Statistical Methods & Applications (2023) 32:1539–1561
https://doi.org/10.1007/s10260-023-00697-y

ORIGINAL PAPER

Constrained optimization for addressing spatial heterogeneity in principal component analysis: an application to composite indicators

Paolo Postiglione[1] · **Alfredo Cartone**[1] · **M. Simona Andreano**[2] · **Roberto Benedetti**[1]

Accepted: 10 April 2023 / Published online: 2 May 2023

Abstract

Principal component analysis, in its standard version, might not be appropriate for the analysis of spatial data. Particularly, the presence of spatial heterogeneity has been recognized as a possible source of misspecification for the derivation of composite indicators using principal component analysis. In recent times, geographically weighted approach to principal component analysis has been used for the treatment of continuous heterogeneity. However, this technique poses problems for the treatment of discrete heterogeneity and the interpretation of the results. The aim of this paper is to present a new approach to consider spatial heterogeneity in principal component analysis by using simulated annealing algorithm. The proposed method is applied for the definition of a composite indicator of local services for 121 municipalities in the province of Rome.

Keywords Simulated annealing · GWPCA · Spatial effects · Local well-being

M. Simona Andreano: Deceased.

The current version of this paper was completed after our dear friend and co-author M. Simona Andreano passed away. We have worked together for a long time, enjoying a wonderful friendship. We really miss her every day. We dedicate this paper to her memory.

✉ Paolo Postiglione
postigli@unich.it

Alfredo Cartone
alfredo.cartone@unich.it

Roberto Benedetti
benedett@unich.it

[1] Department of Economic Studies, "G. d'Annunzio" University of Chieti-Pescara, Viale Pindaro, 42, 65127 Pescara, Italy

[2] Universitas Mercatorum, Rome, Italy

1 Introduction

Principal component analysis (PCA) is a statistical method employed in many applied fields. The statistical and mathematical properties of PCA allow practitioners to derive a set of independent components from large datasets. This collection is obtained by decomposing the eigenstructure of the variance–covariance (VC) matrix, which returns loadings corresponding to eigenvectors and component scores corresponding to new sets of coordinates (Jolliffe 2002).

PCA is also used to synthetise different variables into multidimensional (i.e., composite) indicators (OECD 2008). Component scores from PCA may be considered less *subjective* indicators since weights are not arbitrarily assigned but data driven. De Muro et al. (2011) and Mazziotta and Pareto (2019), among others, criticise this use of PCA, as its weights are defined through a purely statistical technique and do not always reflect the actual relevance of the particular variables. Despite these limitations, PCA is largely used to define composite measures (Decanq and Lugo 2013).

Standard PCA is also mainstream in the analysis of geographically distributed data (Demšar et al. 2013). However, conventional PCA is often applied to spatial objects, while "*geographical effects do not play any role in the PCA itself*" (Demšar et al. 2013; p. 111). Hence, adapting PCA to spatial issues represents a relevant objective for researchers and practitioners.

Indeed, spatial data present particular characteristics that must be considered when applying statistical techniques. In the literature, two different spatial effects can be considered: spatial dependence and spatial heterogeneity (Anselin 1988). While the first implies the analysis of spatial autocorrelation produced by contagion between spatial units (LeSage and Pace 2009), the second considers spatial instabilities in estimated coefficients due to spatial regimes or heteroskedasticity (Anselin 1988; Postiglione et al. 2013; Murakami and Griffith 2019).

In recent decades, the literature has largely focused on the analysis of spatial dependence. Conversely, spatial heterogeneity has been considered only in a smaller and more recent number of contributions. Examples of techniques that account for spatial heterogeneity in regression models are geographically weighted regression (GWR, Fotheringham et al. 2002) and spatially varying coefficients (Wheeler and Calder 2007).

Presumably, the prevalence of studies about spatial dependence can be explained by two factors. Primarily, the analysis of dependence has been explored to avoid potential biases due to spatial autocorrelation, particularly while estimating regressions for geographical data that are popular in theoretical and empirical research. Second, sophisticated algorithms are often needed for modelling spatial heterogeneity, especially in a cross-section framework (Andreano et al. 2017; Billè et al. 2017). This has left room for the analysis of cross-sectional heterogeneity as one of the open challenges for spatial analysts (Postiglione et al. 2013; Murakami and Griffith 2019).

PCA is not exempted from a deeper discussion about spatial effects. In this sense, a first contribution that considered spatial dependence in PCA, denoted as

spatial PCA (i.e., sPCA), was given by Jombart et al. (2008). sPCA defines spatial components through an objective function that combines spatial dependence with standard decomposition of the VC matrix. Essentially, this method finds new composite measures that no longer maximise the variance of the scores (as in PCA) but instead maximises the product of their variance and Moran's I (Moran 1950). For an interesting application of sPCA to the definition of a composite measure of well-being, see Giacalone et al. (2022).

Concerning spatial heterogeneity, this has been mainly investigated for PCA through geographically weighted principal components analysis (GWPCA; Fotheringham et al. 2002). In GWPCA, the local VC matrices are obtained by using a kernel function so that GW components are influenced more by closer observations, and spatial instabilities are approached by obtaining loadings and scores at each unit. However, the loadings and components from GWPCA are not always straightforward, as results must be listed in wide sets of local loadings or mapped in terms of locally relevant variables (i.e., winning variables; Harris et al. 2011).

Following this narrative, this paper contributes to the development of PCA for geographical data (Wartenberg 1985; Jombart et al. 2008; Lloyd 2010; Harris et al. 2011; Sarra and Nissi 2020; Cartone and Postiglione 2021) in the special case of spatial heterogeneity. In this sense, our contribution substantially differs from Jombart et al. (2008) and Giacalone et al. (2022), who used sPCA to examine spatial dependence.

To this end, we consider spatial heterogeneity as a criterion to divide the sample of observations into smaller homogeneous groups, individuating subpopulations as a solution of a combinatorial optimisation problem. We introduce a new algorithm that extends the application of simulated annealing (SA) in spatial regression (Postiglione et al. 2013) to the case of PCA (hereafter, SA-PCA). When solving this optimisation problem, the proposed algorithm decomposes the VC matrix to compute eigenvalues, eigenvectors, and component scores for a parsimonious number of regimes and overcomes the hypothesis of spatial homogeneity.

In this paper, we also note that the SA-PCA algorithm can be an alternative to the use of cluster techniques on geographical units when using coordinates and/or other attributes (for example, when using k-means) and computation of PCA to each cluster. This method may be denoted as a cluster-specific PCA. In a different manner, Libório et al. (2022) have already extended the use of k-means clustering to PCA for computing piecewise composite indicators.

However, two main differences exist between a cluster-specific PCA and SA-PCA. First, by using k-means, clusters are calculated at the first stage, and after, components are obtained by applying PCA on groups. Conversely, local loadings in SA-PCA are individuated by directly modelling instability in the VC matrix, and information from PCA is used step by step in the algorithm for the determination of the groups. Second, in our method, spatial information is explicitly accounted for. Thus, while cluster-specific PCA is meant to individuate components without considering any geographical issue, our proposal aims to tackle unobserved spatial heterogeneity by considering space directly in the objective function.

Furthermore, this contribution differs from GWPCA for two main reasons. First, SA-PCA does not address nonstationarity in continuous space, as it directly considers

the problem of structural differences due to multiple regimes. In fact, spatial heterogeneity can also be present in the form of discrete heterogeneity for which parameters vary between spatial regimes or groups of regions (Anselin 2010). Second, in this paper, heterogeneous eigenvectors are individuated for a limited number of clusters. Thus, in the presence of these features, SA-PCA addresses some critical aspects of GWPCA, especially to improve the interpretability of its results.

Since the use of PCA for composite indicators has been extensively explored (Pampalon and Raymond 2000; Havard et al. 2008), we start from the definition of a composite measure to evaluate the consequences of spatial heterogeneity. Then, a multidimensional indicator of local services for 121 municipalities in the province of Rome is defined. This province is selected as it is one of the most densely populated in Italy.

In the application, various aspects are discussed. SA-PCA, GWPCA and cluster-specific PCA are performed to treat spatial heterogeneity, and the results are compared. The results show that spatial heterogeneity is generally relevant in PCA and that SA-PCA can be used to efficiently individuate endogenous groups and to capture local characteristics. This may enrich the insight in terms of ad hoc policies at the local level.

Last, as highlighted by Jombart et al. (2008), it is meaningless to compare sPCA eigenvalues to the sum of all eigenvalues as in PCA. Additionally, the percentage of the total criterion associated with an eigenvalue in sPCA cannot be used as a direct rule to choose the number of components in terms of the explained variance. Furthermore, since eigenvalues may be both positive and negative in sPCA, it can be difficult to select more representative components for PCA. By considering this last drawback and our focus on spatial heterogeneity, we limited the comparison of SA-PCA to standard PCA, GWPCA, and cluster-specific PCA.

The layout of the paper is as follows. Section 2 is devoted to summarising the methodological contribution of the paper, with a review of the main characteristics of PCA and applying SA to identify zones of local stationarity for eigenvalues and eigenvectors. Section 3 contains the description of the data collected from the ISTAT Statistical Atlas of Municipalities and the results of the composite indicator. In this section, we also apply our algorithm to compare the results of SA-PCA to those of PCA, GWPCA, and cluster-specific PCA. Finally, Sect. 4 presents some concluding remarks and outlines the future research agenda.

2 Methodology

PCA is based on the analysis of a centred matrix $\mathbf{X}_{nm}$ with n statistical units and m variables. The essential idea of PCA is the representation of units in q-dimensional subspaces (with $q < m$), retaining the maximum amount of statistical information. The reduction in data dimensionality allows easier interpretative analysis.

A primary result in PCA is (Jolliffe 2002):

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^t \tag{1}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues, $\mathbf{A}$ is the corresponding matrix of loadings (i.e., the eigenvectors), and $\mathbf{\Sigma}$ is the VC matrix. The eigenvalue λ_j in $\mathbf{\Lambda}$ represents the variance of the principal component $\mathbf{Y}_j$ defined as:

$$\mathbf{Y}_j = \mathbf{X}\mathbf{a}_j \tag{2}$$

where $\mathbf{a}_j$ is the j-th column of the loading matrix $\mathbf{A}$ of $\mathbf{\Sigma}$ and represents the contribution of each variable in $\mathbf{X}$ to the j-th principal component $\mathbf{Y}_j$.

Basically, for an adequate q, the component scores related to components $q+1$ to m represent the Euclidean distances alongside the axes of the corresponding orthogonal vectors to a q-dimensional linear subspace. The first q components are chosen so that this subspace contains the highest proportion of the total variance. The first q components are described by:

$$\mathbf{Y}_q = \mathbf{X}\mathbf{A}_q \tag{3}$$

where $\mathbf{Y}_q$ is the $n \times q$ score matrix and $\mathbf{A}_q$ is the $m \times q$ loading matrix with only the first q columns of $\mathbf{A}$.

Jolliffe (2002) describes an important property that is very useful for the definition of our algorithm. Making use of singular value decomposition theorem, data matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = \sum_{j=1}^{r} \mathbf{d}_j l_j^{1/2} \mathbf{a}_j^t = \sum_{j=1}^{r} l_j^{-1/2} \mathbf{X}\mathbf{a}_j l_j^{1/2} \mathbf{a}_j^t = \sum_{j=1}^{r} \mathbf{X}\mathbf{a}_j \mathbf{a}_j^t \tag{4}$$

where $\mathbf{d}_j$ is the j-th column of the matrix $\mathbf{D}$ (with $\mathbf{D}^t\mathbf{D} = \mathbf{I}_r$), l_j denotes the j-th eigenvalue of $\mathbf{X}^t\mathbf{X}$ (with $l_j = \frac{\lambda_j}{n-1}$) and $\mathbf{a}_j$ is the j-th column of the matrix $\mathbf{A}$(with $\mathbf{A}^t\mathbf{A} = \mathbf{I}_r$). The rank of $\mathbf{X}$ is supposed to be r, and thus, $l_j = 0$ (and $\mathbf{X}\mathbf{a}_j = 0$) for $j = (r+1), (r+2), \ldots, m$.

Equation (4) can be written element by element as:

$$x_{il} = \sum_{j=1}^{r} d_{ij} l_j^{1/2} a_{lj} \tag{5}$$

where d_{ij} is the (i,j)-th entry of $\mathbf{D}$ and a_{lj} is the (l,j)-th entry of $\mathbf{A}$.

Retaining a number of q components (with $q < r$), Eq. (5) can be written as:

$${}_q\tilde{x}_{il} = \sum_{j=1}^{q} d_{ij} l_j^{1/2} a_{lj} \tag{6}$$

where $d_{ij} l_j^{1/2} a_{lj}$ is the part of x_{il} corresponding to the j-th component for $j = 1, 2, \ldots, q$. Householder and Young (1938) and Gabriel (1978) highlighted that ${}_q\tilde{x}_{il}$ represents the best rank q approximation to x_{il}, in the sense that ${}_q\tilde{x}_{il}$ minimises the following function:

$$\sum_{i=1}^{n}\sum_{l=1}^{m}({}_{q}x_{il}-x_{il})^2 \tag{7}$$

where ${}_{q}x_{il}$ is any possible rank q approximation to x_{il}. According to (6) and (7), an $n \times m$ residual matrix $\mathbf{S}$ can be defined as the difference between the data matrix $\mathbf{X}$ and the best rank q approximation ${}_{q}\tilde{\mathbf{X}}$ as:

$$\mathbf{S}=\mathbf{X}-{}_{q}\tilde{\mathbf{X}} \tag{8}$$

where ${}_{q}\tilde{\mathbf{X}}=\sum_{j=1}^{q}\mathbf{X}\mathbf{a}_j\mathbf{a}_j^t=\mathbf{X}\mathbf{A}_q\mathbf{A}_q^t$ (see Harris et al. 2015). Equation (8) represents the core function of our proposed algorithm.

In standard PCA, the implicit hypothesis is that the VC structure of the process is homogenous throughout the geographical area under investigation. This assumption is obviously not always realistic for geographically distributed data (Harris et al. 2015). Therefore, it is necessary to relax this hypothesis to consider potential heterogeneity in principal components.

The first appropriate PCA technique for spatial data is represented by GWPCA (Fotheringham et al. 2002; Harris et al. 2011). In this approach, Eq. (1) can be generalised as (Harris et al. 2015):

$$\mathbf{\Sigma}\left(g_i,h_i\right)=\mathbf{A}\left(g_i,h_i\right)\mathbf{\Lambda}\left(g_i,h_i\right)\mathbf{A}\left(g_i,h_i\right)^t \tag{9}$$

where $\mathbf{\Lambda}\left(g_i,h_i\right)$ is the diagonal matrix of local eigenvalues, $\mathbf{A}\left(g_i,h_i\right)$ is the corresponding matrix of local eigenvectors, $\mathbf{\Sigma}\left(g_i,h_i\right)$ is the local VC matrix, and $\left(g_i,h_i\right)$ are the geographical coordinates of spatial unit i. The corresponding local component scores $\mathbf{Y}\left(g_i,h_i\right)$ are:

$$\mathbf{Y}\left(g_i,h_i\right)=\mathbf{X}\mathbf{A}\left(g_i,h_i\right) \tag{10}$$

GWPCA represents a valid tool to model continuous spatial heterogeneity in PCA. The output of GWPCA consists of different loadings and component scores defined for each spatial unit. For example, if Eq. (10) identifies composite indicators, GWPCA defines completely different indices for each spatial unit as a function of distinct loadings. This may produce remarkable difficulties in the interpretation of the results.

To simplify the interpretation of the phenomena, we propose applying SA to PCA to identify groups of spatial units that share the same eigenstructure (i.e., identifying the same composite indicators). This approach recalls Postiglione et al. (2013) and the enhancements proposed by Postiglione et al. (2017) and Billè et al. (2017). Our contribution is described in the next subsection.

2.1 The proposed algorithm

The main idea of the methodology is that the appropriate treatment of spatial heterogeneity in PCA is substantially equivalent to partitioning an area into groups

of geographical zones that are not necessarily conterminous and have similar component scores. Following this, the output is not represented by different loadings at each spatial unit as for GWPCA but by distinct loadings for every group of regions identified by the SA algorithm.

SA is a stochastic relaxation algorithm that was originally introduced in statistical mechanics by Metropolis et al. (1953) and Kirkpatrick et al. (1983). SA is a random-search technique that is based on the analogy between the way in which a metal freezes into a minimum energy crystalline structure (i.e., the annealing process) and the search for a minimum in a more general system. This approach constitutes the basis of an optimisation technique for the solution of many combinatorial problems.

Geman et al. (1990) observed that a spatial combinatorial optimisation problem might be described through a Markov random field (MRF). The probability measure of an MRF by using Gibbs distribution is defined through the energy function $U(\mathbf{X}, \mathbf{k})$, which in our algorithm represents the objective function to be minimised, and a control parameter, T (see Geman and Geman 1984; Postiglione et al. 2013). $U(\mathbf{X}, \mathbf{k})$ depends on observed data $\mathbf{X}$ and the label vector $\mathbf{k} = \left(k_1, k_2, \ldots, k_i, \ldots, k_n\right)$, which categorises the heterogeneous zones, identifying clusters of regions.

$U(\mathbf{X}, \mathbf{k})$ is defined by considering two different effects: a measure of the goodness of fit of the model and a proximity constraint that describes the extent of aggregation of the spatial units. At the c-th iteration of the procedure, the energy function is defined as:

$$U(\mathbf{X}, \mathbf{k}) = \beta I_c(\mathbf{X}, \mathbf{k}) - (1 - \beta) V_c(\mathbf{k}) \tag{11}$$

In (11), I_c is the interaction term at iteration c calculated as:

$$I_c(\mathbf{X}, \mathbf{k}) = \sum_{i=1}^{m} \sum_{l=1}^{n} s_{il}^2 \tag{12}$$

where s_{il} are the entries of the residual's matrix $\mathbf{S}$ defined according to (8). The second term of (12) is a penalty constraint defined through a Potts model as:

$V_c(\mathbf{k}) = \sum_{r=1}^{n} \sum_{z=1}^{n} b_{rz} 1_{(k(l)_r = k(l)_z)}$ (13).

Specifically, b_{rz} is the element (r, z) of a binary contiguity matrix, $1_{(k(l)_r = k(l)_z)}$ is the indicator function of the event and $k(l)_r = k(l)_z$, and $(1 - \beta)$ is a parameter that discourages configurations with nonconterminous units. The parameter $(1 - \beta)$ is chosen by the researcher and models the importance of the proximity between the spatial units. We note that, unlike Postiglione et al. (2013), the two parts of the energy function (11) are balanced with complementary weights to better control the cooling process.

At the initial value of control parameter T_0, each unit i is randomly classified as $k_{i,0}$, where $k_{i,0} \in \{1, 2, \ldots, K\}$, and K is the number of clusters. This step defines the initial configuration F_0. At the $(c + 1) - th$ iteration, given a current configuration F_c, a different configuration $F_c \neq F_{c+1}$ is randomly chosen, defining a new energy function $U(F_{c+1})$ that is compared with the previous one $U(F_c)$. The old configuration F_c is substituted by the new F_{c+1} in accordance with the probability:

$$Pr_{c,c+1} = \min\left\{1, \exp\left(-\frac{U(F_{c+1}) - U(F_c)}{T_c}\right)\right\} \tag{14}$$

Probability (14) avoids local minima by defining a positive probability for the change in configuration also when the objective function $U(F)$ increases. In essence, more likely patterns (i.e., configurations with lower states of energy) are always accepted, but it is also possible to accept a poorer configuration.

Another relevant issue that should be addressed before applying our SA-PCA algorithm is the choice of the number of groups of regions to be considered. Many criteria exist to determine the optimal number of clusters (Gordon 1999). As highlighted by Harris et al. (2015), the variance levels of the components of **S** are a measure of the "*goodness of fit*" of the projected subplanes. Hence, it seems reasonable to use these in our approach to determine the optimal number of groups. Following the idea by Krzanowski and Lai (1988), we consider the pooled within-group covariance matrix $\mathbf{W}_K$ for the components of **S** calculated for any number of partitions of the dataset and, in particular, $W_K = trace(\mathbf{W}_K)$. The optimal number of groups K maximises the following function:

$$KL(K) = \left|\frac{Diff(K)}{Diff(K+1)}\right| \tag{15}$$

with:

$$Diff(K) = (K-1)^{2/p} W_{K-1} - K^{2/p} W_K \tag{16}$$

where K is the number of groups and W_{K-1} and W_K are the traces of the pooled within-group covariance matrix for the components of **S** for $K-1$ and K, respectively. This approach is used in our empirical application.

The main steps of our algorithm for considering spatial discrete heterogeneity in PCA are summarised in Appendix 1.

3 Empirical results

In this case study, we calculated a composite indicator of local services in the province of Rome to evaluate the effects of spatial heterogeneity and to assess the capacity of SA-PCA. In recent studies, there has been rising interest in composite indicators of well-being and development at the local level (Salvati and Carlucci 2014; Fusco et al. 2018). Even in Italy, a discussion on well-being under a multivariate perspective is justified by the attention given by official statistics (see, for a broader discussion, Mazziotta and Pareto 2019). Therefore, in this paper, we have calculated an indicator for targeting disparities in access to private and public services as well as for assisting policy-makers in evaluating development at the municipality level.

In the construction of the index, we have briefly considered some theoretical aspects suggested by OECD (2008) to better understand the phenomenon under investigation. Since there is no unique and generally accepted definition of a composite indicator to evaluate local governance, we have taken inspiration from recent

literature in the field of local services and well-being (D'Inverno and De Witte 2020; Tommaselli et al. 2021). Accordingly, various domains at the local level, such as economy, education, health, social care services, environment, local mobility, and public transport, have been considered.

Moreover, other variables inspired by recent literature for the case of Italy have been added. We have used the number of businesses as a relevant indicator for territorial development and potential in wealth generation (Scaccabarozzi et al. 2022). We have included the number of accommodation facilities, as they are able to increase local attractiveness and preserve cultural heritage (see, for Italy, Cracolici and Nijkamp 2009). A variable for local production of quality agricultural goods (Protected Geographical Indicator Agriculture, PGI) has been added, as promoting those businesses can support socially and environmentally sustainable development (Calcagnini and Perugini 2019; ISTAT 2022).

The province of Rome has been selected as it is one of the most populated areas in Italy but also as a region that presents a significant amount of spatial heterogeneity. In fact, due to the wide range of densely urban areas and rural or inner municipalities in the region, it seems particularly suitable to observe the effects of spatial heterogeneity on composite indicators (Salvati et al. 2019; Cartone and Postiglione 2021).

Official data from the Statistical Atlas of Municipalities[1] by the Italian National Statistical Institute (ISTAT) have been employed in the study. Specifically, we have considered the 121 municipalities ("Comuni") in the province. Depending on data availability, data were collected for 2015, except for two variables available only in the database for 2014. Hence, through this choice, we can consider a sufficient number of variables for the phenomenon and in a more recent period than 2011, the last census year. In Table 1, the various domains as well as the variables included in the indicators are reported.

As a first step, standard PCA is applied after data have been standardised to zero mean and unit variance. Complete results from PCA suggest that all first five components have eigenvalues above one and should be accounted for following Kaiser's rule (see Appendix 2a). However, for simplicity, only the first two components are more deeply explored in Table 2.

From Table 2, we observe that loadings of the first component are characterised by a higher magnitude of variables related to water, business and social welfare, all negative in sign.

By looking at the results, two issues can be raised with reference to global PCA. On one hand, the first two components account for less than half of the total variance. Hence, by considering a *whole map* estimation, the representativeness is quite low. Cartone and Postiglione (2021) noted that this situation may be linked to spatial heterogeneity, as discarding spatial instabilities may sometimes result in a poor specification of composite indicators.

Furthermore, there may be difficulties in gaining evidence from loadings of global estimation. Global PCA returns averaged weights over the whole area under

[1] https://www.istat.it/it/archivio/227189.

Table 1 Variables used for the derivation of a composite indicator of local services from the statistical atlas of italian municipalities (ISTAT)

Domain	Indicator	Name	Year	Description
Agriculture and food	Number of PGI businesses	*PGI*	2015	Number of PGI businesses on total resident population
Business	Number of local businesses	*Business*	2015	Number of total businesses on total resident population
Childcare and education	Number of kindergartens	*Kindergartens*	2014	Number of kindergartens on total resident population
Credit and Banking	Number of bank offices	*Banks*	2015	Number of bank offices on total resident population
Culture	Number of museums	*Museums*	2015	Number of museums on total resident population
Environment	Supplied water (in m^3)	*Water*	2015	M3 of supplied water on total resident population
Health	Number of beds in health care facilities	*Health*	2015	Number of beds in health care facilities on total population
Social Welfare	Social services expenses	*Welfare*	2014	Social services per user on total resident population
Tourism	Number of businesses in tourism	*Tourism*	2015	Number of businesses in tourism on total population
Public transport	Number of vehicles for public transport	*Public transport*	2015	Number of vehicles for public transport on total population

Table 2 Loadings, eigenvalues, proportion of variance, and cumulative proportion associated with the first five components of PCA on 121 municipalities in the Province of Rome

	PC 1	PC 2
PGI	−0.3490	0.2662
Business	−0.4750	−0.3829
Kindergartens	0.0688	0.1102
Banks	−0.0572	−0.4642
Museums	−0.1571	0.5936
Water	−0.4804	−0.1809
Health	0.0545	−0.3208
Welfare	−0.4790	−0.0064
Tourism	−0.3900	0.2558
Public transport	0.0672	−0.0386
Eigenvalues	1.6757	1.1771
Proportion of variance	0.2808	0.1386
Cumulative proportion	0.2808	0.4194

Fig. 1 Quantile maps of scores for global PCA. First component scores are mapped on the left **A** while second component is on the right **B** and relative Getis—Ord clusters are mapped in **C** and **D**

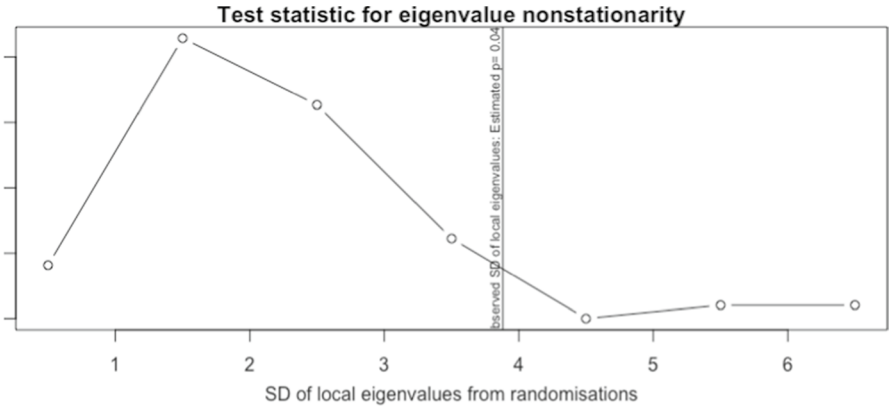


Fig. 2 Monte Carlo test for the stationarity of eigenvalues

investigation, and this simplification can preclude a detailed assessment of local services at the municipality level. Accordingly, the scores for the first two global PCs are mapped in Fig. 1A and B, while the other selected components are reported in Appendix 2b. The extent of spatial instabilities for the first two components is also evaluated by using local G statistics (Ord and Getis 1995).[2] To do so, a k-nearest contiguity matrix accounting for neighbours of the five closest municipalities is used (see Fig. 1C and D).

In fact, although some characteristics may be common to any municipality in the province, the relevance of local services may vary depending on various factors, such as history and tradition, as well as geographical features (see, for public services, Narbón-Perpiñá and De Witte 2018). Hence, to locally examine profiles of municipalities, we explicitly account for spatial heterogeneity.

One alternative in order to relax spatial homogeneity is to use GWPCA. As mentioned before, this technique allows loadings to locally change by computing the VC matrix by using a kernel function, and it can also be applied to wider datasets (Kallio et al. 2018; Trogu and Campagna 2018). As a starting point for GWPCA, we perform a Monte Carlo (MC) test to verify the significant nonstationarity of eigenvalues, as suggested by Harris et al. (2011). Figure 2 reports the results for the test, highlighting the significant nonstationarity of eigenvalues ($p\ value = 0.041$).

According to Fotheringham et al. (2002), GW technique results are relatively insensitive to the choice of the kernel, while bandwidth selection is a crucial step. For this reason, bandwidth is carefully investigated in this paper by exploring cross-validation (CV) functions for various components and bandwidth levels.

[2] The G_i statistic is a measure of local association calculated for each $i = 1, \ldots, n$, as $G_i = \frac{\sum_z b_{iz} y_z}{\sum_z y_z}$ with $z \neq i$. Here, b_{iz}s are entries of a $n \times n$ spatial weight matrix $\mathbf{B}$. Typically, the diagonal elements of $\mathbf{B}$ are zero, while for $z \neq i$, $b_{iz} = 0$ if locations i and z are not neighbours and $b_{iz} = 1$ if i and z are neighbours according to a proximity criterion.

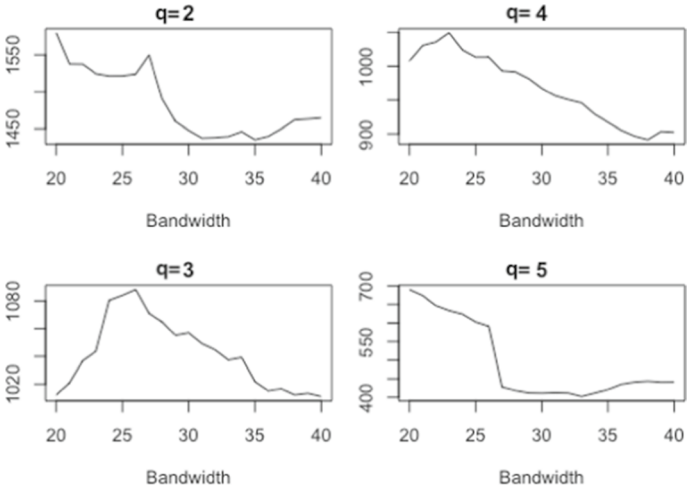


Fig. 3 Cross-validation scores for an adaptive bisquare kernel

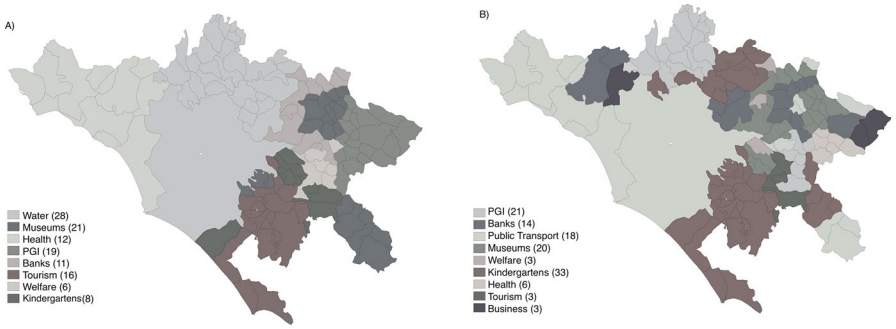


Fig. 4 Map of the winning variables for the first **A** and second **B** local component obtained with GWPCA

Again, the CV function is calculated equivalently to Harris et al. (2011) and by using the residuals for various values of the bandwidth based on an adaptive bisquare kernel. The choice of an adaptive kernel is justified by the presence of irregularities in the spatial configuration.

In Fig. 3, the CV functions calculated for each component are shown. For the number of components, $q = 2, 4$, and 5, a minimum of the CV function is obtained. However, to make results from GWPCA comparable to those of global PCA, we consider $q = 5$ and a bandwidth of 33 according to the minimum of the CV function.

An often-used method to visualise results from GWPCA is to map the winning variables, i.e., indicating the variable corresponding to loadings with higher magnitude in absolute value at each location. For the phenomenon under study, we

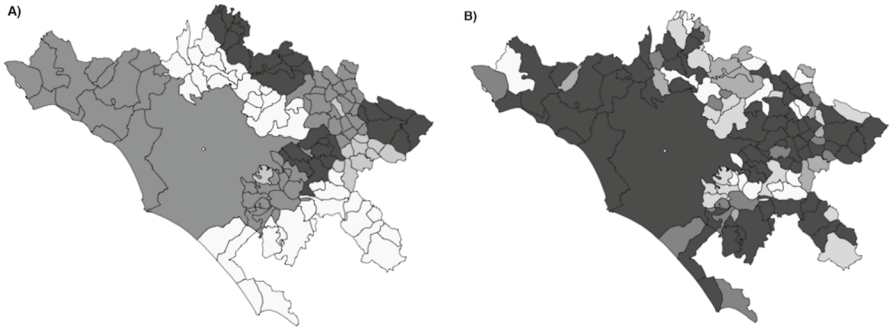


Fig. 5 Spatial configuration obtained by SA-PCA for 5 groups with different levels of $1-\beta$, (A) $1-\beta=0.45$, (B) $1-\beta=0.05$

observe that the winning variables change considerably, a result in line with the performed MC test.

In Fig. 4, the maps of the winning variables of the first two components are shown. According to the first component, water supply can be considered the winning variable in the municipality of Rome and in the northeast, while the number of businesses in tourism is the winning variable in the south. The second component is also characterised by significant differences in terms of winning variables, with the north being more influenced by public transportation and the south by childcare structures.

If maps of winning variables give evidence on the differences across the study area, they cannot be considered themselves as maps of strictly homogeneous groups. Indeed, in GWPCA, the structure of eigenvectors changes in any spatial unit by construction. Thus, using this technique, it is not directly possible to identify spatial regimes. To overcome this shortcoming, we employ our novel algorithm SA-PCA to endogenously individuate subgroups of spatial units. This methodology relaxes the assumption of homogeneity of the VC matrix, and it provides endogenous spatial regimes.

Parameters related to the cooling process have been set according to previous successful experiences in the literature (Stander and Silverman 1994; Fouskakis and Draper 2002; Postiglione et al. 2013). To avoid falling into local minima, the cooling rate should be set in the interval between 0.80 and 0.99. Hence, the level of θ is chosen to be equal to 0.95. Additionally, the level of the initial temperature T_0 is set at approximately 0.05, like Postiglione et al. (2013). After studying the behaviour of the energy function (11) for various levels of the two parameters, this combination ensures the convergence of the algorithm while preventing entrapment in local minima at the same time.

Regarding the choice of $1-\beta$, we investigate various spatial configurations for certain levels of the parameter. Some examples of these spatial configurations are reported in Fig. 5. We observed, for 5 groups, that choosing a value of $1-\beta$ lower than 0.50 (Fig. 5A) leads to very agglomerated spatial groups. In fact, increasing the penalty term forces contiguous spatial units to be part of the same group. This may

Table 3 Criterion values for different numbers of groups K obtained by SA-PCA

Number of sub-groups	W_K	$K^{2/p}W_K$	$KL(K)$
2	117.41	104.29	0.00122
3	95.22	78.27	0.81095
4	81.86	63.74	0.85971
5	75.06	56.71	0.92915
6	69.10	49.80	0.90852

result in a poorer optimisation process, and this circumstance may not be preferable in a context such as the province of Rome, which presents local pockets that are hard to capture by strictly conterminous groups. Conversely, when the contiguity constraint is set very low (Fig. 5B), the interpretability of groups decreases, as units of the same cluster tend to be scattered. Hence, in this application, our choice consists of a slight level of 0.20 for $1-\beta$, which reveals only partially scattered groups and good adaptation to the phenomenon under investigation.

Another relevant choice to be addressed before applying SA-PCA is the number of groups. Although in composite indicators the choice of groups can often be linked to specific research aims (Nardo et al. 2005), certain selections may lead to suboptimal solutions. Hence, this choice must be properly considered by researchers to obtain more reliable and interpretable configurations.

Nevertheless, the proposed SA-PCA algorithm is studied to trim the objective function by a constraint that discourages nonconterminous groups. As mentioned before, this feature encourages more conterminous spatial configurations that tend to be more robust to outliers in the geographical area under study (Benedetti et al. 2013). Additionally, robustness to spatial outliers obtained by using constraints may help the stability of the grouping procedures (García-Escudero et al. 2010). In addition to being less sensitive to outliers, groups obtained by SA-PCA preserve a certain amount of proximity between units despite the number of groups allowed, which can help interpret underlying spatial features over the study area.

In the application, we first obtain partitions as solutions of the SA algorithm and then calculate the $KL(K)$ function for various levels of K. The results for the criterion are summarised in Table 3. By maximising the KL statistic on the residuals obtained by the solution to the optimisation problem, we find that a level of $K = 5$ is the most convenient to apply SA-PCA.

Figure 6 presents the map of the clusters of units obtained when using our SA-PCA algorithm. Group 1 includes the City of Rome and neighbours' municipalities mainly in the centre of the province. A sizeable number of southern municipalities are in Group 2. Group 4 is largely composed of areas located in the west of the province, while Group 3 consists of two smaller pockets in the east of the province. The rest of the eastern municipalities are mostly included in Group 5.

In terms of local characteristics, Group 1 includes densely populated municipalities in the centre of the province, while Group 2 is comprised of coastal municipalities largely linked to the city of Rome (such as *Anzio*, *Nettuno*, and *Velletri*). Group

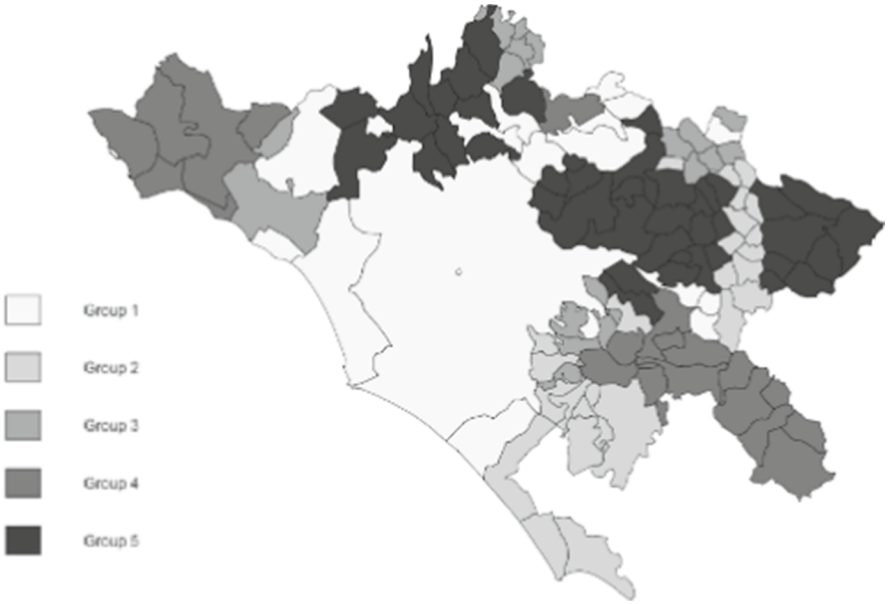


Fig. 6 Groups of spatial units obtained by SA-PCA

Table 4 Loadings of the first component for each group obtained with SA-PCA

Variables	Group 1	Group 2	Group 3	Group 4	Group 5
PGI	0.2440	0.7270	−0.0291	−0.0618	0.4664
Business	0.5145	0.7856	0.1849	−0.1017	−0.2732
Kindergartens	−0.1870	0.0005	−0.1964	−0.2130	−0.0003
Banks	−0.0929	0.2332	0.2367	0.1569	−0.6164
Museums	−0.0091	0.1565	−0.1366	−0.5747	0.4712
Water	0.6457	−0.0227	0.0209	−0.0880	0.0044
Health	−0.4337	−0.2751	0.8975	0.0112	−0.0078
Welfare	0.4084	0.4064	−0.0862	−0.0825	−0.1315
Tourism	0.2581	0.3671	−0.0313	−0.7555	0.1981
Public transport	−0.0521	−0.1013	0.1909	0.0071	0.2220
POV of first component	0.7072	0.4030	0.3664	0.3669	0.3035

3 consists of small and scarcely populated municipalities in the mountain area in the east (e.g., *Arsoli*, *Subiaco*, and *Vallinfreda*). Last, Group 4 includes mainly rural municipalities on the hills at the borders, especially those in the south, known as *Castelli,* and Group 5 includes a variety of larger municipalities often considered to be Roman suburbs (e.g., *Guidonia Montecelio* and *Tivoli*).

As expected, the five clusters show substantial differences in the underlying phenomenon. The compositions of eigenvectors for the first and second components are shown in Tables 4 and 5, respectively. From the results, we see that

Table 5 Loadings of the second component for each group obtained with SA-PCA

Variables	Group 1	Group 2	Group 3	Group 4	Group 5
PGI	0.0108	0.7270	−0.1320	0.1548	−0.5041
Business	0.0325	−0.2605	−0.2164	0.1521	0.0702
Kindergartens	0.9357	0.0002	0.3769	−0.1106	0.0002
Banks	0.2890	−0.1854	−0.4049	0.4010	0.1890
Museums	−0.0269	−0.0967	−0.3593	0.1387	0.7259
Water	0.1078	0.0147	−0.0465	−0.0489	0.0060
Health	0.0001	−0.8941	0.2626	0.0014	0.0029
Welfare	0.0399	−0.1500	−0.2788	0.4274	0.3241
Tourism	0.1412	0.1585	−0.1936	−0.0722	0.0664
Public transport	0.0771	−0.1525	−0.5606	−0.7552	0.2623
Cumulated POV of second component	0.8506	0.5879	0.6175	0.5893	0.5263

considering spatial heterogeneity leads to differences resulting from local instabilities in the loadings. From Table 4, those differences among the five different groups can be further appreciated.

In Group 1, *business* and *water* are the major drivers together with the presence of welfare aids. The value of *tourism* is also consistent with the presence of tourists that come yearly across the city of Rome and Fiumicino airport. For Group 2, *business* is important and positively correlated with the phenomenon under investigation. Additionally, *PGI* and *welfare* are positively connected to the phenomenon. For Group 3, *health, banks,* and *public transport* represent the most relevant loadings, contributing to an increase in the first component scores. For Group 4, the level of local service in those municipalities is mainly related to *banks*. In Group 5, the increasing presence of *museum* and *PGI* mainly contribute to the positive development of local services.

Table 5 indicates the loadings for the second components of the five groups generated by SA-PCA. Once again, the evidence emphasises structural heterogeneity suggested by diverse values of eigenvectors. For instance, the second component of Group 1 indicates the need for facilities to take care of children at an early stage (high magnitude of *kindergartens*).

In Fig. 7, performances for global PCA (dashes), cluster-specific PCA (dots), GWPCA (dashes and dots) and SA-PCA (line) for the first five components are reported in terms of average proportion of variance accounted, while the plotted points indicate unit-by-unit fit for GWPCA. In the application, it also appears that SA-PCA helps to specify components according to structural differences in the VC matrix. This last feature is justified by a major increase in the proportion of variance accounted when compared to that of global PCA. Even when compared to that of GWPCA, representativeness increases by approximately five percent in terms of the average proportion of variance explained. Finally, our SA-PCA presents higher performance also when compared with that of cluster-specific PCA,

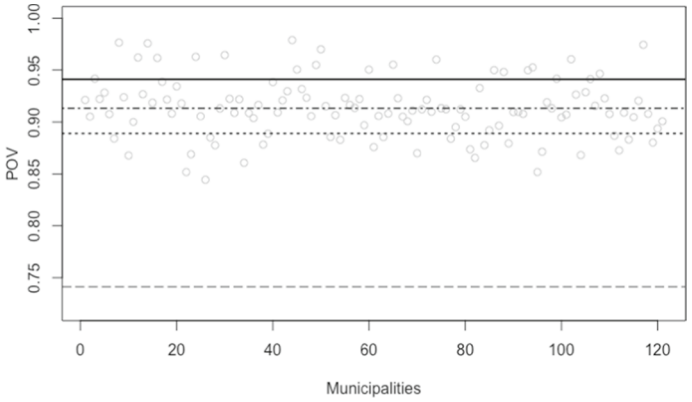


Fig. 7 Performance (average proportion of variance of the first five components) of global PCA (dashes), cluster-specific PCA (dots), GWPCA (dashes and dots), and SA-PCA (line). Grey circles show the by-unit GWPCA performance of 121 municipalities of Rome

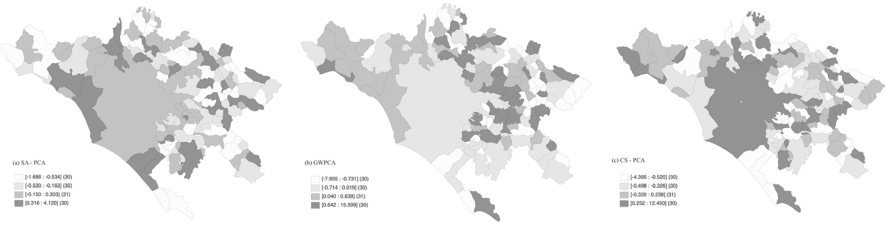


Fig. 8 First component composite indicator for SA-PCA **a**, GWPCA **b**, and cluster-specific PCA **c**

calculated by using the *k*-means algorithm in the first step and standard PCA, for each cluster, in the second step.

In Fig. 8, the first component indicators are shown for SA-PCA (a), GWPCA (b), and cluster-specific PCA (c) to highlight some differences between the various definitions. SA-PCA reports higher levels for the composite indicator mainly for municipalities on the coast and in the northern part of the province. GWPCA shows higher levels for large municipalities in the immediate surroundings in eastern Rome. Moreover, cluster-specific PCA returns higher achievements mainly in Rome, at the northeast, and on the southeast side. Those differences may be connected to the diverse treatment of spatial issues. In SA-PCA, for instance, consideration of discrete heterogeneity slightly shifts the performance of Rome compared to when using cluster-specific PCA, where spatial effects are not explicitly considered. Conversely, this feature seems to favour seaside towns connected to the main city.

Finally, in addition to improving fit performance, the empirical application shows that SA-PCA can directly address nonstationarity by identifying endogenous clusters of units. Therefore, the components (i.e., indicators) obtained from SA-PCA

are perfectly comparable within groups, as they are directly obtained by loadings from the same spatial cluster. This feature may be of help for regional policy-makers in two ways. On the one hand, instead of relying on unit level indicators as in GWPCA, SA-PCA can define a limited number of clusters to model discrete heterogeneity, which is in many cases a reasonable assumption (Anselin 1988, 2010). Moreover, being easily interpretable, indicators from SA-PCA allow for local policies to be more accurately set, providing better support multilevel governance.

4 Concluding remarks

Spatial heterogeneity is a relevant issue when standard statistical techniques are applied. In this paper, this spatial effect has been investigated deeper to add to the previous literature in the field of PCA. To this end, we developed a novel methodology to consider heterogeneity in the form of various spatial groups. SA-PCA offers a novel approach to tackle unobserved spatial heterogeneity by using a spatially constrained algorithm.

As seen above, while GWPCA considers continuous nonstationarity, SA-PCA relaxes the hypothesis of spatial homogeneity by providing groups as a solution to a combinatorial problem. By identifying spatial clusters of units, SA-PCA allows us to individuate loadings and scores for units that share the same structure of the VC matrix in the same group.

In this paper, we benchmark this new methodology against other options while defining composite indicators. SA-PCA, GWPCA and cluster-specific PCA were applied to calculate a composite indicator of deprivation in 121 municipalities in the province of Rome. Here, SA-PCA generally had better representativeness than that of GWPCA and cluster-specific PCA in terms of the average proportion of variance explained. In this sense, SA-PCA offers a plausible alternative in the case of spatial heterogeneity. Additionally, SA-PCA allows for a more straightforward interpretation than does GWPCA because SA-PCA utilizes a limited number of different loadings and score sets.

Last, we note that the algorithm shows some limitations when applied to very large datasets, as a wider number of observations and variables can add to the computational burden. In future studies, the application to large datasets can be deepened by considering alternative algorithms extended to PCA for the same purpose.

Appendix 1—Modified SA algorithm for spatial heterogeneity in PCA (SA-PCA)

Main steps:

(0) A number of groups K is chosen.
(1) The initial control parameter T_0 is specified and each geographical unit i, for $i = 1, 2, ..., n$, is randomly classified as $k_{i,0}$, where $k_{i,0} \in \{1, 2, ..., , K\}$ is the label of the assigned cluster.

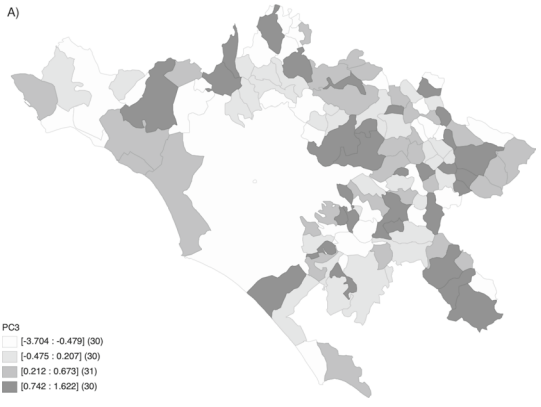
(2) For each partition $\in \{1, 2, ..., , K\}$ a principal component analysis is computed, and the number of components is chosen. Residuals are calculated.
(3) At $(c + 1) - th$ iteration for each geographical unit $i = 1, 2, ..., n$, a new candidate label $k_{i,c+1} \in \{1, 2, \dots , k\} \backslash \{k_{i,c}\}$ is randomly selected. The energy function $U(\dots , k_{i,c+1}, \dots)$ (i.e., the objective function) is computed and compared with the current energy $U(\dots , k_{i,c}, \dots)$ function. If $U(\dots , k_{i,c+1}, \dots) < U(\dots , k_{i,c}, \dots)$, the label $k_{i,c}$ is replaced with $k_{i,c+1}$. Otherwise, the label $k_{i,c}$ is replaced by $k_{i,c+1}$ with probability $p = \exp(-[U(\dots , k_{i,c+1}, \dots) - U(\dots , k_{i,c}, \dots)]/T_c)$. Update the control parameter according to the schedule $T_c = T_0 \cdot \theta^{c-1}$.
(4) Consider $\mathbf{k}_c = (k_{1,c}, k_{2,c}, ..., k_{i,c}, ..., k_{n,c})$ as the label vector at the end of the $c - th$ iteration. The algorithm will stop if $\mathbf{k}_{c+1} \equiv \mathbf{k}_c$ holds true.
(5) Principal component analysis is calculated for each of the cluster of regions obtained.

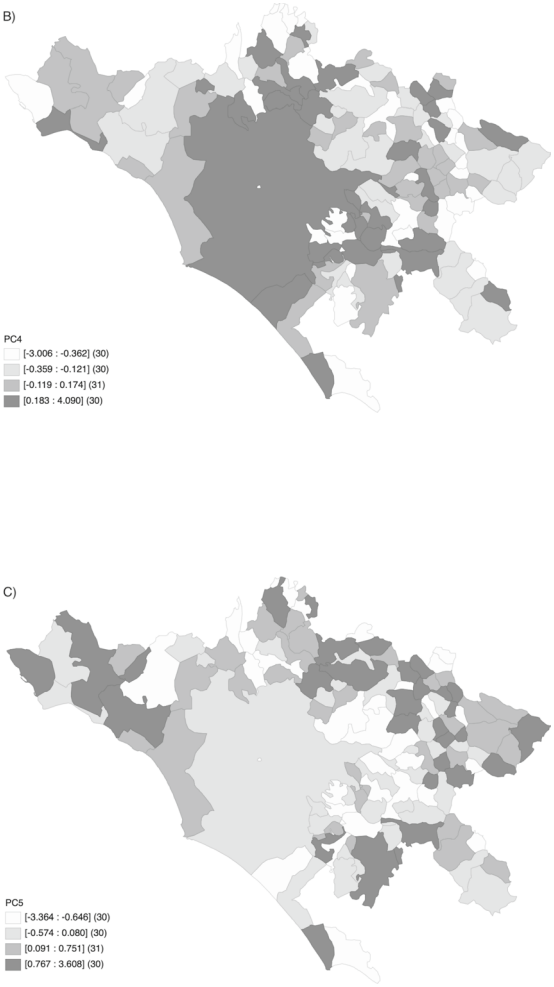
Appendix 2

(a) Table of eigenvalues, proportion of variance, and cumulative proportion of variance for first five components for PCA.

	PC 1	PC 2	PC 3	PC 4	PC 5
Eigenvalues	1.6757	1.1771	1.0571	1.0463	1.0036
POV	0.2808	0.1386	0.1117	0.1095	0.1007
Cumulated POV	0.2808	0.4194	0.5311	0.6406	0.7413

(b) Maps of scores for third (A), fourth (B), and fifth (C) component for PCA.





Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti Pescara within the CRUI-CARE Agreement. Research contracts on innovation and green topics" FSE-REACT-EU Project by the European Commission—National Operational Program "PON Ricerca & Innovazione 2014–2020—DM 1062/2021 (D25F21001470007) Alfredo Cartone.

References

Andreano MS, Benedetti R, Postiglione P (2017) Spatial regimes in regional European growth: an iterated spatially weighted regression approach. Qual Quant 51:2665–2684
Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht
Anselin L (2010) Thirty years of spatial econometrics. Pap Reg Sci 89:3–25
Benedetti R, Pratesi M, Salvati N (2013) Local stationarity in small area estimation models. Stat Method Appl 22:81–95
Billé AG, Benedetti R, Postiglione P (2017) A two-step approach to account for unobserved spatial heterogeneity. Spat Econ Anal 12:452–471
Calcagnini G, Perugini F (2019) Social capital and well-being in the Italian provinces. Socio Econ Plan Sci 68:100668
Cartone A, Postiglione P (2021) Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators. Spat Econ Anal 16:126–147
Cracolici MF, Nijkamp P (2009) The attractiveness and competitiveness of tourist destinations: a study of Southern Italian regions. Tour Manag 30:336–344
D'Inverno G, De Witte K (2020) Service level provision in municipalities: a flexible directional distance composite indicator. Eur J Oper Res 286:1129–1141
De Muro P, Mazziotta M, Pareto A (2011) Composite indices of development and poverty: an application to MDGs. Soc Indic Res 104:1–18
Decancq K, Lugo MA (2013) Weights in multidimensional indices of wellbeing: an overview. Econ Rev 32:7–34
Demšar U, Harris P, Brunsdon C, Fotheringham AS, McLoone S (2013) Principal component analysis on spatial data: an overview. Ann Assoc Am Geogr 103:106–128
Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression–the analysis of spatially varying relationships. Wiley, Chichester
Fouskakis D, Draper D (2002) Stochastic optimization: a review. Int Stat Rev 70:315–349
Fusco E, Vidoli F, Sahoo BK (2018) Spatial heterogeneity in composite indicator: a methodological proposal. Omega 77:1–14
Gabriel KR (1978) Least squares approximation of matrices by additive and multiplicative models. J Roy Stat Soc Ser B 40:186–196
García-Escudero LA, Gordaliza A, Matrán C, Mayo-Iscar A (2010) A review of robust clustering methods. Adv Data Anal Classif 4:89–109
Geman S, Geman D (1984) Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6:721–741
Geman D, Geman S, Graffigne C, Dong P (1990) Boundary detection by constrained optimization. IEEE Trans Pattern Anal Mach Intell 12:609–628
Giacalone M, Mattera R, Nissi E (2022) Well-being analysis of Italian provinces with spatial principal components. Socio Econ Plan Sci 84:101377
Gordon A (1999) Classification, 2nd edn. Chapman and Hall/CRC Press, London
Harris P, Brunsdon C, Charlton M (2011) Geographically weighted principal components analysis. Int J Geogr Inf Sci 25:1717–1736
Harris P, Clarke A, Juggins S, Brunsdon C, Charlton M (2015) Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. Geogr Anal 47:146–172
Havard S, Deguen S, Bodin J, Louis K, Laurent O, Bard D (2008) A small-area index of socioeconomic deprivation to capture health inequalities in France. Soc Sci Med 67:2007–2016
Householder AS, Young G (1938) Matrix approximation and latent roots. Am Math Mon 45:165–171
ISTAT (2022) Rapporto BES 2021: il benessere equo e sostenibile in Italia

Jolliffe IT (2002) Principal component analysis. Springer, Berlin
Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Hered 101:92–103
Kallio M, Guillaume JH, Kummu M, Virrantaus K (2018) Spatial variation in seasonal water poverty index for Laos: an application of geographically weighted principal component analysis. Soc Ind Res 140:1131–1157
Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Sci 220:671–680
Krzanowski WJ, Lai YT (1988) A criterion for determining the number of groups in a data set using sum-of-squares clustering. Biom 44:23–34
LeSage J, Pace RK (2009) Introduction to spatial econometrics. Chapman and Hall/CRC, London
Libório MP, Martinuci ODS, Machado AMC, Lyrio RDM, Bernardes P (2022) Time-space analysis of multidimensional phenomena: a composite indicator of social exclusion through k-Means. Soc Ind Res 159:569–591
Lloyd CD (2010) Exploring population spatial concentrations in Northern Ireland by community background and other characteristics: an application of geographically weighted spatial statistics. Int J Geogr Inf Sci 24:1193–1221
Mazziotta M, Pareto A (2019) Use and misuse of PCA for measuring well-being. Soc Ind Res 142:451–476
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092
Moran PA (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23
Murakami D, Griffith DA (2019) Spatially varying coefficient modeling for large datasets: eliminating N from spatial regressions. Spat Stat 30:39–64
Narbón-Perpiñá I, De Witte K (2018) Local governments' efficiency: a systematic literature review—part I. Int Trans Oper Res 25:431–468
Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for composite indicators building. EUR 21682 EN, JRC31473
OECD (2008) Handbook on constructing composite indicators: methodology and user guide
Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geog Anal 27:286–306
Pampalon R, Raymond G (2000) A deprivation index for health and welfare planning in Quebec. Chronic Dis Can 21:104–113
Postiglione P, Andreano MS, Benedetti R (2013) Using constrained optimization for the identification of convergence clubs. Comput Econ 42:151–174
Postiglione P, Andreano MS, Benedetti R (2017) Spatial clusters in EU productivity growth. Growth Chang 48:40–60
Salvati L, Carlucci M (2014) A composite index of sustainable development at the local scale: Italy as a case study. Ecol Ind 43:162–171
Salvati L, Ciommi MT, Serra P, Chelli FM (2019) Exploring the spatial structure of housing prices under economic expansion and stagnation: the role of socio-demographic factors in metropolitan Rome, Italy. Land Use Policy 81:143–152
Sarra A, Nissi E (2020) A spatial composite indicator for human and ecosystem well-being in the Italian urban areas. Soc Ind Res 148:353–377
Scaccabarozzi A, Mazziotta M, Bianchi A (2022) Measuring competitiveness: a composite indicator for Italian municipalities. Soc Ind Res. https://doi.org/10.1007/s11205-022-02990-x
Stander J, Silverman BW (1994) Temperature schedules for simulated annealing. Stat Comput 4:21–32
Tomaselli V, Fordellone M, Vichi M (2021) Building well-being composite indicator for micro-territorial areas through PLS-SEM and K-means approach. Soc Ind Res 153:407–429
Trogu D, Campagna M (2018) Towards spatial composite indicators: a case study on sardinian landscape. Sustainability 10(5):1369
Wartenberg D (1985) Multivariate spatial correlation: a method for exploratory geographical analysis. Geogr Anal 17:263–283
Wheeler DC, Calder CA (2007) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. J Geogr Syst 9:145–166