


FULL ARTICLE



Measuring spatial concentration: A transportation problem approach

Mauro Ferrante¹  | Giovanni Luca Lo Magno¹ | Stefano De Cantis² | Geoffrey J.D. Hewings³

¹ Department of Culture and Society,
University of Palermo, Palermo, Italy

² Department of Economics Business and
Statistics, University of Palermo, Palermo, Italy

³ Regional Economic Application Laboratory,
University of Illinois, Urbana, Illinois

Correspondence

Mauro Ferrante, Assistant Professor,
Department of Culture and Society, University
of Palermo, Viale delle Scienze, ed. 15, 90128
– Palermo, Italy.
Email: mauro.ferrante@unipa.it

Abstract

This paper will propose an index for measuring spatial concentration, which is based on the solution of a particular transportation problem. This approach extends a new index proposed for the measurement of seasonality to the spatial case, and it takes into account the level of the concentration of the phenomenon of interest according to its spatial distribution. Various properties of the proposed index, which make it a desirable measure for spatial concentration, will also be described. An empirical application, using data from selected European countries, will be provided, and the results derived from the proposed index will be compared with those derived from currently used indices.

KEYWORDS

European regions, inequality measures, Spatial concentration, transportation problem

JEL CLASSIFICATION

R12; D63; C60

1 | INTRODUCTION

The measurement of spatial concentration has received significant attention in recent decades. As the volume of research within the sphere of the *New Economic Geography* has increased, *concentration*, *agglomeration* and *inequality* have become familiar themes in the attempt to provide insights in the spatial structure of economies on different spatial scales. Beyond a major interest in measuring the spatial concentration of economic activities (Arbia & Piras, 2009; Ellison & Glaeser, 1997; Krugman, 1991; Liu, 2014), many other disciplines are concerned with phenomena



distributed across space (see for example, Anselin, Cohen, Cook, Gorr, & Tita, 2000; Freeman, Grogger, & Sonstelie, 1996; Reardon & O'Sullivan, 2004).

Many solutions have been proposed for the measurement of the spatial concentration of spatial point patterns (Duranton & Overman, 2005; Marcon & Puech, 2003), all of which require the availability of micro-level information (e.g. location of the industry). Nonetheless, it is more common for regional scientists to only have access to spatially aggregated, areal data. The majority of concentration statistics currently used are generally non-spatial measures; examples of these include the Gini, Herfindahl or the Ellison and Glaeser indices. This can lead to a lack of sensitivity to permutations in the spatial position of regions. Although most of the challenges affecting the currently used indices for measuring spatial concentration have been well documented in the literature (Arbia, 2001; Kopczewska, 2018), only a few papers (Arbia & Piras, 2009; Bickenbach & Bode, 2008; Guimarães, Figueiredo, & Woodward, 2011; Sohn, 2014) have attempted to overcome these challenges; in many cases, the proposed solutions are generally *ad hoc* solutions. These approaches generally aim at combining non-spatial concentration indices (e.g. the Herfindahl or Ellison and Glaeser indices) with spatial autocorrelation indices (e.g. Moran's *I*). The aim of such a combination is to address two facets of spatially concentrated phenomena, namely: the (non-spatial) degree of inequality in the distribution of the phenomenon of interest among the considered units, and its distribution in space (Arbia & Piras, 2009; Guimarães et al., 2011; Lafourcade & Mion, 2007).

Departing from these considerations, rather than combining non-spatial and spatial measures, this paper will present a measurement approach, in which the degree of spatial concentration is evaluated via the solution of an appropriately defined transportation problem (Winston & Goldberg, 2004). This paper extends a recent proposal by Lo Magno, Ferrante, and De Cantis (2017) to the spatial case; its aim is to propose a spatial concentration index based on the minimization of the cost of transferring the amount of the phenomenon of interest from over-average to under-average regions, in order to achieve an even distribution of the phenomenon of interest among the regions under study. The shift of the focus from the temporal to the spatial context will be achieved through an appropriate specification of the cost matrix, which incorporates the structure of the relationship among the units. It is proposed that the latter should be related to location and distance in a geographic space or in a more general economic space. Despite the similarities between temporal and spatial concentration, the latter is a more complex concept due to the relationships which exist among the spatial units (Dubin, 1998). This raises several considerations which, to the best of our knowledge, have not been previously discussed in the wider literature dealing with spatially concentrated phenomena.

After outlining the main limitations arising from the measuring of spatial concentration with the more traditional indices and recent proposals, and exemplifying peculiar spatially concentrated patterns, a real data example will be provided for illustrative purposes. The aim of this is to demonstrate how the proposed approach captures various intuitive features of spatial concentration, which are overlooked by other indices currently in use. By considering the wide range of disciplines and applications related to spatially concentrated phenomena, this paper contends that this proposed approach merits attention and from a variety of perspectives.

2 | MAIN LIMITS OF SPATIAL CONCENTRATION INDICES

2.1 | General Considerations and Prior Research

In measuring spatial concentration, not only the level of the phenomenon of interests is of relevance, but also its distribution across space. Notwithstanding the renewed interest in regionally-distributed phenomena, such as regional economic growth and regional income inequality (see McCann, 2016; Ganong & Shoag, 2017), it is the opinion of many researchers (eg. Rey & Janikas, 2005) that the geographical dimension characterizing these phenomena has been overlooked. That is, many methods and measures developed within the various disciplines of the social science have been applied to spatial data without paying adequate attention to the spatial dimension (Chakravorty, 1996).

In measuring spatial concentration, traditional concentration measures have generally borrowed directly from the income inequality literature. Examples of such measures include the Gini and Herfindahl indices, both of which are insensitive to the distribution of the phenomenon across considered spatial units. Due to the *anonymity* property (Sen, 1973), these measures are invariant to permutations of the phenomenon across spatial units. Thus, significantly different spatial patterns would produce the same result (Arbia, 2001). From the time when Openshaw (1983) raised the issue, subsequently known as the *modifiable areal unit problem* (MAUP), a number of approaches have been explored to deal with this problem in a variety of different contexts and applications (see, for example, Horner & Murray, 2002; Wong, 2009; Menon, 2012).

The MAUP was again discussed by Lafourcade and Mion (2007), and, more recently, Kopczewska (2018) provided an in-depth analysis of the cluster-based measures of the geographical and sectoral concentration offered in the regional science literature, in which the MAUP had been emphasized. Other issues related to the limits in identifying spatial agglomeration through raw concentration measures have been discussed by Carlei and Nuccio (2014). And Chain, Santos, Castro, and Prado (2018) have provided a bibliometric analysis of the methods used for the measurement of industrial clusters.

On the other hand, pure spatial correlation measures, such as Moran's *I* or the Getis-Ord statistic, are not suitable as concentration indices: they fail to appropriately take into account the distribution of the phenomenon of interest. This point has also been exemplified with illustrative examples and empirical applications by Guimarães et al. (2011). Furthermore, correlation and concentration are clearly two different concepts.

In analyzing spatial point data (given, for example, by the availability of geo-referenced microdata), Duranton and Overman (2005, 2008) and Marcon and Puech, (2003, 2017) have proposed concentration measures using functions based on Ripley's *K* function. Such approaches may overcome the challenges related to the checkerboard problem and the MAUP. Nonetheless, the unavailability of detailed geo-coded data and the computational difficulties, which are related to the analysis of very large spatial databases, markedly limits the widespread application of these approaches (Duranton & Overman, 2005; Ochojski, Polko, & Churski, 2017). Consequently, the measurement of spatial concentration, in the context of areal data (the most common case for regional scientists) is still an open issue.

2.2 | An Illustrative Example

In order to highlight the main differences among spatial and non-spatial concentration measures, let us first consider the spatial patterns in Figure 1 with the results of traditional concentration indices, as reported in Table 1 and as first proposed by Arbia (2001).

This example highlights the well-known checkerboard problem, which arises when the spatial distribution of the phenomenon of interest is not taken into account in the measurement of concentration. It has been already highlighted that the level of spatial concentration is the highest in Situation 1.a, and higher in Situation 1.b. than in Situation 1.c. None of the considered non-spatial indices, namely the Herfindahl index (*H*), the Gini index (*G*) and the Ellison and Glaeser (*γ*) index, are capable of distinguishing among these hypothetical situations.

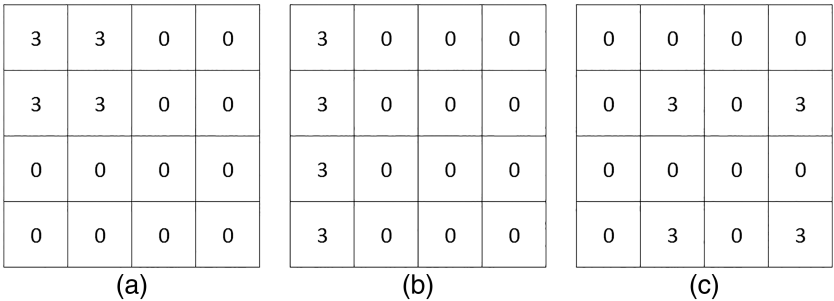


FIGURE 1 Hypothetical spatial distribution of 12 firms



TABLE 1 Concentration measures for the example in Figure 1

Index	1.a	1.b	1.c
H	0.2500	0.2500	0.2500
G	0.1875	0.1875	0.1875
γ	0.1273	0.1273	0.1273

Departing from this consideration, Arbia and Piras (2009), Guimarães et al. (2011), among others, have proposed an alternative approach in order to take into account the spatial structure of the area of interest. These proposed approaches share the notion of combining traditional inequality measures with various statistics regarding spatial association, such as Moran's I or the Getis Ord statistic (Bickenbach & Bode, 2008). In this regard, Guimarães et al. (2011) have proposed a spatially weighted version of the Herfindahl index, as follows:

$$H_S = H + \mathbf{x}'\mathbf{W}\mathbf{x}, \tag{1}$$

in which H is the original version of the Herfindahl index, $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ is the vector containing the regional shares of the phenomenon of interest distributed across the n regions, and \mathbf{W} is the conventional row standardized contiguity matrix. Similarly, a spatially weighted version of the Gini index, and the Ellison and Glaeser index have been derived (Guimarães et al., 2011) in the same work. Nonetheless, the proposed approaches may be seen as the combination of two concepts, namely, spatial autocorrelation and (non-spatial) concentration; these concepts may produce counter- intuitive results in the measurement of spatial concentration.

3 | MEASURING SPATIAL CONCENTRATION: A TRANSPORTATION PROBLEM APPROACH

Whilst theoretically sound, it is not easy to understand the results obtained by combining spatial autocorrelation with non-spatial indices of concentration in many empirical and hypothetical situations; in some cases, this may also be counter-intuitive. For example, let us consider the patterns in Figure 2. It seems intuitive that a measure of spatial concentration should be able to identify Figure 2.a as being more concentrated than that in Figure 2.b. However, these situations would be equally evaluated by the indices proposed by both Arbia and Piras (2009), and by Guimarães et al. (2011). This is also true for all the intermediate situations under consideration in Figures 2.a and 2.b.

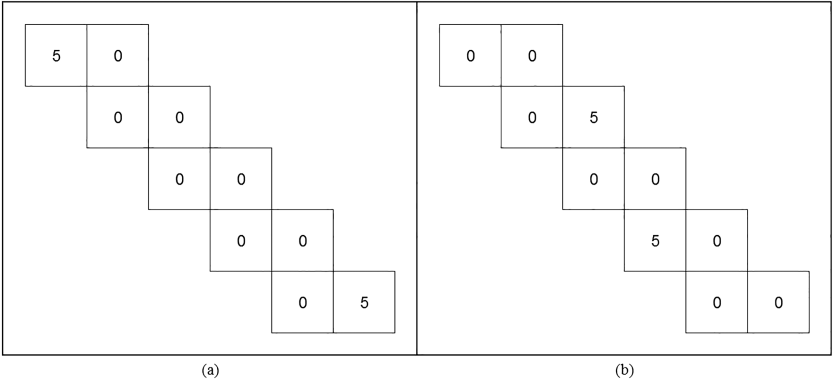


FIGURE 2 Two hypothetical distributions of 10 firms



Bearing in mind these considerations, the approach proposed herein for measuring spatial concentration is based on the solution of an appropriately defined transportation problem. This transportation problem is a well-known linear minimization problem: the goal is to minimize the cost of transferring of a single commodity from a set of origins to a set of destinations, thereby satisfying the constraints given by the available resources and the requested demands (Hillier & Lieberman, 2014). Referring to the minimization problem discussed in this paper, origins and destinations are represented by the spatial units (for which the observed value is over and under the average respectively) while the constraints are such as to ensure that any imbalance in the phenomenon level across the units is eliminated after the transfers. The minimum cost required for eliminating this imbalance can be considered as an absolute measure for spatial concentration (i.e. expressed in the same unit of measurement as the cost), and as the basis for constructing a corresponding relative, α -dimensional and normalized, measure, whose values range from 0 to 1. This approach has been recently used by Lo Magno et al. (2017) for the measurement of seasonality (i.e. concentration over time). It has here been adapted to the spatial case in which the more complex structure relationships between the spatial units (when compared to the temporal case) raises several methodological considerations regarding the measurement of spatial concentration. This is beyond the conceptual differences between temporal and spatial concentration.

Consider a set \mathcal{R} of n regional (or any spatial) units $\mathcal{R} = \{1, 2, \dots, n\}$ for which the spatial concentration of the phenomenon of interest is to be evaluated. The observed values of the variable are represented by the vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$, and all the values are assumed to be non-negative. The total amount of the observed phenomenon is $T = \sum_{i=1}^n x_i$

and the observed average is $\mu = T/n$. We can assume the pattern $\mathbf{x}_0(T, n) = (T/n, T/n, \dots, T/n)$ to be the extreme case of an absence of spatial concentration which, whilst agreeing with the notion of entropy (Vranken, Baudry, Aubinet, Visser, & Bogaert, 2015; Wilson, 1970, 2010), would correspond to an even distribution of the phenomenon in the area under consideration. The proposed measure of spatial concentration can be derived as the minimum cost, which is required to transform a generic n -dimensional pattern \mathbf{x} , with a total amount T , into a pattern $\mathbf{x}_0(T, n)$. This will occur where the pattern transformation occurs by transferring the phenomenon from over-average regions to under-average regions. Further, each transfer will have a cost which depends on the amount transferred and the spatial location of the regions of origin and destination.

The challenge of locating that minimum cost can be treated as a particular case of the well-known transportation cost problem; this is a classical problem in linear programming (Ferguson & Dantzig, 1955; Hillier & Lieberman, 2014). The unitary transportation costs, which are used to posit the problem, are given by a symmetric $n \times n$ cost matrix $\mathbf{C} = [c_{ij}]$, in which the generic element c_{ij} represents the cost for transferring one unit from region i to region j . Although the diagonal elements are not used in the definition of the problem, a value of zero can be imposed on all these elements for mathematical convenience. Furthermore, the off-diagonal elements must be all non-negative.

According to the approach proposed in this paper, a simple and natural candidate for the cost matrix may be given by a Shimbil matrix (Shimbil, 1951). Ideally with empirical applications, a spatial proximity function should depend on the underlying research aims in order to be able to capture the specificities of the economic sector or social phenomenon under consideration. This may be based on physical distance or on other parameters, such as a social distance or any quantity which measures the cost associated with the formation of a link between any two regions (Barthélemy, 2011; Getis, 2009). However, various generic properties will be presented which do not always depend on the cost matrix chosen, although the choice of this matrix would evidently have implications for various properties and results of the proposed index.

In order to define the transportation problem, the above average and below average regions need to be identified. The above average regions are given by the set $\mathcal{A} = \{i \in \mathcal{R} : x_i > \mu\}$, whereas the below average regions are represented by the set $\mathcal{B} = \{j \in \mathcal{R} : x_j < \mu\}$. For each region $i \in \mathcal{A}$, the surplus $a_i = x_i - \mu$ is defined; similarly a deficit $b_j = \mu - x_j$ is defined for each region $j \in \mathcal{B}$. The value a_i represents the amount which has to be transferred from region i to the other regions in \mathcal{B} , while b_j is the amount which region j has to receive from the set of regions in \mathcal{A} . It holds that



$\sum_{i \in \mathcal{A}} a_i = \sum_{j \in \mathcal{B}} b_j$, where each member in this equality represents the amount which has to be transferred in order to achieve a pattern of *no spatial concentration* $\mathbf{x}_0(T, n)$.

Following the proposal by Lo Magno et al. (2017) in the context of seasonality, the transportation problem for measuring concentration can be defined as:

$$\begin{aligned} \text{minc} &= \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} c_{ij} t_{ij} \\ \text{s.t.:} \\ \sum_{j \in \mathcal{B}} t_{ij} &= a_i, \quad \forall i \in \mathcal{A} \\ \sum_{i \in \mathcal{A}} t_{ij} &= b_j, \quad \forall j \in \mathcal{B} \\ t_{ij} &\geq 0, \quad \forall i, j: i \in \mathcal{A} \wedge j \in \mathcal{B}, \end{aligned} \quad (2)$$

where c is the total cost of eliminating spatial concentration and the $|\mathcal{A} \times \mathcal{B}|$ variables t_{ij} , representing the transfers, are the decision variables of the problem. The first set of $|\mathcal{A}|$ constraints ensures that the surplus a_i is transferred from each region $i \in \mathcal{A}$; the second set of $|\mathcal{B}|$ constraints ensures that the deficit b_j is received by each region $j \in \mathcal{B}$; and the third set of $|\mathcal{A} \times \mathcal{B}|$ constraints guarantees that all the transfers are non-negative. An optimal solution always exists for a transportation problem (Dantzig & Thapa, 2003), and this can be found by using the simplex algorithm (Dantzig, 1963).

The optimal solution is given by the $|\mathcal{A} \times \mathcal{B}|$ values t_{ij}^* and the corresponding minimum cost c^* , given \mathbf{x} and \mathbf{C} , represents an absolute measure of spatial concentration, namely:

$$S_A(\mathbf{x}, \mathbf{C}) = c^* = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} c_{ij} t_{ij}^*. \quad (3)$$

For the generic pattern $\mathbf{x}_0(T, n)$ the sets \mathcal{A} and \mathcal{B} are void, thus the problem (2) cannot be defined; in this case, the value of the index is conventionally set equal to 0, which represents its minimum value.

The proposed index follows a similar approach to that proposed by Morgan (1983), which in turn is based on the work by Jakubs (1981) in the context of spatial segregation. This context has long been characterized by a debate regarding the most appropriate way for measuring spatial segregation (Folch & Rey, 2016; Harris, 2016). In the context of seasonality, Lo Magno et al. (2017) demonstrated that, given \mathbf{C} and holding constant T , the maximum value of the index in (3) for an n -dimensional pattern is given by:

$$S_{\text{MAX}}(\mu, \mathbf{C}) = \mu \max_{i \in \mathcal{R}} \left\{ \sum_{j \in \mathcal{R}} c_{ij} \right\}. \quad (4)$$

Thus, the maximum value depends on the mean value of the phenomenon of interest and the maximum value of the row sum in the cost matrix.

In the context of seasonality, as discussed in Lo Magno et al. (2017), a particular cost structure was assumed to take into account the cyclical distance between time periods. This assumption determines an equal value of the row sum of the cost matrix for all the rows. That is, the maximum degree of seasonality is achieved when the totality of the phenomenon of interest is concentrated in only one month, independently of the month. Considering the Shimbel distance matrix as a cost matrix in the spatial case, the row sum of the matrix is known as the Shimbel index (Pirie, 1979); it is a concise measure of the degree of accessibility of each region when compared to another (Taaffe, Gauthier, & O'Kelly, 1996). Consequently, the maximum value of spatial concentration is achieved when the total amount of the phenomenon of interest is concentrated in the least accessible region, i.e. that most isolated. Indeed, the Shimbel index may be seen as a concise measure of the accessibility of the units with lower values, denoting a higher degree of accessibility of the unit under consideration (Pitts, 1965).



Having identified the value S_{MAX} , it is possible to derive a relative spatial concentration index:

$$S_R(\mathbf{x}, \mathbf{C}) = \frac{S_A(\mathbf{x}, \mathbf{C})}{S_{MAX}(\mu, \mathbf{C})}. \quad (5)$$

S_R is bounded in the interval $[0,1]$; it may be used for comparing the degree of spatial concentration in situations in which the total volume of the phenomenon of interest is different (eg. a comparison of the different phenomena related to the same country, a comparison of the same phenomenon for the same country over different time periods, and comparisons among countries).

For $\lambda > 0$ the following set of properties can be derived, which specifies the differences between S_A and S_R (Lo Magno et al., 2017):

- Property 1. $S_A(\lambda \mathbf{x}, \mathbf{C}) = \lambda S_A(\mathbf{x}, \mathbf{C})$, i.e., linearity according to the phenomenon
- Property 2. $S_A(\mathbf{x}, \lambda \mathbf{C}) = \lambda S_A(\mathbf{x}, \mathbf{C})$, i.e., linearity according to costs
- Property 3. $S_R(\lambda \mathbf{x}, \mathbf{C}) = S_R(\mathbf{x}, \mathbf{C})$, i.e., scale invariance, according to the phenomenon
- Property 4. $S_R(\mathbf{x}, \lambda \mathbf{C}) = S_R(\mathbf{x}, \mathbf{C})$, i.e., scale invariance according to the cost matrix.

Properties 1 and 2 state that scale transformations of \mathbf{x} or \mathbf{C} multiplicatively affect S_A by the same scale transformation coefficient λ . On the contrary, Properties 3 and 4 reveal that S_R is insensitive to the scale transformations. From Property 3, it follows that, similar to many other inequality indices, that which effectively affects S_R is the *share* of the phenomenon located in each region, namely the relative – and not the absolute distribution – of the phenomenon. Finally, from Property 4 it can be shown that, although many cost matrices could be chosen, S_R does not change for all the infinite number of matrices, which can be obtained by multiplying a given matrix \mathbf{C} by λ . Demonstrations of these properties may be found in Lo Magno et al. (2017).

In order to facilitate the calculation of the proposed index in the spatial case, the authors of this paper have developed an R-package, called “spatialct”. This allows for the calculation of S_A and S_R indices for the distribution of a spatial variable, given a user-defined cost matrix. Alternatively, the software can receive an adjacency matrix as an input and generate the Shimbel matrix, the latter which can be used for the computation of the indices. The package is open-source and it can be freely downloaded from <https://sourceforge.net/projects/spatialct>.

4 | EXPLORING THE INDEX

Several patterns will be presented in this section to explore how the proposed index evaluates spatial concentration. These examples have been organized into four pattern sequences, each highlighting a particular aspect of the index. Each pattern in a sequence can be conceived as one which is derived from the previous one by transferring units among regions. Useful comparisons can, therefore, be made among the patterns in a sequence, as the total number of the observed phenomena is constant across the patterns.

The same topological structure was used for all the patterns. In this structure there are 37 regions, each with the same regular hexagonal shape. The cost matrix used in the following application is the Shimbel distance matrix, which can be derived directly from the adjacency matrix. Thus, the distance between two regions is defined as the minimum number of regions connecting the two regions. The structure of this cost matrix would facilitate a direct comparison of the results deriving from the proposed index with those deriving from other approaches proposed in the literature (Guimarães et al., 2011), and for direct comparison with the standard *Moran's I* index of spatial autocorrelation. However, it is well known in the literature that the selection of an appropriate spatial weight matrix is crucial for the results, and there exists an extensive debate regarding this topic (see, e.g., Anselin & Bera, 1998; Anselin, 2002; Getis, 2009; Griffith, 1996), albeit beyond the scope of this study. Nonetheless, the implementation of the proposed index



will permit a choice of the appropriate cost matrix, which should depend on the context under analysis. Indeed, this issue is complementary to that related to the specification of the spatial weight matrix, and several approaches have been proposed for its estimation (Ahrens & Bhattacharjee, 2015; Beenstock & Felsenstein, 2012; Chen, 2012; Getis & Aldstadt, 2004; Harris, Moffat, & Kravtsova, 2011).

Departing from these premises, the first sequence to be discussed is displayed in Figure 3, where the total amount of a hypothetical phenomenon (e.g., number of firms) is equal to 4,921 and is distributed among the regions. The distribution of the phenomenon in this sequence is gradually concentrated towards the regions at the center. The pattern displayed in Figure 3.a shows an equal distribution of the phenomenon under analysis among the 37 regions. This is clearly a case in which any concentration is absent and, indeed, the value for S_R is zero. The second pattern, Figure 3.b, has been obtained by equally transferring all the units from the peripheral regions to the inner regions. The total amount in the resulting pattern is equally distributed among the non-zero regions. As the same amount is now concentrated over a narrower area, the value for S_R increases to 0.184. The pattern displayed in Figure 3.c has once again been obtained by narrowing the area, in which the phenomenon is concentrated. S_R for this pattern rises to 0.367. Finally, the pattern in Figure 3.d is one in which the total amount of the phenomenon under analysis has been concentrated in only one region; $S_R = 0.604$ for this pattern.

The sequence of patterns displayed in Figure 3 demonstrates that S_R increases as the phenomenon progressively concentrates in an ever narrowing area. This would seem to be a basic requirement for a spatial concentration index, although, as will be seen later, this does not conclude the analysis. On the other hand, the behavior of H_5 and *Moran's I* is far from clear. There is a decrease in the concentration values, as measured by H_5 from pattern 3.a. to pattern 3.b., and a fluctuating behavior of *Moran's I* from pattern 3.a. to pattern 3.d. The only index showing an increase in concentration is H , which however is insensitive to the spatial distribution of the phenomenon under analysis. That is, any permutation in space of the proposed pattern would produce the same value as expressed by the H index.

All the pattern sequences in Figure 4 refer to a total amount of 1, and the phenomenon for each pattern is always concentrated in only one region. In this sequence, the non-zero region gradually approaches the center from the

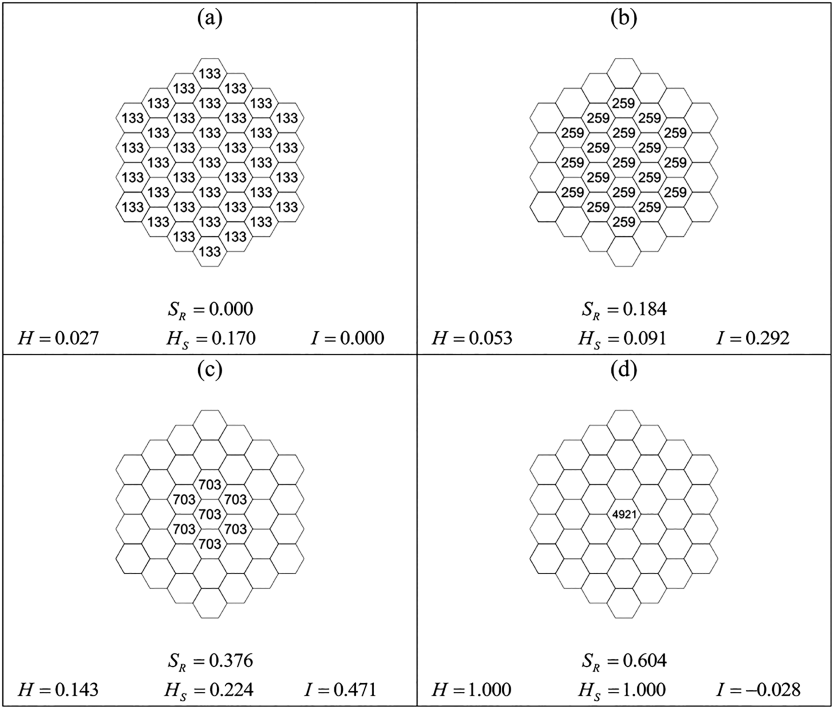


FIGURE 3 Sequence of progressively concentrated patterns

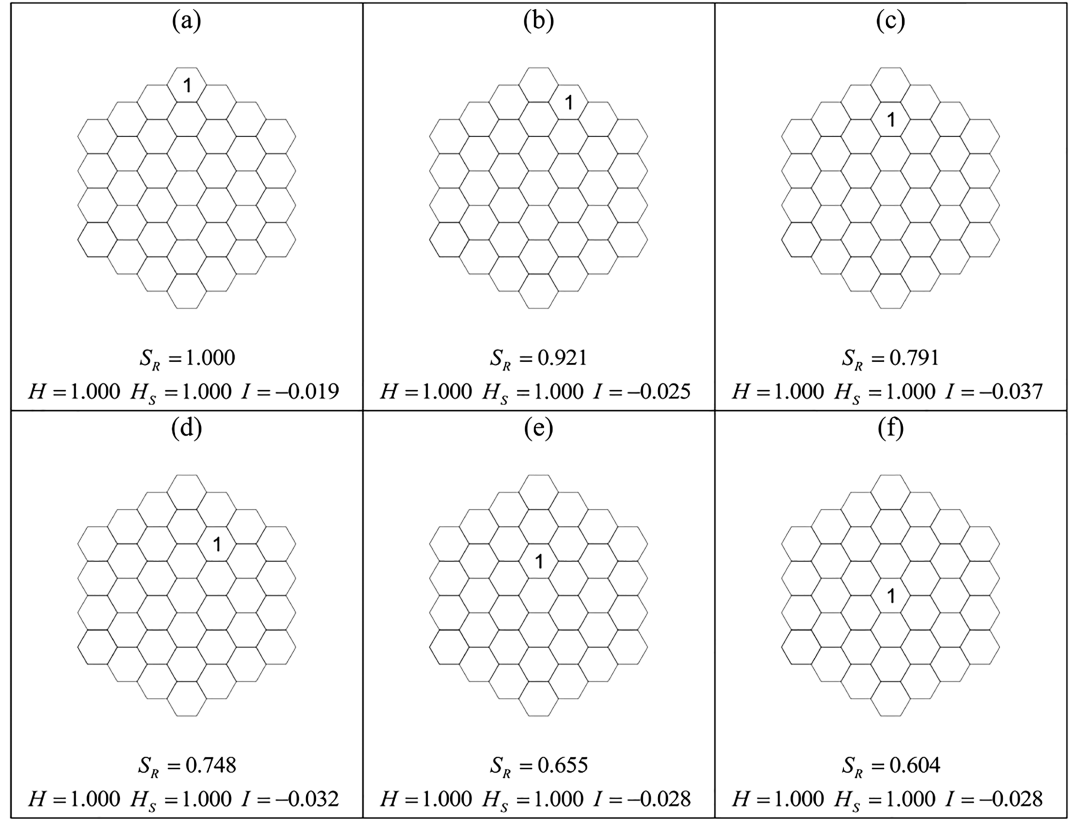


FIGURE 4 Sequence of patterns with a unique non-zero region approaching from the periphery to the center

periphery. Although the phenomenon is always concentrated in only one region, the value for S_R gradually decreases from 1 to 0.604. This behavior reveals an apparent counter-intuitive aspect of S_R , which is markedly differentiated from that temporal. The latter, discussed in Lo Magno et al. (2017), due to the particular structure of the cost matrix, obtains its maximum value when the phenomenon is concentrated in only one month, independently from the month. Thus, also in the spatial case it could be reasonable to expect that the six patterns displayed in Figure 4 should all be examples of maximum concentration. This is what results, for example, from the analysis of H and H_S . Nevertheless, it must be considered that S_R is sensitive to the isolation of the regions. That is, the more remote is/are the region/s (in which the amount/s to be distributed is/are located), the higher will be the concentration of the distribution.

A metaphor can be used to clarify why this sensitivity may be a reasonable expectation for evaluating concentration. Let us suppose a policy maker aims to reduce the concentration of tourist arrivals, which are represented by the pattern in Figure 4.a. 100% of tourists (for example, 1 million tourists) chose only one region to visit, as represented by this pattern. In order to reduce tourist concentration, the policy maker will want to promote other tourist destinations through targeted marketing policies. It can be assumed that tourists have a utility function, which decreases when the destination region departs from the one that is most preferred. If the marketing effort is oriented at compensating for the decrease in the tourists' utility, then the total cost of eliminating tourist concentration would be greater for pattern 4.a, as compared with pattern 4.f.

As a complementary metaphor, the spillover effect can also be considered. Let us suppose that the pattern in Figure 4.a refers to the concentration of an industry, whose total revenues are concentrated in the northeast region of a given country. The northeast region is isolated in that it is not well connected to other regions; therefore, fewer regions would benefit from a spillover effect. A clearer explanation is provided by the pattern in Figure 4.f, in which



the industry is concentrated in the most accessible region and, consequently, more regions would benefit from a spill-over effect. Extending the metaphor, it can be said that the best place for positioning a lamp is in the center of the room!

Formal results will now be introduced with the aim of improving the exposition of the aforementioned. Let us calculate the value for S_A in the case of concentration in only one region r . This value would be:

$$S_{Ar} = \mu \sum_{j=1}^n c_{rj}, \quad (6)$$

because an amount μ has to be transferred from region r to all the other regions. Thus, the value of S_{Ar} is obtained by multiplying the amount μ to be transferred, multiplied by the sum of the costs from the origin r to all the other $n-1$ regions.

The value for S_R in the case of concentration in only one region r is obtained by dividing (6) by its maximum isolation measure for region r can be defined as the average unitary cost of transfer from r to all the other regions. This is the average of all the unitary cost in the r^{th} row of the cost matrix:

$$\omega_r = \frac{\sum_{j=1}^n c_{rj}}{n}. \quad (7)$$

This result can be used to reformulate S_{Rr} in terms of isolation values ω_r , which can be obtained for all the regions:

$$S_{Rr} = \frac{\omega_r}{\max_{i \in \mathcal{R}} \{\omega_i\}}. \quad (8)$$

Thus S_{Rr} can be interpreted as the relative isolation of r , compared to the maximum isolated region. The values for S_{Rr} are provided in Figure 5. They regard the 37 hexagonal regions of the structure which was used in the examples throughout this paper.

Figure 6 shows a sequence of patterns in which the two non-zero regions approach each other. Another main feature of this sequence is that non-zero regions always have a value of 1 regarding their relative isolation. It can be observed in all these sequences that S_R increases (i.e. higher spatial concentration) as the two non-zero regions approach each other, while the average level of S_R depends on the relative isolation of the regions. Also, neither H nor H_5 is able to measure changes occurring to the degree of spatial concentration.

The final sequence to be discussed in this section is Figure 7, representing two regions, each with the same amount of phenomenon of interest. The latter are initially located in the extremes of the diameter of the spatial

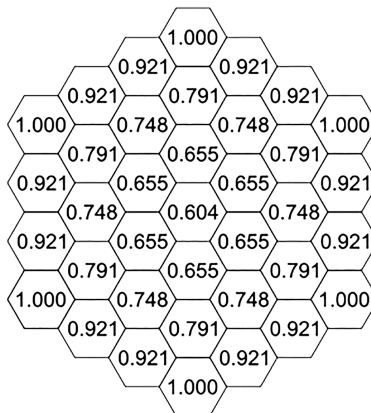


FIGURE 5 Values of relative isolation S_{Rr} for each region in a 37 units hexagonal grid

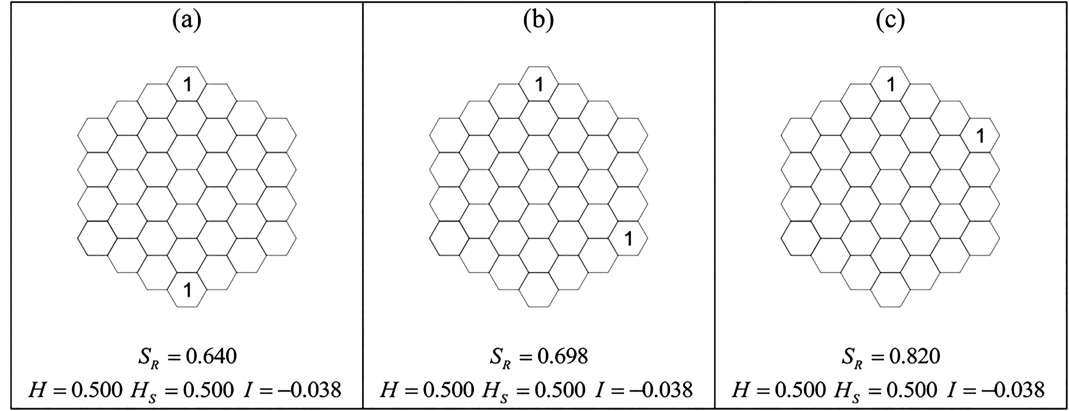


FIGURE 6 Approaching non-zero regions with a relative isolation value of 1

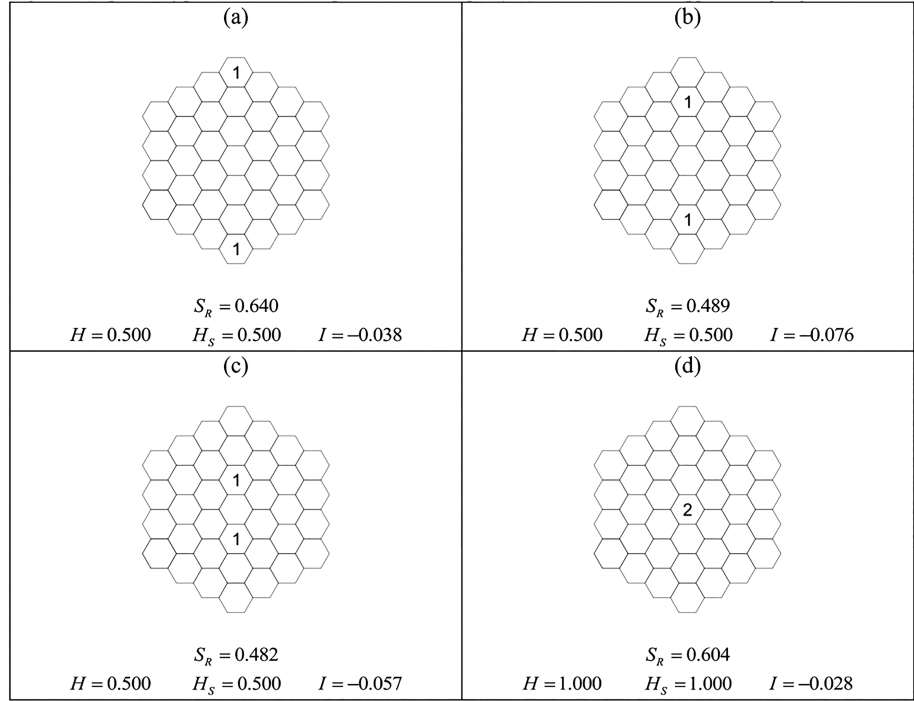


FIGURE 7 Sequence of patterns where the phenomenon is equally distributed over two approaching regions

structure. They then progressively approach each other until they collapse into a single region. Note that S_R always decreases with the exception of the final step. This is another counter-intuitive aspect of the index, and it can be explained by highlighting that there are two forces determining the value of the index:

- an increase in (non-spatial) concentration (i.e. a more unequal, non-spatial, distribution of the phenomenon), which has a positive influence on the index and
- a departure from the peripheral to internal regions, which has a negative influence on the index. This is due to the lower degree of the internal regions' isolation, compared to those in the periphery.



Each observed value for S_R is an admixture of these two components: the non-spatial distribution of the phenomenon under investigation, and its spatial distribution across the regions under analysis.

5 | EMPIRICAL APPLICATION

In order to illustrate the behavior of the S_R index, compared to other traditional measurements in real data situations, an analysis of the NUTS3 level for three European countries – France, Spain and Italy – will now be presented for illustrative purposes; this includes a selection of economic indicators (GPD and hotel bed places).

The data was obtained from the Eurostat databank, and the two indicators were selected for a variety of reasons. The spatial concentration of per-capita GDP was analyzed in various contexts (Yamamoto, 2007), especially before and after European monetary union, as documented in the wider literature pertaining to convergence among member states (Arbia, 2001; Geppert & Stephan, 2008; Rey & Janikas, 2005). An evaluation of the effectiveness of European policy was also undertaken (Bosker, 2009; Crescenzi, 2009). Similar considerations also hold for the sector of tourism, in which the topic of spatial concentration of tourism activities has been at the center of research agenda since the very beginning of tourism studies (Oppermann, 1993; Pearce, 1998). Spatial concentration of tourism activities continues to attract the attention of researchers (Lau, Koo, & Wu, 2019; Luo & Yang, 2013; Sarrión-Gavilán, Benítez-Márquez, & Mora-Rangel, 2015; Yang & Wong, 2013) due to the implications of a destination management perspective (Mawson et al., 1995).

The three countries of France, Spain and Italy have been selected on account of their similarity, and thus comparability, in terms of population, and economic and political contexts. Beyond providing an application to relevant issues – GPD is one of the most widely used measures of welfare, and tourism is one of the fastest growing industries in the world – the aim of Section 5 is to demonstrate how the proposed approach can perform in real data applications. This is done by examining territorial units which are characterized by different topological structures, and in which marked differences emerge as a result of a comparison of traditional indices and the proposed index.

Departing from these premises, the non-spatial distribution of per-capita GDP at NUTS3 level shows a similar degree of inequality, as measured by the Gini index. The distribution of per-capita GDP at current prices in France, Spain and Italy in 2014 is reported in Figure 8. Here, classes have been determined according to quintile criterion. From a spatial perspective, the results in Figure 8 show marked differences in the distribution of per-capita GDP among these three countries. Both Spain and Italy display a clear north–south axis, with higher values in the Northern provinces and values below average for southern provinces. Excluding the province of Rome, all the wealthiest provinces are located in the northern part of the country, whereas all the southern provinces belong to the fourth and fifth distribution quintiles. Similarly, all the south-west provinces of Spain display very low values of per-capita GPD, whereas all the richer provinces are located in the north-eastern part of Spain, excluding the province of Madrid.

However, the spatial distribution of per-capita GDP regarding France appears to be less concentrated, with provinces belonging to the first and fifth quintiles and which are well distributed all over the country. As a first step, the weighted Herfindahl index (H_S) (as proposed by Guimarães et al., 2011, and described in (1)), and S_R were derived, by using an adjacency matrix and the Shimbil matrix respectively in order to facilitate a direct comparison of the results. Thereafter, the proposed concentration index (S_R), together with other potential candidates concentration indices (namely *Gini* and H_S), was plotted for all the years under consideration (analytical results are reported in tables A.1–A.3 in Appendix). These results demonstrated that, whilst the three countries exhibited similar values of inequality in terms of the Gini index (non-spatial measure), they revealed marked differences when the spatial dimension considered. Indeed, the relative spatial concentration index highlights a higher spatial concentration of per-capita GDP regarding Italy, followed by Spain; France exhibits the lowest values of the spatial concentration of per-capita GDP. In this example, the ability of the index S_R to take into account the geographical dimensions of the phenomenon being analyzed (given a similar, non-spatial distribution, as shown by the Gini index) seems very compelling.

By considering the temporal dimension, all the considered indices (*Gini*, H_S and S_R) highlight an increase in the concentration of per-capita GDP. However, differences between the various indices can also be observed. Indeed, an

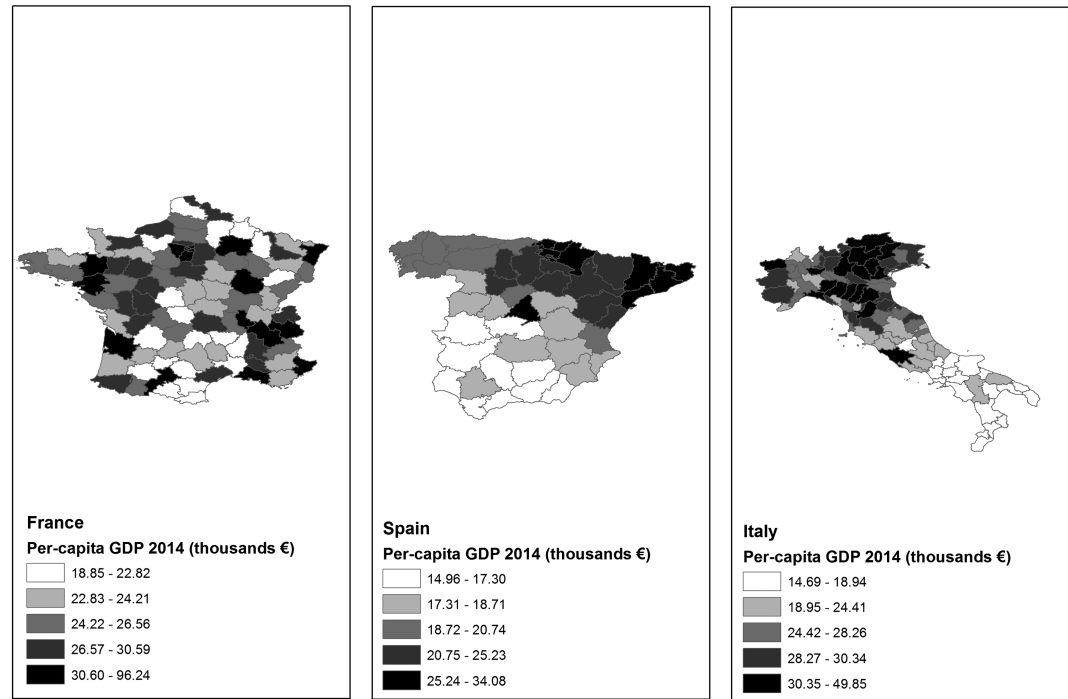


FIGURE 8 Distribution of per-capita GDP at current market prices for France, Spain and Italy at NUTS3 regional level in 2014

increase in concentration between 2006 and 2007 and 2008–2014 can be observed in the Gini and the S_R indices, whereas H_S does not reveal marked variations for the time period under consideration (Figure 9).

Let us now consider a second indicator available at the NUTS 3 level for Italy, Spain and France, namely, hotel beds per 1,000 residents. In this case, the non-spatial distribution indicator, as measured by the Gini index, displays a similar degree of inequality for France and Spain, whereas a higher degree of inequality characterizes its distribution in Italy. However, an inspection of the maps in Figure 10 produces a different situation to the spatial distribution of per-capita GDP: there seems to be a more spatially-concentrated distribution of bed places in France, with higher values in the southern part of the country, whereas, less spatial concentration seems to characterize Spain and Italy, as compared to the distribution of per capita GDP.

By considering the results in terms of non-spatial and spatial concentration indices, as reported in Figure 11, several comments can be made (the analytical results are reported in Tables A.4-A.6 in the *Appendix*). First, a much higher Gini index value relating to the distribution of hotel beds in Italy, Spain and France may be observed, as compared to the distribution of per-capita GDP. The Gini index values (relatively stable from year to year), as stated above,

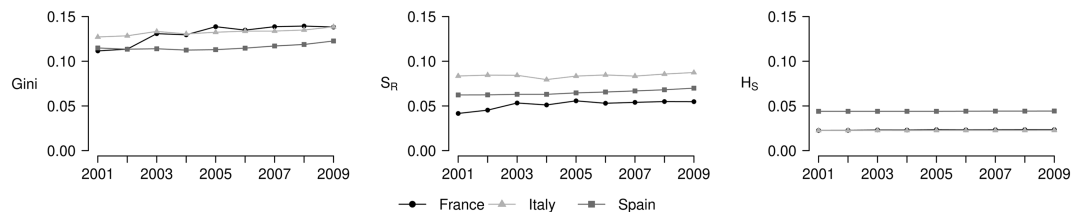


FIGURE 9 Values of G_{ini} , S_R , and H_S for per-capita GDP in France, Italy and Spain, by NUTS3 regions, years 2006–2014

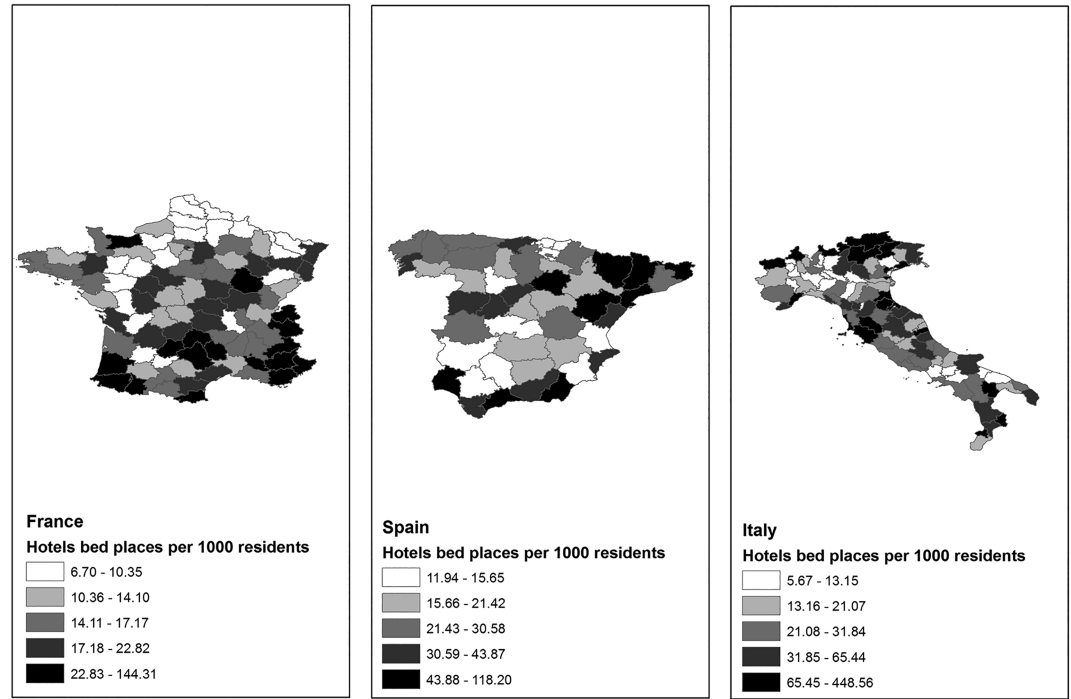


FIGURE 10 Distribution of hotels bed places per 1000 residents for France, Spain and Italy at NUTS3 regional level in 2011

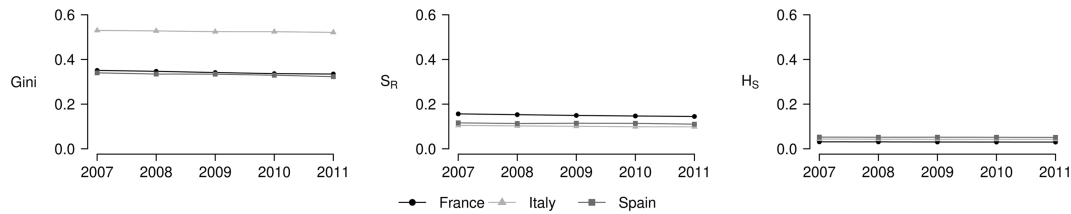


FIGURE 11 Values of G_{ini} , S_R , and H_S for hotel beds per 1,000 residents in France, Italy and Spain, by NUTS3 regions, years 2007-2011

approximate to those of Spain and France, whilst being much higher for Italy. However, when the spatial dimension is taken into account, of the S_R index produces different results in the analysis: Italy is characterized by a lower degree of spatial concentration of hotel beds, whereas France exhibits a much higher degree of spatial concentration. None of the non-spatial measures, nor the weighted Herfindahl (H_S) index would have provided such insights.

In order to shed light on how the proposed index performed in the applications under consideration, and to separate the non-spatial component from that spatial, 50,000 permutations of the observed distribution across the regions under study were generated, and the value of S_R calculated for each distribution. This approach permits the evaluation of index behavior (given the same non-spatial distribution of the phenomenon), according to differences only in its spatial distribution across the regions. The distributions of the 50,000 index values for both per-capita GDP and for hotel beds for France, Spain and Italy are reported in Figures 12 and 13. Also included are the empirical values of the index for the last available year (2014 and 2011 for per-capita GPD and hotel beds respectively).

The S_R values for per-capita GPD vary from 0.0197 to 0.0676 for France with an empirical value equal to 0.0548; from 0.0180 to 0.0634 for Spain with an empirical value equal to 0.0699; and from 0.0100 to 0.0535 for Italy with an

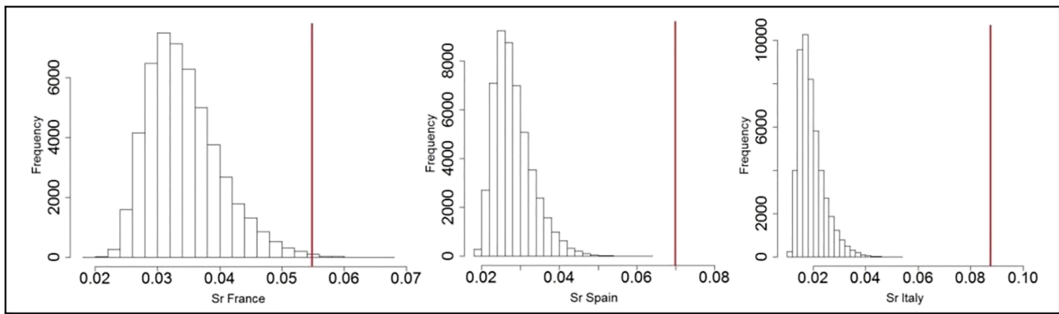


FIGURE 12 Values of S_R regarding the distribution of per-capita GDP for France, Spain and Italy corresponding to 50,000 permutations of the observed distribution over the NUTS3 regions, 2014

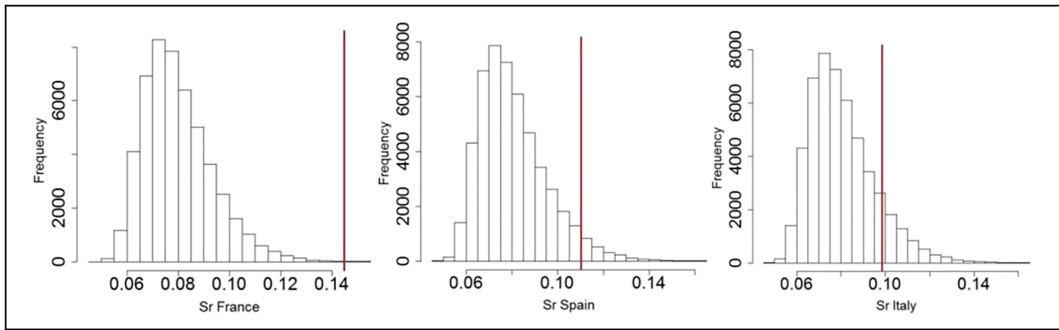


FIGURE 13 Values of S_R regarding the distribution of hotel beds per 1,000 residents for France, Spain and Italy, corresponding to 50,000 permutations of the observed distribution over the NUTS3 regions, 2011

empirical value equal to 0.0874. Regarding the *hotel beds indicator*, the 50,000 S_R values ranged from: 0.0494 and 0.1510 for France with an empirical value equal to 0.1447; 0.0491 and 0.1631 for Spain with an empirical value equal to 0.1102; and 0.0497 and 0.2632 for Italy with an empirical value equal to 0.0984.

These results easily lead to the conclusion that the empirical distribution of per-capita GDP across Italian provinces is much more spatially concentrated, compared to all the other situations; the findings could be verified by randomly assigning the observed distribution to the various provinces. That is, none of the randomly selected 50,000 distributions produced index values greater or equal to 0.0874. This highlights the fact that, given the non-spatial distribution indicator and the cost matrix, such a degree of spatial concentration is unlikely to be random, with higher values of the indicator concentrated in the northern regions and lower values in the southern regions of Italy. Similar considerations can be made regarding Spain. On the other hand, the observed spatial distribution indicator for France could be determined by random factors, given that the empirical value of S_R within the range in which its value is more probable.

The situation changes when the second indicator is considered, namely hotel beds per 10,000 residents. In this case the empirical value observed for France is quite far from the most probable values, which would have been observed by chance. That is, the observed value of 0.1447 highlights the high degree of spatial concentration of hotel bed places in southern regions compared to the northern ones. A different situation can be observed for Spain and even more so for Italy, in which the empirical values fall within the range of the simulated values, by highlighting a low degree of spatial concentration of hotel bed places in these countries.

Finally, using the same indicators (per-capita GDP and hotel beds), S_R , H_S and Moran's I indices were compared for the same 50,000 permutations in terms of their correlation and relationships; the results are displayed in Figures 14 and 15 and Table 2. The results reveal the scant existence of any relationship between S_R and H_S in all the considered



situations, whereas a certain degree of correlation appears between S_R and Moran's I , and, most of all between H_S and Moran's I . Nonetheless, the situations in which Moran's I approximates to values close to zero (no spatial auto-correlation), which are associated to high values of S_R , are not negligible. This result highlights differences in the behavior of the two indices in different empirical situations. Furthermore, the inadequacies of Moran's I for the measurement of spatial concentration have been already reported by several authors (e.g., Arbia, 2001; Guimarães et al., 2011). Finally, different types of relationships among the considered indicators are produced in relation to the different country under consideration (i.e. different structure of the cost matrix), as well as in relation to the non-spatial distribution of the phenomenon for the same country (i.e. keeping constant the cost matrix), thus highlighting an influence of both these components on the behavior of the indices.

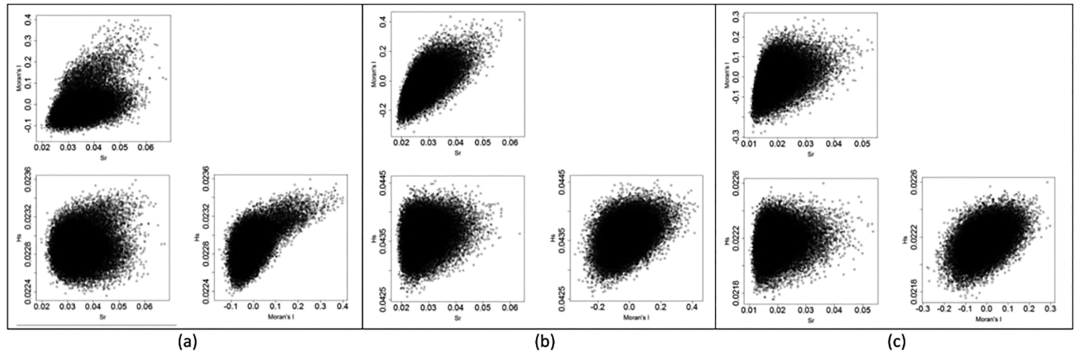


FIGURE 14 Values of S_R , H_S and Moran's I regarding the distribution of per-capita GDP for France (a), Spain (b) and Italy (c), corresponding to 50,000 permutations of the observed distribution over the NUTS3 regions, 2014

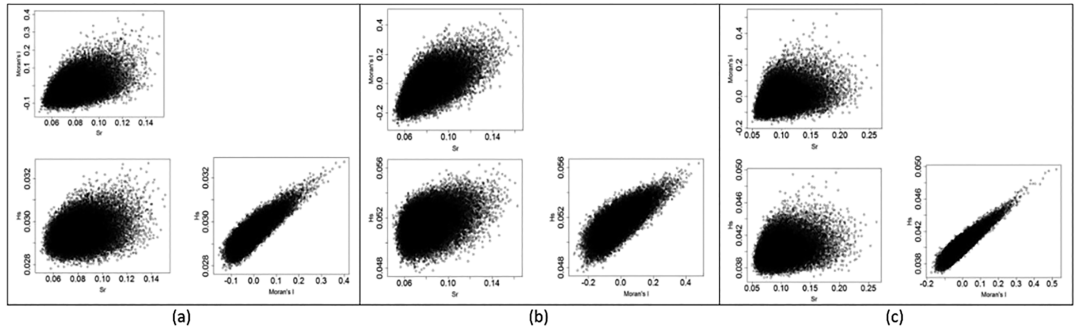


FIGURE 15 Values of S_R , H_S and Moran's I regarding the distribution of hotels bed per 1,000 residents for France (a), Spain (b) and Italy (c), corresponding to 50,000 permutations of the observed distribution over the NUTS3 regions, 2011

TABLE 2 Correlation coefficients among S_R , H_S and Moran's I regarding the distributions of per-capita GDP in 2014 and hotel beds in 2011 corresponding to 50,000 permutations of the observed distribution over the NUTS3 regions for France, Spain and Italy

	Per-capita GDP						Hotel beds					
	France		Spain		Italy		France		Spain		Italy	
	S_R	I	S_R	I	S_R	I	S_R	I	S_R	I	S_R	I
I	0.414		0.641		0.397		0.450		0.598		0.323	
H_S	0.092	0.558	-0.009	0.378	0.176	0.453	0.230	0.848	0.381	0.752	0.276	0.921



These results point to the need for an enhanced understanding of the indices currently used for the measurement of spatial concentration, in terms of their formal properties and interpretation of the results. They may also demonstrate that different results are obtained in terms of concentration according to the different index chosen.

6 | CONCLUSIONS

In conducting an exploratory data analysis of spatial phenomena, it is important to select an array of different measures to ensure that as many possible dimensions as possible have been included. The results presented in this paper suggest that the new spatial concentration index adds an additional dimension to the work of analyzing spatial distributions. By highlighting the latter, this index will yield more illuminating insights into distributions, which merit further analytical investigation. The S_R index is characterized by clear, conceptual simplicity: it is capable of measuring the degree of spatial concentration as a function of the cost of redistributing the phenomenon of interest in order to achieve an even distribution across the regions. This total cost will depend on three components: the first depends directly on the (non-spatial) concentration of the phenomenon. Given the same spatial structure, the degree of concentration will depend on the distribution of the phenomenon across the regions with a maximum value achieved when the phenomenon is concentrated in only one region, more specifically, relating to the most isolated region. The second component of the index relates to the spatial structure of the units being analyzed. A higher isolation of the regions among each other, or for simplicity, a higher distance value among the various regions, will determine an increase in the spatial concentration due to the higher costs required in achieving an even distribution of the phenomenon of interest. This characterizes the intrinsic spatial nature of the S_R index that is also made explicit in the determination of the cost matrix, that is, the third main component of the proposed approach. Indeed, the criteria determining the costs associated with transfers of the phenomenon under analysis from one region to another may be determined according to various criteria (Getis, 2009). The latter could include distance, transportation costs (Combes & Lafourcade, 2005) or other criteria in relation to the phenomenon under investigation (Copus, 1999; Rakshit, Nair, & Baddeley, 2017). This characteristic provides greater flexibility to the index, rendering it suitable for a wide field of application. Moreover, the proposed approach could also mitigate the MAUP problem through an appropriate specification of interregional costs. As a consequence, further research is required in order to analyze the impact of different cost specifications on the index behavior.

The empirical application of the proposed index, as presented in this paper, has highlighted that spatial concentration strongly characterizes both the indicators under consideration, although with characteristics and intensity which vary according to the indicator (GDP and hotel beds) and the specific country being considered. Whilst the GDP of Italy, Spain and France share a degree of inequality as measured by non-spatial indices, the implementation of the proposed index revealed a much higher degree of spatial concentration in the case of Italy, as compared to Spain and France. This may demonstrate that regions with a similar level of GDP per capita may have marked differences in terms of their growth potential if they belong to a spatially concentrated context or not (Crescenzi, 2009). This, therefore, demonstrates the importance of an accurate measurement of spatial concentration for regional policy considerations and impact evaluation. Similarly, the phenomenon of tourism in France, Spain and Italy (i.e. same topological structure) displayed marked differences in terms of their spatial concentration of hotel beds, when compared to GDP per capita. Thus, the different driving forces of spatial concentration according to the phenomenon under consideration is highlighted. The availability of natural and cultural resources, as well as facilities and tourism services, encompass factors which may explain the degree of spatial concentration in tourism activity. This in turn may determine an increase in tourism pressure with consequences on the environment and the resident population (Koen, Postma, & Papp, 2018). Marked differences also emerged as regards the measurement of spatial concentration by comparing the results of S_R with those derived from traditional indices.

In conclusion, the theoretical and empirical evidence provided in this paper suggests that the inclusion of the spatial dimension in the measurement of concentration may well affect the results of research in this field, thereby capturing elements of spatial concentration in a more effective manner as compared with alternative approaches.



ACKNOWLEDGMENTS

We gratefully acknowledge the invaluable comments of the Editor, the anonymous Reviewers, and the participants of conferences and EU-REAL meetings in Coimbra, Jerusalem, Palermo, Seville, and Vienna. We are also grateful to Jane O Davies for her editorial support.

ORCID

Mauro Ferrante  <https://orcid.org/0000-0003-1287-5851>

REFERENCES

- Ahrens, A., & Bhattacharjee, A. (2015). Two-step Lasso estimation of the spatial weights matrix. *Econometrics*, 3(1), 128–155. <https://doi.org/10.3390/econometrics3010128>
- Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267. <https://doi.org/10.1111/j.1574-0862.2002.tb00120.x>
- Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah, & D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics* (pp. 237–289). New York: Marcel Dekker.
- Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Spatial analyses of crime. *Criminal Justice*, 4(2), 213–262.
- Arbia, G. (2001). The Role of Spatial Effects on the Empirical Analysis of Regional Concentration. *Journal of Geographical Systems*, 3(3), 271–281. <https://doi.org/10.1007/PL00011480>
- Arbia, G., & Piras, G. (2009). A new class of spatial concentration measures. *Computational Statistics & Data Analysis*, 53(12), 4471–4481. <https://doi.org/10.1016/j.csda.2009.07.003>
- Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1–3), 1–101. <https://doi.org/10.1016/j.physrep.2010.11.002>
- Beenstock, M., & Felsenstein, D. (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis*, 44(4), 386–397. <https://doi.org/10.1111/j.1538-4632.2012.00851.x>
- Bickenbach, F., & Bode, E. (2008). Disproportionality measures of concentration, specialization, and localization. *International Regional Science Review*, 31(4), 359–388. <https://doi.org/10.1177/0160017608319589>
- Bosker, M. (2009). The spatial evolution of regional GDP disparities in the 'old' and the 'new' Europe. *Papers in Regional Science*, 88(1), 3–27. <https://doi.org/10.1111/j.1435-5957.2008.00183.x>
- Carlei, V., & Nuccio, M. (2014). Mapping industrial patterns in spatial agglomeration: A SOM approach to Italian industrial districts. *Pattern Recognition Letters*, 40, 1–10. <https://doi.org/10.1016/j.patrec.2013.11.023>
- Chain, C. P., Santos, A. C. D., Castro, L. G. D., & Prado, J. W. D. (2018). Bibliometric analysis of the quantitative methods applied to the measurement of industrial clusters. *Journal of Economic Surveys* (on-line), 33(1), 60–84. <https://doi.org/10.1111/joes.12267>
- Chakravorty, S. (1996). A measurement of spatial disparity: The case of income inequality. *Urban Studies*, 33(9), 1671–1686. <https://doi.org/10.1080/0042098966556>
- Chen, Y. (2012). On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5(2), 65–72. <https://doi.org/10.1007/s12076-011-0076-6>
- Combes, P. P., & Lafourcade, M. (2005). Transport costs: measures, determinants, and regional policy implications for France. *Journal of Economic Geography*, 5(3), 319–349. <https://doi.org/10.1093/jnlecg/lbh062>
- Copus, A. (1999). *A new peripherality index for the NUTS III regions of the European Union*. ERDF/FEDER Study 98. 00/27/130. European Commission.
- Crescenzi, R. (2009). Undermining the principle of concentration? European Union regional policy and the socio-economic disadvantage of European regions. *Regional Studies*, 43(1), 111–133. <https://doi.org/10.1080/00343400801932276>
- Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton: Princeton University Press.
- Dantzig, G. B., & Thapa, M. N. (2003). Decomposition of Large-Scale Systems. In *Linear Programming. Springer Series in Operations Research and Financial Engineering*. New York: Springer. doi: 10.1007/0-387-21569-7_10
- Dubin, R. A. (1998). Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7(4), 304–327. <https://doi.org/10.1006/jhec.1998.0236>
- Durantón, G., & Overman, H. (2005). Testing for Localisation Using Micro-Geographic Data. *The Review of Economic Studies*, 72(4), 1077–1106. <https://doi.org/10.1111/0034-6527.00362>
- Durantón, G., & Overman, H. G. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science*, 48(1), 213–243. <https://doi.org/10.1111/j.1365-2966.2006.0547.x>



- Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *The Journal of Political Economy*, 105(5), 889–927. <https://doi.org/10.1086/262098>
- Ferguson, A. R., & Dantzig, G. B. (1955). The Problem of Routing Aircraft. *Aeronautical Engineering Review*, April, 14(4), 51–55.
- Folch, D. C., & Rey, S. J. (2016). The centralization index: A measure of local spatial segregation. *Papers in Regional Science*, 95(3), 555–576. <https://doi.org/10.1111/pirs.12145>
- Freeman, S., Grogger, J., & Sonstelie, J. (1996). The spatial concentration of crime. *Journal of Urban Economics*, 40(2), 216–231. <https://doi.org/10.1006/juec.1996.0030>
- Ganong, P., & Shoag, D. (2017). Why has regional income convergence in the US declined? *Journal of Urban Economics*, 102, 76–90. <https://doi.org/10.1016/j.jue.2017.07.002>
- Geppert, K., & Stephan, A. (2008). Regional disparities in the European Union: Convergence and agglomeration. *Papers in Regional Science*, 87(2), 193–217. <https://doi.org/10.1111/j.1435-5957.2007.00161.x>
- Getis, A. (2009). Spatial weights matrices. *Geographical Analysis*, 41(4), 404–410. <https://doi.org/10.1111/j.1538-4632.2009.00768.x>
- Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, 36(2), 90–104. <https://doi.org/10.1111/j.1538-4632.2004.tb01127.x>
- Griffith, D. A. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In S. L. Arlinghaus (Ed.), *Practical Handbook of Spatial Statistics* (pp. 65–82). Boca Raton, FL: CRC.
- Guimarães, P., Figueiredo, O., & Woodward, D. (2011). Accounting for neighboring effects in measures of spatial concentration. *Journal of Regional Science*, 51(4), 678–693. <https://doi.org/10.1111/j.1467-9787.2011.00723.x>
- Harris, R. (2016). Measuring segregation as a spatial optimisation problem, revisited: a case study of London, 1991–2011. *International Journal of Geographical Information Science*, 30(3), 474–493. <https://doi.org/10.1080/13658816.2015.1032973>
- Harris, R., Moffat, J., & Kravtsova, V. (2011). In search of 'W'. *Spatial Economic Analysis*, 6(3), 249–270. <https://doi.org/10.1080/17421772.2011.586721>
- Hillier, F. S., & Lieberman, G. J. (2014). *Introduction To Operations Research*. Boston: McGrawHill.
- Horner, M. W., & Murray, A. T. (2002). Excess commuting and the modifiable areal unit problem. *Urban Studies*, 39(1), 131–139. <https://doi.org/10.1080/00420980220099113>
- Jakubs, J. F. (1981). A distance-based segregation index. *Socio-Economic Planning Sciences*, 15(3), 129–136. [https://doi.org/10.1016/0038-0121\(81\)90028-8](https://doi.org/10.1016/0038-0121(81)90028-8)
- Koens, K., Postma, A., & Papp, B. (2018). Is Overtourism Overused? Understanding the Impact of Tourism in a City Context. *Sustainability*, 10(12), 4384. <https://doi.org/10.3390/su10124384>
- Kopczewska, K. (2018). Cluster-based measures of regional concentration. Critical overview. *Spatial Statistics* (in press), 27, 31–57. <https://doi.org/10.1016/j.spasta.2018.07.008>
- Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.
- Lafourcade, M., & Mion, G. (2007). Concentration, Agglomeration and the Size of Plants. *Regional Science and Urban Economics*, 37(1), 46–68. <https://doi.org/10.1016/j.regsciurbeco.2006.04.004>
- Lau, P. L., Koo, T. T., & Wu, C. L. (2019). Spatial Distribution of Tourism Activities: A Polya Urn Process Model of Rank-Size Distribution. *Journal of Travel Research*. <https://doi.org/10.1177/0047287519829258>
- Liu, Z. (2014). Global and local: Measuring geographical concentration of China's manufacturing industries. *The Professional Geographer*, 66(2), 284–297. <https://doi.org/10.1080/00330124.2013.784953>
- Lo Magno, G. L., Ferrante, M., & De Cantis, S. (2017). A new index for measuring seasonality: a transportation cost approach. *Mathematical Social Sciences*, 88, 55–65. <https://doi.org/10.1016/j.mathsocsci.2017.05.002>
- Luo, H., & Yang, Y. (2013). Spatial pattern of hotel distribution in China. *Tourism and Hospitality Research*, 13(1), 3–15. <https://doi.org/10.1177/1467358413508579>
- Marcon, E., & Puech, F. (2003). Evaluating the Geographic Concentration of Industries using Distance-based Methods. *Journal of Economic Geography*, 3(4), 409–428. <https://doi.org/10.1093/jeg/lbg016>
- Marcon, E., & Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, 62, 56–67.
- Mawson, J., Chisholm, M., Pack, A., Clewer, A., Sinclair, M. T., & Jones, M. R. (1995). Policy Review Section. *Regional Studies*, 29(6), 563–580. <https://doi.org/10.1080/00343409512331349183>
- McCann, P. (2016). *The UK Regional National Problem*. Oxford: Routledge/Taylor & Francis.



- Menon, C. (2012). The bright side of MAUP: defining new measures of industrial agglomeration. *Papers in Regional Science*, 91(1), 3–28. <https://doi.org/10.1111/j.1435-5957.2011.00350.x>
- Morgan, B. S. (1983). An Alternate Approach to the Development of a Distance-Based Measure of Racial Segregation. *American Journal of Sociology*, 88(6), 1237–1249. <https://doi.org/10.1086/227802>
- Ochojski, A., Polko, A., & Churski, P. (2017). Theoretical Foundations of Specialization, Agglomeration and Concentration. In K. Kopczewska, P. Churski, A. Ochojski, & A. Polko (Eds.), *Measuring Regional Specialization. A New Approach* (pp. 1–57). Cham: Palgrave Macmillan.
- Openshaw, S. (1983). *The modifiable areal unit problem*. Norwick: Geo Books.
- Oppermann, M. (1993). Tourism space in developing countries. *Annals of Tourism Research*, 20(3), 535–556. [https://doi.org/10.1016/0160-7383\(93\)90008-Q](https://doi.org/10.1016/0160-7383(93)90008-Q)
- Pearce, D. G. (1998). Tourist districts in Paris: structure and functions. *Tourism Management*, 19(1), 49–65. [https://doi.org/10.1016/S0261-5177\(97\)00095-2](https://doi.org/10.1016/S0261-5177(97)00095-2)
- Pirie, G. H. (1979). Measuring Accessibility: A Review and a Proposal. *Environment and Planning A*, 11, 299–312. <https://doi.org/10.1068/a110299>
- Pitts, F. R. (1965). A graph theoretic approach to historical geography. *The Professional Geographer*, 17(5), 15–20. https://doi.org/10.1111/j.0033-0124.1965.015_m.x
- Rakshit, S., Nair, G., & Baddeley, A. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22, 129–154. <https://doi.org/10.1016/j.spasta.2017.10.002>
- Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34(1), 121–162. <https://doi.org/10.1111/j.0081-1750.2004.00150.x>
- Rey, S. J., & Janikas, M. V. (2005). Regional convergence, inequality, and space. *Journal of Economic Geography*, 5(2), 155–176. <https://doi.org/10.1093/jnlecg/lbh044>
- Sarrión-Gavilán, M. D., Benítez-Márquez, M. D., & Mora-Rangel, E. O. (2015). Spatial distribution of tourism supply in Andalusia. *Tourism Management Perspectives*, 15, 29–45. <https://doi.org/10.1016/j.tmp.2015.03.008>
- Sen, A. (1973). *On Economic Inequality*. Oxford: Oxford University Press.
- Shimbel, A. (1951). Applications of matrix algebra to communication nets. *The Bulletin of Mathematical Biophysics*, 13(3), 165–178. <https://doi.org/10.1007/BF02478225>
- Sohn, J. (2014). Industry classification considering spatial distribution of manufacturing activities. *Area*, 46(1), 101–110.
- Taaffe, E., Gauthier, H., & O'Kelly, M. (1996). *Geography of Transportation*. Englewood Cliffs, NJ: Prentice-Hall.
- Vranken, I., Baudry, J., Aubinet, M., Visser, M., & Bogaert, J. (2015). A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. *Landscape Ecology*, 30(1), 51–65. <https://doi.org/10.1007/s10980-014-0105-0>
- Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.
- Wilson, A. G. (2010). Entropy in Urban and Regional Modelling: Retrospect and Prospect. *Geographical Analysis*, 42(4), 364–394. <https://doi.org/10.1111/j.1538-4632.2010.00799.x>
- Winston, W. L., & Goldberg, J. B. (2004). *Operations research: applications and algorithms* (Vol. 3). Boston: Duxbury press.
- Wong, D. (2009). The modifiable areal unit problem (MAUP). In A. S. Fotheringham, & P. Rogerson (Eds.), *The SAGE handbook of spatial analysis* (pp. 105–124). Los Angeles: Sage.
- Yamamoto, D. (2007). Scales of regional income disparities in the USA, 1955–2003. *Journal of Economic Geography*, 8(1), 79–103. <https://doi.org/10.1093/jeg/lbm044>
- Yang, Y., & Wong, K. K. (2013). Spatial distribution of tourist flows to China's cities. *Tourism Geographies*, 15(2), 338–363. <https://doi.org/10.1080/14616688.2012.675511>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ferrante M, Lo Magno GL, De Cantis S, Hewings GJD. Measuring spatial concentration: A transportation problem approach. *Pap Reg Sci*. 2020;99:663–682. <https://doi.org/10.1111/pirs.12485>