Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 14-2012

Reinhold Kosfeld and Jorgen Lauridsen

Identifying Clusters within R&D Intensive Industries Using Local Spatial Methods

This paper can be downloaded from
https://www.uni-marburg.de/fb02/makro/forschung/magkspapers

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Identifying Clusters within R&D Intensive Industries Using Local Spatial Methods

Reinhold Kosfeld[1] and Jorgen Lauridsen[2]

Abstract. More recently, there has been a renewed interest in cluster policies for supporting industrial and regional development. By virtue of the linkage between growth and innovation, R&D intensive industries play a crucial role in cluster development strategies. Empirical cluster research has to contribute to the understanding the process of cluster formation. Some experiences with the use of local spatial methods like local Moran's I_i and Getis-Ord G_i tests in pattern recognition are already available. However, up to now the utilisation of spatial scan techniques in detecting economic clusters is largely ignored (Kang, 2010). In this paper, the performance of the above-mentioned local spatial methods in identifying German R&D clusters is studied. Differences in cluster detection across the tests are traced. In particular, the contribution of Kulldorff's spatial scan test in detecting industry clusters is critically assessed.

Keywords: Spatial Clusters, R&D Intensive Industries, Local Spatial Methods, Spatial Scan Test

JEL: R12, R15

1. Introduction

In Porter's sense a cluster is a geographically concentrated group of companies of related branches often forming linkages and alliances (Porter, 1998, 2000). In his papers Porter emphasises the role of clusters in regional competition. He shows in which way clusters can positively affect competition by increasing productivity and innovation. Because of the linkage between growth and innovation, R&D intensive industries play a crucial role in cluster development strategies. As clusters are credited with the creation of tangible economic benefits, an increasing number of researchers plead in favour of active cluster policy (European Commission, 2008). While there is a far-reaching consensus that the emergence of clusters depends on many factors which may differ from industry to industry, there is a dispute on the stability and growth effects arising from geographic concentration of firms producing in related branches (see e.g. Litzenberger, 2007).

Empirical cluster research has to contribute to the understanding the process of cluster formation. In particular for developing profound clusters strategies and assessing the limits cluster policy, knowledge of existing structures and tendencies is necessary. In these strategies, high-tech and research-intensive industries play a crucial role. Audretsch and Feldman (1996) and Feldman and Audretsch (1999) argue that industries with high innovation activity tend to cluster for exploiting benefits from tacit knowledge flows. In their view, spatial clusters primarily emerge from the rise of new economic knowledge. Because of economic knowledge with R&D, a skilled labour pool and the size of pool of basic science, industries where knowledge spillovers are relevant, are expected to concentrate more than

[1] Institute of Economics, University of Kassel, Germany
[2] Centre of Health Economics Research (COHERE), Southern University of Denmark, Denmark

Corresponding Author:
Reinhold Kosfeld, University of Kassel, Instítute of Economics, Nora-Platiel-Str. 4, D-34127 Kassel, Germany
Email: rkosfeld@wirtschaft.uni-kassel.de

other industries. The propensity to clustering of R&D intensive industries can be viewed as a special case of localisation economies arising from Marshall-Arrow-Romer (MAR) spillovers (see e.g. Neffke et al., 2008).

Krugman (1991) stresses that information flows and knowledge spillovers may be sensitive to geographic impediments. Since obstacles tend to rise with increasing distance, spatial clusters may be localised. If, however, geographic barriers are less relevant, the reach of tacit knowledge flows may be much larger. For regional policy the geographical level, at which clusters occur, is of prominent interest. While clusters on a small spatial scale are often primarily promoted by local governments and institutions, favourite development strategies of clusters on larger spatial scales may demand interregional cooperation.

Traditional concentration indices like the Gini coefficient, Theils's inequalitiy index or the Ellison-Glaeser index are 'aspatial' by construction (see e.g. Feser, 2000; Südekum, 2006; Südekum, 2006; Bickenbach and Bode, 2008). This means that these indices disregard relevant spatial information on the distribution of a geo-referenced variable. In particular, attribute values of adjacent regions are completely ignored. Moreover, the spatial scale of clustering formation is not taken into account.

Some experiences with local spatial methods in pattern recognition are already available. Le Gallo and Ertur (2003) utilise local indicators of spatial association to analyse the distribution of regional GDP per capita in Europe. Galloway and Robison (2008) identify of knowledge and innovation clusters using Getis-Ord G_i statistics. Feser et al. (2005), Lafourcade and Mion (2007) and Kies et al. (2009) demonstrate the potential of local spatial methods in identifying economic clusters and spatial heterogeneity in geographical space. However, while usually local Moran's I and Getis-Ord G_i statistics are applied in detecting economic clusters, up to now, spatial scan techniques are largely ignored (Kang, 2010). In this paper, the performance of the above local spatial methods in identifying German R&D clusters is studied. Differences in cluster detection across the tests are traced. In particular, the contribution of Kulldorff's spatial scan test in detecting industry clusters is critically assessed.

The paper is organised as follows. In section 2, cluster detection methods are presented. Section 3 deals with data issues. In section 4, the clustering trends in R&D intensive industries are examined at different spatial scales. Main results of local spatial data analysis in identifying German R&D clusters are outlined in section 5. Section 6 discusses the results and concludes.

2. Cluster Detection Methods

Global tests of spatial autocorrelation like Moran's I and Geary's c[3] or spatial association[4] like the Getis-Ord G statistic can reveal overall spatial trends, but not the existence and location of spatial clusters. A matching of locational similarity and attribute similarity gives reason for positive spatial autocorrelation. In this case, some clustering of high or low values of the attribute variable will occur across space. By contrast, negative spatial autocorrelation arises from dissimilar values of an attribute in nearby regions. When values of a geo-

[3] As Geary's c is strongly linked to Moran's I, in this study only the latter autocorrelation coefficient is considered.

[4] While measures of spatial autocorrelation are based on second-order moments of the distribution of a geo-referenced variable, indicators of spatial association may be defined more generally.

referenced variable at a given location do not depend on values observed in nearby regions, space does not matter. This independence of values of an attribute occurring in regional arrangements indicates spatial randomness.

Although global indicators of spatial association are eligible to whether mapped data exhibit an organised pattern, care must be taken in interpreting the results. The global trend of spatial autocorrelation may mask spatial heterogeneity. Not only the strength but even the direction of spatial dependency can vary significantly across space. Atypical regions may exert considerable influence on the overall picture. Spatial outliers occur when in regions dissimilar values compared to their neighbourhoods are observed. Also in case of positive global spatial autocorrelation spatial clustering of high values ("hot spots") and low values ("cold spots") may occur in different areas. In this study, global spatial autocorrelation analysis is mainly conducted to establish the scale at which formation of clusters most likely takes place.

Local spatial indicators like Moran I_i and Getis-Ord G_i^* statistic make use of the possible range of spatial interaction. This applies similarly to Kulldorff's spatial scan statistic where the maximal size of the scanning window needs to be fixed.

The local Moran coefficient I_i ,

$$(1)\quad I_i(d) = \frac{1}{s^{*2}} \cdot (x_i - \bar{x}) \cdot \sum_{j=1}^{n} w_{ij}(d) \cdot (x_j - \bar{x}),$$

compares the observed value of an attribute variable in region i with the weighted sum of values in its surrounding (Anselin, 1995). s^{*2} is the descriptive variance (with factor 1/n) of the whole sample. All spatial units within a given distance d from the geographic centroid define the surrounding of a region i. The weights $w_{ij}(d)$ of these regions are assigned the value 1 and 0 for all other regions:[5]

$$(2)\quad w_{ij}(d) = \begin{cases} 1, \text{ if } d_{ij} < d \text{ and } i \neq j \\ 0, \text{ otherwise} \end{cases} .$$

Usually the weights are row-standardised: $\tilde{w}_{ij}(d) = w_{ij}(d)/\sum_h w_{ih}$. The weighted sum then becomes a weighted average.

Local industry concentration presupposes positive I_i values for the regions of a contiguous area. However, on the basis of the local Moran coefficients alone one cannot differentiate between hot and cold spots as positive I_i's indicate spatial clustering of similar values (high or low values). This can be done by using the classification of the Moran scatterplot. The program GeoDa enables an identification of hot spots by local Moran tests.[6]

The identification of spatial clusters presupposes significant deviations of the observed I_i values from the expected I_i values $E[I_i(d)] = -W_i^*/(n-1)$ with $W_i^* = \sum_j w_{ij}^*(d)$. However, as the distribution of I_i is unknown and does not approach the normal distribution, test of significance are usually based on Monte Carlo methods. For this, Anselin (1995) proposed a conditional randomisation approach where the attribute value of the i*th* region is held constant, while all other data values are permutated over the remaining n-1 regions. In case of

[5] Instead of the distance concept, the weights can alternatively be based on the concept of contiguity (Anselin, 1988, pp. 17).
[6] GeoDa is used here for cluster detection with local Moran and Getis-Ord statistics (Anselin, 2003).

K more extreme I_i values than the observed one in S permutations, an approximate significance level is given by (K+1)/(S+1).The permutation method has to be employed for all n spatial units of the study area.

By exploiting information from the Moran scatterplot, spatial clusters identified by the I_i statistics can be classified as with the Getis-Ord G_i^* statistics as hot and cold spots. There are, however, differences between the I_i's and G_i^*'s in identifying HH (high-high) and LL (low-low) clusters. In contrast to the I_i's the attribute value of the considered region is treated with G_i^*'s in the same way as the neighbouring values. While the new Gets-Ord G_i^* indicators are standardised, local Moran I_i statistics are not.

The Getis-Ord G_i and G_i^* statistics differ from each other with respect to the treatment of the i*th* region. While the i*th* region's attribute value is included in G_i^* it is not in G_i. In measuring local industry concentration, the G_i^* statistic provides the relevant concept as employment in the i*th* region and its surrounding contributes to clustering. Thus, although G_i is closer to the global Getis-Ord G statistic, we only consider G_i^* for identifying spatial clusters.

The original local Getis-Ord G_i^* statistic (Getis and Ord, 1992),

$$(3)\quad G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}^*(d) \cdot x_j}{\sum_{j=1}^{n} x_j},$$

is like the global Getis-Ord G statistic restricted to geo-referenced variables with a natural origin and positive values. The binary spatial weights $w_{ij}^*(d)$ are defined according to (2) but with $w_{ii}^*(d) = 1$ instead of $w_{ii}(d) = 0$. The G_i^* statistic gives the sum of attribute values in i*th* region and the surrounding regions within a distance of d kilometres relative to the sum of all values of the considered variable. Significant deviations of the G_i^* values from their expected value $E[G_i^*(d)] = W_i^*/n$ with $W_i^* = \sum_j w_{ij}^*(d)$ indicate local spatial clustering. If the deviation is significantly positive, the spatial cluster is called hot spot.

In their 1995 paper, Ord and Getis redefined G_i and G_i^* statistics. The new indicators of spatial association are more general as they not restricted to positive variables with a natural origin. Moreover, they can also be used with non-binary spatial weights. More precisely is the new G_i^* statistics a standardised variate of the form

$$(4)\quad G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}^*(d) \cdot x_j - W_i^* \cdot \bar{x}}{s \cdot [(n \cdot S_{1i}^* - W_i^{*2})/(n-1)]^{1/2}}$$

with $W_i = \sum_j w_{ij}(d)$ and $S_i = \sum_j w_{ij}^2(d)$. The statistics $\bar{x}$ and s denote the mean and the standard deviation of the whole sample. Significant positive values of the new G_i^* statistics identify hot spots. In the case of the ordinarily observed skewed distribution of the concentration variable, the G_i^* statistics are asymptotically normally distributed. The normal approximation improves with an increasing number of neighbours. In GeoDa significance of the G_i^* statistics is assessed by Monte Carlo simulation.

Kulldorff's spatial scan test (Kulldorff and Nagarwall, 1995; Kulldorff, 1997) determines the most likely cluster as well as secondary clusters by a likelihood ratio approach. The test statistic is obtained by scanning the surroundings of each centroid of a region (e.g. district, county, travel-to-work area) for cases (e.g. employment). To ensure comparability with local Getis-Ord tests we assume circular scanning windows that are increased from zero until a given threshold distance is reached. This variant is preferable in identification of economic clusters when knowledge on the strength of spatial interaction is available.[7]

Let M_z be the number of observed cases and N_z the population size in a circular zone Z. Further the total number of cases and population in the study area are denoted by M and N, respectively. Under the assumption that the events are generated by a Poisson process, the likelihood ratio is given by

$$(5) \quad LR_Z \propto \left(\frac{M_Z}{\hat{\lambda}\cdot N_Z}\right)^{M_Z} \cdot \left(\frac{M-M_Z}{M-\hat{\lambda}\cdot N_Z}\right)^{M-M_Z} \cdot I\left(M_Z > \hat{\lambda}\cdot N_Z\right).$$

with $\hat{\lambda} = M/N$ as the estimated incidence rate under the null hypothesis of no spatial clustering. The indicator function I takes the value 1 if the observed number of cases, M_z, exceeds the expected number of cases, $\hat{\lambda}\cdot N_Z$, inside zone Z. In this case the relative risk RR_z of an event occurring within the circle,

$$(6) \quad RR_Z = \frac{M_Z}{\hat{\lambda}\cdot N_Z}$$

is larger than one. Thus, the specification of I initiates a scan for high-value clusters (hot spots) instead of a test for either high- or low-value clusters.

For fixed M and N the likelihood ratio LR_z is an increasing function of the number of cases in zone Z. The most likely cluster is achieved by maximizing LR_z over all possible zones and centroids of the areal units. With area data, the number of windows to be scanned for each location is usually considerably lower than the number of regions as all events are assigned to the regional centroids. Each secondary cluster is obtained conditional to the clusters detected in the previous stages. In this way, the problem of dependency in multiple testing procedures present in predecessors like Openshaw's Geographical Analysis Maschine (GAM) (Openshaw

[7] Usually, an upper limit for the size of the scanning window is specified in form of the maximal percentage of the population of risk. However, such a choice seems to be quite arbitrary. While Kulldorff and Nagarwall, (1995) suggest to include maximal 20 per cent of the population, the SatScan manual recommends a threshold of 50 per cent (Kulldorff, 2003). Here the scan test is conducted with SatScan by specifying the threshold distance obtained from global spatial autocorrelation analysis.

et a., 1987) or Turnbull's Cluster Evaluation Permutation Procedure (CEPP) (Turnbull et al., 1990) is avoided (Kulldorff and Nagarwalla, 1995).

Testing for significance of the maximised likelihood ratio LR_z is done by Monte Carlo simulation. The scan statistic is the likelihood ratio maximised over all zones with different sets of events of all regional centroids in study region up to a given threshold. The distribution of the test statistic is obtained by multinomial randomisation under the null hypothesis. With K+1 as the rank of the maximised likelihood ratio of the real data set in a large number of random replication S, the p value of the test is (K+1)/(S+1). A potential industry cluster is located, if the p value is lower than the nominal significance level α. Overlapping clusters are usually excluded. If they exist, the exact boundaries of a cluster are difficult to establish.

3. Data

We explore spatial patterns of German R&D intensive industries using 2006 employment data from the regional data base of the Federal Statistical Office Germany. The regional data base comprises the number of employees subject to social security obligations for various levels of regional and sectoral disaggregation. In particular for identifying local industry clusters, highly regionally disaggregated data are required. Employment data are available at the district level. However, because of secrecy, the number of employed are only reported for districts where three or more firms of the industry are located. Missing data are estimated by the average employees of the branch in the state. This method is also applied for completing fragmentary employment data in the electrical industry for the districts in the state of Baden-Württemberg.

In all industrial sectors, firms spend a part of their revenue on research and development (R&D). Most of the almost 52 billion € German R&D expenses in 2006 come from large companies. Only an estimated share of 9 per cent goes on small and medium enterprises (SME) (Grenzmann et al., 2009). Four industries account for roughly two thirds of the private R&D expenses. The sector automobile manufacturing is clearly dominating with a share of about one third. It is followed up by the electrical industry with 20 per cent, the chemical industry with 17 per cent and the mechanical engineering industry with 9 per cent. Because the individual contributions of expenses on research and development of these sectors are distinctly larger than those of all other branches, they are called F&D intensive industries.

The study region consists of 439 German districts that vary considerably in size. The sizes of the districts range from 35.63 km^2 (city of Schweinfurt) to 3058.23 km^2 (rural district Uckermark). In view of these differences in size, spatial employment patterns in R&D industries can easily become distorted on the basis of the original count data. When the employees were randomly distributed across the study region, local clusters may be erroneously detected in districts whose area is, for instance, twice or thrice of the territorial average. Favouring large areal units can be avoided by converting count data into ratios. In the special case of a density indicator, count data are related to the territorial sizes of the regions. With E_{ik} as the observed number of employed persons in region i and industry k and A_i region's size in km^2 the employment density is defined by $ED_{ik} = E_{ik}/A_i$.[8] In context of

[8] Haining (2003, pp. 194) discusses also standardised rates defined by the ratio of some the number of events and a special concept of the population at risk. With an appropriate choice of the population at risk, such rates can be interpreted as location coefficients suitable for establishing regional specialisation.

the spatial scan procedure, the quantity A_i represents the population at risk (cf. Coulston and Riitters, 2003).

The regional database of the Federal Statistical Office Germany, CD "Statistik regional 2010", includes data on the number of plants and employees subject to social security contribution in 439 German districts. All four R&D intensive sectors belong to the manufacturing industry (section D) of the German Classification of Economic Activities (WZ 2003). Up to the four-digit sectors, this classification corresponds with the NACE Rev. 1.1 classification[9] that is based on the International Standard Industrial Classification of all Economic Activities (ISIC Rev. 3.1) of the United Nations. Table 1 summarises descriptive statistics of the variables used in this study.

Table 1: Descriptive statistics

2006	Mean	Stand. deviation	Minimum	Maximum
Employment DG24	1035.4	2574.7	0	39322
Employment DK29	2149.7	2776.3	0	22319
Employment DL	1890.3	1810.5	0	27503
Employment DM34	1550.5	3308.7	0	40627
Territorial size	813.2	597.4	35.3	3058.2
Empl. Density DG24	4.221	26.068	0	505.205
Empl. Density DK29	6.670	18.691	0	315.717
Empl. Density DL	6.193	13.246	0	143.499
Empl. Density DM34	4.688	13.463	0	195.925

Notes:
Employment data: Number of employees subject to social insurance contributions in 439 German districts
Source: Employment and territorial size: CD "Statistik regional 2010", German Federal Statistical Office; employment density: Own calculations
DG24: Chemical industry, DK29: Mechanical engineering industry, DL: Electrical industry, DM34: Automotive industry

4. Clustering trends in R&D intensive industries

In order to explore overall spatial dependence and clustering in F&D intensive industries Moran's I and Getis-Ord G statistic is employed for a range of distances. Specifically the tests on global spatial association are conducted within a distance band from 20 to 100 km by increments of 5 km. The assessment of significance is always based on 999 Monte Carlo replications. The testing results must be interpreted cautiously for distances lower or equal 40 km due to the occurrence of empty neighbourhood sets. While the values of the Moran coefficient are comparable across industries and distances because of its unchanged expected value,[10] the expectation of the Getis-Ord G statistic varies considerably. For that reason only the standardised values are reported for the latter measure. All tests are done using regional employment densities as the relevant attribute for detecting spatial clusters in RD intensive industries.

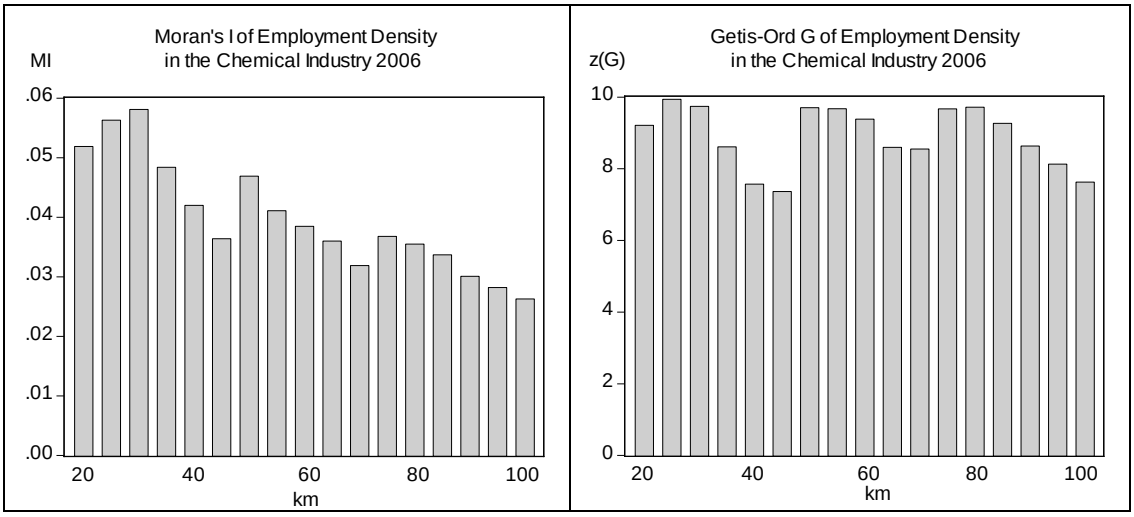
Significant positive Moran coefficients indicate that high or low employment within an industry tends to cluster in space. However, whether significant values of Moran's I have to

[9] Nomenclature des Activités Economiques dans les Communautés Européennes (NACE).
[10] In the case of missing neighbours, however, the expected value of Moran's I is computed with the "reduced" sample size.

be ascribed to hot spots, cold spots or both can be signified by the local counterparts. The existence of hot spots in R&D activity can be directly inferred from the outcomes of the Getis-Ord G test.

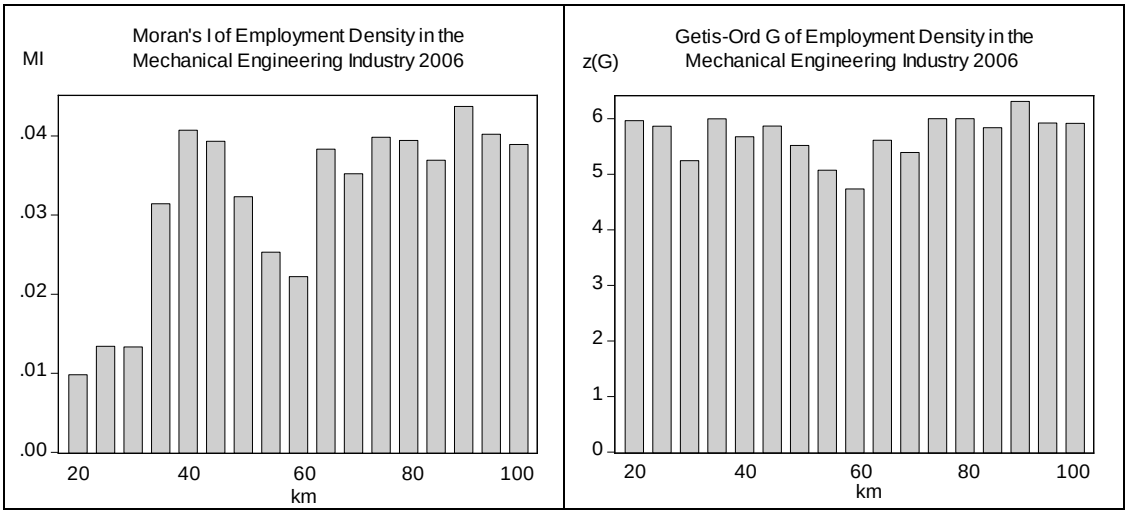
Figure 1 displays Moran's I and the Getis-Ord G statistic of employment density for the chemical industry at different spatial scales. Note that the maximal value Moran coefficient at a distance of 30 km is based on a substantial loss of degrees of freedom as with this radius 60 regions stay without any neighbourship. As a consequence, the maximal Moran coefficient of 0.047 under the condition of a non-empty neighbourhood set at a distance of 50 km is of higher significance.

Figure 1: Moran's I and Getis-Ord G for employment in the chemical industry



While Moran's I tends to decrease with increasing distance, the standardised Getis-Ord G statistic shows no clear pattern. However, the maximal z(G) value as well arises at a distance where missing neighbours occur for a lot of regions. When each region is assigned at least one neighbour, the highest and second highest significance for the G statistic is reached at d = 80 and d = 50, respectively.

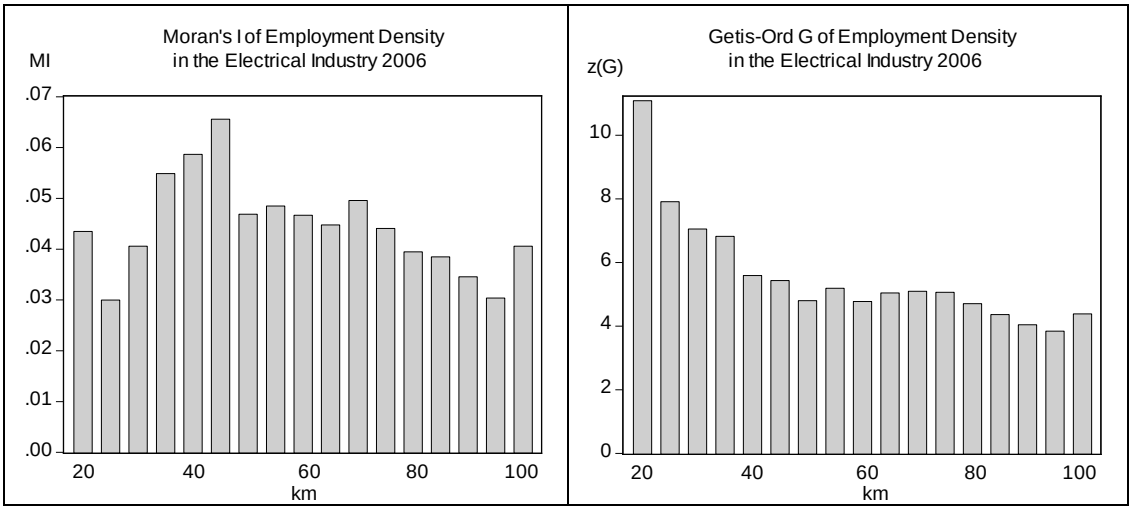
Figure 2: Moran's I and Getis-Ord G for employment in the mechanical engineering industry



A completely different pattern emerges in the mechanical engineering industry (Figure 2). Here Moran's I tends to increase with growing distance. The maximal and most significant

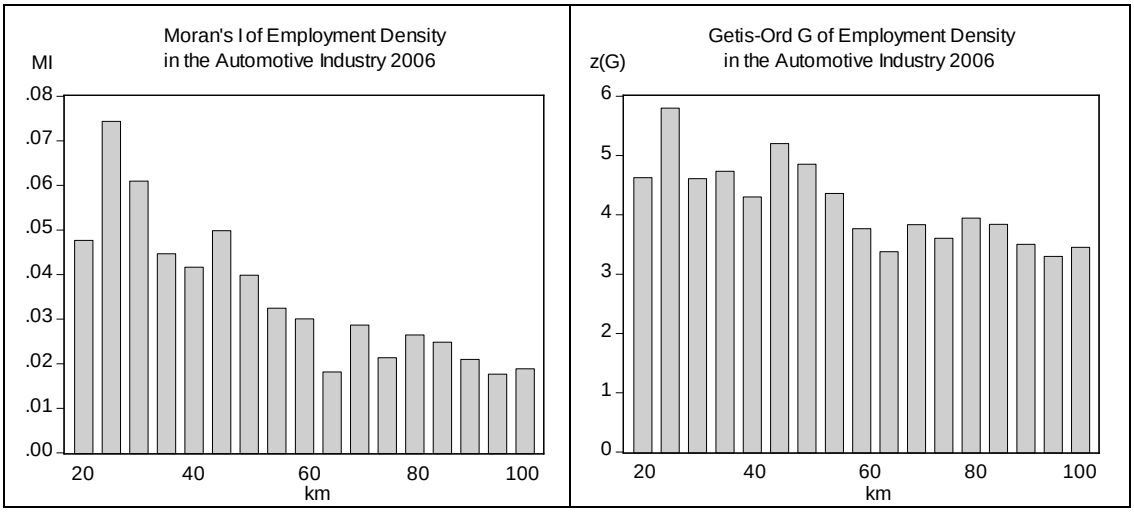
value of 0.044 is reached at a distance of 90 km. Although the standardised Getis-Ord G statistic is relatively stable, its highest value is achieved at the same distance. We will consider this distance for finding spatial clusters of activity in the mechanical engineering industry.

Figure 3: Moran's I and Getis-Ord G for employment in the electrical industry



For employment density in the electrical industry, the MI values with distances lower or equal than 30 km are nonsignificant. The highest Moran coefficient of 0.066 is measured at a distance of 45 km This outcome matches well with the testing result for the Getis-Ord G statistic within the range $45 \leq d \leq 100$ (Figure 3). The highest z(G) value at a distance of 20 km is not well grounded as it is based on less than a half of the regions.

Figure 4: Moran's I and Getis-Ord G for employment in the automotive industry



In the automotive industry, the Moran statistics tend to taper off with increasing distance (Figure 4). The highest values of Moran's I are observed at distances of 25 and 30 km. However, because of the reduction of effective sample size by nearly one third, significance fails to be proved at the 5% level. In the restricted range from 45 to 100 km, the maximal MI value of 0.05 occurs with a distance of 45 km. The same preferable spatial scale for the manufacture of motor vehicles and trailers is obtained from the global Getis tests.

Similar adverse effects in testing for spatial autocorrelation with lower threshold distances (15 – 35 km) are also reported by Kies et al. (2009). Apart from the cases of isolated regions, Moran's I is significant at the 5% level for all R&D-intensive industries. This means that high or low employment within the industries tend to cluster in space. In particular we wish to discover high employment clusters. Because the G statistics are significant and positive for all R&D intensive industries, the spatially autocorrelated attribute variable at least partly reflects the presence of hot spots. In all cases the testing results clearly reject the hypothesis of a completely spatially random (CSR) distribution of R&D employment. The spatial processes generating specific clustering patterns in R&D intensive industries seem to be at work at different scales. The diminishing strength of spatial autocorrelation observed in all but the mechanical engineering industry may be indicative for highly localised spillover effects.

5. Spatial clusters in R&D intensive industries

As the hypothesis of spatial randomness is clearly rejected for both employment indicators, we now take a closer look at the spatial patterns of employment in R&D intensive industries. In particular we are interested in identifying hot spots of R&D activity. Thus we test for the existence of local clusters in the spatial distribution of employment in innovative branches. The knowledge of spatial employment patterns in R&D industry is a core requirement for policymakers in shaping regional and cluster policy.

In principle, regional clusters of R&D activity could be identified at a broad range of spatial scales. However, global spatial analysis has revealed varying tendencies to cluster across industries as well as at different spatial scales. The extent of the neighbourhoods of the regions affects the strength of spatial autocorrelation of the attribute variable. In our local spatial analysis of employment distribution we will concentrate on preferable industry-specific scales suggested by Moran's I and the global Getis-Ord G statistic.

Depending on the industry, with Kulldorff's approach the number of significant R&D clusters varies between 50 and 70. In cluster research it is argued that only clusters with a critical mass contribute to regional growth and development (Wares and Hadley, 2008). Thus only secondary clusters with high industry-specific employment are portrayed. The threshold is fixed by the factor ten. Usually this requirement is met for the most significant secondary clusters.

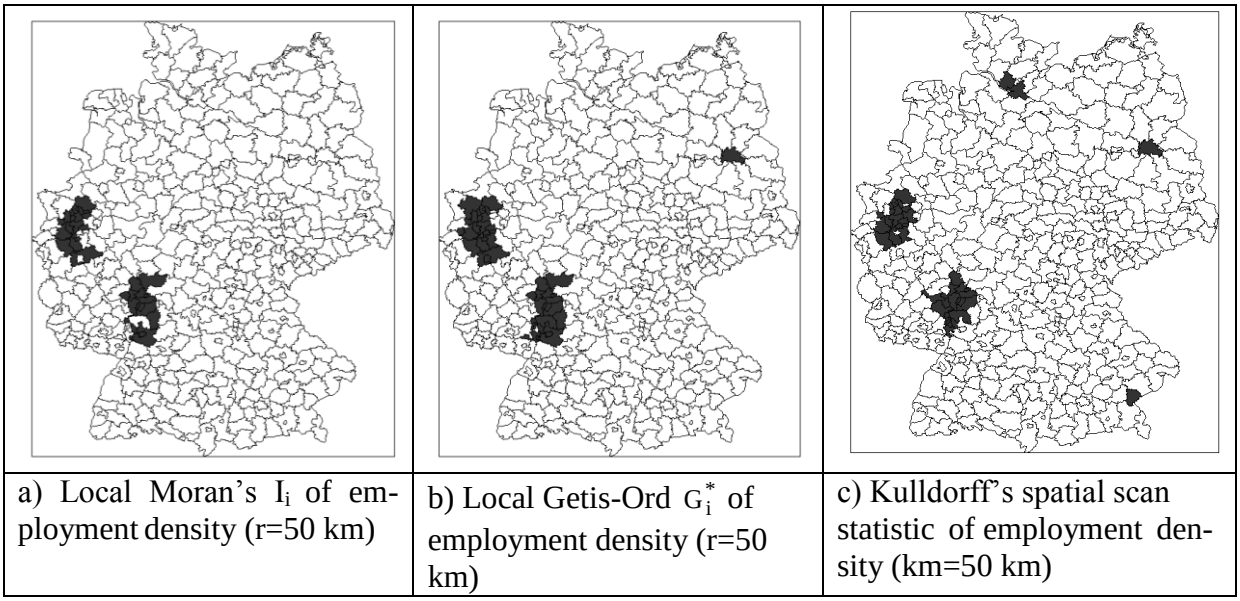
According to global trend analysis we choose a radius of 50 km around the regional centres as the preferable spatial scale for identifying local clusters in the chemical industry (DG24). Although the number of interconnected high density areas discovered by the local tests is not unique, two comparably large clusters of dense employment are detected with all methods.[11] The cluster in North Rhine-Westphalia located in the Rhine-Ruhr area comprises at 61,000 (15%) of total employment in the chemical industry. The southern cluster that extends from the Rhine-Main area to south-east Rhineland-Palatinate and north-west Baden-Württemberg is of comparable size. About 56,000 (13%) of the total employees in the industry are concentrated in this area.

The Rhine-Ruhr cluster is discovered as the most likely cluster with Kulldorff's spatial scan tests (log LR=122335.3, p=0.0000). It is presented by diversified cities like Cologne,

[11] With Kulldorff's approach, depending on Only secondary clusters with a critical mass are portrayed with Kulldorff's scan statistic.

Düsseldorf and Essen, but also mainly chemical locations like Leverkusen and Neuss. The relative risk of 14.1 indicates that the likelihood of engagement in chemical inside this area is about fourteen times higher than outside. There is a conspicuous overlapping with I_i-based cluster. Both clusters are only somewhat smaller than the cluster found by local G_i^* tests. In particular, the large-scaled district of Wesel is not enclosed in the former sets of regions. The Rhine-Main area is the secondary cluster with the second highest log likelihood ratio (log LR=79517.1, p=0.0000) and a relative risk of 10.0. Here also diversified cities (Frankfurt/ Main, Darmstadt and Mannheim) coexist with more specialised ones (Ludwigshafen and surroundings).

Figure 5: Spatial employment patterns in the Chemical Industry



a) Local Moran's I_i of employment density (r=50 km)	b) Local Getis-Ord G_i^* of employment density (r=50 km)	c) Kulldorff's spatial scan statistic of employment density (km=50 km)

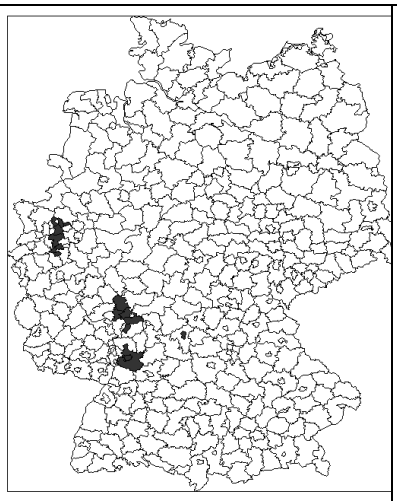
In both clusters the headquarters of international companies specialised in manufacture of coke, refined petroleum products or nuclear fuel are located. The state of Berlin is additionally identified as a high density area of employment in the chemical industry by the G_i^* test. Berlin, the extended state of Hamburg and the Bavarian district of Altötting are disclosed as highly significant secondary clusters by the spatial scan tests. These areas may at least be viewed as important chemical locations as they are dense with more than ten times as much employees compared with an average.

Clustering processes in the mechanical engineering industry (DK29) seem to take place at a larger spatial scale than in the chemical industry. Both global association measures, Moran's I and the Getis-Ord G statistic, indicate strongest spatial dependence for neighbourhoods within 90 km circles around the regional centres. Seemingly three high employment clusters are identified with all local tests. They differ considerably in size and partly as well in location.

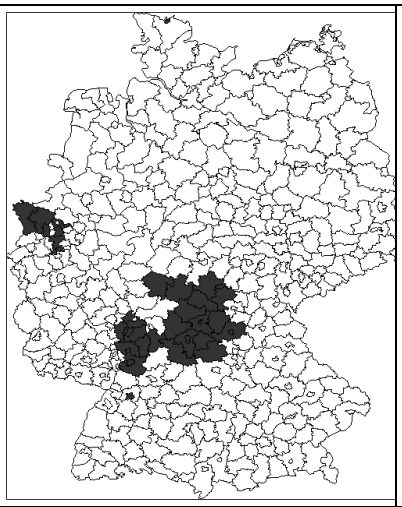
Actually, there are two separated and one combined engineering cluster revealed by Kulldorff's spatial scan statistics. The large southern area of dense employment consists of a northern part extending from southern Rhineland-Palatinate to the Black Forest in Baden-Wuerttemberg (Stuttgart) and a southern part ranging from eastern Baden-Wuertemberg to the southeast Bavaria (Augsburg). While the plants in this branch belonging to the northern part employ nearly 160,000 (17%) of total workers in the mechanical engineering industry, about 12,500 (13%) employees are occupied in companies of the southern cluster. The risk factors of 4.2 and 2.4 for subclusters are only moderate. Although the northern part of the southern

cluster is found as the most significant cluster with the scan tests (log LR=100309.1, p=0.0000), only a small section of this area is identified by the I_i and G_i^* statistics as a hot spot.

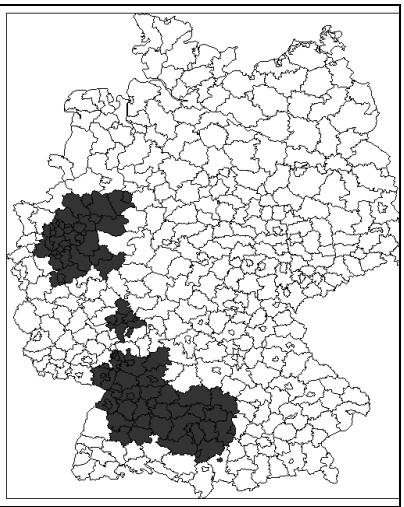
Figure 6: Spatial employment patterns in the Mechanical Engineering Industry



a) Local Moran's I_i of employment density (r=90 km)



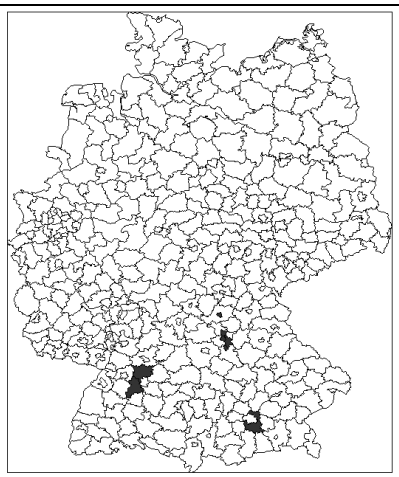
b) Local Getis-Ord G_i^* of employment density (r=90 km)



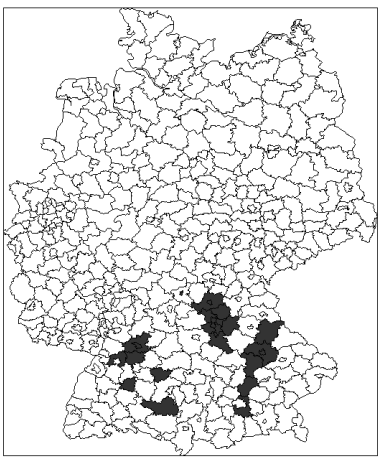
c) Kulldorff's spatial scan statistic of employment density (km=90 km)

The most significant secondary cluster with about 12,500 (13%) workers in the mechanical engineering sector is located in the Rhine-Ruhr area (log LR=63893.5, p=0.0000). A small part of it is as well identified by the other local tests. This also holds for the Rhine-Main cluster (log LR=14098.6, p=0.0000) which is part of a larger G_i^*-based cluster extending to the south of Hesse. The relative risk of mechanical engineering activity in these clusters is quantified between 3.5 and 4.0. An additional engineering cluster extending from Middle Hesse to Northern Bavaria identified by the G_i^* tests is neither confirmed by the I_i nor by the scan tests.

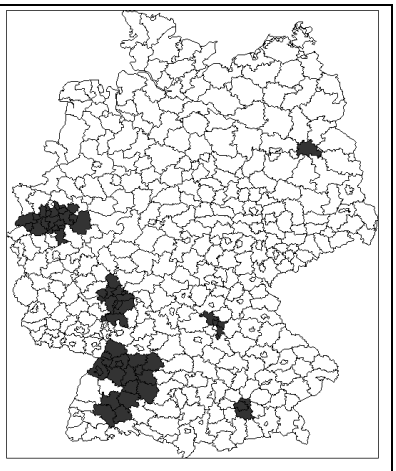
Figure 7: Spatial employment patterns in the Electrical Industry



a) Local Moran's I_i of employment density (r=45 km)



b) Local Getis-Ord G_i^* of employment density (r=45 km)

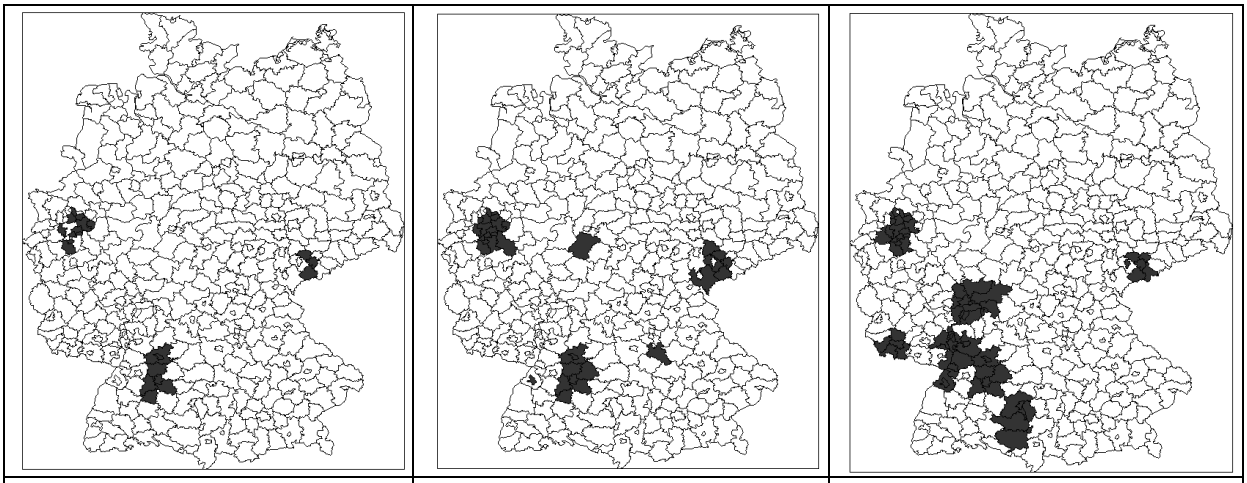


c) Kulldorff's spatial scan statistic of employment density (km=45 km)

Divergent high employment clusters are discovered with different approaches in the electrical industry (DL30-DL33). The most likely cluster of Munich and its surrounding (log LR=61407.3, p=0.0000) is found with all methods. It has a high relative risk of 15.5. Electrical companies in this compact cluster employ about 35,000 (4%) workers. The most significant secondary cluster accruing from the scan tests is as well located in Bavaria around Nuremberg (log LR=47215.5, p=0.0000) with a risk factor of 14.0. The existence of this cluster is confirmed by the I_i and G_i^* tests, though in the latter case as part of a larger Bavarian area of high employment density. An additional hot spot of electrical activity in Middle Bavaria arises only from local G_i^* tests.

Although the electrical clusters of Berlin, Rhine-Ruhr and Rhine-Main are important secondary clusters according to Kulldorf's spatial scan statistic, they are not uncovered by the other local tests. Centres of manufacture of electrical and optical instruments are discovered with all methods in Baden-Wuerttemberg. The area of high employment density delineated by the scan tests comprises about 83,000 (10%) workers employed in this sector. With the local Moran test, the cluster shrinks to two districts (Ludwigsburg, Böblingen) in the vicinity of Stuttgart. From the local Getis-Ord tests, four unconnected hot spots in Baden-Wuerttemberg arise which overlap in large part with the scan-based cluster.

Figure 8: Spatial employment patterns in the Automotive Industry



a) Local Moran's I_i of employment density (r=45 km)	b) Local Getis-Ord G_i^* of employment density (r=45 km)	c) Kulldorff's spatial scan statistic of employment density (km=45 km)

The three automotive clusters Rhine-Ruhr, Saxony and Baden-Wuerttemberg identified by the local Moran tests are as well found with two other tests. Additional hot spots detected by the G_i^* and scan tests turn out to be method specific. While local Getis-Ord tests point to additional clusters in Bavaria Hesse in the surroundings of Nuremberg and Kassel, Kulldorff's scan tests disclose some larger clusters in the southwest of Germany. High activities in manufacture of automotive vehicles, trailers and semi-trailers (DM34) Baden-Wuerttemberg are reflected by the existence of three automotive clusters. One further cluster is located in the Saarland and another in the Rhine-Main area.

Stuttgart and its surroundings shows up as the most likely cluster (log LR=151204.2, p=0.0000). Nearly 100,000 (15%) automotive workers are concentrated in this area. In the most significant secondary Rhine-Ruhr cluster employment in this sector amounts only to 27,000 (4%) workers (log LR=25636.3, p=0.000). Passing from the former to the second cluster is accompanied with a decrease of relative risk from 13.6 to 5.0. Somewhat less important is the Rhine-Main cluster where about 19,000 (3%) employees are occupied in car manufacture. This also holds for the two other high density areas in Baden-Wuertemberg. By contrast, automotive employment is in the single East German cluster around Chemnitz slightly larger.

6. Discussion

The use of local tests for cluster detection in German R&D intensive industries shows that different concepts can result in diverging conclusions on the size and location of spatial clusters. Hot spots discovered by the local Getis-Ord test may be reduced to a core of high attribute regions delineated on the basis of the I_i coefficients. Such patterns occur, when employment density within a coherent area of high-high (HH) and low-high (LH) regions is significant higher than outside. This delineation feature is observed in our study for the Rhine-Ruhr and Rhine-Main clusters in the chemical and mechanical engineering industry. It also emerges for the former cluster in the electrical industry. The extreme case where G_i^* tests may classify a low-low (LL) district between two medium or high density centres as a hot spot does not occur for the R&D intensive industries.

Apart from the chemical industry the largest cluster sizes are identified by Kulldorff's spatial scan statistics. This is not an artefact of a larger scanning window as all methods are implemented with optimal distances derived from global spatial autocorrelation analysis. An explanation may be a higher power of the scan tests compared to the local Getis-Ord and local Moran tests. In this case, the probability of extending a cluster is larger for the former than the latter test when the alternative hypothesis of clustering is true for regions in question. However, the obvious higher rejection rate of the CSR hypothesis may also be due to the testing design. Simulation studies could give insight in cluster detection capabilities of the different approaches. Waller et al. (2006) and Dai et al. (2010) show a sensitivity of Kulldorff's spatial scan tests with respect to the location of suspected clusters. Up to now, however, comparative studies on the statistical performance of cluster detection tests are missing.

Identified cluster patterns are not independent from the definition of neighbourhoods. With rising distance from a regional centre an existing cluster of medium or large size has a better chance of being detected by local tests. This finding can be ascribed to the increased power of the test with growing sample sizes (cf. Huang et al., 2009). However, parts of clusters may be undiscovered in case of large thresholds when they are not allowed to overlap. Chen et al. (2008) examined the effects of an increasing the maximal scanning window from 1 to 50% on number of identified clusters as well as their location and size. They established instability of the SatScan clusters. When the maximum window size is large, artificial heterogeneous clusters are identified possibly due to some core clusters located within their boundaries. On the other hand, with a too low maximum distances, clusters of medium size may be undiscovered.

With an arbitrary choice of the maximum-size parameter, existing cluster patterns may be masked. As Kosfeld et al. (2011) have shown, spatial clustering of industrial activity can emerge at varying spatial scales. Thus, it is important to establish the spatial scale at which clustering formation in an industry takes place. For R&D intensive industries this is done here by global spatial autocorrelation analysis using Moran's I and the Getis-Ord G statistic. For the electrical and automotive industry the strength of spillovers seem to decrease after reaching the maximal interaction intensity at 45 km. For the mechanical engineering industry spatial interaction tends to be strongest at a larger spatial scale of 90 km. While these ranges are uniquely inferred from both global measures, different indications arise for the chemical industry. Moran's I suggests an optimal distance of 50 km and Getis-Ord's G a range of 80 km. As the optimal choice by the former coefficient turns out to be the second best by the latter, we preferred the lower distance. Local tests may respond differently to a change of the spatial scale. Whereas the I_i-base and scan-based clusters do not change noticeably, both G_i^*-based clusters in the Rhine-Ruhr and Rhine-Main area would increase considerably with a threshold um 80 km instead of 50 km.

For distance-based spatial weights neighbourhoods are ordinary defined by circular windows around the centroids of the areal units. The spatial lags of the I_i and G_i^* statistics in the local Moran and Getis-Ord tests are formed for such surroundings. In order to ensure comparability, circular windows are likewise used with Kulldorff's spatial scan test. Although SatScan is extended to search circular and elliptical clusters (Kulldorff et al., 2006), the circular scan statistic is able to detect the latter ones (Pfeiffer et al., 2008, p. 51). This is especially expected in case of smaller window sizes. Particular in the chemical industry, elliptical-shaped clusters are identified by all local methods. More general, real clusters may exhibit complex irregular shapes. A simulation study could reveal the contribution of the flexibly shaped scan statistic developed by Tango and Takahashi (2005) to cluster detection. In case of substantive improvements in the validity and reliability of cluster detection, irregular shaped neighbourhoods should as well be considered for the local Moran and Getis-Ord test.

An open question with Kulldorff's spatial scan test is further the treatment of secondary clusters. Depending on the industry, the number of significant R&D clusters varies in this study between 50 and 70. In empirical cluster research, often additional to the primary cluster two or three most significant secondary clusters are interpreted. However, in cluster theory it is argued that existing clusters must have reached a critical mass in size and/or diversity of operation in order to promote regional growth and development (Wares and Hadley, 2008). We have addressed this issue by imposing a threshold for the size of the clusters. The contribution of Kulldorff's approach to economic cluster research will not least depend on a satisfactory solution of this issue.

References

Anselin, L. (1988), Spatial Econometrics: Methods and Models, Kluwer, Boston.

Anselin, L. (1995), Local Indicators of Spatial Association – LISA. Geographical Analysis 27: 93-115.

Anselin, L. (2003), GeoDa™ 0.9 User's Guide, http://geodacenter.asu.edu/software/documentation.

Audretsch, D.B. Feldman, M.P. (1996), R&D Spillovers and the Geography of Innovation and Production, American Economic Review 86, 630-640.

Bailey, T.C., Gatrell, A.C. (1995), Interactive Spatial Data Analysis, Prentice Hall, Harlow, England.

Bertinelli, L., Nicolini, R. (2005), R&D investments and the spatial dimension: evidence from firm level data. Review of Regional Studies 35, 206–230.

Bickenbach, F., Bode, E. (2008), Disproportionality Measures of Concentration, Specialization, and Localization, International Regional Science Review 31, 359-388.

Chen, J., Roth, R.E., Naito, A.T., Lengerich, E.J., MacEachren, A.M. (2008), Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality, International Journal of Health Geographics 7:57.

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

Coulston, J.W., Riitters, K.H. (2003), Geographic Analysis of Forest Health Indicators Using Spatial Scan Statistics, Environmental Management 31, 764-773.

Dai, J., Chen, F., Sahu, S., Naphade. M. (2010), Regional Behavior Change Detection via Local Spatial Scan, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM 2010, New York, 490-493.

Feldman, M.P., Audretsch, D.B. (1999), Innovation in cities: Science-based diversity, specialization and localized competition', European Economic Review 43, 409-429.

Feser, E. (2000), On the Ellison-Glaeser geographic concentration index, Discussion Paper, University of North Carolina.

Feser, E., Sweeney, S., Renski, H. (2005), A descriptive analysis of discrete U.S. industrial complexes, Journal of Regional Science 45, 395-419.

Galloway, H., Robison, H. (2008), Identification of knowledge and innovation clusters: A GIS application of concentration, co-existence, and correlation, Discussion Paper, EMSI - Economic Modeling Systems Inc.

Getis, A., Ord, J.K. (1992), The Analysis of Spatial Association by Use of Distance Statistics, Geographical Analysis 24, 189-206.

Glaeser, E., Kallal, H., Sheinkman, J., Shleifer, A. (1992), Growth in Cities, Journal of Political Economy 100, 1126-1152.

Grenzmann, C., Kladroba, A., Kreuels, B. (2009), FuE Datenreport 2009, Wissenschaftsstatistik, Stifterverband für die Deutsche Wissenschaft, Essen.

Henderson, R., Jaffé, A.B., Trajtenberg, M. (1998), Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988, Review of Economics and Statistics 80, 119-127.

Huang, L., Stinchcomb, D.G., Picle, L.W., Dill, J., Berrigan, D. (2009), Identifying Clusters of Active Transportation Using Spatial Scan Statistics, American Journal of Preventive Medicine 37, 157-166.

Kies, U., Mrosek, T., Schulte, A. (2009), Spatial Analysis of Regional Industrial Clusters in the German Forest Sector, International Forestry Review 11, 38-51-

Kosfeld, R., Eckey, H.-R., Lauridsen, J. (2011), Spatial Point Pattern Analysis and Industry Concentration, Annals of Regional Science 47, Special Issue: Advanced Methods and Applications in Regional Science, 311-328.

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

Kulldorff, M. (1997), A Spatial Scan Statistic, Communications in Statistics – Theory and Methods 26, 1481-1496.

Kulldorff, M. (2010), SatScan™ User Guide for version 9.0, http://www.satscan.org/.

Kulldorff, M., Nagarwalla, N. (1995), Spatial disease clusters: detection an inference, Statistics in Medicine 14, 799-810.

Kulldorff, M., Hunang, L., Pickle, L., Duczmal, L. (2006), An elliptical spatial scan statistic, Statistics in Medicine 25, 3929-3943.

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

Le Gallo, J., Ertur, C. (2005), Exploratory Spatial Data analysis of the Distribution of Regional per capita GDP in Europe, 1980-1995, Papers in Regional Science 82, 175-201.

Litzenberger, T. (2007): Cluster und die New Economic Geography, Frankfurt/M.

Molle, W. (1997), The regional structure of the European Union: an analysis of long-term developments, in: Peschel, K. (ed.), Regional growth and regional policy within the framework of European integration, Physica, Heidelberg.

Neffke F.M.H., Svensson Henning M., Boschma R.A., Lundquist K.-J., Olander L.-O., 2008, Who Needs Agglomeration? Varying Agglomeration Externalities and the Industry Life Cycle, Papers in Evolutionary Economic Geography (PEEG) 0808, Utrecht University, Netherland.

Openshaw, S., Charlton, M., Wymer, C., Craft, A. W. (1987). A mark I geographical analysis machine for the automated analysis of point data sets. International Journal of Geographical Information Systems 1, 335-358.

Ord, J.K., Getis, A. (1995), Local Spatial Autocorrelation Statistics: Distributional Issues and an Application, Geographical Analysis 27, 286-306.

Oxford Research (2008), Regional clusters in Europe, Europe INNOVA Cluster Mapping Project, Kristiansand, Norway.

Pfeiffer, D.U., Robinson, T.P., Stevenson, M., Stevens, K.B., Rogers, D.J., Clements. A.C.A. (2008), Spatial Analysis in Epidemiology, Oxford University Press, Oxford, U.K.

Porter, M.E. (1998), Clusters and the new economics of competition, Harvard Business Review, November-December, 77-90.

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

Porter, M.E. (2008), Clusters and Economic Policy: Aligning Public Policy with the New Economics of Competition, Discussion Paper, Harvard Business School, Cambridge, Mass.

Südekum, J. (2006), Concentration and Specialization Trends in Germany since Re-unification, Regional Studies 40, 861-873.

Tango, T., Takahashi, K.A. (2005), A flexibly shaped spatial scan statistic for detecting clusters, International Journal of Health Geographics 4:11.

Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L.,Clark, L.C. (1990), Monitoring for clusters of disease: Applications to leukemia incidence in upstate NewYork., American Journal of Epidemiology 132, 136-143.

Waller, L.A., Hill, E.G., Rudd, R.A. (2006), The Geography of Power: Statistical Performance of Tests of Clusters and Clustering in Heterogeneous Populations, Statistics in Medicine 25, 853-865.

Wares, A.C., Hadley, S.J. (2008), The Cluster Approach to Economic Development, Technical Brief No. 7, Office of Economic Growth of EGAT/USAID.