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.
Community mobility in the European regions during COVID-19 pandemic: A partitioning around medoids with noise cluster based on space–time autoregressive models

Pierpaolo D’Urso a,*, Massimo Mucciardi b, Edoardo Otranto c, Vincenzina Vitale a

a Department of Social and Economic Sciences, Sapienza University of Rome, Italy
b Department of Cognitive Science, Education and Cultural Studies, University of Messina, Italy
c Department of Economics, University of Messina, Italy

ARTICLE INFO

Article history:
Received 22 May 2021
Received in revised form 7 July 2021
Accepted 8 July 2021
Available online 17 July 2021

Keywords:
COVID-19 outbreak
STAR model
Fuzzy C-medoids clustering
Robust clustering
Google COVID-19 community mobility report
NUTS 2

ABSTRACT

In this paper we propose a robust fuzzy clustering model, the STAR-based Fuzzy C-Medoids Clustering model with Noise Cluster, to define territorial partitions of the European regions (NUTS2) according to the workplaces mobility trends for places of work provided by Google with reference to the whole COVID-19 pandemic period. The clustering model takes into account both temporal and spatial information by means of the autoregressive temporal and spatial coefficients of the STAR model. The proposed clustering model through the noise cluster is capable of neutralizing the negative effects of noisy data. The main empirical results regard the expected direct relationship between the Community mobility trend and the lockdown periods, and a clear spatial interaction effect among neighboring regions.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The COVID-19 pandemic has affected the dynamics of all social and economic variables, very often with the conditioning of policy actions dedicated to the attempt to stem the contagion. One of the variables mainly affected by these blocking policies actions has been mobility. These

* Corresponding author.
E-mail addresses: pierpaolo.durso@uniroma1.it (P. D’Urso), massimo.mucciardi@unime.it (M. Mucciardi), edoardo.otranto@unime.it (E. Otranto), vincenzina.vitale@uniroma1.it (V. Vitale).

https://doi.org/10.1016/j.spasta.2021.100531
2211-6753/© 2021 Elsevier B.V. All rights reserved.
policies have encouraged work at home, thus reducing the number of commuters on the assumption that social distancing reduces the risk of infection for travelers and other people in their areas of residence and work (Francetic and Munford, 2021). So the most significant difference in behavior was staying at home and then working remotely (i.e. smart working). As a result, mobility for work purposes in many countries has been limited. Authorities continued to require their employees to work at home to avoid crowding of the roads and public transport networks (Lozzi et al., 2020). At the European level, the epidemic disease has produced a diversity of policy responses in the countries. These actions ranged from very drastic lockdown policies implemented in Southern Europe to less stringent approaches implemented elsewhere (e.g. Sweden) (Mendolia et al., 2020). On the other hand, it seems that the data on mobility do not follow the dynamics of contagion, with different behaviors among countries. This is what emerges from Google’s Mobility Report (Anon, 2021) relating to European countries, which highlights, for example, how the United Kingdom, Italy, most of Germany, the Netherlands show a net reduction in mobility due to local restriction rules, while Spain, many regions of France and Switzerland have an idiosyncratic behavior.

In the literature, recently, some interesting studies analyze the association between community mobility and COVID-19 outbreak (Carteni et al., 2020; Saha et al., 2020; Ophir et al., 2021; Sulyok and Walker, 2020; Lapatinas, 2020).

An interesting aspect in the mobility dynamics is to understand if they are affected by spatial interactions due to the policy actions of the neighboring countries or the level of contagion of neighbors (also within the same country at a local regional level).

In this paper we propose to characterize the 451 European regions (NUTS2) identified by the classification ISO alpha 2 (ISO 3166) by the International Organization for Standardization with a set of coefficients derived from a simple Space–Time AutoRegressive (STAR) model, relative to both the time and space dimensions, and cluster the spatial units using a model-based fuzzy clustering algorithm. From a methodological point of view, this work belongs to the model-based clustering approach, generally adopted in a time series framework (see, D’Urso, 2016; Maharaj et al., 2019), in particular to the class of methods for panel data classification, where the clustering is based on two dimensions (space and time in our case), see Frühwirth-Schnatter (2011). The idea of this methodology is to compare the estimated coefficients of similar models referring to different series to verify which of them are characterized by the same data generating process. Starting from the AR distance for time series processes (Piccolo, 1990), several authors have proposed clustering algorithms for time series based on distance measure between coefficients (see, for example, Maharaj, 1999; Otranto, 2008, 2010; D’Urso et al., 2013a,b, 2015, 2016, 2017). Notice that, in the literature, other clustering approaches which manage the spatial–temporal information in a different manner have been proposed by Disegna et al. (2017), D’Urso et al. (2019), D’Urso and Vitale (2020), D’Urso et al. (2021).

Considering the time series of the 451 European spatial units for the period from February 15, 2020 to April 18, 2021, we clustered them based on the STAR coefficients relative to each spatial unit and detecting spatial patterns with groups with similar coefficients, that can be considered as units with similar space–time dynamics. Moreover, we split the time span in four subsets to verify how the classification changes in different phases of the pandemic period.

In Section 2 we present the methodology adopted (model and clustering approach) to derive the proposed classification. In Section 3, after describing the dataset used in this analysis, the clustering results are presented; some final considerations conclude the paper.

2. Classification of spatial units: methodology

2.1. The space–time model

Our proposal consists in characterizing each spatial unit with a set of coefficients derived from a space–time model. The reference model is a classical STAR model (see, for example, Anselin, 1988) but with the properties of flexibility contained in Otranto and Mucciardi (2019). Their Flexible STAR (FSTAR) approach consists in the specification of a STAR model with the same coefficients for groups of spatial units; the final specification is obtained with a recursive procedure, which starts from
an overparameterized model with different coefficients for each single spatial unit and converging toward a parsimonious model with a small set of parameters, obtained through statistical tests to verify the equality of coefficients of different spatial units.

Let us call $y_t = (y_{i,t})$ ($i = 1, \ldots, n; t = 1, \ldots, T$) the $n \times 1$ vector containing the variable of interest for each spatial unit; the general STAR model adopted is:

$$
y_t = \sum_{p=1}^{P} \phi_p \odot y_{t-p} + \sum_{q=0}^{Q} \psi_q \odot (Wy_{t-q})
$$

(1)

where $W = \{w_{ij}\}, i, j = 1, \ldots, n$ with $w_{ii} = 0$ for each $i$, is the $n \times n$ weight matrix representing the spatial connection between each pair of spatial units, $\phi_p = (\phi_{1,p}, \ldots, \phi_{n,p})'$ is a $n \times 1$ vector of coefficients (one for each spatial unit) referred to the $p$-th time lag, $\psi_q = (\psi_{1,q}, \ldots, \psi_{n,q})'$ is the $n \times 1$ vector of spatial coefficients at time lag $q$, $\odot$ represents the Hadamard (element-by-element) product. Not dealing with a forecasting framework, we consider the possibility that there is a simultaneous spatial dependence in the model ($q = 0$).

Each equation in (1) has the following form:

$$
y_{i,t} = \sum_{p=1}^{P} \phi_{i,p} y_{i,t-p} + \sum_{q=0}^{Q} \psi_{i,q} \sum_{j=1}^{n} w_{ij} y_{j,t-q} \quad i = 1, \ldots, n.
$$

(2)

The set of $n$ equations (2) is the starting point of the Otranto and Mucciardi (2019) procedure to obtain a flexible but parsimonious STAR model. The particular structure of our dataset, with the number of spatial units $n$ largely greater than the number of time periods $T$, can be analyzed unit by unit with $n$ univariate models as (2), whereas a simultaneous estimation, also considering groups of spatial units with the same $\phi_p$ and $\psi_q$ coefficients, as in model (1), is not feasible for collinearity problems. However, we can consider each spatial unit as characterized by the set of coefficients $\theta_i = (\phi_{i,1}, \ldots, \phi_{i,P}, \psi_{i,0}, \ldots, \psi_{i,Q})'$, with the $\phi_{i,p}$ ($p = 1, \ldots, P$) coefficients representing the dynamic behavior of the unit $i$ along time, and the $\psi_{i,q}$ ($q = 0, \ldots, Q$) coefficients representing the spatial dependence elements of the same spatial unit.

Each set of coefficients $\theta_i$ can be used in a clustering algorithm to detect patterns of similar units in terms of space–time features. In our empirical application we identify $P = 2$ and $Q = 0$. Among the several possible alternative for the exogenous weighted matrix $W$, we select the Gaussian kernel matrix (Otranto et al., 2016), where $w_{ij}$ is obtained by row-standardizing the matrix with elements:

$$
w^*_{ij} = \exp \left[ -0.5 \left( \frac{d_{ij}}{h} \right)^2 \right]
$$

(3)

such that each row sums up to one; in (3), $d_{ij}$ is the Euclidean distance between the centroids $i$ and $j$ of two regions identified by their spatial coordinates longitude and latitude and $h = \text{Maxmin}(d_{ij})$ is the selected kernel bandwidth. We recall that varying the bandwidth results in a different exponential decay profile, which in turn produces weights that vary more or less rapidly over space. As a consequence the spatial weight matrix has the important property of being sensitive to topological transformations of the territory. The Arcgis 10 software was used to compute the matrix of the distances between the territorial barycentres of the European regions to derive the spatial weight matrix $W$.

---

1. When the number of spatial units is large we have problems of ill-conditioning of the covariance matrix of the estimators, with a near-to-zero determinant. See Otranto and Gallo (1994) for a simulation study to detect balanced space-to-time dimensions.

2. In this paper we set $h = \text{Maxmin}(d_{ij}) = 459$ km. The Maxmin bandwidth is chosen in such a way that the following relationship is satisfied: $h_{\text{Maxmin}} = \max(e_1, e_2, \ldots, e_i, \ldots, e_n)$ where $e_i$ represents the minimum distance of the generic spatial unit $i$ with the other units $j$ (with $i \neq j$). As a consequence each spatial unit is connected to all the others (Mucciardi and Bertuccelli, 2012).
2.2. The STAR-based Fuzzy C-Medoids Clustering with Noise Cluster

STAR-based Fuzzy C-Medoids Clustering model with Noise Cluster (STAR-FCMdC-NC) deals with contaminating spatial time series units by introducing an “artificial” cluster, called noise cluster, in addition to the C real clusters, that collects all objects distant from the prototypes of the C groups.

Notice that the STAR-FCMdC-NC model is based on the Partitioning Around Medoids (PAM) approach: the partitioning process defines clusters in which the cluster prototypes, called medoids, are observed units, i.e. the actual objects in the cluster whose average dissimilarity to all the objects in the cluster is minimal. Medoids can provide more intuitive information about the dataset than the centroids (means) used in the Fuzzy C-Means clustering; in fact the centroids usually do not physically exist, meanwhile the medoids are the most (actual) representative objects in the dataset.

Given \( n \) spatial time series of length \( T \), let \( \Theta = \{ \theta_i : i = 1, \ldots, n \} \) be the matrix of model coefficients of dimension \( n \times (P + Q + 1) \) and let there be \( C + 1 \) clusters, the \( (C+1) \)-th representing the noise cluster (Davé, 1991); the STAR-FCMdC-NC model is, then, based on the minimization of the following objective function:

\[
\begin{align*}
\min & : \sum_{i=1}^{n} \sum_{c=1}^{C} u_{i,c}^m ||\theta_i - \bar{\theta}_c||^2 + \sum_{c=1}^{C} u_{i,C+1}^m \delta^2 = \\
= & \sum_{i=1}^{n} \sum_{c=1}^{C} u_{i,c}^m \left[ \sum_{p=1}^{P} (\phi_{i,p} - \bar{\phi}_c,p)^2 + \sum_{q=0}^{Q} (\psi_{i,q} - \bar{\psi}_c,q)^2 \right] + \sum_{c=1}^{C+1} u_{i,c+1}^m \delta^2 \\
\text{s.t.:} & \sum_{c=1}^{C} u_{i,c} = 1, \quad u_{i,c} \geq 0.
\end{align*}
\]

where \( u_{i,c} \) represents the fuzzy membership of the \( i \)-th units in the \( c \)-th cluster while \( m > 1 \) is the fuzziness parameter; the greater is its value, the higher is the fuzziness of the solution. The \( C \) medoids are denoted by \( \bar{\theta}_1, \ldots, \bar{\theta}_C \).

The \( \delta \) parameter is the noise distance whose role is that of giving, for higher values, more emphasis to the “noise component” in the objective function. A common setting for \( \delta^2 \) is

\[
\delta^2 = \rho (nC)^{-1} \sum_{i=1}^{n} \sum_{c=1}^{C} u_{i,c}^m \left[ \sum_{p=1}^{P} (\phi_{i,p} - \bar{\phi}_c,p)^2 + \sum_{q=0}^{Q} (\psi_{i,q} - \bar{\psi}_c,q)^2 \right],
\]

where \( \rho \) may range between 0.05 and 0.5, even if the results are in general not very sensitive to its values (Davé, 1991).

The iterative solutions of \( 4 \) are:

\[
u_{i,c} = \frac{1}{\sum_{c'=1}^{C} \left[ \sum_{p=1}^{P} (\phi_{i,p} - \bar{\phi}_c,p)^2 + \sum_{q=0}^{Q} (\psi_{i,q} - \bar{\psi}_c,q)^2 \right]^{\frac{1}{m-1}} + \left[ \frac{1}{\delta^2} \right]^{\frac{1}{m-1}}},
\]

and

\[
u_{i,(C+1)} = \frac{1}{\sum_{c'=1}^{C} \left[ \sum_{p=1}^{P} (\phi_{i,p} - \bar{\phi}_c,p)^2 + \sum_{q=0}^{Q} (\psi_{i,q} - \bar{\psi}_c,q)^2 \right]^{\frac{1}{m-1}} + \left[ \frac{1}{\delta^2} \right]^{\frac{1}{m-1}}},
\]

The computational steps of the proposed robust clustering model are described in Algorithm 1.
Algorithm 1 STAR-based Fuzzy C-Medoids Clustering with Noise Cluster (STAR-FCMdC-NC) algorithm

1: Fix C, δ, nstart and max.iter;
2: Set iter = 0;
3: Randomly initialize $U = \{u_1, \ldots, u_n\}$ with $u_i = (u_{i,1}, \ldots, u_{i,C})$;
4: Store the C initial medoids: $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_C$;
5: repeat
6: Store the current medoids $(\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_C)_{OLD} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_C)$;
7: Compute $u_i (i = 1, \ldots, n)$ by using (5) and (6);
8: Select the new medoids $\theta_c$ for $c = 1, \ldots, C$;
9: for $c = 1$ to $C$ do
10: \hspace{1cm} $q = \arg \min_{1 \leq p \leq n} \sum_{i=1}^{n} u_i^{m}_{i,c} \left[ \sum_{p=1}^{m} (\phi^{c,p} - \phi^{q,p})^2 + \sum_{q=0}^{Q} (\psi^{p,q} - \psi^{q,q})^2 \right]$. \\
11: \hspace{1cm} return $\tilde{\theta}_c = \theta_q$
12: end for
13: iter $\leftarrow$ iter + 1;
14: until $(\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_C)_{OLD} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_C)$ or iter = max.iter
15: Repeat initialization: Start nstart times from step 2 and provide as final output of the algorithm those parameters yielding the smallest value of the target function out of these nstart random initializations.

Summing up the proposed clustering model (4), we argue that it inherits different benefits by its structural features. Firstly, following the PAM (Partitioning Around Medoids) approach, the cluster prototypes (i.e., medoids) are not “fictitious” units like “centroids” but observed ones (Bezdek, 1981) making clusters interpretation easier (Kaufman and Rousseeuw, 2005). Then, the fuzzy approach could be considered a more attractive and suitable solution when there are no clear boundaries among clusters (McBratney and Moore, 1985; Wedel and Kamakura, 2000); by the matrix of membership degrees, a second-best cluster almost as good as the best cluster could also be identified (Everitt et al., 2011). As stated in García-Escudero and Gordaliza (1999) and García-Escudero et al. (2010), since PAM procedure provides only a “timid robustification” of the C-Means clustering, the model formulation based on noise cluster, in a fuzzy context, allows to handle with outliers neutralizing their negative effects on groups structure.

Furthermore, in our model, we take into account both temporal and spatial information which have been embedded by means of $nP$ autoregressive temporal coefficients and $n(Q + 1)$ autoregressive spatial coefficients; in this manner, the spatio-temporal information is preserved considering only $n(P + Q + 1)$ values instead of all time series observations leading to a sensible computational gain.

2.3. Validity measure

To choose the optimal number of groups, the Fuzzy Silhouette (FS) index has been adopted, that is one of most well known internal validity criteria proposed by Campello and Hruschka (2006). It is based on the weighted average of the individual silhouettees width, $\lambda_i$, as follows:

$$FS = \frac{\sum_{i=1}^{n}(u_{i,c} - u_{i,q})^2 \cdot \lambda_i}{\sum_{i=1}^{n}(u_{i,c} - u_{i,c})^2}.$$

$$\lambda_i = \frac{(b_{i,c} - a_{i,c})}{\max(b_{i,c}, a_{i,c})},$$

where $a_{i,c}$ is the average distance of object $i$ to all other objects belonging to the same cluster $c$ ($c = 1, \ldots, C$).\(^3\) If $d_{i,q}$ is the average distance of the object $i$ to all objects belonging to another

\(^3\) An object $i$ belongs to the $c$th fuzzy cluster if its membership degree to this cluster, $u_{i,c}$, is higher than the membership of this object to any other fuzzy cluster.
cluster, say \( q \), with \( q \neq c \), then \( b_{i,c} \) is the minimum average distance \( d_{i,q} \) computed over the \( q \) clusters: i.e. the dissimilarity of object \( i \) to its closest neighboring cluster. \((u_{i,c} - u_{i,q})^\alpha\) is the weight of each \( \lambda_i \), where \( u_{i,c} \) and \( u_{i,q} \) are the first and second largest elements of the \( i \)th row of the fuzzy partition matrix \( U \), respectively; it implies that an object nearer to the cluster prototype is more important than a fuzzy one. \( \alpha \geq 0 \) is an optional user defined weighting coefficient that, when set equal to 0, reduces the FS to its crisp version.

Ranging from \([-1, 1]\), the higher the value of FS, the better is the assignment of the units to the clusters implying that, simultaneously, the intra-cluster distance \((a_{i,c})\) is minimized while the inter-cluster distance \((b_{i,c})\) is maximized.

3. On clustering European regions in terms of community mobility and COVID-19 outbreak

3.1. Data description

Data on human mobility by country are drawn from the Google Community Mobility Reports (GCMR) released by Google for 133 countries. These reports are created with aggregated, anonymized sets of data from users’ mobile device and are classified into six location categories: retail and recreation, grocery and pharmacy, parks, public transports, workplaces and residential areas. The GCMR dataset shows how visitors to (or time spent in) categorized places change compared to baseline days (Anon, 2021). Data are provided as percentage variations in the number of visits or time spent in each category relative to a pre-COVID-19 baseline. The baseline day is the median value in the 5-week period from January 3rd to February 6th, 2020. However, it should be noted that the GCMR data does not represent a perfect random sample of the target population as users of Android smartphones and tablets may differ from the general population in terms of various demographic, social and economic characteristics. The time span we consider covers the period from February 15th, 2020 to April 21th, 2021 (429 daily data) and refer to 451 European regions. Moreover we consider four subperiods, chosen according to the main outbreak waves. The time windows considered are: First — from 02-15-2020 to 05-31-2020 (first wave of the pandemic); Second — from 06-01-2020 to 08-31-2020 (summer with a clear decrease of contagion); Third — from 09-01-2020 to 12-31-2020 (signals for a second wave of contagion); Fourth — from 01-01-2021 to 04-18-2021 (explosion of the second wave of the pandemic); Whole — from 02-15-2020 to 04-18-2021 (full dataset). Some European regions were not considered due to partial lack of data. To erase the presence of weekly seasonality patterns we transformed time series by moving averages of 7 terms.

Table 1 provides the descriptive statistics in the four subsets and for the whole period for workplace mobility category in Europe. The entries refer to the average values of each spatial units within the time span considered. The decrease in workplace mobility was on average around 28 percent with peaks of 33 percent in the first period. The great impact of the pandemic in the first subperiod, the attenuation of this effect in the second and third subperiod and a resurgence of the phenomenon in the fourth subperiod are evident. The greatest percentage decreases compared to the baseline (−92%) were recorded in the Lombardy region (Italy) and in the Community of Madrid (Spain) in April 2020. Specifically Fig. 1 shows the spatial distribution of workplace mobility in European regions at the beginning (first wave) of the COVID-19 pandemic (March 2020).
3.2. Clustering of the European regions

The proposed STAR-FCMdC-NC model has been applied to the matrix of model coefficients of dimension $451 \times 3$ (where $P = 2$ and $Q = 0$), for different time windows. The identification of the orders $P$ and $Q$ is based on the following procedure. We first identified, over the full time span, for each unit, the order $P$ by comparing three alternative $AR(P)$ models with $P = 1, 2, 3$; the comparison is made in terms of minimum BIC. In all the 451 spatial units the order $P = 2$ was selected. Then we compared, for each unit, three models (2) with $P = 2$ and $Q = 0, 1, 2$; in this case the result is a bit puzzling: BIC selects $Q = 0$ in 33.7% of units, $Q = 1$ in 26.6%, $Q = 2$ in 39.7%. However, repeating the same experiment for each sub-period, we see a clear preference for $Q = 0$ (see Table 2). For the sake of parsimony and to facilitate the interpretation and comparison of the models for clustering, we decide to fix $Q = 0$ for all models.

For each period, the number of groups has been chosen according to the combination of the $m$ and $C$ values that maximize the Fuzzy Silhouette (FS) index (Campello and Hruschka, 2006). To this purpose, the algorithm has been run for $m \in \{1.3, 1.5, 1.7\}$ and $C \in \{2, \ldots , 10\}$ leading to the solutions, for each time window, shown in Table 3. The FS values on varying $C$ have been shown in Fig. 2 for each time window. For each period, besides the noise cluster, two groups (whose medoids have been shown in Table 3) are always well identified as we can see by inspecting the membership degrees represented by means of the ternary plots of Figs. 3–4. Indeed, most of units...
Fig. 2. FS values on varying $C$, for each time window.
Table 3
The best combination of C and m based on the Fuzzy Silhouette (FS) index for each time window (Table includes also the Medoids of the two clusters).

| Time Window | C   | m   | FS     | Medoid Cluster 1 | Medoid Cluster 2 |
|-------------|-----|-----|--------|------------------|------------------|
| First       | 2   | 1.7 | 0.743  | Vwstra Galand county | LikaSenj county  |
| Second      | 2   | 1.7 | 0.765  | Aberdeen city     | Community of Madrid |
| Third       | 2   | 1.7 | 0.770  | Angus council     | Vaud             |
| Fourth      | 2   | 1.5 | 0.764  | Kooiceregion      | County Carlow    |
| Whole       | 2   | 1.7 | 0.781  | Komgrom–Esztergom | Geneva           |

Table 4
Clusters Centroids (the centroids have been computed also for the three groups of fuzzy units).

| Time Window | Cluster 1 | Cluster 2 | Fuzzy units clusters 1–2 | Fuzzy units clusters 1-Noise | Fuzzy units clusters 2-Noise |
|-------------|-----------|-----------|--------------------------|-----------------------------|-----------------------------|
| First       | 0.025     | 0.031     | 0.005                    | 0.008                       | 0.006                       |
|             | −0.077    | −0.130    | −0.106                   | −0.103                      | −0.152                      |
|             | 1.450     | 0.680     | 1.072                    | 2.127                       | −0.019                      |
| Second      | 0.023     | 0.027     | 0.022                    | 0.033                       | 0.016                       |
|             | −0.082    | −0.147    | −0.119                   | −0.034                      | −0.150                      |
|             | 1.355     | 0.545     | 0.979                    | 2.037                       | −0.097                      |
| Third       | 0.027     | 0.025     | 0.030                    | 0.003                       | 0.021                       |
|             | −0.087    | −0.122    | −0.071                   | −0.108                      | −0.156                      |
|             | 1.408     | 0.593     | 1.018                    | 2.097                       | −0.074                      |
| Fourth      | 0.027     | 0.016     | 0.016                    | −0.009                      | 0.022                       |
|             | −0.084    | −0.143    | −0.093                   | −0.097                      | −0.148                      |
|             | 1.334     | 0.557     | 0.954                    | 2.053                       | −0.147                      |
| Whole       | 0.031     | 0.021     | 0.034                    | 0.018                       | 0.017                       |
|             | −0.078    | −0.130    | −0.087                   | −0.072                      | −0.145                      |
|             | 1.355     | 0.604     | 0.975                    | 2.046                       | −0.064                      |

belongs to one of the two groups with high membership degree while lesser is the number of fuzzy units (distinguished among fuzzy units between the two natural clusters, fuzzy units between the first cluster and the noise one, fuzzy units between the second cluster and the noise one).

To improve clusters interpretation, we assigned each region to a specific group setting a cut-off value according to the ith time series is assigned to the cth cluster if $u_{i,c} \geq 0.6$. Taking into account the obtained crisp partition, the multidimensional scaling projection on two dimensions of the model coefficients and their representation on the tridimensional space have been generated, as shown in Figs. 5–9. By looking at these graphs, the structure in two separated groups becomes fairly evident together with the following important consideration, arisen above all from the 3D graph, i.e. groups have been mainly discriminated by the spatial coefficient. The violin plots of Figs. 10–14 strongly confirm this evidence. Indeed, looking at the kernel density of the model coefficients, we can conclude that the spatial component is the only one able to differentiate among observations, detecting a structure in two main groups while temporal autoregressive component has a very marginal role in the clustering process. On the other hand, we can argue that the spread of the distribution of the spatial coefficient is much wider than that of the temporal ones.

In Table 4 we show the average of each coefficient within the detected two clusters and the different groups of fuzzy data (excluding the noise cluster, which contains extreme values, both small and high) for the entire period and each subperiod. It is confirmed that the time coefficients $\phi_1$ and $\phi_2$ do not seem to characterize the cluster, while some regularity can be seen in the spatial coefficient $\psi_0$. In particular, it appears that these five groups can be interpreted in terms of the increasing level of spatial dependence in the following order: Fuzzy units cluster 2-Noise (very small spatial dependence), Cluster 2 (small), Fuzzy units Clusters 1–2 (medium), Cluster 1 (high), Fuzzy units Cluster 1-Noise (very high).
Given this interpretation, considering the entire period, we can see a low presence of very high spatial dependence (blue zones) and a high prevalence of dependence in Italy, a part of Spain, the United Kingdom and Eastern Europe (Fig. 15). By dividing the time span into the four
subperiods, we note a general small spatial dependence during the first pandemic wave (Fig. 16), excluding parts of the United Kingdom and Northern Italy (where the European pandemic began). Probably the beginning of the COVID-19 pandemic in Lombardy (Italy) could have had an immediate impact on workplaces mobility trends in the surrounding European regions due to the strong economic, cultural and geographical links between these regions. The summer of 2020, with the hope of the end of the spread of the pandemic, is denoted by a sharp increase in the small spatial dependence (Fig. 17). This trend seems to be confirmed in the third period (Fig. 18) with a diffusion of violet areas, but also large areas characterized by noise. The second pandemic wave of 2021 is characterized by new increases in space dependence (Fig. 19), especially in Italy and in the Balkan regions, due to the new restrictions adopted in particular by these countries. In summary, space dependence appears to be directly related to the actions of pandemic waves and the protective measures adopted by European nations.

Fig. 4. Ternary plots of membership degrees for clustering results referred to the fourth time window and to the whole one.
It might be interesting to link our results with some recent comments from the Spanish newspaper El Pais. In their purely descriptive analysis, they note that in 2021 the citizens of countries like Spain and France begun moving freely without taking into account the danger of contagion, while countries like Italy are blocked by internal restrictions. If we think of high spatial dependence as a uniformity of people’s behavior due to restrictions and low spatial dependence as idiosyncratic behavior of citizens, this picture is consistent with our analysis. It can be clearly seen how, in general (Fig. 15), Spain is characterized by a prevalence of gray areas (poor spatial relationship) with only a few black areas (high spatial dependence in the Center and North); France is clearly characterized by low spatial dependence (gray and purple zones); Northern Italy, where the European contagion started, is all black.
Also following the evolution over time (Figs. 16–19) we note how Spain has increased its spatial dependence from the second to the third period (growth of black and blue areas), but in the fourth there is again the prevalence of gray areas.

In France, only Normandy has a certain degree of spatial dependence (perhaps due to its proximity to England which has shown high levels of mortality due to the pandemic), but the rest of the country always remains in the gray and purple zone, with some cases of noise.

To this purpose, it is worth mentioning the two French regions Hauts-de-France and Grand Est (the latter involved in the early diffusion of pandemic due to the cluster formed after the gathering of the Christian Open Door Church in February 2020 in the Mulhouse city), characterized by rates of positives and deaths amongst the highest in France and classified, in almost all partitions, as outliers, with a negative sign of their associated spatial coefficient. A possible explanation could be based on their geographic position. The Grand Est region is located in Northeastern France and can be seen
as a “region gateway” being the only French region to border more than two countries. Indeed, it shares borders with the Wallonia region (Belgium), the Cantons of Esch-sur-Alzette and Remich (Luxembourg) on the North, with Germany on the East and Northeast and Switzerland on the Southeast. Within France, its neighbors are Bourgogne-Franche-Comté on the South, Île-de-France, seriously hit by pandemic on the West, and Hauts-de-France on the Northwest.
**Fig. 11.** Violin plots for Clusters 1 e 2 — Second time window.

**Fig. 12.** Violin plots for Clusters 1 e 2 — Third time window.
**Fig. 13.** Violin plots for Clusters 1 and 2 — Fourth time window.

**Fig. 14.** Violin plots for Clusters 1 and 2 — Whole time window.
As far as Hauts-de-France is concerned, it is the Northernmost region of France and the second most densely populated in metropolitan France after its neighbor Île-de-France. It also borders Grand Est and Normandy. More important, outside France, it is connected to England via the Channel Tunnel sharing borders with the Flanders and Wallonia (Belgium), overlooking the North Sea to the North.

Therefore, the two regions are neighbors and have a strategic geographic position, sharing borders with regions belonging to other countries. They figure as outliers, with a negative spatial dependence, probably because of the effect of the lack of uniformity of restriction policies that varied in time and intensity across the European countries.

Northern Italy always remains black, reflecting the strong restrictions imposed in this country.

4. Concluding remarks

Social restrictions and local lockdowns have been implemented in many Europe countries to reduce viral transmission during the COVID-19 pandemic. Consequently these restrictions had a big impact in human mobility. Specifically to evaluate changes in workplace mobility we use GCMR data and propose an innovative strategy analysis. So we estimate a robust fuzzy partitioning
around medoids model with noise cluster for clustering the European regions (NUTS2) -identified by the classification ISO alpha 2 (ISO 3166) by the International Organization for Standardization — considering spatial and temporal information and using community mobility during COVID-19 pandemic. In particular, considering the time series of the 451 European spatial units for the period from February 15th, 2020 to April 18th, 2021, we cluster them based on a set of estimated coefficients derived from a simple Space-Time AutoRegressive (STAR) model, relative to each spatial unit, and detecting spatial patterns with cluster with similar coefficients that can be considered as units with similar space–time dynamics. To cluster the European regions we proposed a model-based fuzzy clustering algorithm, i.e. the STAR-based Fuzzy C-Medoids Clustering model with Noise Cluster (STAR-FCMdC-NC). In addition, we split the time span in four subsets to verify how the classification changes in different phases of the COVID pandemic period. Clustering results are mainly influenced by spatial component that, in turn, as expected, seems to reflect the different regimes of restrictive measures adopted by the European Governments to face the outbreak waves. The effects of the total and local lockdowns across the different phases of pandemic and across regions translate into a spatial autocorrelation more than into a temporal one. The results seem consistent with the phases of the pandemic (first wave, illusion of the end of the pandemic, second wave), the blocking actions of European nations, the different times of spread of the pandemic
among the European countries. To conclude, in this study we observed that there is a significant spatial dependence in workplace mobility reduction that relates with the incidence of the pandemic, demonstrating (with some exceptions) an impact of restrictions on mobility proportional to the epidemiological situation. We believe that the spatial–temporal approach of human mobility data can help to inform the European government in order to implement common policies capable of balancing the economic and social effects and controlling the spread of the pandemic. In the future, we will focus our attention on other possible methodological and empirical investigations connected to the territorial partitioning procedure based on complex data structures and, in particular, related to community mobility during the COVID-19 pandemic. In particular, from a methodological point of view, we will investigate the possibility to add another source of uncertainty in the partitioning process, i.e. the statistical uncertainty associated to the estimates of the temporal and spatial parameters of the STAR models; furthermore, we will define a weighting system for tuning in an objective manner the influence of the temporal and spatial information in the clustering process and we will explore other kinds of robust clustering procedures for spatial–temporal univariate and multivariate data.

Fig. 17. Map of clustering results — Second time window.
In our empirical study we analyzed workplaces mobility trends for places of work. In the future, we will extend our empirical study adding in the dataset the Google mobility trends data related to other place categories; i.e. grocery & pharmacy (mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies), parks (mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens), transit stations (mobility trends for places like public transport hubs such as subway, bus, and train stations), retail & recreation (mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters) and residential (mobility trends for places of residence) (Anon, 2021).

**Acknowledgments**

The authors thank the Editor and the referees for their useful comments and suggestions which helped to improve the quality and presentation of this manuscript.
Fig. 19. Map of clustering results — Fourth time window.

References

Anon, 2021. Google COVID-19 community mobility reports. Retrieved from: https://www.google.com/covid19/mobility/april2021.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht.

Bezdek, J., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic, Norwell, MA, USA.

Campello, R.J.G.B., Hruschka, E.R., 2006. A fuzzy extension of the silhouette width criterion for cluster analysis. Fuzzy Sets and Systems 157, 2858–2875.

Cartenì, A., Di Francesco, L., Martino, M., 2020. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. Sci. Total Environ. 741, 140489.

Davé, R.N., 1991. Characterization and detection of noise in clustering. Pattern Recognit. Lett. 12, 657–664.

Disegna, M., D’Urso, P., Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. Spatial Statist. 21, 209–225.

D’Urso, P., 2016. Fuzzy clustering. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.), Handbook of Cluster Analysis. Chapman and Hall, pp. 241–263.

D’Urso, P., De Giovanni, L., Disegna, M., Massari, R., 2019. Fuzzy clustering with spatial-temporal information. Spatial 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., De Giovanni, L., Massari, R., 2016. GARCH-based robust clustering of time series. Fuzzy Sets and Systems 305, 1–28.
D’Urso, P., De Giovanni, L., Massari, R., Di Lallo, D., 2013a. Noise fuzzy clustering of time series by autoregressive metric. Metron 71 (3), 217–243.

D’Urso, P., De Giovanni, L., Vitale, V., 2021. Spatial robust fuzzy clustering of COVID 19 time series based on B-splines. Spatial Statist. 100518.

D’Urso, P., Di Lallo, D., Maharaj, E.A., 2013b. Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks. Soft Comput. 17 (1), 83–131.

D’Urso, P., Massari, R., Cappelli, C., De Giovanni, L., 2017. Autoregressive metric-based trimmed fuzzy clustering with an application to PM10 time series. Chemometr. Intell. Lab. Syst. 161, 15–26.

D’Urso, P., Vitale, V., 2020. A robust hierarchical clustering for georeferenced data. Spatial Statist. 35, 100407.

Everitt, B., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis, fifth ed. John Wiley & Sons, Ltd, London.

Francetic, I., Munford, L., 2021. Corona and coffee on your commute: a spatial analysis of COVID-19 mortality and commuting flows in England in 2020. Eur. J. Public Health 1–7.

Frühwirth-Schnatter, S., 2011. Panel data analysis: a survey on model–based clustering of time series. Adv. Data Anal. Classif. 5, 251–280.

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.

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

Lapatinas, A., 2020. The Effect of COVID-19 Confinement Policies on Community Mobility Trends in the EU, EUR 30258 EN. Publications Office of the European Union, Luxembourg.

Lozzi, G., Rodrigues, M., Marcucci, E., Teod, T., Gatta, V., Pacelli, V., 2020. Research for TRAN Committee – COVID-19 and urban mobility: impacts and perspectives. In: European Parliament. Policy Department for Structural and Cohesion Policies, Brussels, pp. 1–24.

Maharaj, E.A., 1999. Comparison and classification of stationary multivariate time series. Pattern Recognit. 32, 1129–1138.

Maharaj, E.A., D’Urso, P., Caiado, J., 2019. Time Series Clustering and Classification. Chapman and Hall, http://dx.doi.org/10.1201/9780429058264.

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

Mendolia, S., Stavrunova, O., Yerokhin, O., 2020. Determinants of the community mobility during the COVID-19 epidemic: The role of government regulations and information. Discussion Paper Series - IZA Institute of Labor Economics. pp. 1–27.

Mucciardi, M., Bertuccelli, P., 2012. The impact of the weight matrix on the local indicators of spatial association: an application to per-capita value added in Italy. Int. J. Trade Glob. Mark. 5, 133–141.

Ophir, Y., Walter, D., Arnon, D., Lokmanoglu, A., Tizzoni, M., Carota, J., D’Antiga, L., Nicastro, E., 2021. The framing of COVID-19 in Italian media and its relationship with community mobility: A mixed-method approach. J. Health Commun. 1–13.

Otranto, E., 2008. Clustering heteroskedastic time series by model-based procedures. Comput. Statist. Data Anal. 52 (10), 4685–4698.

Otranto, E., 2010. Identifying financial time series with similar dynamic conditional correlation. Comput. Statist. Data Anal. 54 (1), 1–15.

Otranto, E., Gallo, G.M., 1994. Regression diagnostic techniques to detect space–to–time ratios in STARMA models. Metron 52, 129–145.

Otranto, E., Mucciardi, M., 2019. Clustering space-time series: FSTAR as a flexible STAR approach. Adv. Data Anal. Classif. 13 (1), 175–199.

Piccolo, D., 1990. A distance measure for classifying ARIMA models. J. Time Series Anal. 11 (2), 153–164.

Saha, J., Barman, B., Chouhan, P., 2020. Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. Child. Youth Serv. Rev. 116, 105160.

Sulyok, M., Walker, M., 2020. Community movement and COVID-19: a global study using Google’s Community Mobility Reports. Epidemiol. Infect. 148.

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