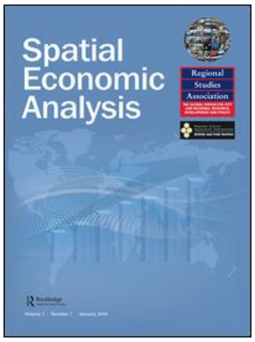

Spatial Economic Analysis

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rsea20

Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators

Alfredo Cartone & Paolo Postiglione

To cite this article: Alfredo Cartone & Paolo Postiglione (2020): Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators, Spatial Economic Analysis, DOI: 10.1080/17421772.2020.1775876

To link to this article: https://doi.org/10.1080/17421772.2020.1775876

Published online: 26 Jun 2020.

SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2020.1775876

Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators

Alfredo Cartone[a] and Paolo Postiglione[b]

ABSTRACT
This paper investigates the role of spatial dependence, spatial heterogeneity and spatial scale in principal component analysis for geographically distributed data. It considers spatial heterogeneity by adopting geographically weighted principal component analysis at a fine spatial resolution. Moreover, it focuses on dependence by introducing a novel approach based on spatial filtering. These methods are applied in order to derive a composite indicator of socioeconomic deprivation in the Italian province of Rome while considering two spatial scales: municipalities and localities. The results show that considering spatial information uncovers a range of issues, including neighbourhood effects, which are useful in order to improve local policies.

KEYWORDS
spatial filtering, spatial dependence, geographically weighted principal component analysis (GWPCA), modifiable areal unit problem (MAUP), deprivation index

JEL C1, C21, C43
HISTORY Received 9 July 2019; in revised form 19 May 2020

INTRODUCTION

Nowadays, the increasing availability of large data sets and Big Data is shining the spotlight onto multivariate statistical techniques. To simplify the interpretation of this large amount of data, many researchers try to describe complex phenomena by combining sets of different variables. Synthesizing multivariate phenomena is indicated as a proper way to represent economic, social and environmental problems in the world (OECD, 1993).

Particular attention is given to the derivation of composite indicators in a policy perspective because of the idea that those indicators can be used to rank geographical units (Kuc-Czarnecka et al., 2020; Nardo et al., 2005). Ideally, a composite indicator should be founded on a theoretical framework, which permits single indicators to be selected.

Principal component analysis (PCA) (Jolliffe, 2002) has often been adopted to obtain composite indicators. Demšar et al. (2013) point out that one of the future challenges for this technique will be the explicit consideration of spatial effects (Anselin, 1988). From a methodological point of

CONTACT
[a] alfredo.cartone@unich.it Department of Economic Studies, 'G. d'Annunzio' University of Chieti-Pescara, Pescara, Italy.
[b] **(Corresponding author)** postigli@unich.it
Department of Economic Studies, 'G. d'Annunzio' University of Chieti-Pescara, Pescara, Italy.
Supplemental data for this article can be accessed at https://doi.org/10.1080/17421772.2020.1775876

view, discarding spatial effects from the analysis may lead to misspecification and inappropriate interpretation of the phenomenon under investigation (Jombart et al., 2008). Moreover, taking into account the effects of heterogeneity and spatial scale makes possible the definition of more accurate local actions (Rodríguez-Pose, 2018).

The present paper directly addresses the issues connected to the spatial nature of the data in the definition of a composite indicator using PCA. We introduce a novel approach to PCA for geographically distributed data that is based on the use of spatial filtering (Getis & Griffith, 2002). This innovative methodology is discussed in detail in the third section, with clear reference to different filtering approaches.

In order to evaluate the influence of spatial effects and spatial scale in the definition of composite measures, we derive a deprivation index for local territories in the province of Rome, the capital city of Italy. Deprivation may be defined 'as a state of observable and demonstrable disadvantage relative to the local community or the wider society or nation to which an individual, family or group belongs' (Townsend, 1987, p. 125), and it can be calculated on a refined scale to return information on disparities affecting each area (Caranci et al., 2010). Local deprivation is selected here as an appealing example to show that only an analysis based on the consideration of the spatial effects can lead to an effective examination of the multivariate phenomenon under investigation. This also adds value to the existing literature studying the effects of space in measures for disparities (Márquez et al., 2019; Panzera & Postiglione, 2020), but in a multivariate scenario.

The paper is structured as follows. The next section overviews the problems related to the use of PCA to define composite indicators for spatial data. Sections three summarizes the methodologies adopted in the paper to investigate the spatial effects in composite indicators. The fourth section reports the results of an application in the province of Rome performed at two levels of spatial resolution: municipalities and localities. Finally, the fifth section concludes.

COMPOSITE INDICATORS AND PRINCIPAL COMPONENT ANALYSIS FOR GEOGRAPHICAL DATA

A composite indicator evaluates multidimensional concepts that cannot be captured by a single indicator. Different methods can be used to calculate a composite indicator. For an overview of these techniques, see Decancq and Lugo (2013).

PCA is often used to define composite indicators (e.g., Greyling & Tregenna, 2016; McGillivray, 2005). In this case, the loadings represent the weights of the composite measure. De Muro et al. (2011) criticize the use of PCA for building composite indicators because weights are based on a 'pure' statistical technique and may not reflect the relevance of single variables in the underlying phenomenon. However, weights from PCA may be less 'subjective' because these are data driven and not assigned by the researcher, in contrast to the case of the 'normative' weights (Bellandi & Ruiz-Fuensanta, 2010).

Despite its popularity in geographical data applications, an aspect that has been only recently studied with PCA is related to the presence of spatial effects. For example, ad hoc analyses for units located in different areas should be considered (Lloyd, 2010). Moreover, it is necessary to assess the influence of neighbours on component scores. For these reasons, the spatial effects should be further explored, relaxing the assumptions of homogeneity and independence at different spatial scales of analyses.

Three different types of spatial information should be considered in the definition of composite indicators when using PCA: heterogeneity, dependence and scale of analysis.

Spatial heterogeneity may cause the misspecification of the model, leading to problems in the interpretation of the results. This issue has been considered in linear regression models only in recent years (Postiglione et al., 2013). The geographically weighted approach (Fotheringham et al., 2002) is extended to PCA, as geographically weighted PCA (GWPCA), to consider the presence of spatial heterogeneity (Harris et al., 2011).

The problem of spatial dependence has been faced in factorial analyses using multivariate kriging (Wackernagel, 2003). However, multivariate factorial kriging represent a model-based approach that is different from PCA.

The possible presence of spatial dependence cannot be ignored in analyses when using PCA for geographically distributed data.

Spatial autocorrelation in PCA has been discussed by Jombart et al. (2008) through an approach called spatial PCA (i.e., sPCA). This method is used to investigate spatial patterns by an objective function obtained combining Moran's I with standard PCA decomposition of the variance matrix. This function is highly positive when the variable has a large variance and positive spatial autocorrelation; conversely, it is largely negative when the variable has a high variance and shows negative spatial autocorrelation.

This tool may be considered an advancement of multivariate spatial techniques (Wartenberg, 1985). By using sPCA, the positive definite structure of the variance–covariance matrix is modified, and Wartenberg (1985) points out that techniques based on Moran's I matrix decomposition may lead to negative eigenvalues. Eigenvectors related to positive eigenvalues describe states of positive autocorrelation, while eigenvectors associated with negative eigenvalues explain cases of negative autocorrelation. Therefore, the main aim of sPCA is to help the researcher to define patterns in the spatial structure of the phenomenon.

Another important issue in geographical studies is the choice of an appropriate spatial scale for the analysis (Fernández-Vazquez & Rubiera-Morollòn, 2012). In many cases, empirical applications are performed at a coarse spatial resolution. While analysing different spatial levels, the effects of the modifiable areal unit problem (MAUP) (Fotheringham & Wong, 1991; Openshaw & Taylor, 1979) on statistical analyses can be severe (for the case of spatial econometric models, see Arbia & Petrarca, 2011). Hence, the effect of the MAUP can bring puzzling statistical results, and it often leads to difficulties in interpreting the phenomenon under investigation (Purtauf et al., 2005).

Since these features are commonly ignored in the use of PCA, in this paper we contribute to the scientific debate about the importance of using spatial information when PCA is applied to geographically distributed data, particularly in the derivation of composite indicators.

The first step of this study is based on reconsidering the use of GWPCA in spatial composite indicators (Harris et al., 2015). This tool is widely used as a spatial approach to PCA (Saib et al., 2015). However, this method explicitly addresses only spatial heterogeneity (in the sense of non-stationarity), neglecting the role of spatial dependence.

In order to analyse and display spatial dependence, we introduce a new methodology based on spatial filtering techniques (Getis & Griffith, 2002). All variables are split into a spatial component and in an idiosyncratic (i.e., filtered) component by spatial filtering. The spatial components are multiplied for the loading values obtained by overall PCA to calculate the spatial scores linked to neighbourhood effects.

Finally, the problem of the spatial scale is effectively considered by analysing the phenomenon under investigation at two different spatial scales to address the potential effects of the MAUP at a very fine spatial resolution (Bonneu & Thomas-Agnan, 2015). The selection of the spatial scale can produce major changes in the components scores as well in the loadings. Our analysis highlights the need to consider refined scale (i.e., finer than NUTS 3), when data are available, in order to allow a deeper knowledge of local peculiarities and to identify ad hoc policies.

Note that our approach for handling spatial dependence in PCA differs in many aspects from sPCA. First, our methodology, in contrast to that proposed by Jombart et al. (2008), allows direct comparisons with standard PCA useful for the purpose of ranking different spatial units in an intuitive way for regional scientists. In fact, eigenvalues from sPCA may be positive as well negative. Therefore, this may bring the practitioners to face problems in the selection of the components under investigation. In PCA, those problems have not to be faced thanks to the structure of the variance–covariance matrix which leads to a descending representativeness of components in terms of accounted variance (Jolliffe, 2002). Another problem is related to the definition of the composite indicator itself. In fact, if sPCA could perform well in an explorative sense, this technique does not allow one to split the spatial and non-spatial part from the global indicator. sPCA individuates spatial indicators that cannot be easily compared with the overall indicator as these rely on a different system of eigenvectors (i.e., weights) from the overall PCA. Lastly, sPCA does not decompose the total variance into decreasing additive components as PCA.

One of the main purposes of this paper is to highlight the importance of spatial component in the definition of composite indicators obtained through PCA. In particular, we aim to assess how much of the composite indicator obtained is due to the effects of spatial contagion between geographical units. For these reasons, to manage spatial dependence in PCA, we choose a new approach based on spatial filtering. This methodology may complement an analysis based on the use of GWPCA, and offers a new tool to evaluate the presence of spatial dependence in composite indicators.

METHODOLOGY

PCA is based on the analysis of a data matrix X_{ij}, where $i = 1, \ldots, n$ denotes the statistical units and $j = 1, \ldots, p$ denotes the variables. The variance–covariance matrix Σ may be decomposed into its eigenstructure as (Jolliffe, 2002):

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

where Λ is the diagonal matrix of eigenvalues; A is the corresponding matrix of loadings (i.e., the eigenvectors); Σ is the variance–covariance matrix; and the superscript t indicates the transpose. The eigenvalues in Λ represent the variance of the principal component Y_r defined as:

$$Y_r = XA_r \tag{2}$$

where A_r is the r-th column of the loading matrix A of Σ and represents the contribution of each variable in X to the r-th principal component Y_r. The entries of Y_r are defined as scores and represent our composite indicator. In PCA not all principal components are usually considered. Hence, the essence of data reduction is to select a certain number of components that account for the largest part of the variance.

As pointed out by (1) and (2), the composite indicator is synthesized according to the loadings of the variance–covariance matrix Σ. It follows that the relevance of each phenomenon is based on the structure of the whole-map variance–covariance matrix. Standard PCA is not appropriate in order to analyse geographically distributed data. Considering the presence of spatial heterogeneity, we can use GWPCA, as described below.

Reappraisal of geographically weighted principal component analysis (GWPCA)

GWPCA (Fotheringham et al., 2002; Harris et al., 2011; Lloyd, 2010) may be considered as a valid alternative for localized sets of weights used in the building of composite indicators, considering spatial heterogeneity. It allows non-stationarity of eigenvalues and eigenvectors across spatial units. In GWPCA, principal components are calculated using a local variance–covariance matrix. Given the (u_i, z_i) coordinates for each location i, the local variance–covariance matrix $\Sigma(u_i, z_i)$ is:

$$\Sigma(u_i, z_i) = X^t W(u_i, z_i)X \tag{3}$$

where $W(u_i, z_i)$ is a diagonal matrix of spatial weights that depends on the number of observations considered as a neighbour of a particular unit (u_i, z_i) and are calculated according to a convenient kernel function. Equation (1) can be generalized for the local case as:

$$\Sigma(u_i, z_i) = A(u_i, z_i)\Lambda(u_i, z_i)A(u_i, z_i)^t \quad (4)$$

where $A(u_i, z_i)$ is the matrix of the local loadings; and $\Lambda(u_i, z_i)$ is the matrix of the local eigenvalues.

In this paper, in order to calculate geographical weights, we use a Gaussian kernel:

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{\gamma}\right)^2\right) \quad (5)$$

where w_{ij} is the entry of the matrix $W(u_i, z_i)$; d_{ij} is the geographical distance between the units; and γ is the bandwidth. Generally, in geographically weighted approaches, the results are robust to the choice of the kernel. Nevertheless, great care should be given to bandwidth selection (Fotheringham et al., 2002). To this end, we use a cross-validation (CV) function built considering the residuals from the data (Harris et al., 2011). For further details about the definition of the goodness-of-fit (GOF) measure used in the CV function, see Harris et al. (2015).

The interpretation of GWPCA results is challenging due to the very large volume of information produced. Harris et al. (2011, 2015) suggest some possible visualization methods to support the interpretation: mapping the proportion of variance explained by the local PCs, mapping the *winning variable* of each local PC, and using multivariate glyphs to represent the loadings at all locations.

In equation (2), it is evident that standard PCA defines composite indicators using loadings as global weights (i.e., defined following the same rule for all spatial units). To consider the peculiarity of the spatial information inherent in geographically distributed data, we may use a weighting scheme that changes locally (i.e., different loadings for each spatial unit) derived through GWPCA.

Following this idea, we define a first principal component composite local indicator, for unit i, as:

$$y_{1i} = X_i^t A_1(u_i, z_i) \quad (6)$$

where X_i represents the vector of p variables measured at location i; and $A_1(u_i, z_i)$ is the first column of the local loading matrix $A(u_i, z_i)$. Our approach to spatial composite indicators with GWPCA is similar to that of Trogu and Campagna (2018) and Kallio et al. (2018).

Unfortunately, GWPCA only considers the influence of spatial heterogeneity. To evaluate the extent of the spatial dependence in composite indicators for geographically distributed data, we need to move a step ahead by proposing a novel methodology based on spatial filtering.

Novel approach for screening the spatial dependence in PCA

In this section we define a measure to account for neighbourhood effects in component score. In standard PCA, the score for the first principal component (i.e., the first composite indicator), for a location i, is defined as:

$$y_{1i} = X_i^t A_1 \quad (7)$$

where X_i represents the vector of the p variables measured at location i; and A_1 is the first column of the loading matrix A. The variables included in X may be characterized by a certain degree of spatial autocorrelation, which may be different from a variable to another.

To capture the spatial component of each variable, it seems appropriate to use spatial filtering techniques (Getis & Griffith, 2002). Among others, two different approaches can be developed for screening the spatial autocorrelation by using spatial filtering. The first approach, introduced by Griffith (2000, 2003), exploits an eigenvector decomposition associated with Moran's I to remove nuisances due to spatial dependence. The second approach is based on a multistep procedure based on the Getis' statistic G_i (Getis, 1990, 1995).

The Griffith's filtering is based on the eigenvectors decomposition of the matrix:

$$\left(I - \frac{11^t}{n}\right)C\left(I - \frac{11^t}{n}\right)$$

where I is an $n \times n$ identity matrix; 1 is $n \times 1$ vector of ones; and C is a binary spatial weight matrix (Griffith, 2008; Griffith & Chun, 2014). Using this methodology, the spatial part of the variable is identified in order to minimize the degree of spatial autocorrelation in the model residuals. The spatial autocorrelation model for the generic variable X[1] is defined as:

$$x_i = \mu_X + V_i^t \beta_V + \epsilon_i \tag{8}$$

where x_i is the variable at each location i; μ_X is the mean of the variable X; V_is are the values associated to unit i included into the matrix of the $q < n$ selected eigenvectors V of the matrix $\left(I - 11^t/n\right)C\left(I - 11^t/n\right)$; β_V is the set of parameters associated to the eigenvectors; and ε_i is an independent and identically distributed error term. Tiefelsdorf and Boots (1995) demonstrates that each of the eigenvalues of $\left(I - 11^t/n\right)C\left(I - 11^t/n\right)$ is a Moran index value. The eigenvectors Vs describe latent spatial autocorrelation included in each of variable.

Following Getis and Griffith (2002), the spatially filtered variable obtained with the eigenvectors approach is:

$$x_{iE}^* = e_i \tag{9}$$

with $e_i = x_i - \hat{x}_i$, where the model prediction $\hat{x}_i$ is obtained estimating the parameters of equation (8).

For every spatial location i, each variable can be rewritten as the sum of the filtered part x_{iE}^* and its spatial component s_{iE} as:

$$x_i = x_{iE}^* + s_{iE} \tag{10}$$

where: $s_{iE} = x_i - x_{iE}^* = x_i - e_i = \hat{x}_i$.

In other words, the predicted values from the regression (8) $\hat{x}_i$ constitute the spatial component for the variable, while the residuals e_i are the spatially filtered component.

Consider now, for each spatial location i, a vector of the p filtered variables, X_{iE}^*, and the vector of the spatial components of the p variables, S_{iE}, and plug equation (10), expressed in matrix notation, into equation (7) the following decomposition for first component:

$$y_{1i} = \left(X_{iE}^* + S_{iE}\right)^t A_1 \tag{11}$$

$$y_{1i} = X_{iE}^{*t} A_1 + S_{iE}^t A_1 \tag{12}$$

The first term on the right side of equation (12) is the filtered first component score $X_{iE}^{*t} A_1$ and the second part $S_{iE}^t A_1$ accounts for the presence of spatial dependence.

Using the Getis' approach, for each spatial location i, every variable can be filtered out by:

$$x_{iG}^* = x_i \left[\frac{C_i}{n-1}\right] / G_i(d) \tag{13}$$

where the x_{iG}^* represents the filtered part of the variable at each observation i, C_i is the sum of the geographical connections c_{il} (for every location i, with $l = 1, \ldots, n$) that are the elements of a binary contiguity matrix C, and $G_i(d)$ is the Getis–Ord statistic (Getis & Ord, 1992) measured for every variable as:

$$G_i(d) = \frac{\sum_{l \neq i} c_{il}(d) x_l}{\sum_{l \neq i} x_l} \tag{14}$$

In equation (13), $C_i/(n-1)$ is the expected value of equation (14), so that the filter is obtained from the ratio between the expected value of the $G_i(d)$ statistic and the value of the statistic itself.

To measure spatial connectivity throughout the paper, we follow standard practices for all techniques. In the GW approaches proximity is mostly defined according to a distance-decay function that allow to smoothly model variance–covariance structure and the use of binary weights is less frequent (Fotheringham et al., 2002). Conversely, adopting the filtering approaches, the use of binary contiguity matrix is common (Getis & Griffith, 2002; Griffith, 2008). In our study, for simplicity, the proximity is defined according to a k-nearest rule for both filtering approaches.

Also in this case, for every spatial location i, each variable can be rewritten as the sum of the filtered part x^*_{iG} and its spatial component s_{iG} as follows (Getis & Griffith, 2002):

$$x_i = x^*_{iG} + s_{iG} \tag{15}$$

The spatial component s_{iG} assumes positive values when the difference $x_i - x^*_{iG}$ is positive, a circumstance linked to the presence of autocorrelation among high values of the variable. When the spatial association is due to autocorrelation among low values, the spatial component s_{iG} is expected to assume negative values.

By plugging (15), expressed in matrix notation, into (7), we obtain the following equivalence for the first composite indicator:

$$y_{1i} = (X^*_{iG} + S_{iG})^t A_1 \tag{16}$$

$$y_{1i} = X^{*t}_{iG} A_1 + S^t_{iG} A_1 \tag{17}$$

The interpretation of the terms on the right side of equation (17) is analogous to those of equation (12). Thus, the score components can be split into two parts that sum the initial composite indicator obtained from standard PCA. In equation (17), $X^{*t}_{iG} A_1$ is the filtered first composite indicator and $S^t_{iG} A_1$ is the first spatial composite indicator.

Following the rationale of Getis and Griffith (2002), the magnitude of the spatial indicator will increase in two circumstances. The first occurs when spatial associations are present in high values (i.e., hot spots) for variables positively linked to the multivariate phenomenon (i.e., positive values of the weights). The second one is linked to a concentration of low values (i.e., cold spots) for variables negatively linked to the composite indicator by negative weights. Conversely, the spatial indicator will tend to assume negative values. In this sense, the spatial composite indicator is a measure of the effect of spatial associations on the multivariate phenomenon that will differ from zero when the concentration of high (or low) values of the variable (i.e., x_i) have the relevant effect on the composite indicator.

It is noteworthy that the Getis (1990) approach is based on the use of positive variables that have a natural origin. Since, in the present study, each variable respect these conditions, this method can be correctly implemented to extract the spatial part of each variable.

The properties of the two filtering approaches and the properties of these estimators are further explored in Griffith (2003, ch. 4). In the context of the screening spatial dependence in PCA, the Getis' approach seems to be preferable for many reasons. First, the spatial indicator obtained is directly connected to the extent of the spatial association exhibited by the geographical units, making the interpretation easier. Second, the Getis' filtering approach could be easily implemented, and it presents lower computational costs. In fact, in large data set applications, the Griffith's methodology could be performed at a high computational efforts due to the time needed to select the relevant eigenvectors. Finally, the Getis' approach offers a practical rule for the decision about the weighting matrix (Getis & Griffith, 2002). Hence, practitioners may simply try different numbers of neighbourhoods, until the filtered component shows no autocorrelation.

Sensitivity analysis for composite indicators

Defining a composite indicator involves a number of steps in which the researcher deals with subjective decisions regarding the variables to include and how to aggregate them (Kuc-Czarnecka et al., 2020). Generally, the first step is devoted to the definition of a conceptual framework on which the composite indicator is based, while the second step concerns the aggregation of the simple variables (Becker et al., 2017).

In the aggregation phase, two main aspects should be considered. First, the users have to define the aggregation method to synthetize the variables. In this step, inputs are combined through a simple or weighted average. Second, users have to choose values of the weights.

Besides the choice of a robust theoretical framework, a deeper understanding of the importance of each variable for the composite indicator is relevant, especially in the weighting process. This issue can be investigated using the sensitivity analysis (Saisana et al., 2005; Saltelli et al., 2004). If several measures can be defined to estimate sensitivity, here we rely on a popular approach, which uses the dependence of the composite indicator Y on each of the input variables X_j.

Following Becker et al. (2017), it is possible to define a measure of the importance of the input variable X_j on the composite indicator Y using a correlation ratio $S_j, j = 1, \ldots, p$, that can be estimated as follows:

$$\hat{S}_j = \frac{\sum_{i=1}^{n} (m_{ij} - \bar{m}_j)^2}{\sum_{i=1}^{n} (y_i - \bar{Y})^2} \tag{18}$$

where $m_{ij} = \hat{f}_j(x_{ij})$ is the fitted value of a regression of Y on X_j (for a particular point of X_j); and $\bar{m}_j = n^{-1} \sum_{i=1}^{n} (m_{ij})$.

Therefore, the correlation ratio S_j is an index of dependence of Y on X_j and it can be estimated by fitting a regression of Y on X_j, taking the variance of this regression, and dividing it by the variance of $\mathbf{Y}$. The statistic (18) is also known as the *first-order sensitivity index*. If the form of the regression function $f_j(x_{ij})$ is linear, the index $\hat{S}_j$ reduces to the well-known R-squared index.

In this paper, a sensitivity analysis is developed to evaluate the importance that each variable has on the indicator. In fact, if the set of variables used here is able to include different aspects in the dimensions of socioeconomic deprivation, it is also useful to assess the importance that each used variable may have on the final composite indicator. Particularly, we estimate if the obtained weight (i.e., the loadings) considerably differ from the relevance attributed by sensitivity analysis. Furthermore, besides the linear case, the relevance of omitted non-linear relationships between the composite indicators and the input variables is also considered through a non-parametric local non-linear kernel regression (Becker et al., 2017).

AN APPLICATION IN THE PROVINCE OF ROME

The empirical application considers two different levels of spatial resolution, to investigate differences in terms of correlation structures, interpretation of components, and spatial patterns. In fact, if aggregation is expected to increase representativeness, it could potentially lead to a difficulty of individuation of local pockets of deprivation that can only be captured by a lower spatial scale.

The spatial scales adopted are municipalities and localities, both based on administrative partitions. Localities are defined by the Italian Statistical Institute (ISTAT) as 'more or less large areas where inhabited houses are situated' and they represent a level between census tracts and municipalities. We use data from the 2011 General Census of Population and Housing with reference to 797 localities and 121 municipalities in the province of Rome.[2]

The data set is inspired by Pampalon and Raymond (2000) and Pampalon et al. (2009), but it has been integrated with additional variables similarly to Havard et al. (2008). In fact, these authors suggest using a small set of variables capable of capturing socioeconomic deprivation. Accordingly, the rate of people without a high school diploma (*School*), the rate of individuals living alone (*Mono*), the proportion of individuals separated, divorced or widowed (*SDV*), and the rate of single-parent families (*Sin Par*) are included. The proportion of people living in a rented house (*No House Prop*) is added as a proxy for house property. The unemployment rate (*Unemp*) and the population aged 70 and above on the total population (*Age*) are also used. We do not consider income and productivity, as this information is not provided at our fine spatial scale.

In the first step, standard PCA is used to obtain composite indicators at municipality and locality levels. Table 1 reports the cumulative proportion of variance for five PCs for both spatial scales.

At municipality level, the first component explains about 67% of the overall variance, while the first two components describe around the 80% of the total information. At the lower level, the first component accounts for the 50%, while including the second one representativeness reaches the 69% of the total variance. Those features confirm how many multivariate analyses tend to worsen representativeness at more refined scales (Fotheringham & Wong, 1991). The first two components are considered as indicators of socioeconomic deprivation at both levels. Table 2 presents the loadings of the first two components at municipality level.

First component suggests that the main effect is due to the number of people living alone (*Mono*) and the share of older people (*Age*). At the same time, economic variables tend to be less relevant. Additionally, the variable *No House Prop* has a negative sign. This combination points out how the first component is an indicator more connected to a social and demographic sphere. The second component indicator suggests a positive link to *No House Prop*. Moreover, this component is negatively related to *Age* and *School*. Not surprisingly, the picture changes at locality level because of the scaling effects. Table 3 shows loadings for localities.

In the first component indicator at locality level, *School* is negatively related to deprivation, together with the strong influence of social variables (i.e., *Mono* and *Sin Par*), which show negative association. This result appears in line with previous studies that point out how an increasing number of people living especially in middle-class and educated families could have been suffering over the last years, especially in the peripheral areas (De Muro et al., 2011). The second component is negatively related to *Mono*, but positively related to a lack of school attainment.

We also perform a sensitivity analysis in order to assess what relevance each variable has on the final output. The measure S_j of sensitivity is calculated for two different functional forms f_j. The first one $\hat{S}_{j,Lin}$ is obtained by simple linear regression, while the second one, $\hat{S}_{j,LocLin}$, for a non-linear local regression (Becker et al., 2017). Table 4 reports the results for the first component (see Appendix A in the supplemental data online for the sensitivity analysis of the second components)

Table 1. Cumulative proportion of the variance of the first five principal components for 121 municipalities and 797 localities in the province of Rome.

	PC1	PC2	PC3	PC4	PC5
Municipality level	0.674	0.795	0.873	0.919	0.963
Locality level	0.503	0.685	0.793	0.867	0.932

Table 2. Loadings for the first and second principal components for 121 municipalities in the province of Rome.

Variables	PC 1	PC 2
Unemp	0.024	−0.038
School	0.150	−0.422
No House Prop	−0.210	0.772
Age	0.357	−0.245
Mono	0.862	0.348
Sin Par	−0.147	0.130
SDV	0.203	0.162

which confirm, in absolute values, variable relevance from PCA loadings. This circumstance does not happen for all weighting techniques, as for example in equal weighting methods (Becker et al., 2017). Additionally, with only a few exceptions, the importance that each variable has for the values of sensitivity $\hat{S}_{j,Lin}$ in the linear case do not differ from $\hat{S}_{j,LocLin}$, calculated by local linear regression. Only slighter discrepancies between linear and non-linear cases emerge at locality level.

Figure 1 shows the spatial distributions of the two first indices of deprivation at municipality scale obtained with standard PCA. We can observe that the municipality of Rome presents a higher level of the two indicators (i.e., dark grey and black colour in the centre of the map). This counterintuitive evidence can be explained by the fact that very different situations are considered (and aggregated) within the municipality of Rome, including the peripheries and the downtown. Therefore, indicators at this level cannot highlight differences within the municipality.

To obtain a better analysis of the phenomenon under investigation, we consider the locality scale. Figure 2 outlines the spatial distribution of the two composite indicators at locality level obtained with standard PCA. The indices tend to worsen in some of the peripheral areas of the province and they improve in the downtown area of Rome (in the centre of the map). This aspect shows a dualism between downtown and periphery and suggests that *going spatially deeper* can reveal information that was hidden at a coarse spatial resolution.

The first spatial effect considered in this paper is spatial heterogeneity. Before investigating spatial heterogeneity by GWPCA, we perform a Monte Carlo test to verify significant non-stationarity of the eigenvalues (Harris et al., 2011). Figure 3 summarizes the result. Looking at the distribution obtained from Monte Carlo randomization, the hypothesis of stationarity is rejected at the lower spatial scale (p-value = 0.040). Instead, the hypothesis of stationarity of the eigenvalues is accepted at the municipality level (p-value = 0.200).

Table 3. Loadings for the first and second principal components for 797 localities in the province of Rome.

Variables	PC 1	PC 2
Unemp	−0.146	0.020
School	−0.751	0.395
No House Prop	−0.135	−0.111
Age	−0.172	−0.017
Mono	−0.397	−0.842
Sin Par	−0.422	0.232
SDV	−0.174	−0.260

Table 4. Results from sensitivity analysis for first component at municipality and locality level.

Variable	**Weight (PCA)**	$\hat{S}_{j,Lin}$	$\hat{S}_{j,LocLin}$
Municipality *level*			
Unemp	0.024	0.009	0.009
School	0.150	0.189	0.191
No House Prop	−0.210	0.234	0.483
Age	0.357	0.683	0.692
Mono	0.862	0.961	0.963
Sin Par	−0.147	0.246	0.253
SDV	0.203	0.491	0.493
Locality level			
Unemp	−0.146	0.151	0.181
School	−0.751	0.863	0.905
No House Prop	−0.135	0.086	0.090
Age	−0.172	0.221	0.260
Mono	−0.397	0.370	0.463
Sin Par	−0.422	0.583	0.650
SDV	−0.174	0.271	0.301

Therefore, the evidence from Figure 3 suggests exploring spatial heterogeneity only at locality level.

In applying GWPCA, the research on the optimal bandwidth is crucial. Figure 4 shows the CV functions for different numbers of components at locality level.

Owing to the presence of an irregular spatial configuration of the province of Rome, an adaptive kernel is preferred in this study (Fotheringham et al., 2002). As can be observed from the plot, for $q = 5$ the CV function shows a clear global minimum of $\gamma = 40$. This combination is chosen in the analysis. However, coherently with the rest of the analysis, only the first two components are reported in this section.

Figure 1. Quantile maps of the first (A) and second (B) socioeconomic deprivation index for 121 municipalities in the province of Rome.

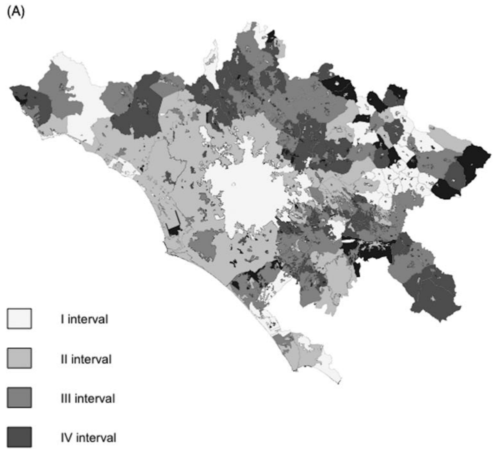

Figure 2. Quantile maps of the first (A) and second (B) socioeconomic deprivation index for 797 localities in the province of Rome.

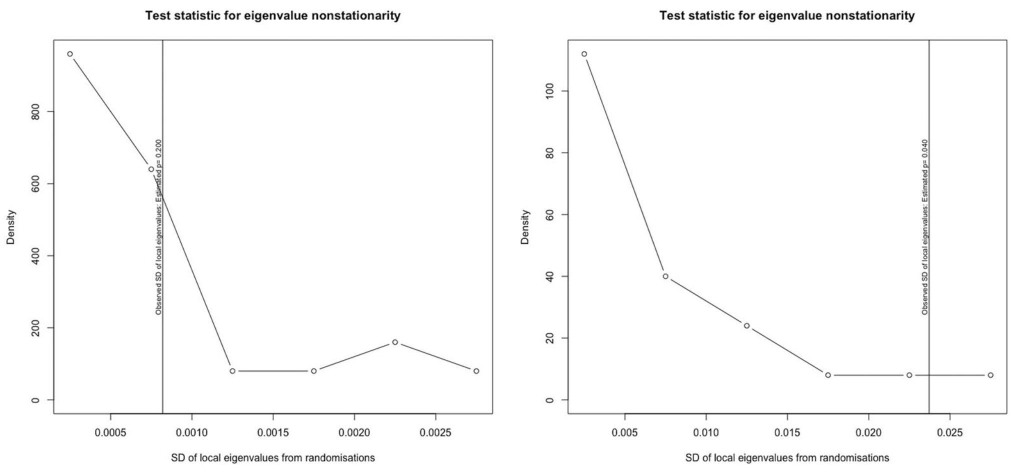

Figure 3. Monte Carlo test for the stationarity of eigenvalues for municipalities (left) and localities (right).

Results from GWPCA are often summarized using the winning variables (Harris et al., 2011) that can be defined as the variables indicating the highest local loading in absolute value. In terms of composite indicators, the winning variables suggest what is the most important variable influencing the final indicator. Figure 5 includes maps of the winning variables at locality level for the first two components. The effects of different kernels choice (i.e., exponential and bi-square) on the winning variables map has been also considered and reported in Appendix B in the supplemental data online. The similarities in the first component patterns confirm the robustness of the results to the choice of the kernel.

Figure 5 highlights local instabilities. For the first composite indicator, the lack of school attainment largely influences socioeconomic deprivation. This result is very similar to that obtained with standard PCA. However, the map highlights the importance of *Mono* and *No House Prop* in several localities at the northern border. In the second composite indicator, the

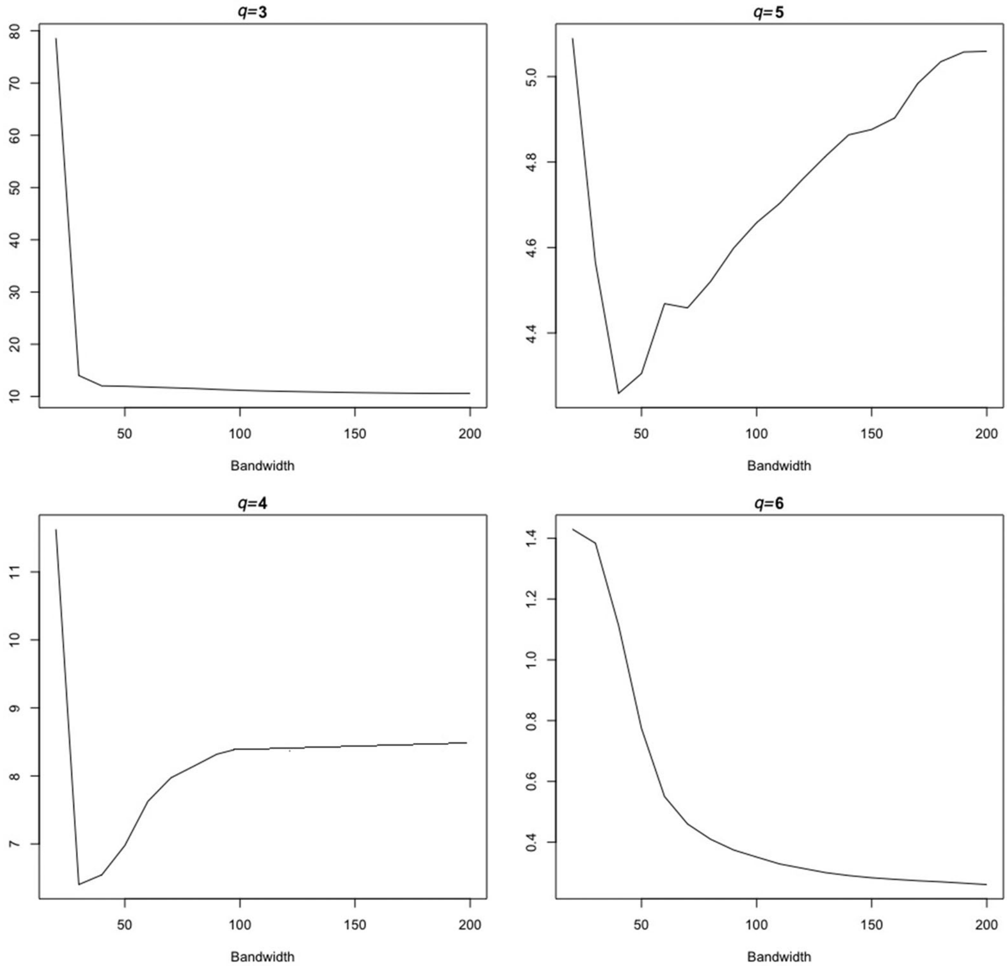

Figure 4. Cross-validation (CV) functions in the case of adaptive bandwidth for several choices of retained components at locality level.

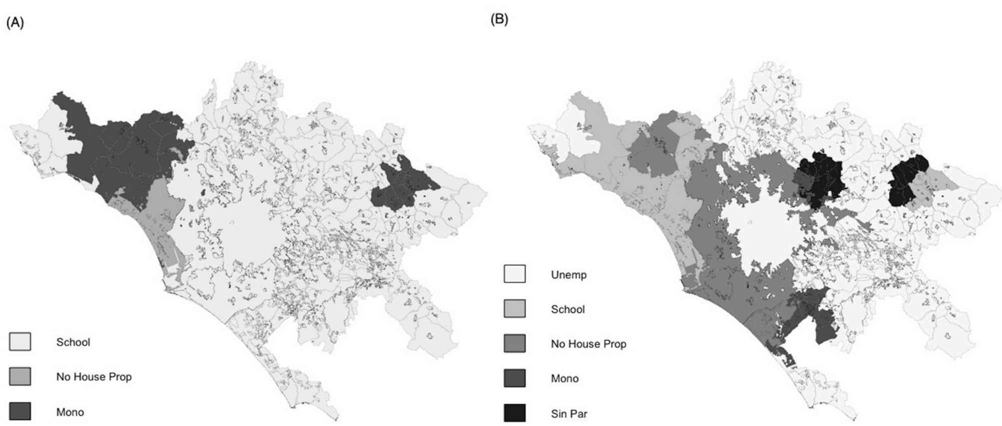

Figure 5. Map of the winning variables for the first (A) and second (B) socioeconomic deprivation index calculated by the GWPCA, with a Gaussian kernel, for 797 localities in the province of Rome.

relevance of the house property affects the peripheral areas of the municipality of Rome, while unemployment is important at the east border.

Considering potential heterogeneity produces a general increase in the proportion of variance accounted. The performance of GWPCA (Figure 6), in terms of the variance explained by the first five components, is better in 680 out of 797 localities (i.e., 85% of the units under study) compared with that obtained with standard PCA.

Investigating local deprivation using GWPCA cannot directly evidence the amount of the neighbourhood effects. This information might help practitioners to evaluate spatial dependence in the composite indicator (Atkinson & Kintrea, 2001).

The first and second terms of the right side of equations (12) and (17) show the calculation of filtered indicators of deprivation and their spatial counterparts. Results allows us to evaluate spatial effects in multivariate deprivation by different filtering techniques. In both the filtering processes, different connectivity matrices were analysed and a $k = 5$ is finally used at both levels.[3]

Figure 7 shows spatial and filtered indicators at municipality scale by Griffith's filtering procedure, while Figure 8 reports spatial and filtered indicators obtained using Getis' technique.

In both Figures 7 and 8, the spatially filtered components (A and B) attribute an idiosyncratic level (not influenced by dependence) of deprivation to the city of Rome. This result suffers from the impossibility to detect in both components differences between peripheries and downtown.

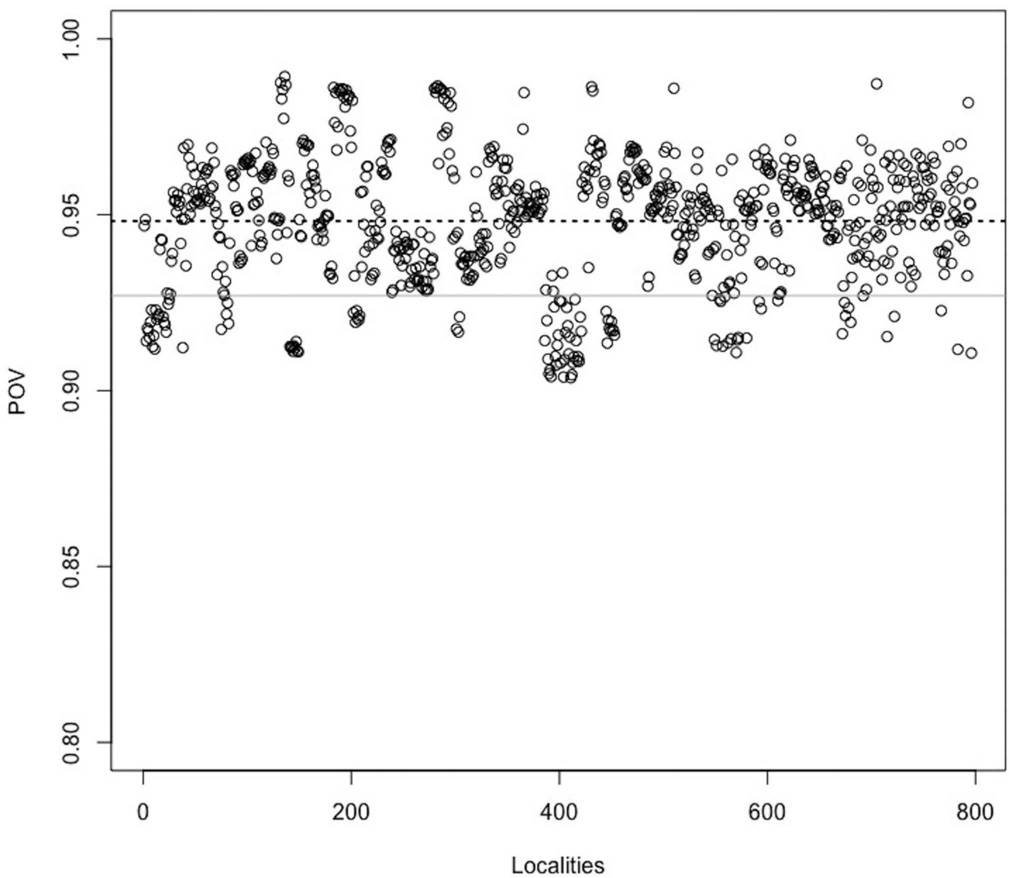

Figure 6. Proportion of variance (POV) of the first five components for the GWPCA (black points), overall standard PCA (line) and the GWPCA mean (dotted) in 797 localities in the province of Rome.

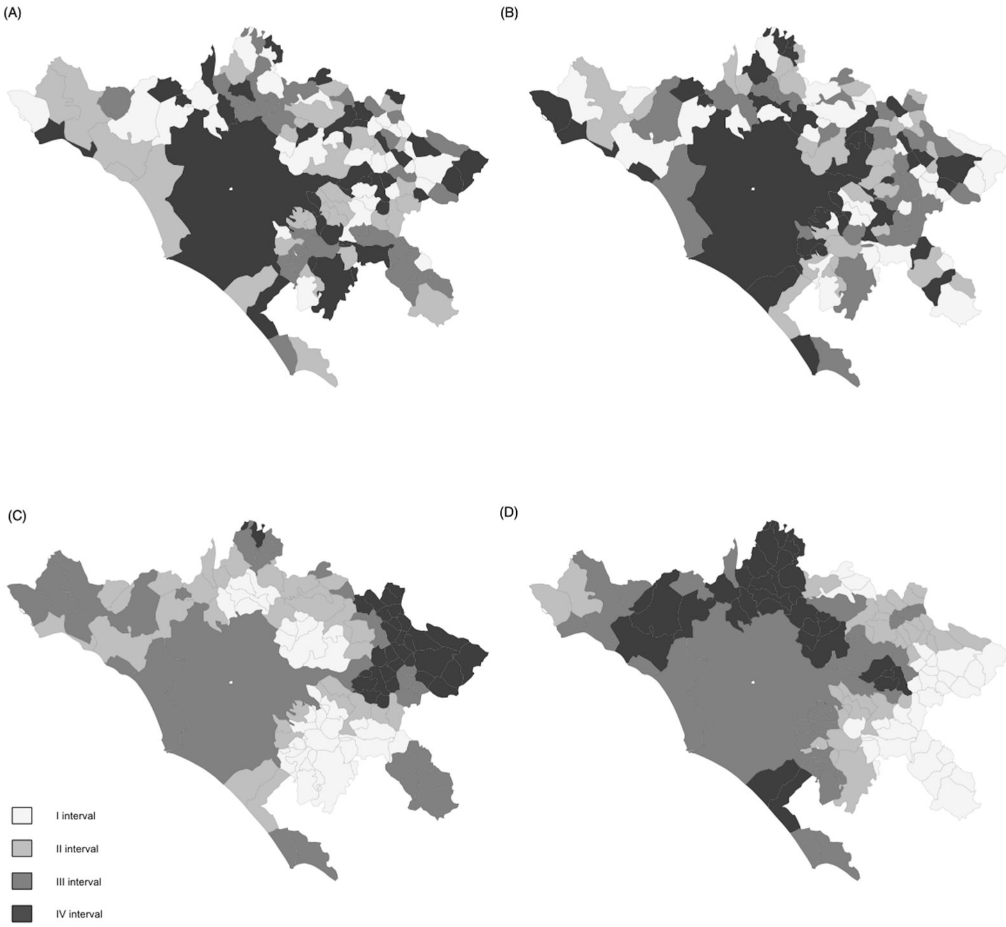

Figure 7. Quantile maps of the filtered (A) and spatial part (C) of the first and of the filtered (B) and spatial part (D) of the second socioeconomic deprivation index for 121 municipalities in the province of Rome adopting Griffith's filtering.

However, at this level it is already clear that spatial dependence has consequences on the effects of the global deprivation.

Figure 8(C, D) shows spatial terms for the first two component obtained by the Getis' filtering procedure. For the first component (Figure 8C), we have that areas out of the municipality of Rome are characterized by higher values of the spatial part. In the second component (Figure 8D), we note that spatial dependence has effects on Rome and its surrounding municipalities. For Griffith's filtering (Figure 7D), the spatial term of second component returns lower levels in Rome than in the conterminous areas.

By comparing Figures 7 and 8, we can note differences between the two approaches. Maps from Getis' filtering seems to evidence very accurate patterns of the spatial terms. This aspect is justified by a clear east–west polarization for the Getis' filtering between the mountain municipalities and the more densely inhabited areas closer to the sea.

Figures 9 and 10 summarize the results at locality level. At this level, we appreciate how lower scale helps us to regain a more detailed picture. For the first component, both filtering approaches (Figures 9A and 10A) show that downtown Rome is affected by lower levels of idiosyncratic deprivation, which increases in the peripheral areas. In the second component, the filtered part

Figure 8. Quantile maps of the filtered (A) and spatial part (C) of the first and of the filtered (B) and spatial part (D) of the second socioeconomic deprivation index for 121 municipalities in the province of Rome adopting Getis' filtering.

from Getis' (Figure 10B) is higher in the downtown, differently from the level indicated by the use of Griffith's filtering (Figure 9B).

Moving to the spatial terms, spatial dependence influences many areas of the province. Local pockets where the effects of spatial dependence are better highlighted by the Getis' filtering (Figure 10C) than in the Griffith's procedure (Figure 9C). Moreover, in both solutions, the spatial indicator is also very relevant for a set of small localities situated at the East and South. In the second component, the spatial indicator changes its pattern from the Griffith's (Figure 9D) to the Getis' approach (Figure 10D). In fact, the last pattern points out again a core–periphery pattern. This issue is more tenuous using Griffith's filtering.

The results at this level present more evident changes between the two filtering techniques. Again, by the use of Griffith's approach is more difficult to spot very local features and differences, as the pattern is more homogenous throughout the whole map. This evidence confirms that, also at locality scale, the maps from the Getis' filtering allows a clearer interpretation of spatial components.

The use of the two filtering techniques also suggests that, for a larger number of units and several variables, the Griffith's filtering for PCA tends to slow down due to computational efforts for

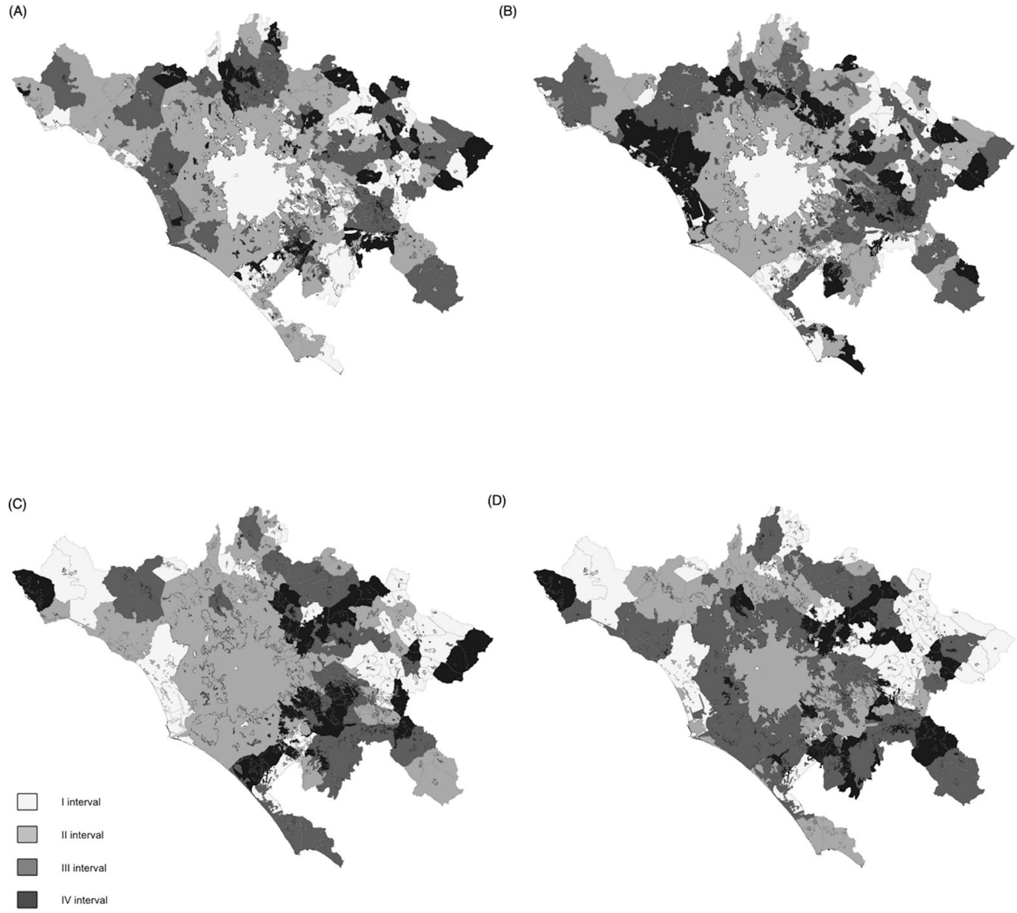

Figure 9. Quantile maps of the filtered (A) and spatial part (C) of the first and of the filtered (B) and spatial part (D) of the second socioeconomic deprivation index for 797 localities in the province of Rome adopting Griffith's filtering.

eigenvectors selection. Furthermore, at locality level, the filtered indicators from the Getis' procedure are less connected in terms of Pearson correlation to the composite indicator obtained by PCA. In fact, the Griffith's filtered indicators are more correlated to the ones from standard PCA (Pearson correlation is about 0.80 for the two components). Hence, the filtered indicators obtained with the Getis' approach could be considered as more informative. Those empirical issues lead us to prefer the Getis' approach.

Compared with GWPCA, splitting spatial and non-spatial components helps the practitioners to isolate the presence of spatial effects and it enlarges considerations on the spatial context. Therefore, the spatial indicators might help us to individuate the high relevance that spatial component have on global indicators and they may improve the effectiveness of local policies.

Table 5 reports Moran's I for the first two components and their filtered parts at the two spatial scales. Both procedures filter out spatial dependence at the selected levels. This feature is justified by not statistically significant Moran's I of the filtered components (second and third columns in Table 5).

Table 5 gives evidence about the changes of the spatial autocorrelation that occurs when the level of aggregation shifts. The severe effect of MAUP is consistent with other analyses that

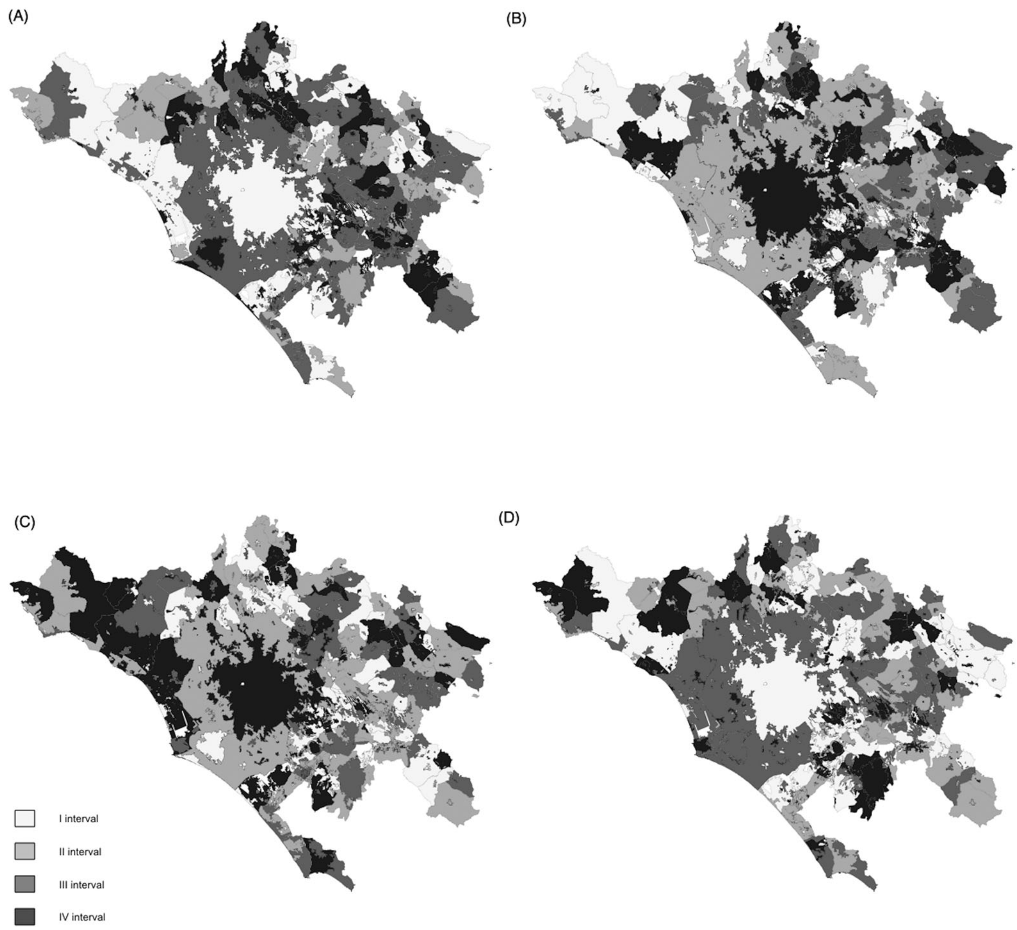

Figure 10. Quantile maps of the filtered (A) and spatial part (C) of the first and of the filtered (B) and spatial part (D) of the second socioeconomic deprivation index for 797 localities in the province of Rome adopting Getis' filtering.

support the idea that the correlation between variables increases while the variance decreases at aggregated spatial resolutions. This situation occurs in other statistical methods when a coarse spatial resolution is considered (Arbia et al., 1996; Wu, 2004).[4] For this reason, the use of a coarse spatial scale may greatly affect not only global results but also lead to problems in the interpretation of spatial effects.

CONCLUSIONS

In this paper we investigate the influence of spatial effects and spatial scale in PCA for the derivation of composite indicators. PCA calculates composite indicators using the same as weights for all spatial units under investigation. This assumption seems inaccurate to treat geographical data. To address this problem, researchers may use local weights derived from GWPCA. Unfortunately, this methodology does not consider the presence of dependence that often affects spatial data.

In order to properly model dependence for composite indicators, we propose a novel methodology based on spatial filtering approaches (Getis & Griffith, 2002). This framework is considered

Table 5. Moran's *I*-statistics for the first and second components and for their filtered part calculated at municipalities and localities in the province of Rome.

	Global indicator	Spatially filtered indicator (Getis)	Spatially filtered indicator (Griffith)
Municipality level			
First component	0.469	−0.126	−0.108
$k = 5$ contiguity matrix	(0.000)	(0.987)	(0.973)
Second component	0.399	−0.186	−0.089
$k = 5$ contiguity matrix	(0.000)	(0.996)	(0.937)
Locality level			
First component	0.051	−0.208	−0.033
$k = 5$ contiguity matrix	(0.020)	(0.980)	(0.930)
Second component	0.212	−0.170	0.012
$k = 5$ contiguity matrix	(0.000)	(0.980)	(0.260)

Note: *p*-values are given in parentheses.

more helpful than sPCA to decompose the composite indicator obtained with PCA in a spatial and an idiosyncratic (non-spatial or filtered) part.

Looking at the results, the Getis' filtering approach seems to be preferable for the lower computational effort and for the easier interpretation of the results. Besides, our evidence highlights that the MAUP dramatically affects the results obtained with all spatial approaches to PCA. The use of an aggregate spatial scale is a common circumstance, but this choice should be made with particular caution to interpret the phenomenon in an appropriate way.

In addition, investigating spatial effects helps us to increase the informative potential of composite indicators. Isolating the spatial components through our novel method quantifies the magnitude of the neighbourhood component of material deprivation.

A real and broader comprehension of the role of space seems to be useful in the analysis of multivariate phenomena. The assessment of spatial effects appears to be a crucial element in promoting policies and political agendas that consider interactions for facilitating development and addressing disparities (Barca et al., 2012).

Further, we are aware that the problem of spatial dependence effects could be also addressed according to the definition of a nested hierarchy (Chung & Hewings, 2015). Interactions between variables that affect units and the groups they belong represent a reality for a certain number of spatial processes (Corrado & Fingleton, 2012). For this reason, a future research line might explore the connection between hierarchical models and the spatial econometric tools which have been mainly considered in the present study.

Lastly, future research might involve the modelling of discrete spatial heterogeneity, facing the problem of structural differences between spatial units directly due to the presence of multiple clusters. This research line may supplement the use of GWPCA that analyses heterogeneity in the continuous space.

NOTES

[1] Note that in this case, for simplicity of the narrative, we omitted the subscript $j=1,\ldots,p$ in the definition of the variable X.
[2] Data at the local level from the General Census of Population and Housing are openly available at https://www.istat.it/it/archivio/104317#accordions.
[3] The level of the nearest neighbours was set for the two spatial levels according to Getis and Griffith (2002).
[4] Arbia et al. (1996, p. 124) claim that 'everything is related to everything else, but things observed at a coarse spatial resolution are more related than things observed at a finer resolution'.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This study was financed by the project H2020 IMAJINE which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 726950.

ORCID

Paolo Postiglione http://orcid.org/0000-0002-2798-4519

REFERENCES

Anselin, L. (1988). *Spatial econometric: Methods and models*. Kluwer Academic Publisher.
Arbia, G., Benedetti, R., & Espa, G. (1996). Effects of the MAUP on image classification. *Geographical Systems*, *3*, 123–141.
Arbia, G., & Petrarca, F. (2011). Effects of MAUP on spatial econometric models. *Letters in Spatial and Resource Sciences*, *4*(3), 173–185. https://doi.org/10.1007/s12076-011-0065-9
Atkinson, R., & Kintrea, K. (2001). Disentangling area effects: Evidence from deprived and non-deprived neighbourhoods. *Urban Studies*, *38*(12), 2277–2298. https://doi.org/10.1080/00420980120087162
Barca, F., McCann, P., & Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. *Journal of Regional Science*, *52*(1), 134–152. https://doi.org/10.1111/j.1467-9787.2011.00756.x
Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056
Bellandi, M., & Ruiz-Fuensanta, M. J. (2010). An empirical analysis of district external economies based on a structure-conduct-performance framework. *Papers in Regional Science*, *89*(4), 801–818. https://doi.org/10.1111/j.1435-5957.2009.00272.x
Bonneu, F., & Thomas-Agnan, C. (2015). Measuring and testing spatial mass concentration with micro-geographic data. *Spatial Economic Analysis*, *10*(3), 289–316. https://doi.org/10.1080/17421772.2015.1062124
Caranci, N., Biggeri, A., Grisotto, L., Pacelli, B., Spadea, T., & Costa, G. (2010). L'indice di deprivazione italiano a livello di sezione di censimento: definizione, descrizione e associazione con la mortalità. *Epidemiologia e Prevenzione*, *34*, 167–176.

Chung, S., & Hewings, G. J. D. (2015). Competitive and complementary relationship between regional economies: A study of the Great Lake States. *Spatial Economic Analysis*, *10*(2), 205–229. https://doi.org/10.1080/17421772.2015.1027252

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, *52*(2), 210–239. https://doi.org/10.1111/j.1467-9787.2011.00726.x

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, *32*(1), 7–34. https://doi.org/10.1080/07474938.2012.690641

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A. S., & McLoone, S. (2013). Principal component analysis on spatial data: An overview. *Annals of the Association of American Geographers*, *103*(1), 106–128. https://doi.org/10.1080/00045608.2012.689236

De Muro, P., Mazziotta, M., & Pareto, A. (2011). Composite indices of development and poverty: An application to MDGs. *Social Indicators Research*, *104*(1), 1–18. https://doi.org/10.1007/s11205-010-9727-z

Fernández-Vazquez, E., & Rubiera-Morollòn, F. (2012). *Defining the spatial scale in modern regional analysis: New challenges from data at local level*. Springer.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression – The analysis of spatially varying relationships*. Wiley.

Fotheringham, A. S., & Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space*, *23*(7), 1025–1044. https://doi.org/10.1068/a231025

Getis, A. (1990). Screening for spatial dependence in regression analysis. *Papers of the Regional Science Association*, *69*(1), 69–81. https://doi.org/10.1007/BF01933897

Getis, A. (1995). Spatial filtering in a regression framework: Experiments on regional inequality, government expenditures, and urban crime. In L. Anselin & R. J. G. M. Florax (Eds.), *New directions in spatial econometrics* (pp. 172–185). Springer.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical Analysis*, *34*(2), 130–140. https://doi.org/10.1111/j.1538-4632.2002.tb01080.x

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Greyling, T., & Tregenna, F. (2016). Construction and analysis of a composite quality of life index for a region of South Africa. *Social Indicators Research*, *131*(3), 887–930. https://doi.org/10.1007/s11205-016-1294-5

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, *2*(2), 141–156. https://doi.org/10.1007/PL00011451

Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Springer Science & Business Media.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A: Economy and Space*, *40*(11), 2751–2769. https://doi.org/10.1068/a38218

Griffith, D. A., & Chun, Y. (2014). Spatial autocorrelation and spatial filtering. In N. Fischer M. (Ed.), *Handbook of regional science* (pp. 1479–1507). Springer.

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, *25*(10), 1717–1736. https://doi.org/10.1080/13658816.2011.554838

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. *Geographical Analysis*, *47*(2), 146–172. https://doi.org/10.1111/gean.12048

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. *Social Science & Medicine*, *67*(12), 2007–2016. https://doi.org/10.1016/j.socscimed.2008.09.031

Jolliffe, I. T. (2002). *Principal component analysis*. Springer.

Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, *101*(1), 92–103. https://doi.org/10.1038/hdy.2008.34

Kallio, M., Guillaume, J. H., Kummu, M., & Virrantaus, K. (2018). Spatial variation in seasonal water poverty index for Laos: An application of geographically weighted principal component analysis. *Social Indicators Research*, *140*(3), 1131–1157. https://doi.org/10.1007/s11205-017-1819-6

Kuc-Czarnecka, M., Lo Piano, S., & Saltelli, A. (2020). Quantitative storytelling in the making of a composite indicator. *Social Indicators Research*. https://doi.org/10.1007/s11205-020-02276-0

Lloyd, C. D. (2010). Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, *34*(5), 389–399. https://doi.org/10.1016/j.compenvurbsys.2010.02.005

Márquez, M. A., Lasarte, E., & Lufin, M. (2019). The role of neighborhood in the analysis of spatial economic inequality. *Social Indicators Research*, *141*(1), 245–273. https://doi.org/10.1007/s11205-017-1814-y

McGillivray, M. (2005). Measuring non-economic well-being achievement. *Review of Income and Wealth*, *51*(2), 337–364. https://doi.org/10.1111/j.1475-4991.2005.00157.x

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicators building* (Report JRC31473). European Commission.

OECD. (1993). *OECD Core set of indicators for environmental performance reviews* (Environment Monographs n. 83).

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in spatial sciences* (pp. 127–144). Pion.

Pampalon, R., Hamel, D., Gamache, P., & Raymond, G. (2009). A deprivation index for health planning in Canada. *Chronic Disease in Canada*, *29*, 178–191.

Pampalon, R., & Raymond, G. (2000). A deprivation index for health and welfare planning in Quebec. *Chronic Disease in Canada*, *21*, 104–113.

Panzera, D., & Postiglione, P. (2020). Measuring the spatial dimension of regional inequality: An approach based on the Gini correlation measure. *Social Indicators Research*, *148*(2), 379–394. https://doi.org/10.1007/s11205-019-02208-7

Postiglione, P., Andreano, M. S., & Benedetti, R. (2013). Using constrained optimization for the identification of convergence clubs. *Computational Economics*, *42*(2), 151–174. https://doi.org/10.1007/s10614-012-9325-z

Purtauf, T., Thies, C., Ekschmitt, K., Wolters, V., & Dauber, J. (2005). Scaling properties of multivariate landscape structure. *Ecological Indicators*, *5*(4), 295–304. https://doi.org/10.1016/j.ecolind.2005.03.016

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, *11*(1), 189–209. https://doi.org/10.1093/cjres/rsx024

Saib, M. S., Caudeville, J., Beauchamp, M., Carré, F., Ganry, O., Trugeon, A., & Cicolella, A. (2015). Building spatial composite indicators to analyze environmental health inequalities on a regional scale. *Environmental Health*, *14*(1), 68. https://doi.org/10.1186/s12940-015-0054-3

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice. A guide to assessing scientific models*. Wiley.

Tiefelsdorf, M., & Boots, B. (1995). The exact distribution of Moran's I. *Environment and Planning A: Economy and Space*, *27*(6), 985–999. https://doi.org/10.1068/a270985

Townsend, P. (1987). Deprivation. *Journal of Social Policy*, *16*(2), 125–146. https://doi.org/10.1017/S0047279400020341

Trogu, D., & Campagna, M. (2018). Towards spatial composite indicators: A case study on Sardinian landscape. *Sustainability*, *10*(5), 1369. https://doi.org/10.3390/su10051369

Wackernagel, H. (2003). *Multivariate geostatistics* (3rd ed.). Springer.

Wartenberg, D. (1985). Multivariate spatial correlation: A method for exploratory geographical analysis. *Geographical Analysis*, *17*(4), 263–283. https://doi.org/10.1111/j.1538-4632.1985.tb00849.x

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*(2), 125–138. https://doi.org/10.1023/B:LAND.0000021711.40074.ae