

상업적 토지이용 패턴의 시공간 변화 탐색을 위한 공간통계 기법 적용 연구

신정엽* · 이경주**

Research on Application of Spatial Statistics for Exploring Spatio-Temporal Changes in Patterns of Commercial Landuse

Jungyeop Shin* and Gyoungju Lee**

요약 : 많은 지리적 현상은 시간 변화에 따라 동적인 공간 패턴을 보이며, 이러한 동적인 공간 패턴을 탐색하기 위한 연구들이 수행 되어왔다. 그러나 기존의 많은 연구는 시간의 흐름에 따른 공간 패턴의 변화를 연속 또는 누적 측면에서 다루기 보다는 특정 시점이 나 기간 동안의 정적인 공간 패턴 분석에 초점을 두고 있다. 따라서 시간 변화 과정에서 수반되는 공간 프로세스의 관성(inertia)을 효과적으로 파악할 필요가 있다. 이러한 측면을 고려하여, 본 연구의 목적은 지리현상의 공간패턴을 탐색하는 새로운 공간통계 탐색 방법을 제안하고, 이를 사례연구에 적용하는데 있다. 즉, 새로운 공간통계량을 제안하고, 몬테카를로 시뮬레이션(Monte Carlo Simulation)을 통해 새로운 통계량의 z 값을 산출한 뒤, 시간 변화에 따른 공간 패턴의 변화를 누적 방식으로 탐색하는 방법을 소개 하고자 한다. 이를 위해 공간 패턴을 측정하는 J 통계량과 $CUSUM$ 통계량이 결합된 방법을 제안하고, 사례연구로 최근 200년 동안 미국 뉴욕 주의 이리 카운티(Erie County)의 상업적 토지이용의 공간 패턴 변화를 살펴 보았다. 이러한 시공간 패턴 변화 탐색 방법을 통하여 새로 구성된 공간통계량을 단위시간마다 누적적으로 반영하여 공간패턴의 연속적인 변화추이의 효과적인 탐색이 가능하였다

주요어 : 상업적 토지이용, 시공간 변화, J 공간통계량, 몬테카를로 시뮬레이션, $CUSUM$ 통계량

Abstract : Lots of geographic phenomena have dynamic spatial patterns with time changes, and there have been lots of researches on exploring these dynamic spatial patterns. However, most of these researches focused on the static pattern analysis in a given period, rather than dealing with dynamic changes in the spatial pattern over time with the continual or cumulative perspective. For this reason, investigation of the inertia of spatial process in terms of temporal changes is needed. From this background, the purpose of this paper is to propose the methodology to explore the changes in spatial pattern cumulatively by considering the inertia of the spatial statistics over time, and to apply it to the case study. That is, we introduce the new spatial statistic, and produce the z -values of the statistic using Monte Carlo Simulation, and then to explore the changes in spatial patterns over time cumulatively. To do this, the method to combine the J statistic with $CUSUM$ statistic for exploring spatial patterns, and to apply it to the changes in the commercial landuse in Erie County, New York State. Through the proposed method for spatio-temporal patterns, we could explore continual changes effectively in the spatial patterns reflecting the statistics by temporal spot cumulatively.

* 서울대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Seoul National University), geoshin@snu.ac.kr

** 뉴욕주립대(버팔로) 지리학과 박사과정(Ph.D. Student, Department of Geography, State University of New York at Buffalo), gjlee@buffalo.edu

Key Words : Commercial landuse, Spatio-temporal changes, J spatial statistic, Monte Carlo Simulation, *CUSUM* (Cumulative sum) statistic

1. 서론

도시공간 내의 많은 지리적 현상 (예, 상업 입지선정, 거주지 선정, 범죄나 교통사고 발생, 전염병 발생 등)은 시공간 차원에서 유의미한 패턴의 변화를 가지는 경우가 많다. 그런데 이러한 많은 지리적 현상은 특정 시점에서 우연히 발생하였기 보다는, 연속적인 시공간 프로세스를 통해 형성되는 경우가 많으며, 따라서 이러한 동적인 지리 현상을 올바르게 이해하기 위해서는 연속적인 시간의 흐름 속에서 공간 패턴의 변화를 탐색하는 것이 바람직하다. 이러한 동적인 공간 패턴을 탐색하기 위해 여러 공간통계 연구들이 수행되어 왔다. 그런데 많은 기존 연구의 전제는 시간 흐름에 따라 지리적 현상이 확률적 프로세스(stochastic process)를 통해 발생하며, 이 프로세스는 시공간 자기상관(spatio-temporal autocorrelation)의 특성을 가진다는 것이다. 즉, 특정 시점의 공간패턴은 이전까지 누적된 공간패턴의 특성과 관성과 밀접한 관련성을 가지며, 따라서 이를 잘 반영하는 확률변수를 공간패턴 탐색 과정에서 고려할 필요성이 제기된다.

이러한 측면에서 수행된 상당한 연구는 전역적 공간 패턴을 위한 탐색 연구이며, 이외에도 국지적 공간패턴 탐색을 위해 국지적 통계량(local statistic)을 이용한 연구도 수행되어 왔다(Anselin, 1995; Getis and Ord, 1992 등). 특히 국지적 통계량은 인접한 공간 단위간의 공간 상호작용을 반영한 변량의 공간 효과를 국지적으로 탐색할 수 있는 장점을 가진다. 그런데 공간분석의 수행 측면에서 전역적 탐색방법과 국지적 탐색방법은 서로 보완역할을 한다. 즉, 전체적인 공간패턴을 파악하는 데에는 전역적 탐색방법이, 그리고 국지적 탐색 방법은 전체 공간패턴에서 특징적인 국지적 공간패턴을 탐색하는데 유용하다.

그런데 기존의 많은 공간패턴 탐색 연구는 사용가능한 시계열 데이터의 부족 등의 이유로 연속적인 시공

간 패턴의 탐색보다는, 주로 특정 기간이나 시점에서의 정적인 공간패턴 탐색에 초점을 둔 경향이 있다. 연속적인 시간 흐름에 따른 공간 패턴의 변화를 탐색하기 위해서는 단위 시기별로 공간패턴을 누적적으로 반영하여 설명할 수 있는 방법이 필요하며, 이러한 맥락에서 공간패턴의 변화를 모니터링하기 위한 연구들이 수행되어 왔다¹⁾. 특히 공간통계량을 구성하고 매 시점에서의 독립적인 공간패턴의 변화과정을 시간 자기상관(temporal autocorrelation)을 구분하여 살펴보는 것은 매우 의미가 있다(Rogerson, 1997).

이러한 배경을 토대로, 본 연구는 상업적 토지이용의 시공간적 패턴변화를 살펴보기 위해 공간 클러스터를 탐색하는 새로운 공간통계량을, 시계열 통계량과 결합하여 제안하고자 한다. 즉, 공간 클러스터 정도를 간단하고 효율적으로 나타낼수 있는 J 공간 통계량을 구성하고, 이를 *CUSUM*(Cumulative Sum) 통계량에 결합하여 상업적 토지이용의 공간패턴의 변화과정을 탐색하기 위한 방법으로 제시하고자 한다. 이후 이 방법을 미국 뉴욕주의 이리 카운티 (Erie County)에 적용하여 실제적인 적용가능성을 살펴보는 것이 본 연구의 목적이다.

본 논문은 다음과 같이 구성된다. 2장에서는 공간패턴 탐색을 위한 공간통계량에 대한 사례 연구와 본 논문에서 새로이 제안될 J 통계량에 대해 기술하고, 다음으로 시계열 변화를 반영하는 통계량의 하나인 *CUSUM* 통계량에 대한 이론 검토와 적용이 이루어질 것이다. 그런 다음, J 통계량과 *CUSUM* 통계량을 결합하는 방법을 소개한 뒤, 3장에서는 실증적 사례연구와 더불어 그 결과를 논의하였다. 마지막으로 이러한 방법론적인 절차에 대한 확장 및 적용 가능성과 향후 연구과제에 대해 논의하고자 한다.

2. 공간통계량 및 시공간 모니터링 연구방법

여기서는 지리현상의 공간패턴을 계량화 하기위한 새로운 J 공간통계량을 고안하고, 시계열 변화를 측정하는 $CUSUM$ 통계량에 대해서 검토하고자 한다. 마지막으로, 상업적 토지이용의 시공간 패턴 변화를 탐색하기 위해 이 두 통계량을 결합하는 과정에 관하여 설명한다.

1) J 통계량

지금까지 공간 자기상관(spatial autocorrelation) 통계량을 이용하여 지리적 현상의 패턴을 탐색하는 공간통계 연구들이 있다. 이러한 연구들은 공간검색 범위에 따라 전역적 접근(global approach)과 국지적 접근(local approach)으로 구분되며, 전역적 접근은 연구지역 전반에 걸친 공간패턴 경향을, 국지적 접근은 특징적인 국지적인 공간 패턴을 탐색한다. 대표적인 전역적 방법으로는 Moran's I (Moran, 1948; Cliff and Ord, 1981), Getis G , Geary C , K -function 등이 있으며, 국지적 방법으로는 Local Moran's I 통계량 (Anselin, 1995), Getis' local G 통계량(Getis and Ord, 1992) 등이 있다.

공간통계량을 이용하여 토지이용 및 도시구조의 공간패턴 변화를 탐색한 많은 사례 연구는 주로 인문 현상의 클러스터 존재 유무와 공간집중 정도를 통계적 유의성(statistical significance)으로 측정한다²⁾. 대표적인 연구로는 도시 인구분포의 공간 패턴 변화 탐색 연구(Pacheco and Tyrrell, 2002)³⁾, 도시내 산업체 입지의 공간패턴을 탐색한 연구(Paci and Usai, 1999; Sweeney and Feser, 1998; Barff, 1987), 농촌지역의 산업체 공간패턴 탐색 연구(Ceccato and Persson, 2002)⁴⁾ 등이 있다. 특히 토지이용의 시공간 변화 연구와 관련하여 Cuthbert and Anderson의 연구(2002)는 K -function과 커널밀도방법(Kernel Estimate Method) 등을 이용하여 미국 핼리팩스(Halifax) 지역의 토지이용 패턴 변화를 1971년에서 1996년까지 살펴보았다. 그러나 이 연구는 공간패턴의 변화를 연속적인 시간범

위에서 살펴보지 못하고 특정 시점에서의 공간패턴 결과를 비교하는데 초점을 둬으로써, 시간 흐름에 따른 통계량 변화를 중점적으로 다루지 않았다. 국내 연구에서도 시공간 측면에서 공간 패턴과 분포에 대한 여러 연구가 수행되어 왔다. 이러한 연구들은 주로 특정 시기 동안의 변화량 계산을 통한 공간 패턴 연구(Choi, 2005; 권영아, 2006), 두 시점에서의 공간 패턴의 비교 연구(이민부 외, 2003), 시공간 상에서의 경로 접근성에 대한 연구(Kim, 2005; 박기호 외, 2005) 등에 초점을 두어 진행되었다.

본 연구에서는, 지리적 현상의 전체적인 공간군집정도를 나타내는 전역적 공간통계량은 $J(d)$ 로, 이를 구성하는 국지적 통계량은 $J_i(d)$ 로 하여 제시하고자 한다. 이들 공간통계량은 식 (1)-(2)에서 정의된다.

$$J(d) = \frac{\sum_i \sum_j w_{ij}(d)x_i x_j}{\sum_j w_{ij}(d)} \quad \text{식(1)}$$

$$J_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j w_{ij}(d)} \quad \text{식(2)}$$

식 (1)-(2)에서, $J(d)$ 통계량은 전역적 공간패턴을 나타내며, 이 통계량은 국지적 공간패턴을 나타내는 $J_i(d)$ 통계량의 합의 상수(constant) 배로 나타낼 수 있다. 이러한 특성은 국지적 통계량의 구성에 있어서 Anselin(1995)이 제시한 필요 조건에 부합한다. 그런데 전역적 통계량은 대상 지역의 전반적인 공간패턴 경향을 탐색하지만, 세부적인 국지적 공간패턴을 탐색하지 못하므로 이를 보완하는 필요성이 제기되어 왔으며 (Anselin, 1995), 이에 대한 대응으로 본 연구에서도 국지적 탐색 기법인 $J_i(d)$ 를 제안하였다⁵⁾. 여기서, d 는 근린 공간단위의 인접성 또는 지리적 영향권을 설정하는 공간검색 범위 패러미터이다. w_{ij} 는 공간객체 간의 지리적 영향력을 수치화한 공간가중행렬의 i 행 j 열의 요소를 나타낸다. 이 행렬은 두 공간객체를 각각 행과 열로 배열하고, 두 객체가 거리 d 내에 있으면 1을, 그렇지 않으면 0을 부여한다. 두 공간객체의 거리는 두 중심점(centroid) 간의 유클리디안(Euclidian) 거리를 사용한다.

즉, 개념적으로 $J_i(d)$ 통계량은 지리적으로 근접한 관찰값의 평균으로 볼 수 있다. 예를 들어, 지역 i 주변에 높은 관찰값을 가진 공간객체들이 지리적으로 클러스터되어 분포하면(중심점간의 거리가 d 보다 작은 경우), 이 지역 i 를 중심으로 이 범위(d) 내에 속한 주변지역 관찰값들과의 평균($J_i(d)$)은 높은 수치를, 반대의 경우는 낮은 수치를 보일 것으로 예상된다. 만약 이러한 관찰값들이 공간적으로 무작위(randomly) 분포한다면, 지리상의 위치변화에 따른 평균들간의 차이는 통계적으로 우연히 발생한 것으로 간주될 수 있다. 한가지 주목할 점은, 통계치 구성 방식에서 이 통계량은 Knox(1964)가 고안한 통계량과 유사하다. 즉, Knox 통계량은 시공간 인접성에 따라 특정 시공간 범위내에 속하는 공간객체의 쌍(pair)의 수로 정의되는 반면에, $J_i(d)$ 통계량은 공간객체간의 공간 접근성만을 기준으로 묶은 관찰값들의 평균으로 정의된다. $J(d)$ 통계량은 이러한 국지적 클러스터 현상을 계량화하는 $J_i(d)$ 통계량의 합으로 구성된다. $J(d)$ 는 연구지역 전반에 걸친 공간 패턴의 정도를 나타내는 반면, $J_i(d)$ 는 단위지역 i 를 중심으로 한 국지적 공간 패턴을 탐색한다⁶⁾. 만약 특정 지역에서 높은 값들의 국지적인 클러스터 현상이 강하면, 이 지역은 $J(d)$ 통계량의 증가에 양적(positive)으로 기여할 것이고, 이러한 $J_i(d)$ 통계량들이 전반적으로 우세하면, $J(d)$ 통계량은 높은 값을 가질 것으로 예상된다. 이러한 $J(d)$ 및 $J_i(d)$ 통계량은 전역적 또는 국지적 공간 패턴을 파악하는데 간편한 통계적 방법을 제공하며, 본 연구에서 파악하고자 하는 상업적 토지 이용의 공간패턴파악에 적합할 것으로 판단된다.

본 논문의 실증 사례연구는 상업적 토지이용의 공간 패턴 경향을 통계적으로 탐색하기 위해 $J(d)$ 통계량을 이용하였다. 분석의 결과로 $J(d)$ 통계량 값이 크면 상업적 토지이용 패턴이 공간적으로 집중됨을 예상할 수 있는 반면, $J(d)$ 통계량 값이 작더라도 $J_i(d)$ 통계량이 큰 공간단위들이 국지적으로 클러스터된 경우를 탐색하는데 유용할 것으로 판단된다. 즉, 전역적 통계기법과 국지적 통계기법을 함께 보완적으로 사용함으로써 상업적 토지이용 패턴을 탐색하는데 보완 역할이 기대된다.

그런데 한가지 유의할 점은 $J(d)$ 통계값 만으로는 공

간패턴의 통계적 유의성을 판단하기 힘들다. 즉 $J(d)$ 값으로 공간패턴의 경향을 파악할 수 있지만, 그 자체로는 결과의 통계적 유의성 정보를 제공하지 않는다. 대표적인 다른 공간 통계량들(예를 들어, Moran's I 나 Geary's C , 또는 Getis' G 통계량)의 경우는 각 통계량의 분석적인 평균과 분산값이 주어지고, 이들을 이용하여 구해진 통계량을 표준정규변수로 표준화하여 공간패턴의 통계적 유의성을 검증하고 해석하는 절차가 일반적이다. 그러나 본 연구에서 제시한 J 통계량의 경우는 위의 공간 통계량들처럼 분석적인 평균과 분산의 유도과 관련한 추가적인 검증절차에 대한 연구는 수행하지 않았다. 이는 본 연구의 주 목적이 새로운 공간통계량을 고안하고 그 통계적 특성을 상세히 다루기보다는, 연구목적에 따라 계량화가 필요한 개념을 통계치로 구성하여 활용하는 과정을 실험적으로 제시하는 데에 있기 때문이다. 또한 새로 구성되는 공간통계량에 대해서 분석적인 평균과 분산의 산출이 쉬운 과정은 아니며, 따라서 고안된 통계량의 유의성 검정을 위해서 수치적(numerical) 방법 또는 몬테카를로 시뮬레이션(Monte Carlo simulation)을 이용하여 p -값을 구하는 경우(예를 들어, Kulldorff의 Spatial Scan Statistic)가 많다.

몬테카를로 시뮬레이션(Monte Carlo simulation)은 특정한 귀무가설의 조건하에서 무작위로 추출된 자료에 대한 통계치를 구하고, 이러한 과정을 반복적으로 수행하여 경험적 통계분포를 산출한뒤, 이를 활용하여 관찰된 통계값의 통계적 유의성 검증이나 신뢰구간 등을 구하는 방법을 포괄적으로 지칭한다(Manly, 1998). 이 중에서 많이 활용되는 대표적인 방법들은 무작위 재배열 접근방식(random permutation approach), 부트스트랩(bootstrap) 방법이나 잭나이프(Jackknife) 추정치 등이 있다. 이러한 방법들은 새로 제안된 통계량의 이론적인 모수들의 특성을 유도하기 어려운 경우에 효과적이고 편리하게 활용될 수 있다(Besag and Diggle, 1977; Findlay and Findlay, 1984).

본 논문은 다음 두 가지 이유로 몬테카를로 시뮬레이션 방법의 하나인 무작위 재배열 접근방식을 채택하였다. 첫째, 무작위 재배열 접근방식을 활용함으로써, 관찰값을 통해 그 자료가 추출된 모집단의 평균, 분산

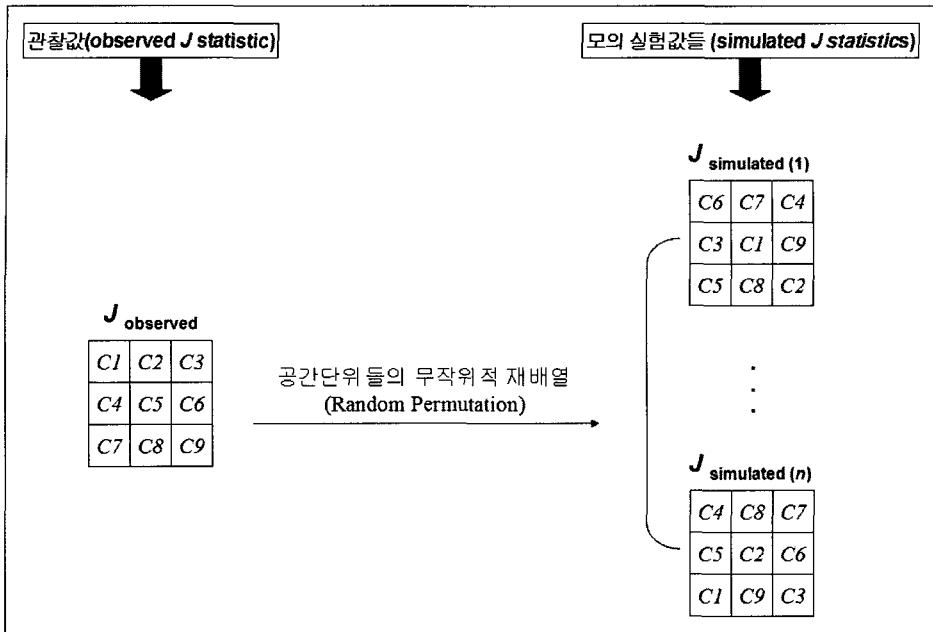


그림 1. 시점 t 에서의 평균과 분산을 구하기 위한 몬테카를로 시뮬레이션 과정

및 신뢰구간(confidence interval)을 효과적으로 구할 수 있으며, 추정과정도 쉽게 구현할 수 있다(Manly, 1998). 둘째, 전역적 공간 자기상관이 유의미하게 존재할 경우, 공간상에서 공간 사상의 발생과 관련된 우도의 동일성(equal likelihood)의 가정이 위반될 가능성이 높다. 따라서 국지적 공간 통계량의 이론적 모수의 값(평균과 분산)이 상당히 왜곡될 수 있으며, 결과적으로 통계적 유의성 검정이나 국지적 통계량의 해석에 있어서 오류를 범할 위험성이 지적되고 있다(Anselin, 1995). 이러한 경우 Anselin(1995)은 무작위적 재배열 접근방식을 활용한 통계적 유의성 검정을 실용적인 대안으로 제안하였다. 이 접근 방식은 Moran's I 와 같은 공간자기상관을 나타내는 통계량의 p -값을 구하는데 많이 활용되고 있으며, Getis's G (1992) 통계량에 대한 적용가능성도 논의되고 있다.

이러를 고려하여 본 논문에서는, 무작위 재배열 접근방식을 활용하여 제안된 전역적, 국지적 J 통계량의 정규모수들을 추정하여 표준정규변수를 산출하는 과정을 제안하였다. 이러한 시뮬레이션 과정은 그림 1에 예시적으로 설명되고 있다.

그림 1에서처럼 특정 시점에서 관찰되고 있는 공간

단위(여기서는 9개)별 상업적 토지이용의 수($C1, \dots, C9$)를 무작위로 재배열하는 반복 과정(n 회)을 통해 $J(d)$ 통계량의 모의 실험값을 구할 수 있다. 이렇게 구해진 $J(d)$ 통계량 (표본평균에 해당)의 모의실험값의 분포는 중심극한정리(Central Limit Theorem)에 의해 정규분포를 이룬다. 본 연구에서는 단위지역의 수가 243 개로서 30개 이상이며, 따라서 $J(d)$ 통계량은 정규분포의 성격에 잘 부합될 것으로 예상된다. 국지적 통계량인 $J_i(d)$ 에 대해서도 같은 방식으로 정규화하였다. 국지적인 통계량인 $J_i(d)$ 의 경우 Getis and Ord(1992)가 지적한 바와 같이 공간 범위가 확대됨에 따라서 합하는 관찰값의 수가 증가될 것이며 이에 따라 점차 정규 분포에 가까워질 것이다.

이를 바탕으로, J 통계량들의 평균 및 분산은 식 (3)-(4)로 표현되며, 이들을 이용하여 $J(d)$ 통계량의 표준화된 값(z -value)은 식 (5)로 나타낼 수 있다. 한편, $J_i(d)$ 통계량의 평균 및 분산은 식 (6)-(7)로 표현되며, 이들은 이용하여 $J_i(d)$ 통계량의 z -값은 식 (8)로 나타낼 수 있다.

$$E[J(d)] = \frac{\sum_{k=1}^n J(d)_{\text{simulated}(k)}}{n} \quad \text{식(3)}$$

$$\text{Var}[J(d)] = \frac{\sum_{k=1}^n (J(d)_{\text{simulated}(k)} - E[J(d)])^2}{n-1} \quad \text{식(4)}$$

$$Z[J(d)] = \frac{J(d) - E[J(d)]}{\sqrt{\text{Var}[J(d)]}} \quad \text{식(5)}$$

$$E[J_i(d)] = \frac{\sum_{k=1}^n J_i(d)_{\text{simulated}(k)}}{n} \quad \text{식(6)}$$

$$\text{Var}[J_i(d)] = \frac{\sum_{k=1}^n (J_i(d)_{\text{simulated}(k)} - E[J_i(d)])^2}{n-1} \quad \text{식(7)}$$

$$Z[J_i(d)] = \frac{J_i(d) - E[J_i(d)]}{\sqrt{\text{Var}[J_i(d)]}} \quad \text{식(8)}$$

식 (3)-(4)와 (6)-(7)에서 n 은 몬테카를로 시뮬레이션의 횟수를 나타내며, n 이 클수록 평균과 분산의 무작위 오차 (random error)가 작아질 것으로 예상된다. 본 논문에서는 시뮬레이션의 횟수인 n 을 999회로 정하였으며, 이를 통하여 산출된 z -값들은 소수점 2자리 이하에서 약간의 무작위 오차만을 보였다.

위의 식을 이용하여 공간통계량과 z -값의 관계를 살펴보면 높은 $J(d)$ 값은 z -값이 높을 것으로 기대된다. 그리고 $Z[J(d)]$ 값의 p -값이 설정된 유의수준(예, $\alpha=0.05$) 보다 작으면, 속성값이 큰 공간 클러스터 경향이 통계적으로 유의하다고 판단된다. 국지적 통계량의 경우도 $J_i(d)$ 값 자체보다는 그것의 표준정규분포 값인 $Z[J_i(d)]$ 를 탐색에 활용하는 것이 가능하다⁷⁾. 한가지 주목할 것은, 국지적인 z -값에 대한 통계적 유의성과 관련된 해석인데, 전역적 통계량의 해석과는 달리 국지적 통계량은 유의성 평가가 단위지역의 수 만큼 동시에 이루어지므로, 다중가설검정 (multiple hypothesis test)과 관련한 문제가 발생할 수 있다. 물론 Bonferroni 조정과정 등을 통해 이 문제를 어느 정도 다룰 수 있지만, 국지적 통계량은 통계적 유의성 검증수단으로서의 유용성보다는, 공간 클러스터 지역을 z -값이나 p -값을 지도화하여 탐색하는 유용성이 더 강조되고 있다. 따라서 본 연구에서는 공간탐색 측면에서 $Z[J_i(d)]$ 를 지도화 하여 활용하였다.

J 통계량 제안과정의 시사점으로는 다양한 공간현상의 특성을 계량화하기 위해 새로운 공간통계량의 구성이 필요한 경우 유용한 대안으로 활용될 수 있다는 것이다. 즉 연구자가 연구목적에 부합하는 통계값을 새로 정의하여 활용하는 경우, 통계값의 분석적인 통계 분포의 모수를 산출하기는 쉬운 과정이 아니며, 따라서 본 논문에서 활용한 시뮬레이션을 통한 모수추정방법이 실제적인 하나의 대안으로 활용될 수 있을 것으로 판단된다. 본 논문의 경우, 기존의 통계량 (예를 들어, 이론적인 분포와 관련된 모수가 정의되어 있는 G 통계량)을 통하여 상업적 토지이용의 시공간적 패턴변화 과정을 살펴볼 수도 있겠지만, 새로운 통계량을 정의하고 시뮬레이션 과정을 실험적으로 적용해보는 것은 큰 의미가 있는 것으로 생각되어 새로운 통계량을 정의하여 활용하였다.

지금까지 살펴본 $J(d)$ 통계량의 z -값은 특정 시점 또는 기간 동안의 단편적 패턴만을 설명해준다. 예를 들어 1800년부터 1998년까지 200년 동안 발생한 상점입지의 공간 패턴을 탐색하는데 표준화된 $J(d)$ 값만으로는 전체 기간 중 어느 시점에서 상점입지가 집중 또는 분산되었는지 구체적으로 파악하기 어렵다. 이러한 이유 중의 하나는 각 시점별로 연속적인 공간패턴변화의 관성에 기인한다. 따라서 공간패턴을 나타내는 변량의 통계적인 중복을 적절히 다루기 위한 방법론적인 탐색 기법이 요구된다⁸⁾. 이러한 맥락에서 다음 절에서는 특정 시점의 공간패턴 탐색이 아니라, 시간 흐름에 따라 확률변수의 변화과정을 누적적으로 반영하기 위해 고안된 $CUSUM$ 통계량을 소개하고, 다음절에서는 $J(d)$ 통계량이 어떻게 $CUSUM$ 통계량과 결합하여 시공간 변화패턴을 연속적으로 탐색 가능한지를 살펴보고자 한다.

2) $CUSUM$ (cumulative sum) 통계량

원래 $CUSUM$ 통계량은 산업공학 분야에서 통계적 품질관리를 목적으로, 특정 제품 생산공정에서 발생하는 오차중에서 통계적으로 유의미한 변이만을 누적적으로 반영하여 변량의 평균값의 체계적인 변화를 신속하게 탐지하는 방법으로 사용되어 왔으며 다음과 같이

정의된다(Rogerson and Sun, 2001).

$$S_t = \max(0, S_{t-1} + z - k) \quad \text{식(9)}$$

위의 식에서 S_t 는 바로 전 시점($t-1$)에서 S_{t-1} 에 $z-k$ 만큼 누적된 값이다. 여기서, z 는 현 시점 t 에서의 표준화된 정규변수 값이고, k 는 변화의 기준이 되는 임계값(threshold)이다. 본 연구에서는 표준화된 $J(d)$ 통계량이 z 값에 해당한다. 만약, t 시점에서의 z 값이 임계값 k (k 는 보통 정규화된 변수 z 에 대해서 통상적으로 0.5로 설정된다)보다 크면 z 와 k 의 차이가 S_{t-1} 의 값이 증가하는 데 기여하면서 누적되고, 반대의 경우는 감소하는데 기여하면서 누적된다. 이렇게 누적된 값이 또 다른 임계값 b 를 초과하면, 이는 대상 변량의 기대값이 이 시점에 이르러 통계적 의미에서 근본적, 체계적으로 변화가 발생했음을 알려준다(Rogerson and Sun, 2001). 한편 임계값 b 는 일반적으로 연구 대상의 변량에 따라 경험적으로 정의되며, b 를 계산하기 위하여 Siegmund(1985)는 아래와 같은 식을 제시하였다.

$$ARL_o = \frac{\exp(2k(b+1,166)) - 2k(b+1,166) - 1}{2k^2} \quad \text{식(10)}$$

위의 식에서 ARL_o (In-control Average Run Length)

은 S_t 가 임계값 b 에 도달할 때까지의 평균적인 시행단위 또는 시간의 길이이다. 설탕 공정을 사례로 들면(Hawkins and Olwell, 1998), 공정 관리자는 설탕 완제품의 무게를 400g으로 설정하고 제조 기계를 가동시킨다. 그러나 실제 공정이 진행됨에 따라 여러 오차에 의해 개별 설탕 완제품은 정확히 400g으로 생산되지 않을 것이다. 그런데 오차가 무시할 수 있는 수준이라면 생산공정은 계속 진행되지만, 체계적인 오차(systematic error)에 의해서 설탕 완제품의 무게가 계속 400g을 초과하거나 미달되면, 공정상의 오차를 파악하고 조치를 취할 필요가 있다. 이때 관리자는 이러한 오차를 가능한 한 빨리 발견할 필요가 있으며, 이때 과거의 경험을 바탕으로 제조 공정의 체계적인 오차의 발생 여부를 판단한다. ARL_o 는 체계적인 실제 오차(b)에 도달하기까지 평균적인 설탕 제조공정 횟수라고 할 수 있다.

그런데 이 과정에서 필연적으로 판단과 관련된 오류가 수반된다. 즉 설탕제조 공정에서 실제로 체계적인 오차가 발생하지 않았는데도 기계를 조정하는 경우가 있는 반면, 실제 체계적인 오차가 발생했지만 기계의 조정 없이 공정을 계속 진행하는 경우도 있다. 실제로 체계적인 오차가 발생하였는데 관리자가 설정한 임계값에 도달하지 않아서 공정이 계속 진행되는 평균적인 시행시간의 길이는 ARL_i (Average out-of-control Run

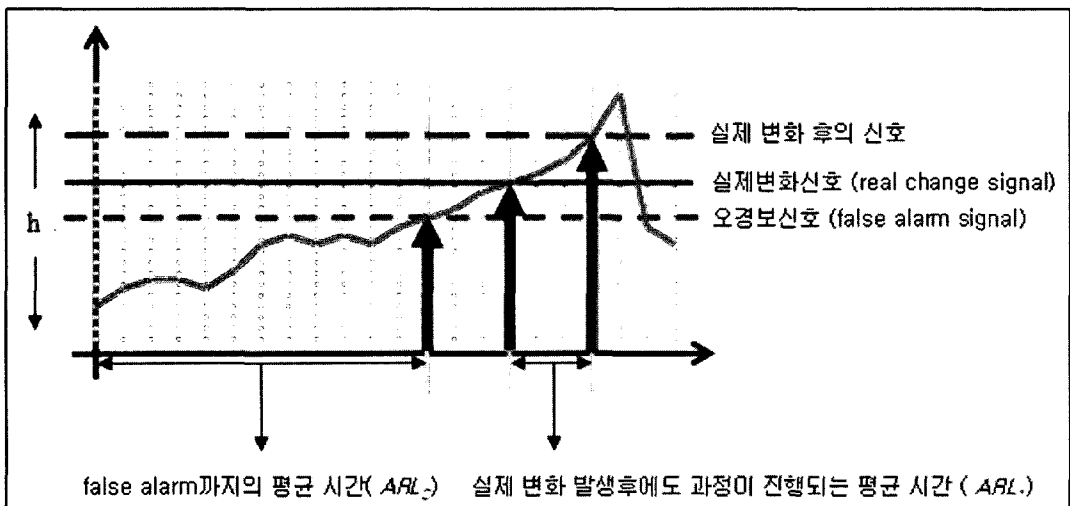


그림 2. CUSUM 차트의 및 ARL_o , ARL_i , h 의 개념도

Length)로 표시된다⁹⁾. 관리자 입장에서는 두 경우 모두 바람직하지 않다. 보통 임계값 b 는 ARL_0 과 관련된 경험적 지식을 바탕으로 위의 식을 이용하여 계산되며 연구 대상에 맞게 이를 설정할 필요가 있다. 즉 의사결정자 입장에서는 체계적인 오류가 발생하는 시점까지의 시간을 가능한 최대로 하여 오경보율(false alarm rate)을 최소화 하는 동시에 실제 변화를 탐지하는 확률을 최대화(또는 ARL_1 을 최소화) 하도록 ARL_0 (결과적으로 b)값을 맥락에 맞게 경험적으로 설정해야 한다¹⁰⁾. 그림 2는 임계값 b 와 ARL_0 , ARL_1 와의 상호 관련성을 나타낸다.

본 연구에서 임계값 b 관련하여 변화 추이를 나타내는 변량은 공간통계량 $J(d)$ 의 표준화된 정규변수이다. 만일 시간 변화에 따라 상업적 토지이용의 발생이 국지적으로 클러스터되고 전역적으로도 이러한 경향이 우세하다면, 표준화된 z 값은 증가할 것이다. 이때 어느 시점에서 설정된 임계값보다 커질때, 과거로부터의 공간 클러스터 과정이 체계적으로 진행되어 그 시점에 이르렀음을 나타내준다. 이를 바탕으로 다음 절에서는 $J(d)$ 통계량과 $CUSUM$ 통계량을 결합하는 과정을 설명한다.

3) 표준화된 $J(d)$ 와 $CUSUM$ 통계량의 결합에 관한 방법론적 고찰

$J(d)$ 통계량을 $CUSUM$ 통계량의 변수로 사용하기 위해서는 우선 $J(d)$ 통계량을 z 값으로 표준화 시킬 필요가 있으며 이를 위해 몬테카를로 시뮬레이션을 이용하였다. $J(d)$ 통계량을 표준화하여 $CUSUM$ 방법과 결합할 때, 과거의 공간패턴의 관성을 배제하여 현 시점 t 에서의 변화된 부분만을 반영해야 한다. 즉 어떻게 새로운 변화만을 $CUSUM$ 값에 누적적으로 반영할 것인가를 주의해야 하며, 이는 $CUSUM$ 방법의 전제조건인 하나인 관찰 값들 간의 시간 독립성과 관련된다. 앞서 언급한 것처럼 공간 패턴화 과정은 시간 흐름에 따른 관성을 내포하며, 이전 시점까지 형성된 공간 패턴이 바로 다음 시점의 공간패턴에 영향을 줄 수 있으며, 결과적으로 시간 독립성 가정이 무시되면 결과가 왜곡될 수 있다. 이 점을 고려하여 매 시점마다 그 시점에서

발생한 변화만을 분리할 필요가 있다. 시점 t 에서의 독립 변량 z_t 는 아래와 같이 표현된다(Rogerson and Sun, 2001; Lee and Rogerson, 2007).

$$Z_t J(d) = \frac{J_t(d) - E[J_t(d) | J_{t-1}(d)]}{\sqrt{\text{Var}[J_t(d) | J_{t-1}(d)]}} \quad \text{식(11)}$$

본 연구에서는 매 시점에서의 독립적인 공간패턴의 변화를 반영하기 위해 다음의 방법을 사용하고자 한다. 즉 시점 t 에서의 독립적인 변화를 $CUSUM$ 통계량에 반영하기 위해, $t-1$ 부터 t 시점까지에 대해서만 새로 발생된 상업적 용도의 필지 수를 변수로 선정하여 $J(d)$ 의 기대값과 표준편차를 산출한다. 그리고 이 값을 표준화 ($z[J(d)]$)하면 매 시점에서의 독립적인 z 값을 구할 수 있다. 시점 t 에서의 변량 z_t 는 S_{t-1} 에 누적되어 현 시점 t 까지의 공간패턴(분산 또는 클러스터)에 반영된다.

$$S_t = \max(0, S_{t-1} + z_t [J(d) - k]) \quad \text{식(12)}$$

이러한 방식으로 매 시점에서 새로이 계산된 z 값은 S_t 에 누적되어 그 시점까지의 공간패턴의 변화를 반영한다. 이 과정을 개념화 하면 그림 3과 같다.

즉, 그림 3의 (나)에서처럼 시점 $t-1$ 에서 시점 t 로의 시기 동안 변화가 수반된 셀은 음영된 오른쪽 하단의 4개 셀로 국한되며, 변화된 셀에 대해서 $J(d)$ 값이 계산되면 시점 t 에서 변화된 공간패턴을 독립적으로 탐색할 수 있다. 반대로 앞의 방법을 사용하지 않고 시점 t 에서의 $J(d)$ 값을 계산하면(그림 3의 다), 시점 $t-1$ 까지의 관성으로 인한 $J(d)$ 값 간의 독립성이 위반되어 $CUSUM$ 값이 왜곡될 수 있다.

마지막으로 $CUSUM$ 통계량 방법과 관련하여 다음 두 가지 점을 유의할 필요가 있다. 첫째는 임계값 b 에 대한 해석이다. 원래 $CUSUM$ 연구는 예기치 않은 지리적 현상의 시공간 발생을 사전에 신속하게 탐지하기 위한 취지로 지리학 연구에 적용되었다(Rogerson and Sun, 2001). 그러나 본 연구는 이러한 시공간 패턴의 사전 모니터링 측면보다는 $CUSUM$ 방법이 가지는 장점(과거의 공간과정의 추이를 체계적으로 누적 반영할 수 있는 방법론적 특성)과 관련한 적용가능성에 대한 실험적 실증 연구로서의 취지로 수행되었으므로 임계

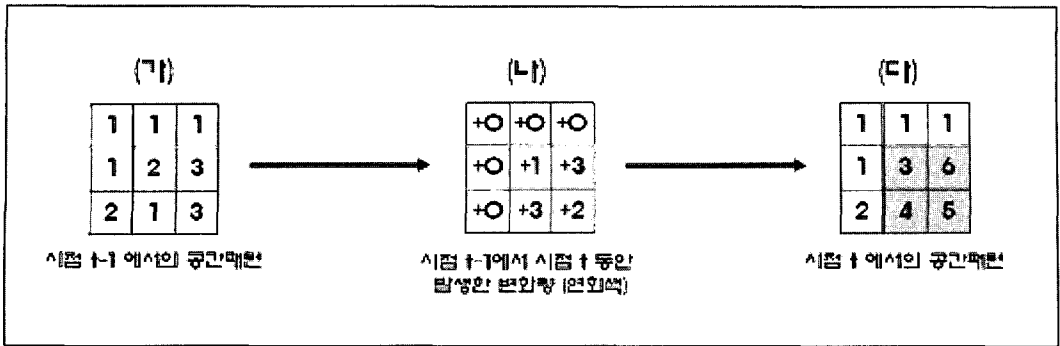


그림 3. 각 시점에서의 독립적인 공간패턴 변화량 산정 과정

값을 다른 방식으로 다루었다. 이러한 제안의 하나는 공간패턴의 변화를 CUSUM 차트에 표시하고, 이 차트의 최고점 (peak)에 대응하는 시점을 포함한 변화 정보를 토대로 공간패턴 결과를 해석하는 것이다. 최고점에 해당하는 시점에서는 상업적 토지이용의 연속적인 공간 클러스터 경향성이 최고조에 달하는 것으로 해석할 수 있으며, 최고점 이후 공간패턴이 점차 분산되는 경향으로 해석할 수 있다.

둘째는 임계값 b 를 넘어선 시점 이후의 재조정 과정에 대한 사항이다. 질병 또는 범죄 발생의 모니터링과 같은 기존 연구에서처럼, 체계적인 변화가 탐지된 후에는 일련의 조치를 통해(예, 질병 예방/ 치료 프로그램 가동 및 순찰강화 조치) 탐지된 공간패턴 변화가 재조정되어 무작위 공간 패턴으로 돌아갈 수 있으며, 이러한 과정이 반복되면서 사전 감시체계가 가능하다. 그러나 본 연구와 관련하여 상업적 토지이용의 공간패턴 변화가 바람직하지 못한 패턴으로 간주되어 재조정이 반드시 필요한 것은 아니며, 이러한 맥락에서 재조정의 해석 의미와 관련한 후속 논의가 필요하다. 따라서 본 연구는 임계값 설정과 체계적인 공간패턴변화에 대한 경보 발생 보다는, 시간의 흐름에 따른 공간패턴 변화를 통계적 모델 속에서 누적 반영함으로써 시간 변화에 따른 공간 패턴 변화에 대한 해석에 보다 초점을 두고자 한다. 예를 들어 상업적 토지이용이 어느 시점에서 공간 집중 경향을 보이고 어느 시기에서 분산 패턴을 보이는지 등을 모니터링하고, 어디서 국지적인 클러스터가 발생하는 지를 탐색한다.

이를 위해 본 연구는 CUSUM 방법을 통해 전역적으

로 유의미한 공간패턴 변화를 보이는 시기 또는 시점을 탐색하고, 해당 시기/시점에서의 $J(d)$ 통계량의 국지적 변량인 $J_i(d)$ 통계량의 z -값을 지도화 한 뒤, 국지적인 상업입지 패턴의 공간적 유사성을 탐색하였다. 이러한 탐색 과정을 통해 시간 추이에 따라 상업적 토지이용의 공간 패턴이 무작위 변화보다는 시공간상에서 체계적인 변화를 수반하는 지에 대한 실증적 고찰이 가능하게 될 것으로 예상된다.

3. 실증적 사례 분석 및 결과

본 연구는 앞서 살펴본 방법론의 실증적 사례연구를 위해 미국 뉴욕 주, 이리 카운티 (Erie County) 지역의 상업적 토지 이용패턴을 연구 대상으로 선정하였다. 사용된 시계열의 공간 데이터는 미국 뉴욕 주에서 과세를 위해 조사되는 지가감정 데이터로 ORPS(Office of Real Property Services)에서 제작하고 있다. 이 데이터베이스에는 필지(parcel)의 주요 정보에는 필지의 중심점(centroid)의 공간좌표(x-y coordinate), 해당 용도의 설정 연도 등이 포함된다. 사례지역인 뉴욕 주 이리 카운티 (Erie County)는 약 36만개의 필지로 구성되어 있으며(시간흐름에 따른 필지의 물리적 변화가 수반됨), GIS shapefile 포맷으로 제공된다. 사례 연구의 시간 범위는 보다 오랜 시기 동안의 공간패턴을 보기 위해 1800년부터 1998년까지 최근 약 200여년을 설정하였으며 이에 적용된 시간 해상도는 1년 단위로 공간패턴의 변화를 파악할 수 있다¹¹⁾.

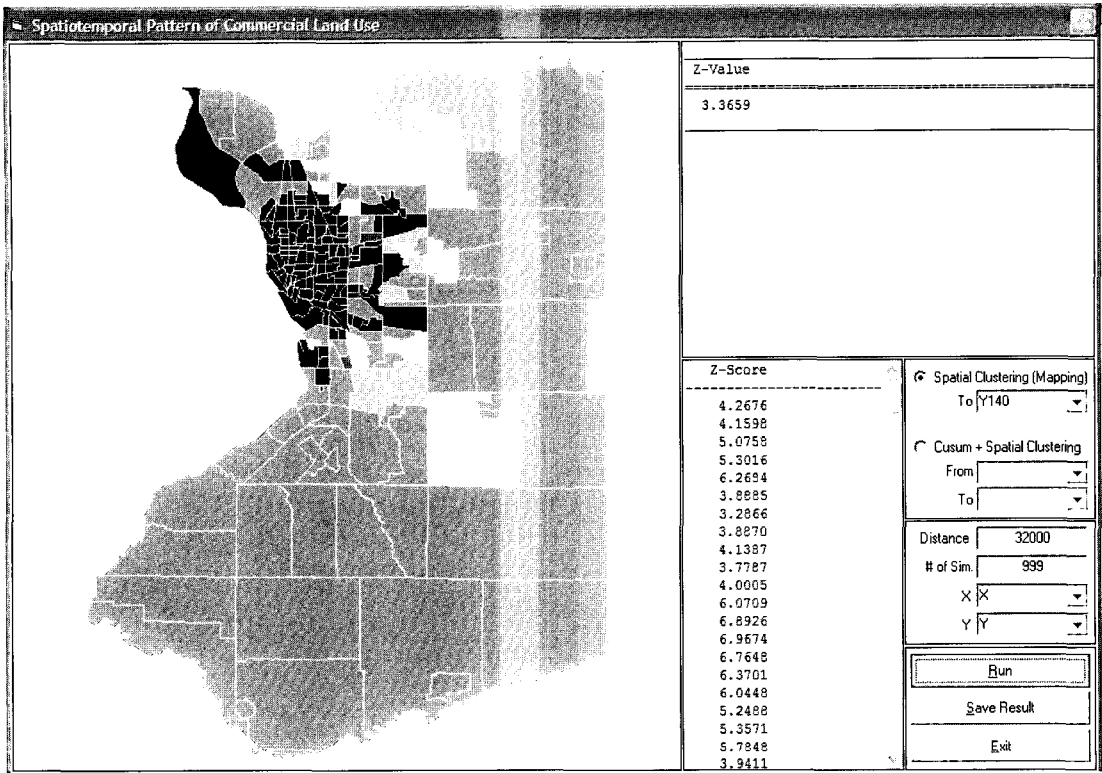


그림 4. CUSUM과 $J(d)$ 를 결합한 새로운 통계방법론의 프로그램 구현 GUI

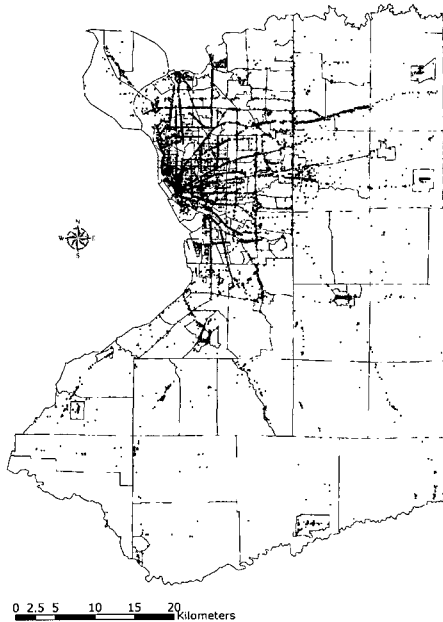


그림 5. 이리 카운티 지역의 상업적 토지 이용 패턴 (1800~1998)

$J(d)$ 통계량과 CUSUM 통계량의 결합된 알고리즘의 구현은 Visual Basic을 통해 수행되었다(그림 4). $J(d)$ 통계량과 그것의 CUSUM 통계량과의 결합 방법의 구현은 기존의 GIS 소프트웨어 패키지에서 제공되지 않고 있다. 따라서 Visual Basic을 이용한 프로그래밍 과정을 통해 본 제안 알고리즘을 구현하였으며, 분석 결과로 도출된 연속된 CUSUM 값은 MS Excel 프로그램에서 선 그래프로 시각화 하였다. 그리고 CUSUM 값이 최고치로 나타난 시점까지의 누적된 상업적 토지이용의 필지 수에 대한 $J(d)$ 값의 분포는 ArcGIS를 이용하여 지도화 하였다. 그림 5는 1800년부터 1998년까지 이루어진 사례 지역의 상업적 토지이용 분포를 보여주고 있다.

그림 5에서 볼 수 있는 것처럼 상당한 부분의 상업용 토지 필지는 중심업무지역(CBD)(지도상에서 북서부 해안 지역 근처)을 중심으로 집중적으로 분포하고 있음을 알 수 있으며, 이 CBD를 중심으로 주요 간선도로를 따라 방사상으로 분포하고 있음을 알 수 있다. 즉,

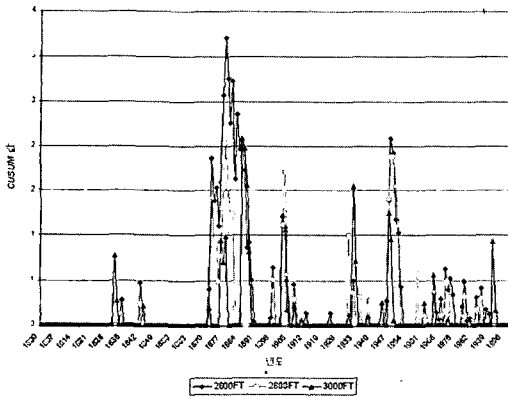


그림 6. 상업적 토지이용의 CUSUM 값 추이
($d=2600, 2800, 3000$ feet)

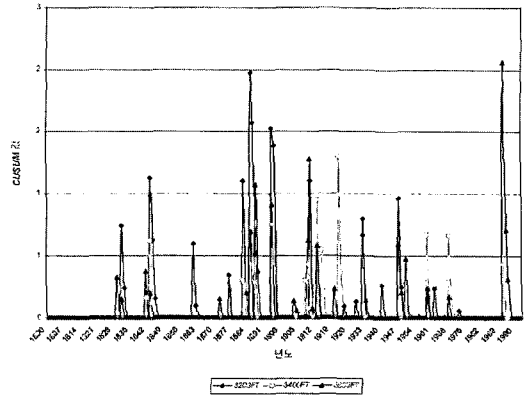


그림 7. 상업적 토지이용의 CUSUM 값 추이
($d=3200, 3400, 3600$ feet)

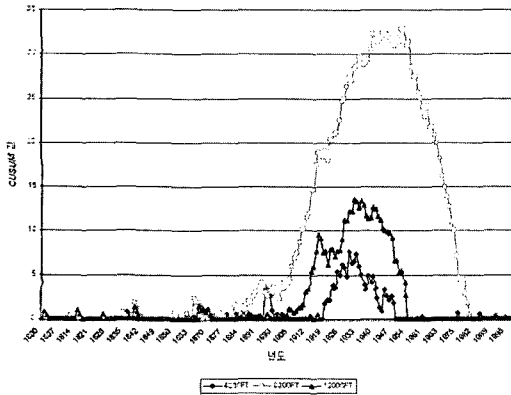


그림 8. 상업적 토지이용의 CUSUM 값 추이
($d=4000, 8000, 12000$ feet)

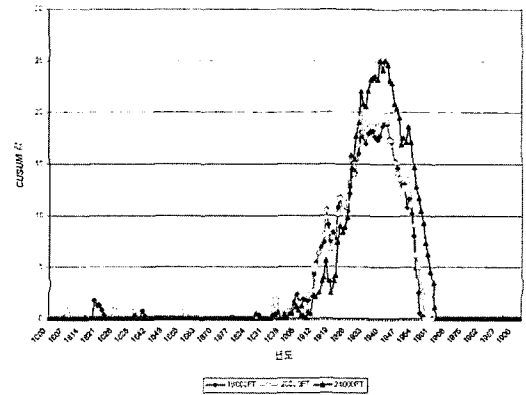


그림 9. 상업적 토지이용의 CUSUM 값 추이
($d=16000, 20000, 24000$ feet)

지도를 통해 CBD 지역 중심의 클러스터를 경험적으로 파악할 수 있으며, 따라서 분석 결과로 높은 $J(d)$ 값이 예상된다. 그러나 이 공간 패턴은 오랜 시기 동안 형성되었으며, 본 연구의 분석설계에 따라 200 여년 동안 1년 단위로 공간패턴 변화를 누적하여 CUSUM 값을 탐색하면, 상업적 토지이용 패턴 변화를 시간 흐름에 따른 형성과정 측면에서 보다 세밀하게 살펴볼 수 있을 것이다. 즉 CUSUM 방법을 통해 어느 시기에 전역적인 공간 클러스터가 형성되어 왔으며 이러한 클러스터 현상이 심화 또는 완화되는 지 등 공간패턴의 변화에 대한 탐색이 가능할 것이다.

이때 유의할 점은 $J(d)$ 통계량이 공간검색 범위(d)에 따라 결과가 달라진다는 것이다²⁾. 즉, 시공간 범위가 동일한 데이터에 대해 공간검색 범위를 달리 적용하면

상이한 공간 패턴이 탐색된다. 예를 들면 공간검색 범위가 커짐에 따라, $J(d)$ 는 점차 증가하다가 어느 시점에 이르러 다시 감소하거나 일정한 값으로 수렴할 수 있다. 이때 $J(d)$ 이 최대 값을 가지는 스케일에서 클러스터 현상이 가장 의미있게 또는 활발하게 나타난 것으로 해석할 수 있다. 이를 위해 본 연구는 여러 공간검색 범위를 적용하여 각 스케일에서의 공간패턴의 변화를 비교하고자 하였다.

이를 위해 여러 공간검색 범위가 실험, 적용되었는데, 작은 스케일 범위는 2,600 feet(약 790 meter)에서 3,600 feet(약 1,220 meter)까지 200 feet 단위로, 그리고 큰 범위의 경우는 4,000 feet부터 24,000 feet까지 4,000 feet 단위, 그리고 마지막으로 32,000 feet에서 56,000 feet까지 12,000 feet 단위로 CUSUM 값을 산출

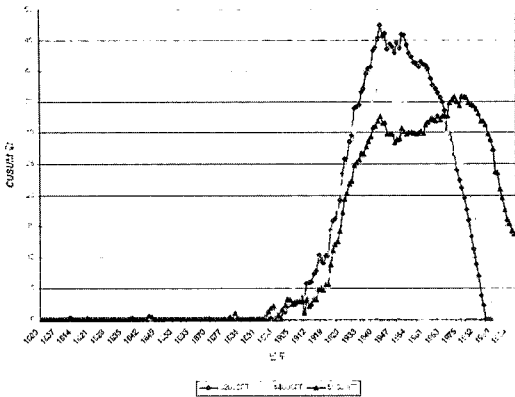


그림 10. 상업적 토지이용의 CUSUM 값 추이
($d=32000, 44000, 56000$ feet)

하여 그 변화량을 비교하였다. 선정된 스케일 패러미터는 연구자가 사례지역의 검토를 통해 경험적으로 설정하였으며 공간스케일 변화에 따라 상이한 공간패턴이 탐색되었다³⁾. 그림 6에서 10까지는 이러한 공간검색 범위에 따른 공간패턴의 누적된 변화양상의 결과를 나타낸다.

작은 공간범위 스케일의 경우(2,600~3,600 feet)는 시기별 공간 클러스터 패턴이 불규칙 또는 주기적 특성을 보이고 있다. 즉, 이 경우 J 통계량 산출을 위해 포함되는 근린 공간단위들이 상대적으로 적으므로, $J(d)$ 의 변이정도가 커서 각 시기별 공간패턴의 안정성이 낮은 것에 기인한 것으로 사료된다. 따라서 작은 공간스케일의 경우 관찰되는 뚜렷한 상업적 토지이용의 클러스터 패턴이 나타나지 않는 것으로 볼 수 있다. 반면 공간스케일이 4,000 feet 이상인 경우에는, 시기별 공간패턴의 세부적인 차이는 있지만, 공통적으로 1900년대 초반을 전후하여 상업적 토지이용의 공간 클러스터 정도가 증가하기 시작해서 1930~40년대에 이르러 최고치를 보이고 있다. 즉, $Z_i(d)$ 값이 1880년 후반부터 1920년대 후반에 이르는 시기동안 급격히 증가하는 경향을 가지며, 결과적으로 이러한 높은 수치들이 누적되어 1930~1940년 시기동안 공간 클러스터 경향이 극대화 됨을 보여준다. 이후 클러스터 경향이 점차 완화되면서 CUSUM 값도 줄어드는 경향을 보인다.

한편 그림 8에서는 이러한 공간 클러스터 현상이 4,000 feet에서 8,000 feet 사이에서 급격히 증가했음

을 알 수 있으며, 그 후로 12,000 feet에 이르는 동안 다시 감소하여 왔음을 볼 수 있다. 또한 그림 8, 9, 10에서 볼 수 있는 바와 같이 공간 클러스터 정도가 8,000 feet 이후에 16,000, 20,000, 24,000 feet에 이르는 동안 다시 소폭으로 증가하며 32,000 feet에 이르는 동안 최고치에 도달한 후 점차 감소하고 있다. 따라서 이리 카운티의 과거 200년간의 상업적 토지이용은 1880년대 후반부터의 급격한 지역경제 발전에 힘입어, 1930~40년대 초반에 이르기까지 공간 클러스터화 경향이 최고조에 이르렀음을 알 수 있다. 이때 그 클러스터 범위는 약 32,000 feet (약 9.8km) 정도에서 최대치에 도달했음을 유추해 볼 수 있다. 그러나 이 시기 이후 상업적 토지이용의 클러스터 경향이 점차 완화되고 있으며, 그 완화정도(클러스터 경향에서 분산화로의 추이)나 시기도 각 공간범위별로 차이를 보이고 있다. 즉, 모든 시기에 동일한 변화 정도와 속도로 공간패턴이 클러스터화 또는 분산화가 이루어지는 것은 아니며, 시기에 따른 패턴의 변화 양상의 파악이 CUSUM 값을 통해 가능하다. 각 공간범위별로 어느 특정 시점에 이르러서는 CUSUM 값이 계속 감소하거나, 수렴(0으로)하는 것을 볼 수 있다. 이는 1930~40년대에 이르러 상업적 토지이용의 공간 클러스터 정도가 극에 달했을 때 최대화 되었던 CUSUM 값이, 그 이후에 $z-k$ (식9) 값이 지속적으로 음(-)의 값이 산출되어 왔음을 볼 수 있다. 이러한 $z-k$ 의 음의 값은 공간적으로 분산화된 상업적 토지이용패턴을 나타낸다. 이로서 1940년대까지 이르러 상업적 토지이용의 공간 클러스터가 극에 달했다가, 이후의 상업적 토지이용은 분산화된 패턴으로 변화하고 있음을 알 수 있다.

더불어 이러한 상업적 토지 이용의 공간 클러스터가 최대에 이르는 시기(1930-40년대경)와 공간 스케일(32,000 feet)에서의 국지적 패턴을 탐색적으로 살펴보기 위해, $Z_i(d)$ 통계량을 산출하여 지도화 하였다(그림 11). $Z_i(d)$ 통계량이 크게 나타나는 지역은 공간 검색 스케일 d 내에 있는 주변 지역에 높은 값(여기서는 센서스 트랙(census tract) 내의 상업적 토지로 이용된 필지 수)을 가진다는 것을 의미한다. 그림 11에서처럼, 1940년까지의 대부분의 상업적 토지이용은 CBD를 중심으로 이루어져 왔음을 알 수 있다. 그 외에

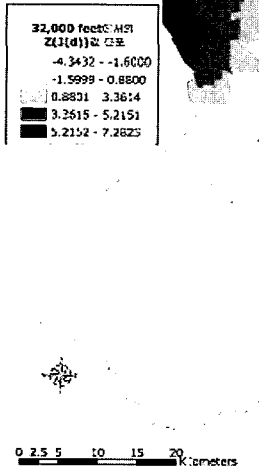


그림 11. 1940년까지의 상업적 토지이용의 국지적 공간패턴($Z(J(d))$, $d=32,000$ feet)

CBD에서 떨어진 도시 교외 지역 일부에서도 상업적 토지이용이 이 시기에 이르러 상당부분 진행되어 왔음을 유추해 볼 수 있다.

4. 결론 및 토의

많은 지리적 현상의 변화는 시공간 측면에서 동적인 프로세스를 통해 형성되며, 이를 파악하기 위한 적합한 탐색 기법이 요구된다. 그러나 정적인 지리현상의 패턴의 이해보다는 시간변화에 따른 공간패턴의 역동적인 변화를 탐색하는 방법의 중요성에 비해, 이에 대한 연구가 적합한 데이터의 부재 등 여러 가지 이유로 아직까지 토지이용과 관련된 연구분야에 본격적으로 활성화되고 있지 않다. 즉 지리적 현상의 변화에 대한 시공간 측면의 중요성에 불구하고 이러한 통계적 탐색 방법에 대한 실증 연구가 많이 수행되지 못하였다. 이러한 배경에서 본 연구는 공간 패턴을 측정하는 공간 통계량의 시간 흐름에 따른 관성을 고려하여, 공간 패턴의 변화를 시간 추이에 따른 관성을 배제하여 누적

적으로 탐색할 수 있는 방법을 고찰하였다. 그리고 이러한 방법론을 사례연구에 적용하여, 미국 뉴욕주, 이리 카운티 (Erie County) 지역의 과거 약 200년 동안의 상업적 토지이용의 공간 패턴의 변화 분석을 수행하였다. 이를 위해, 본 논문에서 제안한 $J(d)$ 통계량과 시간의 흐름에 따른 변량의 통계적인 근본 변화를 감지하는 *CUSUM* 방법을 결합하여, 대상기간 동안 다양한 공간스케일에서 상업적 토지이용의 공간패턴 변화를 탐색하였다.

사례 지역인 이리 카운티는 1940년대에 이르러 상업적 토지이용의 클러스터 패턴이 모든 공간 스케일에서 최고치를 가지는 경향을 보이며, 그 이후 점차 분산화 패턴 경향을 보이고 있다. 또한 공간 스케일이 32,000 feet(약 9.8km)일 때 클러스터화가 최대로 이루어졌음을 알 수 있었다. 이와 더불어 이 시기에 CBD를 중심으로 상업적 토지이용의 국지적 클러스터가 집중적으로 형성되었으며, 그 결과 단핵 도시구조 패턴을 구성해 왔음을 알 수 있었다. 그리고 결과의 해석과 관련하여, 공간패턴의 통계량을 누적하면서 과거 200년 동안의 패턴을 계속 누적한 결과 상당히 높은 *CUSUM* 값이 산출되었으며, 이는 통계적인 유의성 검정절차와 관련하여 해석을 어렵게 하는 원인으로 생각된다.

본 연구는 $J(d)$ 공간통계량과 시계열 변화를 누적적으로 반영하는 *CUSUM* 통계량을 결합하여 토지이용과정의 시공간적 패턴 탐색과정을 고찰하였다. 이 과정을 통해 기존에 많이 수행되었던 정적인 공간탐색방법에서는 밝혀내기 어려운 상업적 토지이용의 단위시간별 공간패턴의 영향을 누적적으로 파악할 수 있었다. 또한 기존의 정적인 공간통계 탐색방법에서는 오랜 시기 동안에 유의미한 시점에서의 공간 패턴의 파악과 해당 시점에서의 구체적인 공간패턴의 결과를 파악하는데 한계가 있었으나, 본 연구는 오랜 시기 중에서 유의미한 공간패턴을 가지는 시점의 파악에 도움을 주고 있다.

요약하면, 본 연구는 시간자기상관의 효과를 구분하고 시간흐름에 따른 독립적인 공간패턴의 변화를 누적적으로 반영하고 이를 상업적 토지이용의 시공간 패턴 변화의 탐색과정에 활용한다는 점에서 의의가 있다. 그러나 시공간 탐색 차원을 넘어 유의미한 이상적인

공간패턴의 변화를 특정한 임계값을 기준으로 자동화하여 탐색하지 못한 것은 본 연구의 한계라 할 수 있다. 또한 사례지역에 대한 경험적인 정보를 토대로 공간검색 범위 임계값이 선정되어 적용되었다. 그러나 많은 다양한 사례지역에 적용할 수 있는 일반화된 임계값 선정 연구가 요구된다. 즉 토지이용의 시간의 흐름에 따른 지리적 변화를 탐색하는 과정에서 유의미한 이상 변화부분(구간)을 찾아낼 필요성이 있으며, 이를 위한 후속 연구가 필요하다. 그러나 어떤 임계값을 설정할 것인지에 대한 이론적, 통계적 접근이 아직 정립되지 않은 상태이며, 아직까지는 연구자가 연구대상지의 파악을 통한 경험적 정보에 의존하는 경우가 많다. 따라서 이에 대한 보편적인 임계값 설정에 대한 연구가 앞으로 필요할 것으로 생각된다. 또한 공간적 관련성을 정의하기 위한 적절한 시간 범위에 관한 연구가 필요할 것으로 보인다.

註

- 1) 사례 연구로서, 유방암 같은 질병의 공간패턴의 변화를 신속하게 탐지하기 위해, 다양한 기존의 공간통계량(Moran's I 나 Score Test 등)의 시계열적 변화과정을 $CUSUM$ 통계량을 통해서 모니터링하는 연구들이 수행되었다(Rogerson, 2001a; Lee and Rogerson, 2007).
- 2) 공간통계기법을 이용한 인문 현상의 공간 패턴에 대한 연구 동향은 Shin(2005)에 언급되어 있다.
- 3) Pacheco and Tyrrell(2002)은 Moran's I 와 Local Moran's I 를 함께 사용함으로써 전역적, 국지적 탐색 모두를 수행하였다.
- 4) Ceccato and Persson(2002)는 global G 와 local G 를 함께 사용하였다.
- 5) 전역적, 국지적 J 통계량의 해석에 유의할 점이 있다. 예를 들어 높은 지가의 공간단위들이 국지화적으로 클러스터된 경우는 전역적인 J 통계량이 증가하는 데 기여하지만, 낮은 지가의 공간단위들이 클러스터 되는 경우는 반대의 영향을 미친다. 또한 양(+)의 값을 가진 공간단위가 클러스터 되더라도, 지역 전반에 걸쳐 지가 분포가 낮게 유지된다면 전역적 J 통계량은 낮게 나타날 수 있다. 반대의 경우도 마찬가지로 예상된다. 즉 전역적 J 통계량 자체 만으로는 국지적 공간 패턴의 탐색에는 한계가 있으며, 둘다 함께 활용될 필요가 있다.
- 6) 본 논문에서 이 통계량은 폴리곤 유형의 공간패턴분석을 위

- 해 제안 되었으나, 또한 포인트 유형 데이터에 대한 공간 패턴분석에도 잠재적인 활용 가능성이 있을 것으로 생각된다.
- 7) 이에 관련한 사항은 Getis and Ord(1992)가 상세히 다루고 있다.
 - 8) 이는 시간 자기상관(temporal autocorrelation)과 관련된다. 즉 시점 t 과 $t-1$ 에서의 공간패턴은 완전히 독립적이지 않다. 시점 t 에서의 공간패턴은 시점 $t-1$ 에서의 수행 결과를 어느 정도 반영한다. 이는 공간 과정의 관성에 상당부분 기인하며, 이로 인한 정보의 중복을 구분하지 않으면 공간패턴의 통계적 유의성이 왜곡될 가능성이 있다. 이와 관련하여 Rogerson(2001a)은 이러한 왜곡 가능성에 대해 논의하였다.
 - 9) ARL_0 의 값은 오경보율(실제 오류가 발생하지 않았는데 발생한 것으로 탐지하는 경우와 관련됨)과 밀접한 관련이 있으며, ARL_0 값이 클수록 이러한 오경보율은 감소될 가능성이 커지는 반면, 실제 통계적 변화를 탐지할 가능성이 줄어든다. 즉 ARL_1 이 커질 가능성이 증가하게 된다.
 - 10) 사례연구로, Rogerson and Sun(2001)은 미국 뉴욕 주 버팔로 지역의 범죄 발생의 공간 패턴을 $CUSUM$ 방법론을 이용해서 분석하였는데, 범죄의 클러스터 정도를 측정하는 공간통계량으로 최근린 통계량(nearest neighbor statistic)을 $CUSUM$ 통계량과 결합함으로써, 범죄발생의 공간패턴을 모니터링하였다.
 - 11) ORPS 데이터베이스는 모든 시점에서의 상업적 토지이용의 발생과 소멸을 모두 하지 않으며, 1998년 현재 시점에서 상업적 토지이용의 설정년도를 토대로 각 시점의 공간패턴을 역추산하였다. 따라서 이러한 데이터베이스의 한계로 보다 정확한 사례지역의 동적인 변화를 실증적으로 탐색하는데 일부 제약이 예상된다.
 - 12) 공간검색 범위(d)는 인접한 관찰값을 통합적으로 재편성하여 공간 자기상관적 특성을 명시적으로 나타내기 위한 지리적 범위(extent)라고 볼 수 있다. 즉, 개별 공간단위들 속에서 관찰되는 상업적 토지이용의 정도는 각각 다를 수 있지만, 주변 공간단위들의 영향속에서 관찰값들이 재조정되는 경우, 공간 자기상관의 특성을 명시화 할 수 있으며, 이러한 상호 영향력의 범위를 공간검색 범위(d)라고 해석할 수 있다. 이는, 커널(kernel) 밀도추정함수에서 개별 관찰값의 독특한 특성 때문에 전체적인 특성들을 파악하지 어려운 경우 국지적 노이즈(noise)를 완화(smoothing)하기 위해 연구자가 명시하는 대역폭(bandwidth)에 비유될 수 있다. 한 공간단위에서 관찰되는 상업적 토지이용의 정도가 유의미한 특성을 갖지 못하는 경우라도, 주변 값들과의 연계속에서 묶어서 관찰하는 경우 유의미한 값을 보일 수 있다.
 - 13) 대상지역의 공간단위(census tract)의 수는 243개이며, 모든 공간단위들의 면적을 합하여 공간단위의 개수로 나눈

뒤에 제곱근을 계산하면 대상지역의 공간단위들의 평균 길이가 산출된다. 이 평균길이의 값은 3,336 feet이며, 이 값보다 작은 공간단위를 고려하여 2,600 feet를 공간검색 범위의 시작점으로 설정하였다. 실제로, Rogerson(2001b)은 국지적 공간 통계량(*M-statistic*)의 구성을 위한 가중행렬을 산출하는 데 있어서 이와 유사한 방법을 사용하였다. 공간 스케일의 상위 임계점은, 대상지역의 면적의 제곱근(52,005 feet)을 기준으로 이와 유사한 범위(56,000 feet)로 설정하였다.

文獻

- 권영아, 2006, “최근 한국의 서리 현상의 공간 분포와 시계열 변화 경향,” *대한지리학회지*, 41(3), 361-372.
- 박기호 · 안재성 · 이양원, 2005, “시공간 개인통행자료의 지리적 시각화,” *대한지리학회지*, 40(3), 310-320.
- 신정엽, 2004, “VCEC(Variable Clumping method for Economic Clusters)을 이용한 도시내 경제 클러스터 탐색 방법에 대한 연구,” *지리교육논집*, 48, 63-72.
- 이민부 · 김남신 · 최한성 · 신근하, 2003, “GIS와 RS를 이용한 토지피복 및 식생 분포의 시공간적 변화: 평안북도 서부 지역을 중심으로,” *대한지리학회지*, 38(5), 835-848.
- Anselin, L., 1995, Local indicators of spatial association - LISA, *Geographical Analysis*, 27, 93-115.
- Barff, R., 1987, Industrial clustering and the organization of production: a point pattern analysis of manufacturing in Cincinnati, Ohio, *Annals of the Association of American Geographers*, 77(1), 89-103.
- Besag, J. and Diggle, P. J., 1977, Simple Monte Carlo tests for spatial pattern. *Applied Statistics*, 26(3), 327-333.
- 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.
- Choi, Y., 2005, Temporal and spatial variability of heating and cooling degree-days in South Korea, 1973-2002, *Journal of the Korean Geographical Society*, 40(5), 584-593.
- Cliff, A. and Ord, J.K., 1981, *Spatial Processes: Models and Applications*, Pion, London.
- Cuthbert, A. and Anderson, W., 2002, Using spatial statistics to examine the pattern of urban land development in Halifax-Dartmouth, *Professional Geographer*, 54(4), 521-532.
- Findlay, A. and A. Findlay., 1984, A Monte Carlo approach to estimating the significance of segregation. *Environment and Planning A*, 16, 225-231.
- Fotheringham, A. S., Brunsdon, C., and Charlton, M., 2000, *Quantitative Geography: Perspectives on Spatial Data Analysis*, Sage Publications, London.
- Getis, A., and Ord, J. K., 1992, The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24, 189-206.
- Hawkins, D. M. and Olwell, D. H., 1998, *Cumulative Sum Charts and Charting for Quality Improvement*, Springer-Verlag, London.
- Kim, H-M., 2005, A GIS-based Analysis of spatial patterns of individual accessibility: A critical examination of spatial accessibility measures, *Journal of the Korean Geographical Society*, 40(5), 514-532.
- Knox, G., 1964, The detection of space-time interactions. *Applied Statistics*, 13(1), 25-29.
- Lee, G. and Rogerson, P., 2007, Monitoring global spatial statistics, *Stochastic Environmental Research and Risk Assessment (SERRA)*, 21(5), 545-553.
- Manly, B. F. J., 1998, *Randomization, Bootstrap and Monte Carlo Methods in Biology*, CHAPMAN & HALL.
- Moran, P., 1948, The interpretation of statistical maps, *Journal of the Royal Statistical Society B*, 10, 243-251.
- Openshaw, S., 1977, A Geographical solution to scale and aggregation problems in region-building, partitioning, and spatial modelling, *Transactions of the Institute of British Geographers*, 2, 459-475.

- Openshaw, S., 1984, The Modifiable Areal Unit Problem: *Concepts and Techniques in Modern Geography*, 38, Geo Books, Norwich.
- Openshaw, S. and Taylor, P.J., 1979, A million or so correlation coefficients: three experiments on the modifiable areal unit problem, in Wrigley, N.(ed.), *Statistical Applications in the Spatial Sciences*, Pion, London, 127-144.
- Pacheco, A. and Tyrrell, T., 2002, Testing spatial patterns and growth spillover effects in clusters of cities, *Journal of Geographic Systems*, 4, 275-285.
- Paci, R. and Usai, S., 1999, Exetrnalities, knowledge spillovers and the spatial distribution of innovation, *GeoJournal*, 49, 381-390.
- Rogerson, P., 1997, Surveillance systems for monitoring the development of spatial patterns, *Statistics in Medicine*, 16, 2081-2093.
- Rogerson, P., 2001a, Monitoring point patterns for the development of space-time clusters, *Journal of Royal Statistical Society*, A, 164, 87-96.
- Rogerson, P., 2001b, A statistical method for the detection of geographic clustering, *Geographical Analysis*, 33, 215-227.
- Rogerson, P., 2006a, Statistical methods for the detection of spatial clustering in case-control data, *Statistics in Medicine*, 25, 811-823.
- Rogerson, P., 2006b, Formulas for the design of CUSUM quality control charts, *Communications in Statistics: Theory and Methods*, 35, 373-383.
- Rogerson, P. and Sun, Y., 2001, Spatial monitoring of geographic patterns: an application to crime analysis. *Computers, Environment and Urban Systems*, 25, 539-556.
- Rogerson, P. and Yamada, I., 2004, Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches, *Statistics in Medicine*, 23, 2195-2214.
- Shin, J., 2005, The statistically and economically significant clustering method for economic clusters in an urban region, *Journal of the Korean Geographical Society*, 40(1), 63-77.
- Siegmund, D. O., 1985, *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.
- Sweeney, S. and Feser, E., 1998, Plant size and clustering of manufacturing activity, *Geographic Analysis*, 30(1), 45-64.
- 교신: 신정엽, 151-742, 서울시 관악구 신림동 산56-1, 서울대학교 사범대학 지리교육과(이메일: geoshin@snu.ac.kr, 전화: 02-880-4060, 팩스: 02-882-9873)
- Correspondence: Jungyeop Shin, Department of Geography Education, Seoul National University, San56-1, Shillim-dong, Kwanak-gu, Seoul, 151-742, Korea (e-mail: geoshin@snu.ac.kr, phone: 02-880-4060, fax: 02-882-9873)
- 최초투고일 07. 09. 08.
최종접수일 07. 09. 17.