x_i(t) : t = 1, \ldots, T \} \quad i = 1, \ldots, I
\]  \hspace{1cm} (2)

The spatial proximity between the \( I \) objects is described by means of the contiguity matrix \( P_{I \times I} \). It is a symmetric matrix with zero diagonal elements and with off-diagonal elements given by \( (i, i') = 1, \ldots, I, \ i \neq i' \):

\[
p_{ii'} = \begin{cases} 
1 & \text{if object } i \text{ is contiguous to object } i', \\
0 & \text{otherwise.}
\end{cases}
\]

Contiguity, on its turn, can be specified in several ways, for instance, two spatial units can be contiguous: if they are adjacent (neighbours); if they belong to the same macro-area, even if they are not adjacent. In both cases, a binary index \( 0 – 1 \) can be created where 1 is assigned to contiguous spatial units, 0 otherwise. Different definitions of connectivity and contiguity can be embedded in the clustering procedure.
2.1. B-splines

A time series \(\{(t, x_i(t))\}\) is the result of collecting a variable \(X\) on unit \(i\) at the \(T\) times \(\{t = 1, \ldots, T\}\). Given a \(p\)-dimensional functional basis \(\{B_s(\cdot)\}_{s=1}^{p}\), we can model that time series by a simple linear least-squares fit as:

\[
x_i(t) = \sum_{s=1}^{p} b_s^i B_s(t) + \epsilon_i, \quad t = 1, \ldots, T
\]

Thus, for \(I\) time series \(x_i, i = 1, \ldots, I\), we will obtain \(I\) vectors of fitted coefficients \(b_i = (b_1^i, \ldots, b_s^i, \ldots, b_p^i)'\), \(i = 1, \ldots, I\). Notice that these \(b_s^i\) parameters can be determined by using ordinary least squares regression. We consider the cubic B-spline bases. Notice that, considering \(S\) interior knots \(\xi_1, \xi_2, \ldots, \xi_S\), a cubic B-spline basis will be made up of \(p = S + 4\) basis elements. To fit a cubic spline to a data set with \(S\) knots we perform least square regression with an intercept and \(S + 3\) predictors, of the form \(X(t), X^2(t), X^3(t)\) and \(S\) truncated cubic functions (one per knot).

Other functional bases may be used, Fourier basis when the data curves exhibit periodical nature, wavelets basis that allows multiresolution analysis (in time and frequency scale). Cubic B-spline bases have good mathematical properties and are easy to implement (de Boor, 2001). Notice also that cubic B-spline bases have an attractive property in that they are locally sensitive to data (each parameter is nonzero over a span of at most five distinct nodes).

One option for the choice of the number of knots is to try out different numbers of knots and see which produces the best looking curve. A more objective approach is to assess the accuracy of the model using the Residual Sum of Squares (RSS) between observed values and predicted values. A portion of the data is removed, a spline with a certain number of knots is fitted to the remaining data, and then the splines are used to make predictions for the held-out portion. The process is repeated multiple times until each observation has been left out once, and then RSS is computed. The procedure is repeated for different number of knots and the value corresponding to the smallest RSS is chosen.

2.2. Exponential distance

Let \(b_i\) and \(b_{i'}\) be the vectors of fitted spline coefficients of time series \(i\) and \(i'\). In order to deal with outlier time series the complementary exponential distance is introduced. The exponential distance between two vectors of coefficients \(b_i, b_{i'}\) is defined as (Wu and Yang, 2002; Zhang and Chen, 2004):

\[
d_{\text{exp}}(b_i, b_{i'}) = \left[1 - \exp(-\beta \|b_i - b_{i'}\|^2)\right]^\frac{1}{2}
\]

The exponential distance gives a small weight to those noisy points or outliers and a large weight to those compact points in the data set. First, it should be noted that the exponential distance is bounded by 1. Second, as the value of \(\beta\) increases, the distance tends more rapidly to its maximum value.

The parameter \(\beta\) plays an important role in the distance. Hence, if \(\beta\) is too high, in the classification process each time series is a singleton, since it has no neighbours. Fig. 1 shows the effect of increasing values of \(\beta\) on the exponential distance. Following Wu and Yang (2002), \(\beta\) is set as the inverse of the variability of the data:

\[
\beta = \left[\frac{\sum_{i=1}^{I} d^2(\tilde{b}_q, \tilde{b}_i)}{I}\right]^{-1}
\]

where \(\tilde{b}_q : q = \arg\min_{1 \leq i \leq I} \sum_{i=1}^{I} d^2(\tilde{b}_q, \tilde{b}_i)\), i.e. the vector of coefficients closest to all other vectors.

The value of \(\beta\) appropriately tunes the distance according to the variability of the data so that in the presence of low variability moderate distances are heavily magnified.

See Wu and Yang (2002) for further insights on the robustness of the exponential distance.
3. The BS-Exp-FCMd-S clustering model with spatial penalty term

In this study, a suitable version of the well-known fuzzy C-Medoids (FCMd) algorithm is proposed. More specifically, the FCMd with spatial penalty term and squared Exponential distance is adopted, after transforming the time series onto the coefficients of the projection on a B-splines functional basis. As known, the FCMd technique allows to identify non-fictitious representative time series, the so-called “medoids” favouring the interpretation of the final partition (Kaufman and Rousseeuw, 2005b).

Its main advantage is related to a series of computational aspects: it is more efficient since the distance matrix needs to be computed once at the beginning of the iterative process and it is less affected by getting stuck in a local optima or by convergence problems (Everitt et al., 2001; Hwang et al., 2007).

The BS-Exp-FCMd-S clustering model can be defined as follows:

$$
\min \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^m [1 - \exp(-\beta \| b_i - \tilde{b}_c \|^2)] + \frac{\gamma}{2} \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^m \sum_{i'=1}^{I} \sum_{c' \in C_c} p_{ii'} u_{i'c'}^m
$$

s.t. $\sum_{c=1}^{C} u_{ic}^m = 1, \; u_{ic}^m \geq 0$

where $b_i$ and $\tilde{b}_c$ are the vectors of coefficients of the B-spline representation of the $i$th spatial time series and of the $c$th spatial medoid ($c = 1, \ldots, C$) respectively, while $m > 1$ is well-known fuzziness parameter.

As far as the spatial penalty term is concerned:

$$\frac{\gamma}{2} \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^m \sum_{i'=1}^{I} \sum_{c' \in C_c} p_{ii'} u_{i'c'}^m$$

it is worth noting that $\gamma$ has to be seen as the tuning parameter of spatial information; $p_{ii'}$ is the generic element of the “contiguity” matrix $P_{I \times I}$ while $C_c$ is the set of the $C$ clusters but cluster $c$. 

Fig. 1. Effect of the parameter $\beta$ on the exponential distance (3).
The \( u_{ic} \) is the membership degree of the unit \( i \) belonging to the cluster \( c \), that is equal to:

\[
 u_{ic} = \frac{\left[ 1 - \exp(-\beta \|b_i - \overline{b}_c\|^2) \right] + \gamma \sum_{i' = 1}^{l} \sum_{c' \in C_c} p_{i'i'} u_{i'i'}^{m}}{\sum_{c' = 1}^{C} \left[ 1 - \exp(-\beta \|b_i - \overline{b}_{c'}\|^2) \right] + \gamma \sum_{i' = 1}^{l} \sum_{c' \in C_c} p_{i'i'} u_{i'i'}^{m}}^{-1/\gamma} \tag{6}
\]

We argue again that the role of the spatial penalty term is of increasing, for all spatial units contiguous to \( i \) unit, their final membership degrees to the cluster to which \( i \) belongs.

In the next paragraph, we introduce the validity measure used in the application to choose the best partition.

### 3.1. Validity measure

In order to choose the best solution in terms of number of groups, in this study we adopt the Fuzzy Silhouette (FS) index (Campello and Hruschka, 2006), one of the most known cluster internal validity criteria based on the weighted average of the individual silhouettes width, \( \lambda_i \), as follows:

\[
 FS = \frac{\sum_{i=1}^{l}(u_{pi} - u_{qi})^\alpha \cdot \lambda_i}{\sum_{i=1}^{l}(u_{pi} - u_{qi})^\alpha}, \quad \lambda_i = \frac{(l_{pi} - a_{pi})}{\max\{l_{pi}, a_{pi}\}} \tag{7}
\]

where \( a_{pi} \) is the average distance of object \( i \) to all other objects belonging to the same cluster \( p \) \((p = 1, \ldots, C)\) and \( l_{pi} \) is the minimum (over clusters) average distance of the \( i \)th unit to all units belonging to the cluster \( q \) with \( q \neq p \). \((u_{pi} - u_{qi})^\alpha \) is the weight of each \( \lambda_i \), where \( u_{pi} \) and \( u_{qi} \) correspond to the first and the second largest element of the \( i \)th column of the fuzzy partition matrix \( U \), respectively; \( \alpha \geq 0 \) is an optional user defined weighting coefficient. Setting \( \alpha = 0 \), it reduces to the crisp Silhouette measure.

A higher value of \( FS \) means a better assignment of the units to the clusters implying that, simultaneously, the intra-cluster distance is minimized while the inter-cluster distance is maximized.

### 3.2. Spatial autocorrelation

As deeply analyzed in Coppi et al. (2010), the optimal choice of the value of the parameter \( \gamma \) is a very complex issue. It has to be set exogenously by means of a heuristic procedure based on the spatial autocorrelation measure introduced in Coppi et al. (2010), that could be seen as a generalization of the Moran’s index (Moran, 1950a). For a chosen value of \( C \) and \( m \), the algorithm is run for increasing values of \( \gamma \) (chosen in a suitable interval): the optimal \( \gamma \) value is that maximizes the within cluster spatial autocorrelation. Properly, it maximizes the Global Moran overall spatial autocorrelation measure \( \rho_{\text{overall}} \) that, for a given partition, is computed as follows:

\[
 \rho_{\text{overall}} = \frac{\sum_{c=1}^{C} \rho_c s_c}{l} \tag{8}
\]

where \( s_c = \sum_{i=1}^{l} u_{ic} \).

The \( \rho_c \), the spatial autocorrelation measure for the \( c \)th cluster, is computed as:

\[
 \rho_c = \frac{\overline{x}^T \overline{U}_c^\frac{1}{2} \overline{P} \overline{U}_c^\frac{1}{2} \overline{x}}{\overline{x}^T \overline{U}_c^\frac{1}{2} \text{diag}(P) \overline{U}_c^\frac{1}{2} \overline{x}} \tag{9}
\]

where \( \overline{x} \) is the centred “compromise” vector (mean of the \( T \) vectors \( x_i \)); \( \overline{U}_c \) is the square diagonal matrix (of order \( l \)) of the membership degrees of cluster \( c \), and \( P \) is the spatial contiguity matrix. The operator \( \text{diag}(\cdot) \) creates a diagonal matrix whose elements in the main diagonal are the same as those of the square matrix in the argument. If \( P \) is a contiguity matrix with 0/1 values, every diagonal element contains the number of neighbouring units for the associated spatial unit. It is
important to observe that (9) is very similar to the measure proposed by Moran (1950b). The main difference concerns the matrix $U_c$, which tunes the contribution of the neighbours.

The vector $\bar{b}$ (mean of the $p$ vectors $b_i$) was also used to compute (9) with the same results.

As for Moran's index, also for $\rho_{\text{overall}}$, a value of 1 ($-1$) identifies a perfect positive (negative) autocorrelation, while 0 indicates the absence of autocorrelation. Therefore, to higher values of the $\rho_{\text{overall}}$ corresponds a better spatial assignment of the units to the clusters.

### 3.3. The BS-Exp-FCMd clustering model

With the choice $\gamma = 0$ the spatial regularization term is not taken into account. In this case only the temporal information is taken into account in the clustering process. Then, we obtain the BS-Exp-FCMd clustering model formalized as follows:

$$\min \sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic}^m [1 - \exp(-\beta \|b_i - \bar{b}_c\|^2)]$$

s.t. $\sum_{c=1}^{C} u_{ic} = 1, \ u_{ic} \geq 0.$

(10)

The membership degree $u_{ic}$ of the unit $i$ belonging to the cluster $c$, is equal to:

$$u_{ic} = \frac{[1 - \exp(-\beta \|b_i - \bar{b}_c\|^2)]^{-\frac{1}{m-1}}}{\sum_{c=1}^{C} [1 - \exp(-\beta \|b_i - \bar{b}_c\|^2)]^{-\frac{1}{m-1}}}.$$

(11)

### 4. Clustering of Italian regions — COVID-19 data

In this study, we proposed the BS-Exp-FCMd clustering model with spatial penalty term (BS-Exp-FCMd-S) to COVID-19 data collected during the pandemic in order to cluster Italian regions. Specifically, the data refer to variables observed on $I = 20$ Italian regions during $T = 351$ times represented by the days from 2020-02-24 to 2021-02-08. Dealing with Italian regions, we expect that there would be a strong highly positive spatial autocorrelation between units, that motivates the use and application of a space–time series clustering model.

To apply the proposed BS-Exp-FCMd-S clustering model the following procedure has been taken into account (Coppi et al., 2010). The optimal number of clusters has been identified running the model, with $\gamma = 0$ and $m = 1.5$, and choosing the number of groups that maximizes the Fuzzy Silhouette index described in Section 3.1. The optimal value of the tuning parameter $\gamma$ is chosen according to the procedure described in Section 3.2; in detail, we ran the algorithm, with fixed number of groups $C$ selected by the Fuzzy Silhouette index and $m = 1.5$, for all values of $\gamma$ in the range $[0, 2]$ with increasing steps equal to 0.01. To assign each region to a specific cluster we have set cut-off values. In two (three) cluster partitions the $i$th time series is assigned to the $c$th cluster if its fuzzy membership degree was $u_{ic} \geq 0.7$ ($\geq 0.6$). Notice that the chosen cut-off point for the membership degrees is compatible with the indications suggested in literature; see, e.g., D'Urso and Maharaj (2009), Maharaj et al. (2010), Maharaj and D’Urso (2011), Dembéle and Kastner (2003) and Lafuente-Rego et al. (2018). Time series membership degrees in the interval $[0.3, 0.7]$ $([0.3, 0.6])$ define fuzzy time series. One notices that, in the contiguity matrix, Calabria and Sicily have been considered contiguous since they are separated only by the Strait of Messina and the coasts are very close, with very frequent ferry connections. Moreover, the number of knots has been set to 50 according to the number of weeks in the reference period of time.

In the following sections, the clustering results will be described in detail with reference to the total cases over population (per 10 000 inhabitants), the total deaths over population (per 10 000 inhabitants).
and the total cases over swabs, respectively, spanning from 2020-02-24 to 2021-02-08. One notices that total cases include active cases (infected patients) and closed cases (deaths and recovery/discharged). As far as the cumulative cases over monitored cases (over swabs) are concerned, we argue that the quality of clustering results is strictly related to the strategies of contact tracing and to the number of available daily swabs that are variable among regions and over time making this kind of data much less reliable.

4.1. Total cases over population

The total number of cases over population (per 10,000 inhabitants), by region, is shown in Fig. 2 while the daily number of new infections over population (per 10,000 inhabitants), by region, is shown in Fig. 3, spanning from 2020-02-24 to 2021-02-08. Both graphs well highlight regional heterogeneity and, looking at the second graph, in particular, we can clearly distinguish two peaks in correspondence of the first and second wave; the latter have soon surpassed the incidence rate of the former essentially because contact tracing has been more accurate and faster, carrying out more swab tests daily.

The BS-Exp-FCMd-S clustering model with spatial penalty term has been applied with reference to the total number of cases over population (per 10,000 inhabitants). The Fuzzy Silhouette index selected three groups showing a value of 0.52. Fig. 4 shows the values of the spatial correlation index (9) over $\gamma$. The optimum value of $\gamma$ is 0.12 related to a value of Global Moran index $\rho_{overall} = 0.65$. The membership degrees for the partition without and with penalty term are shown in Table 1; the latter is also reported in the ternary plot of Fig. 5. The corresponding partition is plotted in both Figs. 6 and 7 to emphasize the clustering results in time and space, respectively.

Focusing on the partition with spatial penalty term, three clusters are clearly identified. The first cluster, with medoid Piedmont, also includes Aosta Valley, Lombardy, Trentino–South Tyrol, Veneto, Friuli–Venezia Giulia, Liguria and Emilia–Romagna, all the Northern regions characterized by a higher number of total infections over time that firstly faced the second wave too. The second cluster, with medoid Lazio, also collects Tuscany, Umbria, Marche, Abruzzo, Molise, Campania and
Apulia; they are central-southern regions for which the number of total infections over time was, however, lower and that faced the second wave later than the first group. The third cluster, with medoid Calabria, also includes Sardinia with a high membership degree, Sicily and Basilicata, that can be classified as fuzzy being their membership degree less than 0.6. Calabria and Sardinia are characterized by the lowest number of total cases over time that also faced the second wave later with respect to all other Italian regions. It is worth noticing that while Basilicata is clearly a fuzzy unit sharing features belonging to the clusters 2 and 3, the low membership of Sicily is due to an anomalous increase of infections during the second wave, as we can see by looking at Fig. 3 showing the twofold sharp increase of new infections in Sicily, that exhibits amongst the highest values.

With respect to the partition without spatial penalty term, the Molise region moves from the cluster with medoid Basilicata, not contiguous with the former, to the cluster with medoid Lazio, contiguous instead. As far as Sicily is concerned, with the spatial penalty term, its membership in cluster 1 is fuzzier than it was without penalty term (0.376 versus 0.414) and it is also comparable to that of the cluster 3, whose medoid is Calabria, its contiguous region. The memberships of the regions considering the spatial penalty term are improved.

4.2. Total deaths over population

The total deaths over population due to COVID-19 disease, by region, is shown in Fig. 8, spanning from 2020-02-24 to 2021-02-08. The trends are comparable with those referred, in the previous section, to the number of infections. Even if they are cumulative values, the time series give fair evidence of the effects of two pandemic waves, with the Northern regions more hit by the outbreak, especially in the first phase. Sicily deserves attention being the only Southern region with a worrying upward trend of deaths in the last period, similar to that of the Northern areas.

To this purpose, a little focus on the main results presented in the last available official report jointly produced by the Italian National Institute of Statistics (Istat) and the Italian National Institute of Health (ISS) (Istat and Iss, 2020) about total excess mortality during the period January–November 2020, is particularly useful.
In the report, one has been confirmed that the fatality rate for people aged under 50 was about 1% for both males and females while people aged over 80 were the most hit, males more than females, representing the 60% of all COVID-19 deaths.

In general, COVID-19 deaths weighted the 13% out of the total mortality, in the first wave (February–May 2020), to rise to 16% in the second wave (end of September–November 2020).

During the first pandemic wave, from March to May 2020, the number of total deaths was more than 211 000, 50 000 more than the average recorded in the same period with reference to the years 2015–2019. Moreover, 45 000 of these were residents in the Northern regions; in particular, the same areas recorded an increase of deaths equal to +94.5% in March 2020 and to +75.0% in April 2020, if compared to the average value recorded in the same month of the years 2015–2019.
The highest price has been paid by Lombardy (+82%) while the other Northern regions recorded a rate between the 28% and 38%; only Veneto and Friuli-Venezia Giulia recorded lower rates, +13.5% and +4.8%, respectively. Among Central regions, it is worth noting the case of Marche with an increment of +19.5%, considerably higher if compared with the same average increment recorded by the geographic partition to which it belongs (+4.6%).

The number of total deaths, for all regions, became comparable with those of the reference period 2015–2019 only during the summer increasing, once again, in correspondence on the second pandemic wave, beginning at the end of September 2020. The excess mortality, in November 2020, was +61.4% in the North, +39.3% in the Centre, and +34.7% in the South. Many Northern regions
Fig. 7. Total cases over population (per 10000 inhabitants) — partition map.

Fig. 8. Total deaths over population by region (per 10000 inhabitants).
Table 2

| Region              | Model with no spatial penalty ($\gamma = 0$) | Model with spatial penalty ($\gamma = \gamma_{opt}$) |
|---------------------|---------------------------------------------|---------------------------------------------------|
|                     | Piedmont         | Veneto       | Campania | Trentino–South Tyrol | Lazio | Calabria |
| 1 Piedmont          | 1.000            | 0.000        | 0.000    | 0.956               | 0.023 | 0.021    |
| 2 Aosta Valley      | 0.452            | 0.280        | 0.267    | 0.563               | 0.218 | 0.218    |
| 3 Lombardy          | 0.433            | 0.287        | 0.281    | 0.812               | 0.094 | 0.094    |
| 4 Trentino–South Tyrol | 0.496        | 0.480        | 0.024    | **1.000**            | 0.000 | 0.000    |
| 5 Veneto            | 0.000            | **1.000**    | 0.000    | 0.975               | 0.014 | 0.011    |
| 6 Friuli-Venezia Giulia | 0.005        | 0.988        | 0.006    | 0.850               | 0.095 | 0.055    |
| 7 Liguria           | 0.997            | 0.002        | 0.001    | 0.834               | 0.097 | 0.070    |
| 8 Emilia-Romagna    | 0.997            | 0.002        | 0.001    | 0.846               | 0.093 | 0.061    |
| 9 Tuscany           | 0.042            | 0.606        | 0.352    | 0.194               | 0.708 | 0.097    |
| 10 Umbria           | 0.002            | 0.008        | 0.990    | 0.004               | 0.990 | 0.007    |
| 11 Marche           | 0.165            | 0.744        | 0.091    | 0.173               | 0.748 | 0.079    |
| 12 Lazio            | 0.001            | 0.005        | 0.994    | 0.000               | 1.000 | 0.000    |
| 13 Abruzzo          | 0.046            | 0.621        | 0.333    | 0.011               | 0.978 | 0.011    |
| 14 Molise           | 0.001            | 0.003        | 0.997    | 0.000               | 1.000 | 0.000    |
| 15 Campania         | 0.000            | 0.000        | **1.000**| 0.001               | 0.997 | 0.002    |
| 16 Apulia           | 0.001            | 0.002        | 0.998    | 0.001               | 0.998 | 0.001    |
| 17 Basilicata       | 0.000            | 0.000        | 1.000    | 0.042               | 0.744 | 0.215    |
| 18 Calabria         | 0.004            | 0.007        | 0.989    | 0.000               | 1.000 | 0.000    |
| 19 Sicily           | 0.037            | 0.923        | 0.040    | 0.563               | 0.298 | 0.139    |
| 20 Sardinia         | 0.013            | 0.018        | 0.969    | 0.000               | 0.001 | 0.999    |

exceeded the peak recorded in March–April 2020: Aosta Valley (+139.0% versus +71.0% in April 2020), Piedmont (+98.0% versus +77% in April 2020), Veneto (+42.8% versus +30.8% in April 2020) and Friuli-Venezia Giulia (+46.9% versus +21.1% in April 2020).

In counter-trend there were only Lombardy (+66% versus +192% in March 2020 and +118% in April 2020) and Emilia-Romagna (+34.5% versus +69% in March 2020).

Moreover, in the summer and in the second wave, the median age class of deaths rose both for male and females (85–89 years for females versus 80–84 years in the first wave; 80–84 years for males versus 75–79 years in the first wave); moreover, the higher age of deaths was also accompanied by a higher severity of comorbidities in the population of deaths.

In this second outbreak wave, many regions belonging to the Centre and South of Italy experienced, for the first time, an increase of mortality for all causes. Even if the excess mortality by age was similar to that of the Northern regions, their increments were however considerably lower than those recorded by Northern regions in March and April 2020 with reference to the same age classes. The Fuzzy Silhouette index identified three groups showing a value of 0.63 while the optimum value of $\gamma$ is 0.5 related to a value of the Global Moran index $\rho_{overall} = 0.56$. The membership degrees for the partition without and with penalty term are shown in Table 2; the latter is also reported in the ternary plot of Fig. 9. The corresponding partition is plotted in both Figs. 10 and 11 to emphasize the clustering results in time and space, respectively.

Focusing on the partition based on spatial penalty, the first cluster, with medoid Trentino–South Tyrol, includes all Northern regions while the second cluster, with medoid Lazio, all its neighbour Central regions (Tuscany, Marche and Umbria) and all the Southern regions but Calabria, Sicily and Sardinia. It is worth noting that Tuscany, Marche and Basilicata belong to the second cluster with a membership degree considerably lower if compared with the other units of the same cluster. The membership of Marche could be explained taking into account that it was the only region, in the first wave, with an excess of deaths considerably higher if compared to that of its neighbour regions.

The third group is a niche cluster composed only by Calabria, the medoid, and Sardinia, both characterized by the lowest values of mortality rates.

Two fuzzy units have been identified, that are Aosta Valley and Sicily; the former is a clear global outlier, as one can see by inspecting its trend with an abnormal upward step in the last period; the latter is a clear local outlier since its upward trend, in the second pandemic wave, does not match
with those of the other Southern regions. The anomalous trend of Aosta Valley had been already observed in the last November whose excess of total mortality went up to +139.0%, if compared with the average value recorded in the same month of period 2015–2019, as already mentioned.

The three clusters identified three profile, from that which paid the highest price in terms of deaths to that which paid the least price; indeed, the first cluster included all regions with the highest number, and also the highest increase in time, of the mortality due COVID-19 pandemic.
The model with the spatial penalty had a substantially different partition from that for the model without the spatial penalty; all fuzzy units in the latter with the exception of Aosta Valley became non-fuzzy in the former, and they were assigned to the groups of its own geographically nearest units.

Furthermore, not all units belonging to the second group in the model with no spatial penalty, with medoid Veneto, were assigned to the same group in the model with spatial penalty. Indeed, the Veneto region was included in the cluster 1, whose new medoid was the contiguous region Trentino–South Tyrol, together with Friuli-Venezia Giulia, that is contiguous with Veneto. Tuscany, Marche and Abruzzo, instead, were assigned to the group of its contiguous regions (cluster 2 of the model with spatial penalty). Abruzzo, in particular, moved to the cluster with medoid Lazio with a high membership degree.

The added value produced by the robust spatial model is evident when we consider Sicily. Its membership degree (0.923), associated to the first cluster with medoid Veneto, had been considerably reduced so that it became a fuzzy unit in the spatial model. In other words, the model with penalty term and squared exponential distance is able to detect the presence of local outliers of particular interest in spatial statistics.

4.3. Total cases over swabs

In October 2020 Health Minister announced quick swabs in pharmacy as already carried out in some regions and an agreement with general practitioners to carry out the rapid swabs. The total cases over swabs are presented by region in Fig. 12 spanning from 2020-02-24 to 2021-02-08. The regional heterogeneity on the onset of an epidemic outbreak, the occurrence of possible recurrent epidemic waves and the policy about swabs can be analysed in terms of the range of the variable and the time of occurrence of changes.
The Fuzzy Silhouette index selected three groups showing a value of 0.42. The optimum value of the spatial correlation index \( (9) \) over \( \gamma \) is 0.10 related to a value of Global Moran index \( \rho_{\text{overall}} = 0.30 \). The membership degrees for the partition without and with penalty term are shown in Table 3; the latter is also reported in the ternary plot of Fig. 13. The corresponding partition is plotted in both Figs. 14 and 15 to emphasize the clustering results in time and space, respectively.

Focusing on the partition with spatial penalty, three clusters are clearly identified. The first cluster, with medoid Piedmont, collects the Northern regions Aosta Valley, Lombardy, Liguria, Emilia-Romagna. With respect to the other regions, they show higher values of the variable total cases over swabs, that stand in the second wave. The second cluster, with medoid Sardinia, collects the Central and Southern regions of Tuscany, Campania, Apulia and the northern region of Trentino–South Tyrol. The values of the variable total cases over swabs are lower and stand lower with respect to cluster 1. The third cluster, with medoid Friuli-Venezia Giulia, collects Northern, Central and Southern regions: Veneto, Lazio, Basilicata, Calabria, Sicily. The values of the variable total cases over swabs are the lowest. Four regions, Umbria, Marche, Abruzzo and Molise are classified as fuzzy as their membership to the cluster is less than 0.6. They share the characteristics of two clusters. The interpretation is facilitated by considering the variable cumulative swabs, showing the policies of monitored cases and swabs (Fig. 16).

With respect to the partition without spatial penalty term, the membership of Lazio and Sicily to the cluster with medoid the non contiguous region Friuli-Venezia Giulia decreases.

5. Conclusions

In this study, the Exponential distance-based Fuzzy C-Medoids clustering algorithm based on B-splines with spatial penalty term (BS-Exp-FCMd-S) is applied to clustering of time series related to COVID-19 pandemic.

Data reduction is obtained by the use of B-splines, robustness is obtained by the use of the exponential distance while spatial information is taken into account by the use of a penalty term in the objective function.
Table 3
Total cases over swabs by region — 3 clusters memberships.

| Region                | Model with no spatial penalty ($\gamma = 0$) | Model with spatial penalty ($\gamma = \gamma_{opt}$) |
|-----------------------|---------------------------------------------|---------------------------------------------------|
|                       | Piedmont Tuscany Friuli-Venezia Giulia       | Piedmont Sardinia Friuli-Venezia Giulia            |
| 1 Piedmont            | 1.000 0.000 0.000                            | 1.000 0.000 0.000                                 |
| 2 Aosta Valley        | 0.836 0.086 0.078                            | 0.845 0.080 0.075                                 |
| 3 Lombardy            | 0.905 0.048 0.047                            | 0.890 0.055 0.055                                 |
| 4 Trentino–South Tyrol| 0.046 0.773 0.181                            | 0.065 0.693 0.242                                 |
| 5 Veneto              | 0.006 0.107 0.887                            | 0.010 0.251 0.739                                 |
| 6 Friuli-Venezia Giulia| 0.000 0.000 1.000                          | 0.000 0.000 1.000                                 |
| 7 Liguria             | 0.877 0.065 0.058                            | 0.874 0.065 0.061                                 |
| 8 Emilia-Romagna      | 0.572 0.237 0.191                            | 0.600 0.206 0.194                                 |
| 9 Tuscany             | 0.000 1.000 0.000                            | 0.012 0.713 0.275                                 |
| 10 Umbria             | 0.028 0.512 0.460                            | 0.037 0.424 0.539                                 |
| 11 Marche             | 0.486 0.263 0.251                            | 0.469 0.268 0.263                                 |
| 12 Lazio              | 0.005 0.085 0.911                            | 0.009 0.285 0.706                                 |
| 13 Abruzzo            | 0.016 0.778 0.206                            | 0.033 0.576 0.392                                 |
| 14 Molise             | 0.153 0.389 0.458                            | 0.158 0.375 0.467                                 |
| 15 Campania           | 0.028 0.784 0.188                            | 0.046 0.640 0.313                                 |
| 16 Apulia             | 0.011 0.604 0.324                            | 0.010 0.740 0.250                                 |
| 17 Basilicata         | 0.002 0.040 0.957                            | 0.004 0.199 0.797                                 |
| 18 Calabria           | 0.009 0.081 0.910                            | 0.009 0.141 0.850                                 |
| 19 Sicily             | 0.006 0.167 0.827                            | 0.005 0.356 0.640                                 |
| 20 Sardinia           | 0.003 0.658 0.338                            | 0.000 1.000 0.000                                 |

Fig. 13. Total cases over swabs — ternary plot.

We obtain on the entire period almost the same partition into yellow, orange and red area obtained on the basis of the 21 indicators considered by the Ministry of Health.

The results show that the heterogeneity among regions along with the spatial contiguity is essential to understand the spread of the pandemic and to design effective policies to mitigate the effects. In particular the clustering of the regions taking into account the spatial proximity along with socio-demographic profiling of the clusters might lead to similar data driven policies of prevention for regions sharing the same cluster.
Fig. 14. Total cases over swabs — partition of the daily time-series.

Fig. 15. Total cases over swabs — partition map.
Fig. 16. Cumulative swabs.

References

Basford, K., McLachlan, G., 1985. The mixture method of clustering applied to three-way data. J. Classification 2, 109–125.
Bezdek, J., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic, Norwell, MA, USA.
Birant, D., Kut, A., 2007. ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data Knowl. Eng. 60, 208–221.
de Boor, C., 2001. A Practical Guide to Splines. Springer.
Campello, R., Hruschka, E., 2006. A fuzzy extension of the silhouette width criterion for cluster analysis. Fuzzy Sets and Systems 157, 2858–2875.
Coppi, R., D'Urso, P., Giordani, P., 2010. A fuzzy clustering model for multivariate spatial time series. J. Classification 27 (1), 54–88.
Dembélé, D., Kastner, P., 2003. Fuzzy C-means method for clustering microarray data. Bioinformatics 19, 973–980.
Disegna, M., D'Urso, P., Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. Spat. Statist. 21, 209–225.
D'Urso, P., 2000. Dissimilarity measures for time trajectories. Stat. Methods Appl. 9 (1–3), 53–83.
D'Urso, P., 2004. Fuzzy C-means clustering models for multivariate time-varying data: Different approaches. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 12 (3), 287–326.
D'Urso, P., 2005. Fuzzy clustering for data time arrays with inlier and outlier time trajectories. IEEE Trans. Fuzzy Syst. 13 (5), 583–604.
D'Urso, P., 2015. Fuzzy clustering. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.), Handbook of Cluster Analysis. Chapman and Hall, pp. 545–573.
D'Urso, P., De Giovanni, L., 2014. Robust clustering of imprecise data. Chemometr. Intell. Lab. Syst. 136, 58–80.
D'Urso, P., De Giovanni, L., Disegna, M., Massari, R., 2019. Fuzzy clustering with spatial-temporal information. Spat. Statist. 30, 71–102.
D'Urso, P., De Giovanni, L., Massari, R., 2015. Time series clustering by a robust autoregressive metric with application to air pollution. Chemometr. Intell. Lab. Syst. 141, 107–124.
D'Urso, P., Maharaj, E.A., 2009. Autocorrelation-based fuzzy clustering of time series. Fuzzy Sets and Systems 160 (24), 3565–3589.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. In: KDD’96, AAAI Press, pp. 226–231.
Everitt, B., Landau, S., Leese, M., 2001. Cluster Analysis, forth ed. Arnold Press, London.
Everitt, B., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis, fifth ed. John Wiley & Sons, Ltd. London.
Gao, X., Yu, F., 2016. Fuzzy C-means with spatiotemporal constraints. In: Proceedings - 2016 IEEE International Symposium on Computer, Consumer and Control, IS3C 2016. pp. 337–340.
García-Escudero, L.A., Gordaliza, A., 1999. Robustness properties of k means and trimmed k means. J. Amer. Statist. Assoc. 94, 956–969.

García-Escudero, L.A., Gordaliza, A., Matrán, C., Mayo-Iscar, A., 2010. A review of robust clustering methods. Adv. Data Anal. Classif. 4, 89–109.

Heiser, W., Groenen, P., 1997. Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. Psychometrika 62 (1), 63–83.

Hu, T., Sung, S., 2006. A hybrid EM approach to spatial clustering. Comput. Statist. Data Anal. 50 (5), 1188–1205.

Hwang, H., Desarbo, W., Takane, Y., 2007. Fuzzy clusterwise generalized structured component analysis. Psychometrika 72 (2), 181–198.

Ienco, D., Bordogna, G., 2016. Fuzzy extensions of the DBScan clustering algorithm. Soft Comput. 1–12.

Iss, 2021. Characteristics of SARS-CoV-2 Patients Dying in Italy Report Based on Available Data on January 27th, 2021. Tech. rep., URL https://www.epicentro.iss.it/en/coronavirus/bollettino/Report-COVID-2019_27_january_2021.pdf.

Istat, Iss, 2020. Impatto Dell’Epidemia COVID-19 Sulla Mortalità Totale della Popolazione Residente Periodo Gennaio-Novembre 2020. Tech. rep., URL https://www.istat.it/it/files//2020/12/Rapp_Istat_Iss.pdf.

Izakian, H., Pedrycz, W., Jamal, I., 2013. Clustering spatiotemporal data: An augmented fuzzy C-means. IEEE Trans. Fuzzy Syst. 21 (5), 855–868.

Kaufman, L., Rousseeuw, P., 2005a. Finding Groups in Data: An Introduction to Cluster Analysis. WileyBlackwell.

Kaufman, L., Rousseeuw, P., 2005b. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons.

Lafuente-Rego, B., D’Urso, P., Vilar, J., 2018. Robust fuzzy clustering based on quantile autocovariances. Statist. Papers 1–56.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New Engl. J. Med. 382 (13), 1199–1207. http://dx.doi.org/10.1056/NEJMoa2001316.

Maharaj, E.A., D’Urso, P., 2011. Fuzzy clustering of time series in the frequency domain. Inform. Sci. 181 (7), 1187–1211.

Maharaj, E., D’Urso, P., Galagedera, D., 2010. Wavelet-based fuzzy clustering of time series. J. Classification 27, 231–275.

McBratney, A., Moore, A., 1985. Application of fuzzy sets to climatic classification. Agricult. Forest Meteorol. 35 (1-4), 165–185.

Moran, P., 1950a. A test for the serial independence of residuals. Biometrika 37, 178–181.

Moran, P., 1950b. A test for the serial independence of residuals. Biometrika 37 (1-2), 178–181.

Torabi, M., 2014. Spatial generalized linear mixed models with multivariate CAR models for areal data. Spat. Statist. 10, 12–26.

Torabi, M., 2016. Hierarchical multivariate mixture generalized linear models for the analysis of spatial data: An application to disease mapping. Biom. J. 58 (5), 1138–1150.

Viroli, C., 2011. Finite mixtures of matrix normal distributions for classifying three-way data. Stat. Comput. 21 (4), 511–522.

Wang, M., Wang, A., Li, A., 2006. Mining spatial-temporal clusters from geo-databases. In: Li, X., Zaiane, O.R., Li, Z. (Eds.), Advanced Data Mining and Applications: Second International Conference, ADMA 2006, Xi’an, China, August 14-16, 2006 Proceedings. In: Advanced Data Mining and Applications, Springer, Berlin, Germany, pp. 263–270.

Wedel, M., Kamakura, W., 2000. Market Segmentation: Conceptual and Methodological Foundations, Vol. 8. Springer.

Wu, K.-L., Yang, M.-S., 2002. Alternative c-means clustering algorithms. Pattern Recognit. 35 (10), 2267–2278.

Xie, J., Gao, H., Xie, W., Liu, X., Grant, P., 2016. Robust clustering by detecting density peaks and assigning points based on fuzzy weighted K-nearest neighbors. Inform. Sci. 354, 19–40.

Zhang, D.-Q., Chen, S.-C., 2004. A comment on “Alternative c-means clustering algorithm”. Pattern Recognit. 37 (2), 173–174.