# MEASURING AGGLOMERATION: AN EXPLORATORY SPATIAL ANALYSIS APPROACH APPLIED TO THE CASE OF PARIS AND ITS SURROUNDINGS *

**Rachel Guillain**

LEG-UMR 5118 Université de Bourgogne, Pôle d'Economie et de Gestion, BP 21611, 21066 DIJON CEDEX (FRANCE) guillain@u-bourgogne.fr

**Julie Le Gallo** **

CRESE, Université de Franche-Comté, 45D, Avenue de l'Observatoire, 25030 BESANCON CEDEX (FRANCE) legallo@u-bordeaux4.fr

**Abstract:** This paper suggests a methodology allowing the measurement of the degree of spatial agglomeration and the identification of location patterns of economic sectors. We develop an approach combining the locational Gini index with the tools of Exploratory Spatial Data Analysis. Applying this methodology on Paris and its surroundings for 26 manufacturing and services sectors in 1999, we find that the Gini coefficient and the global Moran's *I* provide different but complementary information about the spatial agglomeration of the sectors considered. Moran scatterplots and LISA statistics reveal a high level of diversity in location patterns across sectors.

**Keywords**: Agglomeration, Exploratory Spatial Data Analysis, Location Patterns, Spatial Autocorrelation

**JEL Classification**: C12, R14, R30

---

* Previous versions of this paper have been presented at the 51st North American Meeting of the Regional Science Association International, Seattle (USA), November 11–13, 2004; at the 45th Congress of the European Regional Science Association, Amsterdam (Netherlands), August 23–27, 2005 and at the International Workshop on Spatial Econometrics and Statistics, Rome (Italy), May 25–27, 2006. We would like to thank M. Copetti, G. Duranton, G.J.D. Hewings, M. Kilkenny, A. Larceneux, F. Puech, P. McCann, H.C. Renski, F.-P. Tourneux and the other participants at these meetings for their valuable comments. This work is part of a study on the economic and geographic evolution of Paris and its surroundings, which is being financed by la Région Picardie (France). The usual disclaimer applies.

** Corresponding author.

## 1. Introduction

Examples of spectacular and famous agglomerations of activities are numerous in the literature, both in manufacture and service sectors: Silicon Valley (California), Route 128 (Boston), Cambridge (UK), Sophia Antipolis (France) for the high-tech industry; Dalton (GA) for the carpet industry; Baden-Württemberg (Germany) for automobiles; Third Italy (Italy) for the ceramics and clothing industries; Wall Street (NY) and the City (UK) for financial services; South East of England (UK) for business services, etc.

Numerous theoretical and empirical studies have been carried out to analyze the determinants for spatial agglomeration of activities (see Fujita and Thisse, 2002 for a theoretical survey; Rosenthal and Strange, 2004 for an empirical survey). Despite this increased interest in the benefits of agglomeration for economic activities, the question of the identification of the spatial limits of agglomeration remains problematic.

The terms “agglomeration” or “cluster” are used to refer to various forms of geographic concentrations (Fujita and Thisse, 2002; Martin and Sunley, 2003; McCann and Sheppard, 2003), as revealed by the examples quoted before: states, regions, cities, districts… This use of the term “agglomeration” in a general sense can be justified since the forces at work in the agglomeration process depend on the spatial scale considered (Anas *et al*., 1998; Rosenthal and Strange, 2001; Fujita and Thisse, 2002) so that the type of agglomeration to which the authors refer has to be specified depending of the topic of the analysis. However, the issue is not so simple since there is still no agreement in the empirical literature as to the geographical limits of the forces at work (Rosenthal and Strange, 2001; Parr *et al*., 2002; O’Donoghue and Gleave, 2004).

A possible starting point for improving the comprehension of agglomeration is to provide a measure of this concept and define clearly its boundaries before analyzing empirically the causes of spatial clustering (O’Donoghue and Gleave, 2004; Duranton and Overman, 2005). This is the focus of the present paper. Indeed, our aim is not to provide explanations of the determinants of agglomeration; on the contrary, we step back from the theoretical and policy concerns and provide instead a method to measure spatial agglomeration.

Since the objective is to evaluate precisely the spatial distributions of activities, the most intuitive approach is to implement methodologies based on a continuous-space approach. In this respect, Feser and Sweeney (2000, 2002), Marcon and Puech (2003, 2006) and Duranton and Overman (2005) test for clustering using different versions of the *K*-density functions in the context of point data analysis. Unfortunately, the data requirements are very demanding since the address of each producing individual establishment is required. Such data are often not available due to confidentiality restrictions (Head and Mayer, 2004; Mori *et al*., 2005; Marcon and Puech, 2006). Generally, data are available at an aggregated geographic level so that a discrete-space approach for evaluating spatial agglomeration of economic activities is often the only option.

Several measures exist to assess geographic concentration in discrete space (see Kim *et al*., 2000; Marcon and Puech, 2003 or Holmes and Stevens, 2004, for a review). However, all these measures share one common weakness: they are a-spatial in the sense that geographic units under study are considered to be spatially independent from each other. The spatial units are treated identically, even if they are neighbors or distant, so that the role of spatial agglomeration can be underestimated. Therefore, two dimensions of agglomeration must be captured by an appropriate empirical methodology; concentration in one spatial unit but also the spatial distribution of these units in the study area. Moreover, another drawback of these methodologies is that they are global in the sense that the local spatial patterns of agglomerations cannot be determined. The location quotient is traditionally used for that purpose but the same problem can be mentioned; the relative position of the spatial units is not taken into account.

In this context, the aim of this paper is to suggest a methodology allowing the measurement of the degree of spatial agglomeration and the identification of location patterns of economic sectors in a discrete space. We develop an approach combining the locational Gini index with the tools of Exploratory Spatial Data Analysis, which presents three advantages. First, the spatial configuration of the data and the spatial autocorrelation in the distribution of employment are explicitly integrated in the analysis. Secondly, the spatial limits of agglomerations are precisely delimited and their spatial patterns are uncovered. Thirdly, no arbitrary cut-offs are needed and the statistical significance of the agglomerations identified can be assessed. This methodology is applied to identify the location employment patterns for 26 manufacturing and services sectors in Paris and its surroundings in 1999 at a fine spatial scale, namely communes (French municipalities).

The paper is organized as follows. In the following section, we discuss the importance of providing a method for measuring agglomeration and present our methodology based on a combination of Gini's index and methods of exploratory spatial data. In section 3, we present the study area, the data and the weights matrix used to perform the analysis. The empirical results are divided into two parts (section 4): first, we compute global measures of spatial agglomeration (Gini coefficient and Moran's *I*) for the sectors considered and show that these measures are complementary, each of them providing interesting insights into the agglomeration process. Secondly, ESDA is used to identify the location and delimitation of agglomerations. The paper concludes with a summary of key findings.

## 2. Exploratory spatial data analysis applied to the measurement and identification of spatial agglomerations

While several contributions have been made to measure agglomeration, there is no general agreement on the criteria that a measure of agglomeration should satisfy[1] (see also Combes and Overman, 2004; Bertinelli and Decrop, 2005). Without providing a definitive answer to this debate, we argue that measuring spatial agglomeration of economic activities in a meaningful way requires first, an evaluation of both the concentration of activities and their location patterns; secondly, an assessment of the statistical significance of these agglomerations and thirdly, accounting explicitly for the spatial dimension of the data.

First, the degree of concentration has to be measured since it provides insights into the relative agglomeration or dispersion of a particular sector and into the respective levels of concentration between sectors. Comparisons between sectors can then be made and differences in the tendency to cluster between sectors can be identified (for example: traditional manufacturing versus high tech sector or manufacturing versus services). While providing interesting information on the propensity of firms to agglomerate, the degree of concentration does not offer any information on the location patterns, that is, where and how the agglomeration process takes place. "Where" refers to the location of the agglomeration process and "how" refers to its form. Different forms of agglomeration can emerge: for example, there could be concentration of sectors in one cluster of spatial units or concentration in several spatial units randomly distributed in all the area or all intermediary forms. Such information has to be uncovered since it also characterizes spatial agglomeration: agglomeration displays several forms, and the forces at work can be different so that the location patterns have to be identified.

Secondly, the significance of spatial agglomeration has to be assessed. Indeed, since the economic activities tend to be naturally unevenly distributed in space, a measurement of agglomeration must be able to separate random from non-random clusters of employment. In

[1] Bertinelli and Decrop (2005) discuss the adequacy of the different measures proposed in the literature to the five criteria proposed by Duranton and Overman (2005) as well as the relevance of these criteria.

other words, it must provide ways to measure exceptional concentration of economic activity (O'Donoghue and Gleave, 2004; Duranton and Overman, 2005).

Thirdly, a fine spatial scale for the spatial units is required to clearly identify the location patterns since agglomeration of a particular sector can appear at the level of a district. Therefore, this measurement has to be performed at the finest spatial unit available with the data. In return, this implies that particular care should be devoted to the issue of spatial autocorrelation, which is even more likely with finer spatial units.

In this section, we present the methodology and its contribution for the identification of agglomeration by discussing the techniques previously used according these criteria. Two steps are necessary: the first step involves measuring the agglomeration while the second step consists of identifying the spatial patterns of agglomerations.

### 2.1 The issue: measuring agglomeration not just concentration

Several global indices have been suggested to measure the spatial concentration of activities: the spatial concentration ratio, the spatial Hirshman-Herfindhal index, the locational Gini coefficients, the Ellison and Glaeser (1997) concentration index, etc. We focus here on one of the most widely used measure because of its ease of computation and its limited data requirements, the locational Gini coefficient. Initially introduced by Krugman (1991) to analyze the relative spatial concentration of U.S. industries, this index can be computed for each sector and measures the relative pattern of this sector in a commune opposed to the same sector in other communes[2]. It is a summary measure of spatial dispersion derived from a spatial Lorenz curve. Formally, the locational Gini coefficient for a sector *m* is calculated as (Kim *et al.*, 2000):

$$G_m = \frac{\Delta}{4\bar{\mu}_x} \tag{1}$$

with: $\Delta = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|$,

$n$ is the number of communes

$i$ and $j$ are indices for communes ($i \neq j$),

$$x_{i(j)} = \frac{\text{Commune } i\text{'s } (j\text{'s}) \text{ share of employment in } m}{\text{Commune } i\text{'s } (j\text{'s}) \text{ share of total employment}} \tag{2}$$

[2] As opposed to the traditional Gini coefficient that focuses on the relative concentration pattern of a sector as opposed to other sectors in a single commune. The advantage of the locational Gini index lies in the fact that the weight of each spatial unit is taken into account. This allows correcting the differences in size between the spatial units which biases the measurement of concentration.

$\overline{\mu}_x$ is the mean of $x_i$ : $\overline{\mu}_x = \sum_{i=1}^{n} x_i \Big/ n$

The locational Gini coefficient has a value of zero if employment in sector *m* is distributed identically to that of total employment (that is if the employment share of sector *m* equals the total employment share), and a value of 0.5 if sector employment is totally concentrated in one commune.

The locational Gini coefficient provides information about the level of concentration of a certain sector and allows comparison between the levels of concentration (dispersion) between sectors. However, only one dimension of the agglomeration is revealed in the case of a high coefficient, namely, the concentration of a sector *m* in a limited number of communes. The geographical pattern of these communes remains unknown: they may be spatially clustered or evenly distributed across the whole area. The value of the locational Gini remains unchanged in both cases whereas the level of agglomeration differs.

Consider for example the two following hypothetical distributions in which the activity is unevenly distributed and the value of the locational Gini is the same:

| | | | |
|---|---|---|---|
| 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 |

| | | | |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 |

The distribution pattern of communes in which employment is concentrated is clearly different: it is evenly distributed on the left and spatially clustered on the right. There is stronger evidence of agglomeration in the second case than in the first case. Thus, the geographical pattern (what Arbia (2001) referred to polarization) of spatial units has to be uncovered in the measurement of agglomeration in addition to concentration. The polarization is low in the first case and high in the second case. In other words, the relative position between the spatial units and the distance between them matters in the measurement of agglomeration, aspects that are not captured by the locational Gini, as is also the case with the other indices used in discrete space. Considering spatial units identically, whether they are geographically distant or neighbors, implies that the measurement of agglomeration is not reliable; if agglomeration effects spill over several neighboring spatial units, the agglomeration is then underestimated. This is even more problematic when the spatial units considered are administratively defined since there is no reason that economic strategies of location coincide with the administrative rules of census. An administrative boundary can split the agglomeration of a sector artificially, a phenomenon that is corrected by taking into account the proximity between spatial units (Head and Mayer, 2004; Sohn, 2004b; Viladecans-Marsal, 2004; Duranton and Overman, 2005). As a consequence, the locational Gini must be complemented by another index that measures the degree of spatial clustering of the distribution, the level of spatial autocorrelation in the distribution.

Spatial autocorrelation can be defined as the coincidence between value similarity and locational similarity (Anselin, 2001). Therefore, there is positive spatial autocorrelation when high or low values of a random variable tend to cluster in space and there is negative spatial autocorrelation when geographical areas tend to be surrounded by neighbors with very dissimilar values. The measurement is usually based on Moran's *I* statistics (Cliff and Ord, 1981). For each sector *m*, this statistic is written in the following form:

$$I_m = \frac{n}{S_0} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \overline{\mu}_x)(x_j - \overline{\mu}_x)}{\sum_{i=1}^{n}(x_i - \overline{\mu}_x)^2} \tag{3}$$

with the same notation as before and $w_{ij}$ is one element of the spatial weights matrix *W*, which indicates the way the region *i* is spatially connected to the region *j* (usually, the diagonal elements $w_{ii}$ are set to zero). These elements are non-stochastic, non-negative and finite. $S_0$ is a scaling factor equal to the sum of all the elements of *W*. In order to normalize the outside influence upon each region, the spatial weights matrix is row-standardized such that the elements $w_{ij}$ in each row sum to 1. In this case, the expression (3) simplifies since for row-standardized weights, $S_0 = n$. Values of *I* larger than the expected value $E(I) = -1/(n-1)$ indicate positive spatial autocorrelation in the distribution of employment in sector *m*, i.e. spatial clustering of similar values of $x_i$, while values of *I* smaller that the expected value indicate negative spatial autocorrelation or spatial clustering of dissimilar values of $x_i$.

The locational Gini and the Moran's *I* have to be considered jointly, since they are complementary to each other (Arbia, 2001; Sohn, 2004b). Moran's *I* is a measure of the covariance between neighboring values (normalized by the total variance in the sample) and indicates whether similar values tend to cluster together. Thus, in the previous hypothetical distributions, even if the value of the locational Gini remains unchanged, Moran's *I* differs indicating some differences in the spatial distribution of the observations. On the other hand, consider now two other hypothetical distributions: in the first, the value 1 in only one square (0 elsewhere) and in the second, the value 10. This difference is reflected in the locational Gini coefficient, which captures the intensity of clustering in one commune, and not in Moran's *I*. As a result, the locational Gini focuses more on the relative distribution pattern among observations while Moran's *I* is more devoted to the spatial pattern of this distribution (Sohn, 2004a, 2004b). Both indices characterize agglomeration in different ways.

While complementary, the locational Gini and Moran's *I* are rarely considered simultaneously in empirical studies, except in Arbia (2001), Viladecans-Marsal (2004), Sohn (2004a, 2004b) and Lafourcade and Mion (2006). A step further can be made. Indeed, these coefficients are global in the sense that only one measure is computed for each sector, so that the local spatial patterns of agglomerations are not completely identified. One can expect to identify the location of agglomerations and their borders in a discrete-space approach as in a continuous-space approach. For that purpose, a spatially disaggregated analysis where one measure or statistic is computed for each commune is required.

### 2.2 Local spatial autocorrelation and spatial patterns of agglomerations

The locational Gini and Moran's *I* can be used to analyze the degree of agglomeration of sectors and allow comparison between the degree of agglomeration or dispersion between sectors. Nevertheless, they fail to identify local spatial patterns of agglomerations; this information is only given about the propensity of sectors to be evenly distributed or to be clustered. No information is given about the location and the distribution of communes in which sectors are located whereas the spatial pattern may differ. For example, a sector can be agglomerated in a single large area, consisting of several contiguous communes or the agglomeration can take place in several areas of geographically contiguous communes with a smaller spatial extent. Moreover, since empirical studies show that the degree of agglomeration varies widely between sectors (see Ellison et Glaeser, 1997; Maurel and Sédillot, 1999; Marcon and Puech, 2003; Bertinelli and Decrop, 2005; Mori *et al.*, 2005; Duranton and Overman, 2005), the spatial patterns of agglomeration are expected to be different according to the sector considered. Since the spatial pattern of agglomerations is informative for the study of the mechanism underlying clustering, it has to be revealed.

The most common approach to spatially delimit agglomeration is to compute a location quotient (LQ) for each sector and each commune. They are given by equation (2) and measure the ratio between the local and the total percentage of employment attributable to a particular sector. A commune is said to be specialized in one sector if it has an LQ over 1. Indeed, in this case, this sector is over-represented within this commune. This property is used to identify agglomerations for a particular sector, which are then defined as the areas with high LQs for that sector.

This methodology raises two main problems. First, several cut-offs have been used in the literature since there is no theoretical or empirical agreement as to how large an LQ should be to indicate clustering (Martin and Sunley, 2003; O'Donoghue and Gleave, 2004); for example, some authors use a cut-off value of 1.25 (Miller *et al.*, 2001) while others prefer a more restrictive definition and use a cut-off of 3 (Isaksen, 1996; Malmbert and Maskell, 2002). The identification of agglomerations is therefore highly dependent upon arbitrary cut-offs. To overcome this problem, O'Donoghue and Gleave (2004) suggest a modification of the LQ index and derive a methodology aimed at testing whether an LQ is significantly high.

Secondly, one measure or statistic is computed for each commune independently from the values of the neighboring communes (Feser and Sweeney, 2002). Therefore, as illustrated above, when spatial autocorrelation characterizes the distribution of employment in the area, such a methodology may not identify properly the locations and delimitations of agglomerations.

To overcome these drawbacks, exploratory spatial measures can be used to identify agglomerations, namely the Moran scatterplots and LISA statistics. These methodologies have been applied to study, *inter alia*, spatial patterns of regional income disparities in European Union (Le Gallo and Ertur, 2003; Dall'Erba, 2005), of homicide rates (Messner *et al.*, 1999) or of urban segregation (Florax *et al.*, 2006). To our knowledge, they have not been applied for the identification of spatial patterns of agglomerations even if two empirical studies made a first step in this direction.

First, Feser and Sweeney (2002) propose the application of Getis-Ord statistics (Getis and Ord, 1992; Ord and Getis, 1995). We argue that the use of Moran scatterplots and LISA statistics is more relevant. Indeed, the former (Getis-Ord) only imply a 2-way split of the sample; the observations are classified either in clusters of high values of the variable considered or in clusters of low values. On the contrary, Moran scatterplots and LISA statistics imply a 4-way split of the sample where not only clusters of high or low values are detected but also "atypical locations" in a sense defined below. Moreover, they overcome the two limitations mentioned above: they are able to identify statistically significant agglomerations without the use of *a priori* and arbitrary cut-offs and explicitly deal with the problem of spatial autocorrelation. Secondly, Lafourcade and Mion (2006) used these tools to detect which Local Labor Systems (Italian spatial nomenclature) contribute the most to the Moran's $I$ significance for the sector "manufacturing of musical instruments." We argue that it is relevant to generalize such a methodology to detect the spatial pattern of agglomerations.

More precisely, the methodology is as follows. Moran scatterplots (Anselin, 1996) plot the spatial lag $Wz$ against the original values $z$ of the variable. Here, $z$ is a vector containing the location quotients defined in (2) in deviation from the mean and, since $W$ is row-standardized, $Wz$ contains the spatially weighted averages of neighboring values for each commune. A Moran scatterplot allows visualizing of four types of local spatial association between an observation and its neighbors, each of them localized in a quadrant of the scatterplot: quadrant HH refers to an observation with a high[3] value surrounded by observations with high values, quadrant LH refers to an observation with a low value surrounded by observation with high values, etc. Quadrants HH and LL (LH and HL) indicate positive (negative) spatial autocorrelation indicating spatial clustering of similar (dissimilar) values. The Moran scatterplot may thus be used to visualize atypical localizations, i.e. communes in quadrant LH or HL.

Since Moran scatterplots do not assess the statistical significance of spatial associations, Local Indicators of Spatial Associations (LISA) statistics will also be computed. Anselin (1995) defines a Local Indicator of Spatial Association (LISA) as any statistic satisfying two criteria: first, the LISA for each observation provides an indication of significant spatial clustering of similar values around that observation; secondly, the sum of the LISA for all observations is proportional to a global indicator of spatial association. The local version of Moran's $I$ statistic for each observation $i$ is written as:

$$I_i = \frac{(x_i - \overline{\mu}_x)}{m_0} \sum_j w_{ij} (x_j - \overline{\mu}_x) \text{ with } m_0 = \sum_i (x_i - \overline{\mu}_x)^2 / n \qquad (4)$$

with the same notation as before and where the summation over $j$ is such that only neighboring values of $j$ are included. A positive value for $I_i$ indicates spatial clustering of similar values (high or low) whereas a negative value indicates spatial clustering of dissimilar values between a zone and its neighbors. Due to the presence of global spatial autocorrelation, inference must be based on the conditional permutation approach. This

[3] High (resp. low) means above (resp. below) the mean.

approach is conditional in the sense that the value $x_i$ at location $i$ is held fixed, while the remaining values are randomly permuted over all locations. The $p$-values obtained for the local Moran's statistics are then pseudo-significance levels (Anselin, 1995).

Finally, combining the information in a Moran scatterplot and the significance of LISA yields the so called "Moran significance map," showing the communes with significant LISA and indicating by a color code the quadrants in the Moran scatterplot to which these communes belong (Anselin and Bao, 1997).

To formally identify agglomerations, we adopt the following definition. For a given sector, an agglomeration is defined by a commune (or a set of neighboring communes) for which the location quotient is significantly higher than the average location quotient. Given our definition of HH and HL communes described above, our definition implies that both the sets of neighboring significant HH communes and the significant HL communes can be considered as agglomerations of a particular sector.

## 2. Study area, data and weights matrix

The study area consists of Ile-de-France, the French capital region, and its surrounding departments. On the basis of the standard division used by public authorities such as DATAR[4], the study area comprises four rings. First, the Ile-de-France region is divided into three rings: the city of Paris, which is also considered administratively as a department, the first ring[5] ("Petite Couronne") and the second ring[6] ("Grande Couronne"). Secondly, the departments immediately surrounding the Ile-de-France region form the third ring[7], which covers several regions of France: Bourgogne, Centre, Champagne-Ardenne, Haute-Normandie and Picardie.

This study area is interesting since it includes the French capital (Paris), which attracts people and economic activities. Even if Ile-de-France, the metropolitan area of Paris, covers only 18% of the study area, it concentrates 73% of the population and 77% of the employment of the study area. Therefore, Ile-de-France is usually perceived as exerting an organizational (centripetal) power of economic activities in a pejorative sense since it seems to have a shadow effect on the surrounding areas; one would thus expect to see fewer activities in these areas. One of the aims in this empirical study is to determine if such an

[4] "Délégation à l'Aménagement du Territoire et à l'Action Régionale".
[5] With the departments of Hauts-de-Seine, Seine-Saint-Denis and Val-de-Marne.
[6] With the departments of Seine-et-Marne, Yvelines, Essonne and Val-d'Oise.
[7] With the departments of Aisne, Aube, Eure, Eure-et-Loir, Loiret, Marne, Oise, Yonne.

effect is observed. For this purpose, two departments in the North (Somme) and North-West (Seine-Maritime) of the study area have been added since they are known to be under Ile-de-France's influence (their major cities, respectively Amiens and Rouen, are only one hour from Paris)[8]. In total, the study area therefore consists of 7,252 communes (French municipalities), which are displayed in maps 1 and 2.

[Maps 1 and 2 about here]

To conduct our empirical analysis, we use the Population Censuses ("Recensement Général de la Population") compiled by INSEE for the year 1999. It provides information about population by place of residence and about public- and private-sector employment by place of work. These data are measured at the communal level. The employment data are classified according to INSEE's industrial classification NAF 700 ("Nomenclature d'Activités Française").

The survey conducted by INSEE is built in two steps. The first exploitation of the questionnaires sent to the households, is exhaustive and concerns general information about population and employment. The second exploitation which concerns, among others, the workers' sectors of activity, is made by a one-quarter survey of households so that the reliability of these data may be questioned. More precisely, a trade-off has to be made between the level of geographical unit on which the analysis is conducted and the level of sectoral disaggregation of activities. For example, it seems to be highly risky to conduct the survey at the communal level by considering all the 700 sectors of activity. Since our analysis requires the use of the smallest geographical unit possible, we choose to aggregate the sectors of activity by following the INSEE hierarchical classification so that 26 sectors are analyzed (see first column of table 1). Moreover, since our aim is to identify high levels of employment, we do not focus on communes in which employment is low, that is in which the risk of measurement error is higher. Therefore, the shortcomings of the data are not really problematic for our study even if the results have to be interpreted with caution.

Contrary to previous empirical studies, which focus only on industrial sectors[9], we focus on both manufacturing and services sectors for two reasons. First, the service sector now accounts for a larger part of economic activities in the developed countries and more particularly in Ile-de-France since 80% of regional employment belongs to this sector, compared to 72% nationwide (IAURIF, 2001). Secondly, the geographic concentration and

---

[8] The Ile-de-France covers 15% of the surface of the entire study area, concentrates 65% of the population and 71% of the total employment.

[9] See for example, Krugman (1991), Audretsch and Feldman (1996), Ellison and Glaeser (1997), Maurel and Sédillot (1999), Marcon and Puech (2003), Devereux *et al.* (2004), Bertinelli and Decrop (2005).

location patterns tend to be different between manufacturing and service sectors so that they have to be analyzed (O'Donoghue and Gleave, 2004). Nevertheless, we neither focus on rural employment since it is not well referenced in the survey, nor on public employment since their location patterns are governed by public decisions.

Looking at the distribution of employment in the study area (see table 1), the supremacy of Ile-de-France in the study area appears clearly but varies according to the sectors of activity. The most concentrated sectors in Ile-de-France are the services sectors (computing industry, R&D, real estate, finance-insurance, high-order services, renting, standard services, consumer services), the wholesale trade, transportation and communication sectors since 70% or more of employment is located in Ile-de-France. The manufacturing sector displays more variation in the degrees of concentration. Two thirds or more of employment in the wood and wood products industry, in the coking, petroleum refining, nuclear industry and in the rubber and plastic industry is located outside Ile-de-France whereas more than 70% of the employment in the sectors of production of electrical and electronic equipments and of production and distribution of electricity, gas and water are concentrated in Ile-de-France.

[Table 1 about here]

This first examination of the distribution of employment suggests differences in the degree of concentration between sectors as in location patterns, which have to be studied more precisely. Before implementing the analysis, we present the spatial weights matrix, upon which part of our empirical analysis rely.

Various spatial weights matrices have been considered in the literature: simple binary contiguity matrices, binary spatial weights matrices with a distance-based critical cut-off above which spatial interactions are assumed to be negligible, generalized distance-based spatial weights matrices. The appropriate choice of a specific weights matrix is still one of the most difficult and controversial methodological issues in spatial statistics and econometrics. From an applied perspective, this choice can be based *inter alia* on the geographical characteristics of the spatial area as the size of the observations in the sample (Le Gallo and Ertur, 2003). We therefore tried several weights matrices: simple contiguity and nearest-neighbors matrices. In the first case, $w_{ij} = 1$ if communes *i* and *j* share a common border and 0 otherwise. In the second case, the weights are computed from the distance between the units' centroids and imply that each spatial unit is connected to the same number

$k$ of neighbors, wherever it is localized. The general form of a $k$-nearest neighbors weights matrix $W(k)$ is defined as following:

$$\begin{cases} w^*_{ij}(k) = 0 \ \text{ if } i = j, \forall k \\ w^*_{ij}(k) = 1 \ \text{ if } d_{ij} \le d_i(k) \\ w^*_{ij}(k) = 0 \ \text{ if } d_{ij} > d_i(k) \end{cases} \quad \text{and} \quad w_{ij}(k) = w^*_{ij}(k) / \sum_j w^*_{ij}(k) \tag{5}$$

where $w^*_{ij}(k)$ is an element of the unstandardized weights matrix; $w_{ij}(k)$ is an element of the standardized weights matrix and $d_i(k)$ is a critical cut-off distance defined for each unit $i$. More precisely, $d_i(k)$ is the $k^{th}$ order smallest distance between unit $i$ and all the other units such that each unit $i$ has exactly $k$ neighbors. Since the average number of neighbors in our sample is 5.80, we used $k = 6$. All our spatial data analysis has been carried out with the simple contiguity weights matrix and the 6 nearest-neighbors. The results will be presented with the nearest-neighbor matrix but they are robust when the contiguity matrix is chosen[10].

## 4. Empirical identification of agglomerations in Paris and its surroundings

We now turn to the examination of agglomerations in Paris and its surroundings. Following the methodology presented in section 2, we begin by providing a global measure of agglomeration by sectors so that we compute in a first step Gini and Moran's *I* coefficients for each of the 26 sectors and compare the results across sectors and measure. We then identify the spatial patterns of agglomerations with the tools of exploratory spatial data analysis: Moran scatterplots and LISA statistics.

### 4.1 Global measures of agglomerations

By analyzing the global measures of agglomerations, we focus on the following questions. First, do the different sectors exhibit the same degree of agglomeration in the study area and if not, which ones are more concentrated or dispersed? Secondly, what information is given about agglomerations by considering jointly the Gini and the Moran's *I* coefficients?

Columns 2 and 3 of table 2 display the Gini coefficients for each sector. It appears that among the most concentrated sectors, three different types of sector are concerned. First, we find the sectors for which location is constrained by natural advantages: coking, petroleum refining and nuclear industry (1st), extraction (2nd). Secondly, R&D is the third most

[10] Complete results are available from the authors upon request.

concentrated sector. Thirdly, there are some traditional industries: production of transport materials (5$^{th}$), wood and wood products industry (6$^{th}$) and non-metallic mineral products industry (7$^{th}$). Conversely, the less concentrated sectors are very diverse: construction (26$^{th}$), consumer services (25$^{th}$), high-order services (24$^{th}$), transportation and communication (23$^{rd}$), wholesale trade (22$^{nd}$).

Columns 4 to 6 display the standardized Moran's *I* statistics for the location quotients for each sector. Inference in this case is based on the permutation approach with 9,999 permutations. Five sectors are not significantly spatially autocorrelated at the 5% level: coking, petroleum refining and nuclear industry, production of machine and equipments, wholesale trade, transportation and communication, renting. This means that globally for these sectors, there is no tendency for clustering of similar values but this does not mean that local pockets of high employment do not exist, as we will show in the following section. All the other sectors are positively and significantly spatially autocorrelated, which indicates that communes with similar values (high or low) of location quotients tend to be spatially clustered in the study area.

[Table 2 about here]

Comparing the rankings of Gini and Moran's *I* coefficients for each sector, it appears that the Kendall and Spearman tests do not reject the null hypothesis of global correlation in rankings[11]. However, considering jointly both coefficients facilitates the identification of three different patterns of agglomerations. First, some sectors tend to agglomerate mainly by concentrating in communes; the Gini coefficient is relatively high whereas the Moran's *I* is relatively low (wood and wood products industry) or not significant (coking, petroleum refining and nuclear industry). This indicates that the agglomeration does not sprawl over a large number of neighboring communes or even tend to be limited at the level of a single commune.

Secondly, for several sectors, agglomeration is characterized by concentration in communes but also by the clustering of communes in which one sector is concentrated: both coefficients are relatively high. This is the case for extraction, textile, clothes, leather and footwear industry, rubber and plastic industry, non-metallic mineral products industry, production of transport materials, computing and R&D.

Thirdly, some sectors present a relatively low value for the locational Gini and a relatively high value for Moran's *I* (farm and food industry, metallurgy and metal

[11] The *p*-values are respectively of 0.21 and 0.23.

transformation, production of electrical and electronic equipments, construction and consumer services). This means that agglomeration tends to sprawl over communes while the degree of concentration in each commune is relatively low.

The sectors present therefore different patterns of agglomerations. Some additional comments are worth noting. First, as for other studies[12], the sectors ‘coking, petroleum refining and nuclear industry’ and ‘extraction,’ the location of which mainly depends on natural resource endowments, are found to be very concentrated (first and second ranking for the locational Gini). However, Moran’s *I* is significant and high for extraction (7th ranking) while not significant for coking, petroleum refining and nuclear industry. This suggests two different patterns of agglomeration, which can be clarified by the analysis of local spatial autocorrelation.

Secondly, the pattern of agglomeration of sectors oriented to the services of population, such as construction or consumer services, seems to be linked with the distribution of final demand, that is population; the value of the locational Gini is relatively low while Moran’s *I* is relatively high. Thirdly, the results for high-order services may appear surprising at first sight since the sector does not appear as agglomerated (low value for both locational Gini and Moran’s *I*) while this sector is well-known for its central nature. However, note that the sectoral desegregation adopted does not offer the possibility of revealing the diversity of location strategies of high-order services. For example, Guillain *et al*. (2004), by using a finer sectoral desegregation for high-order services for Ile-de-France only, show that legal and accounting services tend to be more centrally concentrated whereas data processing and engineering services are more uniformly distributed in Ile-de-France.

Finally, the diversity in terms of patterns of agglomeration is more difficult to explain for several sectors and further investigation about the determinants of agglomeration is required. Even for the traditional sectors, which are known to rank high in terms of agglomeration (Bertinelli and Decrop, 2005), the agglomeration patterns differ strongly: the textile, clothes, leather and footwear industry ranks 10 for the locational Gini and 5 for Moran’s *I* while the metallurgy and metal transformation ranks respectively 19 and 1. This variety of patterns may be explained by the diversity of factors underlying the agglomeration.

These results of global positive spatial autocorrelation must be refined. First, spatial clusterings of high and low values need to be distinguished since we are mainly interested in the former to identify agglomerations. In other words, we need to assess local spatial

[12] See e.g. Maurel and Sédillot (1999), Devereux *et al.* (2004), Bertinelli and Decrop (2005)

autocorrelation in our sample. Secondly, the preceding results revealed not only that the degree of concentration varies across sectors but also that a similar degree of concentration is compatible with different values of spatial autocorrelation, as indicated by Moran's *I* statistic. Therefore, a closer look at the location patterns of the different sectors is necessary.

**4.2 Local spatial autocorrelation and identification of agglomerations**

Columns 2 to 5 of table 3 display the distribution of communes in the quadrants of the Moran scatterplot expressed as percentages of the total number of communes for the 26 sectors. It appears that for all sectors, most communes are characterized by positive spatial association (a majority of communes lying in the HH or LL quadrant) while only a little proportion of the other communes are characterized by negative spatial association (quadrants HL of LH). Therefore, the local spatial pattern is representative of the global positive association in the sample. Note that LL communes overwhelmingly prevail in the distribution of positive spatial autocorrelation (more than 50% of the communes lie in the LL quadrant except for construction, consumer services, transportation and communication, high-order services) and that the deviations from the global trend are dominated by LH communes. For our purpose, we mainly focus on the HH and HL communes.

Moran scatterplots allow us to detect the local spatial instability in our sample and identify the spatial distribution of a particular sector in our study area on a map (Moran scatterplot map). However, they do not allow assessment of the *statistical significance* of such spatial associations. Therefore, only the significant HH or HL communes should be considered as agglomerations of a particular sector. We have computed the LISA statistics for each sector and the distribution of *significant* communes in the quadrants of the Moran scatterplot expressed as percentages of the total number of *significant* communes for the 26 sectors is displayed in columns 6 to 9 of table 3. Interestingly, while the communes were mostly located in the LL quadrant of the Moran scatterplot, they are not significantly so for the majority of sectors. As for the Moran scatterplots, it is possible to map the statistical significant HH and HL communes to identify the location and the form of agglomerations (Moran significance map).

[Table 3 about here]

It is interesting to consider the Moran scatterplot map and the Moran significance map jointly to examine the location patterns. Indeed, since Moran scatterplot shows the spatial distribution of a particular sector, it informs us about the environment in which an agglomeration is located. An agglomeration can be located in a dense cluster of HH communes or surrounded immediately by LL communes. In the first case, it means that an agglomeration is located in a dense environment of high value of LQs. In the second case, it implies that the agglomeration is located in an environment of low value of LQs.

We now turn to a cartographic analysis of our results. We have chosen to display in the paper the Moran scatterplot maps for the following sectors: (i) coking, petroleum refining and nuclear industry (ii) construction (iii) finance-insurance and (iv) R&D.[13] The choice is motivated by the various situations these sectors display. Coking, petroleum refining and nuclear industry has the highest value for the locational Gini while the associated Moran's *I* is not significant; construction presents a relatively low value for the locational Gini and a relatively high value for Moran's *I*; finance-insurance presents relatively low values for both coefficients while R&D has relatively high values for both coefficient. We are interested in the following two aspects: (i) what are the location patterns of these sectors? (ii) is there a shadow effect around Ile-de-France, i.e. are no activities located in the fringes of Ile-de-France?

Looking at the Moran scatterplot map for coking, petroleum refining and nuclear industry (map 3), HH and HL communes are not evenly distributed. This sector is globally not significantly spatially autocorrelated but some very limited local pockets of HH communes appear in Ile-de-France and in the port of the city Le Havre (North-West of the study area). Few HL communes are located randomly in the area outside Ile-de-France. Most of these HH and HL communes remain significant (map 4). The agglomerations of coking, petroleum refining and nuclear industry present a small spatial extent and tend to be immediately surrounded by LL or LH communes.

[Maps 3 and 4 about here]

The location pattern of construction is different (map 5 and 6). In the Moran scatterplot map, HH and HL communes are uniformly distributed in the study area except in Ile-de-France where only LL communes are present. This is indeed a dispersed sector since high levels of LQs are observed in numerous communes. We observe that lot of HH communes are contiguous, the reason why this sector is globally significantly spatially autocorrelated even if HH communes tend to cluster everywhere in the study area. For this sector, Ile-de-France does not have the expected shadow effect since construction is rejected outside. Once the insignificant HH and HL communes are filtered out, agglomerations of construction still appear randomly distributed across the whole study area. Contrary to the coking, petroleum

[13] The Moran scatterplot maps and the Moran significant maps for all sectors are not displayed due to lack of space but they are available from the authors upon request.

refining and nuclear industry, there are more agglomerations in the study area and most of these agglomerations are located in a dense environment of HH communes. This reinforces the idea of a relatively dispersed sector with a specific location pattern that is the clustering in contiguous communes.

[Maps 5 and 6 about here]

Contrary to construction, most of HH communes for finance-insurance are located in Ile-de-France while HL communes are to be found across the whole study area (map 7). Therefore, there seems to be some shadow effect surrounding Ile-de-France since there is a large cluster of HH contiguous communes located in Ile-de-France and clusters of HH contiguous communes outside of Ile-de-France have a smaller spatial extension. This shadow effect is more striking when looking at the Moran significant map (map 8). An agglomeration of finance-insurance formed by contiguous significant HH communes appears clearly in Paris and in communes located in the West of Paris (in the Haut-de-Seine department), an agglomeration located in a very dense environment of HH communes belonging to Ile-de-France. Outside Ile-de-France, there are just few HH significant communes while most HL communes are significant and randomly distributed.

[Maps 7 and 8 about here]

As for finance-insurance, most of HH communes for R & D are located in Ile-de-France (map 9). However, the shadow effect of Ile-de-France is even more striking compared to finance-insurance since the spatial extent of contiguous HH communes is smaller and only two clusters are located outside Ile-de-France. First, in the South, three HH communes belong to the urban area of Montargis where an important center on material, digital simulation, product development and processes is located (center of Hutchinson). Secondly, a cluster of four HH communes is located east in the urban area of Reims. There are less HL communes outside Ile-de-France compared to finance-insurance. The shadow effect is reinforced when looking at the Moran significant map (map 10) since R & D tends to be agglomerated in a cluster of HH significant communes in Paris and in communes located in the West.

[Maps 9 and 10 about here]

Therefore, all these sectors present location patterns widely different in terms of location, spatial extent of agglomeration and environment. Forces at work could explain theses differences. The location pattern of coking, petroleum refining and nuclear industry seems to be linked to the presence of natural endowments while location pattern of construction seems to be related to the distribution of population. The particular location pattern of finance-insurance (a majority of HH significant communes in Ile-de-France and significant HL communes outside) could be explained by the fact that in Ile-de-France, the finance-insurance sector is oriented toward businesses while outside, services for consumers are more likely. The clustering of R & D could be explained by the role of knowledge spillovers, as pointed out by Audretsch and Feldman (1994) in the American context. Further analysis would be required to assess and develop these explanations.

The examination of the other sectors under study corroborates the idea that location patterns depend on the sector considered. For the most agglomerated sectors, the shadow effect of Ile-de-France is especially strong for the computing industry, R&D, real estate, finance-insurance, renting and is moderately pronounced for standard services and for production of electrical and electronic equipments. Moreover, the less agglomerated sectors do not systematically present a uniform pattern of localization in all the study area. For example, wholesale trade, consumer services, transportation and communication, high-order services, farm and food industry, metallurgy and metal transformation display a similar dispersed distribution of HH and HL communes as in construction (expect that for the first four of those sectors, there are HH communes in Ile-de-France). On the contrary, some sectors are only moderately agglomerated such as finance-insurance and standard services but the HH communes are mostly located in Ile-de-France.

The overall picture is one of spatial autocorrelation and diversity of location patterns even though we have not attempted to explain in more depth specific results for each sector. From the applied econometric perspective where the focus would be on the identification of the determinants of agglomeration, these results may have important implications for the proper estimation of the regressions, which consists in linking measures of sector agglomeration to *proxy* variables evaluating agglomeration factors (Head and Mayer, 2004; Rosenthal and Strange, 2004). Indeed, the ESDA results reveal significant spatial autocorrelation. Therefore, spatial autocorrelation of the error term should be systematically tested for in cross-section and panel data specifications and if detected, an appropriate spatial specification (spatial error or spatial lag model) should be estimated using the proper econometric tools to achieve reliable statistical inference (Anselin, 2001, 2006).

## Conclusion

The aim of this paper was to propose and to apply a methodology allowing the measurement of the degree of spatial agglomeration and the identification of location patterns of economic activities. We developed an approach combining the locational Gini index with the tools of Exploratory Spatial Data Analysis. As in previous methodologies suggested in the literature, it provides information about the degree of concentration of the different sectors considered. However, two main advantages of the technique can be pointed out. First, it succeeds in identifying the location patterns by uncovering global and local patterns of spatial autocorrelation. Secondly, the significance of both the agglomeration of activities and their location patterns is assessed.

The technique has been applied to Paris and its surroundings to investigate the employment location patterns of 26 manufacturing and services sectors at a very spatially disaggregated level, i.e. the communes. Our findings show that the Gini coefficient and the global Moran's *I* provide different but complementary information about the concentration of the various sectors. However, these global measures must be enriched by a local analysis using Moran scatterplots and LISA statistics. These tools reveal a high level of diversity in location patterns across sectors. Of course, such diversity has been previously observed but our methodology is further able to delimit precisely the boundaries of the various clusters and to characterize in detail this diversity.

In its current form, this methodology overlooks the problem of the differences in the size distribution of firms. More precisely, by using the Gini index, we are not able to purge spatial concentration from industrial concentration (Bertinelli and Decrop, 2005). This issue can be relevant for identifying agglomeration. Indeed, employment concentration in a commune can be due to the clustering of several small or medium sized interlinked firms or to the presence of a large single firm in the geographical unit. While clusters of different firms suggest that agglomeration forces are at work, the presence of only a large firm does not. Therefore, the plant size distribution within industries has to be taken account to understand the determinants of agglomeration (Ellison and Glaeser, 1997; Maurel and Sedillot, 1999; Martin and Sunley, 2003; Bertinelli and Decrop, 2005; Duranton and Overman, 2005). However, to perform this analysis, the number of establishments in the communes by sectors is required. Unfortunately, this information is not provided by the INSEE survey because of the French laws imposing secrecy of data at the communal level.

Nevertheless, this drawback, due to the use of the Gini index, suggests that a further step can be made for improving the technique by taking into account the size distribution of firms if data becomes available. For example, instead of using the locational Gini coefficient to implement the analysis, the adjusted location quotient proposed by O'Donoghue and Gleave (2004) could be employed, since only employment in small and medium sized firms is taken into account in the computation of the adjusted location quotient instead of total employment.

A further extension of this analysis could be, for example, the study of co-localization. In other words, one can wonder whether agglomerated sub-sectors locate together or separately. For instance, do four-digit sectors belonging to a common two-digit sector agglomerate together or separately? Indeed, it is also informative to consider the extent to which concentrations arise within and between groups of industries and link the results to the various types of agglomeration economies outlined by Parr (2002a, 2002b). In a spatial context, such an analysis could be complemented by the computation of multivariate LISA statistics as suggested in Anselin *et al*. (2002) since they allow evaluating spatial autocorrelation between two different variables.

## References


Anas A., Arnott R., Small K.A. (1998) Urban spatial structure, *Journal of Economic Literature*, 36, 1426-1464.

Anselin L. (1995) Local indicators of spatial association-LISA, *Geographical Analysis*, 27, 93-115.

Anselin L. (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association, in Fisher M., Scholten H.J., Unwin D. (Eds.), *Spatial Analytical Perspectives on GIS*, Taylor and Francis, London.

Anselin L. (2001) Spatial econometrics, in Baltagi B. (Ed.), *Companion to Econometrics*, Basil Blackwell, Oxford.

Anselin L. (2006) Spatial econometrics, in Mills T.C., Patterson K. (Eds.), *Palgrave Handbook of Econometrics: Volume I, Econometric Theory*, Palgrave Macmillan, Basingstoke.

Anselin L., Bao S. (1997) Exploratory spatial data analysis linking SpaceStat and ArcView, in Fisher M., Getis A. (Eds.), *Recent Developments in Spatial Analysis*, Springer Verlag, Berlin.

Anselin L., Syabri I., Smirnov O. (2002) Visualizing multivariate spatial autocorrelation with dynamically linked windows, *REAL Working Paper*, n°02-T-8, University of Illinois at Urbana-Champaign.

Arbia G. (2001) The role of spatial effects in the empirical analysis of regional concentration, *Journal of Geographical Systems*, 3, 271-281.

Audretsch D.B., Feldman M.P. (1996) R&D spillovers and the geography of innovation and production, *American Economic Review*, 86, 630-640.

Bertinelli L., Decrop J. (2005) Geographical agglomeration: Ellison and Glaeser’s index applied to the case of Belgian manufacturing industry, *Regional Studies*, 39, 567-583.

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

Combes P.-P., Overman H.G. (2004) The spatial distribution of economic activities in the European Union, in Henderson J.V., Thisse J.-F. (Eds.), *Handbook of Urban and Regional Economics: Cities and Geography*, Elsevier, Amsterdam.

Dall’Erba S. (2005) Distribution of regional income and regional funds in Europe 1989-1999: an exploratory spatial data analysis, *Annals of Regional Science*, 39, 121-148.

Devereux M.P., Griffith R., Simpson H. (2004) The geographic distribution of production activity in the UK, *Regional Science and Urban Economics*, 34, 533-564.

Duranton G., Overman H.G. (2005) Testing for localization using micro-geographic data, *Review of Economic Studies*, 72, 1077-1106.

Ellison G., Glaeser E.L. (1997) Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105, 889-927.

Feser E.J., Sweeney S.H. (2000) A test for the coincident economic and spatial clustering of business enterprises, *Journal of Geographical Systems*, 2, 349-373.

Feser E.J., Sweeney S.H. (2002) Theory, methods and a cross-metropolitan comparison of business clustering, in McCann P. (Ed.), *Industrial Location Economics*, Edward Elgar, Cheltenham.

Florax R.J.G.M., de Graaff T., Waldorf B.S. (2006) A spatial economic perspective on language acquisition : segregation, networking and assimilation of immigrants, *Environment and Planning A*, forthcoming.

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

Getis A., Ord J.K. (1992) The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24, 189-206.

Guillain R., Le Gallo J., Boiteux-Orain C. (2006) Changes in spatial and sectoral patterns of employment in Ile-de-France, 1978-1997, *Urban Studies*, forthcoming.

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

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

IAURIF (Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France) (2001) *40 ans en Ile-de-France. Rétrospective 1960-2000*, IAURIF, Paris (Etudes et Documents).

Isaksen A. (1996) Towards increased regional specialization? The quantitative importance of new industrial spaces in Norway, 1970-1990, *Norsk Geografisk Tidsskrift*, 50, 113-123.

Kim Y., Barkley D.L., Henry M.S. (2000) Industry characteristics linked to establishment concentrations in nonmetropolitan areas, *Journal of Regional Science*, 40, 231-259.

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

Lafourcade M., Mion G. (2006) Concentration, agglomeration and the size of plants, *Regional Science and Urban Economics*, forthcoming.

Le Gallo J., Ertur C. (2003) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995, *Papers in Regional Science*, 82, 175-201.

Malmberg A., Maskell P. (2002) The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering, *Environment and Planning A*, 34, 429-449.

Marcon E., Puech F. (2003) Evaluating the geographic concentration of industries using distance-based methods, *Journal of Economic Geography*, 3, 409-428.

Marcon E., Puech F. (2006) Measures of the geographic concentration of industries: improving distance-based methods, TEAM mimeo.

Martin R., Sunley P. (2003) Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography*, 3, 5-35.

Maurel F., Sédillot B. (1999) A measure of the geographic concentration in French manufacturing industries, *Regional Science and Urban Economics*, 29, 575-604.

McCann P., Sheppard S. (2003) The rise, fall and rise again of industrial location theory, *Regional Studies*, 37, 649-663.

Messner S., Anselin L., Baller R., Hawkins D., Deane G., Tolnay S. (1999) The spatial patterning of county homicide rates: an application of exploratory spatial data analysis, *Journal of Quantitative Criminology*, 15, 423-450.

Miller P., Botham R., Gibson H., Martin R., Moore B. (2001) *Business clusters in the UK – A first assessment*, report for the Department of Trade and Industry by a consortium led by Trends Business Research.

Mori T., Nishikimi K., Smith T.E. (2005) A divergence statistic for industrial localization, *Review of Economics and Statistics*, 87, 635-651.

O'Donoghue D., Gleave B. (2004) A note on methods for measuring industrial agglomeration, *Regional Studies*, 38, 419-427.

Ord J.K., Getis A. (1995) Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, 27, 286-305.

Parr J.B. (2002a) Agglomeration economies: ambiguities and confusions, *Environment and Planning A*, 34, 717-731.

Parr J.B. (2002b) Missing elements in the analysis of agglomeration economies, *International Regional Science Review*, 25, 151-168.

Parr J.B., Hewings G.J.D., Sohn J., Nazara S. (2002) Agglomeration and trade: some additional perspectives, *Regional Studies*, 36, 675-684.

Rosenthal S., Strange W. (2001) The determinants of agglomeration, *Journal of Urban Economics*, 50, 191-229.

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

Sohn J. (2004a) Do birds of a feather flock together? Economic linkage and geographic proximity, *Annals of Regional Science*, 38, 47-73.

Sohn J. (2004b) Information technology in the 1990s: more footloose or more location-bound? *Papers in Regional Science*, 83, 467-485.

Viladecans-Marsal E. (2004) Agglomeration economies and industrial location: city-level evidence, *Journal of Economic Geography*, 4, 565-582.

# Tables and maps

**Table 1.** Employment in Ile-de-France compared to the whole study area

| | Total in study area | Total in Ile-de-France | % in Ile-de-France |
|---|---|---|---|
| **Total employment** | 4 639 376 | 3 314 495 | 71.44 |
| **Extraction** | 7627 | 4719 | 61.87 |
| **Farm and food industry** | 128410 | 58590 | 45.63 |
| **Textile, clothes, leather and footwear industry** | 70568 | 42877 | 60.76 |
| **Wood and wood products industry** | 13621 | 3746 | 27.50 |
| **Paper and board products industry, publishing, printing** | 128580 | 95761 | 74.48 |
| **Coking, petroleum refining, nuclear industry** | 6028 | 2138 | 35.47 |
| **Chemical industry** | 105646 | 56005 | 53.01 |
| **Rubber and plastic industry** | 53086 | 12660 | 23.85 |
| **Non-metallic mineral products industry** | 37382 | 16236 | 43.43 |
| **Metallurgy and metal transformation** | 126574 | 52850 | 41.75 |
| **Production of machine and equipments** | 82548 | 41723 | 50.54 |
| **Production of electrical and electronic equipments** | 165922 | 118891 | 71.65 |
| **Production of transport materials** | 123259 | 79679 | 64.64 |
| **Various industry** | 47197 | 25269 | 53.54 |
| **Production and distribution of electricity, gas and water** | 68638 | 50209 | 73.15 |
| **Construction** | 360273 | 232777 | 64.61 |
| **Wholesale trade** | 353287 | 274553 | 77.71 |
| **Consumer services** | 773606 | 547927 | 70.83 |
| **Transportation and communication** | 555862 | 416360 | 74.90 |
| **Finance-Insurance** | 305844 | 256205 | 83.77 |
| **Real estate** | 122524 | 102677 | 83.80 |
| **Renting** | 25514 | 20115 | 78.84 |
| **Computing industry** | 143395 | 134000 | 93.45 |
| **R & D** | 65386 | 58966 | 90.18 |
| **High-order services** | 557293 | 444544 | 79.77 |
| **Standard services** | 211306 | 165018 | 78.09 |

**Table 2.** Gini and Moran's *I* coefficients

| | Gini coefficient | Ranking | Standardized Moran's *I* | *p*-value | Ranking |
|---|---|---|---|---|---|
| **Extraction** | 0,492697 | 2 | 5,536232 | 0,0003 | 7 |
| **Farm and food industry** | 0,433559 | 21 | 5,159420 | 0,0001 | 9 |
| **Textile, clothes, leather and footwear industry** | 0,480825 | 10 | 6,044118 | 0,0001 | 5 |
| **Wood and wood products industry** | 0,484626 | 6 | 2,188406 | 0,0275 | 18 |
| **Paper and Board products industry, publishing, printing** | 0,468838 | 14 | 1,884058 | 0,0429 | 19 |
| **Coking, petroleum refining, nuclear industry** | 0,498878 | 1 | *0,585651* | *0,053* | 25 |
| **Chemical industry** | 0,475156 | 11 | 2,405797 | 0,017 | 17 |
| **Rubber and plastic industry** | 0,481168 | 9 | 5,144927 | 0,0003 | 10 |
| **Non-metallic mineral products industry** | 0,483193 | 7 | 6,371428 | 0,0001 | 3 |
| **Metallurgy and metal transformation** | 0,440478 | 19 | 9,449275 | 0,0001 | 1 |
| **Production of machine and equipments** | 0,463818 | 17 | *-0,072464* | *0,4903* | 26 |
| **Production of electrical and electronic equipments** | 0,464257 | 16 | 5,927536 | 0,0001 | 6 |
| **Production of transport materials** | 0,485776 | 5 | 6,289855 | 0,0002 | 4 |
| **Various industry** | 0,469047 | 13 | 4,161764 | 0,0001 | 13 |
| **Production and distribution of electricity, gas and water** | 0,482345 | 8 | 4,115942 | 0,0022 | 14 |
| **Construction** | 0,297233 | 26 | 4,900000 | 0,0001 | 11 |
| **Wholesale trade** | 0,389793 | 22 | *1,550724* | *0,0618* | 23 |
| **Consumer services** | 0,321997 | 25 | 4,782609 | 0,0001 | 12 |
| **Transportation and communication** | 0,379683 | 23 | *0,842857* | *0,1984* | 24 |
| **Finance-Insurance** | 0,455656 | 18 | 3,661764 | 0,0034 | 16 |
| **Real estate** | 0,466710 | 15 | 3,753623 | 0,0043 | 15 |
| **Renting** | 0,488203 | 4 | *1,794118* | *0,0598* | 20 |
| **Computing industry** | 0,471710 | 12 | 5,333333 | 0,0007 | 8 |
| **R & D** | 0,492189 | 3 | 7,597014 | 0,0004 | 2 |
| **High-order services** | 0,375829 | 24 | 1,742857 | 0,0428 | 22 |
| **Standard services** | 0,437488 | 20 | 1,772151 | 0,0301 | 21 |

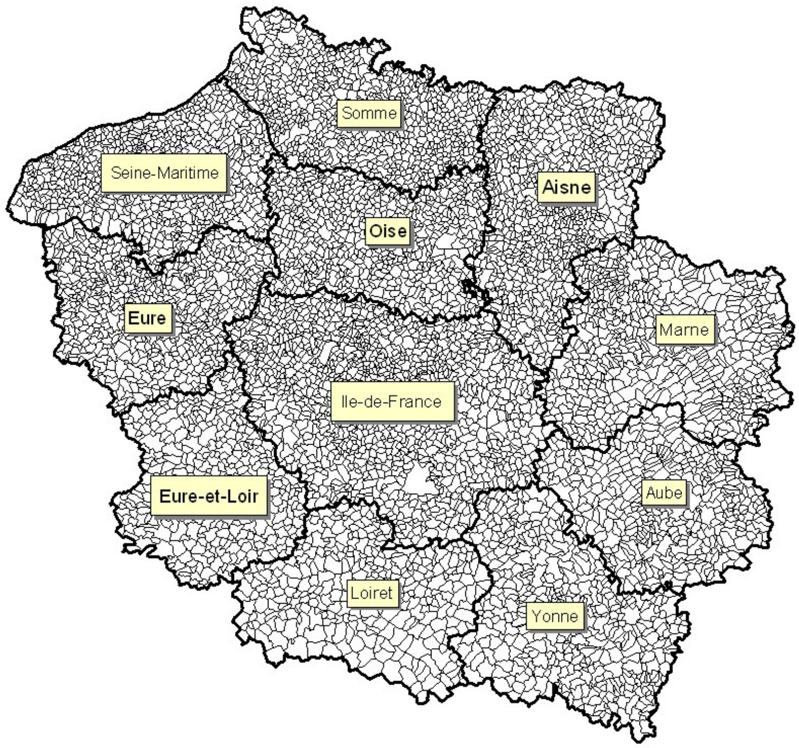
*Notes:* Inference for Moran's *I* is based on the conditional permutation approach with 9 999 permutations. The non-significant sectors at 5% are written in italics.

**Table 3.** Distribution in percentage of communes in the quadrants of the Moran scatterplots and LISA statistics

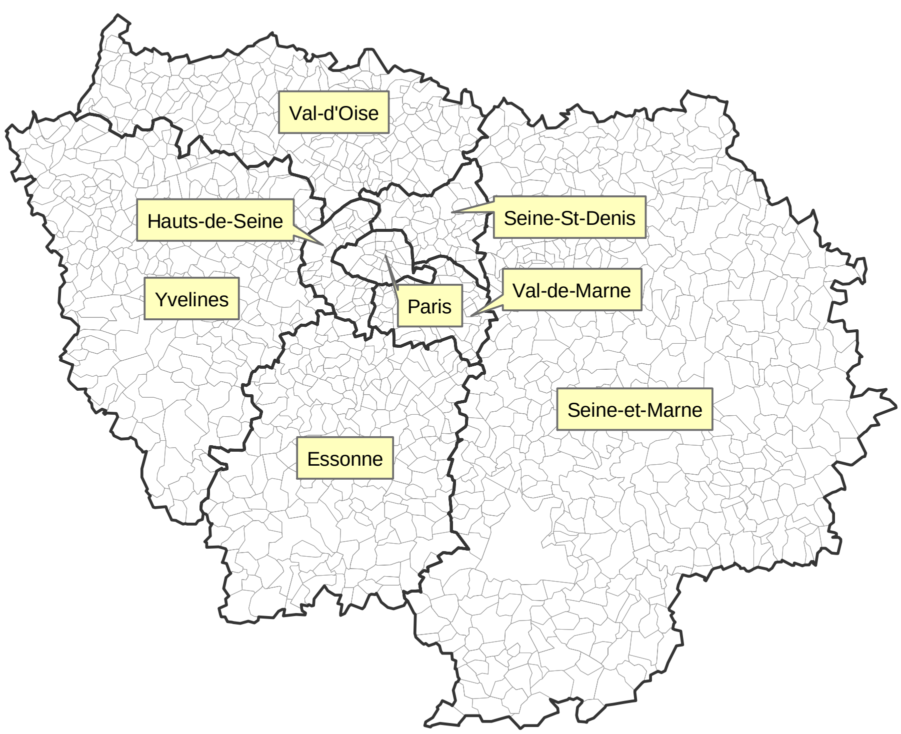
| | Moran scatterplots | | | | LISA statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | *HH* | *LL* | *HL* | *LH* | *HH* | *LL* | *HL* | *LH* |
| **Extraction** | 1,30 | 85,49 | 2,12 | 11,09 | 7,37 | 0 | 24 | 68,63 |
| **Farm and food industry** | 6,77 | 56,77 | 11,27 | 25,19 | 7,37 | 0 | 24 | 68,63 |
| **Textile, clothes, leather and footwear industry** | 2,63 | 73,81 | 5,16 | 18,39 | 8,53 | 0 | 37,54 | 53,92 |
| **Wood and wood products industry** | 1,65 | 77,37 | 4,62 | 16,35 | 5,85 | 0 | 37,69 | 56,45 |
| **Paper and Board products industry, publishing, printing** | 5,57 | 66,57 | 6,44 | 21,41 | 8,81 | 0 | 33,39 | 57,80 |
| **Coking, petroleum refining, nuclear industry** | 0,32 | 95,30 | 0,92 | 3,46 | 6,01 | 14,10 | 14,36 | 65,54 |
| **Chemical industry** | 3,81 | 69,79 | 5,68 | 20,73 | 9,23 | 0 | 32,82 | 57,95 |
| **Rubber and plastic industry** | 2,33 | 74,48 | 5,12 | 18,08 | 6,87 | 0 | 34,68 | 58,45 |
| **Non-metallic mineral products industry** | 2,16 | 75,70 | 4,78 | 17,35 | 7,22 | 0 | 35,20 | 57,58 |
| **Metallurgy and metal transformation** | 7,42 | 57,23 | 9,74 | 25,62 | 16,95 | 0 | 33,44 | 49,61 |
| **Production of machine and equipments** | 3,54 | 62,92 | 8,88 | 24,66 | 6,40 | 0 | 43,20 | 50,40 |
| **Production of electrical and electronic equipments** | 6,83 | 66,23 | 6,05 | 20,89 | 13,58 | 0 | 29,23 | 57,19 |
| **Production of transport materials** | 2,45 | 79,29 | 3,86 | 14,40 | 13,33 | 0 | 29,80 | 56,86 |
| **Various industry** | 3,23 | 66,59 | 7,57 | 22,61 | 8,87 | 0 | 39,52 | 51,61 |
| **Production and distribution of electricity, gas and water** | 3,28 | 75,99 | 5,03 | 15,69 | 7,10 | 0 | 56,66 | 36,23 |
| **Construction** | 17,17 | 37,11 | 17,33 | 28,39 | 21,11 | 36,46 | 16,58 | 25,85 |
| **Wholesale trade** | 12,59 | 44,17 | 13,57 | 29,67 | 6,03 | 60,25 | 12,44 | 21,28 |
| **Consumer services** | 17,95 | 37,60 | 17,25 | 27,19 | 14,04 | 47,37 | 15,88 | 22,71 |
| **Transportation and communication** | 12,34 | 44,08 | 15,10 | 28,47 | 9,73 | 53,40 | 13,26 | 23,61 |
| **Finance-Insurance** | 6,77 | 62,60 | 9,29 | 21,33 | 13,04 | 0 | 47,15 | 39,81 |
| **Real estate** | 7,67 | 69,99 | 5,98 | 16,35 | 12,98 | 0 | 33,22 | 53,79 |
| **Renting** | 2,96 | 80,56 | 3,81 | 12,67 | 5,78 | 0 | 26,89 | 67,33 |
| **Computing industry** | 7,34 | 72,38 | 4,10 | 16,19 | 24,51 | 0 | 23,60 | 51,89 |
| **R & D** | 1,78 | 85,70 | 2,59 | 9,93 | 14,76 | 0 | 25,57 | 59,67 |
| **High-order services** | 14,59 | 43,79 | 14,86 | 26,75 | 6,85 | 62,56 | 12,72 | 17,86 |
| **Standard services** | 10,62 | 55,60 | 9,29 | 24,49 | 11,55 | 0 | 35,74 | 52,71 |

*Notes*: HH denote the High-High communes, LL denote the Low-Low communes, HL denote the High-Low communes and LH denote the Low-High communes. The distribution of communes in the quadrants of the Moran scatterplot is expressed in percentage of the total number of communes. The distribution of significant communes in the quadrants of the Moran scatterplot is expressed in percentage of the total significant communes.

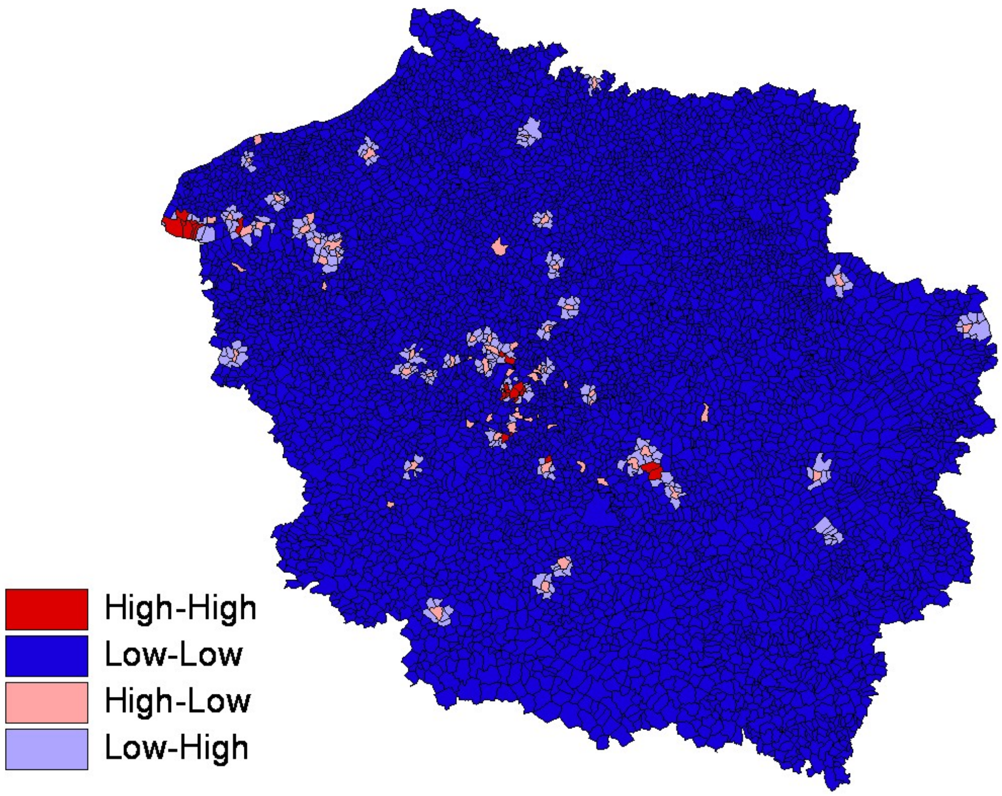
**Map 1:** The departments and communes in the study area



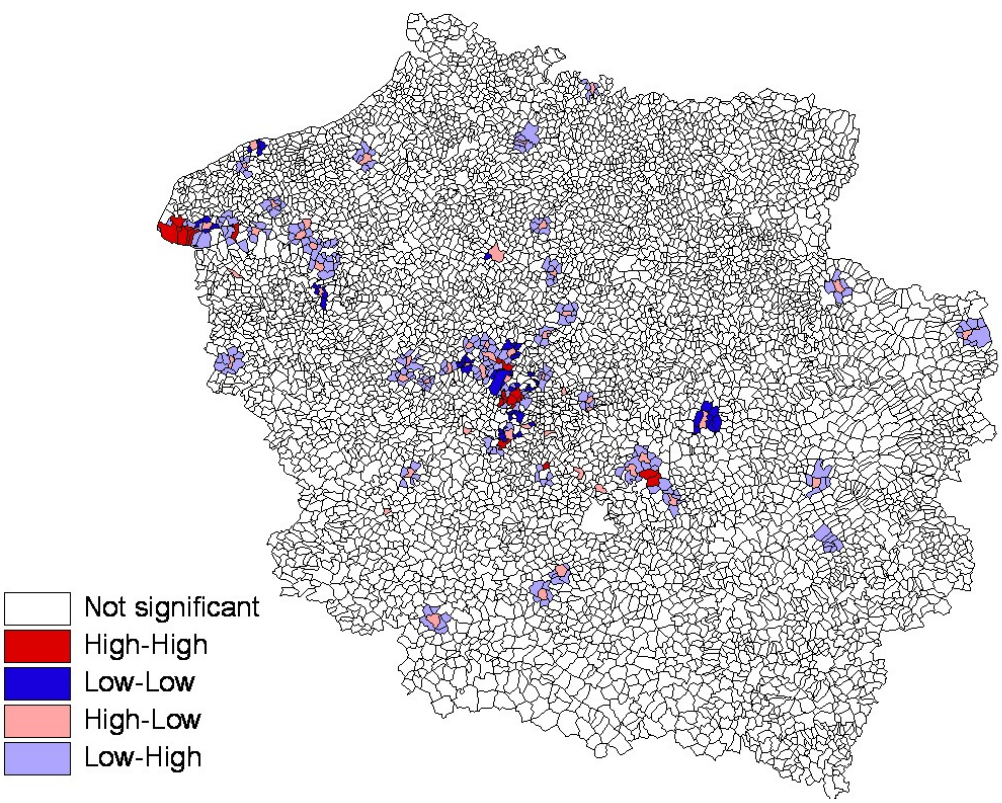

**Map 2:** The departments and communes in Ile-de-France



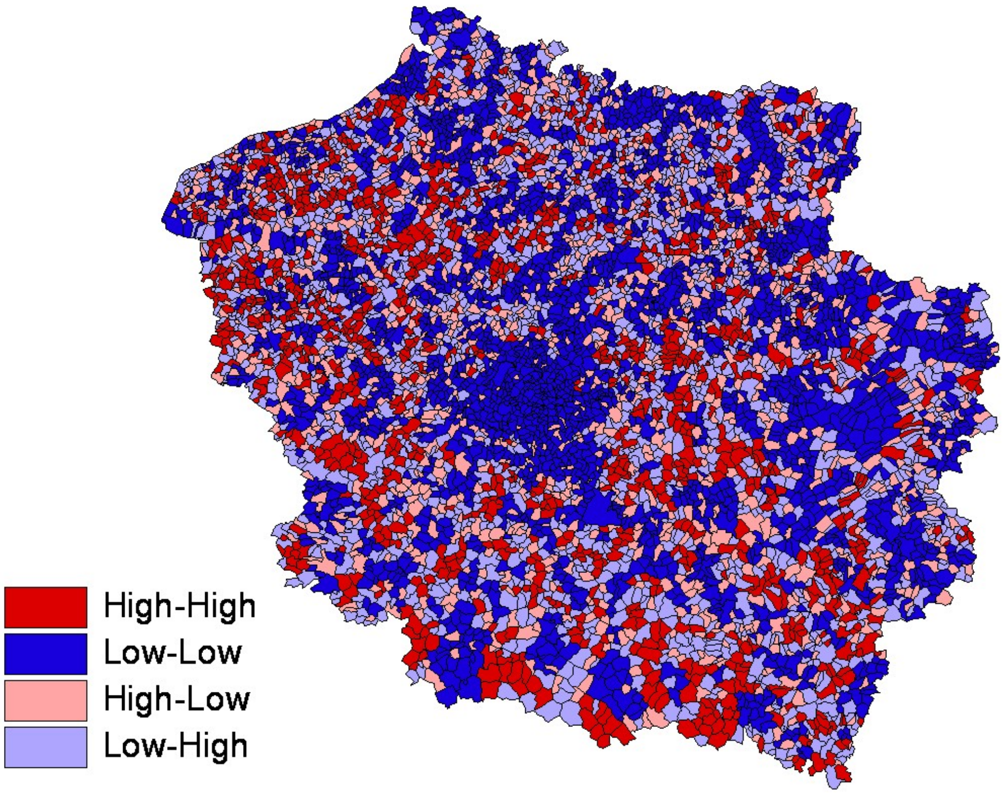
**Map3:** Moran scatterplot map
for coking, petroleum refining and nuclear industry



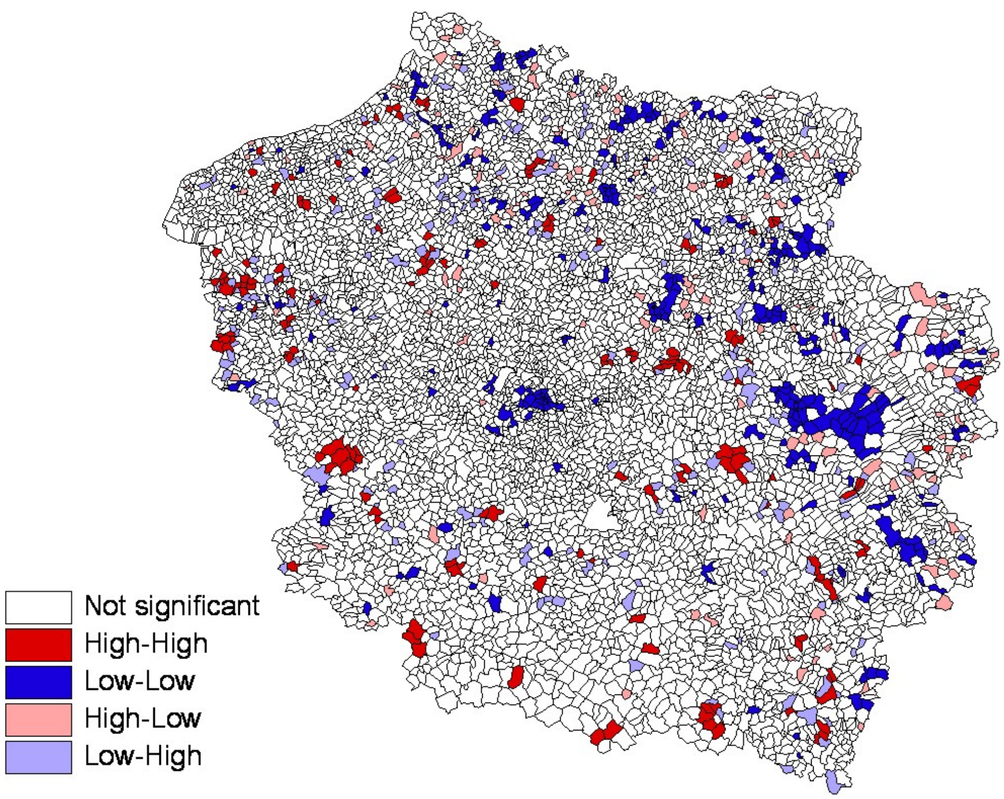

**Map 4:** Moran significance map
for coking, petroleum refining and nuclear industry



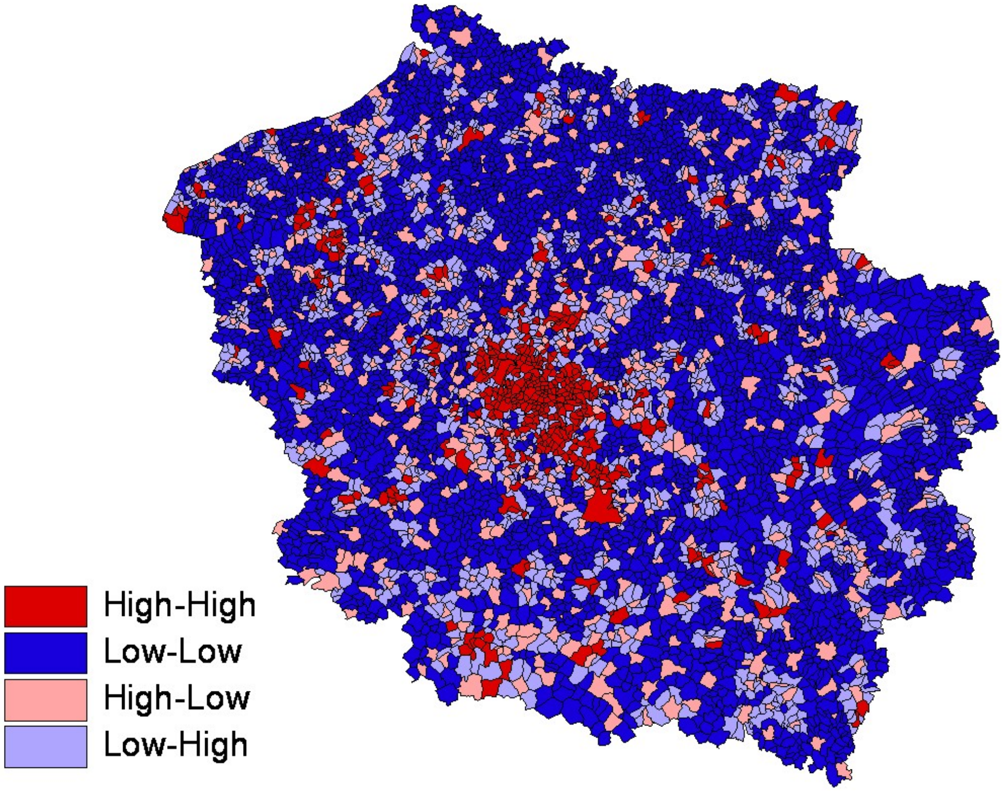
**Map 5:** Moran scatterplot map for construction



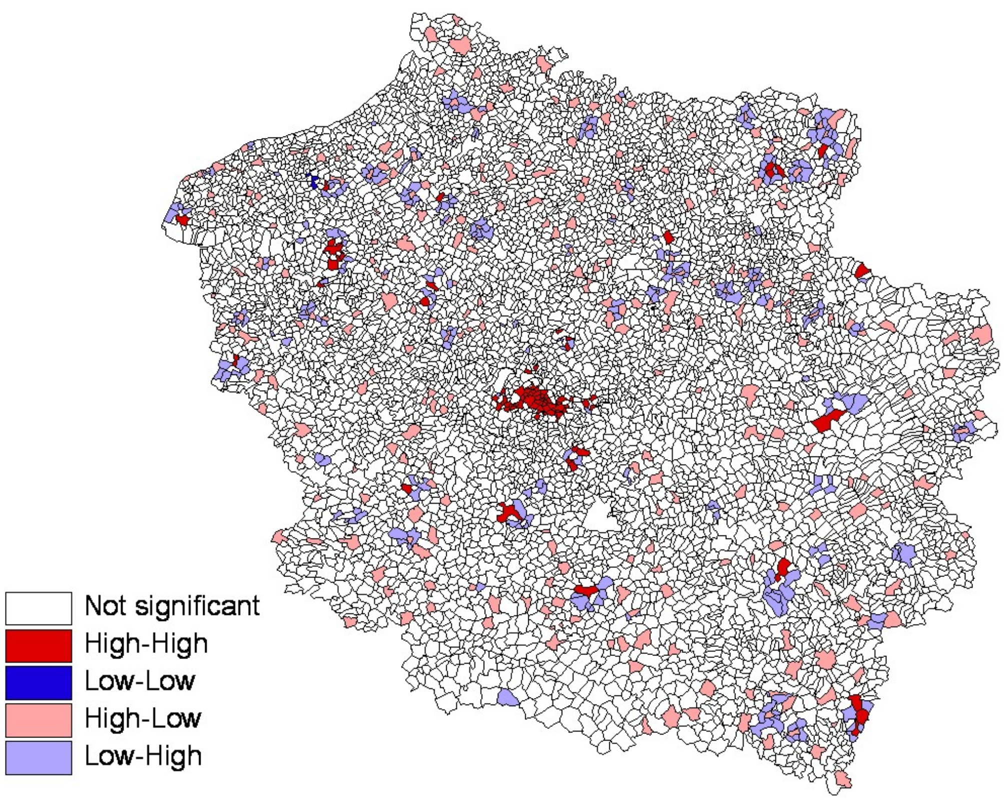

**Map 6:** Moran significance map for construction



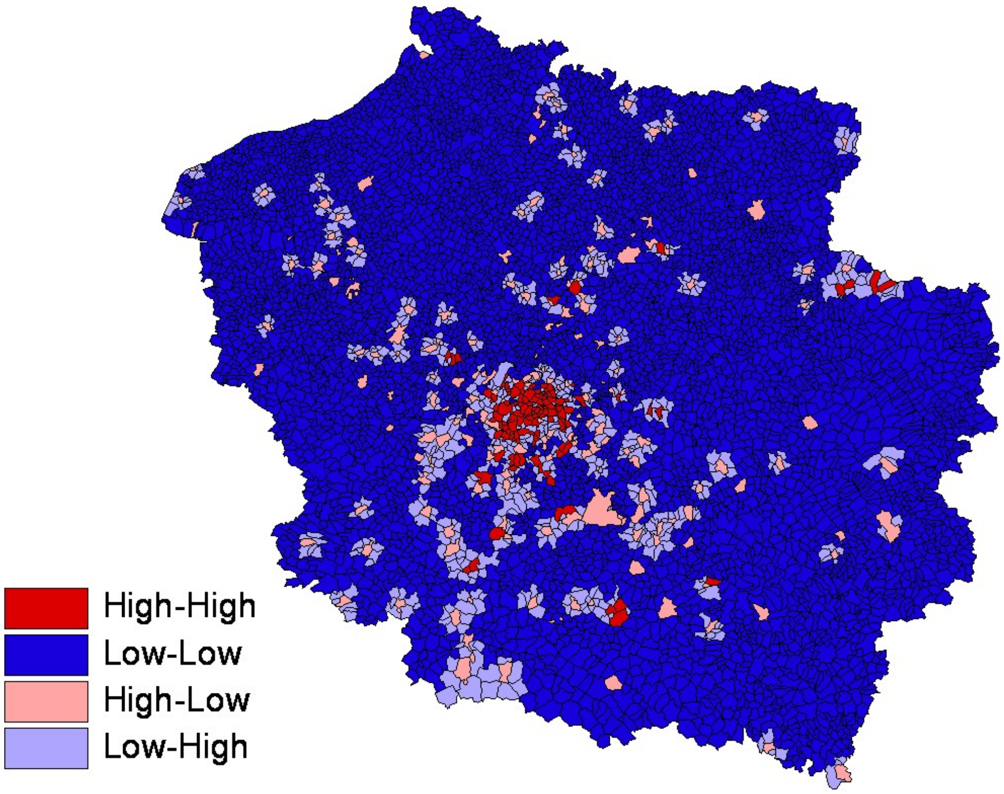
**Map 7:** Moran scatterplot map for Finance-Insurance




**Map 8:** Moran significance map for Finance-Insurance



**Map 9:** Moran scatterplot map for R & D




**Map 10:** Moran significance map for R & D