The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/1059-5422.htm

Cluster identification

A joint application of industry concentration analysis and exploratory spatial data analysis (ESDA)

Reinaldo Belickas Manzini
Department of Management, Escola Superior de Propaganda e Marketing, Sao Paulo, Brazil, and
Di Serio Carlos Luiz
School of Business Administration, Fundacao Getulio Vargas, Sao Paulo, Brazil

Received 9 January 2018
Revised 24 August 2018
Accepted 22 October 2018

Abstract

Purpose – This paper aims to contribute to the approaches based on traditional industry concentration statistics for identifying clusters by complementing them with the techniques of exploratory spatial data analysis (ESDA).

Design/methodology/approach – Using a sample with 34,500 observations retrieved from the social information annual report released by Brazil Ministry of Labor and Employment, the methodology was designed to make a comparison between the application of industry concentration statistics and ESDA statistics.

Findings – As the results show, the geographic distribution measures proved to be fundamental for longitudinal studies on regional dynamics and industrial agglomerations, and the local indicator of spatial association statistic tends to overcome the limitation of the industry concentration approach.

Research limitations/implications – In the period considered, due to economic, structural and circumstantial questions, activities linked to the transformation industry have been losing ground in the value creation process in Brazil. In this sense, the study of other industries may generate other types of insights that should be considered in the process of regional development.

Originality/value – This paper offers a critical analysis of empirical approaches and methodological advances with an emphasis on the treatment of special effects: spatial dependence, spatial heterogeneity and spatial scale. However, the regional dynamic presents a temporal dimension and a spatial dimension. The role of space has increasingly attracted attention in the analysis of economic changes. This work has identified opportunities for incorporating spatial effects in regional analysis over time.

Keywords Clusters, Exploratory Spatial data analysis, Industry concentration analysis

Paper type Research paper

Introduction

Many studies on industrial agglomeration and specialization, regional development, firm performance in clusters and on the economic, social and institutional processes derived from them have been produced over the past decades. According to Porter (1998), clusters are made up of interconnected companies and associated institutions linked by their common aspects or by their complementarities, whose colocalization fosters the formation and amplifies the value creation benefits that emerge from the networks of firms. Krugman (1991a, 1991b) relates competitive markets and agglomeration and describes two forces acting on this:

Competitiveness Review: An International Business Journal
Vol. 29 No. 4, 2019
pp. 401-415

1059-5422
DOI 10.1108/CR-01-2018-0001

(1) centripetal, when the structure of the market favors industrial agglomeration; and
(2) centrifugal, when competitive conditions dissuade industrial agglomeration.

Other authors, such as, for example, Delgado *et al.* (2014), incorporate among others, the role of local demand, the institutional tissue, industry structure and the social networks.

However, as described in the current literature, the traditional indicators of industry concentration analysis have significant limitations when applied to identify potential clusters. Consequently, the question on what level of industrial agglomeration should be considered to indicate the possible existence of a cluster, remains unanswered in accordance with O'Donoghue and Gleave (2004). Additionally, Perry (2005) argues that approaches for identifying clusters are far from achieving consensus given the ambiguity that surrounds the very notion of cluster.

The economic base theory, from where concentration indicators are derived, aims to explain the relationships between a given geographical region and others. Fundamentally, its core is to study and evaluate the impact of economic activities between regions, and most studies on clusters identification uses its statistics, more specifically, the location quotient (LQ), that compares a local economy against a benchmark economy to identify specializations in the former. The logic is clear: the more concentrated an economic activity is, the greater the probability of cluster occurrence. However, on which level of concentration? This answer, usually, depends upon the researcher perceptions. Despite the contributions from Krugman (1991a) and Audretsch and Feldman (1996a, 1996b) with the introduction of the Gini location coefficient (GL), the question about which LQ cut-off level remains unanswered.

Moreover, the LQ nothing reveals, directly, about the geographic proximity between the economic activities (Tartaruga and Sperotto, 2009), the regional dynamics over time, and the correlation of concentration (dispersion) of the variable analyzed between the spatial units in the study. In response to these limitations, Guillain and Le Gallo (2010) aimed to evaluate accurately the spatial distribution of activities. They analyzed the agglomeration patterns for 26 manufacturing and service sectors in Paris and its surroundings in 1999, in a thin spatial scale, i.e. communes (French municipalities). These authors sought to adopt a methodology that allowed the degree of spatial clustering to be measured and for economic sector location patterns in a discrete space to be identified.

This article aims to contribute to the approaches used for identifying clusters by complementing the traditional industry concentration statistics (e.g. the LQ) with the techniques of exploratory spatial data analysis (ESDA) that, by its turn, emphasize aspects of the interdependencies and spatial heterogeneity of the data to describe spatial distributions, discover patterns of spatial association (spatial clustering) and identify atypical observations (outliers).

At last, another question that deserves attention is the unit of analysis. Most of the studies, as Ferreira (2009) shows, delimit concentrations to municipalities and a few to the micro-regions, which, thus, limits the capacity to analyze the results because they reveal almost nothing about those concentrations that cross over geographic boundaries. According to the literature, no study related to traditional approaches for identifying clusters has hitherto adopted metropolitan regions as the unit of analysis. However, the ability to produce and polarize is a relevant characteristic of various metropolitan regions in the contemporary world. Therefore, these metropolitan regions are fundamental units of analysis for the study of regional dynamics (Lemos and Crocco, 2000). Thus, by using the metropolitan regions as the unit of analysis and the municipality as the lower spatial unit, the secondary objective of this work is to present a new perspective of analysis whose results allow valuable insights into agglomerate composition, trends and spatial extent. As a contribution, this different form of analysis can be used for defining economic development strategies and delimiting issues and areas to be addressed in further quantitative or qualitative studies.

Theory and empirical studies

It is important to note that several theoretical and empirical studies have sought to analyze the determinants of spatial clustering of activities (Fujita and Thisse, 2002; Rosenthal and Strange, 2004). So far, researchers have not reached an agreement on geographical boundaries to be considered (Rosenthal and Strange, 2001; Parr *et al.*, 2002; O'donoghue and Gleave, 2004). In this sense, the search for an appropriate measure for the agglomeration concept and a clear definition of its borders will be useful for improving the understanding of the concept of spatial clustering and its causes. Hence, this section first discusses the approach of industry concentration, whose origins go back to the theoretical framework of regional economics, next, presents ESDA and its premises and main statistics, and finally, the notion of metropolitan region (RM).

Industry concentration

The approach of industry concentration has its origins in the pioneering work of regional economics (Andrews, 1953; Isard, 1960). Its central indicator is LQ, which measures the ratio between the number of local jobs and the number of nationwide jobs attributed to a specific economic activity. In this manner, LQ shows the relative degree of concentration of an economic activity, and consequently, stimulates some discussions on the degree of specialization/diversification in a specific region. Therefore, regions with a high LQ indicate the presence of potential clusters (Isard, 1960).

However, LQ has some disadvantages. According to Martin and Sunley (2003), there is no agreement about the degree of spatial concentration of an industry group or sector that constitutes a cluster. Another disadvantage of LQ is that it does not supply information about the absolute size of local economic activities, which makes it possible to find regions with high LQ values for activities that have a reduced number of jobs.

To overcome these limitations, several authors such as Suzigan *et al.* (2001b), Fingleton *et al.* (2002) and Porter (2003) and more recently, Delgado *et al.* (2014, 2016) and Ketels (2017) have proposed use, besides LQ, other variables to better qualify the concentrations[1]. Important applications derived from Delgado's study are those made by Resbeut and Gugler (2016) and Mendoza-Velazquez (2017). However, at the end, as all the proposals starts from LQ, all of them still suffer from the same limitation: the LQ cut-off level.

Exploratory spatial data analysis

ESDA can be defined as a statistical study of phenomena that are manifest in space and so, its unit of analysis is location, area, topology, spatial arrangement, etc. To make the concept operational, observations are referenced in space, i.e. their locations are specified as points, lines or areas, which leads to two different types of spatial effect: dependence and heterogeneity. Dependence draws its point of reference from spatial data, as spatial autocorrelation, while spatial heterogeneity is related to spatial (or regional) differentiation and obeys the intrinsic singularity of each location (Anselin, 1994).

ESDA emphasizes the spatial aspects of data in the sense of their dependence and spatial heterogeneity. The purpose of its techniques, therefore, is to describe spatial distributions, to discover patterns of spatial association (spatial agglomeration), to suggest different spatial regimes or other forms of spatial instability (non-stationary) and to identify atypical observations (outliers). In this sense, ESDA relies on an integrated set of methods that aims for visualization, summarizing and investigating spatial patterns and relationships, and specifying statistical models and estimating parameters.

An ESDA central concept is the autocorrelation that refers to the spatial correlation between attributes of the same random variable at different locations in space (Bailey and Gatrell, 1995). Autocorrelation is defined as the property that the mapped-out data has when they exhibit an organized pattern (Upton and Fingleton, 1985) or when there is a systematic spatial variation in the values of spatial units (Cliff and Ord, 1981). Operationally, there are several indicators to estimate how much the observed value of an attribute in a specific spatial unit is dependent on the values of the same variable in neighboring spatial units. These indicators are divided into global and local, respectively, Global Moran's index and Local Indicator of Spatial Association (LISA).

The former is an overall measure of autocorrelation and indicates the degree of spatial association between space units. Positive values between 0 and +1 indicate to the direct correlation, and negative values between 0 and −1, the inverse correlation. The statistical significance level can be reached by both approaches, a pseudo significance test or an approximate distribution (for more details see Anselin, 1995). As a limitation, Global Moran's index ignores the existence of local patterns of spatial autocorrelation and can misleading results about the existence of spatial autocorrelation in the data. To overcome this limitation, Anselin (1995) proposed the use of the LISA statistics.

As Anselin (1995) argues, a LISA indicator is any statistic that satisfies two requirements: each observation should provide an indication of the extent of significant spatial clusters around this observation, and the sum of LISA indicators of all observations should be proportional to the global indicator of spatial association. Hence, there is a direct proportionality between the value of the global autocorrelation and the values of the local autocorrelations, showing thereby that the LISA indicators allow for decomposition of the global indicators in individual contributions, indicating regions of non-stationarity and identifying significant clusters of similar values around certain localities. The spatial autocorrelation is calculated from the product of the deviations in relation to the mean as a measure of covariance. Statistical significance is established by constructing a pseudo-empirical distribution, by permutation, as in the Global Moran's index (Anselin, 1995).

Guillain and Le Gallo (2010) pioneered the use of ESDA in this research field. In their work "agglomeration and dispersion of economic activities in and around Paris: an ESDA," the authors aimed to evaluate accurately the spatial distribution of activities. They analyzed the agglomeration patterns for 26 manufacturing and service sectors in Paris and its surroundings in 1999, on a thin spatial scale, i.e. communes (French municipalities). According to them, two dimensions of the agglomeration should be captured by an appropriate empirical methodology: concentration in a single spatial unit, and, the spatial distribution of the units in the study area (ARBIA, 2001).

In ESDA, autocorrelation indicators are complemented by position measurements as shown in Table I – an approach that proposes the joint use of spatial representation and visualization techniques with the classic measures of descriptive statistics such as mean, median, standard deviation, etc. (TARTARUGA, 2009).

Metropolitan regions

For Sassen (1991), the urbanization process has intensified after the Second World War through the formation of large metropolitan areas that are gradually becoming global

Measurement	Definition
Mean center	It is equivalent to the average from descriptive statistics and it is obtained by calculating the average of the coordinates of spatial units. Moreover, in descriptive statistics, it is possible to calculate the weighted mean center, which makes it possible to represent the spatial behavior of discrete variables such as employment, population, income, GDP, etc.,
Standard distance	It is equivalent to the standard deviation. It is a measure of the concentration degree or a measure of the spatial distribution's dispersion of the data around the mean center. A larger standard distance means a larger dispersion of the points around the respective spatial center; and, conversely, a smaller the standard distance, means a larger concentration of points around the center. Similarly, as for the mean center, it is possible to calculate the weighted standard distance, which allows for representing the spatial behavior of discrete variables
Standard deviation ellipse	It provides knowledge of the spatial dispersion in two ways: density (or compactness) and guidance. This information allows the researcher to know the nature of the data distribution in its asymmetry. Moreover, it is a technique that is very useful in longitudinal studies, such as the dynamics of concentration/distribution of income of a region in a certain period

Source: Prepared by authors

Table I. Major measures of the spatial distribution

cities, with the advent of the information age as from the 1980s. According to Markusen (1984), while economic activities are being concentrated in large urban areas, they tend to be attracted to or even characterized by the singularities of these areas. Lemos and Crocco (2000) consider that the region tends increasingly to be represented by large urban areas that constitute not only centers of attraction and growth but also places that guarantee the specificity of regional production in a globalized context.

The combination of urbanization, metropolitanization and conurbation processes gave rise to the metropolitan regions as physical and socioeconomic phenomena that might be institutionalized by government, aimed at the integrated management of the municipalities that are part of them (Freitas, 2009).

Method

The method performed five stages:

(1) sampling;
(2) application of industry concentration statistics;
(3) agglomerates selection;
(4) application of ESDA statistics; and
(5) results analysis.

Stage 1: sampling

The sample, with 34,500 observations, which includes jobs and firms, was built from the social information annual report[2] from three selection criteria:

(1) geographical (metropolitan regions of São Paulo, Campinas and Baixada Santista);
(2) sectorial (manufacturing[3]); and
(3) temporal (from the year 2006 to 2010).

Stage 2: application of the industry concentration statistics

At this stage, LQ and GL were calculated from a sector-spatial distribution matrix of employment, which allows the production of different types of specialization and localization measures. The LQ uses the following equation:

$$LQ_{ij} = {}^{P_{ij}}/_{P_{pj}} \left(0 \leq LQ_{ij} \leq +\infty\right)$$

where P_{ij} is the participation of economic activity j in the region I and P_{pj} is the participation of economic activity j in the reference region p. The participations are calculated as following:

$$P_{ij} = {}^{x_{ij}}/_{x_i}$$

where x_{ij} is the value of economic activity j in the region i and x_i is the total value from all activities considered in the region i.

$$P_{pj} = {}^{x_{pj}}/_{x_p}$$

where x_{pj} is the value of economic activity j in the region p and x_p is the total value from all activities considered in the reference region p.

GL is calculated from QL and uses the following equation:

$$GL = \left|1 - \sum_{k=1}^{k=n-1} (X_{k+1} - X_k).(Y_{k+1} + Y_k)\right|$$

where X corresponds to P_{pj}, Y corresponds to P_{ij} and n is the number of observations. It is important to note that LQ measures the ratio between the number of local jobs and the number of jobs in the region in studies attributed to a specific economic activity. GL enables the degree of dispersion of economic activities to be identified and selects those that presumably have a greater tendency to the spatial concentration; the closer it is to zero the more similar the regional distribution of economic activity relative to all activities.

Stage 3: agglomerates selection

Table II shows the criteria by which economic activities and municipalities were selected for subsequent application of the ESDA statistics. The "cut-off levels" proposed by Porter (2003), Delgado *et al.* (2014), Resbeut and Gugler (2016) and Mendoza-Velazquez (2017) are also shown. Although the purpose of these works is different from this, they are important references in this field of study.

The value of industrial transformation[4] (***VIT***) was incorporated into the selection criteria as an added-value variable of economic activities. It was originally calculated as following:

$$VIT = GVIP - IOC$$

where the gross value of industrial production (***GVIP***) corresponds to the value of sales by the local unit, plus the value of the transfer of products made for sale in other local units. The industrial operating costs (***IOC***) corresponds to the sum of costs directly involved in production[5] in the local unit, incurred in the year, except salaries and charges.

Variables	This paper	Porter (2003) and Delgado *et al.* (2014)	Resbeut and Gugler (2016)	Mendoza-Velazquez (2017)
Economic activity				
Value aggregation	VIT	Patenting growth	–	–
% of participation of economic activity employment in total MR employment	The municipality must have, at least, 0.5% of the total employment in the RM for the specified industry	The region must have, at least, 50% of the total employment in the state for the specified industry		
GL	≥0.40		>0.30	
Spatial unit	RM		State	
LQ	>2	≥1		1.5 < and <2.5
Number of firms	Minimum of 5 firms in the municipality for a specified industry	–	In LQ calculus, it is used as the denominator of the number of jobs	–

Source: Prepared by authors

Table II. Selection criteria of economic activities and municipalities

Stage 4: application of the exploratory spatial data analysis statistics

This stage was sub-divided in two analyzes:

1. territorial distribution; and
2. spatial association statistics.

The former considered the model proposed by Tartaruga and Sperotto (2009) to analyze the type of territorial distribution behavior for the selected economic activities. In the latter, the spatial autocorrelation statistics were applied according to the following logic. Given an MR, the estimation of the correlation of the total employment of a given economic activity i in a municipality j in relation to the average employment of its neighboring $n-1$ makes it possible to identify potential clusters in this MR, independently of its political-administrative division.

Concerning to ESDA statistics, neighbors are defined as "spatial units" that have a common border or those with a given critical distance between them (Anselin, 1993), while "neighborhood" is formalized in a matrix of spatial weights related to the set of spatial units.

The incidence of such clusters depends on the statistical significance of the spatial autocorrelation test, which, if positive, shows the existence of spatial spillovers between the contiguous municipalities in the MR. Thus, the existence of certain economic activities in an MR is explained not only by the specific attributes of established firms and the MR *per se* but also by the fact that firms located there are favored by the existence of the same economic activities in neighboring municipalities.

Such neighborhood advantages, overflow effects and chaining, arise from various types of cost reduction in the supply of inputs, regional labor market specialized training and ease access to relevant information and sharing of scale-intensive infrastructures such as transport. These external economies within a specific locality have their effects potentiated from the flow of trade between geographically close localities.

The positive externalities in space are defined, therefore, not only by the existence of a given economic activity in a municipality *j* but also by the capacity of contagion and overflow of the product of that economic activity located in *j* for the $n-1$ neighboring municipalities. Clusters, therefore, express geographic proximity as a centripetal force for regional development.

The analysis of spatial dependence in the data set was performed using the Moran[6] scattering model, which allows to visualize and interpret the linear association between each attribute value z_i (in this case number of jobs) in relation to the average values of the attributes of your neighbors z_m. The model yields four quadrants where each one represents a different type of association between z_i and z_m as shown in Table III.

As sustained below, Types 1 and 2 were considered as those more useful for identifying the potential clusters inside RMs that overflows municipalities. However, Types 3 and 4 can offer valuable information about potential clusters located inside municipalities.

Type 1 High-High (HH) expresses, for a given economic activity, a positive spatial correlation of two or more municipalities with a number of jobs above the data set average suggesting the existence of some economies of agglomeration such as the presence of specialized labor, supply infrastructure, innovation externalities, etc., spread by those municipalities and shared by firms located there. This type was considered a "regional cluster," i.e. an inter-municipality cluster.

Type 2 Low-Low (LL) also shows a positive spatial correlation of two or more municipalities but, differently from Type 1, with a number of jobs below the data set average, which suggests an under-representation of a given economic activity in those spatial units. Or, even, the existence of municipalities with a relevant but not statistically significant number of jobs because the correlation between neighboring under-represented municipalities (LL) has predominated over the correlation between the high value of the municipality and the low value of the average of its neighbors High-Low (HL) in the significance test. The latter, from a minimum number of jobs existing in the municipalities (Table II), was also considered as a "regional cluster."

Table III. Types of linear association

Type	Quadrant	Description	Characteristics	
1	HH	The high number of jobs with a high positive correlation with their neighbors	Positive spatial association	Both z_i and z_m are above average
2	LL	The low number of jobs with a high positive correlation with their neighbors		Both z_i and z_m are below average
3	HL	The high number of jobs with high negative correlation with their neighbors	Negative spatial association	High values are surrounded by low values representing positive z_i and negative z_m
4	LH	The low number of jobs with high negative correlation with their neighbors		Low values are surrounded by high values representing negative z_i and positive z_m

Source: Prepared by the authors

Type 3 (HL) reveals, in turn, the existence of a given economic activity located in only one municipality that was considered as a "local cluster," i.e. an "intra municipality cluster."

Finally, Type 4 Low-High (LH) can reveal two very different phenomena. The former reveals the geographic limits of clusters, indicating the restrictive and excluding nature of a given economic activity in space. In addition, the latter, reveals a phenomenon like Type 2 (HL), i.e. the existence of a given economic activity located in only one municipality, which does not reach the level of expected significance (H), but, by the other side, offers significance for the neighbor of low number of jobs (L). This type was also considered as a "local cluster."

Results

Results are presented by metropolitan regions. First are discussed those relating to industry concentration statistics, and then, those relating to ESDA[7]. Although three metropolitan regions and six economic activities were studied originally, considering the article objectives and the question involving its length, only the most representative results are presented, but, obviously, preserving the quality of the analysis. The potential clusters are presented in Table IV.

In a longitudinal analysis, from 2006 to 2010, PE and GL for manufacture of paints, varnishes, enamels, lacquers and related have proved stable. By its turn, the manufacture of furniture shows an unstable behavior in relation to PE, but a certain stability in relation to its level of concentration.

Relating to position measures, the mean center of manufacture of paints, varnishes, enamels, lacquers and related products, located in the district of Ipiranga, did not change in the period, while the standard-distance (circa 22 km) increased slightly (1.65 per cent). By its turn, the manufacture of furniture, whose mean center in 2006 was in the Northwest of Campinas, in 2010, moves to the central region of Paulínia, with a slight reduction (−5.81 per cent) of the standard distance (circa 11 km).

The directional distribution (standard deviational ellipse) of the selected economic activities is shown in Figure 1. From these measures, in addition to the central tendency, dispersion and directional tendencies, it is also possible to observe the concentration/dispersion dynamic of both, thus corroborating the findings of the standard distance analysis.

For manufacture of paints, varnishes, enamels, lacquers and related products in the metropolitan region of São Paulo, the Global Moran's index[8] equals 0.152288 (p-value =

Economic activity	Municipality	LQ	Number of firms	Number of jobs
Manufacture of paints, varnishes, enamels, lacquers and related products (RM of São Paulo)				
VIT: 5.9 (R$ b)	*Taboão da Serra*	3.0349	9	526
PE: 1.12%	São Bernardo do Campo	2.7428	14	3,028
GL: 0.4171	Mauá	2.7300	7	845
	Guarulhos	2.2693	66	2,813
Manufacture of furniture (RM of Campinas)				
VTI: 7.5 (R$ b)	Paulínia	11.4415	7	969
PE: 1.75%				
GL: 0.7053				

Notes: LQ: locational quotient (the year 2010); VIT: value of industrial transformation (Brazil-wide, the year 2010); PE: % of participation of economic activity jobs in total MR jobs (the year 2006 to 2010 average); and GL: gini locational (the year 2006 to 2010 average)
Source: Prepared by the authors

Table IV.
Potential clusters

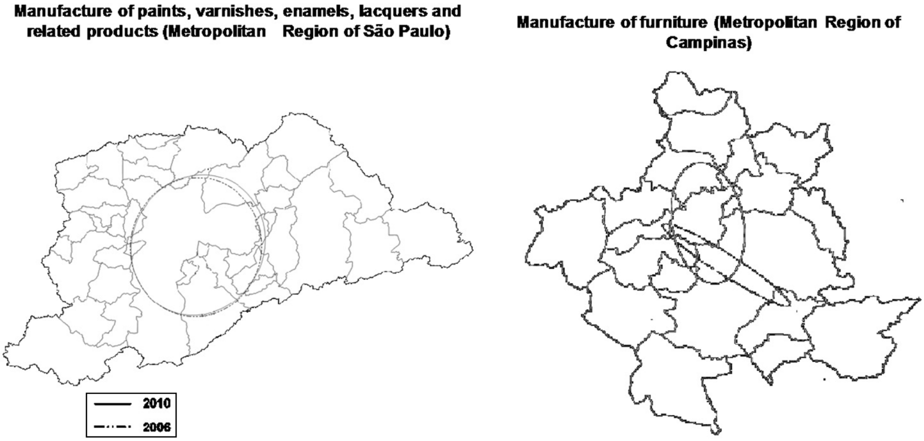

Figure 1. Directional distribution of the selected economic activities

Source: Prepared by the authors

0.003) and revealed the existence of spatial dependence between observed values suggesting the presence of spatial unit clusters. For manufacture of furniture in the metropolitan region of Campinas, the same index equals −0.093463 (p-value = 0.482) and suggests that the data distribution approximates of a random distribution, i.e. there is no spatial dependence between them.

The LISA statistics for manufacture of paints, varnishes, enamels, lacquers and related products indicated the existence of a cluster formed by the municipalities of Guarulhos, São Paulo, Diadema and São Bernardo do Campo, whose indexes are, respectively, 3.6803 (p-value = 0.0426), 10.4484 (p-value = 0.000), 4.4353 (p-value = 0.000) and 9.715875 (p-value = 0.000). For the manufacture of furniture, the LISA statistic indicated the existence of an atypical observation (Paulínia), whose index is −9.155501 (p-value = 0.0021). Both analyzes are shown in Figure 2.

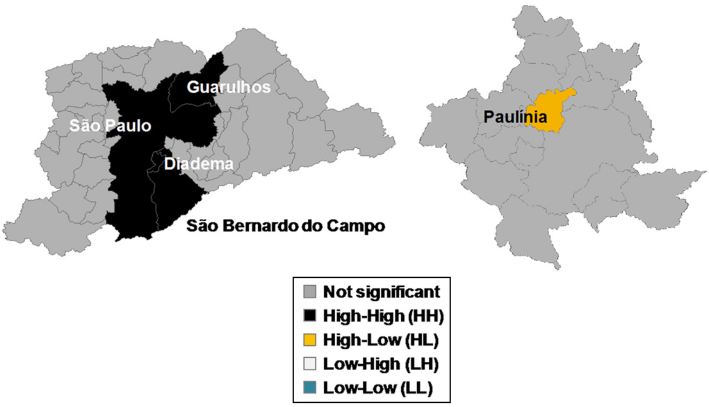

Figure 2. Analysis of clustering and atypical data

Source: Prepared by the authors

The comparison between the results from industry concentration analysis and ESDA applications shows that for manufacture of paints, varnishes, enamels, lacquers and related products, only two, in five municipalities, coincide. From the perspective of the analysis of the concentration of industry, the municipalities of Diadema and São Paulo were not selected because they had LQ below 2, respectively, 1.2408 and 0.3763. On the other hand, from the ESDA perspective, the municipalities of Taboão da Serra and Mauá, although having high employment levels (respectively, 479 and 756), did not reach the level of significance desired by the spatial distribution of the data. For the manufacture of furniture, the municipalities are totally coincident.

Conclusions

Given the ambiguity of the concept, there may be no room for a single methodology or approach that is widely recognized and accepted for cluster identification. However, existing approaches can be improved through the incorporation of perspectives from other fields of science. To improve the robustness of the process of identification and analysis of potential clusters, this research aimed to contribute to such approaches complementing the traditional approach with the techniques of ESDA. It also extended previous studies on the spatial distribution of activities in terms of coverage and requirements for the sample composition to provide greater empirical consistency, and thus, show the practical advantages of using ESDA in industrial clusters studies.

As the results show, the geographic distribution measures (mean center, standard distance and standard deviation ellipse) proved to be fundamental for longitudinal studies on regional dynamics and industrial agglomerations, clearly revealing both, the degree of dispersion and the direction of movement of these agglomerates over time as demonstrated by manufacture of furniture that presented a clear movement in the center-periphery direction.

Regarding the indexes of spatial association, Global Moran's index and LISA, the latter, more notably, was more effective than the former in the process of identifying the units where potential clusters can be found, as it reveals, simultaneously, for a given economic activity, potential clusters that overflows municipalities, i.e. inter cities and those located inside them (intra cities) as demonstrated by both metropolitan regions-economic activities analyzed. The first evidence suggests that the LISA statistic tends to overcome the limitation of the industry concentration approach as observed by Suzigan *et al.* (2001a), who argues that clusters do not necessarily "respect" geographical boundaries in a reference of the limitation of using political-administrative boundaries.

Other aspect that deserve mention is the LISA statistic, which uses the absolute value of the weighting variable (in this case and the number of jobs), and the *pseudo* significance test, which in a certain way, overcome the need of a "cut-off value" for the selection of spatial units. As seen, the application of both methods led to very similar, if not the same results in certain cases. This resolve in part, the limitation of LQ and GL, but it does not completely eliminate the need to going in more deep analysis based on other variables as suggested, for example, by Suzigan *et al.* (2001a), Fingleton *et al.* (2002), Porter (2003), Delgado *et al.* (2014, 2016) and Ketels (2017). However, contributes to reduce the subjectivity observed in studies like this one.

Finally, using the metropolitan region as the unit of analysis and the municipality as the smallest space unit, the results of this work revealed valuable insights about the composition of clusters, trends and spatial extent. Such insights can be used in the development of public policies aimed at transportation, education, etc., and to delimit questions and areas to be addressed in later quantitative or qualitative studies.

Limitations and future research

In the period considered by this work, due to economic, structural and circumstantial questions, activities linked to the transformation industry have been losing ground in the value creation process in Brazil. In this sense, the study of other industries such as services, entertainment, etc., may generate other types of insights that should be considered in the process of regional development.

It is noteworthy that the recent emphasis on the study of convergence and divergence between the different regions has been placed in the underlying spatial dimensions of regional growth processes, both from the theoretical and empirical perspectives, as well as confirmatory and exploratory methodological approaches (Rey and Janikas, 2005). This article offers a critical analysis of empirical approaches and methodological advances with an emphasis on the treatment of special effects: spatial dependence, spatial heterogeneity and spatial scale. However, the regional dynamic presents a temporal dimension and a spatial dimension. The role of space has increasingly attracted attention in the analysis of economic changes. This work has identified opportunities for incorporating spatial effects in regional analysis over time. The use of spatial econometric methods can contribute to an understanding of the effects of spatial dependence and heterogeneity on convergence. However, such opportunities should be addressed in new research when new data should be aggregated to contribute to the analysis proposed.

Finally, it is worth monitoring the development of ESDA methods and tools and persisting in their application and disclosure in the context of planning and regional development as a way of making them better known and more widely accessible to an increasingly greater number of researchers.

Notes

1. It is important to note that the last three works are derivations from Porter (2003).
2. Annually database produced and released by Brazil Ministry of Labor and Employment that aims to control the employment activity in the Country, provide data for employment statistics, and, provide employment market information.
3. National Economic Activity Classification 2.0 (CNAE 2.0) – Section C – Transformation industries and its respective groups (3-Digit).
4. Annual Industrial Survey (PIA 2010).
5. Raw materials, auxiliary materials and components; electricity; fuel; parts and accessories for maintaining and repairing machinery and equipment; industrial services; and maintenance and repair of 3rd parties' machinery and equipment.
6. Anselin (1993) considers the Moran indicator as a linear regression coefficient between z_m and z_i.
7. The software package used was ArcMap^{TM} (v.10.1) from Environmental Systems Research Institute (ESRI®).
8. It should be noted that for the calculation of the Global Moran's Index as well as that of the LISA, the spatial relationship concept used was that of the contiguity of borders and vertices, and the distance method used was that of the Euclidean distance.

References

Anselin, L. (1994), "Exploratory spatial data analysis and geographic information systems", *New Tools for Spatial Analysis*, Vol. 54.

Anselin, L. (1995), "Local indicators of spatial association—LISA", *Geographical Analysis*, Vol. 27 No. 2, pp. 93-115.

Arbia, G. (2001), "The role of spatial effects in the empirical analysis of regional concentration", *Journal of Geographical Systems*, Vol. 3 No. 3, pp. 271-281.

Audretsch, D.B. and Feldman, M.P. (1996a), "Innovative clusters and the industry life cycle", *Review of Industrial Organization*, Vol. 11 No. 2, pp. 253-273.

Audretsch, D.B. and Feldman, M.P. (1996b), "R&D spillovers and the geography of innovation and prod."

Bailey, T.C. and Gatrell, A.C. (1995), *Interactive Spatial Data Analysis*, Vol. 413, Longman Scientific & Technical, Essex.

Cliff, A.D. and Ord, J.K. (1981), *Spatial Processes: Models and Applications*, Vol. 44, Pion, London.

Delgado, M., Porter, M. and Stern, S. (2014), "Clusters, convergence and economic performance", *Research Policy*, Vol. 43 No. 10, pp. 1785-1799.

Delgado, M., Porter, M. and Stern, S. (2016), "Defining clusters of related industries", *Journal of Economic Geography*, Vol. 16 No. 1, pp. 1-38.

Ferreira, F.C.M. (2009), "O efeito da concentração regional das indústrias sobre o desempenho das firmas: uma abordagem multinível", Doctoral dissertation, Escola de Administração de Empresas de São Paulo, Fundação Getulio Vargas, São Paulo.

Freitas, R. (2009), "Regiões metropolitanas: uma abordagem conceitual", *Humanae*, Vol. 1 No. 3, pp. 44-53.

Fujita, M. and Thisse, J.-F. (2002), *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge.

Guillain, R. and Le Gallo, J. (2010), "Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis", *Environment and Planning B: Planning and Design*, Vol. 37 No. 6, pp. 961-981.

Isard, W. (1960), *Methods of Regional Analysis: An Introduction to Regional Science*, MIT Press, Cambridge.

Ketels, C. (2007), *Cluster Mapping as a Tool for Development, Institute for Strategy and Competitiveness*, Harvard Business School.

Krugman, P. (1991a), *Geography and Trade. MIT Press*, MIT Press, Cambridge.

Krugman, P. (1991b), "Increasing returns and economic geography", *JPE*, Vol. 99 No. 3, pp. 483-499.

Lemos, M.B. and Crocco, M.A. (2000), "Competitividade e dinâmica comparativa das regiões metropolitanas brasileiras", *Ufmg/Cedeplar*, 146, Texto para Discussão.

Markusen, J.R. (1984), "Multinationals, multi-plant economies, and the gains from trade", *Journal of International Economics*, Vol. 16 Nos. 3/4, pp. 205-226.

Martin, R. and Sunley, P. (2003), "Deconstructing clusters: chaotic concept or policy panacea?", *Journal of Economic Geography*, Vol. 3 No. 1, pp. 5-35.

Mendoza-Velazquez, A. (2017), "The effect of industrial competition on employment: a porter's approach to the study of industrial clusters in Mexico", *Competitiveness Review: An International Business Journal*, Vol. 27 No. 4, pp. 410-432.

O'Donoghue, D. and Gleave, B. (2004), "A note on methods for measuring industrial agglomeration", *Regional Studies*, Vol. 38 No. 4, pp. 419-427.

Parr, J.B., Hewings, G.J.D., Sohn, j. and Nazara, S. (2002), "Agglomeration and trade: some additional perspectives", *Regional Studies*, Vol. 36 No. 6, pp. 675-684.

Perry, M. (2005), *Business Clusters: An International Perspective. Routledge*, Routledge, New York, NY.

Porter, M. (1998), "Cluster and competition: new agendas for companies, government and institutions", in Porter, M. (Ed.), *On Competition*, Harvard Business School Press, Boston, MA, pp. 197-299.

Porter, M. (2003), "The economics performance of regions", *Regional Studies*, Vol. 37 Nos 6/7, pp. 549-578.

Resbeut, M.P. and Gugler, P. (2016), "Impact of clusters on regional economic performance: a methodological investigation and application in the case of the precision goods sector in Switzerland", *Competitiveness Review*, Vol. 26 No. 2, pp. 188-209.

Rey, S.J. and Janikas, M.V. (2005), "Regional convergence, inequality, and space", *Journal of Economic Geography*, Vol. 5 No. 2, pp. 155-176.

Rosenthal, S. and Strange, W. (2001), "The determinants of agglomeration", *Journal of Urban Economics*, Vol. 50 No. 2, pp. 191-229.

Rosenthal, S. and Strange, W. (2004), "Evidence on the nature and sources of agglomeration economics", in Henderso, N.J.V. and Thisse, J.-F. (Eds), *Handbook of Urban and Regional Economics: Cities and Geography*, Elsevier, Amsterdam, pp. 2119-2172.

Sassen, S. (1991), *The Global City: New York, NY, London, Tokyo*, Princeton University Press.

Suzigan, W., Furtado, J., Garcia, R. and Sampaio, S.E.K. (2001a), "Aglomerações industriais no estado de São Paulo", available at: www.cedeplar.ufmg.br/economia/seminario/wilson_suzigan.pdf (acessado em 21 mar. 2011).

Suzigan, W., Furtado, J., Garcia, R. and Sampaio, S.E.K. (2001b), "Sistemas locais de produção: mapeamento, tipologia e sugestões de políticas", available at: www.anpec.org.br/encontro2003/artigos/E28.pdf (acessado em 21 mar. de 2011).

Tartaruga, I.G.P. (2009), "Análise espacial da centralidade e da dispersão da riqueza gaúcha de 1970 a 2000: notas preliminares", *Ensaios FEE*, Número Especial, Porto Alegre, Vol. 30, pp. 391-426.

Tartaruga, I.G.P. and Sperotto, F.Q. (2009), "A concentração espacial do emprego formal e da massa salarial no Rio Grande Do Sul: metodologia e tipologia", In: *Encontro Nacional Dos Geógrafos. Crise, Práxis e Autonomia: Espaços de Resistência e Esperanças*, 16, Anais. Espaço de Diálogo e Práticas. Porto Alegre.

Upton, G.J.G. and Fingleton, B. (1985), *Spatial Data Analysis by Example: Point Pattern and Quantitative Data. Wiley*, Vol. 1, John Wiley & Sons, UK.

Further reading

BRASIL. Constituição (1988), "Disponível em", available at: www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm (Acesso em set/2012).

BRASIL, Ministério do Planejamento, Gestão, e Orçamento, Instituto Brasileiro de Geografia e, E. and Regiões, M. (2009), "Disponivel em", available at: www.ibge.gov.br/ibgeteen/atlasescolar/mapas_pdf/brasil_regioes_metropolitanas.pdf (acesso em set/2012).

BRASIL. Ministério do Trabalho. Relação Anual de Informações Sociais (RAIS) (2012), "Disponível em", available at: www.mte.gov.br/pdet/index.asp (accessed 14 November 2012).

Ebdon, D. (1985), *Statistics in Geography*, Oxford, Blackwell.

Fingleton, B., Igliori, D.C. and Moore, B. (2002), "Employment growth of small computing services firms and the role of horizontal clusters: evidence from Great Britain 1991–2000", in *European Regional Growth*, Springer, Berlin Heidelberg, pp. 267-291.

Head, K. and Mayer, T. (2004), "The empirics of agglomeration and trade", in Henderson V. and Thisse, J.-F. (Eds), *Handbook of Urban and Regional Economics: Cities and Geography*, 25, Elsevier, Amsterdam pp. 2609-2669.

Holmes, T.J. and Stevens, J.J. (2004), "Spatial distribution of economic activities in North America", in Henderson, V. and Thisse, J.-F. (Eds), *Handbook of Urban and Regional Economics: Cities and Geography*, Elsevier, Amsterdam, pp. 2797-2843.

Mori, T., Nishikimi, K. and Smith, T.E. (2005), "A divergence statistic for industrial localization", *Review of Economics and Statistics*, Vol. 87 No. 4, pp. 635-651.

Ross, J.L.S. (2005), *Geografia Do Brasil. rev. e Ampl*, Edusp, São Paulo.

Tartaruga, I.G.P. (2008), *Análise Espacial da Centralidade e da Dispersão da População e da Riqueza Gaúchas de 1970 a 2000: Metodologias e Notas Preliminaries*, 43, Fundação de Economia e Estatística*, Textos para discussão FEE*, *Porto Alegre*.

Corresponding author
Reinaldo Belickas Manzini can be contacted at: reinaldo.manzini@espm.br