Journal of the Korean Geographical Society, Vol. 40, No. 2, 2005(187~201)

The Statistically and Economically Significant Clustering Method for Economic Clusters in an Urban Region

Jungyeop Shin*

통계적 및 경제적 유의성을 가진 경제 클러스터 탐색방법에 대한 연구

신정엽*

Abstract : With the trend of urban polynucleation, the issue of detecting economic clusters or urban employment centers has been considered as crucial. However, the prior researches had some limitations in detecting economic clusters in the empirical analysis: i.e. inherent inefficiency of density-based clustering methods, difficulty in detecting linear types of spatial clusters and lacks of consideration of economic significance. The purpose of this paper is to propose the clustering method with the procedure of testing statistical and economic significance named as VCEC (Variable Clumping method for Economic Clusters) and to apply it to a case analysis of Erie County, New York, in order to test its validity. By applying a search radius and a total employment as an economic threshold, "the both statistically and economically significant clusters" were detected in the Erie County, and proved to be efficient.
Key Words : spatial cluster, Variable Clumping Method, economic cluster, urban employment center, VCEC (Variable Clumping method for Economic Clusters)

요약 : 경제 클러스터와 도시 고용중심지에 대한 연구는 최근 지리학 분야에서 매우 중요하게 다루어지고 있다. 그러나 경제 클러스터 탐색을 위한 기존 연구들은 탐색방법의 내재적 한계, 선형 클러스터 탐색의 비효율성, 경제적 유의성 검증의 부족등의 문제를 내포하고있다. 본 연구의 목적은 경제 클러스터 탐색방법으로서 통계적, 경제적 유의성을 검증하는 VCEC(Variable Clumping method for Economic Clusters)를 제안하는 것이고, 이를 바탕으로 미국 뉴욕주 이리 카운티(Erie County)의 경제 중심지 탐색을 위한 실증적 경험 사례분석을 하는것이다. 다양한 탐색 반경과 총 고용인구 한계치의 적용을 통해 통계적, 경제적인 유의성을 가진 경제중심지 탐색이 가능하였다.
주요어 : 공간 클러스터, 경제 클러스터, 도시 고용중심지

* Ph. D., Department of Geography, State University of New York, Buffalo, jshin2@buffalo.edu

1. Introduction

There have been abundant prior clustering methods developed in various fields such as economic geography, urban geography, statistics, spatial statistics, epidemiology and ecology. Regarding it, most clustering algorithms were developed with recent advances of statistical, geometric techniques such as hierarchical, partitioning, and searching techniques for neighboring points. From a technical viewpoint, they mainly used the information of location of points, and they calculated further information for clustering such as proximity distance with neighboring points, connectivity with adjacent polygons, creation of hierarchical trees with points with criteria, and calculation of density estimates with neighboring points.

Focusing on the field of urban and economic geography, the issue of economic clusters or urban employment centers in urban polynucleation has been widely considered as important. Among various researches, the most important point is how to detect economic clusters with significance. In regard to it, three major approaches were widely used: (1) the first depended upon the information of the number of employment in order to detect clustered employment centers in terms of an economic context, while (2) the second applied the statistical and geometric algorithm to the urban and economic data in terms of the statistics. The third is the density-based method.

The first approach has several limitations in exploring significant economic clusters. For example, by using rough-scaled data (in many cases, with an aggregated scale), it is not efficient in detecting more detailed results of the clusters, especially at regional or local scale. It also did not consider the statistical significance for the analysis procedure. The second approach has also its limitations. For example, the original clustering methods are inefficient in detecting economic clusters from the linear arrangement of economic activities in the urban region. Second, it did not consider the economic criteria for significant economic clusters. Regarding the third approach, the cluster results are subject to the arbitrary selection of parameters and thresholds during the analysis procedure.

Based on this background and potential problems to be solved, this paper will focus on proper clustering techniques for detecting economic clusters from the linear arrangement of economic clusters in an urban region. The purpose of this paper is, with the critical review of prior researches, to propose a proper clustering method for statistically and economically significant clusters, and to verify it with the case analysis of economic clusters in Erie County, New York.

2. Review of Clustering Methods for Economic activities

1) Researches on Economic Clusters

There are three major categories of prior research on detecting economic clusters: detection of employment centers with economic factors, clustering method with statistical algorithms, and density-based clustering for economic clusters. Each category of the clustering methods will be reviewed and based on it, the proper clustering method for economic clusters will be proposed.

First approach is to detect urban employment centers using economic data and criteria. For detecting major employment centers or economic clusters in an urban region, empirical analysis mainly used economic factors, especially information of employment. With this information of employment, many researches implemented aggregation procedures with the TAZ (Traffic Analysis Zone) data and applied an economic criterion to filter insignificant clusters out. For the criteria, total number of employment and employment density were mostly preferred (see; McDonald and McMillen, 1998; McMillen, 2001 for the related applications).

The most famous criterion was suggested by Giuliano and Small (1991) for detecting employment centers of Metro Los Angeles: 10,000 total employment and at least 15 employees per acre. Bogart and Ferry (1999) adopted 5,000 employment per square mile with 10,000 total employees by modifying the criteria by Giuliano and Small. Cervero and Wu (1997, 1998) also used 7 workers per acre with threshold of 10,000 total employees, and Shearmur & Coffey (2002), Coffey and Shearmur (2002) used 5,000 employments per square mile. Song (1994) used 15 employees per acre with 35,000 total employments.

The second approach was related to the clustering algorithm in spatial statistics. Many researches used the information of spatial autocorrelation to capture neighboring regions, in order to find the economic 'hot spot'. In this context, most common statistics for detecting local spatial clusters are G_i Statistic (Ord and Getis, 1995), Anselin's LISA (Anselin, 1995) including local Moran's I, local Geary's C. More focusing on economic clusters, there were several researches using Local Moran's I (Pacheco & Tyrrell, 2002; Paci & Usai, 1999), Geary's C, Local G statistics (Matisziw & Hipple, 2001; Ceccato & Persson, 2002), K-function (Sweeney & Feser, 1998; Cuthbert & Anderson, 2002; Barff, 1987). Paci & Usai (1999) applied Local Moran's I to case of Italian Local Labor System to find employment clusters.

Regarding G_i stat, Ceccato & Persson (2002) explored employment clusters in rural area of Sweden, and Matisziw & Hipple (2001) used G_i statistic to find clusters of Hog production in Missouri. On the other hand, with K-function technique, Barff(1987) applied it for manufacturing clusters in Cincinnati, Ohio, with three data stratification (less than 15, 16~99, greater than 100), and Cuthbert & Anderson(2002) used parcel-level point data from 1970-1996. Sweeney & Feser (1998) also used it for manufacturing clusters in North Carolina, especially focusing on relationship between plant size and clustering.

The third one used the density functions as a pre-step to filter significant economic clusters. The clustering method using distance searched for neighboring points from a certain point with a range of radius or distance, and calculated how many points are within a distance from a certain point (related to the point density), and applied a threshold value to decide significant clusters. For example, Wallsten(2001) calculated the distance between firms and setup 'density variables' for each observation . the number of other firms within the distance range of from one. One of the major algorithms is to calculate the density, and apply threshold criteria for detecting significant clusters. Density-based clustering method involves the process of density surface creation as a pre-step of detection of clustering. Regarding density surface creation, several techniques are used: trend surface (Wang, 2000,

2001), Smoothing (Craig and Ng, 2001) and IDW. For next step, it needs a cut-off point to decide significant cluster boundaries from the resulting density surfaces. In most cases, some boundaries from local peaks of the density surfaces were defined as clusters. For example, Wang (2001) used '1,096 jobs/km^2' as a threshold for deciding significant clusters from the density surfaces, and Craig and Ng (2001) used a combination of local knowledge and employment statistics as a cut-off point.

2) Critiques for the Prior Clustering Methods

Even though each category of clustering methods has its advantages in detecting clusters, they also had some drawbacks and limitations in their analytical techniques. First, researches on detecting employment centers and clustering method using spatial autocorrelation mainly used the polygon-based data, which may not be proper for regional or smaller scaled study regions. In general, polygon-based data as an aggregated data structure is not proper in detecting more detailed and accurate cluster results, especially at a finer geographic scale. Even though centroids of the polygons are used, the clustering results will be inaccurate.

Second, density-based clustering methods created density surfaces (e.g. employment density or residential density) to detect significant nuclei using various density creating functions and their threshold parameters. However, depending on density creating functions, resulting surfaces to be estimated will be different from the real values, and arbitrary parameter selection and setting will lead to the different results with error propagation, which we may not guarantee validity of the cluster results. For example, resulting density surfaces will be various depending on density methods and parameters, and threshold criteria to define significant nuclei from the surface will be also subject.

The third one is related to the circular search window for neighboring points. Many clustering techniques such as K-function, kernel estimate function, GAM, use circular search window in detecting the neighboring points from a center or point.

The use of the circular search window may mislead clustering results. In the real world, many phenomena such as crime, retail location, especially caused from the human behavior, has linear arrangement or combinations of the linear arrangements: i.e. linear, cross-shaped, star-shaped, cross-shaped clusters. The circular search window can not measure neighboring features for these different types of cluster forms in the real world. Rather, it is likely to mislead clustering results in a different way.

Last, not only the clustering technique, but also the variables to apply for the method are very important to detect the significant clusters or nuclei in a specific application. However, many clustering techniques used only geometric variables such as distance, spatial autocorrelation, without significant variable for the relevant application. For example, for economic clusters, not only geometric variables, but also economic variables are crucial for significant results.

Based on the research review, I propose a proper clustering method for urban economic nuclei: the clumping-based clustering method. The clumping method has strong power to detect various kinds of cluster forms such as linear, irregular shape. In addition to that, the method will use both geometric variable such as distance

between points, and economic variable such as number of employees, which can help to find more significant nuclei.

3. VCEC (Variable Clumping method for Economic Clusters)

1) Methodology

For detecting significant economic clusters, this paper proposes the clustering method with the procedure of testing both statistical and economic significance. The basic idea for detecting spatial clusters was from the Variable Clumping Method (VCM) (Sadahiro, 2003; Okabe and Funamoto, 2000; Okabe, Asami and Miki, 1985 for details). This VCM is useful to detect local clusters with various forms in point distributions, and requires only simple spatial operations such as buffer operation, distance calculation, and point counting, which is widely used in GIS (Sadahiro, 2003). In addition to it, they used the information of the connectivity rather than the density information from the circle search window, which is efficient to capture the accurate cluster results.

The main procedure for clumping method is like below:

"Suppose n points in a region S of area A. The ith point is denoted by P_i. We generate circles of diameter dn centered at the points and regard point pairs whose circles overlap with each other as "neighboring". The distance dn is called the neighborhood distance.

From sets of neighboring points we extract larger ones consisting of more than λ points. The threshold number of points λ is called the minimum cluster size. The points in clusters detected are called cluster points" (Sadahiro, 2003)

This general method is quite simple and easy to understand. From the observed cluster results, in order to eliminate the insignificant clusters from the results, it used expected value of the clusters in the assumption that it is expected that the number of clusters of size i for radius r is significantly larger than the number of clusters that would appear in the distribution of random points. Thus, if the observed number of clusters $n_i(r)$ is greater than the number of clusters that would appear in the distribution of random points, we may say that these clusters are significant clumps. To get the critical number of clumps, Okabe and Funamoto (2000) applied Monte Carlo simulation for 10,000 trials of random point distribution. However, there are several limitations in this method. First, until now, there were researches on the theoretical proposal and empirical analysis with simulated data, but there was no empirical analysis with the real empirical data. The clumping analysis with the simulated data is different from that with the empirical data in terms of analytical procedure and results. Second, the classical VCM succeed in testing the statistical significance of the clustering result, but failed to test the economic significance. In many cases, the resulting economic clusters may have no meaning in terms of the economic criteria. We need to explore "the both statistically and economically significant clusters".

In this paper, I developed this basic VCM by modifying the analysis procedure: the procedure to testify the economic significance is added to the basic procedure. The analytical procedure for

the method is shown below;

Table 1. Analytical procedure of the clustering method

Main Procedure

[Step 1] Define the circles from the points as centers with a given radius

[Step 2] Make chains of the points with overlaying circles for clusters

[Step 3] Implement Monte Carlo Simulation to get random point distribution for 1,000 trials, and get the frequency values of clusters for expected number of clusters

[Step 4] Repeat Step 1 to 3 with increase of the radius value until the radius come up to the threshold value of radius

[Step 5] Create the matrix with the observed number of clusters with variable radius and number of chains

[Step 6] Create the matrix with the expected number of clusters with variable radius and number of chains

[Step 7] Compare observed and expected numbers of clusters at each radius

[Step 8] Get the significant number of clusters in terms of statistics

[Step 9] Calculate total number of employees for each significant cluster

[Step 10] Create the matrix with the significant number of clusters with variable radius and total number of employees

[Step 11] Make the graphs for significant break for the line of number of significant clusters with the increase of radius

[Step 12] Decide the significant number of clusters in terms of statistics and economics

The clustering method as VCEC (Variable Clumping method for Economic Clusters) has several advantages for economic clustering. The target data of this method is point-based dataset, not polygon-based one. With the individual point dataset, we can get more accurate and sophisticated clustering result. Second, VCEC is based on the concepts of membership of clumps and neighboring existence of points, rather than absolute searching distance for neighboring point density. In most cases of spatial phenomena, the membership of cluster and existence of neighboring points much more relevant in that a member of a cluster will be neutral or negative for density of neighboring points, which may lead irrelevant or distort result of cluster. Third, VCEC avoids the circular search window for neighboring points, which are not proper to linear types of clusters that most phenomena in the real world are likely to take. VCEC considers both adjacency and distance for neighboring points, which are important to capture significant neighboring points..

After the comparison of two matrixes (one for observed and the other for expected number of clusters with different radius), we will get the matrix with statistically significant number of

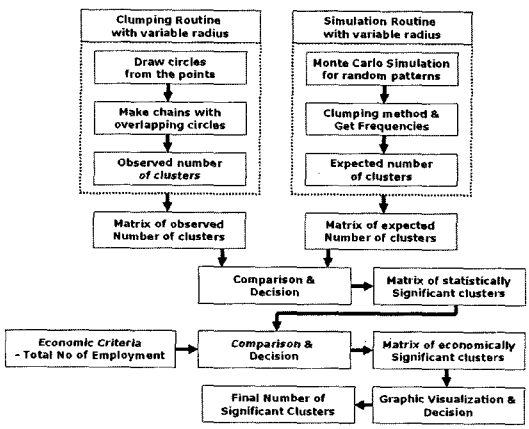

Figure 1. Procedure of VCEC analysis

clusters. However, we need to know if these clusters are still significant in economic context. Even though a cluster will have small number of companies as members, it may be economically significant if it has enough volume of employment, and vice versa. Regarding this, we need an economic criterion, total number of employment, which has been used for the criteria for employment centers.

2) Case Study

The paper applied the proposed VCEC to a case analysis of manufacturing clusters in an urban region. The study area is Erie County, NY, which is one of the typical middle-sized metropolitan regions of United States. The Erie County, as a western part of New York State, consists of 26 municipalities, and has the old CBD in City of Buffalo, and several multi-nuclei in the suburban region. Erie County is one of the typical middle-sized metropolitan regions in the United States, and it has an urban polynucleated form with a live CBD. Located in western New York, Erie County has the CBD in the middle of the region near Erie Lake, and several significant suburban centers such as Buffalo Airport, and SUNY Buffalo North Campus. The transportation networks connected the CBD and suburban areas with two circulation rings, and star-shaped major road networks.

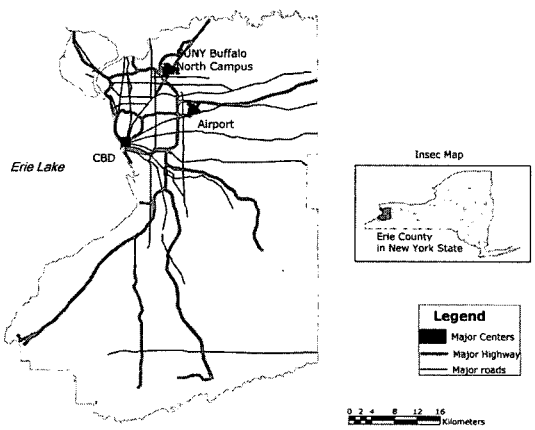

Figure 2. Erie County, New York State as Study Region

The data source is the company directory data in manufacturing industry covering whole Erie County, NY in 2000. From the manufacturing dataset, I used two types of variables: x,y coordinate location with the information of municipality, and number of employees for deciding economically significant clusters. The point data was digitized using geocoding function and converted into GIS format as ArcView shapefile format with UTM coordinate system in meter unit. For VCEC algorithm and mapping, ArcGIS 8.3 and Visual Basic were used for programming and visualization.

From the exploration of the average distance of nearest neighbor points (manufacturing companies), we setup a series of search radius from 400 to 1,000 meter by 100 meter interval from the review of the average distance of the nearest neighbor points, number of manufacturing companies and the local knowledge of Erie County. The average distance of nearest neighbor point in the dataset was 285 meter, and number of points (manufacturing companies) was 1,260. In my opinion, the 100 meter interval is long enough to explore the changes of economic cluster results. The comparison among the cluster results with different radii enabled us to understand the effects of search radius for clustering results, and based on it, we can decide the significant spot or interval of the radius (or radii) as in the K-functions.

Table 2 provides the observed results of the manufacturing clusters at different radius. In

Table 2.The observed number of manufacturing cluster by searching radius

Size* \ Radius	400	500	600	700	800	900	1000
2	71	56	46	39	30	24	18
3	35	23	19	16	11	12	10
4	11	9	12	12	10	8	4
5	16	14	12	10	10	6	5
6	11	6	5	5	4	4	3
7	6	3	2	0	2	4	4
8	4	5	7	5	6	4	2
9	3	2	1	3	3	2	1
10	0	1	3	3	2	3	1
11	3	0	0	0	0	0	1
12	2	2	2	3	1	1	1
13	1	0	0	0	2	1	1
14	2	0	0	1	0	0	0
15	2	0	0	0	1	1	2
16	0	0	0	0	0	0	1
17	1	1	0	0	0	0	0
18	1	2	2	2	1	0	2
19	1	0	0	0	0	1	1
20	1	2	2	1	1	1	0
21~30	1	2	1	2	1	1	1
31~40	2	2	2	1	1	2	1
41~50	1	2	3	3	1	0	1
51~70	1	0	1	0	1	0	0
71~100	1	0	0	0	0	0	0
101~150	0	3	1	0	0	0	0
151~200	0	0	0	1	0	0	0
201~300	0	0	1	1	0	0	0
301~500	0	0	0	0	0	0	0
501~700	0	0	0	0	1	0	0
701~	0	0	0	0	0	1	1
Total No of clusters	177	135	122	108	89	76	61

* size: size of the clusters, which is measured as number of points in a cluster

general, the proportion of the smaller-sized clusters is huge, while that of larger-sized clusters is relatively small. One interesting result is that the total number of the clusters tends to decrease as a search radius increases: from 400 meter to 1,000 meter radius. The huge decreasing proportion of total numbers was caused by decreasing numbers in smaller-sized clusters. The longer a search radius is, the smaller the total number of the clusters.

For the next step, in order to explore statistically significant clusters, the expected number of the clusters resulted from 1,000 Monte Carlo Simulation for each radius. After the comparison of the observed and expected clusters, the statistically significant number of clusters at

Table 3. The statistically significant number of manufacturing cluster by searching radius

Size* \ Radius	400	500	600	700	800	900	1000
2	0	0	0	0	0	0	0
3	35	0	0	0	0	0	0
4	11	0	0	0	0	0	0
5	16	14	12	0	0	0	0
6	11	6	5	0	0	40	0
7	6	3	2	0	0	0	0
8	4	5	7	5	6	0	0
9	3	2	1	3	3	0	0
10	0	1	3	3	2	0	0
11	3	0	0	0	0	0	0
12	2	2	2	3	1	0	0
13	1	0	0	0	2	0	0
14	2	0	0	1	0	0	0
15	2	0	0	0	1	0	0
16	0	0	0	0	0	0	0
17	1	1	0	0	0	0	0
18	1	2	2	2	1	0	2
19	1	0	0	0	0	1	1
20	1	2	2	1	1	1	0
21~30	1	2	1	2	1	1	1
31~40	2	2	2	1	1	2	1
41~50	1	2	3	3	1	0	1
51~70	1	0	1	0	1	0	0
71~100	1	0	0	0	0	0	0
101~150	0	3	1	0	0	0	0
151~200	0	0	0	1	0	0	0
201~300	0	0	1	1	0	0	0
301~500	0	0	0	0	0	0	0
501~700	0	0	0	0	1	0	0
701~	0	0	0	0	0	1	1
Total No of clusters	106	47	45	26	22	6	8

* size: size of the clusters, which is measured as number of points in a cluster

different radius was listed (see Table 3).

From the observed number of the clusters, for each radius, huge proportions of the clusters were deleted as insignificant, and the compared clustering results show significant difference. The deleted proportions were mainly from the smaller-sized clusters for each radius, and there was no exclusion of the larger-sized clusters. It represents middle or larger-sized clusters tend to be, in most cases, proven statistically significant, while some smaller-sized clusters tend to be statistically insignificant. For the smaller-size parts, even though some observed clusters were detected, and in the point of statistical view, they may be interpreted as random errors from the major trends.

Figure 3 shows a bar graph to compare the observed and statistically significant clusters. With

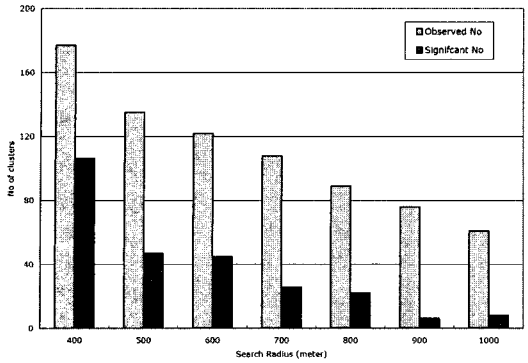

Figure 3. The bar graph on the observed and statistically significant manufacturing clusters by search radius

the decreasing number of the clusters with increasing search radius, the observed clusters show the decreasing trend with relatively constant rate, while the statistically significant clusters shows the decrease in number with relatively breaking points around 500 and 600 meter. Between 400 and 500 meter search radius, the difference in the numbers was great, while after that, they show slow decreasing differences.

One of the main reasons for decreasing numbers with the increasing search radius is that as the search radius gets longer, the existing smaller-sized clusters tend to be merged into the other clusters, no matter what the sizes of the clusters are.

Figure 4 shows differences in number of the clusters at different radius. As you can see the cluster maps, individual manufacturing companies and smaller-sized clusters, especially located in the suburbs, were excluded from the cluster results, which are subject to the size of the search radius. The excluded companies were treated as statistically insignificant.

The common spatial pattern is strong centralization near the CBD and scattered economic clusters in the suburban region (see map (2) ~ (8)). The spatial arrangement of the larger-sized clusters near the CBD takes the linear pattern from the CBD toward two major directions: one to the north along the Erie Lake, and the other to the east along the major local road. The other economic clusters locate along the major transportation networks. One interesting thing for the changes of the clusters with different radius is that as the radius gets longer, the suburban clusters were eliminated or merged into the big ones. For this reason, the number of the suburban clusters decreased with the increasing radius.

In extreme cases, when the search radius is longer enough, the number of the cluster will be eventually one huge cluster in that region. However, in this case we can't differentiate detailed spatial pattern of the economic clusters. In addition to it, the number of the clusters does not always increase with the increasing radius as is shown in Figure 4 (see map (7) and (8)). The total cluster number increased from 6 at 900 meter radius to 8 at 1,000 meter radius. This difference was from the filtering procedure of comparing the observed and expected number of the cluster for each size.

The clustering results with statistical significance are useful in understanding spatial distribution of economic clusters. However, only with these results, we still can't guarantee the economic significance of the clusters. In many cases, all the statistically significant clusters may not be economically significant. In reality, there can be some economic clusters with economic insignificance. By filtering the economic clusters with economic insignificance, we can get "economic clusters with both statistical and economic significance".

In order to test economic significance, one

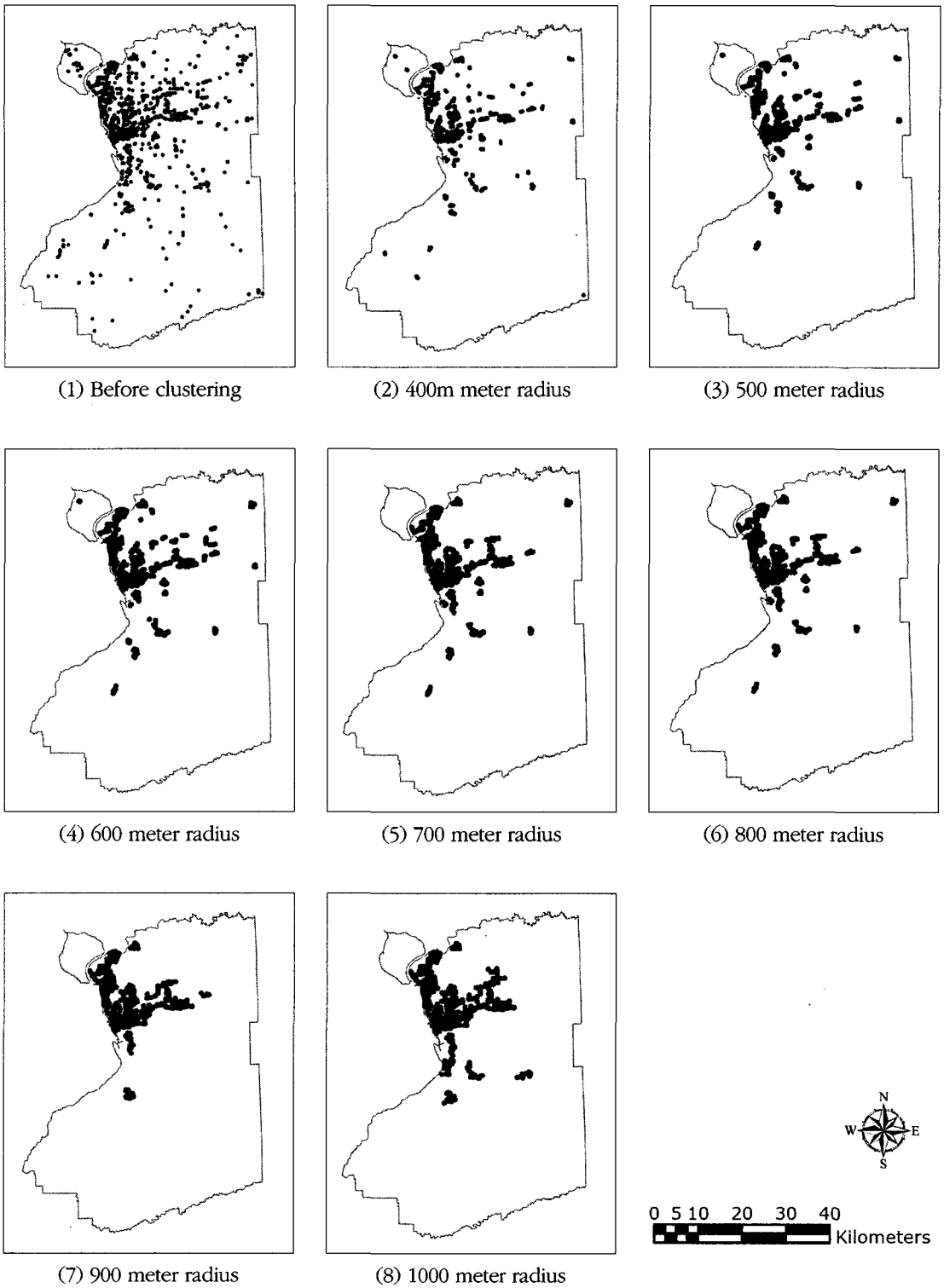

Figure 4. Spatial pattern of the statistically significant cluster by research radius

economic variable was chosen as criteria: total number of the employment in a cluster. From the prior research, the total employment variable was used for detecting economic centers in an urban region, even though the data unit was rough enough to measure accurate cluster boundaries. With the information of point-in-polygon process, we know which points are members of a cluster. Then, total number of the employment was calculated after summating the numbers of the points in a cluster. This analysis choose 600 meter radius, after the review of comparison of the clustering results: with the information of the bar graph (Figure 4) and map (Figure 5). In Figure 4, there was a breaking point in changes of the number by radius, and it is around 500 and 600 meter radius. Furthermore, by comparing the spatial patterns of the economic clusters, we know 600 meter radius is good enough to differentiate location of the clusters in the region.

Table 3 lists the number of manufacturing clusters with statistical and economic significance with different economic threshold. The cluster number in the table shows the number over the economic threshold. For example, with 400 meter radius, and the economic threshold with 150, the cluster number is 63, which means that 63 clusters are over the 150 total employments as threshold. The trend shows the decreasing numbers with increasing economic threshold.

The interesting point is the different decreasing rate across different radius. As the radius gets shorter, the number of clusters with economic significance decreased dramatically, while the numbers at 900 and 1,000 meter radius decreased slowly. For example, at the radius of 400 meter, the number of the statistically significant clusters is 106, and it is 63 with 150 employment threshold, 342 with 300, 24 with 600, and 13 with 1,200 threshold. On the other hand, at the 900 radius, the number was 6, and it was unchanged with 150, 300 and 600 economic threshold, and changed to 4 with 1,200 threshold. Based on the table, it may be argued that smaller-sized clusters are more subject to the economic threshold, which reflects the possibility that at shorter radius, the statistically significant clusters may have the clusters with economic insignificance, while with the longer radius distance, the clusters are not relatively less subject to the changes of the economic threshold.

On the other hand, the statistics shows the

Table 3. Number of manufacturing clusters with statistical and economic significance by search radius

Threshold / Radius	Stat Sig Clusters	The clusters over economic threshold*						
		150	300	600	1200	1800	2400	3000
400 meter	106	63	42	24	13	8	7	5
500 meter	47	39	28	20	11	8	7	7
600 meter	45	34	24	20	11	9	6	6
700 meter	26	24	20	19	9	8	6	5
800 meter	22	19	15	14	5	5	5	4
900 meter	6	6	6	6	4	3	2	2
1000 meter	8	8	7	7	5	4	4	4

* economic threshold: total employment in a cluster

larger-sized clusters tend to be less subject to filtering with economic threshold. One possible reason is that there may be strong correlation between the statistical and economic size of the clusters. However, there is also a possibility for the reverse: there will be small-sized clusters with larger employment, or larger-sized clusters with relatively smaller employment. Focusing on this analysis, the manufacturing clusters look positive correlation between statistical and economic sizes.

From the matrix of the number of the manufacturing clusters with different radius and economic threshold (Table 3), we can choose the specific cell (or cells) based on the purpose of the analysis or the prior information of the study region.

The absolute number of the clusters are subject to the size of the clusters, size of population data (number of points), and decisive radius. It is not easy to get the solid rule to decide the specific number of clusters. In the previous step, I choose the 600 radius as a significant level. For the purpose of exploring the middle or larger sized manufacturing clusters, the 600 economic threshold can be chosen, and the number of the both statistically and economically significant clusters is 20. For another example, for the purpose of exploring larger-sized clusters, the 3,000 threshold will be appropriate, and the number is 6.

4. Conclusion

This paper focused on proper clustering techniques for detecting economic clusters from the linear arrangement of economic clusters in an urban region. For this purpose, the paper proposed VCEC (Variable Clumping method for Economic Clusters) as a proper clustering method for statistically and economically significant clusters, and verified it with the case analysis of economic clusters in Erie County, New York.

With the different search radius, we implemented a series of the observed number of manufacturing clusters, and after the comparison procedure with the expected number of the clusters, the statistically significant clusters. The spatial patterns of the manufacturing clusters showed strong centralization with the CBD and scattered small-sized clusters in the suburban region. During the procedure of comparing the observed and expected number of clusters, with the increasing radius, the numbers decreased dramatically, and especially smaller-sized clusters in the suburbs were excluded. For the next step in order to find the economic significance, the total employment as an economic threshold were selected and applied to the clusters. With the increasing economic threshold, the cluster numbers decreased for all radii. In general trend, the smaller-sized clusters tend to be more subject to the size of economic threshold.

The paper proposed the VCEC in order to detect the economic clusters in the urban region. The scope of the paper focused on the proposal of VCEC and testifying it with the case analysis. For this reason, the paper could not handle the interpretation of the clustering results and implications, and further analysis such as exploring the characteristics of the inter-and intra- clusters. Furthermore, the research on how to decide the significant radius or economic threshold will be important, even though that issue is widely discussed, but unsolved in the field of spatial statistics.

Reference

Anderson, N., Bogart, W., 2001, The structure of sprawl: identifying and characterizing employ-ment centers in polycentric metropolitan area, *The American Journal of Economics and Sociology*, 60(1), 147-169.

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

Bailey, T. and Gatrell, A., 1995, *Interactive Spatial Data Analysis*, Harlow: Longman.

Barff, R., 1987, Industrial clustering and the organization of production: a point pattern analysis of manufacturing in Cincinnati, Ohio, *Annals of Association of American Geographer*, 77(1), 89-103.

Besag, J. and Newell, J., 1991, The detection of clusters in rare diseases, *Journal of the Royal Statistical Society Series*, A, 154, 143-155.

Bogart, W. and Ferry, W., 1999, Employment centers in Greater Cleveland: evidence of evolution in a formerly monocentric city, *Urban Studies*, 36(12), 2099-2110.

Ceccato, V. and Persson, L., 2002, Dynamics of rural areas: an assessment of clusters of employment in Sweden, *Journal of Rural Studies*, 18, 49-63.

Cervero, R. and Wu, K., 1998, Sub-centering and commuting: evidence from the San Francisco Bay Area, 1980-90, *Urban Studies*, 35(7), 1059-1076.

_____, 1997, Polycentricism, commuting, and residential location in the San Francisco Bay Area, *Environment and Planning* A, 29, 865-886.

Coffey, W. and Shearmur, R., 2002, Agglomeration and dispersion of High-order service employment in the Montreal metropolitan region, 1981-96, *Urban Studies*, 39(3), 359-378.

Craig, S. and Ng, P. 2001. Using Quantile smoothing splines to identify employment subcenters in a multicentric urban area, *Journal of Urban Economics*, 49, 100-120.

Cuthbert, A. and Anderson, W., 2002, Using spatial statistics to examine the pattern of urban land development in Halifax-Darmouth, *Professional Geographer*, 54(4), 521-532.

Estivill-Castro, V. and Lee, I., 2002, Argument free cluster for large spatial point-data sets via boundary extraction from Delaunay Diagram, *Computers, Environment and Urban Systems*, 26, 315-334.

Giuliano, G. and Small, K., 1991, Subcenters in the Los Angeles region, *Regional Science and Urban Economics*, 21, 163-182.

Leung, Y., Mei, C. and Zhang, W., 2003, Statistical test for local patterns of spatial association, *Environment and Planning A*, 35, 725-744.

Matisziw, T. and Hipple, J., 2001, Spatial Clustering and state/county legislation: the case of hog production in Missouri, *Regional Studies*, 35(8), 719-730.

McMillen, D., 2003, Identifying sub-centers using contiguity Matrices, *Urban Studies*, 40(1), 57-69.

_____, 2001, Nonparametric employment subcenter identification, *Journal of Urban Economics*, 50, 448-473.

McMillen, D. and McDonald, J., 1998, Suburban subcenters and employment density in Metropolitan Chicago, *Journal of Urban Economics*, 43, 157-180.

Okabe, A., Asami, Y. and Miki, F., 1985, Statistical analysis of the spatial association of convenience-goods stores by use of a random clumping model, *Journal of Regional Sciences*, 25, 11-28.

Okabe, A. and Funamoto, S., 2000, An exploratory method for detecting multilevel clumps in the distribution of points- a computational tool, *Journal of Geographical Systems*, 2, 111-120.

Openshaw, S. Charlton, M. Wymer, C. and Craft, A., 1987, A Mark I Geographical Analysis Machine for the Automated Analysis of Point data sets, *International Journal of Geographical Information Systems*, 1, 359-377.

Openshaw, S., Craft, A. and Charlton, M., 1988,

Searching for Leukemia Clusters using a Geographical Analysis Machine, *Papers of the Regional Science Association*, 64, 95-106.

Ord, J. and Getis, A., 1995, Local spatial autocorrelation statistics distribution issues and an application, *Geographical Analysis*, 27, 286-306.

Paci, R. and Usai, S. 1999. Externalities, knowledge spillovers and the spatial distribution of innovation, *GeoJournal*, 49, 381-390.

Pacheco, A. and Tyrrell, T., 2002, Testing spatial patterns and growth spillover effects in clusters of cities, *Journal of Geographical Systems*, 4, 275-285.

Sadahiro, Y., 2003, Cluster detection in uncertain point distributions: a comparison of four methods, *Computers, Environment and Urban Systems*, 29(1), 33-52.

Shearmur, R. and Coffey, W., 2002, A tale of four cities: intrametropolitan employment distribution in Toronto, Montreal, Vancouver, and Ottawa-Hull, 1981-1996, *Environment and Planning A*, 34, 575-598.

Song, S., 1994, Modeling worder residence distribution in the Los Angeles region, *Urborn Stnckes*, 31, 1533-1544

Sweeny, S. and Feser, E., 1998, Plant size and clustering of manufacturing activity, *Geographical Analysis*, 30(1), 45-64.

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

Wang, F., 2001, Explaining intraurban variations of commuting by job proximity and workers' characteristics, *Environment and Planning B*, 28, 169-182.

Correspondence: Jungyeop Shin, Department of Geography Education, College of Education, Seoul National University, Seoul 151-748, Korea (jshin2@buffalo.edu)

교신 : 신정엽, 151-748 서울시 관악구 서울대학교 사범대학 지리교육과 (jshin2@buffalo.edu)

Received March 20, 2005

Accepted June 7, 2005