Identification of employment concentration and specialization areas : theory and application

Identification of Employment Concentration and Specialization Areas: Theory and Application

Frank P. van den Heuvel*, Peter W. de Langen,
Karel H. van Donselaar, and Jan C. Fransoo
*School of Industrial Engineering, Eindhoven University of Technology,
P.O. Box 513, 5600 MB, Eindhoven, The Netherlands*

September 17, 2010

Abstract

This paper presents a new method to identify 'Employment Concentration & Specialization areas' for a particular industry, by simultaneously analyzing absolute and relative employment concentration. This allows for analyzing the performance of these areas in relation to different characteristics such as infrastructure availability and the housing and labor market. This is relevant for scientists, corporate decision makers and local governments, as it can support investment decisions related to new plants or infrastructure. The method is developed and applied to five industries in a Dutch province subdivided in 502 areas.

Keywords: Employment concentration, Geographical analysis, Location patterns, Agglomeration economies.
JEL classification: R12, R30, J21

1 Introduction

The spatial concentration of economic activities has been studied extensively, both as a result of theoretical work in new economic geography (Fujita *et al.*, 1999; Fujita and Krugman, 2004), a research stream that integrates trade and location theories, and by studies that identify regional clusters of economic activities as important for regional economic development (see Porter (2003) and for a critical survey Martin and Sunley (2003)). An empirical investigation of the extent of spatial concentration and its evolution over time is relevant for both research streams. Hence, a substantial body of literature on measures of spatial concentration has developed (see Dewhurst and McCann (2002) for an overview and Cutrini (2010) for a recent analysis of spatial concentration in Europe). This literature focuses on the analysis of concentration measures for administrative regions. Most studies use rather large geographical areas, such as economic areas or states in the US and provinces (NUTS 2[1]) in Europe. The empirical studies show that economic activities in general and in specific industries are often spatially concentrated. Localized synergies, called agglomeration economies have been identified as the main explanation of this spatial concentration. These economies can be divided into localization economies, which arise from being in the neighborhood of other firms operating within the same industry, and urbanization economies, which arise from the scale or diversity of local economic activity in general (Henderson, 2003). Localization economies have been researched extensively (e.g. Krugman, 1991; Devereux *et al.*, 1999; Porter, 2000; Malmberg and Maskell, 2002; Rosenthal and Strange, 2003), always

* *Corresponding author, Email address: F.P.v.d.Heuvel@tue.nl*

[1] The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. NUTS 2 regions are identified as basic regions for the application of regional policies (Eurostat, 2010)

(partly) based on the three major types (Marshall, 1956): labor market pooling, inputs sharing, and knowledge spillovers. Urbanization economies are caused by regional diversity in economic activity, since many ideas developed by firms in one particular sector can also be applied in other sectors (Jacobs, 1969).

Agglomeration economies are related to proximity and distance: the closer firms are together, geographically and organizationally (Torre and Rallet, 2005), the more synergies between these firms. This applies not only regionally (say e.g. on the level of a state in the US), but also at lower geographical scales (Van Soest *et al.*, 2006; Arauzo-Carod and Viladecans-Marsal, 2009). For instance, input sharing in the chemicals industry often leads to co-location of chemical firms at the same site, and in logistics, co-location leads to reduced transport costs (Taniguchi *et al.*, 1999). Despite the relevance of co-location in small geographic areas, limited attention has been given to the issue of identifying such areas with concentrations of specific industries. This paper addresses this issue.

Identifying such areas is practically relevant both for location decisions of companies and for policy decisions by public authorities. First, consider a manager of a company that is looking for areas within a particular region to locate a new branch. Given the importance of knowledge spillovers (Wallsten, 2001), the manager may be interested in identifying areas where the industry is concentrated. Second, consider a regional policy maker that acknowledges the importance of agglomeration economies, both for firms and for the region at large. In order to develop effective policies, e.g. concerning land use and investments in infrastructure, these agglomeration economies are highly relevant. In general, strengthening (and perhaps broadening) existing specializations is a good starting point for regional economic policies (Porter, 2003). Thus, the policy maker is interested in areas in the region where the industry is concentrated.

In both situations described above, the challenge is to identify the areas within a given region where a given industry (or set of industries) is concentrated. This paper presents a method to identify areas with concentrations of specific industries within a particular region, which is especially applicable when geographical units are unequal in size (for instance differences in employment numbers, as is generally the case). In such cases, relevant areas need to meet two criteria: First, absolute concentration of that industry's employment in those areas is relevant. However, absolute employment only is not enough, since this would mean that areas with a high employment level in general (e.g. urban areas) are more likely to be selected. Hence, the second criterium is relative concentration of the industry's employment. This indicator suggests that areas are specialized in the industry. Specialization of an area's employment in the industry of interest by itself is also not enough, since areas with a limited number of firms in general may be included when only one firm belongs to the specific industry. A telling example would be a 'shepherd in the desert'. Relative concentration alone may result in the conclusion that the desert area is highly specialized in cattle farming. In practice, this would not be enough for a policy maker to justify policies to stimulate cattle farming in the desert, and not convince a diary producer to locate his plant in this area. Hence, concentration areas of a particular industry under study have to be identified both based on absolute and relative concentration. In this paper, these identified areas are termed 'Employment Concentration & Specialization areas' (ECS-area) of a particular industry i within a region. An alternative term for ECS-areas would be hotspots. This term is used in practice for areas with large concentrations of activities in specific industries. However, the term hotspot is also used with a different meaning (Pouder and St. John, 1996). Hence, the term Employment Concentration & Specialization (ECS-)areas is preferred. Following the arguments above, an ECS-area is defined as an area which both has high absolute employment as well as specialization in industry i.

Scientifically, the identification of ECS-areas in a particular region is relevant for various purposes. With regard to infrastructure availability, it can be tested whether or not the location of (logistics) ECS-areas is related to intermodal facilities, distances to the highway network or both. In various industries, the relation between ECS-areas and specialized education institutions can be analyzed. The relation between ECS-areas for high value industries and housing prices can be studied. Furthermore, it can be tested whether land or industry specific property values are higher in ECS-areas than outside these areas. For instance, the relation between the value of logistics real estate and ECS-areas for logistics can be analyzed. Finally, it can be tested whether or not employment growth in specific industries is different in ECS-areas compared to other areas or not. All these research questions can be addressed with this method to identify ECS-areas.

The proposed method contains two steps. In step 1, the spatial concentration of industries in the region is measured. This analysis is relevant, since in the absence of spatial concentration economic activities are spread out evenly in different areas and no ECS-areas can be identified. This is unlikely to be the case in most industries (Guillain and Le Gallo, 2010), but this analysis also shows the order of magnitude of spatial concentration in an industry, which is relevant for the interpretation of ECS-areas. In step 2 of the method, ECS-areas are identified, based on both absolute and relative concentration (specialization). Many techniques have been developed to measure spatial concentration of firms or employment (see e.g. Bickenbach and Bode (2008) and Fratesi (2008) for general overviews of these measures) that all use relative spatial concentration. This is explained by the goal of these measures to identify differences in concentration between different industries, different regions, or over time. For this purpose, absolute concentration is not relevant. There is some literature that uses absolute spatial concentration (e.g. Wennberg and Lindqvist, 2010), however, no literature was found in which absolute and relative spatial concentration is considered simultaneously.

Complementary to the research described above on spatial concentration measures, many researchers have analyzed spatial clusters (e.g. Braunerhjelm and Carlsson, 1999; Feser and Bergman, 2000; Gordon and McCann, 2000; Porter, 2000). An important starting point of the research on clusters is that firms located in the cluster are interconnected through input-output linkages. In ECS-areas, this does not have to be the case; firms located in ECS-areas may or may not transact with other firms located in their proximity.

The remainder of this paper is organized as follows. Section 2 will start with an overview of the literature on spatial concentration measures. Section 3 presents the newly developed method to identify ECS-areas. Then, the method is applied to five industries in North Brabant, one of the southern provinces of the Netherlands, in section 4, to show the general applicability of the newly developed method. This province is divided in 502 postal code areas, with an average employment of 1708 employees and a size of about ten km^2 per area. Section 5 concludes this paper by discussing the outcomes and presenting opportunities for further research.

2 Literature on the measurement of spatial concentration

In the economic geography literature, many spatial concentration measures have been developed. Extensive overviews of different measures are given by Bickenbach and Bode (2008) and Fratesi (2008). This section gives a short overview of the most important measures of which one is selected to use in step 1 of the method. This part focusses on relative concentration, since examples of industries in which employment is equally spatially distributed relative to all employment abound (e.g. retail and construction, see Guillain and Le Gallo (2010)), whereas an equal distribution of employment in absolute terms (i.e. industry's employment is equal in all areas under study) is too unlikely to consider. In step 2 of the method, absolute and relative spatial concentration will be combined to identify ECS-areas. In that part, the strengths of different existing measures are used to develop a new method to analyze where spatial concentration takes place and hence, which areas can be identified as ECS-areas.

For the brief overview in this paper, it is useful to make a distinction between two categories of measures. The first category consists of measures where the total region under study is divided into K smaller subregions $k \in \{1, \ldots, K\}$ and the spatial concentration is analyzed per subregion. Thus, the spatial relationship, i.e. the distance and shared borders, between the different subregions is ignored. For this reason, the measures in this category are called a-spatial (although these still are spatial concentration measures). In the second category of measures, analogously called spatial measures, this spatial relationship between the subregions is explicitly considered or the region is not even divided into smaller subregions. Well-known and commonly used a-spatial measures are the locational Gini coefficient (Krugman, 1991), the EG-index (Ellison and Glaeser, 1997), the MS-index (Maurel and Sedillot, 1999), and the D-index (Mori *et al.*, 2005). The most commonly used spatial measures are developed by Moran (1950) and Duranton and Overman (2005). This section will describe the characteristics of these measures and evaluate their relevance for the method to identify ECS-areas. The measures will be presented to evaluate spatial concentration of industry $i \in \{1, \ldots, I\}$, with I all relevant industries in the total region that can be subdivided into K different subregions $k \in \{1, \ldots, K\}$.

2.1 A-spatial measures

One of the oldest contributions to the literature of measuring spatial concentration is by Krugman (1991), who developed the locational Gini coefficient. This index makes use of the location quotient: $LQ_{i,k} = \frac{s_{i,k}}{s_k}$, being a measure to analyze specialization, defined as a subregion k's share of industry i's employment ($s_{i,k}$) divided by that subregion's share of total employment ($s_k = \sum_{i=1}^{I} s_{i,k}$). The locational Gini coefficient is widely used due to its ease of computation and its limited data requirements (Bertinelli and Decrop, 2005). A commonly mentioned disadvantage of the locational Gini coefficient is that it does not control for industrial concentration (i.e. concentration of employment in part of the firms in a particular sector; an industry with large industrial concentration consists of e.g. some large firms and lots of small firms).

Ellison and Glaeser (1997) made another major contribution to the theory of spatial concentration measures, by developing an index γ_{EG} based on a location choice model in which plants choose their location to maximize profit and take into account spatial differences, i.e. natural advantages and industrial spillovers, while making this choice. Maurel and Sedillot (1999) developed a similar index γ_{MS} with the only difference that this index is based on a slightly different measure of raw spatial concentration. Both indices have the same properties: the indices are scaled, such that they control for industrial concentration, and the indices are designed to make it possible to compare the concentration of different industries, regions, or over time. However, the major drawback of these indices is that the outcomes are hard to interpret; for both indices, boundaries of 0.02 and 0.05 are used to define regions of no concentration ($\gamma < 0.02$), intermediate concentration ($0.02 \leq \gamma \leq 0.05$), and high concentration ($\gamma > 0.05$). However, these boundaries are very arbitrary (Duranton and Overman, 2005).

Finally, as none of the above measures can be statistically tested, Mori *et al.* (2005) developed a measure especially designed to be statistically testable: the D-index. However, this index has another major limitation: it can only be used for an analysis of spatial concentration of firms, since it needs the independence of the single units, which means that the D-index is not suitable for the measurement of employment concentration (Fratesi, 2008).

All these a-spatial concentration indices share two well-known problems (Bickenbach and Bode, 2008), namely the checkerboard problem and the modifiable areal unit problem (MAUP). Both are related to the division of the total region into subregions. The checkerboard problem arises from neglecting relevant information on the locations of or distances between regions (Arbia, 2001b): the a-spatial measures do not make a difference between two subregions with relatively much employment that are geographically close together or far apart. The MAUP arises from dividing heterogeneous continuous space into subregions (Arbia, 2001a): a group of firms that are located close together can be grouped into one subregion or spread over several subregions, depending on the defined borders. This influences the degree of spatial concentration of these firms.

2.2 Spatial measures

Different authors propose solutions to cope with the problem of dividing the region into several subregions. This is mostly done through spatial association measures, which measure the correlation between similarities in value and location. These measures broadly can be subdivided into two groups, namely distance-based measures and neighborhood measures (Anselin, 1996).

The distance-based approaches use distances between firms to measure spatial concentration. The intuitive idea behind these approaches is that when an industry is concentrated in one or more parts of the region, firms are located on a shorter distance from each other then when they are randomly located over space. One of the commonly used distance-based measures is the one developed by Duranton and Overman (2005). For these measures, the distances between the industry's firms have to be calculated and hence, the exact locations of the firms are needed. To calculate Euclidean distances, the geographical coordinates of all firms in the dataset are needed, and to calculate distances via transport networks, both the exact addresses of the firms, transport network structures, and software to analyze these have to be available to the researcher. Since the necessary data and software are often not available, this is a major disadvantage of these distance-based measures. In addition, even when the data and software are available, these measures need many calculations and simulations to draw conclusions about spatial concentration.

Neighborhood measures, also called spatial autocorrelation measures, are still based on a subdivision of the total region into subregions, but measure the degree of correlation between the values observed in neighboring subregions. Per subregion neighbors are defined, either based on a border particular subregions share or on the distance between the center-points of these subregions. This neighborhood structure is formalized in a spatial weight matrix W, with elements $w_{jk} > 0$ if subregions j and k are neighbors and $w_{jk} = 0$ if these subregions are no neighbors or if $j = k$. The most commonly used neighborhood measure is Moran's I, first described by Moran (1950) and much studied by Cliff and Ord (1973). Guillain and Le Gallo (2010) present the Moran's I based on the LQs per subregion. The main disadvantage of these spatial autocorrelation measures is that these only take into account the spatial pattern and do not consider the differences in spatial concentration per subregion (Arbia, 2001b).

3 The ECS-area identification method

As described above, none of the existing measures combines absolute and relative spatial concentration. Hence, a new method is developed to identify ECS-areas. As mentioned in section 1, this method contains two steps. Step 1 analyzes whether an industry's employment is spatially concentrated in the first place. For this purpose the locational Gini coefficient (Krugman, 1991) will be used. Step 2 of the method again uses three consecutive steps to identify ECS-areas. This section explains the method, visualized in figure 1.

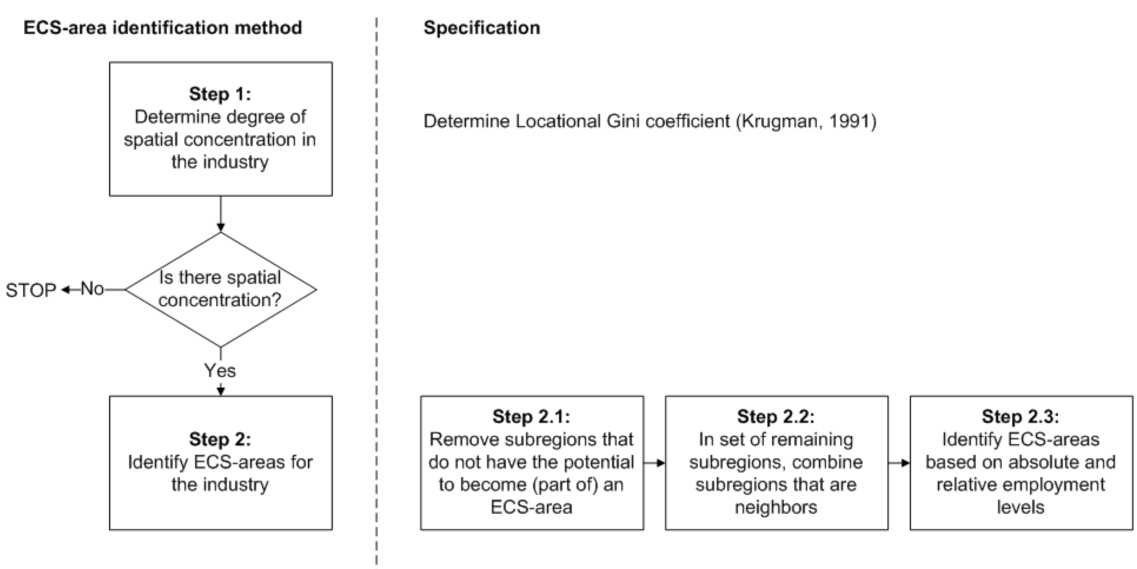


Figure 1: The ECS-area identification method, as developed in this paper

3.1 Step 1: The degree of spatial concentration

The distance-based measures are not used, since, due to the data-requirements and the efforts needed to calculate these measures, these simply are too complex for the purpose of this step of the ECS-area identification method, namely to come up relatively easy with the degree op spatial concentration of a particular industry in a predefined region. Furthermore, similar to the D-index (Mori *et al.*, 2005), these distance-based measures are primarily defined on a firm-level, while the ECS-area identification method will be developed based on employment figures.

Since all other spatial concentration measures divide the total region in smaller subregions, subregion data will be used and consequently the checkerboard problem and the MAUP have to be dealt with. According to Arbia (2001b) and Lafourcade and Mion (2007), the solution is to combine different measures, which then should be interpreted simultaneously to draw sound conclusions. In this paper, following Guillain and Le Gallo (2010), the locational Gini coefficient and the Moran's I are combined. To determine the degree of specialization of subregions in industry i in step 1 of the method, the locational Gini coefficient will be used. In step 2 of the method, this measure will be combined with the Moran's I.

The choice for the locational Gini coefficient is based on several reasons. First, this index is easiest to understand and interpret intuitively. Second, the fact that this measure does not compensate for industrial concentration patterns is actually an advantage for the purposes of this paper. Even if spatial concentration in a specific industry is caused by industrial concentration, it still remains relevant for policy making and for firm location decisions. Third, the locational Gini coefficient uses the LQ per subregion, which can be used in step 2 of the method to determine which subregions are specialized in industry i. When either the EG- or the MS-index is used, it is much harder to determine where spatial concentration takes place.

Based on the locational Gini coefficient, it will be concluded whether an industry's employment is spatially concentrated. For industries where this is not the case, there is no value to identify ECS-areas: ECS-areas for industries for which employment is evenly spread over space are rather arbitrary ECS-areas, since small increases (decreases) of the cut-off values result in much more (less) ECS-areas.

3.2 Step 2: The identification of ECS-areas

After it is concluded *that* employment in industry i is spatially concentrated, it is determined *where* this concentration takes place. Here, the local Moran's I (Anselin, 1995; Guillain and Le Gallo, 2010) is used, which relates a subregion's LQ to the LQ of its neighbors. The ECS-areas can consist of two or more predefined geographical subregions, and hence, spatial concentration per subregion has to be combined with the spatial concentration in neighboring subregions. A positive local Moran's I for subregion k ($I_{LQ,i,k}$) means that there is spatial autocorrelation of similar values between a subregion k and its neighboring subregions, and a negative $I_{LQ,i,k}$ means that there is spatial autocorrelation of dissimilar values.

In addition, the novelty of the method introduced is the combination of absolute and relative concentration to identify ECS areas. A stepwise method is developed to determine where industries concentrate. The method uses industry i's employment, but can also be used with the number of firms. Step 2 of the method itself consists of three steps:

2.1 Remove subregions that do not have the potential to become (part of) an ECS-area.

   (a) Define for both the absolute employment values and the LQ values a cut-off value $e_{min,1}$ and $LQ_{min,1}$ respectively.

   (b) Code all subregions $k \in \{1, \ldots, K\}$ based on their absolute employment level $e_{i,k}$ in industry i and their $LQ_{i,k}$ compared to these cut-off values: $b_{i,k} = 1$ if $e_{i,k} \geq e_{min,1} \wedge LQ_{i,k} \geq LQ_{min,1}$ and $b_{i,k} = 0$ otherwise. Remove all subregions k with $b_{i,k} = 0$.

2.2 In set of remaining subregions, combine subregions that are neighbors.

(a) Determine $B_{i,j,k}$ joins between two neighboring subregions j and k, with both $b_{i,j} = 1$ and $b_{i,k} = 1$: $B_{i,j,k} = w_{jk} b_{i,j} b_{i,k}$, with $w_{jk} = 1$ if subregions j and k are neighbors and $w_{jk} = 0$ otherwise. Furthermore, $w_{jk} = 0$ if $j = k$.

(b) For all j and k for which $B_{i,j,k} = 1$, define a new area $m \in \{K+1, \ldots, M\}$, with employment being the sum of the employment values of the different subregions: $e_{i,m} = e_{i,j} + e_{i,k}$. Area m can exist of more than two earlier defined subregions: e.g. if $B_{i,j,k} = 1 \wedge B_{i,j,l} = 1$, then $e_{i,m} = e_{i,j} + e_{i,k} + e_{i,l}$, etc. In addition, calculate $LQ_{i,m}$. The total number of newly created areas is equal to $(M - K)$.

2.3 Identify ECS-areas based on absolute and relative employment levels.

(a) Define for both the absolute employment values and the LQ values a second cut-off value: $e_{min,2} \geq e_{min,1}$ and $LQ_{min,2} \geq LQ_{min,1}$.

(b) Determine $\beta_{i,c}$ for all areas $c \in \{1, \ldots, M\}$: $\beta_{i,c} = 1$ if $e_{i,c} \geq e_{min,2} \wedge LQ_{i,c} \geq LQ_{min}$, and $\beta_{i,c} = 0$ otherwise. All areas c with $\beta_{i,c} = 1$ are identified as ECS-areas of industry i. If both $\beta_{i,k} = 1$ and $\beta_{i,m} = 1$, with $e_{i,m} = e_{i,k} + e_{i,j}$ for any $j \in \{1, \ldots, M\}$, only the total area m is defined as an ECS-area.

Notice that the method combines the local aspects of the locational Gini coefficient and Moran's I, based on the LQ as well as absolute employment levels. Both levels are used to determine whether a subregion is (part of) an ECS-area for that industry i. Since no clear procedure exist to combine these (absolute and relative) values into one variable, it is not enough to simply calculate the local Moran's I, since this index can only be calculated on one variable with interval or ratio data (Cliff and Ord, 1973). Nevertheless, a binary variable $b_{i,k}$ can be created, containing nominal data per subregion, which indicates whether a particular subregion has the potential to become (part of) an ECS-area. With these nominal data, the most basic measure of spatial autocorrelation, on which Moran's I is based, can be used (Cliff and Ord, 1973, page 4). In step 2.2, subregion k's value of this binary variable is compared to the value of one of its neighbors, subregion j. When the variable is equal to one for both of these two neighboring subregions, it indicates that the probable ECS-area is larger than one subregion. Hence, these subregions are combined into a new area m. With combining two or more of the predefined geographical subregions, the method compensates for the MAUP and checkerboard problem. In step 2.3, all areas, now consisting of one or more predefined subregions, are evaluated against two new cut-off values. These values can be equal to the values defined in step 2.1. However, the cut-off values defined in step 2.1 only determine which predefined subregions are immediately removed from the analysis, since these have a really low (absolute or relative) employment level in industry i. It is possible that two subregions that by itself are no ECS-areas can be one when combined. Thus, when combined areas are created in step 2.2, only the subregions that would not even have potential to become (part of) an ECS-area have to be removed. Later on, when ECS-areas are identified in step 2.3, higher cut-off values are more useful, since then only the areas with very high (absolute and relative) employment in industry i are of interest. Hence, it is advised to use higher cut-off values in step 2.3 than in step 2.1 of the method.

Crucial for the performance of this method is the determination of the cut-off values $e_{min,1}$, $e_{min,2}$, $LQ_{min,1}$, and $LQ_{min,2}$. For the cut-off values used in step 2.1, commonly accepted values are proposed to be used. For the LQs, subregions with an LQ larger than or equal to 1 are selected, since these subregions are commonly stated to be, at least a bit, specialized in that industry i (Guillain and Le Gallo, 2010). For the absolute value, the commonly used cut-off value is the average employment over all subregions, also used in the (local) Moran's I. Since this average is highly influenced by the number of subregions in which the industry's firms are located, this average has to be calculated only based on the subregions in which the industry is present, which makes the method less dependent of the size of the industry and the number of subregions defined.

The cut-off values used in step 2.3, can be chosen differently for different purposes. Consider again the two situations described in the beginning of the paper. In the example of the policy maker, a low number of ECS-areas may be preferred and hence, the cut-off values are chosen relatively high. In the other example, in which ECS-areas are identified to assist a location decision, more ECS-areas may be desirable, since a manager may want to have a complete overview of relevant locations, and the cut-off values can be somewhat lower. In the next section, rather high cut-off values are chosen (in line with the policy perspective) for both the absolute and relative concentration to identify ECS-areas in one of the southern provinces in the Netherlands.

4 Application of the ECS-area identification method

In this section, the ECS-area identification method is applied on five different industries in North Brabant, one of the southern provinces of the Netherlands. First, the dataset used for the analysis is discussed and afterwards the results of the analysis are presented.

4.1 Material used

For this analysis, five industries are used, with a different extent of spatial concentration and of a different size. The industries used in this study are manufacture of chemicals and chemical products (chemical production), research and development (R&D), logistics, construction, and retail trade. Two of these tend to concentrate spatially (chemical production, and R&D) and two others tend to distribute equally over space (construction and retail trade) (Guillain and Le Gallo, 2010). Logistics is included in this list, since this is a relatively large industry in North Brabant, due to the location of this region in Europe, between the two major seaports (Rotterdam and Antwerp) and large consumer markets. Furthermore, this industry is especially relevant from a policy perspective, since logistics activities generate relatively much transport flows and occupy relatively much space. Logistics employment is likely to be less spatially concentrated than the two highly concentrated industries, but more than the two others. To classify firms in different industries, the standard Dutch industry classification, the *Standaard BedrijfsIndeling (SBI)*[2], is used. Appendix A presents a list of the industry codes used in this paper.

For the analysis, the 2008 database containing all business establishments in North Brabant is used. This database contains 153,619 firms in total, and after a standard clean up 150,915 firms. Furthermore, governmental organizations and establishments related to education, health services, culture, and recreation activities were deleted from the database. The remaining database covers 123,611 firms with a total employment of 857,165, which gives an average number of employees per firm of 6.93, with a standard deviation of 43.28 and a median of 1.

To subdivide the total region into subregions, a subdivision based on four-digit postal code areas is used. In total, North Brabant contains 502 four-digit postal code areas. Table 1 presents descriptive statistics of the spread of employment over the different postal code areas. In addition, appendix B presents the distribution of absolute employment and the LQs geographically for the five industries.

4.2 Step 1: The degree of spatial concentration

In this paper, the following definition of the locational Gini coefficient is used (Kim *et al.*, 2000; Guillain and Le Gallo, 2010):

$$G_{LQ,i} = \frac{\Delta_{LQ,i}}{4\overline{LQ_i}} \quad (1)$$

$$\Delta_{LQ,i} = \frac{1}{K(K-1)} \sum_{k=1}^{K} \sum_{j=1}^{K} |LQ_{i,k} - LQ_{i,j}| \quad (2)$$

[2]The SBI is developed by Statistics Netherlands and categorizes economic activities based on five digits: the first four digits correspond with the categorization of the European Union (NACE: *Nomenclature statistique des activits economiques dans la Communaut Europenne*), with a small number of exceptions, and the first two digits of the SBI and the NACE correspond to the categorization of the United Nations (ISIC: *International Standard Industrial Classification of All Economic Activities*).

	Sum	Average	Average if not 0	Standard deviation	Median	Maximum
Total	857,165	1,707.50	1,717.77	2,262.87	889	18,434
Chemical production	17,058	33.98	140.98	245.11	0	4,512
R&D	6,861	13.67	37.09	195.13	0	4,328
Logistics	93,586	186.43	232.80	418.45	33	3,026
Retail trade	114,684	228.45	237.44	374.34	103	3,529
Construction	85,668	170.65	173.42	218.28	85	1,908

Table 1: Descriptive statistics of the number of employees per four-digit postal code area in North Brabant

with $\overline{LQ_i}$ the average LQ of industry i over all subregions K. The locational Gini coefficient is equal to zero when no relative spatial concentration is measured, meaning that the spatial distribution of the industry's employment is equal to the spatial distribution of employment in general ($LQ_{i,k} = 1$ for all k). The maximum value of this index is equal to 0.5, when the industry's employment is completely concentrated in one subregion. In addition, the absolute locational Gini coefficient can be defined similar to the relative locational Gini coefficient, by replacing all LQ_i by s_i.

	Gini (based on LQ values)	Gini (based on absolute values)
Chemical production	0.4730	0.4843
R&D	0.4646	0.4855
Logistics	0.3051	0.3967
Retail trade	0.2285	0.3292
Construction	0.1927	0.2906

Table 2: Locational Gini coefficients based on the employment in North Brabant

	Gini (based on LQ values)
Chemical production	0.4752
R&D	0.4922
Wholesale trade	0.3898
Transportation and communication	0.3797
Retail trade	0.3220
Construction	0.2972

Table 3: Locational Gini coefficients based on employment found by Guillain and Le Gallo (2010)

Table 2 presents the locational Gini coefficients based on the LQ values and the absolute employment values. All coefficients, either based on relative (LQ) or absolute values, lead to the same general conclusions: the chemical and R&D industries are highly spatially concentrated, the construction and retail trade industries are not, and logistics scores somewhere in between, still relatively high.[3] This outcome is comparable to what Guillain and Le Gallo (2010) found based on communes around Paris[4], as can be seen in table 3.

As can be seen in table 2, the coefficients based on the absolute values are generally higher than the coefficients based on the LQs. This is not surprising, due to e.g. zoning schemes or the

[3]In addition, locational Gini coefficients based on number of firms were calculated. Although these coefficients were lower than the locational Gini coefficients based on employment, similar patterns were observed and hence, the analysis can be solely based on employment levels.

[4]French communes have an average surface of 15 km^2, while the four-digit postal code areas in North Brabant have an average surface of 10 km^2, and hence, the studies can be compared.

fact that the subregions are not equally sized. Based on absolute values only, the conclusion could be that spatial concentration is present in a particular subregion, while this is only a result of the size of that subregion. This again shows that absolute and relative concentration have to be considered simultaneously while identifying ECS-areas.

4.3 Step 2: The identification of ECS-areas

Based on the locational Gini coefficients found for the five different industries, only ECS-areas for the chemical, R&D, and logistics industry will be identified. Since there hardly is any spatial concentration in the other two industries, an analysis into the ECS-areas for those industries is not relevant.

In line with section 3, in step 2.1 of the method, the average employment per subregion is used as a cut-off value for the absolute employment level and 1 as a cut-off value for the LQ. The cut-off values used in step 2.3 depend on the purpose of the analysis. In this paper, the policy perspective is chosen and hence, rather high cut-off values are used in step 2.3 of the method. For the absolute employment level, the maximum of the 90th percentile and the bound defined in step 2.1 is used as a cut-off value, meaning that only the top 10 percent subregions can qualify as ECS-areas. For the LQ, the minimum of the 90th percentile and 2 is determined; the LQ cut-off value is equal to the maximum of this value and the cut-off value defined in step 2.1 of the method.[5] As previously argued, all statistics used as bounds are calculated based on subregions in which the employment of the industry is larger than zero. Table 4 presents the cut-off values used for the industries under study.

	$e_{min,1}$	$LQ_{min,1}$	$e_{min,2}$	$LQ_{min,2}$
Chemical production	140.98	1	253.00	2
R&D	37.09	1	37.09	2
Logistics	232.80	1	694.50	2

Table 4: Cut-off values used for the application of the ECS-area identification method

The method applied to chemical production employment is presented step by step. As spatial weight matrix W, a matrix based on the borders shared by the subregions is used: for every row-subregion j, it is determined whether it shares a border with a column-subregion k, if this is the case $w_{jk} = 1$, and $w_{jk} = 0$ otherwise. Furthermore, for clarity reasons, the chemical production industry is defined to be industry $i = 1$, the subregions k are coded according to their four-digit postal code, and the combined areas are coded starting with 1.

2.1 *Remove subregions that do not have the potential to become (part of) an ECS-area.*

   (a) *Define $e_{min,1}$ and $LQ_{min,1}$.*
   These are equal to 140.98 and 1 respectively (see table 4).

   (b) *Code all subregions $k \in \{1, \ldots, K\}$: $b_{i,k} = 1$ if $e_{i,k} \geq e_{min,1} \wedge LQ_{i,k} \geq LQ_{min,1}$ and $b_{i,k} = 0$ otherwise. Remove all subregions k with $b_{i,k} = 0$.*
   Based on this coding, for 18 of the 502 four-digit postal code areas in the province $b_{1,k} = 1$. Only these subregions will be considered in step 2.2 of the method.

2.2 *In set of remaining subregions, combine subregions that are neighbors.*

[5] An LQ equal to 2 means that the share of industry employment is twice as large as the share of total employment in a particular subregion; all subregions that are that specialized in a particular industry should be ECS-area candidates. Instead of the cut-off values based on the 90th percentiles, O'Donoghue and Gleave (2004) propose to convert LQ values to standardized LQ values and use statistical significance to determine which LQ values are extremely high. Using a one-tailed approach (i.e. a cut-off value for the standardized LQs of 1.65) results in LQ cut-off values that do not differ much from the 90th percentiles used in this paper. Since cut-off values on original LQ values are easier to interpret, these are used in the analysis.

(a) *Determine* $B_{i,j,k} = w_{jk} b_{i,j} b_{i,k}$.
There are two joins of both two subregions that share a common border for which for both subregions $b_{1,k} = 1$, which relate to four nonzero B-values: $B_{1,5047,5048} = B_{1,5048,5047} = 1$, and $B_{1,5612,5651} = B_{1,5651,5612} = 1$.

(b) *For all* j *and* k *for which* $B_{i,j,k} = 1$, *define a new area* $m \in \{K+1, \ldots, M\}$, *with employment being the sum of the employment values of the different subregions:* $e_{i,m} = e_{i,j} + e_{i,k}$. *In addition, calculate* $LQ_{i,m}$.
Two combined subregions were created: $e_{1,1} = e_{1,5047} + e_{1,5048}$ and $e_{1,2} = e_{1,5612} + e_{1,5651}$.

2.3 *Identify ECS-areas based on absolute and relative employment levels.*

(a) *Define* $e_{min,2} \geq e_{min,1}$ *and* $LQ_{min,2} \geq LQ_{min,1}$.
These are equal to 253 and 2 respectively (see table 4).

(b) *Determine* $\beta_{i,c}$ *for all areas* $c \in \{1, \ldots, M\}$: $\beta_{i,c} = 1$ *if* $e_{i,c} \geq e_{min,2} \wedge LQ_{i,c} \geq LQ_{min}$, *and* $\beta_{i,c} = 0$ *otherwise. All areas* c *with* $\beta_{i,c} = 1$ *are identified as ECS-areas of industry* i.
Finally, twelve areas were identified as ECS-areas: one of these is the combined area 1, consisting of two four-digit postal code areas, and the other eleven ECS-areas only consist of one subregion each. The combined area 2 created in step 2.2 is not an ECS-area, neither are the four-digit postal code areas it consists of.

This procedure is shown graphically in figure 2. The figure presents (almost) all four-digit postal code areas based on their absolute employment in chemical production (X-axis) and their LQ based on the employment in chemical production (Y-axis). For clarity reasons, one area is not presented in figure 2 (the absolute employment in chemical production in that area is equal to 4512 with an LQ of 32). All blue diamonds represent one of the 502 postal code areas. In step 2.1, all subregions to the left or below the purple line (based on the first cut-off values) are removed from the analysis; the subregions to the right and above the purple line are considered in step 2.2. The combined areas created in step 2.2 are characterized by a blue square in the figure. Finally, the ECS-areas defined in step 2.3 are pictured as red diamonds, based on the second cut-off values (the green line in figure 2).

Figure 3 presents the ECS-areas of the chemical production industry geographically. The twelve ECS-areas, colored red in figure 3, consist of 13 of the 502 (3 percent) four-digit postal code areas and together account for 80 percent of all employment in chemical production in North Brabant.

Besides the red-colored ECS-areas, yellow-colored areas in figure 3 represent the subregions k for which either $LQ_{1,k} \geq LQ_{min,2} \wedge e_{min,1} \leq e_{1,k} < e_{min,2}$ or $e_{1,k} \geq e_{min,2} \wedge LQ_{min,1} \leq LQ_{1,k} < LQ_{min,2}$, characterized in figure 2 by yellow diamonds. Since these yellow-colored areas score high on one of the concentration measures and relatively high on the other one, these can be highly relevant for e.g. policy makers to increase the surface of ECS-areas of chemical production in the region.

The same analysis is conducted for the R&D industry and the logistics industry. Appendix C present the graphical and geographical representations of the ECS-areas for these industries. The nine ECS-areas for R&D account for 83 percent of all R&D employment in North Brabant; the 21 ECS-areas for logistics account for 51 percent of all logistics employment in the province. These numbers correspond with the locational Gini coefficients, which indicate that both the chemical production industry and the R&D industry were more spatially concentrated than the logistics industry.

The maps demonstrate that taking into account both absolute and relative spatial concentration indeed results in different areas than using either one alone. Especially the logistics industry shows large differences. As an example, compare figures 4(a), based on relative concentration only, with figure 4(b), presenting the ECS-areas for logistics. Many postal code areas with an LQ above 2 (colored dark red in figure 4(a)) are not identified as ECS-areas for logistics (colored red in figure 4(b)), due to the cut-off value on the absolute employment level. For example, the western

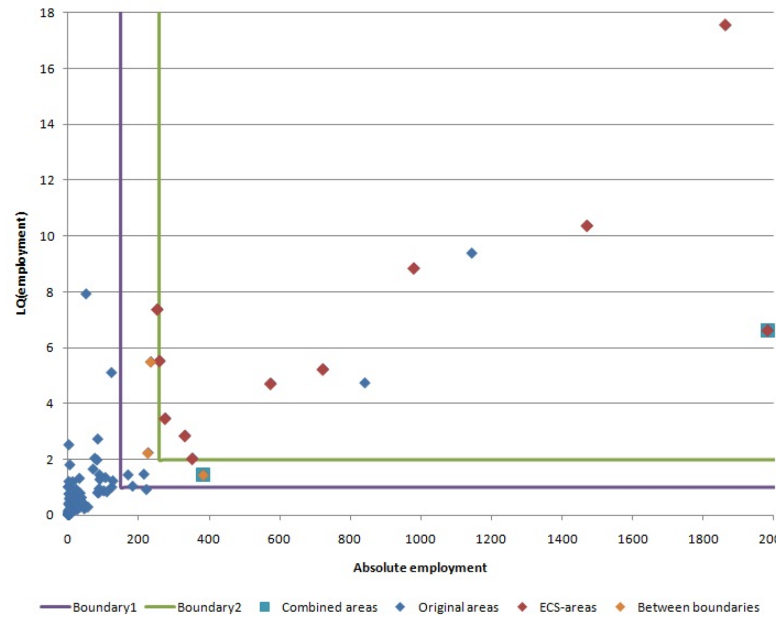


Figure 2: Graphical representation of the ECS-area identification method applied on the chemical production industry (without one area scoring very high on both axes)

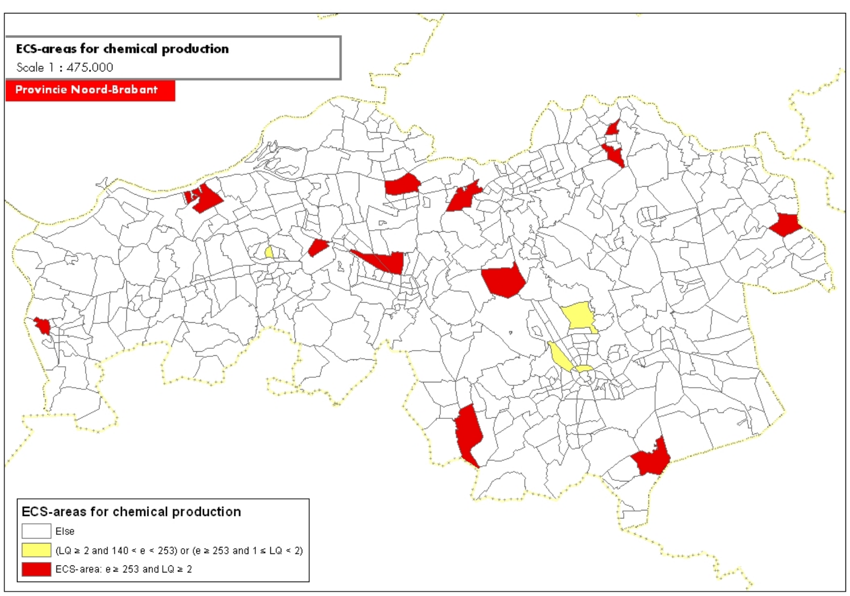
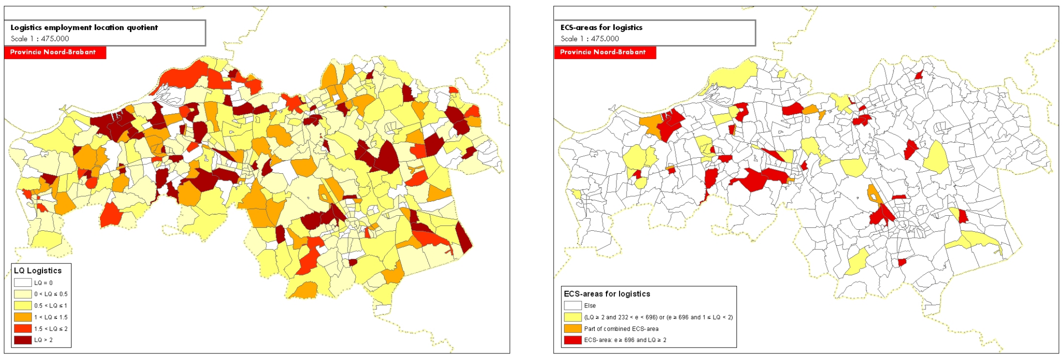


Figure 3: Geographical representation of the ECS-areas for chemical production

part of the province contains at least seven of these areas and does not contain any ECS-area. For absolute employment values, a similar conclusion can be made. The other industries analyzed in this paper show similar patterns, although the differences for the logistics industry are most apparent.



(a) Location quotient of logistics employment

(b) ECS-areas for logistics

Figure 4: Comparison of areas with only high location quotients and ECS-areas for logistics

5 Conclusions and Limitations

In this paper, a method to identify Employment Concentration and Specialization (ECS-) areas is developed. This method is relevant for various purposes: for evaluating whether agglomeration economies described by Marshall (1956), Krugman (1991), and Porter (2000) are also relevant on a low geographical aggregation level, for analyzing relationships between the location of these ECS-areas and other variables such as (intermodal) infrastructure availability, distances to specialized higher education institutions, land value and housing prices, and for analyzing spatial employment development over time. With the identification of these areas, the areas that are most important in terms of an industry's employment are identified. An analysis into the employment development in ECS-areas over time gives valuable insights on the regional dynamics of the economy. In addition, it would be interesting to compare the employment development within these areas with the employment development in other areas in the region to see whether spatial concentration is important for firm and industry growth.

This method takes away two important shortcomings of other available measures. First, where existing measures only consider relative spatial employment concentration (specialization), the newly developed method takes into account both relative and absolute spatial employment concentration. A comparison of ECS-areas and subregions with either a high location quotient or absolute employment shows that combining absolute and relative employment levels filters out subregions that are less relevant both from a policy making and location choice perspective. Second, although it is needed to subdivide the total region into smaller subregions to apply the method, the commonly known problems related to a-spatial measures are dealt with, by combining strengths of different existing methods. An application of the method to five different industries in a province in the Netherlands shows the advantages of the developed method.

An important practical advantage of the ECS-area identification method is that it is very easy to understand and apply. The method can be shown visually by means of a graph that categorizes the subregions based on the absolute and relative employment levels per subregion. This is an advantage in practice; policy makers can use this method to identify the important areas to invest in and substantiate the choice for these areas rather easily, and managers can use this method for location choices.

In addition, since the cut-off values on which the ECS-areas are identified can be chosen flexibly, the method can be adapted to different purposes. In this paper, it is argued that both absolute and relative spatial employment concentration are needed to define these areas, but not whether these are equally important. This may differ per decision maker; for some reasons, some may want to post stricter cut-off values on the absolute employment level than on the relative employment level or vice versa, which can easily be done using the newly developed method.

Although the method identifies ECS-areas based on one or more subregions defined, a limitation of the method still is that it divides the total region into several subregions. Hence, it is advised to only apply the method to predefined subregions, like e.g. postal code areas, due to the highly likely relationship between these areas and ECS-areas, due to e.g. zoning schemes. An alternative would be to use the theory of the distance-based measures to analyze these areas. However, in this paper a-spatial and neighborhood measures are chosen as a base for the developed method, since the distance-based methods have high data requirements and need many calculations and simulations. Nevertheless, it would be interesting to analyze whether the use of distance-based methods would result in other conclusions.

Another limitation of the method developed is that it only takes into account spatial concentration of employment, not of firms. It could be argued that synergies are more related to spatial concentration of firms than to spatial concentration of employment. However, from a policy perspective, employment concentration is highly relevant, since the employment level can be used as an estimate of the size of the firm and hence, of e.g. the transport volumes generated by the firm.

Finally, it has to be emphasized that the method is developed to be applied on a low geographical level. However, although not tested in this paper, the method could also be applied on a higher geographical level, e.g. NUTS 2 areas. Based on the purpose of the analysis, this could be of value, but differs from the applications described in this paper.

Acknowledgements

The authors are grateful to the Provincie Noord-Brabant for financial support, provision of data, and helpful discussions.

Appendices

A Logistics SBI codes

Name (based on ISIC)		**SBI 1993 code**
Chemical production		D: 24
R&D		K: 73
Logistics	Wholesale trade and commission trade (except of motor vehicles and motorcycles)	G: 51
	Freight transport by road (except for removal transport)	I: 60242
	Inland water ways freight transport	I: 61201-61203
	Cargo handling (except for sea transport)	I: 63112
	Storage and warehousing	I: 63121-63123
	Other supporting transport activities	I: 63401-63402
Retail trade		G: 52
Construction		F: 45

Table 5: SBI codes for the industries used in this paper

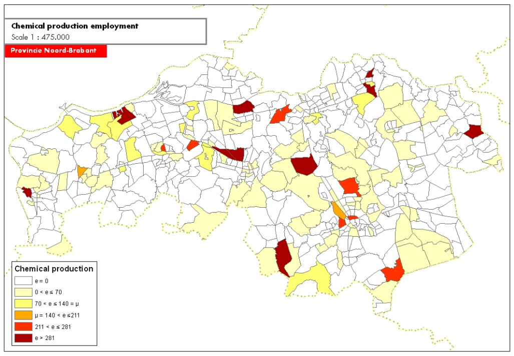
Table 5 presents the SBI1993 codes of the five industries used in this paper. Since no SBI code exists for logistics, several industries are combined to define the logistics industry. In this paper, the logistics industry is defined to consist of the following industries: wholesale trade, freight transport, cargo handling, storage and warehousing, and other supporting transport activities. From all firms characterized as wholesale trade firms based on the SBI code (51), the categories wholesale on a fee or contract basis (SBI = 511), wholesale of live animals (5123), and wholesale in computers, computer peripheral equipment, and software (5184) are excluded. It may be clear that indeed the firms in the first two categories are no logistics firms. In the last category mentioned, the wholesale in software is very dominant, definitely being no logistics, and hence, this category is also excluded from the analysis. In addition, to exclude wholesale trade firms that are only responsible for the administrative part of trade and not the physical part, all wholesale trade firms with less than ten employees were deleted from the database. Administrative trade firms are mostly relatively small and hence, this seems to be a valid method to exclude these firms. Furthermore, all firms in the above-described logistics categories with only one employee were excluded, since for these firms it generally holds that the firm's address is equal to the owner's address, which does not have anything to do with spatial concentration.

B Employment and LQ graphs

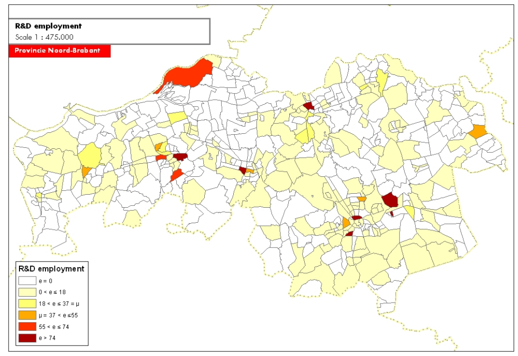
Figure 5 presents the distribution of absolute employment for the five industries geographically. For the categories used in this figure, the average number of employees per postal code area is used (column 4 of table 1; the boundaries are equal to 0.5 times the average, 1 times the average, etc.). In addition, figure 6 present the LQs of industry employment geographically.

C ECS-areas for R&D and logistics in North Brabant

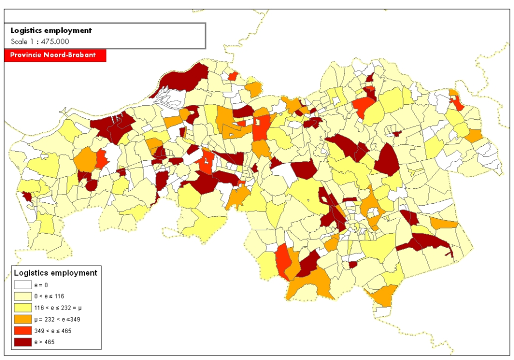
Figures 7 and 8 and figures 9 and 10 respectively present graphically and geographically the ECS-areas for the industries. In figure 7, again one area is left out, since this scores very high on both axes (absolute employment is equal to 4328 with an LQ of 76).



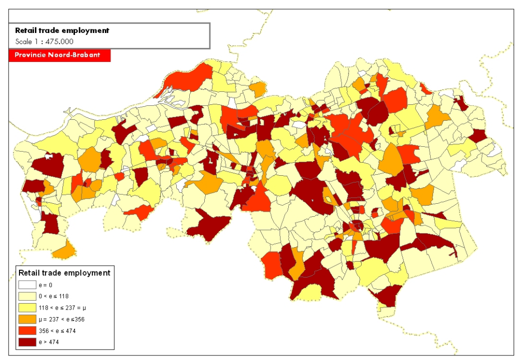
(a) Chemical production



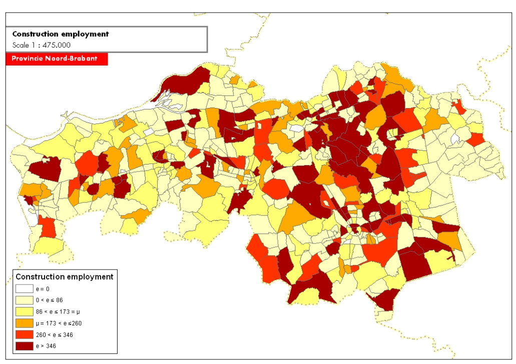
(b) R&D



(c) Logistics

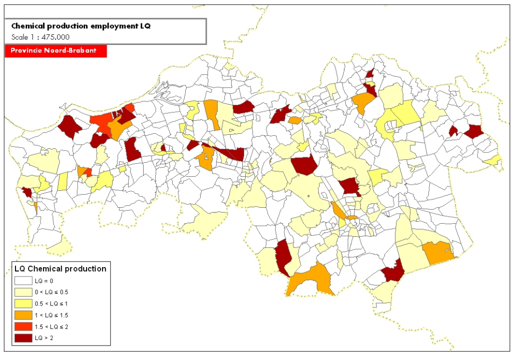


(d) Retail trade

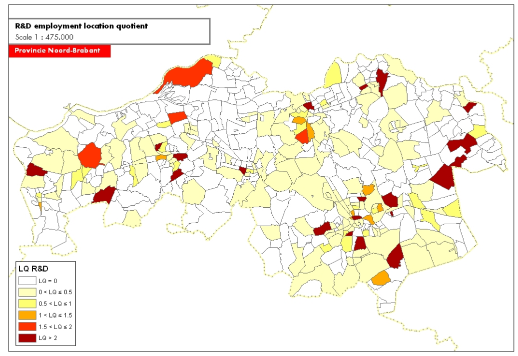


(e) Construction

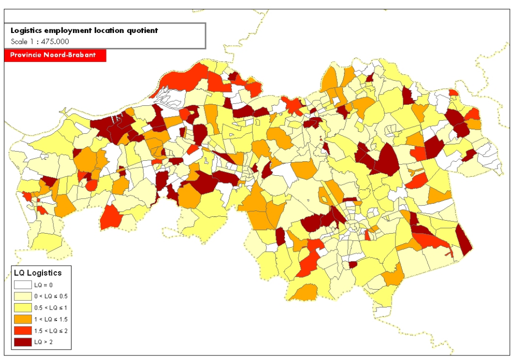
Figure 5: Spread of employment in absolute numbers over four-digit postal code areas



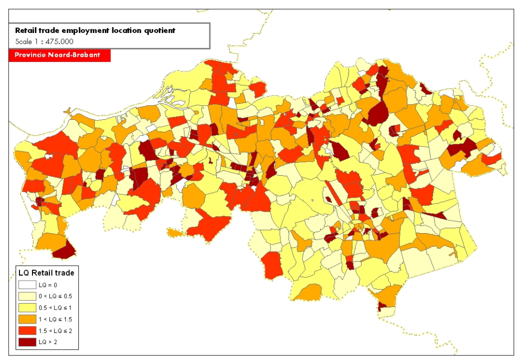
(a) Chemical production



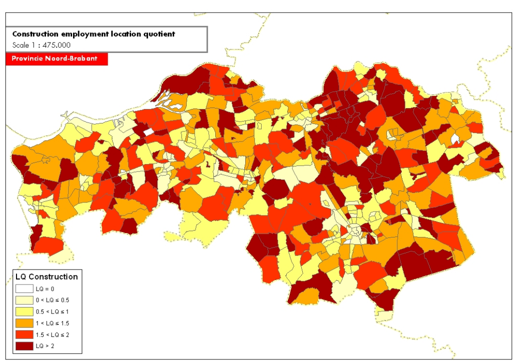
(b) R&D



(c) Logistics (also presented in figure 4(a))



(d) Retail trade



(e) Construction

Figure 6: Location quotient of employment per four-digit postal code area

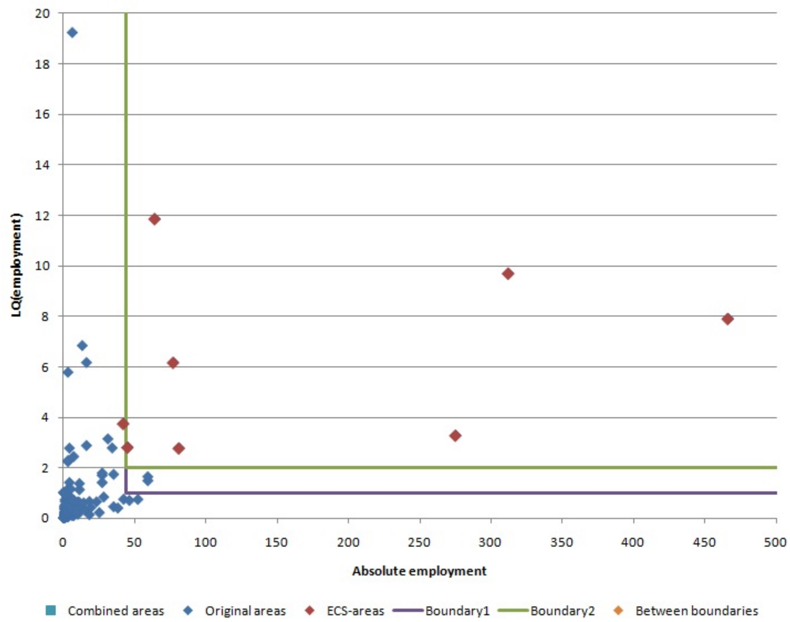


Figure 7: Graphical representation of the ECS-area identification method applied on the R&D industry (without one area scoring very high on both axes)

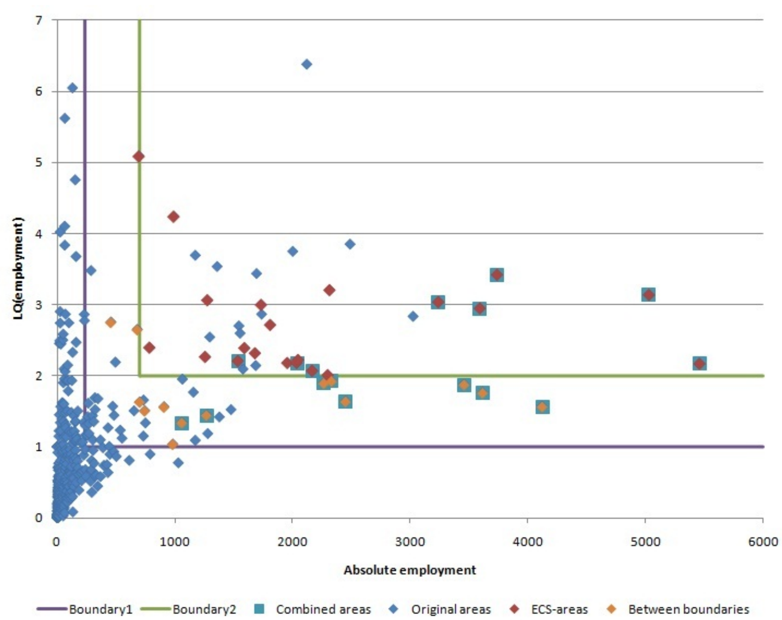


Figure 8: Graphical representation of the ECS-area identification method applied on the logistics industry

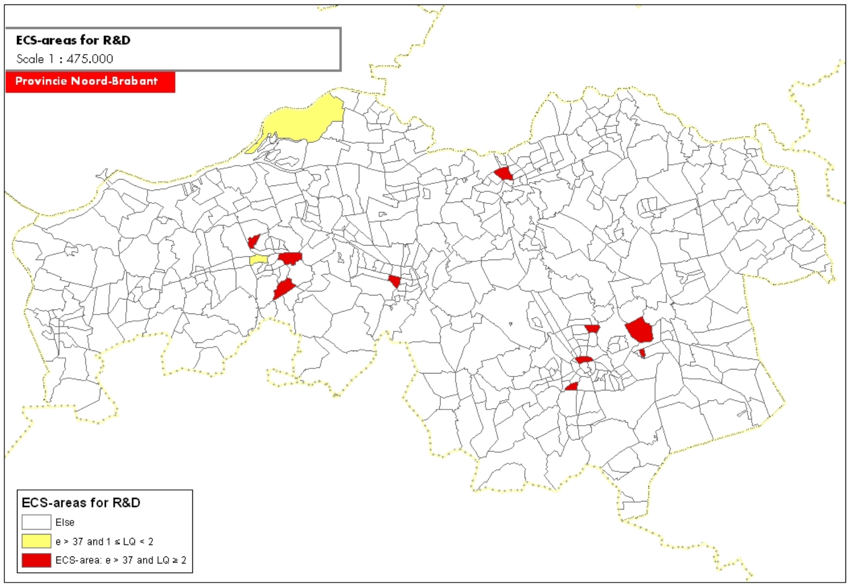


Figure 9: Geographical representation of the ECS-areas for R&D

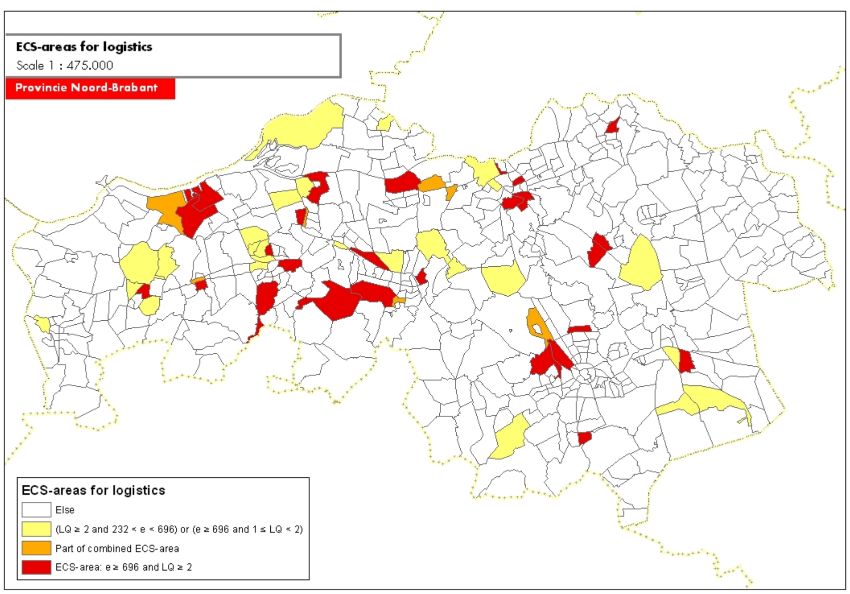


Figure 10: Geographical representation of the ECS-areas for logistics (also presented in figure 4(b))

References

Anselin L. (1995) Local indicators of spatial association - LISA, *Geographical Analysis* **27**(2), 93–118.

Anselin L. (1996) The moran scatterplot as an ESDA tool to assess local instability in spatial association, in Fisher M., Scholten H. and Unwin D. (Eds.) *Spatial analytical perspectives on GIS*, pp. 111–125, Taylor & Francis, London.

Arauzo-Carod J.-M. and Viladecans-Marsal E. (2009) Industrial location at the intra-metropolitan level: the role of agglomeration economies, *Regional Studies* **43**(4), 545–558.

Arbia G. (2001a) Modelling the geography of economic activities on a continuous space, *Papers in Regional Science* **80**, 411–424.

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

Bertinelli L. and Decrop J. (2005) Geographical agglomeration: Ellison and Glaeser's index applied to the case of Belgian manufacturing industry, *Regional Studies* **39**(5), 567–583.

Bickenbach F. and Bode E. (2008) Disproportionality measures of concentration, specialization, and localization, *International Regional Science Review* **31**(4), 359–388.

Braunerhjelm P. and Carlsson B. (1999) Industry clusters in Ohio and Sweden, 1975-1995, *Small Business Economics* **12**, 279–293.

Cliff A. and Ord J. (1973) *Spatial autocorrelation*, Pion, London.

Cutrini E. (2010) Specialization and concentration from a twofold geographical perspective: evidence from Europe, *Regional Studies* **44**(3), 315–336.

Devereux M., Griffith R. and Simpson H. (1999) The geographic distribution of production activity in the UK, *Regional Science and Urban Economics* **34**(5), 533–564.

Dewhurst J. and McCann P. (2002) A comparison of measures of industrial specialization for travel-to-work areas in Great Britain, 1981-1997, *Regional Studies* **36**(5), 541–551.

Duranton G. and Overman H. (2005) Testing for localization using micro-geographic data, *The Review of Economic Studies* **72**(253), 1077–1106.

Ellison G. and Glaeser E. (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach, *Journal of Political Economy* **105**(5), 889–927.

Eurostat (2010) NUTS - Nomenclature of territorial units for statistics, `http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction`.

Feser E. and Bergman E. (2000) National industry cluster templates: a framework for applied regional cluster analysis, *Regional Studies* **34**(1), 1–19.

Fratesi U. (2008) Issues in the measurement of localization, *Environment and Planning A* **40**, 733–758.

Fujita M. and Krugman P. (2004) The new economic geography: past, present and the future, *Papers in Regional Science* **83**, 139–164.

Fujita M., Krugman P. and Venables A. (1999) *The spatial economy. Cities, regions, and international trade*, MIT Press, Cambridge, Massachusetts.

Gordon I. and McCann P. (2000) Industrial clusters: complexes, agglomeration and/or social networks?, *Urban Studies* **37**(3), 513–532.

Guillain R. and Le Gallo J. (2010) Agglomeration and dispersion of economic activities in Paris and its surroundings: An exploratory spatial data analysis, *Environment and Planning B* **forthcoming**.

Henderson J. (2003) Marshall's scale economies, *Journal of Urban Economics* **53**, 1–28.

Jacobs J. (1969) *The economy of cities*, Vintage, New York.

Kim Y., Barkley D. and Henry M. (2000) Industry characteristics linked to establishment concentration in nonmetropolitan areas, *Journal of Regional Science* **40**(2), 231–259.

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

Lafourcade M. and Mion G. (2007) Concentration, agglomeration and the size of plants, *Regional Science and Urban Economics* **37**, 46–68.

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

Marshall A. (1956) *Principles of Economics*, MacMillan & Co, London, eight edn.

Martin R. and Sunley P. (2003) Deconstructing clusters: chaotic concept or policy panacea?, *Journal of Economic Geography* **3**(1), 5–35.

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

Moran P. (1950) Notes on continuous stochastic phenomena, *Biometrika Trust* **37**(1/2), 17–23.

Mori T., Nishikimi K. and Smith T. (2005) A divergence statistic for industrial localization, *The Review of Economics and Statistics* **87**(4), 635–651.

O'Donoghue D. and Gleave B. (2004) A note on methods for measuring industrial agglomeration, *Regional Studies* **38**(4), 419–427.

Porter M. (2000) Location, competition, and economic development: Local clusters in a global economy, *Economic Development Quarterly* **14**(1), 15–34.

Porter M. (2003) The economic performance of regions, *Regional Studies* **37**, 549–578.

Pouder R. and St. John C. (1996) Hot spots and blind spots: geographical clusters of firms and innovation, *Academy of Management Review* **21**(4), 1192–1225.

Rosenthal S. and Strange W. (2003) Geography, industrial organization, and agglomeration, *The Review of Economics and Statistics* **85**(2), 377–393.

Taniguchi E., Noritake M., Yamada T. and Izumitani T. (1999) Optimal size and location planning of public logistics terminals, *Transportation Research Part E* **35**(3), 207–222.

Torre A. and Rallet A. (2005) Proximity and localization, *Regional Studies* **39**(1), 47–59.

Van Soest D., Gerking S. and Van Oort F. (2006) Spatial impacts of agglomeration externalities, *Journal of Regional Science* **46**(5), 881–899.

Wallsten S. (2001) An empirical test of geographical knowlege spillovers using geographic information systems and firm-level data, *Regional Science and Urban Economics* **31**, 571–599.

Wennberg K. and Lindqvist G. (2010) The effect of clusters on the survival and performance of new firms, *Small Business Economics* **34**, 221–241.